Mathematical Optimization Techniques for Resource Allocation in Cognitive Radio Networks

by

Yogachandran Rahulamathavan

A Doctoral Thesis submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy (PhD), at Loughborough University.

October 2011

Advanced Signal Processing Group,
School of Electronic and Electrical Engineering,
Loughborough University, Loughborough
Leicestershire, UK, LE11 3TU.

CERTIFICATE OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgements or in footnotes, and that neither the thesis nor the original work contained therein has been submitted to this or any other institution for a degree.

.... (Signed)

.... (candidate)

I dedicate this thesis to my late parents and brother.

Abstract

Introduction of data intensive multimedia and interactive services together with exponential growth of wireless applications have created a spectrum crisis. Many spectrum occupancy measurements, however, have shown that most of the allocated spectrum are used inefficiently indicating that radically new approaches are required for better utilization of spectrum.

This motivates the concept of opportunistic spectrum sharing or the so-called cognitive radio technology that has great potential to improve spectrum utilization. This technology allows the secondary users to access the spectrum which is allocated to the licensed users in order to transmit their own signal without harmfully affecting the licensed users' communications.

In this thesis, an optimal radio resource allocation algorithm is proposed for an OFDM based underlay cognitive radio networks. The proposed algorithm optimally allocates transmission power and OFDM subchannels to the users at the basestation in order to satisfy the quality of services and interference leakage constraints based on integer linear programming. To reduce the computational complexity, a novel recursive suboptimal algorithm is proposed based on a linear optimization framework. To exploit the spatial diversity, the proposed algorithms are extended to a MIMO-OFDM based cognitive radio network. Finally, a novel spatial multiplexing technique is developed to allocate resources in a cognitive radio network which consists of both the real time and the non-real users. Conditions required for convergence of the proposed algorithm are analytically derived. The performance of all these new algorithms are verified using MATLAB simulation results.

Contents

Statement of Originality

The contributions of this thesis are mainly on the development of various resource allocation algorithms for cognitive radio networks. The novelty of the contributions is supported by the following international journals and conference papers.

In Chapter 3, we propose various optimal resource allocation algorithms for OFDMA based underlay cognitive radio networks. The results have been published in:

1. Y. Rahulamathavan, K. Cumanan, L. Musavian and S. Lambotharan, 'Optimal subcarrier and bit allocation techniques for cognitive radio networks using integer linear programming', in Proc. IEEE SSP, Cardiff, UK, Aug. 2009.

2. Y. Rahulamathavan, K. Cumanan and S. Lambotharan, 'Optimal resource allocation techniques for MIMO-OFDMA based cognitive radio networks using integer linear programming' in Proc. IEEE SPAWC, Morocco, Jun. 2010.

3. Y. Rahulamathavan, K. Cumanan, R. Krishna and S. Lambotharan, 'Adaptive subcarrier and bit allocation techniques for MIMO-OFDMA based uplink cognitive radio networks', in Proc. UKIWCWS, Delhi, India, Dec. 2009.

The contribution of Chapter 4 is on the development of low computational complexity resource allocation algorithms for OFDMA based underlay cognitive radio networks. These works have been presented in:

4. Y. Rahulamathavan, S. Lambotharan, Cenk Toker and A. B. Gershman, 'A Suboptimal Recursive Optimization Framework for Adaptive Resource Allocation in Spectrum Sharing Networks', accepted for publication in IET Proceedings on Signal Processing, Oct. 2011.

5. Y. Rahulamathavan, Yupu Dong, K. Cumanan, S. Lambotharan and A. K. Nandi, 'Adaptive rate balancing techniques in cognitive radio networks', in Proc. IWCR, Bangalore, India, Dec. 2010.

In Chapter 5, a novel mixed QoS resource allocation algorithm for a cognitive radio network using beamforming technique has been presented. The novelty of this work is attested by the following publications:

6. Y. Rahulamathavan, K. Cumanan, and S. Lambotharan, 'A Mixed SINR-Balancing and SINR-Target Constraints based Beamformer Design Technique for Spectrum Sharing Networks', accepted for publication in IEEE Trans. Vehicular Technology, Sep. 2011.

7. Y. Rahulamathavan, K. Cumanan and S. Lambotharan, 'An SINR Balancing based Beamforming Technique for Cognitive Radio Networks With Mixed Quality of Service Requirements', in Proc. IEEE ICC, Kyoto, Japan, Jun. 2011.

8. Y. Rahulamathavan and S. Lambotharan, 'A Rate Balancing Technique for MIMO-Cognitive Radio Network Under a Mixed QoS Requirement', to appear in Proc. IEEE GlobeCom, Houston, TX, Dec. 2011.

Acknowledgements

I AM DEEPLY INDEBTED to my supervisor Professor Sangarapillai Lambotharan for his kind interest, generous support and constant advice throughout the past three years. I have benefitted tremendously from his rare insight, his ample intuition and his exceptional knowledge. This thesis would never have been written without his tireless and patient mentoring. It is my very great privilege to have been one of his research students.

I am extremely thankful to late Professor Alex B. Gershman for providing me with valuable comments on the IET Signal Processing journal. Also, I would like to thank Professor Jonathan Chambers for his support and encouragement.

I would like to express my sincere thanks to Dr. Kanapathippillai Cumanan. His support has greatly improved the quality of my work. When I was losing confidence in my research direction, his warm encouragement and continuous support lifted my sprit and motivated me to run again, which eventually led me to this final stage.

I would like to extend my appreciations to my colleagues Mathavan, Mohsen, Amod, Ranaji, Miao, Jie Tang, Ivan, Gorgia, Jo, Adel, Mustafa, Fahid and Abdulla for making my stay at Loughborough pleasant. I also wish to acknowledge friends in Leicester Belal, Zahir, Swaroop, Zain, Mike, Kazi, Gazi, Nash, Anu, Dev, Tesh, Arif and Mohammed.

I wish to take this opportunity to thank my brothers Rahavan, Kohul, Rajiv and Siva and my uncles Kandasamy, Kunaraj, Deivachandran and Carthigeyan for their moral support throughout my studies. I also want to thank Mr. Peethamparam Divakaran, who is the primary reason for me being in the United Kingdom. I would like to thank Arunakumar and his family for their support throughout my stay in Loughborough.

Last, but most importantly, I wish to express my deepest gratitude and love to my beloved wife, Sutharsini, for her endless love and support. She is the reason for my blissful life at United Kingdom. I would like to dedicate this thesis to my late parents and brother.

Rahul. Y

September, 2011

List of Acronyms

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
AMPS	Advanced Mobile Phone Service
AWGN	Additive White Gaussian Noise
BER	Bit Error rate
BPSK	Binary Phase Shift Keying
CDMA	Code Division Multiple Access
CSI	Channel State Information
EDGE	Enhanced Data rates for GSM Evolution
FDD	Frequency Division Duplexing
FDMA	Frequency Division Multiple Access
FM	Frequency Modulation
GPRS	General Packet Radio Service

GSM	Global System for Mobile
IS	Interim Standard
LTE	Long Term Evolution
M-ary FSK	M-ary Frequency Shift Keying
M-ary PSK	M-ary Frequency Phase Keying
M-ary QAM	M-ary Quadrature Amplitude Modulation
MA	Margin Adaptive
MIMO	Multiple-Input-Multiple-Output
NMT	Nordic Mobile Telephony
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
QoS	Quality-of-Service
QPSK	Quadrature Phase Shift Keying
RA	Rate Adaptive
SC-FDMA	Single-Carrier Frequency Division Multiple Access
SDMA	Space Division Multiple Access
SDP	Semi Definite Programming
SIMO	Single-Input-Multiple-Output
SINR	Signal-to-Noise-plus-Interference Ratio
SISO	Single-Input-Single-Output
SVD	Singular Value Decomposition

TDMA	Time Division Multiple Access
UMTS	Universal Mobile Telecommunications System
WCDMA	Wideband Code Division Multiple Access

List of Symbols

Scalar variables are denoted by plain lower-case letters, (i.e., x), vectors by bold-face lower-case letters, (i.e., $\mathbf{x}$), and matrices by upper-case bold-face letters, (i.e., $\mathbf{X}$). Some frequently used notations are as follows:

$\mathrm{E}\{\cdot\}$	Statistical expectation
$(\cdot)^T$	Transpose
$(\cdot)^H$	Hermitian transpose
$(\cdot)^*$	Complex conjugate
$(\cdot)^{(q)}$	q^{th} iteration
$\|.\|_1$	L_1 norm
$\|.\|_2$	Euclidean norm
$(.)^{-1}$	Matrix inverse
$\mathbf{1}_M$	$M \times 1$ vector of ones
$\mathbf{0}_M$	$M \times 1$ vector of zeros
$\mathbf{1}_{M,N}$	$M \times N$ matrix of ones
$\mathbf{0}_{M,N}$	$M \times N$ matrix of zeros
$\mathrm{Re}\{\cdot\}$	Real part
$\mathrm{Im}\{\cdot\}$	Imaginary part

$\mathrm{Tr}\{\cdot\}$	Trace operator
$\odot$	Hadamard product
$\mathbf{I}$	Identity matrix
$\rho(\mathbf{X})$	Spectral radius of square matrix $\mathbf{X}$
$\mathbf{diag}(\mathbf{x})$	Diagonal matrix with vector $\mathbf{x}$

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Evolution of Wireless Communication Systems

Due to fast development of smart phones and associated multimedia and interactive applications, wireless communication systems have been experiencing an explosive growth. The cellular wireless communication system is the most successful wireless application, nowadays, which is also an important element for globally ubiquitous wireless connections. During the 1950s and 1960s, AT&T Bell laboratories first developed the concept of cellular wireless communications, wherein spectrum within a geographical region can be reused by breaking the region into small cells [2]. Each cell is assigned a set of frequencies, and, these frequencies can be reused by a different cell, when there is a sufficient distance between both cells. Fig. 1.1 shows an example of cellular wireless communication system where adjacent cells do not use the same set of frequencies. This cellular concept coupled with advanced signal processing techniques and developments in reliable radio frequency hardware is a breakthrough for the modern wireless communication developments.

The first international mobile communication system was the analog Nordic Mobile Telephony (NMT) system which was introduced in the Nordic countries in 1981, at the same time as analog Advanced Mobile Phone Service (AMPS) was introduced in North America. These first generation (1G) networks relied on analog Frequency Modulation (FM), where each user was assigned a separate downlink and uplink FM channel. This method

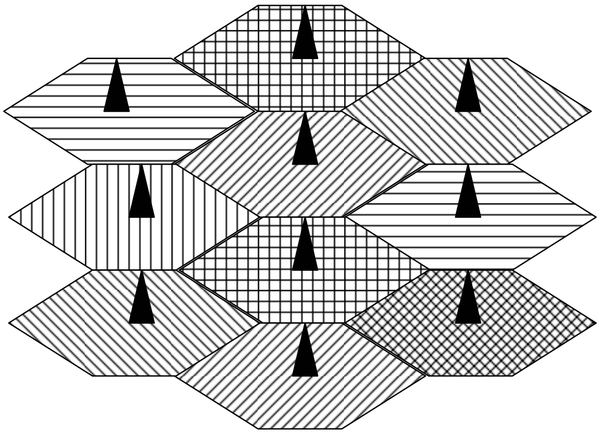

Figure 1.1. Conventional cellular wireless system. Different shading patterns of cells represents different sets of frequency ranges.

of disjoint frequency sharing is called Frequency Division Multiple Access (FDMA) with Frequency Division Duplexing (FDD). Fig. 1.2 shows a typical FDMA with FDD set up, where users are assigned different set of uplink and downlink frequency channels.

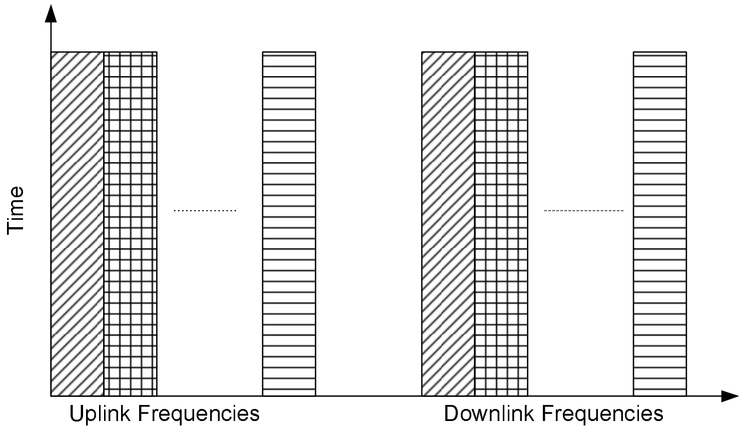

Figure 1.2. FDMA with FDD. Different uplink and downlink frequencies are allocated for different users (i.e., different shading patterns represent occupancy of different users.)

Due to the capacity limitation of 1G cellular systems, these were phased out by the second generation (2G) cellular systems in the early 1990s. There are three major 2G standards, Interim Standard (IS)-95, IS-136 in the United States, and Global System for Mobile (GSM) in Europe. The most widely used 2G standard in the world today, with more than 4.4 billion subscribers, is GSM. Unlike 1G cellular system that relied exclusively on FDMA/FDD and analog FM, 2G standards use digital modulation formats and time division multiple access (TDMA)- FDD (TDMA/FDD) technique. Fig. 1.3 illustrates TDMA with FDD system, where different time slots are allocated for different users in order to share the frequency band. The enhanced versions of 2G GSM standards with higher data-rate are known as General Packet Radio Service (GPRS) and Enhanced Data rates for GSM Evolution (EDGE) for GSM. These improved 2G cellular systems are usually referred to as 2.5G systems [3].

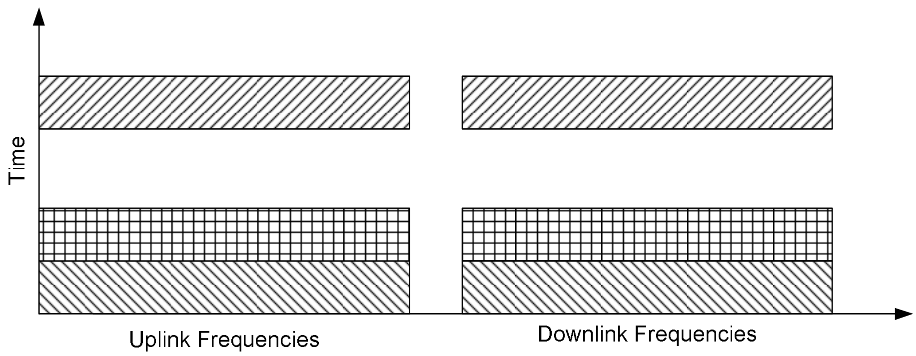

Figure 1.3. TDMA with FDD. Different time slots are allocated for different users (i.e., different shading patterns represent occupancy of different users).

Third generation (3G) systems are designed to enhance the high data rate multimedia communication such as video telephony, and access to information and services on public and private networks with high mobility. The most popular 3G standards are Wideband Code Division Multiple Access (WCDMA) or so-called Universal Mobile Telecommunications System (UMTS) and CDMA2000. These techniques are developed based on Code Division Multiple Access (CDMA). As shown in Fig. 1.4, CDMA allows all users to transmit at the same time and frequency but using different codes. Different frequency bands are used for uplink and downlink. In the standardization forums, WCDMA technology has emerged as the most widely adopted 3G air interface. Its specification has been created in the 3rd Generation Partnership Project (3GPP), which is the joint standardization project of Europe, Japan, Korea, United States and China [4].

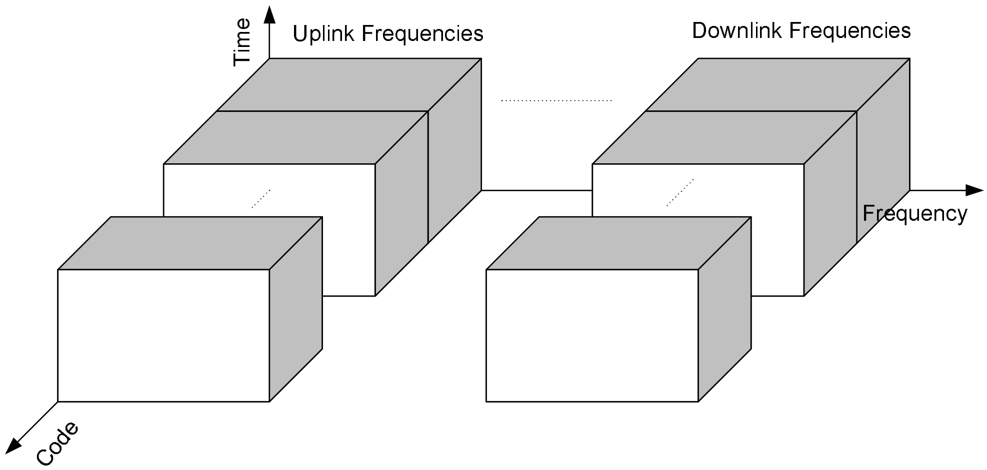

Figure 1.4. CDMA with FDD. All users utilize the whole frequency band and time slots simultaneously for transmission.

3GPP-LTE (Long Term Evolution) is envisioned as the fourth generation (4G) cellular standard, and is aligned with existing 3G deployments. LTE uses Orthogonal Frequency Division Multiple Access (OFDMA) for the downlink, and Single-Carrier FDMA (SC-FDMA) on the uplink. OFDMA and SC-FDMA are the multiple access technologies, wherein users assigned different set of subchannels that effectively divide the wideband frequency spectrum into multiple narrow band subchannels. Fig. 1.5 shows OFDMA using either TDD or FDD, where a wideband channel is divided into multiple narrow band subchannels that are orthogonal to each other. OFDMA is based on modulation method called Orthogonal Frequency Division Multiplexing (OFDM). OFDMA allows for an intelligent scheduling and resource allocation in order to utilized the frequency spectrum efficiently.

The technology development in wireless communication throughout the generations has boosted the growth of wireless applications. However, the limited availability of spectrum resources has brought the necessity of effi-

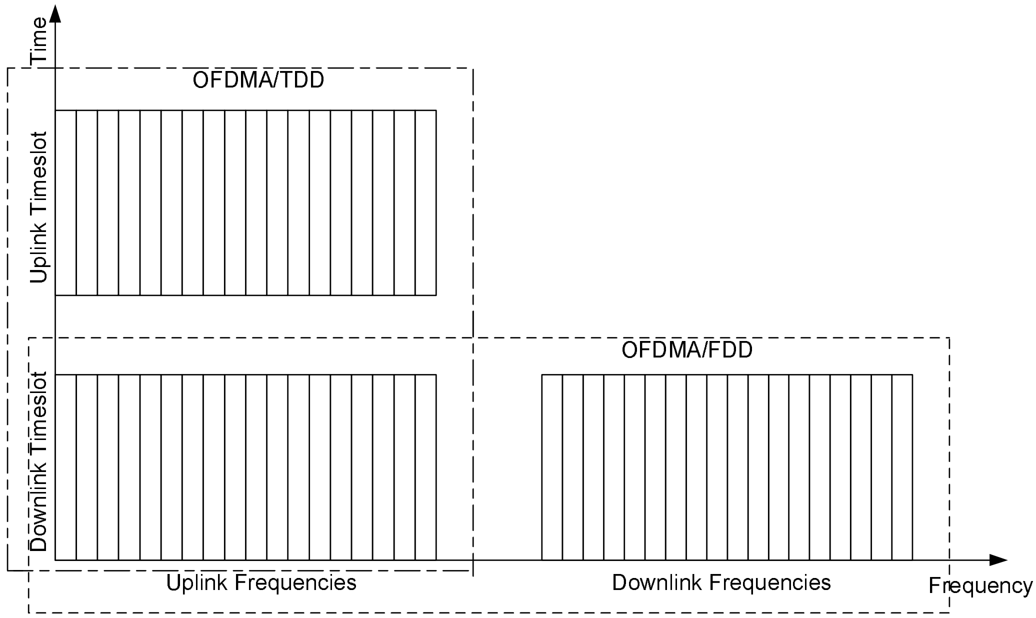

Figure 1.5. OFDMA with TDD and OFDMA with FDD.

cient spectrum utilization in wireless communications [5, 6]. Traditionally, frequency bands are divided into various sub-bands and each sub-band is licensed to operators by spectrum regulatory bodies (i.e., by OFCOM in United Kingdom). The continuous growth in wireless applications have caused spectrum crisis and saturation in the frequency allocation table. Hence, spectrum shortage has become one of the key issues in spectrum allocation. On the other hand, different spectrum measurements showed that most of the time the licensed frequency bands are under-utilized [5, 7]. This motivates the concept of cognitive radio technology that has great capabilities to improve spectrum utilization.

1.2 Cognitive Radio Networks

Cognitive radio network has the ability to increase the spectrum utilization by using the under utilized licensed spectrum for unlicensed users (also known as secondary users) [8]. In cognitive radio networks, secondary users share the radio spectrum bands, that are licensed to primary users, without harmfully affecting the primary user's communication process [8,9]. Sharing the licensed spectrum by secondary users improves the overall spectrum utilization and at the same time the transmission power of secondary user causes interference to primary user. Therefore, secondary user network should be designed in a way to allocate its radio resources to satisfy its own quality of service (QoS) requirements while ensuring that the interference caused to the primary users is below the predefined threshold level. The main functions of a cognitive radio network are spectrum sensing and exploitation of available spectrum by adjusting the transmission parameters such as frequency allocation and transmission power.

In order to use the licensed spectrum, cognitive radio networks should detect the under-utilized licensed frequency bands. The performance of the detection schemes is mainly affected by channel fading and shadowing. There are difficulties in differentiating the attenuated primary signal from a white noise spectrum. This spectrum sensing problem has been widely studied and different spectrum sensing schemes have been proposed to improve the detection performance [10, 11]. Spectrum detection techniques can be classified based on the type of detection techniques employed at the receiver: energy detection [12], coherent detection [13] and cyclostationary feature detection [14]. Energy detection is optimal when the information on the primary signal is limited. Coherent detection can be efficiently employed when the primary pilot signal is known, whereas a cyclostationary detector has the potential to distinguish the primary signal energy from the local noise energy.

There are three different types of spectrum sharing arrangements, namely, interweave, overlay, and underlay, which have been strongly supported for the development of cognitive radio networks [15]. The interweave approach is motivated by the idea of opportunistic communication. Recent spectrum occupancy measurements shows that most of the time licensed spectrum is not utilized efficiently and yields spectrum holes as shown in Fig. 1.6. In addition, these spectrum holes change with time and geographical location and can be used for secondary users communication. In this scheme, cognitive transmitters are required to sense availability of spectrum and transmit signals only when frequency holes are available. This is also known as white space filling.

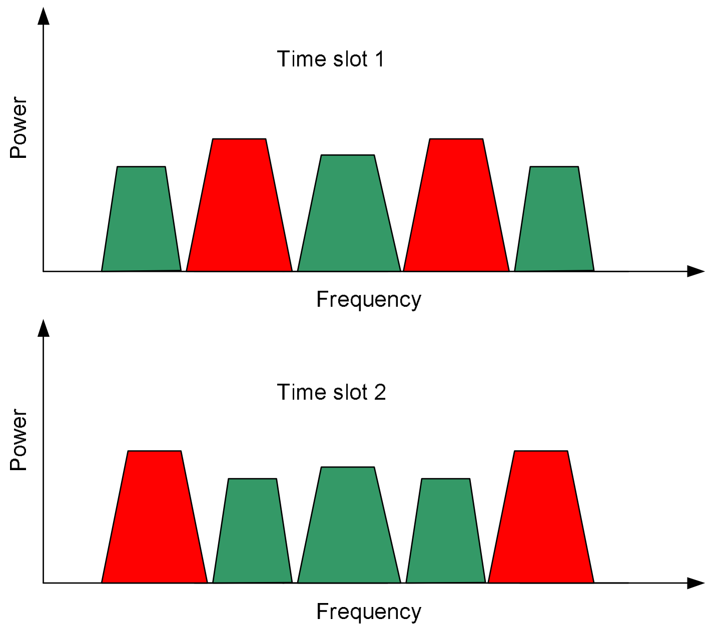

Figure 1.6. Interweave spectrum scheme. Green and red represent the spectrum occupied by the primary users and secondary users respectively.

In the overlay approach, the secondary users coexist with primary users and use part of the transmission power to relay the primary users' signals to the primary receiver. This assistance will offset the interference caused by the secondary user transmissions at the primary users' receiver. Hence, there is no loss in primary users' signal-to-noise ratio by secondary users spectrum access.

In the underlay approach, the secondary users access the licensed spectrum without causing harmful interference to primary users' communications. In this method, the secondary users ensure that interference leakage to the primary users is below an acceptable level as shown in Fig. 1.7.

In the interweave approach, identifying spectrum holes in the absence of cooperation between primary and secondary networks is very challenging. For example, a secondary transmitter could be in the shadow region of the primary transmitter which will falsely indicate (to the secondary transmitter) availability of spectrum. The secondary transmission based on this false indication may harm the primary receivers. This hidden terminal problem is deemed to be very challenging and a limiting factor for the employment of interweave cognitive radio networks. On the other hand, the overlay cognitive radio network is very interesting in terms of its theoretical advantages, however, there are even more challenges in terms of practical implementation as this requires the secondary transmitter to have prior knowledge of the primary user transmitted signal. Hence, the underlay scheme seems more realistic and easy to implement compared to the other schemes. The resource allocation techniques for underlay cognitive network is the main focus of this thesis.

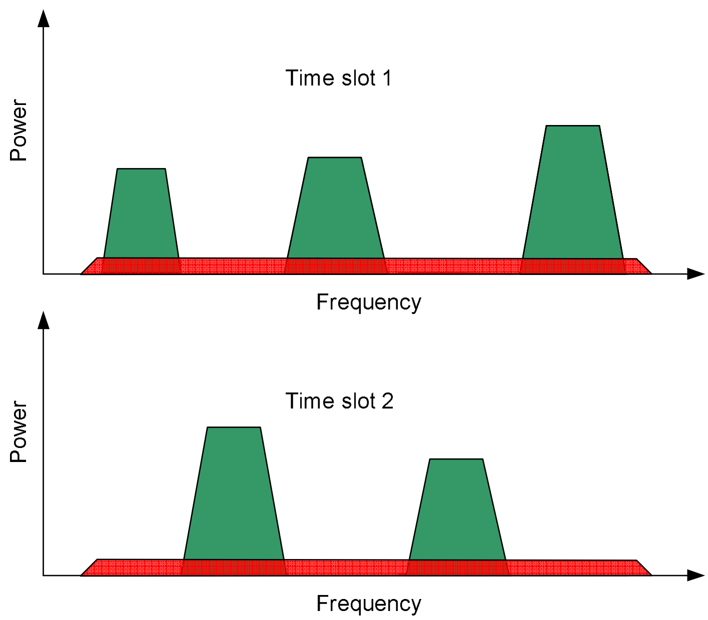

Figure 1.7. Underlay spectrum paradigm. Green and red represent the spectrum occupied by the primary users and the secondary users respectively.

1.3 Thesis Outline

In wireless communications, adaptive resource allocation techniques significantly enhance the spectrum utilization. The techniques developed for conventional wireless networks cannot be directly applied to a cognitive radio network due to the additional interference constraints on the primary users. Hence the work in this thesis mainly focuses on the resource allocation techniques for cognitive radio networks using various mathematical optimization techniques.

The Chapter 2 provides a survey on resource allocation techniques used in conventional wireless networks. Initially, multiple access methods such as FDMA, TDMA and CDMA are discussed. Following on from this, static and adaptive resource allocation techniques are described. Adaptive resource allocation techniques for an OFDM-based wireless system are introduced and different types of adaptive resource allocation techniques followed by resource allocation using multiple antennas techniques such as beamforming and spatial multiplexing are discussed briefly. Finally, recent works related to the cognitive radio network are surveyed.

Contribution Chapters

The novel results of this thesis are given in Chapters 3, 4, and 5. Chapter 3 focuses on optimal resource allocation techniques for downlink transmission of an OFDMA-based cognitive radio network. An algorithm is proposed to allocate OFDM subchannels, transmission power and data bits to various secondary users, according to their QoS requirements, while ensuring the interference leakage to the primary users is below a threshold. This algorithm is then extended to a MIMO-OFDM based cognitive radio network to exploit the spatial domain. Next, MIMO-OFDM based cognitive radio network is considered in the uplink, where various users are allocated in each subchannel based on their spatial correlations. Simulation results validate the performance of the proposed algorithms for different numbers of transmitter and receiver antennas.

The optimal resource allocation techniques based on integer linear programming were provided in Chapter 3. Computational complexity required to determine the optimal solution for OFDMA based resource allocation problem is very high. In order to overcome the computational complexity issue, low complexity algorithms are presented in Chapter 4 based on rate adaptive and rate balancing resource allocation techniques. Algorithms based on rate adaptive and rate balancing techniques are applicable to the wireless network, which supports delay sensitive real time users and delay tolerant non-real time users, respectively. Simulations results and complexity analysis are provided to validate the performance of the algorithms against optimal methods.

Chapter 5 focuses on a cognitive radio network which supports real time users and non-real time users simultaneously. Real time users are delay sensitive and require certain amount of resources all the time regardless of channel conditions. Non-real time users are delay tolerable and resources can be allocated based on channel gain and QoS requirements. A joint resource allocation algorithm is presented in Chapter 5 in order to satisfy both sets of users' QoS requirements simultaneously. Beamformer design framework for consideration of mixed QoS has not been proposed by any researchers before, and this work forms the most important contributions of the thesis. Conclusions and brief summary of this thesis are drawn in Chapter 6.

Chapter 2

RESOURCE ALLOCATION TECHNIQUES FOR WIRELESS COMMUNICATION NETWORKS

2.1 Introduction

The exponential growth of wireless communication systems opened up new challenges in the system design. One of the major challenges for the system design is limited availability of wireless resources such as frequency spectrum and transmission power. In the wireless communications, various resource allocation techniques have been proposed to utilize the scarce resources efficiently. These techniques involve strategies and algorithms for controlling transmission power, frequency allocation, modulation scheme and error coding. The main objective of the resource allocation scheme is to make the best use of the scarce radio resources to increase spectrum efficiency as much as possible [16]. In this chapter, multiple access based resource allocation techniques are introduced, followed by adaptive resource allocation methods for wireless network. Resource allocation techniques for MIMO networks and cognitive radio networks are discussed at end of this chapter.

2.2 Multiple Access Methods

Multiple access method is an essential element in the implementation of resource allocation schemes, which can be classified into several categories. TDMA is a multiple access technique that allows several users to share the same frequency band via transmitting the signals over different time slots. Specifically, different users can transmit one after the other, with each user using his own time slots (see Fig. 1.3). FDMA is another multiple access technique based on frequency division. As shown in Fig. 1.2, FDMA assigns each user one or several non-overlapping frequency bands or subchannels for signal transmission [17]. Apart from the TDMA and the FDMA, CDMA enables several users to transmit information simultaneously over the same frequency range using different codes. To properly multiplex different users, CDMA technique employs the spread-spectrum technology and pseudo-random codes [17]. Recently multiple users have been separated in spatial domain using Space Division Multiple Access (SDMA) technique. By exploiting multiple antennas, SDMA is able to offer significant performance improvement as compared with single-antenna systems [18]. Meanwhile, SDMA can create parallel spatial channels to improve system capacity via spatial multiplexing or diversity.

Traditionally, resources are allocated to various users based on static frequency or time slot allocation. Multiple access techniques FDMA and TDMA have been used to separate the multiple users in the wireless system in frequency domain and time domain respectively. These multiple access schemes allocate fixed frequency channels or time slots for various users. This separation help to avoid the inter-user interference among the users.

After this static user-frequency or user-time allocation, transmission power is optimized at the basestation, according to the users' requirements [19]. However, the resource allocation schemes based on static frequency or static time slot assignments to the users were not optimal, because these schemes were not exploited the user diversity in the frequency domain or in the time domain.

On the other hand, the adaptive resource allocation schemes allocate radio resources to the users, according to the traffic load, channel gain, and QoS requirements, so as to achieve better spectrum utilization as compared with fixed allocation schemes. OFDM technique has been adapted in wireless communication system as a best modulation candidate for adaptive resource allocation [20]. Multiple access scheme based on OFDM modulation, is called OFDMA, and it has been recognized as the potential multiple access scheme for the next generation wireless communication systems (i.e., 3GPP-LTE).

2.2.1 Overview of OFDM

OFDM is a multi carrier modulation technique that has been chosen as the modulation scheme for the current and the next generation broadband communications e.g., IEEE 802.11a/g and 3GPP-LTE [21]. It has been proposed to overcome the technical difficulties in wideband modulation schemes such as WCDMA. Technical issue related to wideband transmission is the increased corruption of the transmitted signal due to time dispersion on the radio channel. Time dispersions occurs when the transmitted signal arrived at the receiver through multiple paths with different delays. In the frequency domain, this time dispersive nature cause non-constant channel frequency response which is called frequency selectivity.

Frequency selectivity of radio channel will corrupt the frequency-domain structure of the transmitted signal and lead to higher error rates. Every radio channel is subject to frequency selectivity, at least to some extent.

However, the extent to which the transmitted signal corrupted is depend on the bandwidth of that particular radio channel, in general, larger impact for large bandwidth [21]. Receiver side equalization has been used to counteract the frequency selectivity nature of channel. Equalization has been shown to provide satisfactory performance with reasonable complexity for smaller bandwidth (i.e., up to WCDMA bandwidth of 5MHz [22]). The target channel bandwidth of 3GPP-LTE is expected to be 20MHz for the downlink transmission. However, if the transmission bandwidth is further increased then complexity of the receiver side equalization is also increased.

On the other hand, OFDM divides the wide band channel into multiple narrow band subchannels as shown in Fig. 2.1. Hence, OFDM converts the frequency selective wide band channel into frequency non-selective multiple subchannels. Another advantage of OFDM is the orthogonality between the subchannels, which allows the OFDM technology to divide the wide band into large number of narrow band subchannels, wherein each subchannel carriers low data rates, which sums up to a high data rate transmission. OFDM can be used for a single user wireless communication system. OFDMA is a multiple access scheme developed based on OFDM modulation, which is capable of serving multiple users simultaneously.

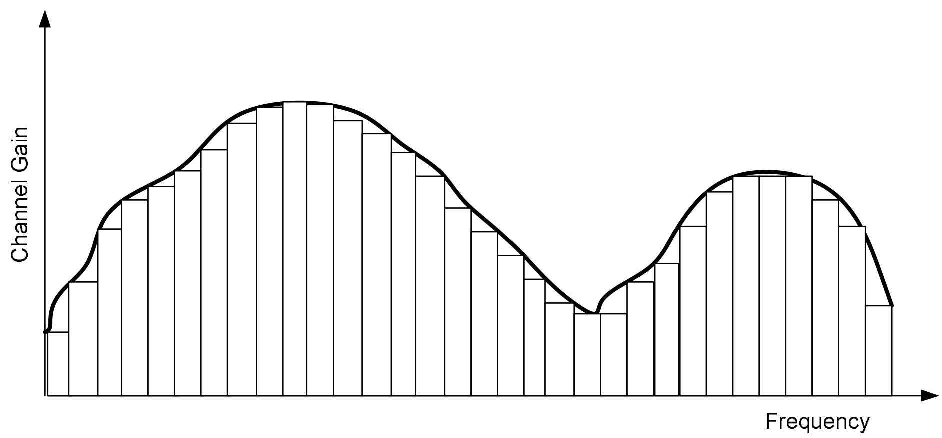

Figure 2.1. Frequency selective wideband channel is divided into multiple frequency non-selective narrow band subchannels.

2.2.2 Overview of OFDMA

Earlier OFDM has been used as a modulation scheme for wireless systems, where all the subchannels of OFDM are assigned to a single user at any given time (i.e., IEEE 802.11a/g). Later TDMA or FDMA has been used with OFDM in order to support multiple users. As mentioned earlier, this kind of static resource allocation cannot provide a good performance. Disadvantage of this static resource allocation is that multi-user diversity is not exploited (i.e., different users have different channel gains on same subchannel). OFDMA has been developed to exploit the multi-user diversity where multiple users are allowed to transmit simultaneously on the different subchannels per OFDM symbol. The probability that all users experience worst channel gain in a particular subchannel is typically quite low. Hence, adaptive resource allocation algorithms can be developed to efficiently allocate resources to multiple users by exploiting the multi user diversity.

2.3 Adaptive Resource Allocation Techniques for OFDMA Based Wireless Networks

Adaptive resource allocation techniques allocate radio resources to various users according to users channel gain and QoS requirements. The problem of assigning the subchannels and transmission power to different users in an OFDMA system has been intensively studied over the past decade (i.e., [20, 23–27] and references therein). All of these studies can be divided into two categories namely Margin Adaptive (MA) and Rate Adaptive (RA) resource allocation problems [20, 23, 24]. The objective of the MA problem is to minimize the total transmission power subject to users' individual data rate constraint, Bit Error Rate (BER) requirements while the objective of the RA is to maximize system data throughput subject to a total transmission power constraint.

2.3.1 MA Resource Allocation Problem

The MA resource allocation problem minimizes the transmission power at the basestation subject to data rate constraints to multiple uses. Since the wireless system is interference limited, the MA resource allocation problem reduces the inter-cell interference by minimizing the transmission power. Pioneering work in [20] developed MA resource allocation algorithm for OFDM-based multi user wireless systems. The proposed iterative algorithm in [20] allocates OFDM subchannels, data bits and corresponding transmission power for various users, according to their QoS requirements.

In order to understand the algorithm in [20], consider a wireless system with K users and N number of OFDM subchannels. Denote the required data rate for the k^{th} user as r_k and required received power at the k^{th} receiver in order to decode $c_{k,n}$ number of bits transmitted via n^{th} subchannel as $f_k(c_{k,n})$. Channel gain between the basestation and the k^{th} receiver in n^{th} subchannel is denoted as $h_{k,n}$. Using these definitions, the mathematical formulation of MA problem can be defined as follows:

$$\min_{c_{k,n}} \quad \sum_{k=1}^{K}\sum_{n=1}^{N} \frac{f_k(c_{k,n})}{h_{k,n}^2}, \tag{2.3.1}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} c_{k,n} \;\geq\; r_k,\; k=1,\ldots,K, \tag{2.3.2}$$

$$\forall n \text{ if } c_{k',n} \;\neq\; 0 \text{ then } c_{k,n}=0 \;\forall k \neq k'. \tag{2.3.3}$$

The objective function in (2.3.1) minimizes the transmission power at the basestation while the constraints in (2.3.2) maintain fairness among the users by satisfying their minimum data requirements. The last constraint in (2.3.3) performs mutual exclusive user allocation in each subchannel to avoid the inter-user interference. Due to the mutual exclusive user allocation, it is possible to allocate more than one subchannels for any user, but more than one users cannot be accompanied together in one subchannel. Due to the mutual exclusive user allocation, the problem in (2.3.1)-(2.3.3) is a combinatorial optimization problem [28].

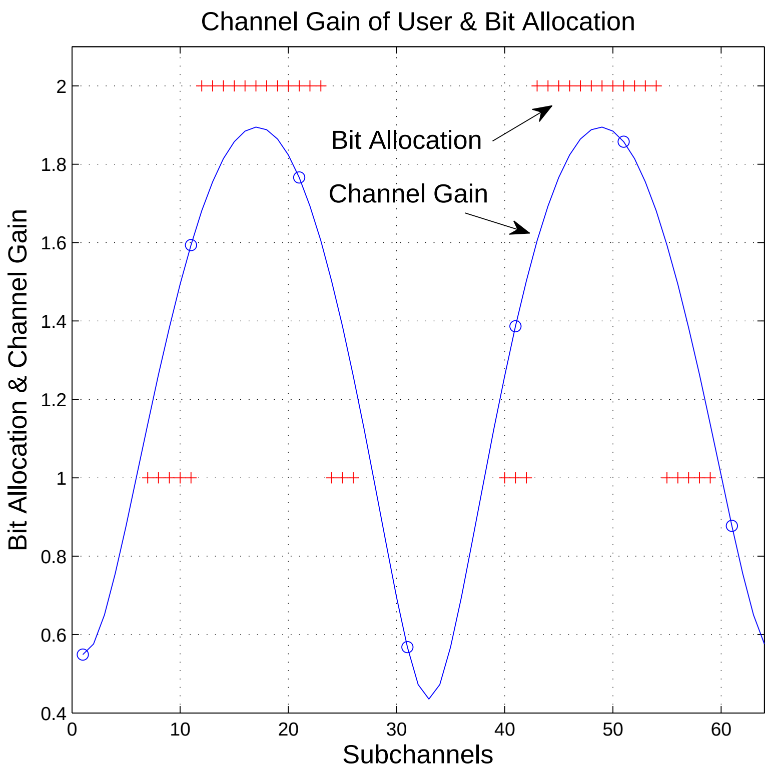

Figure 2.2. Bit allocation for a single user wireless system based on greedy algorithm.

For a single user case this problem can be solved using a greedy algorithm [25]. The greedy algorithm allocates data bits in various subchannels one-by-one, according to the least power consumption. Fig. 2.2 depicts the bit allocation and channel gain against subchannels for a single user wireless system. Also, the number of bits allocated for in each subchannel is indicated using a plus mark, "+". For example, a plus mark at 2 for subchannel 20 means two bits are allocated in the 20^{th} subchannel. Two bits are allocated in subchannels 12-22 and 44-54 due to their higher gain compared to other subchannels. At the same time no data bits are allocated in subchannels

1-5, 29-40 and 61-64 due to poor channel gains. From this figure, it is clear that the greedy algorithm allocates more number of bits to a subchannel which has higher gain compared to other subchannels in order to reduce the transmission power.

The greedy algorithm cannot be used for a multiuser wireless system to obtain an optimal solution due to the user fairness constraint in (2.3.2). It is because, the greedy algorithm always allocates subchannel to a user which has higher gain on that subchannel, compared to other users, in order to reduce the transmission power. However, in order to satisfy user data rates, some subchannels need to be assigned to a user without considering the remaining users channel gain on those subchannels. The optimal solution to this multi user problem has been obtained using an integer linear programming frame work in [26, 27].

The integer linear programming problem can be solved using branch-and-bound method which is used to find a global optimal solution of combinatorial optimization problem. The branch-and-bound method is well known for solving the class of integer linear programming problem and mixed integer programming problem [29]. A typical branch-and-bound algorithm performs two main steps namely branching and bounding. Branching step divides a feasible set of a problem into subsets and formulates the corresponding subproblems with those subsets. Bounding step finds the upper and lower bounds for those subproblems within the corresponding subset. In MA problem, all feasible combinations (i.e., user-subchannel-bit allocation combinations) are used to formulate the subproblems and then global optimal solution is obtained by removing or pruning the branches[1].

[1]Branch-and-bound method with substantially large size variables requires a computing system with high memory capacity. Otherwise, running time required to obtain an optimal solution might be very high or the problem might be terminated in the middle due to the hardware restrictions. Hence, parallel or high-performance computing systems might be suitable to achieve an optimal solution within a reasonable running time [30].

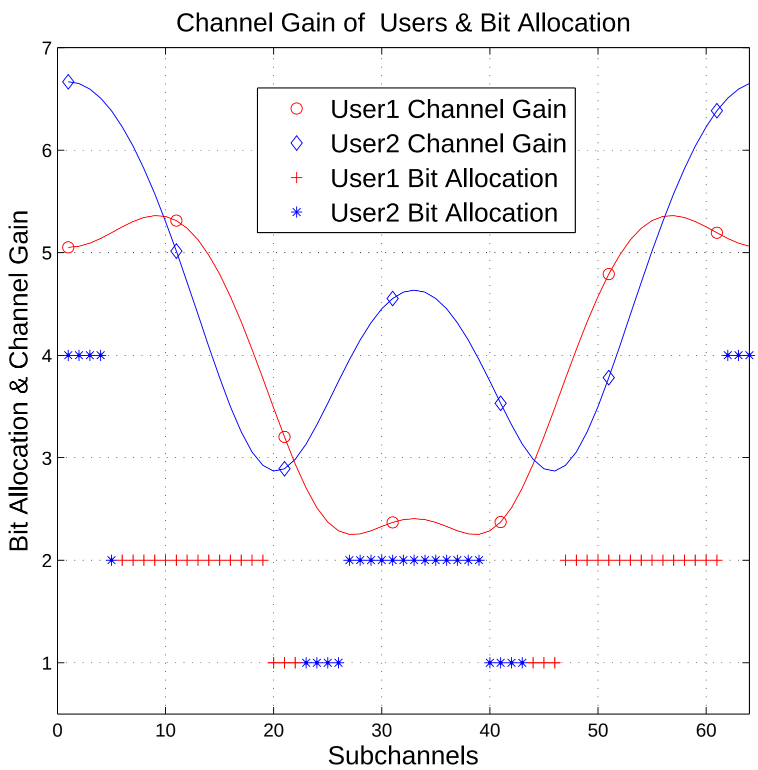

Figure 2.3. Bit allocation for a two user wireless system based on integer linear programming problem. Subchannels 7 to 9 and 55 to 61 are allocated to user 1 even when user 2 has higher gain in these subchannels.

Fig. 2.3 illustrates the optimal user, subchannel and bit allocation for a two user wireless system where each user needs to satisfy data rate of 64 bits per symbol. Number of bits allocated for in each subchannel is indicated using a plus mark, "+", for user1 and using a star mark, "*", for user2. User1 has lower channel gains compared to user2 in most of the subchannels. However, in order to satisfy the user 1s' required data rate, subchannels 7 to 9 and 55 to 61 are allocated to user1 even when user2 has higher gain in these subchannels.

Finding an optimal solution of an integer linear programming problem is computationally expensive. Hence, there are many suboptimal, low complexity algorithms proposed in the literature (i.e., [20, 26, 31, 32] and references therein). The basic concept of all of these suboptimal algorithms is to relax the mutual user allocation constraint in (2.3.3), which in turn implies a time-sharing of each subchannel among users. After this relaxation the original problem turned out to be a convex problem [33] and can be solved efficiently using interior point methods [34]. Solution to this problem can be used to allocate subchannels to users based on the time sharing factor. For example, a subchannel can be assigned to a user which has higher time-sharing factor compared to other users [31]. After the subchannel-user allocation, greedy algorithm can be used to allocate bits to each user separately [31].

2.3.2 RA Resource Allocation Problem

The RA resource allocation problem maximizes the system data throughput subject to total transmission power constraint and individual data rate constraints for multiple users. Mathematically RA problem can be defined as follows:

$$\max_{c_{k,n}} \sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}, \tag{2.3.4}$$

$$\text{s.t.} \sum_{k=1}^{K}\sum_{n=1}^{N} \frac{f_k(c_{k,n})}{h_{k,n}^2} \leq P, \tag{2.3.5}$$

$$\sum_{n=1}^{N} c_{k,n} \geq r_k, \; k = 1, \ldots, K, \tag{2.3.6}$$

$$\forall n \text{ if } c_{k',n} \neq 0 \text{ then } c_{k,n} = 0 \; \forall k \neq k', \tag{2.3.7}$$

where P is the available transmission power at the basestation. The objective function in (2.3.4) maximizes the total data throughput. Total transmission power at the basestation is limited by (2.3.5). The user fairness constraints and mutual exclusive subchannel-user allocation constraint are given in (2.3.6) and (2.3.7), respectively. Due to the mutual exclusive subchannel-user allocation constraint and user fairness constraints, the above problem is a combinatorial optimization problem. Optimal solution of this problem can be obtained using branch-and-bound method. If there is no fairness among the users, RA problem can be solved using greedy method. In [23], RA problem without user fairness was studied. It was shown in [23] that in order to maximize the system data throughput, each subchannel must be allocated to a user which has higher gain on that subchannel compared to other users. After this subchannel-user allocation, greedy algorithm can be used to allocate bits to each user based on minimum transmission power consumption. Hence, a user which has best channel gain will be assigned all the resources, which leaves many users without any chance to transfer any information for a long time.

This fairness issue was addressed in [35] by ensuring that each user achieve minimum data rate. In [35], the approach was a simple greedy algorithm that assume equal power allocation for all subchannel and assigned the best subchannel to each user until all users achieved their minimum data requirement. The remaining subchannels are assigned to the users with the best channel gain in them in order to increase the system throughput.

In [36], data rate balancing criterion is used to maintain the fairness among the users i.e., algorithm maximizes the system throughput while maintaining the equal data rates for the users. The algorithm in [36] composed with two steps: in the first step it allocates best subchannels for each user and in the second step it allocates more subchannels to a user whose data rate is lower than other users. This iteration is continued until each user achieved equal data rates. This method assumes that all the users have the same data rate requirement, which is not the case for practical wireless systems.

In [37], prioritization was enforced among the users using a weighted sum rate maximization. A user that needs higher data rates will be assigned a higher weight. This method exploited the Lagrangian dual decomposition method, which is used decompose the original problem into multiple sub-problems. Each sub-problem is convex and can be solved efficiently. The drawback of this algorithm is that there is no proper way to adjust the weights for each user.

In [38], the total data throughput was maximized under the proportional rate constraint, i.e., the rate of each user should be given according to the predefined target data rate ratio. This algorithm provides the best way of assigning priorities to multiple users, instead of simply assigning arbitrary weights as in [37].

2.4 Resource Allocation Techniques Using Multi-Antenna Techniques

2.4.1 Multi-Antenna Techniques

Multi-antenna techniques can be used in wireless systems to achieve higher spectral efficiency in terms of higher throughput, more users per cell and improved coverage [21]. This spectral efficiency can be achieved by deploying multiple antennas at the transmitter and/or at the receiver. An important characteristic of any multi antenna setup is the distance between any antennas and the mutual correlation between the radio channel experienced by different antennas. Antennas located relatively far from each other, typically implying relatively low mutual correlation and vice versa. Antennas at higher basestation need an order of ten wavelength separation to ensure the low mutual correlation and mobile terminals need that of half a wavelength. The reason for this different is the multi path fading and the arrival angle of signals through these multi paths. For a mobile terminal, signals from different paths will arrive in a wide angle, implying a low fading correlation already with relatively small antenna distance. At the same time signals arrive within a much smaller angle, implying need for relatively large antenna separation to achieve low fading correlation.

Multiple antennas at the transmitter and/or at the receiver can be exploited in different ways to achieve different aims:

- Multiple antennas can be used to provide additional spatial diversity against channel fading. Different antennas separated by larger distance experience different channel gains and signals received by those antennas can be combined in such a way for better performance.
- Instead of larger distance between antennas, different antenna polarization directions or so called polarization diversity can also be exploited to improve the performance.
- Multiple antennas can be used to direct the overall antenna beam, or so called beamforming, towards desired direction in order to suppress the interference.
- The simultaneous availability of transmitter and receiver antennas can be exploited to create parallel transmission subchannels between transmitter and receiver. This is a potential application for high bandwidth utilization. This technique referred as spatial multiplexing in literature.

In this thesis, resource allocation for wireless system based on beamforming technique and spatial multiplexing technique is considered. Hence, brief description about beamforming technique, spatial multiplexing technique and resource allocation based on both techniques are provided in the subsequent subsections.

2.4.2 Beamforming Techniques

Beamforming technique is a signal processing technique used in the physical layer of a communication to control the directionality of transmission or reception of a signal using multiple antennas at the transmission or at the reception [39].

Receiver Beamforming Techniques

In receiver side beamforming, multiple antennas are deployed at the receiver while only one antenna is used at transmitter to transmit data signals. This is referred as Single-Input-Multiple-Output (SIMO) in literature. In the receiver beamforming design, the objective is to estimate the desired signal in the presence of noise and interference. Fig. 2.4 illustrates the receiver side beamformer design. The output of the receiver can be written as

$$y(n) = \mathbf{w}^H \mathbf{r}(n), \tag{2.4.1}$$

where n is the time index, $\mathbf{r}(n) = [r_1(n) \ \cdots \ r_M(n)]^T$ is the $M \times 1$ received

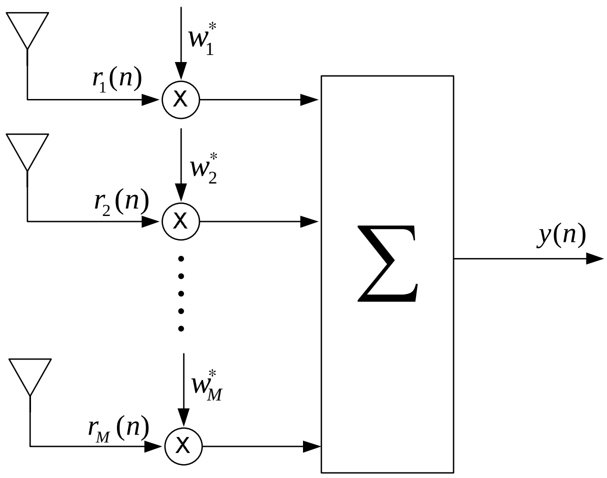

Figure 2.4. A receiver beamformer design.

signal vector and $\mathbf{w} = [w_1 \ \cdots \ w_M]^T$ is the complex beamforming weight vector. The received signal vector is given by

$$\mathbf{r}(n) = \mathbf{d}(n) + \mathbf{i}(n) + \boldsymbol{\eta}(n), \tag{2.4.2}$$

where $\mathbf{d}(n)$, $\mathbf{i}(n)$ and $\boldsymbol{\eta}(n)$ are the desired signal, interference and receiver noise respectively. Hence, the received signal-to-noise-plus-interference ratio (SINR) at the receiver can be given as follows [40]:

$$\text{SINR} = \frac{\mathbf{w}^H \mathbf{R}_d \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}}, \tag{2.4.3}$$

where $\mathbf{R}_d = \text{E}\left\{\mathbf{d}(n)\mathbf{d}(n)^H\right\}$ and $\mathbf{R}_{i+n} = \text{E}\left\{[\mathbf{i}(n) + \boldsymbol{\eta}(n)]\,[\mathbf{i}(n) + \boldsymbol{\eta}(n)]^H\right\}$ are the signal and interference-plus-noise covariance matrices. The optimum beamforming weight vector that maximizes the SINR of received signal can be given as

$$\mathbf{R}_{i+n}^{-1}\mathbf{R}_d\mathbf{w} = \lambda_{\max}\mathbf{w}, \tag{2.4.4}$$

where $\lambda_{\max}$ is the maximum eigenvalue of the matrix $\mathbf{R}_{i+n}^{-1}\mathbf{R}_d$.

Proof: See the Appendix A. ■

Hence, optimal beamformer weight vector is equivalent to generalized eigenvector of the matrices $[\mathbf{R}_d,\ \mathbf{R}_{i+n}]$ [41].

Transmitter Beamforming Techniques

For the transmit beamforming multiple antennas are deployed at the transmitter while the receiver has a single antenna to receive the transmitted signal. Beamforming at the transmitter is substantially different in several aspects from using a beamformer at the receiver. In the latter, the design will only determine the performance of a specific user whereas the transmit beamformer will affect not only the desired user but also all the users in the coverage area. Hence, the transmit beamforming design should ideally take into consideration the system level performance, i.e., all the users in the reception area rather than a specific user. Another fundamental difference is the channel knowledge. For receiver beamformer design, the receiver could estimate the channel coefficients using the training signal. For transmitter beamformer design, the channel knowledge could be made available to the transmitter by sending the estimates of the Channel State Information (CSI) from the receiver through a finite rate feedback channel [42–44] .

The focus of this section is on multiuser transmit beamformers. The transmit beamformers can be designed to satisfy QoS requirements for each user i.e., received SINR for each user. Consider a wireless network basestation equipped with N_t transmit antennas serving K users. Each user is equipped with a single antenna. The signal transmitted by the basestation is given by

$$\mathbf{x}(n) = \mathbf{W}\mathbf{s}(n), \tag{2.4.5}$$

where $s(n) = [s_1(n) \cdots s_K(n)]^T$, $s_k(n) \quad (k = 1, 2, \cdots, K)$ is the symbol intended for the k^{th} user, $\mathbf{W} = [\mathbf{w}_1 \ \cdots \ \mathbf{w}_K]$ and $\mathbf{w}_k \in \mathbb{C}^{N_t \times 1}$ is the transmit beamforming weight vector for the k^{th} user. The received signal at the k^{th} receiver can be written as

$$y_k(n) = \mathbf{h}_k^H \mathbf{x}(n) + \eta_k(n), \tag{2.4.6}$$

where $\mathbf{h}_k$ is the channel coefficient vector between the basestation and the k^{th} user and $\eta_k(n)$ is receiver noise. By defining $\mathbf{R}_k \triangleq \mathbf{h}_k \mathbf{h}_k^H$, the SINR of the k^{th} user can be written as

$$\text{SINR}_k = \frac{\mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k}{\sum_{i \neq k} \mathbf{w}_i^H \mathbf{R}_k \mathbf{w}_i + \sigma_k^2}, \tag{2.4.7}$$

where σ_k^2 is the noise variance at the i^{th} receiver.

The transmit beamforming problem based on SINR requirements can be formulated as minimization of transmitted power at the basestation subject to each user SINR being greater than a target value [45, 46].

$$\begin{aligned} \underset{\mathbf{w}_i}{\text{minimize}} \quad & \sum_{i=1}^{K} \|\mathbf{w}_i\|_2^2 \\ \text{subject to} \quad & \text{SINR}_i \geq \gamma_i \quad i = 1, \cdots, K. \end{aligned} \tag{2.4.8}$$

The problem in (2.4.8) is a quadratically constrained non-convex problem. Nevertheless, this problem can be converted into a Semi Definite Programming (SDP) with Lagrangian relaxation and can be efficiently solved using convex optimization toolboxes [47–49]. However, it is quite difficult to predict in advance whether the problem in (2.4.8) with a given set of target SINRs and total transmit power at the basestation is feasible.

To overcome this infeasibility issue, this problem can be formulated into a more attractive framework based on a max-min fairness approach or where the worst-case user SINR is maximized while using the available total transmission power [41]. This is known as the SINR balancing technique and can be stated as [41, 50–53]

$$\begin{aligned} \underset{\mathbf{U}, \mathbf{p}}{\text{maximize}} \quad & \min_{1 \leq i \leq K} \cdot \quad \frac{\text{SINR}_i(\mathbf{U}, \mathbf{p})}{\gamma_i}, \; i = 1, \ldots, K \\ \text{subject to} \quad & \mathbf{1}^T \mathbf{p} \leq \text{P}_{\max}, \end{aligned} \tag{2.4.9}$$

where $\mathbf{U} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_K]$, $\|\mathbf{u}_k\|_2 = 1$, and $\mathbf{p} = [p_1 \ \cdots \ p_K]^T$. Here $\mathbf{u}_k \in \mathbb{C}^{N_t \times 1}$ and p_k are the transmit beamforming weight vector and the corresponding allocated power for the k^{th} user respectively. In [41], an iterative algorithm has been proposed using uplink-downlink duality, where the solution balances the ratio between the achieved SINR and the target SINR for all users while using all the transmission power available at the basestation.

2.4.3 Spatial Multiplexing Techniques

So far, either receiver beamforming or transmit beamforming has been discussed. The technique using multiple antennas at both the receiver and the transmitter is known as spatial multiplexing or Multiple-Input-Multiple-Output (MIMO) array processing. The MIMO array processing can further improves the system performance by adding additional diversity against fading, compare to the use of only multiple receive antennas or multiple transmit antennas [54–56]. Spatial multiplexing can be obtained by decomposing the MIMO channel matrix into various independent spatial subchannels that are used to transmit different data streams independently. This has the potential to increase the data rate up to a factor that is the same as the rank of the MIMO matrix as compared to the single antenna system [18, 57]. Consider a point-to-point MIMO channel with M_t transmit antennas and M_r receive antennas as shown in Fig. 2.5. The received signal is given by

$$\mathbf{y}(n) = \mathbf{H}\mathbf{x}(n) + \boldsymbol{\eta}(n), \tag{2.4.10}$$

where $\mathbf{y} = [y_1(n) \ \cdots \ y_{M_r}(n)]^T$ and $y_r(n)$ is the received signal at the r^{th} receiver antenna. $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ and h_{ij} is the channel gain between the i^{th} transmitter antenna and j^{th} receiver antenna. $\mathbf{x}(n) \in \mathbb{C}^{M_t \times 1}$ and $\boldsymbol{\eta}(n) \in \mathbb{C}^{M_r \times 1}$ are the transmitted symbol vector and the noise vector at the receiver end respectively. It is assumed that the channel gain matrix $\mathbf{H}$ is known to both the transmitter and the receiver. The MIMO channel matrix $\mathbf{H}$ can be decomposed using the Singular Value Decomposition (SVD) as [58]

$$\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H, \tag{2.4.11}$$

where $\mathbf{U} \in \mathbb{C}^{M_r \times M_r}$ and $\mathbf{V} \in \mathbb{C}^{M_t \times M_t}$ are unitary left and right singular matrices of $\mathbf{H}$. $\boldsymbol{\Sigma} \in \mathbb{R}^{M_r \times M_t}$ is a diagonal matrix with the singular values

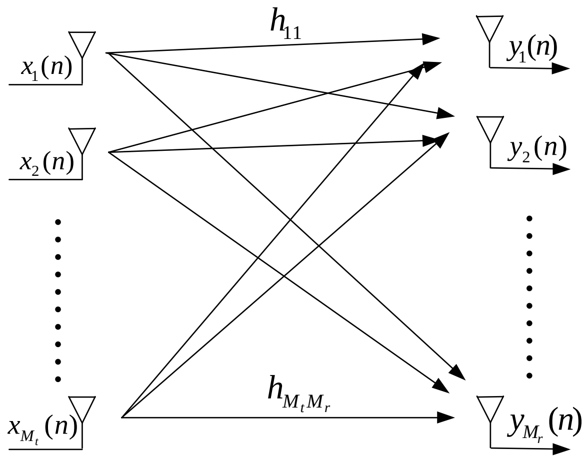

Figure 2.5. A MIMO system with M_t transmit antennas and M_r receive antennas.

(υ_i) of $\mathbf{H}$. R_H number of singular values are nonzero, so that R_H is the rank of matrix $\mathbf{H}$. The singular value satisfies the property $\upsilon_i = \sqrt{\lambda_i}$, where λ_i is the i^{th} eigenvalue of $\mathbf{H}\mathbf{H}^H$. These MIMO spatial subchannels are obtained using linear transformation of the input signal and the output signal through transmit precoding and receiver shaping. In transmit precoding, the modulated symbol stream is precoded as

$$\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}, \tag{2.4.12}$$

where $\tilde{\mathbf{x}}$ is the modulated symbol stream. Similarly, the received signal is shaped as

$$\tilde{\mathbf{y}} = \mathbf{U}^H\mathbf{y} \tag{2.4.13}$$

as shown in Fig. 2.6. Such transmit precoding and receiver shaping de-

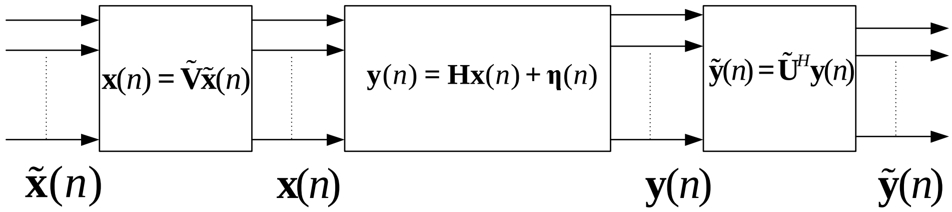

Figure 2.6. Transmit precoding and receiver shaping.

compose the MIMO channel into R_H number of independent single-input single-output (SISO) channels as follows:

$$\begin{aligned}\tilde{\mathbf{y}} &= \mathbf{U}^H(\mathbf{H}\mathbf{x}+\boldsymbol{\eta}) \\ &= \mathbf{U}^H\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{H}}\mathbf{V}\tilde{\mathbf{x}}+\mathbf{U}^H\boldsymbol{\eta} \\ &= \boldsymbol{\Sigma}\tilde{\mathbf{x}}+\tilde{\boldsymbol{\eta}} \qquad (2.4.14)\end{aligned}$$

where $\tilde{\boldsymbol{\eta}} = \mathbf{U}^H\boldsymbol{\eta}$. The resulting parallel spatial subchannels are shown in Fig. 2.7. They are independent from each other in the sense that signals through each spatial subchannels do not interfere with each other. Hence

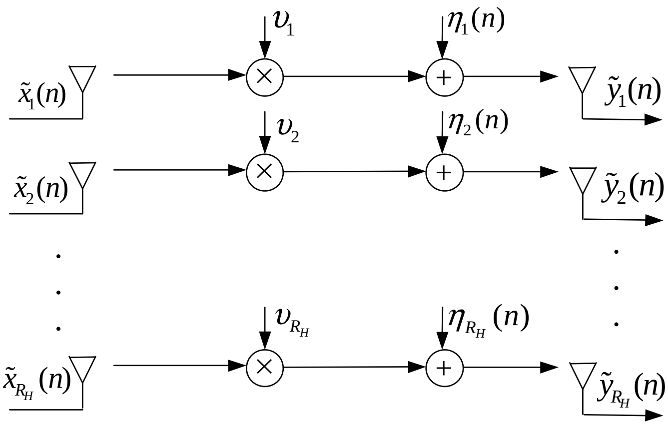

Figure 2.7. Parallel decomposition of the MIMO channel.

this MIMO channel can support up to R_H times the data rate of a SISO channel.

2.4.4 Resource Allocation Techniques Using Multiple Antennas

The power allocation and beamforming problems for multiple users have been widely studied to control interference between users in [45, 46]. In [59], an optimal downlink power assignment technique has been proposed for a given set of beamforming weight vectors. This power allocation problem is formulated into an eigenvector matrix equation and the optimal power allocations have been obtained by finding the eigenvector corresponding the largest eigenvalue of the matrix. The property that all elements of the eigenvector corresponding to the largest eigenvalue of a non-negative matrix are always positive [60], has been exploited in [59]. In [61], an iterative algorithm has been proposed to jointly design the beamforming weight vectors and the power allocation vectors in the uplink and the downlink. This design ensures that SINR of each user is above a threshold while minimizing the total transmission power. The same problem has been formulated into a SDP in [45,46] by using Lagrangian relaxation and it has been solved using interior point methods [34]. This relaxed problem provides a rank-one solution for each user and the optimal beamforming weight vector has been determined by extracting the eigenvector corresponding to the positive eigenvalue of the matrix. In addition it has been proved that the relaxed problem always yields an optimal rank-one solution. In [62], a special scenario has been considered where the transmitter sends the same data to multiple users known as multicasting. In the multicasting setup, the SDP formulation might not always provide a rank-one solution. To overcome this problem, a randomization technique [63] has been recommended to find an optimal solution. The problem of transmit beamforming to multiple cochannel multicast groups is considered in [64, 65], where QoS and the max-min fairness approach have been presented using convex optimization and randomization techniques.

The approaches developed in [61] and [46] use the criterion of minimizing the total transmit power subject to SINR constraints for each user. However, the resulting problem might become infeasible due to tight SINR constraints (higher SINR) or insufficient total transmitter power. To avoid the problem of infeasibility, an SINR balancing technique has been proposed in [41]. Here an iterative algorithm has been developed to maximize the worst-case user SINR, where the beamforming weight vectors are designed in the virtual uplink mode have been employed in the downlink using the principle of uplink-downlink duality [16, 51, 52]. The solution to this iterative algorithm provides a balanced ratio of the individual achieved SINR and the target SINR value for all users.

Resource allocation technique based on combined MIMO and OFDM techniques has been identified as one of the most promising physical-layer transmission technique for its capability to achieve high capacity, high spectral efficiency, and good performance in dispersive channels [66–68]. Multiuser wireless system based on MIMO-OFDM is considered in [66]. In [66] multiple users are allocated in each OFDM subchannel based on their spatial correlation between the users. The spatial correlation between two users were obtained using the vector multiplication of both users' spatial subchannel gains. High spatial correlation between users can cause high inter-user interference. Hence, users with lower spatial correlation are allocated in the same subchannel. After user-subchannel allocation, greedy algorithm is used to allocate transmission power in order to satisfy the QoS requirement of each users.

2.5 Resource Allocation Techniques for Cognitive Radio Networks

The conventional resource allocation techniques discussed so far in this chapter cannot trivially be extended to cognitive radio networks due to the additional interference constraints imposed by primary users. Hence, in literature, various resource allocation techniques have been used to allocate radio resources to secondary users while maintaining the interference leakage to the primary users is below the threshold. In this section, various algorithms proposed for cognitive radio networks based on OFDMA technique and/or multiple antennas technique are discussed.

The OFDMA has been recognized as one of the best candidates for the resource allocation in cognitive radio networks because of its natural ability to utilize different portions of the spectrum [7]. In [69–72], resource allocation techniques have been studied for an OFDMA based cognitive radio networks. In [69], power is allocated to each subchannel, by considering the received interference as a fairness metric. This problem has been solved using Lagrangian dual function. On the other hand, adaptive power loading has been investigated for OFDM-based cognitive radio network in [70]. This scheme is developed based on a non-integer Lagrangian formulation. However, power allocation problem together with subchannel allocation problem will not provide a closed-form solution [20]. Hence, joint subchannel, transmission power and data bit allocation problem for OFDMA based cognitive radio network is studied in [73]. Due to high computational complexity of determining optimal solution for OFDMA based resource allocation problems, low complexity algorithms have been proposed in [71, 72]. The work in [71, 72] consider a single user cognitive radio network and maximizes the cognitive radio network throughput by allocating data bits to various subchannels. Conventional greedy algorithm has been modified in [71, 72] to allocate resources to the cognitive radio networks based on efficiency factors. Low computational complexity resource allocation algorithm for multiuser OFDMA based cognitive radio network is proposed in [74].

In [75], spatial diversity has been exploited in the downlink to improve the throughput of the secondary user, while imposing constraints on the secondary user transmit power and the primary user interference power. A beamforming approach has been proposed to maximize the ratio between the received secondary user signal power and the interference power leakage to the primary users in [76]. In [77], joint beamforming and power allocation techniques have been provided for an uplink cognitive radio network. A multi-level water filling algorithm and a recursive decoupled power allocation algorithm have been presented to maximize the sum-rate of the secondary users. A multicast beamforming technique based on convex optimization has been presented for a QoS aware spectrum sharing underlay cognitive radio network in [78]. In [79], two joint power control and beamforming algorithms have been proposed based on a weighted least squares approach and admission control technique for an underlay cognitive radio network.

Resource allocation based on MIMO-OFDMA techniques have been studied for multiuser cognitive radio network in [80, 81]. The work in [80] optimally allocates spatial beams, subchannels and transmission power to various users based on their QoS requirements. The uplink resource allocation problem is studied in [81], where each subchannel is allocated various users based on their spatial separation to avoid the inter-user interference. After user-subchannel allocation integer linear programming frame work has been used to allocate transmission power and data bits according to each user QoS requirement.

Chapter 3

OPTIMAL RESOURCE ALLOCATION TECHNIQUES FOR OFDMA BASED COGNITIVE RADIO NETWORKS

In this chapter, resource allocation algorithms are proposed for an OFDMA based cognitive radio networks. An OFDMA technology has been widely adopted in wireless communications standards and is one of the best candidates for cognitive radio networks because of its natural ability to utilize different portions of the spectrum. The proposed algorithms optimally allocate OFDMA subchannels, data bits and transmission power to various secondary users at secondary network basestation while ensuring the interference leakage threshold to the primary users below a threshold. In order to exploit the spatial diversity, the proposed schemes are extended to a MIMO system which is discussed later in this chapter.

3.1 Introduction

There are two classes of radio resource allocation problems, namely, the margin adaptive problem and the rate adaptive problem [26]. The objective of margin adaptive problem is to minimize the overall transmission power subject to data rate constraints, whereas the objective of rate adaptive problem is to maximize the overall data throughput subject to a transmission power constraint. The resource allocation algorithms in OFDM and OFDMA-based systems have been studied for many different scenarios in [20,26,66,69,70,82].

In [26] and [20], a joint subchannel, bit and power allocation problem has been studied for OFDMA and OFDM based conventional wireless systems (i.e., without interference constraint). Resource allocation problem in [26] formulated into an integer linear programming framework and solved by branch and bound method. In [20], subchannels and bits are allocated to different users based on greedy algorithm. In [82] subchannels are allocated under the assumption of fixed power constraint for each user. In real scenarios, however, power and frequency allocations are intertwined. A suboptimal resource allocation algorithm is presented for MIMO-OFDM-based uplink system in [66]. The algorithm in [66] schedules the users based on their spatial separability in each subchannel and then allocates bits and power to each user. Grouping the users according to their spatial signature has been shown to reduce inter user interference [66]. Resource allocation algorithms developed for conventional wireless systems cannot be used directly to a cognitive radio network due to the additional interference constraint to primary users. The amount of interference introduced to primary users by a secondary network basestation depends on the power allocated to each subchannels and the corresponding subchannel gain between the secondary network basestation and the primary users (interference subchannel gain).

Resource allocation problem for an OFDM-based cognitive radio network is studied in [69] and [70]. In [69], transmission power is allocated to each subchannel, by considering the received interference as a fairness metric. This problem has been solved using Lagrangian dual function. On the other hand, adaptive power loading has been investigated in [70]. This scheme is developed based on a non-integer Lagrangian formulation. However, disjoint power allocation and subchannel allocation in a resource allocation problem will not provide an optimal solution [20]. Therefore, in this chapter an integer linear programming approach is considered for joint subchannel, bit and power allocation in cognitive radio network, which is different from the work presented in [69] and [70].

3.2 SISO-OFDMA based Downlink Cognitive Radio Network

In this section, a SISO-OFDMA based cognitive radio network is considered in the downlink. An algorithm is proposed to allocate OFDMA subchannels and transmission powers to various secondary users while ensuring interference leakage to the primary users is below a specific value. At the secondary network basestation, transmission power for each subchannel is adapted based on channel gain and different modulation schemes such as Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying, (QPSK) and M-ary Quadrature Amplitude Modulation (M-ary QAM). An integer linear programming technique has been used to formulate the resource allocation problem into a mathematical framework. Simulation results have been provided later in this section to validate the performance of the algorithm.

3.2.1 System Model and Problem Statement

A cognitive radio network with K secondary users and L primary users is considered in an underlay approach. It is assumed that secondary network basestation employs OFDMA scheme and the number of subchannels available in the downlink is denoted by N.

The radio resource allocation algorithm for OFDMA-based cognitive radio network performs subchannel and modulation scheme selection based on bit allocation and the corresponding power allocation for each secondary user at the secondary network basestation. The modulation and subchannel indexes are informed to the secondary users by the secondary network basestation through a separate control channel. It is assumed that the transmission over control channel is error-free and all subchannels are in slow-fading.

Adaptively selecting different modulation schemes significantly improves the performance of OFDMA-based system. Various modulation schemes, such as M-ary Frequency Shift Keying (M-ary FSK), M-ary Phase Shift Keying (M-ary PSK) and M-ary QAM, can be used for data transmission based on bandwidth and power efficiency [2]. Moreover, bandwidth efficiency of a M-ary FSK signal decreases with increasing modulation order. But M-ary PSK and M-QAM keep equal bandwidth efficiency for all modulation orders. In terms of power of efficiency, M-ary QAM is more efficient than M-ary PSK [2].

Four different modulation schemes considered for the problem formulation: BPSK (c =1), 4-QAM (c =2), 16-QAM (c =4) and 64-QAM (c =6) where number of bits per symbol is denoted by c. The minimum signal power required in any subchannel to achieve a given BER, ρ_e at the receiver for BPSK modulation scheme is given by [83]

$$P_{\mathrm{r}}(1, \rho_e) = N_\phi[\mathrm{erfc}^{-1}(2\rho_e)]^2, \tag{3.2.1}$$

where $\mathrm{erfc}(x) = \frac{2}{\sqrt{\Pi}} \int_x^\infty e^{-t^2} dt$, and N_ϕ is the single-sided noise power spectral density which is assumed to be the same for all subchannels. For M-ary QAM (square constellation), the required power for a given BER can be written as [83]

$$P_{\rm r}(c,\rho_e) = \frac{2(M-1)N_\phi}{3}\left[\operatorname{erfc}^{-1}\left(\frac{c\rho_e\sqrt{M}}{2(\sqrt{M}-1)}\right)\right]^2, \tag{3.2.2}$$

where M is the modulation order given by $M = 2^c$. The required power increases as the number of bits per symbol increases for a given BER, ρ_e.

Subchannel gain matrix between the secondary network basestation and the secondary users can be defined as follows $\mathbf{H}_{\rm s} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{K1} & \alpha_{K2} & \dots & \alpha_{KN} \end{bmatrix} \in R^{K\times N}$, where α_{kn} represents the magnitude of the channel gain of the n^{th} subchannel for k^{th} secondary user. The required transmission power at the secondary network basestation (in energy per symbol) to achieve a certain received power at the k^{th} secondary user terminal over n^{th} subchannel for a given modulation scheme and BER is given by the following equation

$$P(k,n) = \frac{P_{\rm r}(c,\rho_e)}{\alpha_{kn}^2}, \tag{3.2.3}$$

where $P_{\rm r}(c,\rho_e)$ is given by (3.2.1) and (3.2.2). Interference subchannel gain matrix between the secondary network basestation and the primary users can be defined as $\mathbf{H}_{\rm p} = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1N} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{L1} & \beta_{L2} & \dots & \beta_{LN} \end{bmatrix} \in R^{L\times N}$, where β_{ln} represents the interference channel gain between the l^{th} primary user and the secondary network basestation over n^{th} subchannel. Denote $c(k,n)$ and $D = \{0,1,2,4,6\}$ as the number of bits allocated to the k^{th} secondary user on the n^{th} subchannel and the set of all possible values for $c(k,n)$ respectively. Denote the total transmission power allocated to the n^{th} subchannel as $\phi(n) = \sum_{k=1}^{K} P(k,n)$. Define the interference leakage threshold to the l^{th} primary user should be less than Υ_l, required for the k^{th} secondary user should be larger than r_k and the total transmission power at the secondary network basestation should be less than P. Using these definitions, the objective of the original resource allocation problem can be stated as maximizing the total data throughput subject to individual data rate constraints to the each secondary users and interference leakage constraints to the primary users as follows:

$$\max_{c(k,n)\in D} \quad \sum_{n=1}^{N}\sum_{k=1}^{K} c(k,n), \tag{3.2.4}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} \phi(n)\beta_{ln}^2 \leq \Upsilon_l, \quad l = 1, \ldots, L, \tag{3.2.5}$$

$$\sum_{n=1}^{N} c(k,n) \geq r_k, \quad k = 1, \ldots, K, \tag{3.2.6}$$

$$\sum_{n=1}^{N} \phi(n) \leq P, \tag{3.2.7}$$

$$c(k,n) = 0 : c(k',n) \neq 0, \forall k \neq k', \; k = 1, \ldots, K, \tag{3.2.8}$$

The objective function in (3.2.4) represents the sum throughput. Interference leakage constraints to the primary users, required data rate constraints to the secondary users and total transmission power constraint at the secondary network basestation are given in (3.2.5) - (3.2.7), respectively. The last constraint (3.2.8) performs mutual exclusive secondary user allocation in each subchannel. Due to the mutual exclusive secondary user allocation, it is possible to allocate more than one subchannel for one secondary user, but more than one secondary user cannot be accompanied together in one subchannel. Allocating radio resources to secondary users without considering the QoS requirement in (3.2.6) may cause unfairness among the secondary users. Because secondary users that are far away from the secondary network basestation may have poor channels and certain data rate should be guaranteed to these users using the constraint in (3.2.6). Due to the non-linear property of (3.2.2), problem introduced in (3.2.4)-(3.2.8) becomes non-linear in nature. Since the solution of this problem should be mutual exclusive secondary user allocation, this problem becomes a non-deterministic polynomial time hard problem [33]. In order to determine the optimal solution of this non-deterministic polynomial time hard problem, an integer linear programming technique is used in the next subsection.

3.2.2 Integer Linear Programming Problem Formulation

To formulate the problem in (3.2.4)-(3.2.8) into an integer linear programming framework, first, a binary vector $\mathbf{x}$ is defined as follows:

$$\mathbf{x} = [\mathbf{x}(1)^T \ \dots \ \mathbf{x}(N)^T]^T \ \in \ \{0,1\}^{NKC\times 1}, \tag{3.2.9}$$

where C is the number of modulation schemes considered in this work (i.e., $C = 4$), $\mathbf{x}(n) = [\mathbf{x}(1,n)^T \ \dots \ \mathbf{x}(K,n)^T]^T$ $\in \ \{0,1\}^{KC\times 1}$ and $\mathbf{x}(k,n) = [x(k,n,1) \ \dots x(k,n,C)]^T \in \{0,1\}^{C\times 1}$. If $x(k,n,c) = 1$ then c number of bits can be transmitted using n^{th} subchannel to the k^{th} secondary user. In order to ensure no more than one user is allocated in each subchannel only one of the element of $\mathbf{x}(n)$ should be equal to one and rest of them should be zeros. Therefore there are the following $KC + 1$ combinations of possible vectors for $\mathbf{x}(n)$

$$\mathbf{x}(n) \in \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}.$$

A modulation order vector $\mathbf{b}$ is defined as

$$\mathbf{b} = [\mathbf{b}(1)^T \ \ldots \ \mathbf{b}(N)^T]^T \ \in \ \mathcal{Z}^{KNC\times 1} \tag{3.2.10}$$

where $\mathbf{b}(n) = [\mathbf{b}(1,n)^T \ \ldots \ \mathbf{b}(K,n)^T]^T \ \in \ \mathcal{Z}^{KC\times 1}$ and $\mathbf{b}(k,n) = [b(k,n,1) \ \ldots \ b(k,n,C)]^T \in \mathcal{Z}^{C\times 1}$. Only four modulation schemes are considered, hence $\mathbf{b}(k,n) = [1 \ 2 \ 4 \ 6\ ]^T$ and $\mathbf{x}(k,n) = [x(k,n,1) \ x(k,n,2) \ x(k,n,3) \ x(k,n,4)]^T$. Using these definitions, the sum throughput in (3.2.4) can be written as $\mathbf{b}^T\mathbf{x}$. Moreover, the transmit power vector $\mathbf{p}$ can be defined as

$$\mathbf{p} = [\mathbf{p}(1)^T \ \ldots \ \mathbf{p}(N)^T]^T \ \in \ \mathcal{R}^{KNC\times 1} \tag{3.2.11}$$

where $\mathbf{p}(n) = [\mathbf{p}(1,n)^T \ \ldots \ \mathbf{p}(K,n)^T]^T \ \in \ \mathcal{R}^{KC\times 1}$ and $\mathbf{p}(k,n) = [p(k,n,1) \ \ldots \ p(k,n,C)]^T \in \mathcal{R}^{C\times 1}$. Referring to (3.2.3), $p(k,n,m)$ is determined as $p(k,n,m) = \frac{P_{\mathrm{r}}(d_m,\rho)}{\alpha_{kn}^2}$. Therefore, the power constraint in (3.2.7) can be written as $\mathbf{p}^T\mathbf{x} \leq P$. In order to write the interference constraint in (3.2.5) using the vector $\mathbf{x}$, define a matrix $\mathbf{A} \ \in \ \{0,1\}^{N\times NKC}$ as follows

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}(1) & \ldots & \mathbf{A}(N) \end{bmatrix} \in \ \{0,1\}^{N\times KNC} \tag{3.2.12}$$

where $\mathbf{A}(n) = \begin{bmatrix} \mathbf{0}_{n-1,KC} \\ \mathbf{1}_{1,KC} \\ \mathbf{0}_{N-n,KC} \end{bmatrix} \in \ \{0,1\}^{N\times KC}$. According to (3.2.12), $\mathbf{A}(\mathbf{p} \odot \mathbf{x})$ is an $N \times 1$ vector whose n^{th} element characterizes the total power used for the n^{th} subchannel. Defining $\boldsymbol{\Upsilon} \triangleq [\Upsilon_1 \ \ldots \ \Upsilon_L]^T \ \in \ \mathcal{R}^{L\times 1}$ and $\mathbf{H}^{\mathrm{p}} = \mathbf{H}_{\mathrm{p}} \odot \mathbf{H}_{\mathrm{p}}$, the interference constraint in (3.2.5) can be written as $\mathbf{H}^{\mathrm{p}}[\mathbf{A}(\mathbf{p} \odot \mathbf{x})] \leq \boldsymbol{\Upsilon}$. The rate for user k can be written as $\sum_{n=1}^{N} \mathbf{b}_{k,n}^T \mathbf{x}_{k,n}$. To write the data rate requirement for all users in the matrix form, define the following matrix $\mathbf{B} = \begin{bmatrix} \mathbf{B}(1) & \ldots & \mathbf{B}(N) \end{bmatrix} \in \ \mathcal{Z}^{K\times KNC}$, where

$\mathbf{B}(n) = \begin{bmatrix} \mathbf{B}(1,n) & \dots & \mathbf{B}(K,n) \end{bmatrix} \in \mathcal{Z}^{K \times KC}$, $\mathbf{B}(k,n) = \begin{bmatrix} \mathbf{0}_{k-1,C} \\ \mathbf{b}_{k,n}^T \\ \mathbf{0}_{K-k,C} \end{bmatrix}$ $\in \mathcal{Z}^{K \times C}$. Defining $\mathbf{r} \triangleq [r_1 \ \dots r_K]^T \in \mathcal{R}^{K \times 1}$, the data rate constraint in (3.2.6) can be written as $\mathbf{Bx} \geq \mathbf{r}$. Therefore the problem (3.2.4)-(3.2.8) can be reformulated in the integer linear programming form as [26]

$$\max_{\mathbf{x}} \quad \mathbf{b}^T\mathbf{x} \tag{3.2.13}$$

$$\text{s.t.} \quad \mathbf{H}^{\text{p}}[\mathbf{A}(\mathbf{p} \odot \mathbf{x})] \leq \mathbf{\Upsilon}, \tag{3.2.14}$$

$$\mathbf{p}^T\mathbf{x} \leq P, \tag{3.2.15}$$

$$\mathbf{Bx} \geq \mathbf{r}, \tag{3.2.16}$$

$$\mathbf{0}_N \leq \mathbf{Ax} \leq \mathbf{1}_N, \tag{3.2.17}$$

$$x_i \in \{0, 1\}, \quad i = 1, \dots, KNC. \tag{3.2.18}$$

The formulated integer linear programming problem in (3.2.13)-(3.2.18) can be efficiently solved using branch and bound method [29].

3.2.3 Simulation Results

In order to validate the performance of the proposed algorithm, an underlay cognitive radio network with two primary users and two secondary users has been considered. Moreover OFDMA system divides the available bandwidth into 64 subchannels ($N = 64$). A random multipath channel of length six is considered secondary network basestation and each secondary user and primary user terminals. The required BER, noise power spectral density and data rate for each secondary user have been set to rate for each secondary user have been set to 0.01, 0.01 and 32 bits per snapshot, respectively. The upper bound on the interference power (summed over 64 subchannels) leaked to the primary user has been set to 20.

The channel gain between the secondary network basestation and pri-

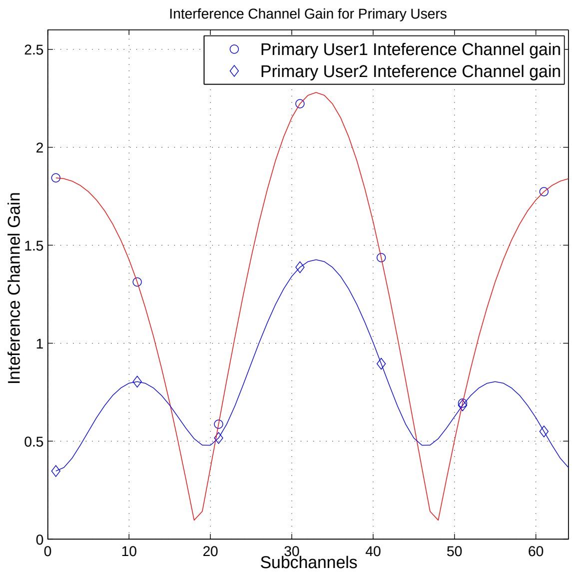

Figure 3.1. Interference subchannel gains between two primary users and the secondary network basestation. For both the primary users, subchannels 18, 19, 20, 21 and 45, 46, 47 from the secondary network basestation are in deep fading compared to other subchannels.

mary users as well as the one between the secondary network basestation and secondary users have been shown in Fig. 3.1 and Fig. 3.2 respectively. Also, number of bits allocated for in each subchannel is indicated in Fig. 3.2 using a plus mark, "+", for secondary user1 and using a star mark, "*", for secondary user2. For example a star mark at 4 for subchannel 20 in Fig. 3.2 means four bits are allocated for secondary user2 in the 20^{th} subchannel. subchannel. Fig. 3.1 and Fig. 3.2 reveal that the secondary network basestation allocates more bits into the subchannels that the channel gain between secondary network basestation and primary user receiver are in deep fade. Specifically, subchannels 18, 19, 20, 21, 45, 46, 47 are in deep fading between primary users and the secondary network basestation. Therefore secondary network basestation allocates symbols with more bits (16-QAM) to the secondary users through these subchannels. It is worth to note that secondary network basestation treats these subchannels as good subchannels and allocate more radio resources to ensure QoS of each secondary user.

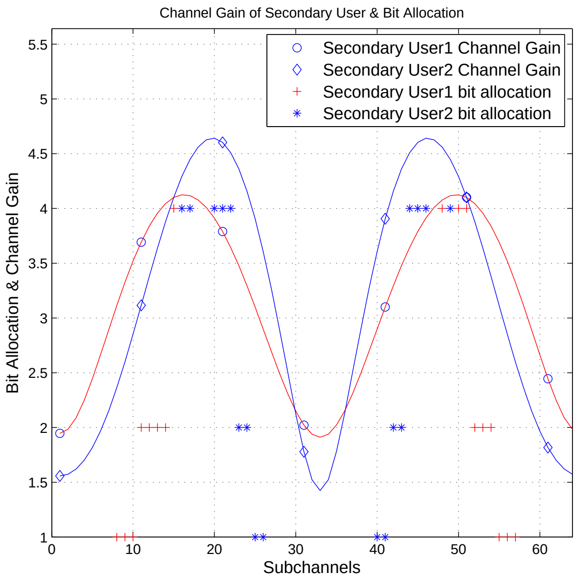

Figure 3.2. subchannel gains between two secondary users and secondary network basestation. Secondary network basestation allocates bits on each subchannel for the secondary users by considering the primary user's current interference subchannel gain in (Fig. 3.1).

On the other hand, the channel gains between secondary network basestation and primary user receiver in subchannels between 1 to 8, 27 to 35 and 59 to 64 are very high compared to other subchannels. Therefore the secondary network basestation, avoids higher order modulations such as 16-QAM and 64-QAM in these subchannels to reduce interference to primary users. This is because, choosing higher order modulation consumes more transmission power and it the primary users occupied in the mentioned subchannels. In the next section, the proposed algorithm is extended to a MIMO-OFDMA based cognitive radio network.

3.3 MIMO-OFDMA based Downlink Cognitive Radio Network

In this section, a MIMO-OFDMA based cognitive radio network is considered in the downlink. A MIMO wireless system contains multiple antennas at both the transmitter and the receiver and has the potential to enhance spatial diversity and data throughput [18]. On the other hand, OFDMA scheme exploits multiuser diversity and perhaps the best candidate technology for cognitive radio network because of its natural ability to utilize different portions of the spectrum [84]. Hence, the resource allocation for cognitive radio network by exploiting the MIMO-OFDMA technologies enhance the spectrum utilizations.

A rate adaptive optimal algorithm is proposed to allocates radio resources to each secondary user using an integer linear programming frame work. The formulated integer linear programming framework has been solved using branch and bound method. The solution of this problem jointly allocates subchannels, spatial beams, powers and bits to each secondary user while satisfying the data rate and BER requirements for each secondary user. It also ensures that the interference leakage to the primary users is always less than a specific value. To null the mutual interference between secondary users, only one secondary user is allowed in each subchannel. However, the approach proposed in this section can be extended to deal with multiple spatial beams from more than one secondary user in each subchannel. In this case the dimension of the subchannel-user allocation vector (to be introduced later) will be increased substantially due to consideration of increased number of combinations. Since each subchannel has several spatial subchannels, the proposed algorithm adaptively assigns optimal number of bits in different subchannels to maximize the data throughput.

3.3.1 System Model and Problem Statement

A downlink cognitive radio network is considered with K secondary users where each secondary user is equipped with N_r receiver antennas. The secondary network basestation in the cognitive radio network consists of N_t transmit antennas. There are L primary users in the primary network and each consists of one receiver antenna[1]. The frequency band is divided into N subchannels. The length of the cyclic prefix is assumed to be longer than the maximum time dispersion of the channels.

The resource allocation algorithm performs subchannel and modulation scheme selection based on the channel gains seen by the secondary network basestation. The modulation schemes and subchannel indices are made known to all the secondary users by the secondary network basestation through a control channel. Hence, each secondary user needs to decode bits only on its assigned subchannel. It is assumed that the transmission over control channel is error-free and all subchannels are in slow-fading.

In order to keep the problem formulation simple, only BPSK (c =1) and 4-QAM (c =2) modulation schemes are considered, where the number of bits per symbol is denoted by c. However, the proposed work can be readily ex-

[1]This work can be extended to a network which has multiple antennas at the primary users' receiver, by aggregating interference at multiple antennas through trace operator. This will not change the overall framework of the optimization.

tended to large number of modulation schemes. For the BPSK and 4-QAM modulation schemes, the minimum signal power required to ensure the BER at the receiver is below a threshold ρ_e can be obtained using (3.2.1) and (3.2.2), respectively. The downlink channel matrix between the secondary network basestation and the k^{th} secondary user at the n^{th} subchannel can be defined as $\mathbf{H}(k,n) = \begin{bmatrix} \alpha_{11}(k,n) & \dots & \alpha_{1N_t}(k,n) \\ \alpha_{21}(k,n) & \dots & \alpha_{2N_t}(k,n) \\ \vdots & \ddots & \vdots \\ \alpha_{N_r1}(k,n) & \dots & \alpha_{N_rN_t}(k,n) \end{bmatrix} \in \mathcal{C}^{N_r \times N_t}$, where $\alpha_{ij}(k,n)$ is the complex channel gain from the j^{th} transmit antenna to the i^{th} receiver antenna of the k^{th} secondary user at the n^{th} subchannel. Using SVD, $\mathbf{H}(k,n)$ can be decomposed into

$$\mathbf{H}(k,n) = \mathbf{U}(k,n)\mathbf{\Sigma}(k,n)\mathbf{V}(k,n)^H, \tag{3.3.1}$$

where $\mathbf{U}(k,n) \in \mathcal{C}^{N_r \times S}$ and $\mathbf{V}(k,n) \in \mathcal{C}^{S \times N_t}$ are the left and right singular matrices, $\mathbf{\Sigma}(k,n) = \mathbf{diag}(\lambda(k,n,1),\dots,\lambda(k,n,S)) \in \mathcal{R}_+^{S\times S}$ is a singular value matrix with singular values in descending order and $S = \min(N_r, N_t)$. The MIMO channel can be decomposed into S number of independent spatial sub-channels with channel gains $\lambda(k,n,1),\dots,\lambda(k,n,S)$ when $\mathbf{V}(k,n)$ and $\mathbf{U}(k,n)^H$ are used as transmit and receive beamformers respectively. The required transmission power at the secondary network basestation to transmit c bits to the k^{th} secondary user over the s^{th} spatial subchannel at the n^{th} subchannel for a given modulation scheme and BER target ρ is given by the following equation,

$$p(k,n,s,c) = \frac{p_{\mathrm{r}}(c,\rho)}{\lambda^2(k,n,s)}. \tag{3.3.2}$$

The interference channel vector between secondary network basestation and the l^{th} primary user at the n^{th} subchannel is defined as

$$\mathbf{h}(l,n) = \left[\begin{array}{ccc} \gamma_1(l,n) & \dots & \gamma_{N_t}(l,n) \end{array} \right] \in \mathcal{C}^{1\times N_t},$$

where $\gamma_i(l,n)$ is the complex channel gain from the i^{th} transmit antenna to the l^{th} primary user at the n^{th} subchannel. Denote $c(k,n,s)$ and $D = \{0,\ d_1,\dots,d_C\} = \{0,\ 1,\ 2\}$ as number of bits allocated to the k^{th} secondary user on the s^{th} spatial subchannel of n^{th} subchannel and the set of all possible values for $c(k,n,s)$. Define the interference leakage threshold to the l^{th} primary user should be less than Υ_l, required data rate for the k^{th} secondary user should be larger than r_k and transmission power at the secondary network basestation should be less than P. Denote the s^{th} column of matrix $\mathbf{V}(k,n)$ as $\mathbf{v}(k,n,s)$. From these definitions, the rate adaptive resource allocation problem for MIMO-OFDMA based cognitive radio networks can be expressed as

$$\max_{c(k,n,s)\in D} \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{s=1}^{S} c(k,n,s) \tag{3.3.3}$$

$$\text{s.t.} \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{s=1}^{S} \mid \mathbf{h}(l,n)\mathbf{v}(k,n,s) \mid^2 p(k,n,s,c) \le \Upsilon_l, l = 1,\dots,L, \tag{3.3.4}$$

$$\sum_{n=1}^{N}\sum_{s=1}^{S} c(k,n,s) \ge r_k, \quad k = 1,\dots,K, \tag{3.3.5}$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{s=1}^{S} p(k,n,s,c) \le P, \tag{3.3.6}$$

$$c(k,n,s) = 0 \text{ if } c(k',n,s) \ne 0,\ \forall k \ne k', k = 1,\dots,K,\ s = 1,\dots,S. \tag{3.3.7}$$

The objective function in (3.3.3) represents the sum throughput, where $c(k,n,s) = 0$ indicates that the k^{th} user does not use the s^{th} spatial subchannel for the transmission at the n^{th} subchannel. Interference leakage constraints to the primary users, required data rate constraints to the secondary users and transmission power constraint at secondary network basestation are given in (3.3.4)-(3.3.6), respectively. The constraint in (3.3.7) performs mutual exclusive secondary user allocation in each subchannel. Due to the non-linear property of (3.2.2), rate adaptive problem introduced in (3.3.3)-(3.3.7) becomes non-linear in nature. Since the solution of this problem should be mutual exclusive secondary user allocation, this problem becomes a combinatorial optimization problem. In the next section, this combinatorial optimization problem formulated into an integer linear programming framework in order to obtain the optimal solution.

3.3.2 Integer Linear programming

To convert the problem into the integer linear programming form, a bit combination matrix is defined as $\mathbf{R}(k,n) \in \mathcal{Z}^{S\times\varsigma^S}$ for the k^{th} secondary user at the n^{th} subchannel, where ς is cardinality of D. As an example, consider a system with only two spatial beams (i.e., $S = 2$) then the $\mathbf{R}(k,n)$ matrix can be expressed as $\mathbf{R}(k,n) = \begin{bmatrix} 2 & 2 & 2 & 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 2 & 1 & 0 & 2 & 1 & 0 \end{bmatrix} \in \mathcal{Z}^{2\times 3^2}$, where the numbers 0, 1 and 2 represent the number of bits in the modulation schemes considered in this section. For an example, the first column of the above matrix denotes that both spatial subchannel at the n^{th} subchannel are used to transmit two bits (i.e., 4-QAM). The sixth column denotes that first spatial subchannel transmits one bit (i.e., BPSK) while the second spatial subchannel does not transmits any bit. In general, there are ς^S possible combinations. The bit combination matrix for all K secondary users at the n^{th} subchannel is denoted by $\bar{\mathbf{R}}(n) = [\mathbf{R}(1,n) \ldots \mathbf{R}(K,n)] \in \mathcal{Z}^{S\times\varsigma^S K}$. The proposed algorithm optimally chooses only one column from $\bar{\mathbf{R}}(n)$ for the n^{th} subchannel. Similarly a transmit power matrix for the k^{th} secondary user at the n^{th} subchannel can be defined as $\mathbf{P}(k,n) \in \mathcal{R}_{+}^{S\times\varsigma^S}$, i.e., $P_{i,j}(k,n)$ denotes the required transmit power in order to transmit $R_{i,j}(k,n)$ number of bits to the the k^{th} secondary user at the n^{th} subchannel. Referring to

(3.3.2), $P_{i,j}(k,n)$ is determined as

$$P_{i,j}(k,n) = p(k,n,i,R_{i,j}(k,n)) = \frac{p_{\text{r}}(R_{i,j}(k,n),\rho)}{\lambda^2(k,n,i)}. \quad (3.3.8)$$

The power matrix for all K secondary users at the n^{th} subchannel is denoted by $\bar{\mathbf{P}}(n) = [\mathbf{P}(1,n)\ldots\mathbf{P}(K,n)] \in \mathcal{R}_+^{S\times\varsigma^S K}$. Using the integer linear programming approach in [26] define a subchannel-user allocation vector $\hat{\mathbf{x}}$ as $\hat{\mathbf{x}} = [\bar{\mathbf{x}}(1)^T\ldots\bar{\mathbf{x}}(N)^T]^T \in \{0,1\}^{\varsigma^S KN\times 1}$, where $\bar{\mathbf{x}}(n) = [\mathbf{x}(1,n)^T\ldots\mathbf{x}(K,n)^T]^T \in \{0,1\}^{\varsigma^S K\times 1}$ represents the allocation for n^{th} subchannel and $\mathbf{x}(k,n) \in \{0,1\}^{\varsigma^S\times 1}$. According to the example above, $\mathbf{x}(k,n)$ is a 9×1 vector. For example, a one in the second element of $\mathbf{x}(k,n)$ and zeros else where means, the k^{th} secondary user uses first subchannel to transmit second subchannel to transmit BPSK using n^{th} subchannel. Due to the mutual secondary user allocation, only one element in $\bar{\mathbf{x}}(n)$ is equal to one and all others are equal to zeros.

Using these definitions, the sum throughput in (3.3.3) can be written as $\mathbf{1}_S^T\hat{\mathbf{R}}\hat{\mathbf{x}}$, where $\hat{\mathbf{R}} = [\bar{\mathbf{R}}(1)\ldots\bar{\mathbf{R}}(N)] \in \mathcal{Z}^{S\times\varsigma^S KN}$. The power constraint in (3.3.6) can be written as $\mathbf{1}_S^T\hat{\mathbf{P}}\hat{\mathbf{x}} \leq P$, where $\hat{\mathbf{P}} = [\bar{\mathbf{P}}(1)\ldots\bar{\mathbf{P}}(N)] \in \mathcal{R}_+^{S\times\varsigma^S KN}$. The data rate constraint in (3.3.5) can be written as

$$\mathbf{A}_u\hat{\mathbf{x}} \leq \mathbf{r},$$

where

$$\mathbf{A}_u = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_1 & \ldots & \mathbf{a}_1 \\ \mathbf{a}_2 & \mathbf{a}_2 & \ldots & \mathbf{a}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_K & \mathbf{a}_K & \ldots & \mathbf{a}_K \end{bmatrix} \in \mathcal{Z}^{K\times\varsigma^S KN},$$

$$\begin{array}{l} \mathbf{a}_1 = [\mathbf{a}^T \quad \mathbf{0}_{\varsigma^S}^T \quad \ldots \quad \mathbf{0}_{\varsigma^S}^T] \in \mathcal{Z}^{1\times\varsigma^S K}, \\ \mathbf{a}_2 = [\mathbf{0}_{\varsigma^S}^T \quad \mathbf{a}^T \quad \ldots \quad \mathbf{0}_{\varsigma^S}^T] \in \mathcal{Z}^{1\times\varsigma^S K}, \\ \quad\vdots \qquad\quad \vdots \qquad \vdots \qquad\qquad \vdots \\ \mathbf{a}_K = [\mathbf{0}_{\varsigma^S}^T \quad \mathbf{0}_{\varsigma^S}^T \quad \ldots \quad \mathbf{a}^T] \in \mathcal{Z}^{1\times\varsigma^S K} \end{array},$$

$\mathbf{a} = \left[\mathbf{1}_{\varsigma}^{T}\mathbf{R}(k,n)\right]^{T}$ and $\mathbf{r} = [\mathbf{r}_1\ \mathbf{r}_2\ \ldots\ \mathbf{r}_K]^T$. Mutual exclusive SU allocation constraint in (3.2.8) can be defined as

$$\mathbf{0}_N \ \leq\ \mathbf{A}_c\hat{\mathbf{x}} \ \leq\ \mathbf{1}_N,$$

where

$$\mathbf{A}_c = \begin{bmatrix} \mathbf{1}_{\varsigma^S K}^T & \mathbf{0}_{\varsigma^S K}^T & \cdots & \mathbf{0}_{\varsigma^S K}^T \\ \mathbf{0}_{\varsigma^S K}^T & \mathbf{1}_{\varsigma^S K}^T & \cdots & \mathbf{0}_{\varsigma^S K}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{\varsigma^S K}^T & \mathbf{0}_{\varsigma^S K}^T & \cdots & \mathbf{1}_{\varsigma^S K}^T \end{bmatrix} \in \{0,1\}^{N\times\varsigma^S KN}.$$

Define interference gain matrix to the primary users at the n^{th} subchannel by the k^{th} secondary user as $\mathbf{\Phi}(k,n) \in \mathcal{R}_+^{L\times\varsigma^S}$ where $\Phi_{l,j}(k,n) = [|\ \mathbf{h}(l,n)\mathbf{V}(k,n)\ | \odot |\ \mathbf{h}(l,n)\mathbf{V}(k,n)\ |]\,P^j(k,n)$ and $P^j(k,n)$ denotes the j^{th} column of $\mathbf{P}(k,n)$. Hence, interference constraints in (3.3.4) can be written as $\hat{\mathbf{\Phi}}\hat{\mathbf{x}} \leq \mathbf{\Upsilon}$ where $\hat{\mathbf{\Phi}} = [\bar{\mathbf{\Phi}}(1)\ldots\bar{\mathbf{\Phi}}(N)] \in R_+^{L\times\varsigma^S KN}$, $\bar{\mathbf{\Phi}}(n) = [\mathbf{\Phi}(1,n)\ldots\mathbf{\Phi}(K,n)] \in R_+^{L\times\varsigma^S K}$ and $\mathbf{\Upsilon} = [\Upsilon_1\ldots\Upsilon_L]^T \in \mathcal{R}_+^{L\times 1}$.

Therefore the problem defined by (3.3.3)-(3.3.7) can be formulated in an integer linear programming framework as

$$\max_{\hat{\mathbf{x}}} \quad \mathbf{1}_S^T\hat{\mathbf{R}}\hat{\mathbf{x}} \tag{3.3.9}$$

$$\text{s.t.} \quad \hat{\mathbf{\Phi}}\hat{\mathbf{x}} \ \leq\ \mathbf{\Upsilon}, \tag{3.3.10}$$

$$\mathbf{A}_u\hat{\mathbf{x}} \ \leq\ \mathbf{r}, \tag{3.3.11}$$

$$\mathbf{1}_S^T\hat{\mathbf{P}}\hat{\mathbf{x}} \ \leq\ P, \tag{3.3.12}$$

$$\mathbf{0}_N \ \leq\ \mathbf{A}_c\hat{\mathbf{x}} \ \leq\ \mathbf{1}_N, \tag{3.3.13}$$

$$\hat{x}_i \ \in \{0,1\}, i = 1,\ldots,\varsigma^S KN. \tag{3.3.14}$$

The above integer linear programming problem can be efficiently solved using branch and bound method [29].

3.3.3 Simulation Results

An OFDMA-based cognitive radio network with $K = 2$ secondary users, $L = 2$ primary users and $N = 16$ subchannels is considered. The minimum data rate requirement for each secondary user has been set to 8 bits/user. The BER requirement for the secondary users has been set to $\rho = 0.01$. The interference leakage to primary user is required to be less than 0.01mW. Channels were generated for primary users and secondary users using statistically independent Gaussian random variables. The average channel gain between the secondary network basestation and primary users is equal to 0.1 while such a gain between the secondary network basestation and secondary users is equal to 1. The interference caused by the primary user transmission is considered as noise at the secondary user receiver and the noise power spectral density at the secondary user is 0.01 mW/subchannel. All simulation results in this section were generated using were generated using 100 randomly generated channel-pairs $\{\mathbf{H}(k,n), \mathbf{h}(l,n)\}$.

In Fig. 3.3, the average total throughput versus P is displayed for a MIMO-OFDMA based cognitive radio network, where, for example, 3×3 denotes the number of antennas at the transmitter and the receiver (i.e., $N_r = 3,\ N_t = 3$). SISO system is denoted as 1×1 because it has only one antenna at both sides. As can be seen from Fig. 3.3, the throughput achieved by MIMO based cognitive radio networks are higher than that of a SISO system. Note that, the total data throughput increases with the number of antennas for a given transmitted power.

The outage probability that the problem in (3.3.3)-(3.3.7) is infeasible has been depicted in Fig. 3.4. When the value of P is very small, the probability that the problem becomes infeasible is large as seen in Fig. 3.4. For an example, at -10dBm, single antenna based cognitive radio network is un-

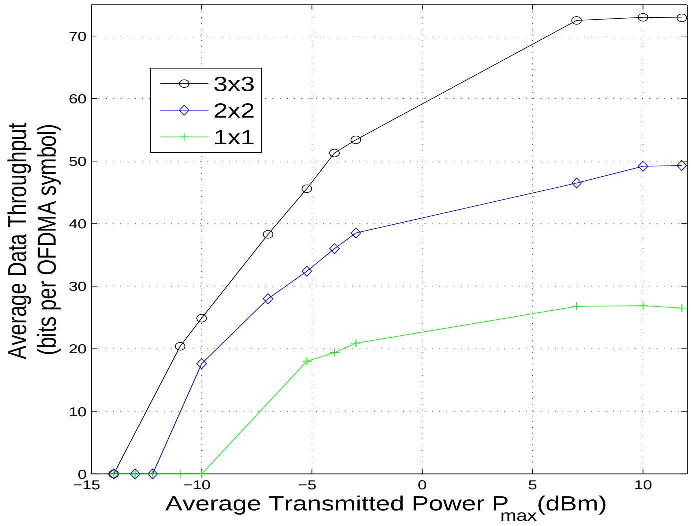

Figure 3.3. Total average number of bits per OFDMA symbol as a function of available transmitted power at the secondary network basestation. The interference value and the minimum data rate requirement for each secondary user have been set to -20 dBm and 8 bits/user respectively.

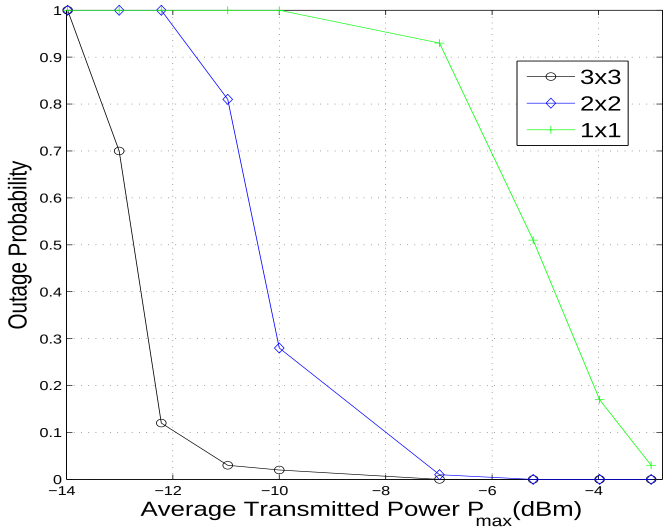

Figure 3.4. The outage probability that the problem in (3.3.3)-(3.3.7) becomes infeasible for various values of transmitted power. The interference value and the minimum data rate requirement for each secondary user have been set to -20 dBm and 8 bits/user respectively.

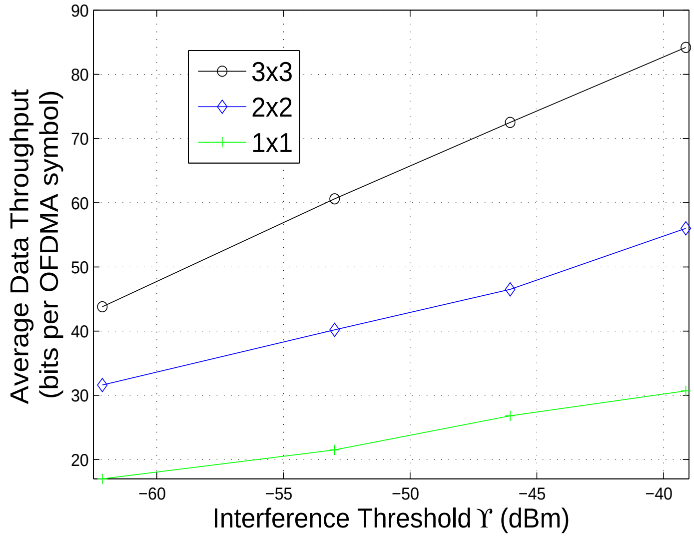

Figure 3.5. Total average number of bits per OFDMA symbol as a function of interference leakage values. The total transmitted power and the minimum data rate requirement for each secondary user have been set to 6.9 dBm and 8 bits/user respectively.

able to achieve the required data rate. On the other hand, for 3×3 MIMO based cognitive radio network, the data rate is satisfied with an of outage probability of less than 0.01.

In Fig. 3.5, the average total throughput versus interference leakage threshold Υ is depicted for 3×3 and 2×2 MIMO system and compared with a SISO system. The total data throughput increases as the bound on interference leakage is relaxed for all three antenna configurations. However, for increased antenna numbers, the slope changes faster and provide higher throughput.

3.4 MIMO-OFDMA based Uplink Cognitive Radio Network

In this section an adaptive radio resource allocation algorithm is proposed for a MIMO-OFDMA based uplink cognitive radio network. The cognitive radio network has multiple secondary users coexisting with multiple primary users. The aim is to admit as many secondary users as possible in various subchannels while ensuring no interference is leaked to the primary users. This is achieved by letting the secondary users to transmit signals through the null-space of the channels seen between secondary users and the primary network basestation. Subchannels are allocated based on the correlation coefficient of the left singular vector of the MIMO channels seen between various secondary users and the secondary network basestation. Once the secondary users are allocated in various subchannels, the radio resource allocation in terms of power and bit allocation is performed on a per user basis using an integer linear programming framework.

3.4.1 System Model and User scheduling

An uplink cognitive radio network with K secondary users is considered where each secondary user is equipped with N_t ($N_t > 1$) transmit antennas. The secondary network basestation in the cognitive radio network consists of N_r ($N_r > 1$) receiver antennas. Without appropriate preprocessing, the signals transmitted by the secondary users could reach the receivers of the primary network basestation. There are L primary network basestations in the primary network and each primary network basestation consists of N_p receiver antennas. It is assumed that $L \times N_p < N_t$ in order to annihilate secondary user interference. The frequency band is divided into N subchannels. The length of the cyclic prefix is assumed to be longer than the maximum time dispersion of the channels. The channel in each subchannel is therefore frequency nonselective. Denote the uplink channel matrix between the k^{th} secondary user and the secondary network basestation on the n^{th} subchannel as

$$\mathbf{H}_{\mathrm{s}}(k,n) = \begin{bmatrix} \alpha_{11}(k,n) & \alpha_{12}(k,n) & \dots & \alpha_{1Nt}(k,n) \\ \alpha_{21}(k,n) & \alpha_{22}(k,n) & \dots & \alpha_{2Nt}(k,n) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{Nr1}(k,n) & \alpha_{Nr1}(k,n) & \dots & \alpha_{NrNt}(k,n) \end{bmatrix} \in \mathcal{C}^{N_r \times N_t},$$

where $\alpha_{n_r n_t}(k,n)$ is the complex channel gain from the k^{th} secondary n_t^{th} transmit antenna to the secondary network basestation's n_r^{th} receive antenna. Define the channel matrix between secondary user k and the primary network basestation l on subchannel n as

$$\mathbf{H}_{\mathrm{p}}(l,k,n) = \begin{bmatrix} \gamma_{11}(l,k,n) & \gamma_{12}(l,k,n) & \ldots & \gamma_{1N_t}(l,k,n) \\ \gamma_{21}(l,k,n) & \gamma_{22}(l,k,n) & \ldots & \gamma_{2N_t}(l,k,n) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N_p1}(l,k,n) & \gamma_{N_p2}(l,k,n) & \ldots & \gamma_{N_pN_t}(l,k,n) \end{bmatrix} \in \mathcal{C}^{N_p \times N_t},$$

where $\gamma_{n_p n_t}(l,k,n)$ is the complex channel gain from the k^{th} secondary user's n_t^{th} transmit antenna to l^{th} primary receive antenna. Define the composed channel matrix between k^{th} secondary user and all primary network basestations on the n^{th} subchannel as

$$\mathbf{H}_{\mathrm{c}}(k,n) = [\mathbf{H}_{\mathrm{p}}(1,k,n)^T \ldots \mathbf{H}_{\mathrm{p}}(L,k,n)^T]^T \;\in\; \mathcal{C}^{(L\times N_p)\times N_t}. \tag{3.4.1}$$

The null-space of the composed channel matrix in (3.4.1) can be obtained to determine the preprocessing vector for the k^{th} secondary user on the n^{th} subchannel so that the interference leakage to primary network basestation is annihilated. Using SVD, $\mathbf{H}_{\mathrm{c}}(k,n)$ can be decomposed into

$$\begin{aligned} \mathbf{H}_{\mathrm{c}}(k,n) &= \mathbf{U}_{\mathrm{c}}(k,n)\Sigma_{\mathrm{c}}(k,n)\mathbf{V}_{\mathrm{c}}(k,n)^H \\ &= \mathbf{U}_{\mathrm{c}}(k,n)\Sigma_{\mathrm{c}}(k,n)[\mathbf{V}_{\mathrm{c,S}}(k,n)\ \mathbf{V}_{\mathrm{c,N}}(k,n)]^H \end{aligned} \tag{3.4.2}$$

where $\mathbf{U}_{\mathrm{c}}(k,n)$ and $\mathbf{V}_{\mathrm{c}}(k,n)$ are unitary matrices whose columns are the left and right singular vectors of $\mathbf{H}_{\mathrm{c}}(k,n)$. The matrix $\Sigma_{\mathrm{c}}(k,n)$ contains singular values of $\mathbf{H}_{\mathrm{c}}(k,n)$. The columns of $\mathbf{V}_{\mathrm{c,S}}(k,n)$ span the signal space and the columns of $\mathbf{V}_{\mathrm{c,N}}(k,n)$ span the null space. In order to ensure that the interference is not leaked to primary network basestation, the transmitter weight vector for the secondary user is chosen as a linear combination of that span the null space of $\mathbf{H}_{\mathrm{c}}(k,n)$. Therefore, the k^{th} secondary users' transmitter weight vector on n^{th} subchannel is written as

$$\mathbf{f}(k,n) = \mathbf{V}_{\mathrm{c,N}}(k,n)\mathbf{w}(k,n), \ \in \ \mathcal{C}^{N_t \times 1}, \tag{3.4.3}$$

where $\mathbf{w}(k,n)$ is the vector that linearly combines the orthogonal basis of the null-space $\mathbf{V}_{\mathrm{c,N}}(k,n)$. Denote the effective channel matrix between the k^{th} secondary user and the secondary network basestation as

$$\mathbf{H}_{\mathrm{e}}(k,n) = \mathbf{H}_{\mathrm{s}}(k,n)\mathbf{V}_{\mathrm{c,N}}(k,n) \ \in \ \mathcal{C}^{N_r \times [N_t - rank(\mathbf{H}_{\mathrm{c}}(k,n))]}. \tag{3.4.4}$$

Therefore the received signal component at the secondary network basestation due to the k^{th} secondary user on the n^{th} subchannel can be expressed as,

$$\begin{aligned}
\mathbf{z}(k,n) &= \mathbf{H}_{\mathrm{s}}(k,n)\mathbf{f}(k,n)\sqrt{p(k,n)}s(k,n) \\
&= \mathbf{H}_{\mathrm{s}}(k,n)\mathbf{V}_{\mathrm{c,N}}(k,n)\mathbf{w}(k,n)\sqrt{p(k,n)}s(k,n) \\
&= \mathbf{H}_{\mathrm{e}}(k,n)\mathbf{w}(k,n)\sqrt{p(k,n)}s(k,n)
\end{aligned} \tag{3.4.5}$$

where $s(k,n)$ is the transmitted data symbol, $\mathbf{z}(k,n) \ \in \ \mathcal{C}^{N_r \times 1}$ is the received signal vector at the secondary network basestation, and $p(k,n)$ is the allocated power for the k^{th} secondary user on the n^{th} subchannel. The power allocation method will be provided later in this section. The SVD of effective matrix $\mathbf{H}_{\mathrm{e}}(k,n)$ can be performed as

$$\mathbf{H}_{\mathrm{e}}(k,n) = \mathbf{U}_{\mathrm{e}}(k,n)\Sigma_{\mathrm{e}}(k,n)\mathbf{V}_{\mathrm{e}}(k,n)^H, \tag{3.4.6}$$

where $\mathbf{U}_{\mathrm{e}}(k,n) = [\mathbf{u}_{\mathrm{e}}(k,n,1)\ldots\mathbf{u}_{\mathrm{e}}(k,n,Nr)]$, and $\mathbf{V}_{\mathrm{e}}(k,n) = [\mathbf{v}_{\mathrm{e}}(k,n,1)$ $\ldots\mathbf{v}_{\mathrm{e}}(k,n,N_t - rank(\mathbf{H}_{\mathrm{c}}(k,n)))]$ are unitary matrices consisting of left and right singular vectors of $\mathbf{H}_{\mathrm{e}}(k,n)$. Denote the left and right singular vectors corresponding to the largest singular value of $\mathbf{H}_{\mathrm{e}}(k,n)$ by $\mathbf{u}_{\mathrm{e}}(k,n,1) \ \in \ \mathcal{C}^{N_r \times 1}$

$\mathbf{v}_{\mathrm{e}}(k,n,1) \in \mathcal{C}^{[N_t - rank(\mathbf{H}_{\mathrm{c}}(k,n))]\times 1}$ respectively. The matrix $\Sigma_{\mathrm{e}}(k,n) \in \mathcal{C}^{N_r \times [N_t - rank(\mathbf{H}_{\mathrm{c}}(k,n))]}$, contains the singular values of $\mathbf{H}_{\mathrm{e}}(k,n)$. When the channels are correlated (for example for a non-rich scattering environment with direct path of arrivals), the first singular value of $\mathbf{H}_{\mathrm{e}}(k,n)$ is likely to be much higher than the rest of the singular values [66]. A similar channel environment is considered for this work. In this case, communications through only the largest singular mode of the MIMO channel is considered. Therefore, the linear combination weight vector, $\mathbf{w}(k,n)$ at the transmitter, and the receiver weight vector, $\mathbf{q}(k,n)$ at the secondary network basestation are chosen as the right and left singular vectors corresponding to the largest singular value of $\mathbf{H}_{\mathrm{e}}(k,n)$, i.e. $\mathbf{w}(k,n) = \mathbf{v}_{\mathrm{e}}(k,n,1)$, $\quad \mathbf{q}(k,n) = \mathbf{u}_{\mathrm{e}}(k,n,1)$. Therefore the received signal due to the k^{th} secondary user in (3.4.5) becomes

$$\begin{aligned}\mathbf{z}(k,n) &= \mathbf{H}_{\mathrm{e}}(k,n)\mathbf{v}_{\mathrm{e}}(k,n,1)\sqrt{p(k,n)}s(k,n), \\ &= \mathbf{u}_{\mathrm{e}}(k,n,1)\lambda_{\mathrm{e}}(k,n,1)\sqrt{p(k,n)} \times s(k,n), \qquad (3.4.7)\end{aligned}$$

where $\lambda_{\mathrm{e}}(k,n,1)$ is the largest singular value of $\mathbf{H}_{\mathrm{e}}(k,n)$. The received signal at the secondary network basestation on the n^{th} subchannel, $\mathbf{r}_n$ can be written as,

$$\mathbf{r}_n = \sum_{i=1}^{K} \mathbf{z}(i,n) + \mathbf{n}_n, \qquad (3.4.8)$$

where $\mathbf{n}_n \in \mathcal{C}^{N_r \times 1}$ is the AWGN component with unity variance. The k^{th} secondary users' signal on the n^{th} subchannel is retrieved by using receiver weight vector, $\mathbf{q}(k,n)$, as

$$\begin{aligned}y(k,n) &= \mathbf{q}(k,n)^H \mathbf{r}_n, \\ &= \mathbf{u}_{\mathrm{e}}(k,n,1)^H \mathbf{z}(k,n) + \mathbf{u}_{\mathrm{e}}(k,n,1)^H \left(\sum_{i=1, i\neq k}^{K} \mathbf{z}(i,n) + \mathbf{n}_n \right), \\ &= \lambda_{\mathrm{e}}(k,n,1)\sqrt{p(k,n)}s(k,n) + \mathbf{IUI} + \mathbf{u}_{\mathrm{e}}(k,n,1)^H \mathbf{n}_n, \qquad (3.4.9)\end{aligned}$$

$$\text{where } \mathbf{IUI} = \mathbf{u}_{\mathrm{e}}(k,n,1)^H \sum_{i=1, i\neq k}^{\mathrm{K}} \mathbf{z}(i,n),$$
$$= \mathbf{u}_{\mathrm{e}}(k,n,1)^H \sum_{i=1, i\neq k}^{K} \mathbf{u}_{\mathrm{e}}(i,n,1)\lambda_{\mathrm{e}}(i,n,1)\sqrt{p(i,n)}s(i,n). \tag{3.4.10}$$

Define the spatial correlation between $k1^{th}$ secondary user and $k2^{th}$ secondary user on the n^{th} subchannel as,

$$\rho(k1,k2,n) = |\mathbf{u}_{\mathrm{e}}(k1,n,1)^H \mathbf{u}_{\mathrm{e}}(k2,n,1)|. \tag{3.4.11}$$

In this work, sub optimum approach proposed in [66] has been adopted to admit secondary users in each subchannel. The correlation of the spatial signature as in (3.4.11) is computed for all possible secondary users, and those secondary users that provide a correlation value smaller than a specific threshold ρ_{TH} is admitted. This will ensure secondary users with small inter user interference are admitted. Any remaining residual inter user interference can be mitigated using multi user detection (MUD) techniques based on for example maximum likelihood estimator, minimum mean square error (MMSE) estimator or parallel and serial interference cancelers. Once secondary users are allocated in each subchannel according to the above method, power allocation and bit loading are performed on a per user basis on those allocated subchannels for each admitted secondary user.

3.4.2 Power Allocation and Bit Loading

A margin adaptive resource allocation algorithm is proposed to allocate power and bits to each secondary user on their allocated subchannels. The proposed scheme performs modulation scheme selection based on bit allocation and the corresponding power allocation. Then the modulation indexes the secondary users by the secondary network basestation through a separate control channel. It is assumed that the transmission over control channel is error-free and all subchannels are in slow fading. In this problem formulation, BPSK, 4-QAM, 16-QAM, 64-QAM and 256-QAM modulation schemes are considered. Assuming zero inter user interference, the minimum signal power required for the k^{th} secondary user to transmit b number of bits through n^{th} subchannel to achieve a given $\text{BER}_{\text{Target}}$ at the secondary network basestation is given by [66]

$$p(k,n) = \frac{\sigma^2}{[\lambda_{\text{e}}(k,n,1)]^2} \ln\left[\frac{1}{5BER_{\text{Target}}}\right]\frac{2^b-1}{1.5}. \tag{3.4.12}$$

Denote the required data rate and the set of subchannels allocated for the k^{th} secondary user as r_k and $\mathcal{N}_k$ respectively. Using these definitions, margin adaptive resource allocation problem for the k^{th} secondary user is given by

$$\min_{c(k,n)\in\{0,1,2,4,6,8\}} \quad \sum_{n\in\mathcal{N}_k} p(k,n), \tag{3.4.13}$$

$$\sum_{n\in\mathcal{N}_k} c(k,n) \geq r_k, \tag{3.4.14}$$

The objective function in (3.4.13) minimizes the transmission power over all subchannels allocated to the k^{th} where $c(k,n)$ denotes the number of bits allocated for the k^{th} secondary user on the n^{th} subchannel. The constraint (3.4.14) ensures that the k^{th} secondary user achieves a minimum data rate. Since the solution to the bit allocation problem is integer valued, this problem is non-convex [33] and the optimal solution can be obtained using integer linear programming technique [29].

3.4.3 Integer Linear Programming Problem Formulation

First, define a binary indicator vector $\mathbf{x}(k)$ as follows:

$$\mathbf{x}_k = [\mathbf{x}(k,1)^T \ldots \mathbf{x}(k, \mid \mathcal{N}_k \mid)^T]^T \ \in \ \{0,1\}^{|\mathcal{N}_k|\text{C}\times 1}, \tag{3.4.15}$$

where $\mathbf{x}(k,n) = [x(k,n,1)\ldots\ldots x(k,n,C)]^T \in \{0,1\}^{\mathrm{C}\times 1}$ where $x(k,n,c)$ equal to one means that the k^{th} secondary user transmits c number of bits per symbol on the n^{th} subchannel. All values of $p(k,n)$ for each possible $c(k,n)$ values are stored in the power vector $\mathbf{p}(k)$ as follows,

$$\mathbf{p}(k) = [\mathbf{p}(k,1)^T \ldots \mathbf{p}(k, \mid \mathcal{N}_k \mid)^T]^T \in \mathcal{R}^{|\mathcal{N}_k|\mathrm{C}\times 1}, \tag{3.4.16}$$

where $\mathbf{p}(k,n) = [p(k,n,1)\ldots p(k,n,C)]^T \in \mathcal{R}^{\mathrm{C}\times 1}$ where $p(k,n,c)$ is the required transmission power for k^{th} secondary user to transmit c number of bits over n^{th} subchannel. Denote $\mathbf{a} = [1\ 2\ 4\ 6\ 8]^T \in Z^{5\times 1}, \mathbf{b} = [\mathbf{a}^T\ \mathbf{a}^T..\ \mathbf{a}^T]^T \in Z^{|\mathcal{N}_k|\mathrm{C}\times 1}$ and $\mathbf{A} = \begin{bmatrix} \mathbf{1}_\mathrm{C}^\mathrm{T} & \mathbf{0}_\mathrm{C}^\mathrm{T} & \ldots & \mathbf{0}_\mathrm{C}^\mathrm{T} \\ \mathbf{0}_\mathrm{C}^\mathrm{T} & \mathbf{1}_\mathrm{C}^\mathrm{T} & \ldots & \mathbf{0}_\mathrm{C}^\mathrm{T} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_\mathrm{C}^\mathrm{T} & \mathbf{0}_\mathrm{C}^\mathrm{T} & \ldots & \mathbf{1}_\mathrm{C}^\mathrm{T} \end{bmatrix} \in \{0,1\}^{|\mathcal{N}_k|\times\mathrm{C}}$.

Using these definitions, the original problem in (3.4.13)-(3.4.14) can now be formulated into integer linear programming as follows [26]

$$\begin{aligned} \min_{\mathbf{x}(k)} \quad & \mathbf{p}(k)^T\mathbf{x}(k), \\ \text{s.t}: \quad & \mathbf{b}^T\mathbf{x}(k) \geq r_k, \\ & \mathbf{0}_{|\mathcal{N}_k|} \leq \mathbf{A}\mathbf{x}(k) \leq \mathbf{1}_{|\mathcal{N}_k|}, \\ & x_i(k) \in \{0,1\}^{|\mathcal{N}_k|\mathrm{C}}. \end{aligned} \tag{3.4.17}$$

The above integer linear programming problem can be efficiently solved using branch and bound method [29].

3.5 Simulation Results

In order to validate the proposed algorithm and assess the performance, a MIMO-OFDMA based cognitive radio network with $K = 25$ secondary users and $N = 64$ subchannels is considered. Each secondary user has $N_t = 3$

transmit antennas and the secondary network basestation consists of $N_r = 4$ receiver antennas. Only one primary network basestation with single receiver antenna ($N_p = 1, L \times N_p = 1$) is considered. The data rate requirement for each secondary user has been set to 16 bits/user. 50 randomly generated channel-pairs $\{\mathbf{H}_{\mathrm{s}}(k,n), \mathbf{H}_{\mathrm{p}}(l,k,n)\}$ have been used for the simulation study.

Average number of secondary users allocated for a particular subchannel for various ρ_{TH} values ($\rho_{\mathrm{TH}} = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 0.9, 1$) is shown in Fig. 3.6. From this figure, it can be observed that the average number of secondary users allocated in each subchannel is increasing with the spatial correlation threshold. Please note that for larger values of ρ_{TH}, each subchannel will tend to get all secondary users in the spatial correlation based user scheduling stage. However, once resource allocation is performed as in (3.4.17), some users will be dropped as soon as they attain their target data rates. This is the reason why the average number of users allocated in each subchannel drops after $\rho_{\mathrm{TH}} = 0.7$ in Fig. 3.6.

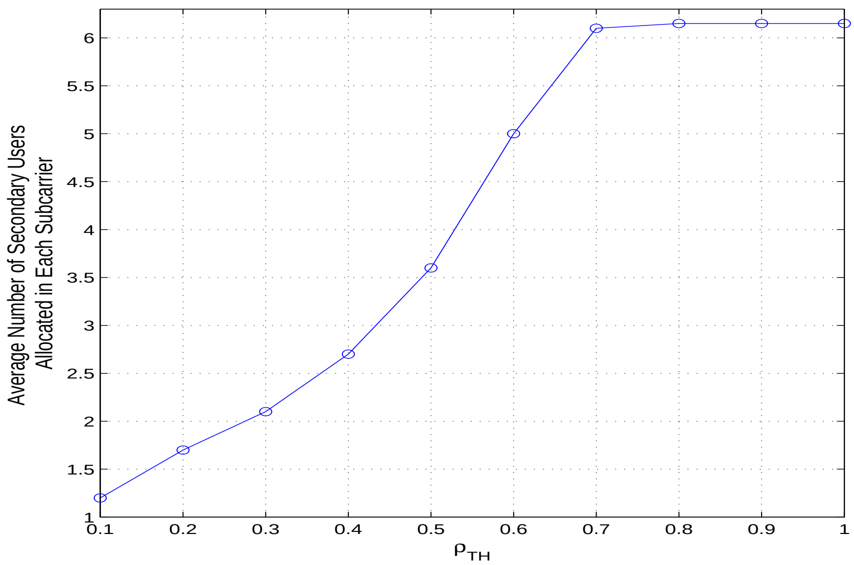

Figure 3.6. Performance comparison of average number of secondary users allocated in each subchannel for different spatial correlation values, ρ_{th}

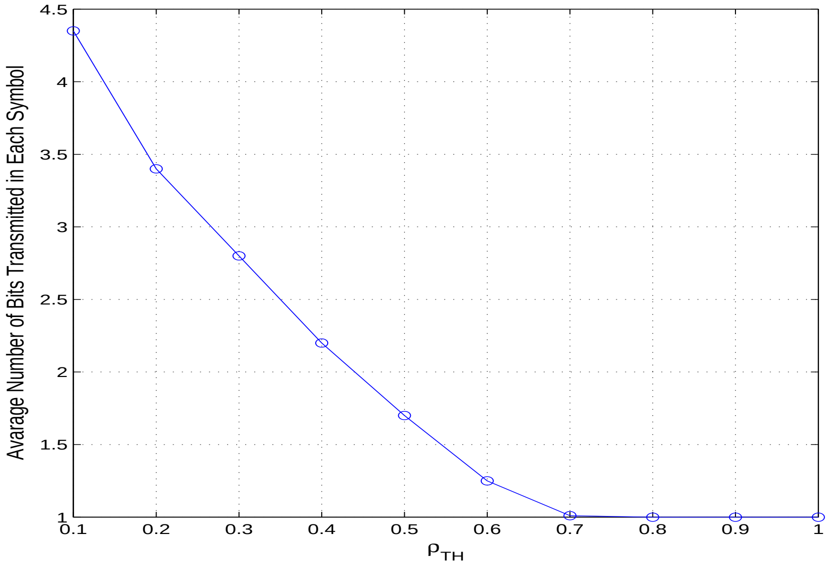

Figure 3.7. Performance comparison of average number of bits allocated in each symbol transmitted for different spatial correlation values, $\rho_{\rm TH}$

Fig. 3.7 compares the average number of bits per symbol transmitted for different values of $\rho_{\rm TH}$. When the value of $\rho_{\rm TH}$ is low, higher number of bits are transmitted per symbol in order to satisfy the data rate requirement in (3.4.17). This is because for low $\rho_{\rm TH}$, the number of subchannels allocated to each secondary users would be small.

When the value of $\rho_{\rm TH}$, is very small, the number of subchannels allocated to each secondary user in the spatial correlation based user scheduling stage is reduced as in Fig. 3.6. Therefore, during the second optimization stage as in (3.4.17), there are possibilities that some users may not get their target data rate and the problem in (3.4.17) might turn out to be infeasible. The outage probability that the problem is infeasible has been computed and depicted in Fig. 3.8.

The results in Fig. 3.6 - Fig. 3.8 demonstrate that the spatial correlation value $\rho_{\rm TH}$ should not be chosen too small as the optimization problem might turn out be infeasible. On the other hand, $\rho_{\rm TH}$ should not be chosen

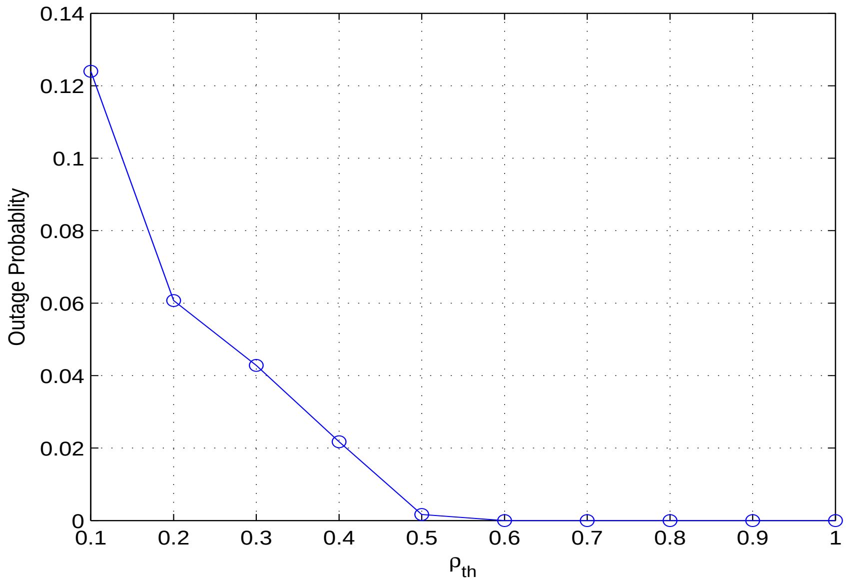

Figure 3.8. The outage probability that the problem in (16) is infeasible for different spatial correlation values, ρ_{TH}

too high as this will introduce inter user interference and the model used in (3.4.12) will turn out be invalid. Therefore, selection of an optimum ρ_{TH} is challenging, and it will depend on the channel scenario such as number of secondary users and the number of available subchannels. It can be determined using a MAC level simulation considering error control coding and assessing the total data throughput for various values of ρ_{TH}.

3.6 Conclusion

In this chapter, first, a resource allocation technique has been proposed for an OFDMA based cognitive radio network in the downlink and then it was extended to a MIMO-OFDMA based cognitive radio network. The proposed algorithms optimally assign subchannels, spatial beams, power and bits for the secondary users based on rate adaptive optimization. The rate adaptive optimization maximizes the total data throughput under interference power constraint to the primary users, individual data rate constraints for the secondary users, and the total transmission power constraint at the secondary network basestation. Finally, resource allocation problem is analyzed for a MIMO-OFDMA cognitive radio network in the uplink. Based on spatial separation between secondary users, various secondary users were allocated in each subchannel. A margin adaptive optimization was considered to allocate required data rate to each secondary user on the allocated subchannels while minimizing the total transmission power. All of these algorithms were formulated using an integer linear programming framework and branch and bound method has been used to determine the optimal solutions. Simulation results have been provided to validate the performance of the proposed algorithms.

Chapter 4

SUBOPTIMAL RESOURCE ALLOCATION TECHNIQUES FOR OFDMA BASED COGNITIVE RADIO NETWORKS

In this chapter, low complexity resource allocation algorithms are proposed for an OFDMA based cognitive radio networks. Since the subchannels and bits are discrete, the multiuser resource allocation problem becomes combinatorial optimization problem. Determining an optimal solution of a combinatorial optimization problem is computationally expensive. Hence, suboptimal low complexity algorithms are proposed for an OFDMA based cognitive radio network in order to reduce the computational complexity. The proposed algorithms allocate subchannels, bits and power to various secondary users while ensuring the interference leakage to the primary users below a threshold. Simulation results and complexity analysis are provided to validate the performance of the algorithms.

4.1 Introduction

Optimal resource allocation algorithms based on integer linear programming have been studied in Chapter 3. In this chapter, suboptimal resource allocation algorithms are proposed in order to reduce the computational complexity. There are two different low complexity algorithms proposed based on rate adaptive optimization technique and data rate balancing technique. The rate adaptive optimization technique maximizes the total data throughput subject to individual data rate constraints to the secondary users, interference leakage constraints to the primary users and total transmission power constraint at the secondary network basestation. This technique has been studied for a conventional OFDMA based wireless networks in [20,26,27,31,85–87] and these works cannot be extended directly into cognitive radio network due to the additional interference leakage constraints to primary users. In [20], a multiuser combinatorial resource allocation problem has been solved using a suboptimal greedy algorithm. In order to find the optimal solution for this combinatorial optimization problem, an integer linear programming framework has been used in [26]. As the problem dimension (for example the number of subchannels and users) grows, the complexity of the integer linear programming algorithm also increases rapidly. In order to overcome this complexity issue, an integer relaxation based suboptimal algorithm has been proposed in [31] for a conventional wireless network which determines the best subchannels for each user by relaxing the integer constraint and then performs adaptive bit loading and power allocation for each user separately using a greedy algorithm. Under the transmission power constraint, the greedy algorithm allocates one bit at a time to various subchannels in the order of least power consumption. It should be noted that the greedy algorithm has limitations in a cognitive radio network scenario due to multiple constraints. Hence, a novel recursive algorithm is proposed for a cognitive radio network in order to satisfy individual data rates constraints for the secondary users, interference leakage constraints for the primary users and total transmission power constraint.

However, the problem based on rate adaptive optimization technique might turn out to be an infeasible problem. Because, it is not possible to satisfy the required data rate to the users all the time for some channel realizations. Hence, the rate balancing technique has been adopted in a conventional network to avoid the infeasibility issue [36, 38, 88, 89]. The rate balancing technique has been used to maintain the fairness among the users. The proposed rate balancing algorithm maximizes the data throughput while maintaining equal data rate for all the secondary users subject to interference leakage constraints to the primary users and total transmission power constraint at the secondary network basestation. The next section describes the proposed rate adaptive based suboptimal algorithm.

4.2 A Rate Adaptive Technique based Suboptimal Resource Allocation Algorithm

A suboptimal algorithm for adaptive subchannel, bit and power allocation for OFDMA based cognitive radio network is proposed. This problem in its original form (i.e., (3.2.13)-(3.2.18)) is non-convex and may be solved using greedy algorithms or integer linear programming techniques. However, the computational complexity of the latter techniques is quite high, while the suboptimal greedy algorithms are not very well suited for cognitive radio network due to multiple constraints on the transmission power, interference leakage and individual user data rate. Therefore, a novel recursion-based linear optimization framework is proposed to provides a solution that is very close to the optimal one and that has the ability to perform adaptive subchannel, bit and power allocation for multiple users in the presence of multiple constraints. Due to the convexity of the proposed algorithm at each recursion, its overall computational complexity is substantially lower than that of the integer linear programming based solution.

4.2.1 A Novel Recursive Based Linear Optimization Framework

In Section 3.2, an optimal rate adaptive algorithms is proposed for SISO-OFDMA based cognitive radio network. Same system model and problem statement as in Section 3.2 are considered here. To reduce the computational complexity of integer linear programming based optimal method in (3.2.13)-(3.2.18), an integer relaxation approach has been used here. The proposed technique has two important features. First, it involves an integer relaxation step similar to that used in [31] to determine subchannels for various users. Second, formulation of the resource allocation problem to determine number of bits and power for various subchannels is recursive. The essence of the proposed integer relaxation is that, instead of forcing x_i in (3.2.18) to be equal to an element of the set {0, 1}, it is relaxed to take any real value between 0 and 1. Hence, the original problem in (3.2.13)-(3.2.18) can be modified as follows:

$$\max_{\mathbf{x}} \quad \mathbf{b}^T\mathbf{x} \tag{4.2.1}$$

$$\text{s.t.} \quad \mathbf{H}^{\mathrm{p}}[\mathbf{A}(\mathbf{p} \odot \mathbf{x})] \leq \boldsymbol{\Upsilon}, \tag{4.2.2}$$

$$\mathbf{p}^T\mathbf{x} \leq P, \tag{4.2.3}$$

$$\mathbf{B}\mathbf{x} \geq \mathbf{r}, \tag{4.2.4}$$

$$\mathbf{0}_N \leq \mathbf{A}\mathbf{x} \leq \mathbf{1}_N, \tag{4.2.5}$$

$$\mathbf{0}_{KNC} \leq \mathbf{x} \leq \mathbf{1}_{KNC}, \tag{4.2.6}$$

so that the relaxed problem is convex [33]. The subchannels are chosen by using the relaxed optimization framework outlined in (4.2.1)-(4.2.6). Referring

1. Solve for $\mathbf{x}$ using (4.2.1)-(4.2.6).
2. Set subchannel index $n = 0$.
3. **repeat**
4. $n \leftarrow n + 1$
5. **if** $\mathbf{b}(j,n)^T\mathbf{x}(j,n) \ > \ \mathbf{b}(k,n)^T\mathbf{x}(k,n), \forall k \neq j$
6. n^{th} subchannel is allocated to the j^{th} secondary user
7. **endif**
8. **until** $n < N + 1$

Table 4.1. Pseudo code for the subchannel allocation algorithm

(3.4.16), $\mathbf{x}(n) = [\mathbf{x}(1,n)^T \ \ldots \ \mathbf{x}(K,n)^T]^T$ corresponds to the n^{th} subchannel. An entry equal to one in $\mathbf{x}(k,n) = [x(k,n,1) \ x(k,n,2) \ x(k,n,3) \ x(k,n,4)]^T$ means that the k^{th} secondary user is assigned to the n^{th} subchannel. For example, if the second entry of $\mathbf{x}(k,n)$ is equal to one and all the other entries in $\mathbf{x}(n)$ are zero, this means that the n^{th} subchannel is assigned to the k^{th} secondary user with two bits. However, due to integer relaxation, the entries of $\mathbf{x}(k,n)$ will not be exactly 0 or 1. Hence, $\mathbf{b}(k,n)^T\mathbf{x}(k,n)$ has been considered as the data throughput for the k^{th} secondary user at the n^{th} subchannel. Therefore, the n^{th} subchannel is assigned to the j^{th} secondary user if $\mathbf{b}(j,n)^T\mathbf{x}(j,n) \ > \ \mathbf{b}(k,n)^T\mathbf{x}(k,n), \forall k \neq j$. The pseudo code for this subchannel allocation algorithm is provided in the Table 4.1.

A recursive optimization loop will be started once the subchannels are allocated to each secondary user as described below. (Please note that once the subchannels are allocated to various users, the dimension of $\mathbf{x}$ can be reduced to $\mathbf{x} \ \in \ \{0,1\}^{NC\times 1}$. Also, during the recursion as explained below, the dimension of $\mathbf{x}$ can be reduced further. However, for the simplicity of explanation of the algorithm, size of $\mathbf{x}$ is unchanged. Instead, zero values are placed at appropriate locations of vector $\mathbf{b}$.) Referring to (3.2.10),

$\mathbf{b}(n) = [\mathbf{b}(1,n)^T \ \dots \ \mathbf{b}(K,n)^T]^T$ is a vector of size $KC \times 1$ with the entries 1, 2, 4, and 6. Suppose that, according to the subchannel allocation above, the first subchannel is allocated to the second secondary user. Then all the entries of $\mathbf{b}(1)$ will be set to zeros except the entries in $\mathbf{b}(2,1)$, i.e., $\mathbf{b}(1) = [0\ 0\ 0\ 0\ 1\ 2\ 4\ 6\ 0\ 0\ 0\ 0\ \dots\ 0\ 0\ 0\ 0\ ]^T$. Similarly, supposing that the second subchannel is allocated to the first secondary user, set $\mathbf{b}(2) = [1\ 2\ 4\ 6\ 0\ 0\ 0\ 0\ \dots\ 0\ 0\ 0\ 0\ ]^T$ and so on. Once certain elements of $\mathbf{b}$ become zeros, according to the subchannel allocation, the vector $\mathbf{b}$ is renamed as $\mathbf{b}^{(1)}$. The subsequent recursion allocates bits and power for the secondary users according to the assigned subchannels. The optimization problem for the first recursion can be written as

$$\max_{\mathbf{x}^{(1)}} \quad {\mathbf{b}^{(1)}}^T \mathbf{x}^{(1)} \tag{4.2.7}$$

$$\text{s.t.} \quad \mathbf{H}^{\mathrm{p}}\{\mathbf{A}[\mathbf{p} \odot \mathbf{x}^{(1)}]\} \leq \boldsymbol{\Upsilon}, \tag{4.2.8}$$

$$\mathbf{B}\mathbf{x}^{(1)} \geq \mathbf{r}, \tag{4.2.9}$$

$$\mathbf{p}^T\mathbf{x}^{(1)} \leq P_{\max}, \tag{4.2.10}$$

$$\mathbf{0}_N \leq \mathbf{A}\mathbf{x}^{(1)} \leq \mathbf{1}_N, \tag{4.2.11}$$

$$\mathbf{0}_{KNC} \leq \mathbf{x}^{(1)} \leq \mathbf{1}_{KNC}. \tag{4.2.12}$$

After the first recursion, the vector $\mathbf{x}^{(1)}$ is used along with the power vector $\mathbf{p}$ to determine the modulation scheme (number of bits) for each secondary user at various subchannels. As considered in the example before, suppose that the first subchannel is allocated to the second secondary user, and all entries of $\mathbf{x}^{(1)}(1)$ (where $\mathbf{x}^{(1)}(1)$ denotes the vector $\mathbf{x}(1)$ in the first recursion step) are zero except the elements in $\mathbf{x}^{(1)}(2,1)$. The total power allocated to the first subchannel can be computed as $\mathbf{p}(2,1)^T\mathbf{x}^{(1)}(2,1)$. In general, if the n^{th} subchannel is allocated to the j^{th} secondary user, the total power allocated to the n^{th} subchannel can be computed as $\mathbf{p}(j,n)^T\mathbf{x}^{(1)}(j,n)$. The modulation scheme l (i.e., with d_l bits) that can be used without exceeding the power $\mathbf{p}(j,n)^T\mathbf{x}^{(1)}(j,n)$ can be obtained as

$$l = \underset{l}{\operatorname{argmax}}\{l \in \{0,\ 1,\ 2,\ 3,\ 4\} : p(j,n,l)\ \leq\ \mathbf{p}(j,n)^T\mathbf{x}^{(1)}(j,n)\} \quad (4.2.13)$$

i.e., as the largest possible modulation order (in terms of the number of bits) with a power not exceeding $\mathbf{p}(j,n)^T\mathbf{x}^{(1)}(j,n)$. Due to the finite modulation size and finite set of power levels $p(j,n,l)$, once the bits are determined using $\mathbf{x}^{(1)}$ and $\mathbf{p}$, the total power consumed so far will still be less than $P_{\max}$. The interference leakage will also be less than $\boldsymbol{\Upsilon}$. Hence further recursions will be performed to use any residual power left to allocate more bits. In general, the following optimization problem has to be solved in the q^{th} recursion step

$$\max_{\mathbf{x}^{(q)}} \quad \mathbf{b}^{(q)^T}\mathbf{x}^{(q)} \quad (4.2.14)$$

$$\text{s.t.} \quad \mathbf{H}^{\mathrm{p}}[\mathbf{A}(\mathbf{p}^{(q-1)^T} \odot \mathbf{x}^{(q)})] \leq \boldsymbol{\Upsilon} - \mathbf{H}^{\mathrm{p}}\mathbf{v}^{(q-1)}, \quad (4.2.15)$$

$$\mathbf{B}\mathbf{x}^{(q)} \geq [\mathbf{r} - \mathbf{f}^{(q-1)}]^+, \quad (4.2.16)$$

$$\mathbf{p}^{(q-1)^T}\mathbf{x}^{(q)} \leq P_{\max} - \|\mathbf{v}^{(q-1)}\|_1, \quad (4.2.17)$$

$$\mathbf{0}_N \leq \mathbf{B}\mathbf{x}^{(q)} \leq \mathbf{1}_N, \quad (4.2.18)$$

$$\mathbf{0}_{KNC} \leq \mathbf{x}^{(q)} \leq \mathbf{1}_{KNC}. \quad (4.2.19)$$

The above optimization problem for the case of $q = 2$, i.e., in the second recursion step will be explained. Suppose that in the first recursion step, the first subchannel is assigned to the second secondary user using two bits, then $\mathbf{b}^{(1)}(2,1)$ can be modified as $\mathbf{b}^{(2)}(2,1) = [0\ 0\ (4-2)\ (6-2)]^T = [0\ 0\ 2\ 4]^T$. This modification means that two bits have already been allocated to the second secondary user at the first subchannel and the residual power is now used to upgrade it to four or six bits by loading two or four more bits respectively. In the first recursion step, the power $p(2,1,2)$ has been allocated to the second secondary user at the first subchannel. Hence, in order to allocate two or four more bits, only $(p(2,1,3) -$

$p(2,1,2))$ or $(p(2,1,4) - p(2,1,2))$ will be required respectively. Hence, the new power vector $\mathbf{p}^{(1)}(2,1)$ is determined from $\mathbf{p}(2,1)$ as $\mathbf{p}^{(1)}(2,1) = [p(2,1,1)\ p(2,1,2)\ (p(2,1,3) - p(2,1,2))\ (p(2,1,4) - p(2,1,2))]^T$. The vector $\mathbf{p}^{(1)}$ is determined accordingly. Let $v^{(1)}(n)$ denote the power that was allocated (discrete level due to discrete modulation) to the n^{th} subchannel in the first recursion step. Also define $\mathbf{v}^{(1)} \triangleq [v^{(1)}(1)\ \ldots\ v^{(1)}(N)]^T$. Hence $P_{\max} - \sum_{n=1}^{N} v^{(1)}(n) = P_{\max} - \|\mathbf{v}^{(1)}\|_1$ is the residual power that is available for the second recursion step. Also, the interference leakage due to the power allocated in the first recursion step is $\mathbf{H}^{\mathrm{p}}\mathbf{v}^{(1)}$ so that the remaining interference should be less than $\boldsymbol{\Upsilon} - \mathbf{H}^{\mathrm{p}}\mathbf{v}^{(1)}$ in the second recursion step. Let $f^{(1)}(k)$ denote the data rate already allocated for the k^{th} secondary user in the first recursion step. Also define $\mathbf{f}^{(1)} = [f^{(1)}(1)\ \ldots\ f^{(K)}(1)]^T$. Hence, the data rate requirement in the second recursion step can be written as $\mathbf{B}\mathbf{x}^{(2)} \geq [\mathbf{r} - \mathbf{f}^{(1)}]^+$. After the second recursive optimization, the amount of power allocated to the n^{th} subchannel can written as

$$S^{(2)}(n) = {\mathbf{p}^{(1)}}^T(j,n)\mathbf{x}^{(2)}(j,n) + v^{(1)}(n). \qquad (4.2.20)$$

The n^{th} subchannel is upgraded to the l^{th} modulation scheme as

$$l = \underset{l}{\operatorname{argmax}}\{l \in \{0,\ 1,\ 2,\ 3,\ 4\}\ :\ p(j,n,l)\ \leq\ S^{(2)}(n)\}. \qquad (4.2.21)$$

The recursion is repeated until no improvement on the sum throughput is observed. After the q^{th} recursion, the vectors $\mathbf{f}^{(q+1)}$ and $\mathbf{v}^{(q+1)}$ will contain the allocated bits and powers for each subchannel. The pseudo code this recursive bit and power allocations is provided in the Table 4.2.

4.2.2 Computational Complexity Analysis

Complexity of the proposed recursive algorithm is compared with complexity of the integer linear programming problem based optimal algorithm in this

1. Set $n = 0$, $q = 0$, $\mathbf{v}(0) = \mathbf{0}_N$ and $\mathbf{p}^{(0)} = \mathbf{p}$
2. **repeat**
3. $q \leftarrow q + 1$
4. Set $\mathbf{f}^{(q)} = \mathbf{0}_K, \mathbf{S}^{(q)} = \mathbf{0}_N$
5. Solve the problem in (4.2.14) - (4.2.19)
6. **repeat**
7. $n \leftarrow n + 1$
8. $S^{(q)}(n) = \mathbf{p}^{(q-1)}(j,n)^T \mathbf{x}^{(q)}(j,n) + v^{(q-1)}(n)$
9. **if** $l = \arg\max_l \{l \in \{0,\ 1,\ 2,\ 3,\ 4\} : p_{j,n,l} \ \leq\ S^{(q)}(n)\}$ **then**
10. Use modulation scheme l (i.e., with d_l bits) on n^{th} subchannel
11. Set $v^{(q)}(j) = p_{j,n,l}$
12. $f^{(q)}(j) = f^{(q)}(j) + d_l$
13. Set $p^{(q)}(j,n,z) = p(j,n,z) - p(j,n,l), \forall z > l$
14. Set $b^{(q+1)}(j,n,z) = b(j,n,z) - d_l, \forall z > l$
15. Set $b^{(q+1)}(j,n,z) = 0, \forall z \leq l$
16. **endif**
17. **until** $n < N + 1$
18. **until** no improvement on the sum throughput
19. The vectors $\mathbf{f}^{(q+1)}$ and $\mathbf{v}^{(q+1)}$ contain the allocated bits and powers for each subchannel

Table 4.2. Pseudo code for the proposed recursive power and bit allocation algorithm

subsection. The size of the optimization variable $\mathbf{x}$ and the number of linear constraints in (3.2.13)-(3.2.18) are KNC and $K+2N+L+1$ respectively. The most popular method for solving integer linear programming is the branch and bound method [29]. If integer linear programming is composed of x number of variables and y number of constraints, the minimum number of linear programming subproblems required to be solved by branch and bound is $\left(\sqrt{2}\right)^x$ [35]. The number of iterations needed to solve one linear program with x number of variables and y number of constraints is approximately $2(x+y)$. Each iteration requires $(xy-y)$ number of multiplications and $(xy-y)$ number of summations [35]. Based on this, the number of operations needed for the integer linear programming-based optimal algorithm is listed in Table 4.3 [35].

The proposed algorithm involves two parts. The first part of the problem (4.2.1)-(4.2.6) is used to allocate subchannels to each secondary users. After the integer relaxation, only one linear program with KNC variables and $K+2N+L+1$ constraints needs to be solved. Once the solution is obtained, NK number of multiplications are required to allocate subchannels to each user. The second part of the algorithm allocates bits and powers on the preassigned subchannels. In each recursion step, the problem (4.2.14)-(4.2.19) solves only one linear program. It should be stressed here that, the optimal integer linear programming requires solving $\left(\sqrt{2}\right)^{KNC}$ number of linear programming problems. From the simulations, it is observed that the average number of recursions required to converge is four. The size of the variable and the number of constraints in the first recursion is NC and $K+2N+L+1$, respectively. Each recursion needs additional N multiplications and $3N+2NC$ summations to determine bits and power allocations for each subchannel. The number of operations needed for the proposed algorithm is listed in Table 4.3. The total complexity of the proposed algorithm is the sum of the complexities of parts 1 and 2. For instance, consider a

Problem	Complexity
Optimal Method	$\mathcal{O}\{2(KNC+K+2N+L+1)\left(\sqrt{2}\right)^{KNC} \times [2KNC(K+2N+L+1)]\}$
Proposed Algorithm: Part 1	$\mathcal{O}\{2(KNC+K+2N+L+1)1\times [2KNC(K+2N+L+1)+NK]\}$
Proposed Algorithm: Part 2	$\mathcal{O}\{2(NC+K+2N+L+1)4\times [2NC(K+2N+L+1)+(4N+2NC)]\}$

Table 4.3. Complexity comparision

network with $K = 2$, $N = 4$, $L = 2$ and $C = 4$. The complexity of integer linear programming is $\mathcal{O}(5 \times 10^9)$ and the complexities of parts 1 and 2 of the proposed algorithm are $\mathcal{O}(8 \times 10^4)$ and $\mathcal{O}(1 \times 10^5)$, respectively. Hence the overall complexity of the proposed algorithm is substantially lower than that of the optimal integer linear programming approach.

4.2.3 Simulation Results

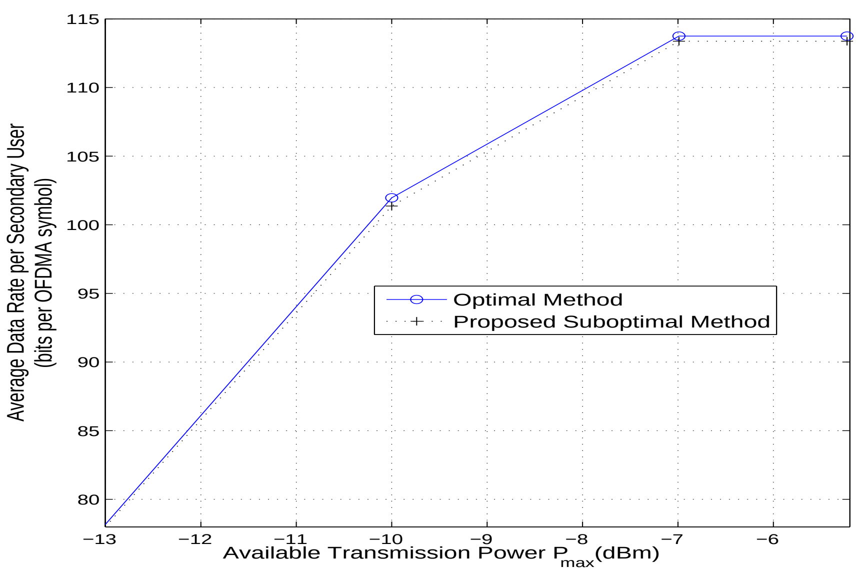

Figure 4.1. Total average number of bits per OFDMA symbol as a function of available transmitted power at the secondary network basestation.

To validate the proposed algorithm and to assess its performance, a cog-

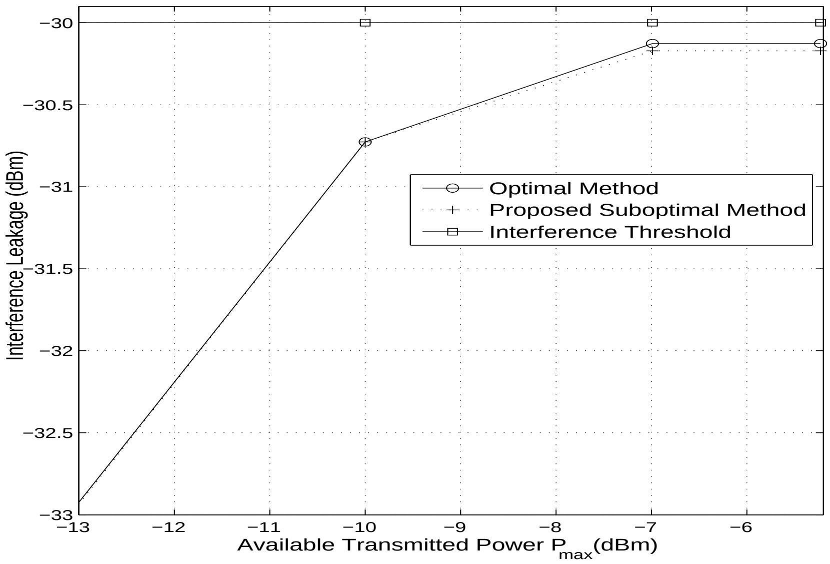

Figure 4.2. Interference to primary user per OFDMA symbol as a function of available transmitted power at the secondary network basestation.

nitive radio network with $K = 2$ secondary users, $L = 2$ primary users and $N = 64$ subchannels is considered. The minimum data rate requirement for each secondary user has been set to 16 bits/user. The BER requirement for the secondary users has been set to $\rho = 0.1$[1]. The interference leakage to primary user is required to be less than 0.001 mW. Multipath channels of length six were generated for primary users and secondary users using statistically independent Gaussian random variables. The average channel gain between the secondary network basestation and primary users is equal to 0.1 while such a gain between the secondary network basestation and secondary users is equal to 1. The interference caused by the primary user transmission is considered as noise at the secondary user receiver and the noise power spectral density at the secondary user is (0.01/64) mW/subchannel. All the simulation results are generated using 100 randomly generated channel-pairs

[1]From (3.2.2), transmission power increases when BER value decreases. Higher transmission power at the secondary network basestation causes higher interference leakage to the primary users. Hence, rate adaptive problem with smaller BER value might turn out to be an infeasible problem due to insufficient resources.

$(\mathbf{H}^{\mathrm{s}}, \mathbf{H}^{\mathrm{p}})$.

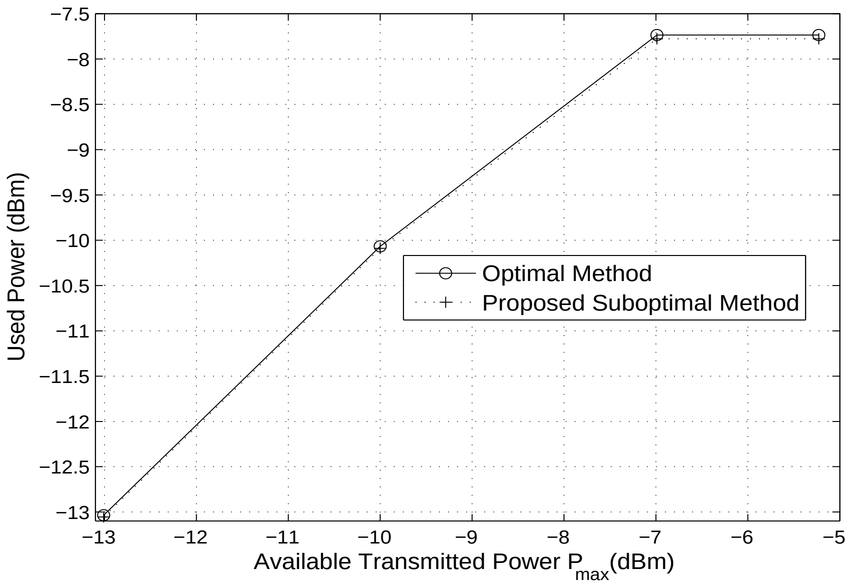

Figure 4.3. Power used by the secondary network basestation versus the available transmitted power at the secondary network basestation.

In Fig. 4.1, the average total throughput versus the available transmission power P is displayed for both the integer linear programming based optimal scheme and the proposed recursive optimization based scheme. As seen in Fig. 4.1, the throughput achieved by the proposed suboptimum method is very close to the optimum integer linear programming technique. The average value of the interference leakage versus P is depicted in Fig. 4.2. Again, it can be observed from Fig. 4.2 that the proposed method performs very close to the integer linear programming approach. It is apparent from Fig. 4.2 that when P is high, the total power may not be fully utilized as the optimization in this case is limited by the interference leakage constraints. The used power versus the available transmitted power is depicted in Fig. 4.3. Note that when P is equal to -13dBm and -10 dBm, the secondary network basestation fully utilizes the available transmission power.

Summarizing, in all the examples the proposed method performs very close to the optimum integer linear programming approach. The average number of recursions observed from the simulation is equal to four. Since at each recursion the problem is convex (linear), the complexity of the proposed scheme is substantially lower than that of the integer linear programming based scheme. As a guidance, our comparison of the CPU times in the simulation examples indicates two orders of savings in the computational complexity.

4.3 Rate Balancing Techniques in Cognitive Radio Networks

A cognitive radio network that employs OFDMA technique has been considered. An algorithm is proposed to allocate bits and power to the secondary users in various subchannels in order to ensure that each secondary user attains the same data rate while ensuring the interference leakage to the primary users is below a threshold. Since, the complexity of the integer linear programming based optimal algorithm is very high, a suboptimal method based on greedy algorithm has been proposed. Recently, a modified greedy algorithm known as greedy Max-Min algorithm has been proposed in [72] and [71] to maximize the data throughput of a single user cognitive radio network. In this section the greedy Max-Min algorithm is extended to a multiuser cognitive radio network. The extended greedy Max-Min algorithm allocates bits and power to the secondary users in a cyclic manner to ensure that each secondary user gets equal data rate.

4.3.1 System Model and Problem Statement

A downlink underlay cognitive radio network is considered with K secondary users and L primary users. All the secondary users are served by the secondary network base station. It is assumed that an OFDMA scheme with N subchannels is employed in the secondary network basestation. The primary users are served by their own base station and their transmission is not necessarily based on OFDMA. The channel gains between the secondary network basestation and the secondary users as well as between the secondary network basestation and the primary users are assumed to be known to the secondary network basestation. The aim of the optimization is to allocate subchannels, transmission power and bits to secondary users while ensuring that the aggregate interference over all the subchannels to the l^{th} primary user is below a certain threshold Υ_l, and each secondary user attains an equal data rate r.

Define the power gain matrix between the secondary network basestation and the secondary users as $\mathbf{H}^{\mathrm{s}} \in \mathcal{R}^{K\times N}$, i.e., $H^{\mathrm{s}}_{k,n}$ denotes the power gain between the secondary network basestation and the k^{th} secondary user at the n^{th} subchannel. The power gain matrix between the secondary network basestation and the primary users is defined as $\mathbf{H}^{\mathrm{p}} \in \mathcal{R}^{L\times N}$, i.e., $H^{\mathrm{p}}_{l,n}$ denotes the power gain between the secondary network basestation and the l^{th} primary user at the n^{th} subchannel. Let $P_{\mathrm{r}}(c,\rho_e)$ be the minimum required received power at any subchannel to ensure that the BER is below the threshold ρ where c is the number of bits per symbol. The transmission power required to ensure that the BER is below the threshold ρ_e will depend on the modulation type used at each subchannel. The minimum required signal power for the BPSK modulation and M-ary QAM (for square constellation) are given in (3.2.1) and (3.2.2). For the 8-QAM, the minimum required power for the BER threshold ρ_e can be written as [90]

$$P_{\mathrm{r}}(c,\rho_e) = 3N_\phi \left[Q^{-1}\left(\frac{3\rho_e}{2}\right)\right]^2. \qquad (4.3.1)$$

The required power at the secondary network basestation to transmit $c_{k,n}$ bits to the k^{th} secondary user at the n^{th} subchannel and to ensure that BER is less than ρ is given by

$$P_{k,n}(c_{k,n},\rho)=\frac{P_{\mathrm{r}}(c_{k,n},\rho_e)}{H^{\mathrm{s}}_{k,n}}. \tag{4.3.2}$$

where $P_{\mathrm{r}}(c_{k,n},\rho_e)$ can be obtained from (3.2.1), (3.2.2) and (4.3.1) for different $c_{k,n}$ values. Letting $P_{k,n}(0,\rho)=0$, the total power allocated to the n^{th} subchannel can be written as

$$\phi_n=\sum_{k=1}^{K}P_{k,n}(c_{k,n},\rho_e). \tag{4.3.3}$$

Denote P as the total power available at the secondary network basestation, $D=\{0,\ d_1,\ldots,d_C\}$ as the set of all possible values for $c_{k,n}$ with C denoting the number of possible modulation types. In order to keep the problem formulation simple, only four modulation types have been considered: BPSK (c =1), 4-QAM (c =2), 8-QAM (c =3), 16-QAM (c =4), i.e., $D=\{0,\ d_1,\ldots,d_C\}=\{0,\ 1,\ 2,\ 3,\ 4\}$. If $c_{k,n}=0$ then the k^{th} secondary user does not use the n^{th} subchannel for transmission. From these definitions, the rate balancing problem for the cognitive radio network can be expressed as

$$\max_{c_{k,n}\in D,r} \quad r \tag{4.3.4}$$

$$\text{s.t.} \quad \sum_{n=1}^{N}c_{k,n}\geq r, \quad k=1,\ldots,K, \tag{4.3.5}$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K}P_{k,n}(c_{k,n},\rho)\leq P, \tag{4.3.6}$$

$$\sum_{n=1}^{N}\phi_n H^{\mathrm{p}}_{l,n}\leq \Upsilon_l, \quad l=1,\ldots,L, \tag{4.3.7}$$

$$c_{k,n}=0 \text{ if } c_{k',n}\neq 0, \forall k\neq k',\ k=1,\ldots,K. \tag{4.3.8}$$

The constraint in (4.3.5) ensures that each secondary user data rate is above r, and r is the optimization variable to be maximized. Hence the cost function (4.3.4) and constraint (4.3.5) aim to maximize the worse case user data rate, i.e., rate balancing is performed. The constraints in (4.3.6) and (4.3.7) ensure that the total transmission is below the available power P and interference leakage to the l^{th} primary user is always below the threshold Υ_l respectively. The constraint in (4.3.8) ensures mutually exclusive secondary user allocation to each subchannel, i.e., no more than one secondary user is allocated to each subchannel. Since the solution of the rate balancing problem introduced in (4.3.4)-(4.3.8) should be mutually exclusive secondary user allocation, this problem is a combinatorial optimization problem [33]. The optimal solution to this combinatorial optimization problem can be obtained using integer linear programming [28]. Using (3.2.13) -(3.2.18), the integer linear programming framework for this rate balancing problem can be formulated as

$$\max_{r,\mathbf{x}} \quad r \tag{4.3.9}$$

$$\text{s.t.} \quad \mathbf{H}^{\mathrm{p}}[\mathbf{A}(\mathbf{p} \odot \mathbf{x})] \leq \boldsymbol{\Upsilon}, \tag{4.3.10}$$

$$\mathbf{p}^T\mathbf{x} \leq P, \tag{4.3.11}$$

$$\mathbf{B}\mathbf{x} \geq r\mathbf{1}_N, \tag{4.3.12}$$

$$\mathbf{0}_N \leq \mathbf{A}\mathbf{x} \leq \mathbf{1}_N, \tag{4.3.13}$$

$$x_i \ \in \ \{0,1\}, \quad i = 1,\ldots,KNC. \tag{4.3.14}$$

The complexity of the integer linear programming approach is rather high and may become prohibitively expensive even for moderate numbers of users and subchannels. Therefore a suboptimal method based on greedy algorithm in order to solve the rate balancing problem is proposed in the next subsection.

4.3.2 Suboptimal Greedy Max-Min Algorithm

A greedy Max-Min algorithm has been proposed in [72] to maximize the system data throughput for a single user cognitive radio network. The algorithm allocates bits, sequentially one bit at a time in various subchannels according to an efficiency figure. In this section, the greedy Max-Min algorithm is extended to a multiuser cognitive radio network. In order to balance the achieved data rate for each secondary user, the extended greedy Max-Min algorithm allocates bits to users one at a time in a cyclic manner.

In order to explain the extended greedy Max-Min algorithm, assume that the n^{th} subchannel is allocated to the k^{th} secondary user. From (4.3.2), the required additional power to add one more bit to the k^{th} secondary user on the n^{th} subchannel can be written as

$$\Delta_{k,n}^{+1} = P_{k,n}(r_n + 1, \rho) - P_{k,n}(r_n, \rho), \qquad (4.3.15)$$

where r_n is the number of bits already allocated in the n^{th} subchannel. The amount of additional interference leakage to the l^{th} primary user due to the addition of one bit to the k^{th} secondary user on the n^{th} subchannel is written as

$$\delta_{k,n,l}^{+1} = \Delta_{k,n}^{+1} H_{l,n}^{\mathrm{p}}. \qquad (4.3.16)$$

Define u_0 and u_l are the costs of resources already allocated as follows:

$$u_0 = \sum_{n=1}^{N} \sum_{k=1}^{K} P_{k,n}(r_n, \rho), \qquad (4.3.17)$$

$$u_l = \sum_{n=1}^{N} \sum_{k=1}^{K} P_{k,n}(r_n, \rho) H_{l,n}^{\mathrm{p}}, \quad l = 1, \ldots, L. \qquad (4.3.18)$$

From these definitions the efficiency capacity of the k^{th} secondary user on the n^{th} subchannel can be defined as [72]

$$e_{k,n}(0) = \frac{P - u_0}{\Delta_{k,n}^{+1}}, \quad (4.3.19)$$

$$e_{k,n}(l) = \frac{\Upsilon_l - u_l}{\delta_{k,n,l}^{+1}}, \quad l = 1, \ldots, L. \quad (4.3.20)$$

From the efficiency capacity values, (4.3.19) accounts for the total power constraint in (4.3.6), whereas (4.3.20) accounts for the set of interference constraints in (4.3.7). The efficiency value of the k^{th} secondary user on the n^{th} subchannel is defined as

$$\Lambda_{k,n} = \min_{l}\{e_{k,n}(l)\}, \quad l = 0, \ldots, L. \quad (4.3.21)$$

A subchannel n is chosen to allocate an additional bit if this subchannel yields the largest efficiency value as follows:

$$\alpha = \operatorname*{argmax}_{n}\{\Lambda_{k,n}\}, \quad n = 1, \ldots, N. \quad (4.3.22)$$

This process of allocating one bit at a time is repeated until one of the constraints becomes tight. Based on the greedy Max-Min algorithm, the pseudo code of the extended greedy Max-Min algorithm for data rate balancing in multiuser cognitive radio network is shown in Table 4.4. In the pseudo code, "$\backslash$" and $[\Omega_k]_i$ denote the set subtraction and the i^{th} element of the set Ω_k respectively. The basic description of the proposed algorithm is as follows. First, one bit is allocated to secondary user 1 in a subchannel that provides the largest efficiency value. Suppose the n_1^{th} subchannel is allocated to the first secondary user, then the algorithm allocates one bit to the second secondary user in one of the $(N-1)$ remaining subchannels. This is repeated until all K secondary users are allocated with one bit. Then the second bit is allocated to the first secondary user in one of the remaining $(N-K)$ subchannels or on the n_1^{th} subchannel according to the efficiency value. This

1) Initialization: $r_n = 0,\ \forall n,\ u_l = 0,\ l = 0, 1, \ldots, L$
2) Set $\Omega_k = \{1, \ldots, N\},\ \forall\ k$
3) **while** $P - u_0\ \geq\ 0$ and $\Upsilon_l - u_l\ \geq\ 0,\ l = 1, \ldots, L$
4) **for** $k = 1$ to K
5) **for** $n = [\Omega_k]_1$ to $[\Omega_k]_{Last}$
6) calculate $e_{k,n}(l),\ l = 0, \ldots, L$ using (4.3.19) and (4.3.20)
7) find $\Lambda_{k,n}$ using (4.3.21)
8) **endfor**
9) choose subchannel using (4.3.22) (denote this subchannel as α)
10) add one bit to user k on the subchannel α
11) update $u_l,\ l = 0, \ldots, L$ using (4.3.17) and (4.3.18)
12) Remove subchannel α from other users $\Omega_j = \Omega_j \setminus \{\alpha\}$, $j = 1, \ldots, K, j \neq k$
13) **endfor**
14) **endwhile**

Table 4.4. Pseudo code of the extended greedy max-min algorithm

procedure is repeated until one of the constraint is met.

4.3.3 Computational Complexity Analysis

The most popular method for solving integer linear programming is the branch and bound method [29]. Using Table 4.3, the number of operations needed to solve the problem in (4.3.9)-(4.3.14) is $\mathcal{O}[2\left(\sqrt{2}\right)^{KNC+1}(KNC + K + 2N + L + 2)]$ [35]. As stated in [72], for the modified greedy algorithm, the number of operations required is given by $\mathcal{O}(RNL)$, where R is the total number of allocated bits. For instance, consider a system with $K = 2$, $N = 4$, $L = 2$ and $C = 4$. The complexity of integer linear programming is $\mathcal{O}(8 \times 10^6)$. In order to find the worst case complexity of the proposed algorithm, assume that the proposed algorithm allocates 4 bits in each subchannel. Hence, complexity of the proposed algorithm is $\mathcal{O}(128)$, which is substantially lower than that of the optimal integer linear programming approach.

4.3.4 Simulation Results

To validate the proposed algorithm and to assess its performance, consider a cognitive radio network with $N = 64$ subchannels. The BER requirement for the secondary users has been set to $\rho_e = 0.01$. Multi path channels of length five were generated for primary users and secondary users using statistically independent Gaussian random variables. The average channel gain between the secondary network basestation and primary users is equal to 0.1 while such a gain between the secondary network basestation and secondary users is equal to 1. The interference caused by the primary user transmission is considered as noise at the secondary user receiver and the noise power spectral density at the secondary user is 0.001.

To validate the proposed extended greedy Max-Min algorithm, the solution obtained from the extended greedy Max-Min algorithm is compared

	Data rate (Number of bits)				
	Sub optimal				Optimal
Channels	secondary user1	secondary user2	secondary user3	secondary user4	r
Channel 1	27	27	27	27	27
Channel 2	24	24	24	24	25
Channel 3	17	17	17	17	18
Channel 4	29	29	29	29	30

Table 4.5. Balanced rate of each secondary user attained by both the proposed suboptimal method and the optimal method

with the optimal solution obtained by integer linear programming.

Table 4.5 shows the balanced data rate achieved by the extended greedy Max-Min algorithm and the integer linear programming based optimal method for four different random channels. The results in Table 4.5 were generated using $K = 4$ and $L = 4$. The total transmission power P and the interference leakage threshold Υ_l, $l = 1, 2, 3, 4$ have been set to 1 and 0.001 respectively. The Table 4.5 reveals that all four secondary users achieve the same date rate which is very closer to the optimum rate.

Fig. 4.4 depicts the averaged balance data rate as a function of the interference threshold when $P = 1$. This result has been generated using 100 random channel-pairs $(\mathbf{H}^{\mathrm{s}}, \mathbf{H}^{\mathrm{p}})$. Two simulation scenarios have been considered. For the first simulation, $K = 2$ and $L = 2$ were used whilst for the second simulation, $K = 4$ and $L = 4$ were used. As seen in Fig. 4.4, the balanced data rate achieved by the proposed method is very close to that of the optimum method.

4.4 Conclusion

Nearly optimal resource allocation algorithms have been proposed for OFDMA based downlink cognitive radio network. There are two different algorithms proposed based on rate adaptive technique and data rate balancing technique. The algorithm based on rate adaptive technique allocates subchan-

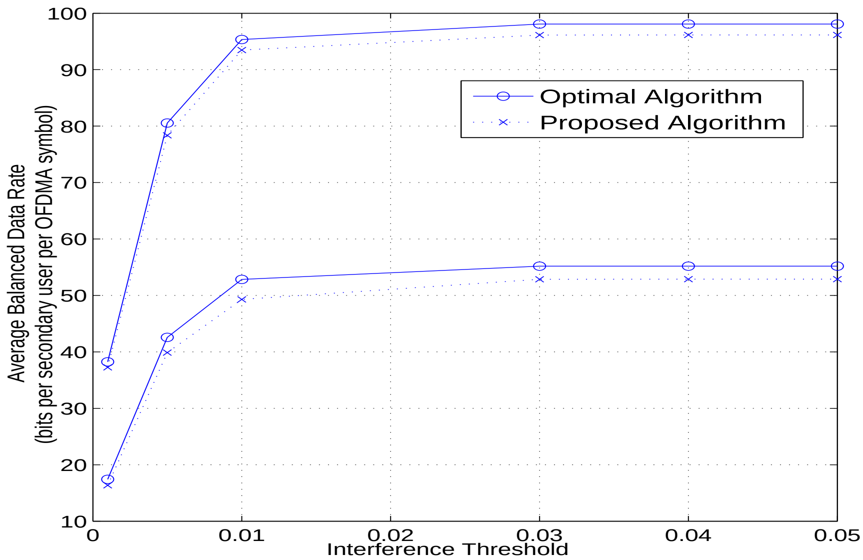

Figure 4.4. Average number of balanced data rate as a function of the interference threshold for the case of $P = 1$.

nels, bits and power adaptively to different secondary users to maximize the data throughput while ensuring that the interference leakage to the primary users is below an interference threshold. However, the solution of this algorithm might turn out to be infeasible for some channel realizations due to insufficient resources. In order to overcome the infeasibility issue, another algorithm is proposed based on data rate balancing. The proposed data rate balancing algorithm maximizes the total data throughput while maintaining equal data rates for all the secondary users subject to interference leakage constraints to the primary users. The complexity of the proposed approaches is substantially lower than that of an optimal linear integer programming method. Simulation results were provided to demonstrate the performance of the proposed methods.

Chapter 5

RESOURCE ALLOCATION AND SPATIAL MULTIPLEXING TECHNIQUES FOR COGNITIVE RADIO NETWORKS

This chapter proposes beamforming based multiuser multiplexing and resource allocation techniques for cognitive radio networks. There are two different sets of secondary users considered: real time users and non-real time users. Real time users are delay sensitive users and require certain level of QoS requirements (i.e., target SINRs or target data rates) all the time regardless of channel conditions. Non-real time users are delay tolerable and resources can be allocated based on their QoS requirements and channel gains. However, beamformers cannot be designed separately to both set of users due to the mutual interference between both set of users. In this chapter, an algorithm is proposed to jointly design beamformers for both set of users simultaneously, while ensuring real time users attains certain target QoS requirements and remaining resources are allocated to the non-real time users based on their QoS requirement and channel conditions. The design framework considered is applicable to both the overlay cognitive radio network and the underlay cognitive radio network, however, the algorithm is described for the example of an underlay cognitive radio network in order to include a very general mathematical framework. Initially an algorithm is proposed for a MISO network and then it is extended to a MIMO network. The condition for convergence is derived analytically. The simulation results are provided to validate the optimality and the convergence of the proposed algorithm.

5.1 Introduction

Introduction of data intensive multimedia and interactive services together with exponential growth of wireless and mobile users have resulted the radio spectrum a scarce resource. There have been extensive research on various techniques for the enhancement of spectrum utilization. The multi-antenna techniques have been widely used to improve spectrum usage. For example, MIMO systems have the ability to enhance the data rate by providing multiple data pipes to users [18, 91]. Employment of multiple antennas at the transmitter and receiver could also facilitate spatial multiplexing [41, 46, 59, 61, 92, 93]. Various spatial multiplexing designs are known in the literature, for example zero-forcing block diagonalization [92], convex optimization based transmitter beamformer design [46] and the uplink-downlink duality based SINR balancing beamformers [41]. The method in [92] is based on the null space of the channel matrices so that the inter-user interference is forced to zero. The method in [46] is based on the optimization of transmitter power and beamformer patterns while constraining the SINR target of each downlink user. This was solved using convex optimization techniques. The method in [41] uses the uplink-downlink duality to design a set of beamformers in order to maximize the SINR of the worst-case user resulting into SINR balancing beamformers, i.e., all users attain equal SINR.

Initially, a downlink MISO based cognitive radio network is considered and beamformers are designed only at the basestation to satisfy the requirements of real time and non-real time users simultaneously. This work is extended to MIMO based cognitive radio network later in this chapter. Since the MIMO system has multiple antennas at both the transmitter and the receiver, transmit and receive beamformers are designed jointly to provide required mixed QoS requirements based on dirty paper coding technique. For the simplicity, a system with only one subchannel is considered throughput this chapter. But this work can be extended to MIMO-OFDM based cognitive radio network by writing the channels in a block diagonal form.

5.2 MISO-based Cognitive Radio Networks

In this section, MISO based cognitive radio network is considered and beamformers for both set of users are designed only at secondary network basestation. Generally, SINR criterion has been used to define the QoS of each user in a MISO based wireless network. Hence, for the real time users, the resources should be allocated so that the users should attain a specific target SINRs all the time regardless of channel conditions. The remaining resources can be allocated to non-real time users in order to maintain fairness (i.e., balancing the achieved SINR to target SINR ratio). The beamformer design techniques based on the satisfaction of target SINRs [1, 45, 46, 61, 94, 95] and the beamformer design techniques based on SINR balancing [41, 77, 96–99] have been treated separately in the literature. However, the beamformer design to include above mentioned both criteria simultaneously is not a trivial extension. This is because these set of beamformers can not be designed separately, as the interference introduced to each other set of beamformers is a function of all the beamformer weight vectors and power allocations. This problem has been tackled by introducing some mathematical manipulations to the design framework as explained later in this section.

The proposed technique has also potential applications in spectrum sharing networks such as the cognitive radio network [8]. The beamformer design considered in this section can be applied to the overlay cognitive radio network. In this scenario the secondary network basestation uses the primary spectrum to transmit signals to its users while relaying the primary user signals to their destination using beamforming techniques. In this case, resources should be allocated (power and beams) such that each primary user receiver should attain a target SINR specified by the primary user network and the remaining resources are used to serve the secondary users. The proposed technique could also be used for an underlay cognitive radio network scenario. In this scenario, the secondary users could transmit signals provided that the interference leakage to the primary user receivers is below a target value. Therefore, additional constraints for interference leakage should also be included in the beamformer design.

5.2.1 System Model and Problem Formulation

The algorithm is described for an underlay cognitive radio network as the optimization framework considers both the transmission power constraint and the interference leakage constraints. The beamformer design techniques for the overlay cognitive radio network and the conventional wireless network can readily be obtained by dropping the interference constraints. There are K secondary users and L primary users and the secondary network basestation consists of N_t transmit antennas and each of the secondary users and primary users is equipped with single antenna. By defining $s_k(n)$, $\mathbf{u}_k \in \mathcal{C}^{N_t \times 1}$

and p_k as the transmitted symbol, transmit beamformer weight vector and the power allocation for the k^{th} secondary user respectively, the signal transmitted by the secondary network basestation is written as

$$\mathbf{x}(n) = \sum_{k=1}^{K} \sqrt{p_k}\mathbf{u}_k s_k(n), \tag{5.2.1}$$

where $\|\mathbf{u}_k\|_2 = 1, \forall k$. The variance of the symbol $s_k(n), \forall k$ is assumed to be unity. The received signal at the k^{th} secondary user can be written as

$$y_k(n) = \mathbf{h}_k^H \mathbf{x}(n) + \eta_k(n), \quad k = 1, \ldots, K, \tag{5.2.2}$$

where $\mathbf{h}_k \in \mathcal{C}^{N_t \times 1}$ is the channel gain vector between the secondary network basestation and the k^{th} secondary user. It is assumed that $\eta_k(n)$ is a zero-mean circularly symmetric AWGN with variance $\sigma_k^2 = 1$. Let $\mathbf{g}_l = \left[\|\tilde{\mathbf{h}}_l^H \mathbf{u}_1\|_2^2 \ldots \|\tilde{\mathbf{h}}_l^H \mathbf{u}_K\|_2^2\right]^T$ and $\mathbf{p} = [p_1 \ \ldots p_K]^T$, where $\tilde{\mathbf{h}}_l \in \mathcal{C}^{N_t \times 1}$ is the channel gain vector between the secondary network basestation and the l^{th} primary user. The interference leakage to the l^{th} primary user due to the secondary user transmission is $\varepsilon_l = \mathbf{g}_l^T \mathbf{p}$. In this chapter, it is assumed that secondary network basestation can determine the channel state information of primary users. As described in the introduction, the proposed scheme has three potential applications:

1. Conventional wireless network (non cognitive radio setting)

2. Overlay cognitive radio network

3. Underlay cognitive radio network

For the overlay cognitive radio network, all the primary users are assumed to be real time users. In this scenario, the secondary basestation exploits the primary users' spectrum to transmit signals to the secondary users while relaying the primary users' signals using beamforming techniques. In this case, resources should be allocated (power and beams) such that each primary user receiver should attain a target SINR specified by the primary network and the remaining resources are used to serve the secondary users. Since secondary basestation is assisting the primary network to relay the primary users signal, a cooperation between both the primary user and the secondary user networks can be expected. In this case, it is possible for the secondary network basestation to acquire the channel state information of the primary users. In the underlay cognitive radio network, the channel state information between primary users and secondary network basestation can be also obtained based on some degree of cooperation. The required protocols to obtain these channel state information are discussed in [77, 100]. The protocol for secondary network basestation is designed as follows: every frame contains sensing sub-frame and data transmission sub-frame. During the sensing sub-frame, secondary network basestation remains silent, and thus the secondary network basestation can decode the primary users code word to obtain the channel state information. In the second frame it can use the acquired channel state information of primary users for its own transmission [77, 100]. Also, as an example, a primary network could sublease the spectrum to a secondary network for monitory purposes. In this case, degree of cooperation between primary network and secondary network can be expected. This kind of scenario has been discussed in [75, 78, 101–104]. By defining $\mathbf{R}_k \triangleq \mathbf{h}_k \mathbf{h}_k^H$, the SINR of the k^{th} secondary user in the downlink can be written as

$$\mathrm{SINR}_k^{\mathrm{DL}} = \frac{p_k \mathbf{u}_k^H \mathbf{R}_k \mathbf{u}_k}{\sum_{i \neq k} p_i \mathbf{u}_i^H \mathbf{R}_k \mathbf{u}_i + \sigma_k^2}. \tag{5.2.3}$$

Consider a general scenario where the first $K1$ secondary users (i.e., real time users) out of the K secondary users employ delay intolerant real time services. Hence a target SINR should be satisfied for these users all the time. The rest of the secondary users (i.e., non-real time users) employ delay tolerant packet data services hence a target SINR is not a priority, however, in order to maintain user fairness, their SINRs should be balanced and maximized. From these requirements, a mixed QoS problem for the beamformer design is formulated as follows:

$$\max_{\mathbf{U},\mathbf{p}} \min_{k} \quad \frac{\text{SINR}_k^{\text{DL}}(\mathbf{U},\mathbf{p})}{\delta_k}, \quad k = K1+1,\ldots,K, \tag{5.2.4}$$

$$\text{s.t.} \quad \text{SINR}_k^{\text{DL}}(\mathbf{U},\mathbf{p}) \geq \gamma_k, \quad k = 1,\ldots,K1, \tag{5.2.5}$$

$$\mathbf{g}_l^T \mathbf{p} \leq P_{\text{int}}^{(l)}, \quad l = 1,\ldots,L, \tag{5.2.6}$$

$$\mathbf{1}_K^T \mathbf{p} \leq P_{\max}, \tag{5.2.7}$$

where $\mathbf{U} = [\mathbf{u}_1 \ldots \mathbf{u}_K]$, δ_k is the preferred target SINR of k^{th} non-real time user, γ_k is the target SINR for the k^{th} real time user, $P_{\text{int}}^{(l)}$ is the interference threshold for the l^{th} primary user and $P_{\max}$ is the available total transmission power at the secondary network basestation. The first set of constraints in (5.2.5) ensure that the real time secondary users should achieve their target SINRs, provided the problem is feasible. The constraints in (5.2.6) and (5.2.7) account for the interference leakage to the primary users and the total transmission power respectively. To consider an overlay network, the first $K1$ users should be treated as primary users whose SINR targets should be satisfied all the time and the SINRs of the remaining users (who are secondary users) can be balanced. The interference constraints in (5.2.6) will also be dropped in this case.

5.2.2 Analytical Framework

The solution to the mixed QoS problem stated in (5.2.4)-(5.2.7) is not trivial. This is because the SINR of each user is a function of the beamformer weight vectors of both the users with target SINRs (considered as real time users) and the users whose SINRs are to be balanced (non-real time users).

Hence beamformers for real time and non-real time users can not be designed separately. This problem is solved by writing the beamformer weight vectors of the real time users as a function of that of the non-real time users and by invoking the uplink-downlink duality. First, the multiple linear constraints in (5.2.6) - (5.2.7) are converted into a single linear constraint by introducing auxiliary variables as described in [98] as follows,

$$\sum_{l=1}^{L} a_l \left(\mathbf{g}_l^T \mathbf{p} - P_{\text{int}}^{(l)} \right) + a_{L+1} \left(\mathbf{1}_K^T \mathbf{p} - P_{\max} \right) \;\leq\; 0, \tag{5.2.8}$$

where a_l, $l = 1, \ldots, L$ and a_{L+1} are the auxiliary dual variables associated with the interference constraints and the power constraint respectively. These auxiliary variables can be updated to find the optimal solution based on the subgradient method [105], as explained later in this chapter. Define vectors $\mathbf{a} = [a_1 \ldots a_L \; a_{L+1}]^T$ and $\mathbf{b} = [\mathbf{g}_1 \ldots \mathbf{g}_L \; \mathbf{1}_K]\,\mathbf{a}$ and a scalar $P := \sum_{l=1}^{L} a_l P_{\text{int}}^{(l)} + a_{L+1} P_{\max}$. Using these definitions, the problem in (5.2.4)-(5.2.7) can be simplified to

$$\max_{\mathbf{U},\mathbf{p}} \min_{k} \quad \frac{\text{SINR}_k^{\text{DL}}(\mathbf{U},\mathbf{p})}{\delta_k}, \quad k = K1+1, \ldots, K, \tag{5.2.9}$$

$$\text{s.t.} \quad \text{SINR}_k^{\text{DL}}(\mathbf{U},\mathbf{p}) \geq \gamma_k, \quad k = 1, \ldots, K1, \tag{5.2.10}$$

$$\mathbf{b}^T \mathbf{p} \;\leq\; P. \tag{5.2.11}$$

All the auxiliary variables should satisfy the non-negative conditions all the time. This will ensure that the problem defined in (5.2.9) -(5.2.11) by introducing auxiliary variables yields an upper-bound solution of the original problem in (5.2.4)-(5.2.7). For a given set of auxiliary variables, the following dual uplink mixed QoS problem can be formulated by modifying the noise covariance and the linear constraint based on the uplink-downlink duality [98].

$$\max_{\mathbf{U},\mathbf{q}} \min_{k} \quad \frac{\mathrm{SINR}_k^{\mathrm{UL}}(\mathbf{u}_k, \mathbf{q})}{\delta_k}, \quad k = K1+1, \ldots, K, \tag{5.2.12}$$

$$\text{s.t.} \quad \mathrm{SINR}_k^{\mathrm{UL}}(\mathbf{u}_k, \mathbf{q}) \geq \gamma_k, \quad k = 1, \ldots, K1, \tag{5.2.13}$$

$$\boldsymbol{\sigma}^T \mathbf{q} \ \leq \ P, \tag{5.2.14}$$

where $\boldsymbol{\sigma} = \left[\sigma_1^2 \ \ldots \sigma_K^2\right]^T$ and $\mathbf{q} = [q_1 \ \ldots q_K]^T$, q_k is the virtual uplink power allocated to the k^{th} secondary user and $\mathrm{SINR}_k^{\mathrm{UL}}(\mathbf{u}_k, \mathbf{q})$ is the virtual uplink SINR of the k^{th} secondary user which is given as

$$\mathrm{SINR}_k^{\mathrm{UL}}(\mathbf{u}_k, \mathbf{q}) = \frac{q_k \mathbf{u}_k^H \mathbf{R}_k \mathbf{u}_k}{\mathbf{u}_k^H \left(\sum_{i \neq k} q_i \mathbf{R}_i + \boldsymbol{\Omega}\right) \mathbf{u}_k}, \tag{5.2.15}$$

where $\boldsymbol{\Omega} = \sum_{l=1}^{L} a_l \tilde{\mathbf{h}}_l \tilde{\mathbf{h}}_l^H + a_{L+1}\mathbf{I}$ is the interference-plus-noise-covariance matrix at the secondary network basestation for the virtual uplink problem. Note that $\mathbf{b} = \left[\mathbf{u}_1^H \boldsymbol{\Omega} \mathbf{u}_1 \ldots \mathbf{u}_K^H \boldsymbol{\Omega} \mathbf{u}_K\right]^T$. In the following subsections the solution to the above problem is described for a fixed set of auxiliary variables. The auxiliary variables will then be updated using a subgradient method.

Uplink Power Allocation for a given Set of Beamformers

First consider the optimal power allocation for a given set of beamformers $\tilde{\mathbf{u}}_k, \forall k$ in the uplink. At the optimal setting, the constraints in (5.2.13)-(5.2.14) must be satisfied with equality. Hence, at optimal power allocation, (5.2.12) - (5.2.14) will satisfy the following set of simultaneous equations

$$\frac{\mathrm{SINR}_k^{\mathrm{UL}}(\tilde{\mathbf{u}}_k, \tilde{\mathbf{q}})}{\delta_k} \ = \ \frac{1}{\lambda}, \quad k = K1+1, \ldots, K, \tag{5.2.16}$$

$$\mathrm{SINR}_k^{\mathrm{UL}}(\tilde{\mathbf{u}}_k, \tilde{\mathbf{q}}) \ = \ \gamma_k, \quad k = 1, 2, \ldots, K1, \tag{5.2.17}$$

$$\boldsymbol{\sigma}^T \tilde{\mathbf{q}} \ = \ P, \tag{5.2.18}$$

where $1/\lambda$ is a balanced SINR of the non-real time secondary users and $\tilde{\mathbf{q}}$ is the optimal power allocation for the given set of beamformers $\tilde{\mathbf{U}}$, where

$\tilde{\mathbf{U}} = [\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_K]$. The equations in (5.2.17) can be modified into

$$\begin{aligned}
\frac{\tilde{q}_k \tilde{\mathbf{u}}_k^H \mathbf{R}_k \tilde{\mathbf{u}}_k}{\tilde{\mathbf{u}}_k^H \left(\sum_{i\neq k, i=1}^{K} \tilde{q}_i \mathbf{R}_i + \mathbf{\Omega}\right) \tilde{\mathbf{u}}_k} &= \gamma_k, \ \ k = 1, \ldots, K1, \\
\Rightarrow \tilde{q}_k &= \gamma_k \frac{\tilde{\mathbf{u}}_k^H \left(\sum_{i\neq k, i=1}^{K} \tilde{q}_i \mathbf{R}_i + \mathbf{\Omega}\right) \tilde{\mathbf{u}}_k}{\tilde{\mathbf{u}}_k^H \mathbf{R}_k \tilde{\mathbf{u}}_k}, \ \ k = 1, \ldots, K1, \\
\Rightarrow \tilde{q}_k &= \frac{\gamma_k}{\tilde{\mathbf{u}}_k^H \mathbf{R}_k \tilde{\mathbf{u}}_k} \left[\tilde{\mathbf{u}}_k^H \left(\sum_{j\neq k, j=1}^{K1} \tilde{q}_j \mathbf{R}_j + \mathbf{\Omega} + \sum_{i=K1+1}^{K} \tilde{q}_i \mathbf{R}_i \right) \tilde{\mathbf{u}}_k \right], \ \ k = 1, \ldots, K1.
\end{aligned} \tag{5.2.19}$$

By rearranging (5.2.19), the optimum power allocation vector $\tilde{\mathbf{q}}_A = [\tilde{q}_1 \ldots \tilde{q}_{K1}]^T$ for the first $K1$ secondary users (i.e., real time secondary users) can be written as:

$$\tilde{\mathbf{q}}_A = \mathbf{D}_A \mathbf{\Psi}_A \tilde{\mathbf{q}}_A + \mathbf{D}_A \mathbf{b}_A + \mathbf{D}_A \mathbf{\Psi}_B \tilde{\mathbf{q}}_B,$$

where $\tilde{\mathbf{q}}_B = [\tilde{q}_{K1+1} \ldots \tilde{q}_K]^T$ is the optimum power allocation vector for the non-real time secondary users and $\mathbf{b}_A = \left[\tilde{\mathbf{u}}_1^H \mathbf{\Omega} \tilde{\mathbf{u}}_1 \ldots \tilde{\mathbf{u}}_{K1}^H \mathbf{\Omega} \tilde{\mathbf{u}}_{K1}\right]^T$, $\mathbf{D}_A = \text{diag}\left[\gamma_1/(\tilde{\mathbf{u}}_1^H \mathbf{R}_1 \tilde{\mathbf{u}}_1) \ldots \gamma_{K1}/(\tilde{\mathbf{u}}_{K1}^H \mathbf{R}_{K1} \tilde{\mathbf{u}}_{K1})\right]$ and

$$[\mathbf{\Psi}_A]_{ki} = \begin{cases} \tilde{\mathbf{u}}_k^H \mathbf{R}_i \tilde{\mathbf{u}}_k, i \neq k, k = 1 \ldots K1, i = 1 \ldots K1, \\ 0, i = k, \end{cases}$$

$$[\mathbf{\Psi}_B]_{ki} = \left\{ \ \tilde{\mathbf{u}}_k^H \mathbf{R}_i \tilde{\mathbf{u}}_k, \quad i = K1 + 1 \ldots K, k = 1 \ldots K1. \right.$$

Similarly, equations in (5.2.16) can be written as follows

$$\begin{aligned}
\frac{\tilde{q}_k \tilde{\mathbf{u}}_k^H \mathbf{R}_k \tilde{\mathbf{u}}_k}{\tilde{\mathbf{u}}_k^H \left(\sum_{i\neq k, i=1}^{K} \tilde{q}_i \mathbf{R}_i + \mathbf{\Omega}\right) \tilde{\mathbf{u}}_k} &= \frac{\delta_k}{\lambda}, \ \ k = K1 + 1, \ldots, K, \\
\Rightarrow \lambda \tilde{q}_k &= \frac{\delta_k \tilde{\mathbf{u}}_k^H \left(\sum_{i\neq k, i=1}^{K} \tilde{q}_i \mathbf{R}_i + \mathbf{\Omega}\right) \tilde{\mathbf{u}}_k}{\tilde{\mathbf{u}}_k^H \mathbf{R}_k \tilde{\mathbf{u}}_k}, \ \ k = K1 + 1, \ldots, K, \\
\Rightarrow \lambda \tilde{q}_k &= \frac{\delta_k}{\tilde{\mathbf{u}}_k^H \mathbf{R}_k \tilde{\mathbf{u}}_k} \left[\tilde{\mathbf{u}}_k^H \left(\sum_{i\neq k, i=K1+1}^{K} \tilde{q}_i \mathbf{R}_i + \mathbf{\Omega} + \sum_{j=1}^{K1} \tilde{q}_j \mathbf{R}_j \right) \tilde{\mathbf{u}}_k \right], \ \ k = K1 + 1, \ldots, K.
\end{aligned} \tag{5.2.20}$$

By rearranging (5.2.20), the following equation is obtained

$$\lambda\tilde{\mathbf{q}}_B = \mathbf{D}_B\boldsymbol{\Psi}_D\tilde{\mathbf{q}}_B + \mathbf{D}_B\mathbf{b}_B + \mathbf{D}_B\boldsymbol{\Psi}_C\tilde{\mathbf{q}}_A,$$

where $\mathbf{D}_B = \text{diag}\left[\delta_{K1+1}/(\tilde{\mathbf{u}}_{K1+1}^H\mathbf{R}_{K1+1}\tilde{\mathbf{u}}_{K1+1})\ldots\delta_K/(\tilde{\mathbf{u}}_K^H\mathbf{R}_K\tilde{\mathbf{u}}_K)\right]$,

$$[\boldsymbol{\Psi}_C]_{ki} = \left\{ \tilde{\mathbf{u}}_k^H\mathbf{R}_i\tilde{\mathbf{u}}_k, \quad k = K1+1\ldots K, i = 1\ldots K1, \right.$$

$$[\boldsymbol{\Psi}_D]_{ki} = \begin{cases} \tilde{\mathbf{u}}_k^H\mathbf{R}_i\tilde{\mathbf{u}}_k, i \neq k, k = K1+1\ldots K, i = K1+1\ldots K, \\ 0, i = k, \end{cases}$$

and $\mathbf{b}_B = \left[\tilde{\mathbf{u}}_{K1+1}^H\boldsymbol{\Omega}\tilde{\mathbf{u}}_{K1+1}\ldots\tilde{\mathbf{u}}_K^H\boldsymbol{\Omega}\tilde{\mathbf{u}}_K\right]^T$. Note that the power allocation $\tilde{\mathbf{q}}$ is composed of the power allocation for the real time users $\tilde{\mathbf{q}}_A$ and the power allocation for the non-real time users $\tilde{\mathbf{q}}_B$, i.e., $\tilde{\mathbf{q}} = [\tilde{\mathbf{q}}_A^T\ \tilde{\mathbf{q}}_B^T]^T$. Hence, the constraint in (5.2.18) can be written as

$$\boldsymbol{\sigma}_A^T\tilde{\mathbf{q}}_A + \boldsymbol{\sigma}_B^T\tilde{\mathbf{q}}_B = P,$$

where $\boldsymbol{\sigma}_A = [\sigma_1\ldots\sigma_{K1}]^T$ and $\boldsymbol{\sigma}_B = [\sigma_{K1+1}\ldots\sigma_K]^T$. Therefore, the equations (5.2.16) - (5.2.18) can be reformulated into the following matrix forms:

$$\lambda\tilde{\mathbf{q}}_B = \mathbf{D}_B\boldsymbol{\Psi}_D\tilde{\mathbf{q}}_B + \mathbf{D}_B\mathbf{b}_B + \mathbf{D}_B\boldsymbol{\Psi}_C\tilde{\mathbf{q}}_A, \tag{5.2.21}$$

$$\tilde{\mathbf{q}}_A = \mathbf{D}_A\boldsymbol{\Psi}_A\tilde{\mathbf{q}}_A + \mathbf{D}_A\mathbf{b}_A + \mathbf{D}_A\boldsymbol{\Psi}_B\tilde{\mathbf{q}}_B, \tag{5.2.22}$$

$$P = \boldsymbol{\sigma}_A^T\tilde{\mathbf{q}}_A + \boldsymbol{\sigma}_B^T\tilde{\mathbf{q}}_B. \tag{5.2.23}$$

Assume that $(\mathbf{I} - \mathbf{D}_A\boldsymbol{\Psi}_A)$ is invertible and $(\mathbf{I} - \mathbf{D}_A\boldsymbol{\Psi}_A)^{-1}$ is a nonnegative matrix. The conditions required to satisfy these assumptions will be provided in the subsequent sections. Using these assumptions and (5.2.22), $\tilde{\mathbf{q}}_A$ can be written in terms of $\tilde{\mathbf{q}}_B$ as,

$$\tilde{\mathbf{q}}_A = (\mathbf{I} - \mathbf{D}_A\boldsymbol{\Psi}_A)^{-1}\mathbf{D}_A\boldsymbol{\Psi}_B\tilde{\mathbf{q}}_B + (\mathbf{I} - \mathbf{D}_A\boldsymbol{\Psi}_A)^{-1}\mathbf{D}_A\mathbf{b}_A. \tag{5.2.24}$$

By substituting (5.2.24) into (5.2.21), the following is obtained,

$$\lambda \tilde{\mathbf{q}}_B = \mathbf{D}\tilde{\mathbf{q}}_B + \mathbf{d}, \tag{5.2.25}$$

where

$$\mathbf{D} = \mathbf{D}_B \boldsymbol{\Psi}_D + \mathbf{D}_B \boldsymbol{\Psi}_C \left(\mathbf{I} - \mathbf{D}_A \boldsymbol{\Psi}_A\right)^{-1} \mathbf{D}_A \boldsymbol{\Psi}_B, \tag{5.2.26}$$

$$\mathbf{d} = \mathbf{D}_B \mathbf{b}_B + \mathbf{D}_B \boldsymbol{\Psi}_C \left(\mathbf{I} - \mathbf{D}_A \boldsymbol{\Psi}_A\right)^{-1} \mathbf{D}_A \mathbf{b}_A. \tag{5.2.27}$$

The constraint in (5.2.23) can also be written in terms of $\tilde{\mathbf{q}}_B$ by substituting (5.2.24) into (5.2.23) as follows:

$$\mathbf{c}^T \tilde{\mathbf{q}}_B = P - c, \tag{5.2.28}$$

where

$$\mathbf{c}^T = \boldsymbol{\sigma}_A^T \left(\mathbf{I} - \mathbf{D}_A \boldsymbol{\Psi}_A\right)^{-1} \mathbf{D}_A \boldsymbol{\Psi}_B + \boldsymbol{\sigma}_B^T, \tag{5.2.29}$$

$$c = \boldsymbol{\sigma}_A^T \left(\mathbf{I} - \mathbf{D}_A \boldsymbol{\Psi}_A\right)^{-1} \mathbf{D}_A \mathbf{b}_A. \tag{5.2.30}$$

Multiplying both sides of (5.2.25) by $\mathbf{c}^T$ and using (5.2.28), the following equation can be obtained

$$\lambda = \frac{1}{P-c}\mathbf{c}^T \mathbf{D}\tilde{\mathbf{q}}_B + \frac{1}{P-c}\mathbf{c}^T \mathbf{d}.$$

Therefore, the uplink problem in (5.2.21)-(5.2.23) for the power allocation can be converted into the determination of power allocation for only the non-real time secondary users, subject to $(\mathbf{I} - \mathbf{D}_A \boldsymbol{\Psi}_A)$ is invertible and $(\mathbf{I} - \mathbf{D}_A \boldsymbol{\Psi}_A)^{-1}$ is a nonnegative matrix, as follows:

$$\begin{aligned} \lambda \tilde{\mathbf{q}}_B &= \mathbf{D}\tilde{\mathbf{q}}_B + \mathbf{d}, \\ \lambda &= \frac{1}{P-c}\mathbf{c}^T \mathbf{D}\tilde{\mathbf{q}}_B + \frac{1}{P-c}\mathbf{c}^T \mathbf{d}. \end{aligned} \tag{5.2.31}$$

These equations can be formulated into the following eigensystem [41]:

$$\lambda \tilde{\mathbf{q}}_{\text{ext}} = \mathbf{\Upsilon}(\tilde{\mathbf{U}})\tilde{\mathbf{q}}_{\text{ext}}. \tag{5.2.32}$$

where $\tilde{\mathbf{q}}_{\text{ext}} = [\tilde{\mathbf{q}}_B^T \ 1]^T$ and

$$\mathbf{\Upsilon}(\tilde{\mathbf{U}}) = \begin{bmatrix} \mathbf{D} & \mathbf{d} \\ \frac{1}{P-c}\mathbf{c}^T\mathbf{D} & \frac{1}{P-c}\mathbf{c}^T\mathbf{d} \end{bmatrix}. \tag{5.2.33}$$

From the Perron-Frobenious theory, the eigenvector corresponding to the largest eigenvalue of a nonnegative matrix is always nonnegative and unique, which is called the Perron vector [60]. There is no nonnegative eigenvector except positive multipliers of Perron vector for a given nonnegative matrix. To this end, the following Lemma is necessary to satisfy the nonnegativity of $\mathbf{\Upsilon}(\tilde{\mathbf{U}})$.

Lemma 1: A sufficient condition to enable $(\mathbf{I} - \mathbf{D}_A\mathbf{\Psi}_A)$ nonsingular and $(\mathbf{I} - \mathbf{D}_A\mathbf{\Psi}_A)^{-1}$ nonnegative is $\rho(\mathbf{D}_A\mathbf{\Psi}_A) \le 1$. Also the conditions

$$\rho(\mathbf{D}_A\mathbf{\Psi}_A) \le 1, \tag{5.2.34}$$

$$c \le P, \tag{5.2.35}$$

will imply that $\mathbf{\Upsilon}(\tilde{\mathbf{U}})$ is a nonnegative matrix.

Proof. See Appendix B. ■

The following Corollary is an immediate consequence of Lemma 1.

Corollary 1: If a given set of beamformers satisfy the conditions in (5.2.34) and (5.2.35), then from (5.2.32), $\tilde{\mathbf{q}}_{\text{ext}}$ is equal to the Perron vector of $\mathbf{\Upsilon}(\tilde{\mathbf{U}})$. Hence, $\tilde{\mathbf{q}}_B$ can be obtained by scaling $\tilde{\mathbf{q}}_{\text{ext}}$ such that the last element of it is equal to one. Once $\tilde{\mathbf{q}}_B$ is determined, $\tilde{\mathbf{q}}_A$ can be obtained from $\tilde{\mathbf{q}}_B$ using (5.2.24).

In the next subsection, how to obtain beamformers for a given power allocation in the uplink has been studied.

Beamformer Design for a given Power Allocation

For a given power allocation, the optimal beamformers for all the secondary users in the virtual uplink can be obtained by maximizing the SINR of each secondary user independently [41]. Hence, the optimal beamformers for all the secondary users in the uplink can be determined by solving the following optimization problem:

$$\tilde{\mathbf{u}}_i = \underset{\mathbf{u}_i}{\operatorname{argmax}} \frac{\mathbf{u}_i^H \mathbf{R}_i \mathbf{u}_i}{\mathbf{u}_i^H \left(\sum_{k=1,k\neq i}^{K} \tilde{q}_k \mathbf{R}_k + \mathbf{\Omega} \right) \mathbf{u}_i}, \text{s.t.} \|\mathbf{u}_i\| = 1, \forall i. \tag{5.2.36}$$

The solution can be obtained by finding the dominant generalized eigenvector of the matrix pairs $\left[\mathbf{R}_i, \left(\sum_{k=1,k\neq i}^{K} \tilde{q}_k \mathbf{R}_k + \mathbf{\Omega}\right)\right], \forall i$. From Corollary 1, these beamformers are required to satisfy the conditions in (5.2.34) and (5.2.35) in order to obtain nonnegative $\mathbf{\Upsilon}(\tilde{\mathbf{U}})$. It is apparent that, the conditions (5.2.34) and (5.2.35) are not necessarily satisfied for any arbitrary power allocation. However, in the subsequent subsections, a method is proposed for an appropriate power initialization so that the conditions in (5.2.34) and (5.2.35) can be satisfied. In the next subsection, an iterative algorithm is proposed to obtain the optimal beamformers and the power allocation for a given set of auxiliary variables with an appropriate power initialization.

Iterative Solution

The virtual uplink power vector is initialized with $\mathbf{q}^{(0)}$ and obtain the corresponding beamformer matrix $\tilde{\mathbf{U}}^{(0)}$ using (5.2.36) and $\mathbf{q}^{(0)}$. Denote the matrices $\mathbf{D}_A$, $\mathbf{\Psi}_A$ and $\mathbf{\Upsilon}(\tilde{\mathbf{U}})$ that are obtained using $\tilde{\mathbf{U}}^{(0)}$ as $\mathbf{D}_A^{(0)}$, $\mathbf{\Psi}_A^{(0)}$ and $\mathbf{\Upsilon}(\tilde{\mathbf{U}}^{(0)})$ respectively. Assume that the beamformers $\tilde{\mathbf{U}}^{(0)}$ obtained from this power initialization $\mathbf{q}^{(0)}$ will satisfy the required conditions in (5.2.34) and (5.2.35) i.e., $\mathbf{\Upsilon}(\tilde{\mathbf{U}}^{(0)})$ in (5.2.33) is non-negative matrix (next subsection will elaborate how to obtain this initial power allocation $\mathbf{q}^{(0)}$). In the first iteration, equation (5.2.32) is given by

$$\lambda^{(1)}\tilde{\mathbf{q}}_{\text{ext}}^{(1)} = \mathbf{\Upsilon}(\tilde{\mathbf{U}}^{(0)})\tilde{\mathbf{q}}_{\text{ext}}^{(1)}. \tag{5.2.37}$$

The superscript associated with each quantity denotes the iteration number. A feasible power allocation $\tilde{\mathbf{q}}_B^{(1)}$ is obtained for a given set of beamformers $\tilde{\mathbf{U}}^{(0)}$ by finding the Perron vector of $\mathbf{\Upsilon}(\tilde{\mathbf{U}}^{(0)})$ and scaling it such that the last element of the Perron vector $\tilde{\mathbf{q}}_{\text{ext}}^{(1)}$ is one. From Lemma 1 (i.e., $(\mathbf{I} - \mathbf{D}_A^{(0)}\mathbf{\Psi}_A^{(0)})^{-1}$ is a nonnegative matrix) and $\tilde{\mathbf{q}}_B^{(1)}$, a valid power allocation $\tilde{\mathbf{q}}_A^{(1)}$ can be obtained using (5.2.24). Power vector $\tilde{\mathbf{q}}^{(1)}$ can be obtained using $\tilde{\mathbf{q}}_A^{(1)}$ and $\tilde{\mathbf{q}}_B^{(1)}$ as $\tilde{\mathbf{q}}^{(1)} = [(\tilde{\mathbf{q}}_A^{(1)})^T \ (\tilde{\mathbf{q}}_B^{(1)})^T]^T$. $\tilde{\mathbf{U}}^{(1)}$ can be obtained for the second iteration using $\tilde{\mathbf{q}}^{(1)}$ and (5.2.36) . Similar to (5.2.37), in the second iteration, the following equation is solved:

$$\lambda^{(2)}\tilde{\mathbf{q}}_{\text{ext}}^{(2)} = \mathbf{\Upsilon}(\tilde{\mathbf{U}}^{(1)})\tilde{\mathbf{q}}_{\text{ext}}^{(2)}. \tag{5.2.38}$$

Consider the following Lemma.

Lemma 2: Matrix $\mathbf{\Upsilon}(\tilde{\mathbf{U}}^{(1)})$ in (5.2.38) is a nonnegative matrix and $\lambda^{(2)} \leq \lambda^{(1)}$.

Proof: See Appendix C. ■

Using these definitions, an iterative algorithm is presented in Table 5.1, namely the Beamformer Allocation (BA) algorithm to obtain the beamformers for a given auxiliary vector $\mathbf{a}$. The quantities associated with the n^{th} iteration are denoted by the superscript (n).

Using Lemma 2 and mathematical induction, it can be can proved that $\mathbf{\Upsilon}(\tilde{\mathbf{U}}^{(n)})$ is a nonnegative matrix and $\lambda^{(n)} \leq \lambda^{(n-1)}$ at the n^{th} iteration (i.e, $n > 2$) for a feasible initial power allocation $\mathbf{q}^{(0)}$. Hence, λ is monotonically decreasing with the iteration number, where $1/\lambda$ is the balanced SINR of the non-real time secondary users. Hence, SINRs of the non-real time secondary users are increasing monotonically with the iteration number. Since, the system is limited by the transmit power constraint, SINR cannot increase beyond a certain value. Hence, λ must converge to a value denoted as λ^*. Denote the corresponding $\tilde{\mathbf{U}}$ as $\tilde{\mathbf{U}}^*$.

1) Initialize $\mathbf{q}^{(0)}$ with an uplink power
2) $n = 0$
3) **repeat**
4) $n \leftarrow n + 1$
5) Solve (5.2.36) using $\mathbf{q}^{(n-1)}$ to obtain $\tilde{\mathbf{U}}^{(n-1)}$
6) Generate $\mathbf{D}_A^{(n-1)}$, $\mathbf{D}_B^{(n-1)}$, $\mathbf{\Psi}_A^{(n-1)}$, $\mathbf{\Psi}_B^{(n-1)}$, $\mathbf{\Psi}_C^{(n-1)}$, $\mathbf{\Psi}_D^{(n-1)}$, $\mathbf{b}_A^{(n-1)}$ and $\mathbf{b}_B^{(n-1)}$ using $\tilde{\mathbf{U}}^{(n-1)}$
7) Solve (5.2.32) and obtain $\lambda^{(n)}$ and $\tilde{\mathbf{q}}_B^{(n)}$
8) Obtain $\tilde{\mathbf{q}}_A^{(n)}$ from $\tilde{\mathbf{q}}_B^{(n)}$ and (5.2.24)
9) Define $\mathbf{q}^{(n)} = \left[\tilde{\mathbf{q}}_A^{(n)^T} \ \tilde{\mathbf{q}}_B^{(n)^T}\right]^T$
10) **until** $\lambda^{(n-1)} - \lambda^{(n)} \leq \epsilon$
11) $\lambda^* = \lambda^{(n)}$ and $\tilde{\mathbf{U}}^* = \tilde{\mathbf{U}}^{(n-1)}$

Table 5.1. Beamformer Allocation (BA) algorithm

From the uplink-downlink duality, identical SINR values can be achieved in both the uplink and the downlink with the same set of beamformers but with a different power allocation. From this, the uplink beamformers $\tilde{\mathbf{U}}^*$ obtained from BA algorithm can be used to achieve the same SINR values in the downlink. Denote the downlink power allocation $\mathbf{p} = \left[\mathbf{p}_A^T \ \mathbf{p}_B^T\right]^T$, where $\mathbf{p}_A$ and $\mathbf{p}_B$ are the downlink power allocation vectors for the real time and the non-real time secondary users respectively. Similar to the uplink equations in (5.2.25)-(5.2.27), the following equations for the power allocation of the non-real time secondary users can be written in the downlink:

$$\lambda^* \tilde{\mathbf{p}}_B^* = \mathbf{D}_D^* \tilde{\mathbf{p}}_B^* + \mathbf{d}_D^*, \tag{5.2.39}$$

where

$$\mathbf{D}_D^* = \mathbf{D}_B^*\mathbf{\Psi}_D^{*T} + \mathbf{D}_B^*\mathbf{\Psi}_B^{*T}(\mathbf{I} - \mathbf{D}_A^*\mathbf{\Psi}_A^{*T})^{-1}\mathbf{D}_A^*\mathbf{\Psi}_C^{*T}, \quad (5.2.40)$$

$$\mathbf{d}_D^* = \mathbf{D}_B^*\boldsymbol{\sigma}_B + \mathbf{D}_B^*\mathbf{\Psi}_B^{*T}(\mathbf{I} - \mathbf{D}_A^*\mathbf{\Psi}_A^{*T})^{-1}\mathbf{D}_A^*\boldsymbol{\sigma}_A. \quad (5.2.41)$$

The matrices $\mathbf{D}_A^*$, $\mathbf{D}_B^*$, $\mathbf{\Psi}_A^*$, $\mathbf{\Psi}_B^*$, $\mathbf{\Psi}_C^*$ and $\mathbf{\Psi}_D^*$ are generated using $\tilde{\mathbf{U}}^*$. To this end, the following Lemma is necessary to determine a feasible $\tilde{\mathbf{p}}_B^*$.

Lemma 3: $(\lambda^*\mathbf{I} - \mathbf{D}_D^*)$ is nonsingular and $(\lambda^*\mathbf{I} - \mathbf{D}_D^*)^{-1}$ is a nonnegative matrix.

Proof: See Appendix D. ■

Using Lemma 3, the power allocation for the non-real time secondary users in the downlink can be obtained from (5.2.39) as follows:

$$\tilde{\mathbf{p}}_B^* = (\lambda^*\mathbf{I} - \mathbf{D}_D^*)^{-1}\mathbf{d}_D^*. \quad (5.2.42)$$

The downlink power allocation for real time secondary users can be obtained similar to the uplink power allocation for the real time secondary users as in (5.2.24), as follows:

$$\tilde{\mathbf{p}}_A^* = (\mathbf{I} - \mathbf{D}_A^*\mathbf{\Psi}_A^{*T})^{-1}\mathbf{D}_A^*\mathbf{\Psi}_C^{*T}\tilde{\mathbf{p}}_B^* + (\mathbf{I} - \mathbf{D}_A^*\mathbf{\Psi}_A^{*T})^{-1}\mathbf{D}_A^*\boldsymbol{\sigma}_A. \quad (5.2.43)$$

For a given set of auxiliary variables, how to obtain the beamformers and the power allocations have been explained so far. For a given auxiliary variable vector $\mathbf{a}$, the beamformers and the corresponding downlink power allocation can be obtained using the BA algorithm in Table 5.1 and equations (5.2.42) and (5.2.43) respectively. Convergence of BA algorithm to a feasible point has been proved for a given set of auxiliary variables. Now, how to update the auxiliary vector using a subgradient method is explained. Elements of the auxiliary vector a_l, $l = 1 \ldots L+1$ are updated via a subgradient

1)	Initialize $\mathbf{a}^{(0)}$, t, ϵ, $m = 0$.
2)	**repeat**
3)	$m \leftarrow m + 1$
4)	Obtain $\tilde{\mathbf{U}}^{*}$ and λ^{*} using BA
5)	Obtain $\tilde{\mathbf{p}}_A$ and $\tilde{\mathbf{p}}_B$ using (5.2.42) and (5.2.43)
6)	$\mathbf{p}^{(m)} = [\tilde{\mathbf{p}}_A^T \; \tilde{\mathbf{p}}_B^T]^T$
7)	Update auxiliary variables using (5.2.44) and (5.2.45)
8)	**until** (5.2.46) and (5.2.47) are met

Table 5.2. Complete algorithm

algorithm [105, 106] according to the downlink power allocation as follows:

$$a_l^{(m+1)} = a_l^{(m)} + t\left(\mathbf{g}_l^T \mathbf{p}^{(m)} - P_{\text{int}}\right), \; l = 1 \ldots L, \tag{5.2.44}$$

$$a_{L+1}^{(m+1)} = a_{L+1}^{(m)} + t\left(\mathbf{1}^T \mathbf{p}^{(m)} - P_{\max}\right), \tag{5.2.45}$$

where t denotes the step-size of the subgradient algorithm. The following stopping criteria are used to terminate the algorithm:

$$\left| a_l^{(m+1)} \left(\mathbf{g}^T \mathbf{p}^{(m)} - P_{\text{int}}\right)\right| \leq \epsilon, \; l = 1 \ldots L, \tag{5.2.46}$$

$$\left| a_{L+1}^{(m+1)} \left(\mathbf{1}^T \mathbf{p}^{(m)} - P_{\max}\right)\right| \leq \epsilon, \tag{5.2.47}$$

where ϵ is a very small value. The complete algorithm is summarized as in Table 5.2[1].

Power Initialization

In the previous subsections, a particular initial power allocation $\mathbf{q}^{(0)}$ is assumed so that $\boldsymbol{\Upsilon}(\tilde{\mathbf{U}}^{(\mathbf{0})})$ is nonnegative and the proposed algorithm converges.

[1] If all the users are non-real time users, then solution of the proposed algorithm will be the same as that of the ordinary cognitive radio SINR balancing problem [96]. At the same time if all the users are real time users, then each real time user achieves an SINR greater or equal to their target SINR [1].

In this subsection, how to obtain this power initialization $\mathbf{q}^{(0)}$ will be discussed. To proceed, first consider only real-time secondary users (i.e., secondary network basestation allocates power only to the real time secondary users). The virtual uplink SINR balancing problem with only the real time secondary users can be written as follows:

$$\begin{aligned} \max_{\mathbf{U}_A,\mathbf{q}_A} \min_k \quad & \frac{\mathrm{SINR}_k^{\mathrm{UL}}(\mathbf{u}_k,\mathbf{q}_A)}{\gamma_k}, \quad k=1,\ldots,K1, && (5.2.48)\\ \text{s.t.} \quad & \boldsymbol{\sigma}_A^T\mathbf{q}_A \ \le \ P, && (5.2.49) \end{aligned}$$

where $\mathbf{U}_A = [\mathbf{u}_1 \ldots \mathbf{u}_{K1}]$ is the matrix containing the beamformers for the real time secondary users. At the global optimal point, for a set of fixed beamformers, $\tilde{\mathbf{U}}_A = [\tilde{\mathbf{u}}_1 \ldots \tilde{\mathbf{u}}_{K1}]$, the following equations can be obtained similar to (5.2.21), (5.2.31), (5.2.33) and (5.2.32):

$$\begin{aligned} \lambda_A\mathbf{q}_A \ &= \ \mathbf{D}_A\boldsymbol{\Psi}_A\tilde{\mathbf{q}}_A + \mathbf{D}_A\mathbf{b}_A, && (5.2.50)\\ \lambda_A \ &= \ \frac{1}{P}\boldsymbol{\sigma}_A^T\mathbf{D}_A\boldsymbol{\Psi}_A\mathbf{q}_A + \frac{1}{P}\boldsymbol{\sigma}_A^T\mathbf{D}_A\mathbf{b}_A, && (5.2.51)\\ \boldsymbol{\Upsilon}_A \ &= \ \begin{bmatrix} \mathbf{D}_A\boldsymbol{\Psi}_A & \mathbf{D}_A\mathbf{b}_A \\ \frac{1}{P}\boldsymbol{\sigma}_A^T\mathbf{D}_A\boldsymbol{\Psi}_A & \frac{1}{P}\boldsymbol{\sigma}_A^T\mathbf{D}_A\mathbf{b}_A \end{bmatrix}, && (5.2.52)\\ \lambda_A\tilde{\mathbf{q}}_{A_{\mathrm{ext}}} \ &= \ \boldsymbol{\Upsilon}_A\tilde{\mathbf{q}}_{A_{\mathrm{ext}}}, && (5.2.53) \end{aligned}$$

where $1/\lambda_A$ is the ratio between the achieved SINR and the target SINR for the real time secondary users and $\tilde{\mathbf{q}}_{A_{\mathrm{ext}}} = [\tilde{\mathbf{q}}_A^T \ 1]^T$ is the corresponding extended power vector. The optimal beamformers $\mathbf{U}_A$ and the power allocation $\mathbf{q}_A$ can be determined using the iterative algorithm proposed in [41]. For any initial power allocation, the beamformers can be obtained using (5.2.36) for the real time secondary users (i.e., $k = 1, \ldots, K1$). For this set of beamformers, a power allocation will be performed again using (5.2.53). This will result into a higher SINR value [41]. This iteration will be continued until the desired accuracy (i.e., there is no significant different in λ_A).

After the convergence of the algorithm, the achieved SINR of the real time secondary user can be written as follows:

$$\mathrm{SINR}_k^{\mathrm{UL}} = \frac{\gamma_k}{\lambda_A^R},\ k = 1, \ldots, K1, \tag{5.2.54}$$

where λ_A^R is the optimum value after the convergence. If $\lambda_A^R \leq 1$, then the problem defined in (5.2.4)-(5.2.7) is feasible i.e., the target SINR for all the real time secondary users can be achieved. After the convergence, (5.2.50) can be written as

$$\lambda_A^R \mathbf{q}_A^R \ = \mathbf{D}_A^R \mathbf{\Psi}_A^R \mathbf{q}_A^R + \mathbf{D}_A^R \mathbf{b}_A^R, \tag{5.2.55}$$

where the superscript R means that the corresponding matrices and vectors are obtained after the convergence. $\mathbf{q}_A^R$ provides the power allocation for the real time secondary users in the absence of non-real time secondary users. If the problem is feasible (i.e., $\lambda_A^R \leq 1$), the initial power allocation for the overall problem is provided as $\mathbf{q}^{(0)} = [\mathbf{q}_A^{R^T}\ \mathbf{0}_{K-K1}{}^T]^T$ (i.e., set the initial power allocation of the real-time secondary users as obtained in (6.2.16) and set the initial power allocation for the non-real time secondary users as zeros). The algorithm in Table 5.1 with this initial power allocation will converge according to the Lemma 4 provided below.

Lemma 4: For a feasible problem $\rho(\mathbf{D}_A^R \mathbf{\Psi}_A^R) \ \leq \ 1$. If $\rho(\mathbf{D}_A^R \mathbf{\Psi}_A^R) \ \leq \ 1$, then the beamformers $\tilde{\mathbf{U}}^{(0)}$ obtained using $\mathbf{q}^{(0)}$ will satisfy the conditions (5.2.34) and (5.2.35), and the algorithm in Table 5.1 will converge.

Proof: See Appendix E. ■

5.2.3 Computational Complexity Analysis

For a given set of auxiliary variables, the complexity of the proposed algorithm in Table 5.1 mainly depends on the complexities of a matrix inversion and the eigenvalue decomposition. For a given $n \times n$ matrix, the required

	Matrix Inversion	Eigenvalue decomposition
Step 5	$K\mathcal{O}(N_t^3)$	$K\mathcal{O}\left[N_t^3 + (N_t\text{log}^2 N_t)\text{log}b\right]$
Step 7	-	$\mathcal{O}\{(N_K+1)^3 + \left[(N_K+1)\text{log}^2(N_K+1)\right]\text{log}b)\}$
Step 8	$\mathcal{O}\{(N_K+1)^3\}$	-

Table 5.3. Required arithmetic operation for the algorithm in Table 5.1

No. of secondary users	4	5	6	7	8	9
Average No. of Iteration	11.61	14.27	17.00	15.62	14.69	14.14

Table 5.4. The average number of iterations required for the algorithm in Table 5.1 to converge

arithmetic operations to determine its inverse and eigenvectors are given by $\mathcal{O}(n^3)$ and $\mathcal{O}(n^3 + (n\text{log}^2 n)\text{log}b))$ respectively, where b is the relative error bound [107]. Based on this, the number of arithmetic operations required per iteration for the proposed algorithm in Table 5.1 is provided in Table 5.3, where N_t and N_K are the number of antennas at the secondary network basestation and the number of non-real time users in the network respectively. Only steps 5, 7 and 8 require the matrix inversion and the eigenvalue decomposition in the algorithm in Table 5.1. Hence the total arithmetic operations required in each iteration is the summation of arithmetic operations needed for matrix inversion and the eigenvalue decomposition. The average number of iterations required is provided in the Table 5.4 for different number of secondary users. The results in Table 5.4 have been averaged over 2000 channel realizations. Hence it is worth to note that the number of iterations needed to converge does not depend on the number of secondary users.

	Used Power by Secondary User				Total	Achieved SINR by Secondary User				Inter. Leakage to Primary User	
Channels	1	2	3	4	Power	1	2	3	4	1	2
Channel 1	1.0286	0.4969	0.2066	0.2678	2	10.0000	5.0000	4.4039	4.4039	0.10	0.10
Channel 2	0.5760	0.2640	0.8025	0.3575	2	10.0000	5.0000	5.4491	5.4491	0.10	0.10
Channel 3	0.2372	0.2850	0.7834	0.6944	2	10.0000	5.0000	9.7150	9.7150	0.10	0.10
Channel 4	0.5308	0.4665	0.2612	0.7415	2	10.0000	5.0000	4.1820	4.1820	0.10	0.10
Channel 5	0.7963	0.4475	0.2917	0.4645	2	10.0000	5.0000	9.6285	9.6285	0.10	0.10

Table 5.5. Power allocations and the achieved SINRs using the proposed method

	Target SINR for SecondaryUser				Total	Used Power by Secondary User				Inter. Leakage to Primary User	
Channels	1	2	3	4	Power	1	2	3	4	1	2
Channel 1	10.0000	5.0000	4.4039	4.4039	2	1.0286	0.4969	0.2066	0.2678	0.10	0.10
Channel 2	10.0000	5.0000	5.4491	5.4491	2	0.5760	0.2640	0.8025	0.3575	0.10	0.10
Channel 3	10.0000	5.0000	9.7150	9.7150	2	0.2372	0.2850	0.7834	0.6944	0.10	0.10
Channel 4	10.0000	5.0000	4.1820	4.1820	2	0.5308	0.4665	0.2612	0.7415	0.10	0.10
Channel 5	10.0000	5.0000	9.6285	9.6285	2	0.7963	0.4475	0.2917	0.4645	0.10	0.10

Table 5.6. Target SINRs and the user power consumptions using the SDP-based method of [1].

5.3 Simulation Results

To validate the optimality of the proposed algorithm, a cognitive radio network with four secondary users and two primary users is considered. The first two secondary users are considered as real time secondary users and they need to achieve their target SINRs all the time whilst the SINRs for the remaining two secondary users should be balanced. The secondary network basestation consists of five antennas. The interference leakage threshold to primary users and the total available transmission power are set to 0.1 and 2 respectively. The channel coefficients between the secondary network basestation and the secondary users as well as those between the secondary network basestation and the primary users are assumed to be known to the secondary network basestation. The channel gains are generated using independent and identically distributed zero-mean circularly symmetric complex Gaussian random variables. The noise power at each secondary user receiver is set to 0.05. The stopping criterion ϵ has been set to 0.001. The auxiliary variables a_l, $l = 1 \ldots L+1$ have been initialized to 0.1 and the step-size t has been set to 0.01. The target SINRs for the first two secondary users have been set to 10 and 5 respectively while preferred target SINRs of non-real times users have been set to 1 ($\delta_k = 1,\ k = K1+1, \ldots, K$). The power allocations for each secondary user and the balanced SINR values obtained using the proposed algorithm are depicted for five different set of random channels in Table 5.5. The first two secondary users achieve their target SINRs whilst the other two users achieve a balanced SINRs. Note that, the interference and the total power constraints are satisfied with equality.

To validate the optimality of the proposed algorithm, the solution shown in Table 5.5 is compared with a result obtained using an SDP approach [1]. The SDP based design will provide optimum results; however, the SINR targets for all four users have to be set. For the same random channels used in Table 5.5, the SINR target of all four users have been set the same as that obtained in Table 5.5. For example, according to Table 5.5, for random channel 1, the SINR targets for all four secondary users are set as [10.0000 5.0000 4.4039 4.4039], whilst the primary users interference threshold has been set to 0.1. The results obtained using the SDP approach of [1] are shown in Table 5.6. Comparing Table 5.5 and Table 5.6, the power allocation obtained using the SDP approach of [1] is the same as the results obtained using the proposed method. The interference leakage value for the primary users is equal 0.1 in both schemes. Also, it has been observed that both the proposed method and the optimum SDP approach provide the same set of beamformers. Therefore, the proposed algorithm yields an optimum solution for the mixed QoS problem considered in this chapter. Note that the SDP-based method of [1] has been used just to demonstrate the optimality of the proposed scheme. However, it should be stressed that the approach of [1] cannot be directly applied to the considered scenario as the maximum achievable balanced SINR values for the non-real time secondary users are not known a priori. In order to verify convergence of λ, the evolution of λ against the iteration number of the BA algorithm is plotted in Fig. 5.1 (a) for a given set of auxiliary variables. The results shown for various target SINRs of secondary user1 and secondary user2 demonstrate monotonically decreasing λ. The inverse of the λ value should provide the balanced SINR value for the secondary user3 and secondary user4. Hence the balanced SINR is monotonically increasing against the iteration number. In Fig. 5.1 (b), the balanced SINR value of secondary user3 and secondary user4 is plotted against the target SINRs of secondary user1 and secondary user2. The target SINRs for secondary user1 and secondary user2 are set to identical value. As expected the balanced SINR value for secondary user3 and secondary user4 is decreasing as the target SINRs of secondary user1 and secondary user2 are increased.

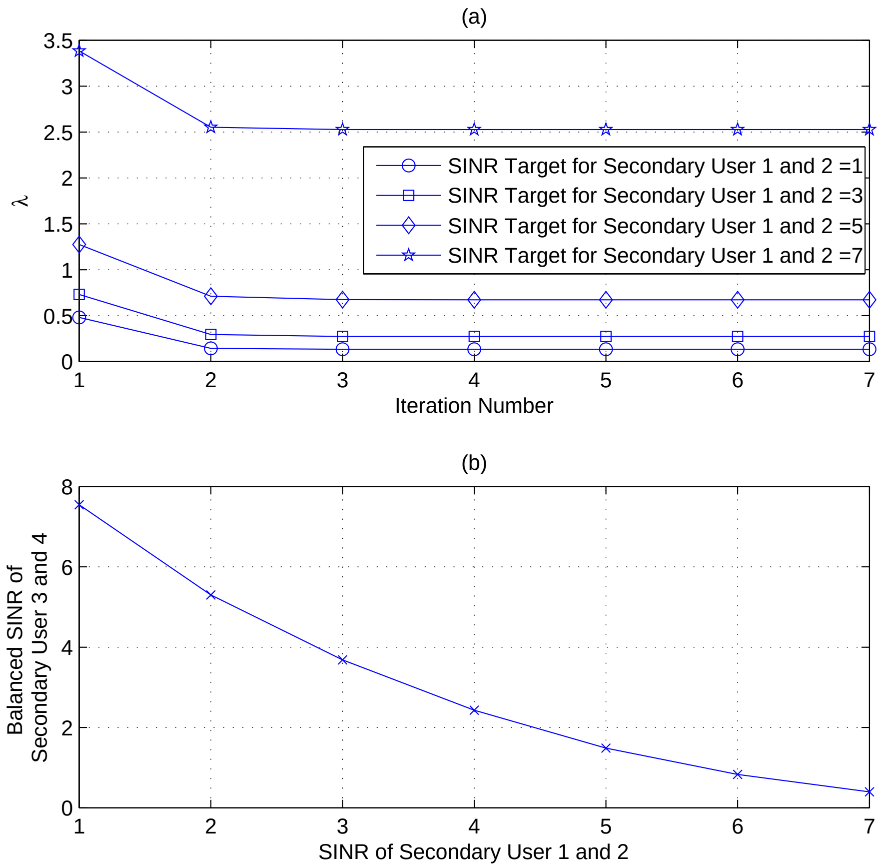

Figure 5.1. The sub-figure (a) depicts the convergence of λ against the iteration number of the BA algorithm for different target SINRs for secondary user1 and secondary user2. The sub-figure (b) depicts the balanced SINR of the secondary user3 and secondary user4 against varying target SINRs for secondary user1 and secondary user2.

Fig. 5.2 reveals the convergence of the complete algorithm in Table 5.2 for three different step-sizes $t = 0.01,\ 0.005$ and 0.001. The sub-figures (a) and (b) show the evolution of interference leakage to the primary users while the sub-figure (c) shows the evolution of the total transmission power against the iteration number. The figures confirm that the interference leakages to primary user1 and primary user2 and the transmission power converge to the value of 0.1 and 2 respectively as set in the constraints.

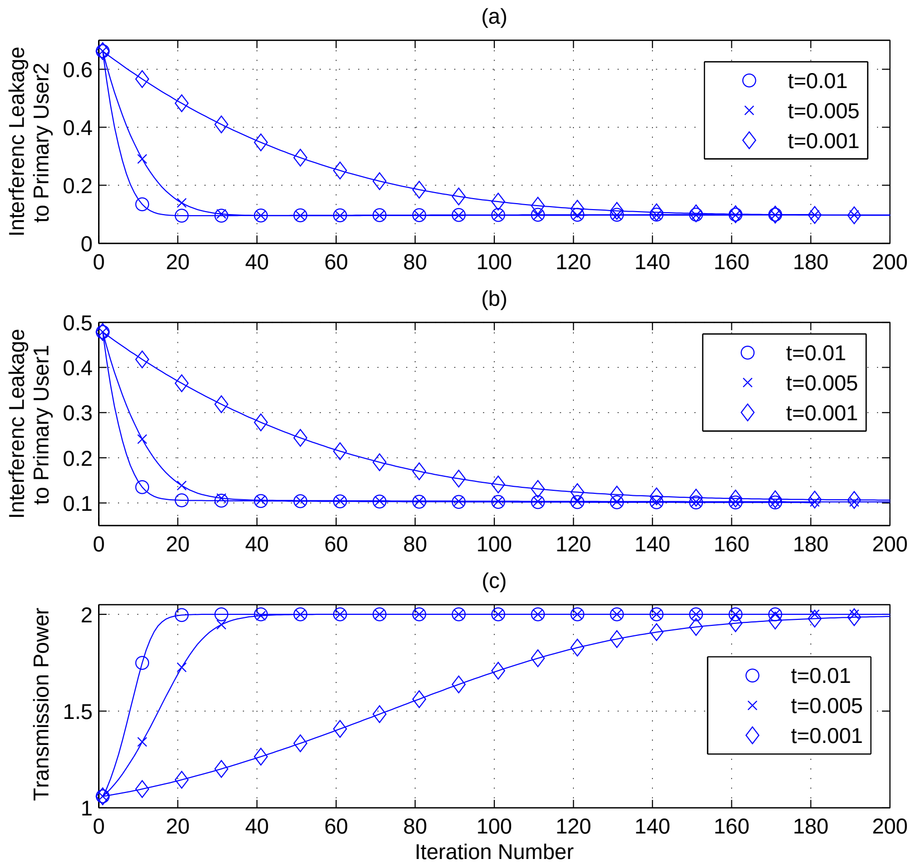

Figure 5.2. The convergence of the complete algorithm in Table 5.2 against the iteration of auxiliary variables. The auxiliary variables are updated using three different step-sizes $t = 0.01,\ 0.005$ and 0.001.

5.4 MIMO-based Cognitive Radio Network

In this section, MIMO based cognitive radio network is considered, where both the secondary network basestation and the secondary users are equipped with multiple antennas. Multiple antennas at both the transmitter and the receiver has the potential to enhance diversity and capacity [18]. The capacity region for the conventional MIMO-BC (i.e., with only the sum power constraint) has been characterized in [108] using the principle of dirty paper coding and successive decoding. The capacity region for the MIMO-BC was obtained by maximizing the weighted sum rate of multiple users subject to a sum power constraint. The weighted sum rate maximization problem for a MIMO-BC is non-convex due to the coupled structure of the transmission covariance matrices. But the same problem in the dual form in terms of MIMO-MAC is convex and it can be solved efficiently. Hence, the key idea to solve the MIMO-BC problem is to transform the MIMO-BC problem into an equivalent convex MIMO-MAC problem using the BC-MAC duality relationship [108].

The weighted sum rate maximization may be a capacity achieving scheme but the rate allocated to each user may not be uniform. The users with good channels will tend to have a higher data rate allocation at the expense of users with weaker channels. In the literature, the rate balancing techniques [109–112] have been considered to maintain fairness among users. The works in [109–112] considered rate balancing techniques in the BC for a conventional wireless network where the basestation is subject to a single sum power constraint. In [111] and [112], the authors obtained a point on the boundary of the capacity region which represents the equal data rate for the users. These works cannot be directly extended to cognitive radio network due to multiple constraints such as the sum power constraint and the interference constraint.

The capacity region for the MIMO-BC cognitive radio network with multiple constraints was obtained in [98]. For this work also, nonlinear encoding and decoding schemes (i.e., DPC at the transmitter and successive decoding at the receiver) were used as in [108]. To balance the data rates among the secondary users in the cognitive radio network, an appropriate point needs to be determined on the boundary of the MIMO-BC capacity region of cognitive radio network. In this section, in contrast to all the work available in the literature, an algorithm is proposed for data rate balancing for a subset of users (i.e., non-real time users) while ensuring the remaining users (i.e., real time users) achieve specific data rate target.

5.4.1 System model and Problem Statement

A downlink underlay MIMO based cognitive radio network with K secondary users and L primary users is considered. It is assumed that the secondary user basestation, k^{th} secondary user and l^{th} primary user consists of n_T transmit antennas, n_k receive antennas (i.e., $k = 1, \ldots, K$) and n_l receive antennas (i.e., $l = 1, \ldots, L$) respectively. The received signal at the k^{th} secondary user, $\mathbf{y}_k \in \mathcal{C}^{n_k \times 1}$, is written as:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{z}_k, \ k = 1, \ldots, K \tag{5.4.1}$$

where $\mathbf{H}_k \in \mathcal{C}^{n_k \times n_T}$ denotes the channel matrix from the secondary user basestation to the k^{th} secondary user, $\mathbf{x} = \sum_{k=1}^{K} \mathbf{x}_k$, $\mathbf{x}_k \in \mathcal{C}^{n_T \times 1}$ is the transmitted signal from secondary user basestation to k^{th} secondary user and $\mathbf{z}_k \in \mathcal{C}^{n_k \times 1}$ is the noise vector at the k^{th} secondary user whose elements are modeled as independent and identically distributed complex Gaussian random variables with zero mean and variance σ_k^2. The transmission signal covariance matrix at the secondary user basestation is denoted as $\mathbf{Q} = \sum_{k=1}^{K} \mathbf{Q}_k \in \mathcal{C}^{n_T \times n_T}$ where $\mathbf{Q}_k = \mathbb{E}\{\mathbf{x}_k \mathbf{x}_k^H\}$ is the transmission covariance matrix for the k^{th} secondary user. With this definition of $\mathbf{Q}$, the interference leakage to the l^{th} primary user is given as:

$$\varepsilon_l = \mathrm{Tr}(\mathbf{G}_l \mathbf{Q} \mathbf{G}_l^H), \ l = 1, \ldots, L \tag{5.4.2}$$

where $\mathbf{G}_l \in \mathcal{C}^{n_l \times n_T}$ denotes the channel matrix from the secondary user basestation to the l^{th} primary user. Assume that dirty paper coding technique is employed at the secondary user basestation and successive interference cancelation is used at each receiver. In the dirty paper coding, the signals for different users are encoded in a sequential manner before the transmission. For example, secondary user1s' signal is encoded first, then secondary user2s' signal, and so on. Thus later users' signals are encoded in such a way to mitigate the known interference from the previous encoded signals. Hence, the rate achieved by the k^{th} secondary user in the BC can be expressed as [98]

$$r_k = \log \frac{|\sigma_k^2 \mathbf{I} + \mathbf{H}_k(\sum_{i=k}^{K} \mathbf{Q}_i)\mathbf{H}_k^H|}{|\sigma_k^2 \mathbf{I} + \mathbf{H}_k(\sum_{i=k+1}^{K} \mathbf{Q}_i)\mathbf{H}_k^H|}, \ k = 1, \ldots, K. \tag{5.4.3}$$

At each receiver, the signals from different users are decoded sequentially by subtracting already decoded signals [98].

Problem Formulation

Since there are two different classes of secondary users considered, assume that the first $K1$ number of secondary users (i.e., $k = 1, \ldots, K1$) out of the K secondary users are real time secondary users and they should satisfy the target data rates all the time. The rest of the secondary users (i.e., $k = K1, \ldots, K$) are non-real time secondary users and in order to maintain fairness, their rates should be balanced. Define $\tilde{r}_k, \quad k = 1, \ldots, K1$, $P_{\max}$ and P_l are the target data rates for the k^{th} real time secondary user, the total transmission power at the secondary user basestation and the interference leakage threshold to the l^{th} primary user respectively. Using these definitions, the rate balancing problem with rate constraints for the real time secondary users can be formulated as follows:

$$\max_{\gamma, \{\mathbf{Q}_k \succeq 0\}} \quad \gamma, \tag{5.4.4}$$

$$\text{s.t.} \quad r_k \geq \tilde{r}_k, \quad k = 1, 2, \ldots, K1, \tag{5.4.5}$$

$$r_k \geq \gamma, \qquad k = K1 + 1, \ldots, K, \tag{5.4.6}$$

$$\text{Tr}(\mathbf{Q}) \leq P_{\max}, \tag{5.4.7}$$

$$\varepsilon_l \leq P_l, \ \forall l, \tag{5.4.8}$$

$$\mathbf{r}, \tilde{\mathbf{r}} \in C(\{H_k\}, \{H_l\}, \{P_l\}, P_{\max}), \tag{5.4.9}$$

where γ is a scalar optimization variable, $\mathbf{r} = [r_1, \ldots, r_k]^T$ and $\tilde{\mathbf{r}} = [\tilde{r}_1, \ldots, \tilde{r}_{K1}]^T$. The first set of constraints in (5.4.5) ensure that the real time secondary users always achieve their target data rates, provided the problem is feasible. In constraints (5.4.6), the optimization variable γ is used to balance the data rate among non-real time secondary users. The constraints in (5.4.7)-(5.4.8) account for the sum power constraint at the secondary user basestation and the interference leakage constraints to the primary users respectively. Finally the target data rates and the balanced data rates must lie in the capacity region $C(\{H_k\}, \{H_l\}, \{P_l\}, P_{\max})$ defined by constraint (5.4.9). It is worth to mention that the capacity region for the MIMO based cognitive radio network is convex [98]. The problem defined in (5.4.4)-(5.4.9) cannot be solved efficiently due to the inclusion of the mixed QoS requirements and the non-convex function in (5.4.3).

5.4.2 Algorithmic Solution

The problem in (5.4.4)-(5.4.9) can be modified in order to facilitate an efficient solution. It is proposed to start the problem by assuming that all secondary users are non-real time secondary users, where none of the secondary users are required to satisfy a specific set of data rate targets. Later, this assumption is modified to account for the real time secondary users with a set of target data rates. Hence, (5.4.4)-(5.4.9) can be modified into a data rate-ratio balancing problem as follows:

$$\max_{\gamma, \{\mathbf{Q}_k \succeq 0\}} \quad \gamma, \tag{5.4.10}$$

$$\text{s.t.} \quad \frac{r_k}{\tilde{r}_k} \geq \gamma, \qquad k = 1, \ldots, K, \tag{5.4.11}$$

$$\mathrm{Tr}(\mathbf{Q}) \leq P_{\max}, \tag{5.4.12}$$

$$\varepsilon_l \leq P_l, \ \forall l, \tag{5.4.13}$$

$$\mathbf{r}, \tilde{\mathbf{r}} \in C(\{H_k\}, \{H_l\}, \{P_l\}, P_{\max}), \tag{5.4.14}$$

where $\tilde{r}_k = 1$ for $k = K1+1, \ldots, K$. In (5.4.11), $\frac{r_k}{\tilde{r}_k}$ is the ratio of the achieved data rate and the target data rate for the k^{th} secondary user. The solution obtained by solving the above problem might not satisfy the required rate of the real time secondary users. However, in the subsequent subsections how this rate balancing problem can be modified to achieve the target data rates for the real time secondary users will be explained.

The problem in (5.4.10)-(5.4.14) is different from the conventional capacity balancing problems in [111] and [112] due to two different classes of constraints in (5.4.12)-(5.4.13). The conventional capacity balancing problems that are based on only a sum power constraint, can be efficiently solved using the BC-MAC duality. It has been shown in [98] that the problem with multiple linear constraints can be written as an equivalent single linear constraint problem with multiple auxiliary variables. By assuming $\mathbf{r}, \tilde{\mathbf{r}} \in C(\{H_k\}, \{H_l\}, \{P_l\}, P_{\max})$, the problem in (5.4.10)-(5.4.14) can be modified into a single linear constraint problem with multiple auxiliary variables as follows [98]:

$$\max_{\gamma, \{\mathbf{Q}_k \succeq 0\}} \quad \gamma, \tag{5.4.15}$$

$$\text{s.t.} \quad \frac{r_k}{\tilde{r}_k} \geq \gamma, \qquad k = 1, \ldots, K, \tag{5.4.16}$$

$$\sum_{l=1}^{L} a_l \varepsilon_l + a_{L+1} \mathrm{Tr}(\mathbf{Q}) \leq P, \tag{5.4.17}$$

where $a_l,\ l = 1, \ldots, L+1$ are the auxiliary variables for the L number of interference constraints and the sum power constraint respectively and $P := \sum_{l=1}^{L} a_l P_l + a_{L+1} P_{\max}$. Using (5.4.2), the inequality constraint in (5.4.17) can be written as

$$\mathrm{Tr}(\mathbf{AQ}) \leq P, \tag{5.4.18}$$

where $\mathbf{A} := \sum_{l=1}^{L} a_l \mathbf{G}_l^H \mathbf{G}_l + a_{L+1} \mathbf{I}$. The auxiliary variables $a_l,\ l = 1, \ldots, L+1$ can be updated using a subgradient method [105]. Since the capacity region for this problem is convex, this problem and its dual problem will hold a strong duality [33]. The Lagrangian dual of the problem in (5.4.15)-(5.4.16) subject to a constraint (5.4.18) can be written as

$$g(\boldsymbol{\lambda}) = \max_{\gamma,\{\mathbf{Q}_k \succeq 0\}} \quad \gamma(1-\sum_{k=1}^{K}\lambda_k) + \sum_{k=1}^{K}\lambda_k \frac{r_k}{\tilde{r}_k}, \tag{5.4.19}$$

$$\text{s.t.} \quad \mathrm{Tr}(\mathbf{AQ}) \le P, \tag{5.4.20}$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_K]$ denotes the Lagrangian variables. The Lagrangian function (5.4.19) is not bounded unless $\sum_{k=1}^{K}\lambda_k = 1$. Hence, (5.4.19)-(5.4.20) can be reformulated to

$$\min_{\boldsymbol{\lambda}} \max_{\{\mathbf{Q}_k \succeq 0\}} \quad \sum_{k=1}^{K}\lambda_k \frac{r_k}{\tilde{r}_k}, \tag{5.4.21}$$

$$\text{s.t.} \quad \mathrm{Tr}(\mathbf{AQ}) \le P, \tag{5.4.22}$$

$$\sum_{k=1}^{K}\lambda_k = 1. \tag{5.4.23}$$

By incorporating the constraint $\sum_{k=1}^{K}\lambda_k = 1$ into the objective function, this problem can be reformulated into

$$\min_{\tilde{\boldsymbol{\lambda}}} \max_{\{\mathbf{Q}_k \succeq 0\}} \quad \frac{r_K}{\tilde{r}_K} + \sum_{k=1}^{K-1}\lambda_k\left(\frac{r_k}{\tilde{r}_k} - \frac{r_K}{\tilde{r}_K}\right), \tag{5.4.24}$$

$$\text{s.t.} \quad \mathrm{Tr}(\mathbf{AQ}) \le P, \tag{5.4.25}$$

where $\tilde{\boldsymbol{\lambda}} = [\lambda_1, \ldots, \lambda_{K-1}]$. For a given $\tilde{\boldsymbol{\lambda}}$, the inner maximization is equal to a weighted sumrate maximization problem in MIMO-BC cognitive radio network [98]. The outer minimization over $\tilde{\boldsymbol{\lambda}}$ is used to balance rate between the secondary users. It has been shown in [111] and [112], $\tilde{\boldsymbol{\lambda}}$ can be updated using a subgradient method.

Weighted Sum Rate Maximization

For a given $\boldsymbol{\lambda}$ with $\sum_{k=1}^{K} \lambda_k = 1$, the weighted sum rate maximization for a MIMO-BC cognitive radio network is given by (i.e., (5.4.21)-(5.4.22))

$$\begin{aligned} \max_{\{\mathbf{Q}_k \succeq 0\}} \quad & \sum_{k=1}^{K} \lambda_k \frac{r_k}{\tilde{r}_k}, \\ \text{s.t.} \quad & \text{Tr}(\mathbf{AQ}) \leq P. \end{aligned} \tag{5.4.26}$$

To exploit the BC-MAC duality, this problem must be converted into an equivalent MAC weighted sum rate maximization. By assuming all secondary users have equal noise variance[2], σ^2, the weighted sum rate maximization in MIMO-MAC cognitive radio network is given by

$$\max_{\{\mathbf{Q}_k^m \succeq 0\}} \quad \sum_{k=1}^{K} \rho_k r_k^m, \tag{5.4.27}$$

$$\text{s.t.} \quad \sigma^2 \text{Tr}(\mathbf{Q}^m) \leq P, \tag{5.4.28}$$

where $\rho_k = \frac{\lambda_k}{\tilde{r}_k}$, r_k^m and $\mathbf{Q}^m = \sum_{k=1}^{K} \mathbf{Q}_k^m$, $\mathbf{Q}_k^m$ are k^{th} secondary user weight, rate and transmission covariance matrix respectively in the MAC. In the dual MAC problem, the decoding order used for successive interference cancelation must be converse to the encoding order of the dirty paper coding in BC [98]. Hence, the dual MAC rate for the k^{th} secondary user is given by $r_k^m = \log \frac{|\mathbf{A} + \sum_{i=1}^{k} \mathbf{H}_i^H \mathbf{Q}_i^m \mathbf{H}_i|}{|\mathbf{A} + \sum_{i=1}^{k-1} \mathbf{H}_i^H \mathbf{Q}_i^m \mathbf{H}_i|}$.

It has been shown in [113], that the users with a higher weight must be decoded prior to the users with lower weights. Hence, assume for a notational simplicity, $\rho_1 > \rho_2 > \ldots \rho_K$. With this assumption, the problem in (5.4.27)-(5.4.28) is equal to

$$\begin{aligned} \max_{\{\mathbf{Q}_k^m \succeq 0\}} \quad & \sum_{k=1}^{K} \eta_k \log |\mathbf{A} + \mathbf{H}_k^H (\sum_{i=1}^{k} \mathbf{Q}_i) \mathbf{H}_k| \\ \text{s.t.} \quad & \sigma^2 \text{Tr}(\mathbf{Q}^m) \leq P, \end{aligned} \tag{5.4.29}$$

[2]If noise variance is different, this factor can be absorbed into the channel coefficient by generating an equivalent BC problem [41].

1) Initialization: ϵ, $n=0$, $\lambda_k^{(0)} = 1/K, \forall k$, $\mathbf{E} = (1-1/K)\mathbf{I}_{K-1}$
2) **repeat**
3) $n \leftarrow n+1$
4) Obtain $\mathbf{Q}_k^m$, $\forall$ k by solving (5.4.29) for a given $\boldsymbol{\lambda}$
5) Obtain $\mathbf{Q}_k$, $\forall$ k using BC-MAC covariance matrix mapping [98]
6) Obtain r_k, $\forall$ k using (5.4.3) and $\mathbf{Q}_k$, $\forall$ k
7) Obtain $\mathbf{s}$ using subgradient of (5.4.24)
8) Define $\tilde{\mathbf{s}} = \sqrt{\frac{\mathbf{s}}{\mathbf{s}^T\mathbf{E}^n\mathbf{s}}}$
9) Update $\tilde{\boldsymbol{\lambda}}^{(n+1)} = \tilde{\boldsymbol{\lambda}}^{(n)} - \frac{1}{K}\mathbf{E}^{(n)}\tilde{\mathbf{s}}$
$\lambda_K^{(n+1)} = 1 - \sum_{k=1}^{K-1}\lambda_k^{(n+1)}$
$\mathbf{E}^{(n+1)} = \frac{(K-1)^2}{(K-1)^2-1}(\mathbf{E}^{(n)} - \frac{2}{K}\mathbf{E}^{(n)}\tilde{\mathbf{s}}\tilde{\mathbf{s}}^T\mathbf{E}^{(n)})$
12) **until** $\lvert\frac{r_k}{\tilde{r}_k} - \frac{r_K}{\tilde{r}_K}\rvert \leq \epsilon,\ k = 1, \ldots, K-1$ **or** $\max(\text{eig}(\mathbf{E})) < \epsilon$

Table 5.7. Pseudo code of the rate ratio balancing algorithm

where $\eta_k = \rho_k - \rho_{k+1}$ and $\rho_{K+1} = 0$. This is a convex problem [33] and it can be solved efficiently to obtain MAC covariance matrices. The obtained MAC covariance matrices must be converted into BC covariance matrices using MAC-BC covariance matrix mapping [98, 112].

Updating $\boldsymbol{\lambda}$ to Balance Rates

The achieved rates for the secondary users in BC can be obtained using (5.4.3) and the BC covariance matrices. Since, the minimization problem in (5.4.24)-(5.4.25) is convex and the subgradient of (5.4.24) with respect to λ_k is equal to $\frac{r_k}{\tilde{r}_k} - \frac{r_K}{\tilde{r}_K}$, $\tilde{\boldsymbol{\lambda}}$ can be updated using the ellipsoidal method [111]. At each iteration, this subgradient is used to compute a new ellipsoid which has a smaller volume than the previous one. This updated $\tilde{\boldsymbol{\lambda}}$ will be used in (5.4.29) to obtain MAC covariance matrices. Define a vector $\mathbf{s} = [s_1, \ldots, s_{K-1}]^T$ where $s_k = \frac{r_k}{\tilde{r}_k} - \frac{r_K}{\tilde{r}_K}$ is the subgradient of (5.4.24) with respect to the λ_k. Defining a matrix $\mathbf{E}$ of size $(K-1)\times(K-1)$ to represents the ellipsoid, the iterative algorithm is provided in Table 5.7. This iteration will continue until all the secondary users attain equal rate ratio with required accuracy i.e., $|\frac{r_k}{\tilde{r}_k} - \frac{r_K}{\tilde{r}_K}| \leq \epsilon,\ k = 1, \ldots, K-1$ or the largest eigenvalue of $\mathbf{E}$ is very small (i.e., $\max(\text{eig}(\mathbf{E})) < \epsilon$).

A similar algorithm is provided in [111] for a rate balancing problem in a conventional network. In [111], the authors showed that the algorithm has the ability to converge. Since the sum power constraint and the interference constraint are coupled into a single linear constraint, the above algorithm will also converge. However, the sum power constraint and the interference constraints might not be satisfied. In the subsequent subsections how to adapt the auxiliary variables in order to satisfy both the constraints will be explained.

The algorithm in Table 5.7 balances the data rate ratio of all the secondary users without considering the real time secondary users' target data rate. The solution obtained from this algorithm may not satisfy the required data rate for each real time secondary users, but this will be resolved in the subsequent subsections.

Rate Balancing with QoS Consideration

Assume the balanced ratio of rates obtained from the algorithm provided in Table 5.7 is equal to γ_0 i.e.,$\frac{r_k}{\tilde{r}_k} = \gamma_0,\ \forall k$. Hence, the rate achieved by the real time secondary users and the non-real time secondary users are given by

$$r_k = \gamma_0 \tilde{r}_k,\ k = 1, \ldots, K1 \tag{5.4.30}$$

$$r_k = \gamma_0,\ k = K1+1, \ldots, K. \tag{5.4.31}$$

1) Initialization: $q = 0$, $\gamma_0^{(0)} = 1$
1) **repeat**
2) $\quad q \leftarrow q + 1$
3) $\quad$ Scale the $\tilde{r}_k,\ k = 1, \ldots, K1$ by $1/\gamma_0^{(q-1)}$
3) $\quad$ Obtain a $\gamma_0^{(q)}$ using algorithm in Table 5.1
4) **until** $\lvert r_k - \tilde{r}_k \rvert \leq \epsilon,\ k = 1, \ldots, K1$

Table 5.8. Pseudo code for rate ratio balancing with QoS inclusion algorithm

The rates of real time secondary users' are satisfied only if $\gamma_0 \geq 1$. Once the balanced ratio of γ_0 is obtained, the target data rate for the real time secondary users must be scaled in order to achieve the target data rate. The new algorithm to include QoS in the rate balancing problem is provided in Table 5.8, where superscript (q) denotes the iteration number. The convergence of this new algorithm is analyzed in the next subsection.

Convergence Analysis

Refering to Table 5.8, assume that, after the first iteration i.e., $q = 1$, the obtained balanced rate ratio $\gamma_0^{(1)}$ is less than 1. Now the target data rate of each real time secondary users is scaled by $1/\gamma_0^{(1)}$ as in step 3. By scaling, the target data rates of real time secondary users are virtually increased. In the second iteration, a smaller balanced data rate ratio $\gamma_0^{(2)}$ can be obtained, i.e., $\gamma_0^{(1)} \geq \gamma_0^{(2)}$. This is because if the target data rates for real time secondary users are increased, from (5.4.30) and (5.4.31), the balanced data rate achieved by non-real time secondary users needs to be decreased. It is clear that if $\gamma_0^{(1)} \leq 1$ then γ_0 will decrease with iteration number. At the same time if the problem is feasible (i.e., $\tilde{\mathbf{r}} \in C(\{H_k\}, \{H_l\}, \{P_l\}, P_{\max})$ then γ_0 cannot decrease below 0 i.e., $1 \geq \gamma_0^1 \geq \gamma_0^2 \geq \gamma_0^3 \geq \ldots \geq 0$. Hence γ_0 is monotonically decreasing with iteration number. After the q^{th} iteration, the

1) Initialization: $a_l^{(0)},\ l=1,\dots,L+1,\ t\ ,p=0,\epsilon$
2) **repeat**
3) $p \leftarrow p+1$
4) Solve the algorithm in Table 5.2 and obtain $\mathbf{Q}$
5) Update $a_l^{(p)},\ l=1,\dots,L+1$ as follows
$a_l^{(p+1)}=a_l^{(p)}+t[\mathrm{Tr}(\mathbf{H}_l\mathbf{Q}\mathbf{H}_l^H)-P_l], l=1,..,L$
$a_{L+1}^{(p+1)}=a_{L+1}^{(p)}+t[\mathrm{Tr}(\mathbf{Q})-P_{\max}]$
6) **until** the following conditions are met
$\lvert a_l^{(p+1)}(\mathrm{Tr}(\mathbf{H}_l\mathbf{Q}\mathbf{H}_l^H)-P_l)\rvert \le \epsilon$
$\lvert a_{L+1}^{(p+1)}(\mathrm{Tr}(\mathbf{Q})-P_{\max})\rvert \le \epsilon$

Table 5.9. Pseudo code of the complete algorithm

data rate achieved by real time and non-real time secondary users are given by

$$r_k = \gamma_0^{(q)}\frac{\tilde{r}_k}{\gamma_0^{(q-1)}},\ k=1,\dots,K1 \tag{5.4.32}$$

$$r_k = \gamma_0^{(q)},\ k=K1+1,\dots,K. \tag{5.4.33}$$

Due to the monotonically decreasing convergence property of γ_0 when $\gamma_0^{(1)} \le 1$, at the convergence, $\gamma_0^{(q)} = \gamma_0^{(q-1)}$. From (5.4.32), the real time secondary users achieve the target data rate at convergence. Similarly the following statement can be proved: if $\gamma_0^{(1)} \ge 1$ then γ_0 will monotonically increase. Since the problem is limited by the sum power and the interference constraints, γ_0 must be less than ∞, i.e., $1 \le \gamma_0^1 \le \gamma_0^2 \le \gamma_0^3 \le \dots < \infty$.

Update Auxiliary Variables $a_l,\ l=1,\dots,L+1$

The algorithm in Table 5.8 will satisfy the data rate for real time secondary users and balance the data rates for the non-real time secondary users for a given set of auxiliary variables. But the obtained rates may not satisfy the sum power and the interference constraints. These constraints can be satisfied by updating the auxiliary variables using a subgradient method [105]. The complete description of the algorithm is provided in Table 5.9.

In Table 5.9, t denotes the step size of the subgradient algorithm and superscript p denotes the p^{th} iteration. In step 5, the auxiliary variables $a_l, l = 1, \ldots, L+1$ are updated via a subgradient algorithm [105] using the transmitter covariance matrices $\mathbf{Q}$ obtained in step 4. This iterative process is repeated until the stopping criteria in step 6 are satisfied.

5.5 Simulation Results

To validate the proposed algorithm and to assess the performance, a MIMO cognitive radio network with four secondary users and two primary users is considered. It is assumed that the secondary user basestation and all users (i.e., secondary users and primary users) are equipped with three transmit and three receive antennas respectively (i.e., $n_T = n_k = n_l = 3, \ \forall \ k, l$). The channel coefficients between the secondary user basestation and the secondary users are assumed to be known to the secondary user basestation and all secondary users. Also the channel between secondary user basestation and the primary users are assumed to be known to the secondary user basestation. Channel gains are generated using independent and identically distributed complex Gaussian random variables.

The first two secondary users are assumed to be real time secondary users and they need to achieve their target data rates all the time whilst the rates of other two secondary users should be balanced. The target rates for the first two secondary users have been set to 1.5. The interference leakage threshold to primary users and the total available transmission power are set to 0.2 and 1 respectively. The noise power at each secondary user receiver is set to 0.1. The stopping criterion ϵ has been set to 0.001. The auxiliary

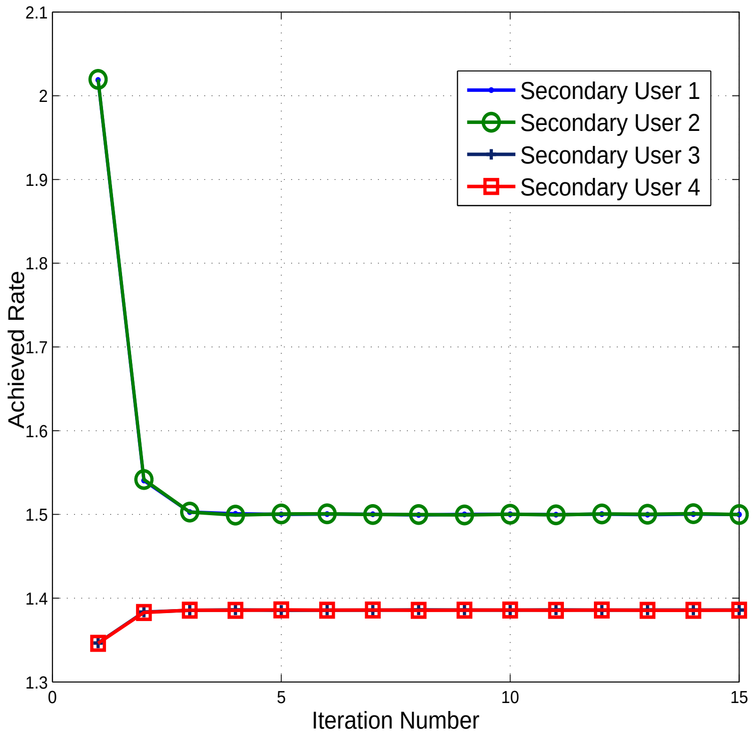

Figure 5.3. Achieved data rate of secondary users against the iteration number.

variables $a_l, \ l = 1, 2, 3$ have been initialized with 10 and the step size t has been set to 0.1 and 0.5.

Fig. 5.3 depicts the convergence of the proposed algorithm in Table 5.8 (i.e., step 4 in Table 5.9). Secondary user1 and secondary user2 achieve their target rates of 1.5 while other two secondary users achieve equal rates at the convergence. Fig. 5.4 demonstrates the convergence of the outer loop of the algorithm proposed in Table 5.9. Top figure in Fig. 5.4 shows the transmission power against the iteration number for two different step sizes. Bottom figure in Fig. 5.4 depicts the interference leakage to the two primary users against the iteration number for different step sizes. For both step sizes, the transmission power approaches its maximum (i.e., 1) while the interference leakage to the primary users are below the threshold values (i.e., 0.14 and 0.145). It is worth to note that when the step size increases,

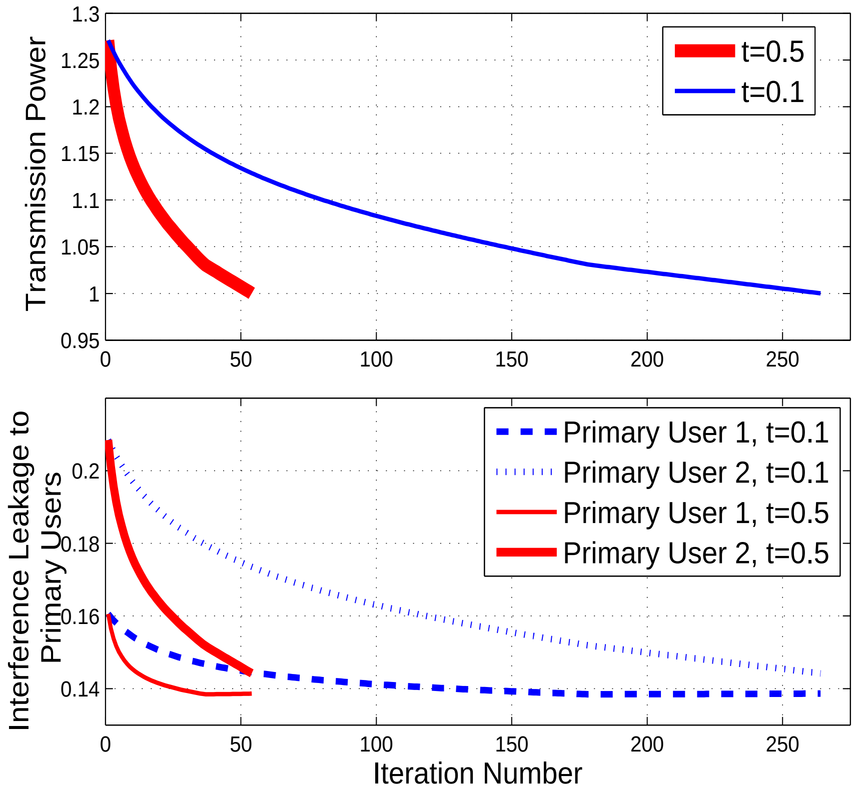

Figure 5.4. Transmission power at the secondary user basestation and the interference leakage to the primary users against the iteration number for different step size t.

the algorithm converges quickly with negligible difference at the final value.

5.6 Conclusion

An underlay cognitive radio network with two different sets of secondary users has been considered (i.e., real time and non-real time users). Initially a MISO based cognitive radio network was considered, where a joint SINR balancing and target SINR provision based beamforming technique has been proposed. The proposed technique optimally designs downlink beamformers and allocate power using the SINR uplink-downlink duality. Then this work has been extended to a MIMO based cognitive radio network using DPC technique. For a MIMO-based cognitive radio network, an algorithm was proposed to obtain transmit covariance matrices subject to constraints that the real time users archive a target data rate and the non-real time users achieve a balanced data rate based on the BC-MAC duality. Both these problems have multiple linear constraints (i.e., transmit power and interference leakage constraints) and these multiple constraints have been combined into a single linear constraint using multiple auxiliary variables. A subgradient method was used to adapt these auxiliary variables in order to satisfy the sum power and the interference leakage constraints. The convergence analysis and the simulation results were provided to validate the proposed algorithms.

Chapter 6

SUMMARY, CONCLUSION AND FUTURE WORK

The thesis has three contributing chapters and the conclusion of each chapter is summarized below, followed by a discussion on future works.

6.1 Summary and Conclusions

The optimal resource allocation algorithms for an OFDMA based underlay cognitive radio network were developed in Chapter 3. The proposed algorithm maximizes the system data throughput while maintaining the interference leakage to the primary users below a threshold, the data rate for each secondary user is above a required data rate and the total transmission power at the secondary network basestation is below a power budget. This algorithm formulated into mathematical framework using integer linear programming and solved using branch-and-bound method. In order to exploit the spatial domain, the proposed method was extended to a MIMO-OFDMA based cognitive radio network. Resource allocation in the uplink for a MIMO-OFDMA based cognitive radio network was also studied in Chapter 3, where multiple secondary users were admitted in each OFDM subchannel based on their spatial separation.

Chapter 4 has provided a low complexity algorithm in order to reduce the computational complexity associated with the optimal algorithms proposed in Chapter 3. There are two different low complexity algorithms developed based on rate adaptive and rate balancing techniques. A novel recursion based linear optimization frame work was proposed for the rate adaptive problem while greedy Max-Min based iterative algorithm was proposed for the rate balancing technique. The complexity required for the proposed technique was substantially lower than that of the optimal methods. Simulations results demonstrated that the proposed algorithm performs very close to the optimal algorithms.

In Chapter 5, a mixed QoS requirement based optimization technique was proposed for a cognitive radio network which consists of both real time and non-real time users. Cognitive radio networks based on MISO and MIMO were considered. Novel algorithms were proposed to satisfy the QoS of both sets of users. The algorithms were developed using the SINR uplink-downlink duality, BC-MAC duality and subgradient methods. The proposed algorithms have applications in conventional wireless networks as well as overlay and underlay cognitive radio networks. In order to have a general framework, the resource allocation algorithm was developed for an underlay cognitive radio network. Convergence property of both the algorithms was derived analytically. Simulations results validated the convergence property of the proposed algorithms.

6.2 Future Work

The potential areas for future research have been recognized. All the algorithms proposed in this thesis, were developed based on the assumption that the secondary network basestation has perfect knowledge of the channel state information of the users. It is important to extend these algorithms for the case when only imperfect channel knowledge is available to the secondary network basestation.

The algorithms in Chapter 5 were developed to design beamformers for mixed QoS requirement problems based on the radio link layer parameters. The target SINR of the non-real time users depends on the channel gain, users' QoS requirement, buffer length and packet queue length at the secondary network basestation. The target SINR must be decided at MAC layer based on these parameters. Hence, the algorithms in Chapter 5 can be extended under a cross-layer design framework.

Appendix

A. Proof: Maximization of SINR

$$\underset{\mathbf{w}}{\text{maximize}} \quad \frac{\mathbf{w}^H\mathbf{R}_d\mathbf{w}}{\mathbf{w}^H\mathbf{R}_{i+n}\mathbf{w}} = \underset{\mathbf{w}}{\text{maximize}} \quad \frac{\mathbf{w}^H\mathbf{R}_d\mathbf{w}}{\mathbf{w}^H\mathbf{R}_{i+n}^{1/2}\mathbf{R}_{i+n}^{1/2}\mathbf{w}} \tag{6.2.1}$$

This maximization is equivalent to

$$\begin{aligned} \underset{\mathbf{u}}{\text{maximize}} \quad & \mathbf{u}^H\mathbf{R}_{i+n}^{-1/2}\mathbf{R}_d\mathbf{R}_{i+n}^{-1/2}\mathbf{u} \\ \text{subject to} \quad & \mathbf{u}^H\mathbf{u} = 1, \end{aligned} \tag{6.2.2}$$

where $\mathbf{u} = \mathbf{R}_{i+n}^{1/2}\mathbf{w}$ and the solution of (6.2.2) will be the eigenvector corresponding to the largest eigenvalue of $\mathbf{R}_{i+n}^{-1/2}\mathbf{R}_d\mathbf{R}_{i+n}^{-1/2}$.

$$\begin{aligned} \mathbf{R}_{i+n}^{-1/2}\mathbf{R}_d\mathbf{R}_{i+n}^{-1/2}\mathbf{u} &= \lambda_{max}\mathbf{u} \\ \Rightarrow \mathbf{R}_{i+n}^{-1}\mathbf{R}_d\mathbf{R}_{i+n}^{-1/2}\mathbf{u} &= \lambda_{max}\mathbf{R}_{i+n}^{-1/2}\mathbf{u} \\ \Rightarrow \mathbf{R}_{i+n}^{-1}\mathbf{R}_d\mathbf{w} &= \lambda_{max}\mathbf{w} \end{aligned} \tag{6.2.3}$$

This completes the proof. ■

B. Proof of Lemma 1

If $\rho(\mathbf{D}_A\mathbf{\Psi}_A) \leq 1$, then $\lim_{n\to\infty}(\mathbf{D}_A\mathbf{\Psi}_A)^n = 0$ [60]. Hence, $(\mathbf{I}-\mathbf{D}_A\mathbf{\Psi}_A)$ is a nonsingular matrix and using Neumann series $(\mathbf{I}-\mathbf{D}_A\mathbf{\Psi}_A)^{-1} = \sum_{n=0}^{\infty}(\mathbf{D}_A\mathbf{\Psi}_A)^n$ which is a nonnegative matrix [60]. When $(\mathbf{I}-\mathbf{D}_A\mathbf{\Psi}_A)^{-1}$ is a nonnegative matrix then matrix $\mathbf{D}$ in (5.2.26) and vector $\mathbf{d}$ in (5.2.27) are nonnegative. If

$c \leq P$ together with nonnegative $\mathbf{D}$ and $\mathbf{d}$ then $\boldsymbol{\Upsilon}(\tilde{\mathbf{U}})$ in (5.2.33) becomes a nonnegative matrix. This completes the proof. ■

C. Proof of Lemma 2

$\tilde{\mathbf{U}}^{(0)}$ (i.e., obtained using $\mathbf{q}^{(0)}$) is assumed to satisfies the conditions in (5.2.34) and (5.2.35). Hence, a feasible power allocation, $\mathbf{q}^{(1)}$, can be obtained at the first iteration. Since there is a feasible solution at the first iteration, real time secondary users achieve their target SINRs, $\gamma_k,\ k = 1 \ldots K1$, and all non-real time secondary users achieve equal SINRs, $1/\lambda^{(1)}$. In the second iteration, $\tilde{\mathbf{q}}^{(1)}$ has been used with (5.2.36) to obtain beamformers $\tilde{\mathbf{U}}^{(1)}$. In a conventional SINR balancing problem, SINR of each user increases monotonically at each iteration [41]. However, some users (i.e., real time secondary users) maintain the same SINR (i.e., target SINR) at each iteration. Hence, it is apparent that, in the second iteration, there is a feasible solution available and the non-real time secondary users achieve SINRs not less than the ones that achieved in the first iteration (i.e., $1/\lambda^{(2)} \geq 1/\lambda^{(1)}$).

At the end of the second iteration, the equation (5.2.22) is updated as

$$\tilde{\mathbf{q}}_A^{(2)} = \mathbf{D}_A^{(1)}\boldsymbol{\Psi}_A^{(1)}\tilde{\mathbf{q}}_A^{(2)} + \mathbf{D}_A^{(1)}\mathbf{b}_A^{(1)} + \mathbf{D}_A^{(1)}\boldsymbol{\Psi}_B^{(1)}\tilde{\mathbf{q}}_B^{(2)}, \tag{6.2.4}$$

Since there exists a feasible solution at the end of the second iteration, $\tilde{\mathbf{q}}_A^{(2)}$, $\tilde{\mathbf{q}}_B^{(2)}$, $\mathbf{D}_A^{(1)}\mathbf{b}_A^{(1)}$ and $\mathbf{D}_A^{(1)}\boldsymbol{\Psi}_B^{(1)}\tilde{\mathbf{q}}_B^{(2)}$ are nonnegative. Hence,

$$\tilde{\mathbf{q}}_A^{(2)} \ \geq\ \mathbf{D}_A^{(1)}\boldsymbol{\Psi}_A^{(1)}\tilde{\mathbf{q}}_A^{(2)}. \tag{6.2.5}$$

The inequality (6.2.5) results into the following inequality for the normalized $\mathbf{q}_A^{(2)}$

$$1 \ \geq\ {\mathbf{q}_A^{(2)}}^T\mathbf{D}_A^{(1)}\boldsymbol{\Psi}_A^{(1)}\mathbf{q}_A^{(2)}, \tag{6.2.6}$$

where $\mathbf{D}_A^{(1)}\mathbf{\Psi}_A^{(1)}$ is a nonnegative matrix. Maximum value of ${\mathbf{q}_A^{(2)}}^T\mathbf{D}_A^{(1)}\mathbf{\Psi}_A^{(1)}\mathbf{q}_A^{(2)}$ is equal to $\rho(\mathbf{D}_A^{(1)}\mathbf{\Psi}_A^{(1)})$ when $\mathbf{q}_A^{(2)}$ is the eigenvector corresponding to the largest eigenvalue of matrix $\mathbf{D}_A^{(1)}\mathbf{\Psi}_A^{(1)}$. From the inequality in (6.2.6),

$$1 \geq \rho(\mathbf{D}_A^{(1)}\mathbf{\Psi}_A^{(1)}). \tag{6.2.7}$$

From equation (6.2.4), the following inequality holds

$$\tilde{\mathbf{q}}_A^{(2)} \geq \mathbf{D}_A^{(1)}\mathbf{\Psi}_A^{(1)}\tilde{\mathbf{q}}_A^{(2)} + \mathbf{D}_A^{(1)}\mathbf{b}_A^{(1)}. \tag{6.2.8}$$

From inequality (6.2.8), we can write the following inequalities using Lemma 1, and the equations (6.2.7), (5.2.23), (5.2.30):

$$\begin{aligned} \tilde{\mathbf{q}}_A^{(2)} &\geq (\mathbf{I} - \mathbf{D}_A^{(1)}\mathbf{\Psi}_A^{(1)})^{-1}\mathbf{D}_A^{(1)}\mathbf{b}_A^{(1)}, \\ P \geq \boldsymbol{\sigma}_A^T\tilde{\mathbf{q}}_A^{(2)} &\geq \boldsymbol{\sigma}_A^T(\mathbf{I} - \mathbf{D}_A^{(1)}\mathbf{\Psi}_A^{(1)})^{-1}\mathbf{D}_A^{(1)}\mathbf{b}_A^{(1)} = c^{(1)}. \end{aligned} \tag{6.2.9}$$

The required conditions are satisfied in (6.2.7) and (6.2.9). From Lemma 1, $\mathbf{\Upsilon}(\tilde{\mathbf{U}})$ is a nonnegative matrix. This completes the proof. ■

D. Proof of Lemma 3

After the convergence of BA described in Table 5.1, $\rho(\mathbf{D}_A^*\mathbf{\Psi}_A^*) \leq 1$. Hence, after the convergence, the vector $\mathbf{d}^*$ and the matrix $\mathbf{D}^*$ are nonnegative in (5.2.25). The following inequality can be obtained from (5.2.25):

$$\begin{aligned} \lambda^*\tilde{\mathbf{q}}_B &\geq \mathbf{D}^*\tilde{\mathbf{q}}_B, \\ \lambda^* &\geq \rho(\mathbf{D}^*). \end{aligned} \tag{6.2.10}$$

Since $\rho(\mathbf{D}_A^*\mathbf{\Psi}_A^*) \leq 1$, the matrix $\mathbf{D}$ in (5.2.26) can be written as follows using the Neumann series:

$$\mathbf{D}^* = \mathbf{D}_B^*\mathbf{\Psi}_D^* + \mathbf{D}_B^*\mathbf{\Psi}_C^*\sum_{n=0}^{\infty}(\mathbf{D}_A^*\mathbf{\Psi}_A^*)^n\mathbf{D}_A^*\mathbf{\Psi}_B^*. \tag{6.2.11}$$

For a diagonal matrix $\mathbf{A}$, $\rho(\mathbf{AB}) = \rho(\mathbf{AB}^T)$ [60]. From this property, the following inequalities can be obtained from equation (6.2.10):

$$\begin{aligned}
\lambda^* &\geq \rho[\mathbf{D}_B^*(\mathbf{\Psi}_D^* + \mathbf{\Psi}_C^*\sum_{n=0}^{\infty}(\mathbf{D}_A^*\mathbf{\Psi}_A^*)^n\mathbf{D}_A^*\mathbf{\Psi}_B^*)], \\
&= \rho[\mathbf{D}_B^*(\mathbf{\Psi}_D^* + \mathbf{\Psi}_C^*\sum_{n=0}^{\infty}(\mathbf{D}_A^*\mathbf{\Psi}_A^*)^n\mathbf{D}_A^*\mathbf{\Psi}_B^*)^T], \\
&= \rho[\mathbf{D}_B^*(\mathbf{\Psi}_D^{*T} + \mathbf{\Psi}_B^{*T}\sum_{n=0}^{\infty}(\mathbf{D}_A^*\mathbf{\Psi}_A^{*T})^n\mathbf{D}_A^*\mathbf{\Psi}_C^*)^T], \\
&= \rho(\mathbf{D}_D^*).
\end{aligned} \tag{6.2.12}$$

However, $\mathbf{D}_D^*$ is a nonnegative matrix. From Lemma 1 and the Neumann series, $(\lambda^*\mathbf{I}-\mathbf{D}_D^*)$ is nonsingular and $(\lambda^*\mathbf{I}-\mathbf{D}_D^*)^{-1} = \sum_{n=0}^{\infty}(\mathbf{D}_D^*)^n$ is a nonnegative matrix. This completes the proof. ■

E. Proof of Lemma 4

In equation (6.2.16), $\mathbf{D}_A^R\mathbf{b}_A^R$ is a nonnegative vector. Hence, the following inequality holds:

$$\lambda_A^R\mathbf{q}_A^R \geq \mathbf{D}_A^R\mathbf{\Psi}_A^R\mathbf{q}_A^R. \tag{6.2.13}$$

From inequality (6.2.13), the following inequality holds for the normalized $\mathbf{q}_A^R$

$$\lambda_A^R \geq \mathbf{q}_A^{R^T}\mathbf{D}_A^R\mathbf{\Psi}_A^R\mathbf{q}_A^R. \tag{6.2.14}$$

$\mathbf{D}_A^R \mathbf{\Psi}_A^R$ is a nonnegative matrix and $\mathbf{q}_A^R$ is a nonnegative power allocation vector. The largest eigenvalue of irreducible nonnegative matrix and the corresponding eigenvector are strictly positive [60]. Largest value of ${\mathbf{q}_A^R}^T \mathbf{D}_A^R \mathbf{\Psi}_A^R \mathbf{q}_A^R$ is equal to $\rho(\mathbf{D}_A^R \mathbf{\Psi}_A^R)$ if and only if $\mathbf{q}_A^R$ is a positive eigenvector corresponding to the largest eigenvalue of the matrix $\mathbf{D}_A^R \mathbf{\Psi}_A^R$. Hence,

$$\lambda_A^R \geq \rho(\mathbf{D}_A^R \mathbf{\Psi}_A^R). \tag{6.2.15}$$

The problem is feasible, only if $\lambda_A^R \leq 1$ which implies that $\rho(\mathbf{D}_A^R \mathbf{\Psi}_A^R) \leq 1$.

Convergence of the BA algorithm in Table 5.2 is proved by assuming $\mathbf{\Upsilon}(\tilde{\mathbf{U}}^{(0)})$ is a non-negative matrix. To complete the proof, the following needs to be proved: initial power allocation $\mathbf{q}^{(0)}$ provide non-negative matrix $\mathbf{\Upsilon}(\tilde{\mathbf{U}}^{(0)})$. For all feasible problems there exists a feasible power allocation $\mathbf{q}_A^R$. Note that, when $\mathbf{q}_A^R$ and $\mathbf{0}_{K-K1}$ are used as an initial power allocation for the real time secondary users and the non-real time secondary users respectively, the same matrices $\mathbf{D}_A^R$ and $\mathbf{\Psi}_A^R$ and vector $\mathbf{b}_A^R$ can be obtained for $\mathbf{D}_A^{(0)}$, $\mathbf{\Psi}_A^{(0)}$ and $\mathbf{b}_A^{(0)}$ at step 6 of the BA algorithm in first iteration. Hence, $\rho(\mathbf{D}_A^{(0)} \mathbf{\Psi}_A^{(0)}) = \rho(\mathbf{D}_A^R \mathbf{\Psi}_A^R) \leq 1$, which satisfies the first sufficient condition (5.2.34) for the non-negativity of $\mathbf{\Upsilon}(\tilde{\mathbf{U}}^{(0)})$. For the feasible problem, equation (6.2.16) satisfies the following inequalities:

$$\begin{aligned} \mathbf{q}_A^R &\geq \mathbf{D}_A^R \mathbf{\Psi}_A^R \mathbf{q}_A^R + \mathbf{D}_A^R \mathbf{b}_A^R, \\ (\mathbf{I} - \mathbf{D}_A^R \mathbf{\Psi}_A^R)\mathbf{q}_A^R &\geq \mathbf{D}_A^R \mathbf{b}_A^R. \end{aligned} \tag{6.2.16}$$

But from Lemma 1 if $\rho(\mathbf{D}_A^R \mathbf{\Psi}_A^R) \leq 1$ then $(\mathbf{I} - \mathbf{D}_A^R \mathbf{\Psi}_A^R)^{-1}$ is a nonnegative matrix. Hence, the inequality (6.2.16) satisfies the following inequality:

$$\mathbf{q}_A^R \geq (\mathbf{I} - \mathbf{D}_A^R \mathbf{\Psi}_A^R)^{-1} \mathbf{D}_A^R \mathbf{b}_A^R.$$

From (5.2.23) and (5.2.30),

$$P \geq \boldsymbol{\sigma}_A^T \mathbf{q}_A^R \ \geq \ \boldsymbol{\sigma}_A^T (\mathbf{I} - \mathbf{D}_A^R \boldsymbol{\Psi}_A^R)^{-1} \mathbf{D}_A^R \mathbf{b}_A^R = c,$$

which satisfies the second sufficient condition (5.2.35) for the non-negativity of $\boldsymbol{\Upsilon}(\tilde{\mathbf{U}}^{(0)})$, and it concludes the proof. ■

References

[1] K. Cumanan, R. Krishna, V. Sharma, and S. Lambotharan, "Robust interference control techniques for multiuser cognitive radios using worst-case performance optimization," in *Proc. Asilomar Conf. Sign., Syst. and Comp.*, pp. 378 –382, Pacific Grove, CA, Oct. 2008.

[2] T. S. Rappaport, *Wireless communications: Principals and Practice.* Upper Saddle River, NJ: Prentice Hall, 2 ed., 1996.

[3] J. Proakis and M. Salehi, *Digital Communications.* McGraw-Hill Science, 5 ed., 2007.

[4] H. Holma and A. Toskala, *WCDMA FOR UMTS - HSPA EVOLUTION AND LTE.* John Wiley & Sons, Ltd, 4 ed., 2007.

[5] Federal Communications Commission, *ET Docket No 03-222 Notice of proposed rule making and order.* Dec. 2003.

[6] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[7] T. A. Weiss and F. K. Jondral, "Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency," *IEEE Commun. Mag.*, vol. 42, no. 3, pp. S8-S14, Mar. 2004.

[8] J. Mitola and G. Q. Maguire, "Cognitive radios: Making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[9] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[10] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 28–40, Feb. 2008.

[11] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1128–1140, Mar. 2009.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory.* NJ, Prentice-Hall: Englewood Cliffs, 1998.

[13] D. Cabric, A. Tkachenco, and R. W. Brodersen, "Experimental study of spectrum sensing based on energy detection and network cooperation," in *Proc. ACM 1st Int. Workshop on Technology and Policy for Accessing Spectrum (TAPAS)*, Aug. 2006.

[14] S. Enserink and D. Cochran, "A cyclostationary feature detector," in *Proc. Asilomar Conf. Sign., Syst. and Comp., Pacific Grove, CA*, pp. 806–810, Oct. 1994.

[15] S. Srinivasa and S. A. Jafar, "The throughput potential of cognitive radio: A theoretical perspective," *IEEE Commun. Mag.*, vol. 45, pp. 73 –79, May 2007.

[16] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications.* Cambridge University Press, 2005.

[17] H. Liu and G. Li, *OFDM-Based Broadband Wireless Networks - Design and Optimization.* NY: Wiley, Nov. 2005.

[18] E. Telatar, "Capacity of multi-antenna Gaussian channels," in *European Transactions on Telecommunications*, vol. 10, pp. 585–595, Nov./Dec. 1999.

[19] E. Lawrey, "Multiuser OFDM," in *Proc. International Symposium on Signal Processing and its Applications*, Brisbane, Australia,1999.

[20] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Ael. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[21] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G Evolution-HSPA and LTE for Mobile Broadband.* ELSEVIER, 2 ed., 2008.

[22] G. Bottomley, T. Ottosson, and Y.-P. Wang, "A generalized RAKE receiver for interference suppression," *IEEE J. Sel. Areas in Commun.*, vol. 18, pp. 1536 –1545, Aug. 2000.

[23] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas in Commun.*, vol. 21, pp. 171 –178, Feb. 2003.

[24] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1150 – 1158, Nov. 2003.

[25] S. K. Lai, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Adaptive trellis coded MQAM and power optimization for OFDM transmission," in *Proc. IEEE Vehicular Technology Conf. (VTC99)*, May Houston, TX, 1999.

[26] Z. Mao and X. Wang, "Efficient optimal and suboptimal radio resource allocation in OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 440–445, Feb. 2008.

[27] I. Kim, I. S. Park, and Y. H. Lee, "Use of linear programming for dynamic subcarrier and bit allocation in multiuser OFDM," *IEEE Trans. Vehic. Tech.*, vol. 55, pp. 1195 – 1207, Jul. 2006.

[28] C. H. Papadimitrou and K. Steiglitz, *Combinatorial Optimization: Algorithm and Complexity.* Prentice Hall, Inc., New Jersey, 1982.

[29] G. L. Nemhauser and L. A. Wolsey, *Integer and combinatorial optimization.* John Wiley & Sons, New york, 1998.

[30] D. A. Bader, W. E. Hart, and C. A. Phillips, *Parallel Algorithm Design for Branch and Bound.* H.J. Greenberg, editor, in Tutorials on Emerging Methodologies and Applications in Operations Research,: Kluwer Academic Press, Chapter 5, pp. 1-44, 2004.

[31] C. Turgu and C. Toker, "A low complexity resource allocation algorithm for OFDMA systems," in *Proc. IEEE Statistical Signal Processing Workshop*, pp. 689–692, Cardiff, UK, Aug. 2009.

[32] B. Krongold, K. Ramchandran, and D. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Trans. Commun.*, vol. 48, pp. 23 –27, Jan. 2000.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, Cambridge, UK, 2004.

[34] Y. Ye, *Interior Point Algorithms. Theory and Analysis.* John Wiley & Sons, 1997.

[35] Y. J. Zhang and K. B. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1566 – 1575, Sep. 2004.

[36] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. Vehic. Tech. Conf.*, pp. 1085 –1089, Tokyo, Japan, 2000.

[37] K. Seong, M. Mohseni, and J. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *Proc. IEEE International Symposium on Information Theory*, pp. 1394 –1398, Jul. 2006.

[38] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726-2737, Nov. 2005.

[39] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[40] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. on Signal Process.*, vol. 51, pp. 2257 – 2269, Sep. 2003.

[41] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. on Vehicular Technol.*, vol. 53, pp. 18 – 28, Jan. 2004.

[42] D. J. Love and R. W. Heath, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[43] T. Pande, D. J. Love, and J. V. Krogmeier, "Reduced feedback MIMO-OFDM precoding and antenna selection," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2284–2293, May 2007.

[44] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Trans. Inform. Theory*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[45] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. 37th Annual Allerton Conference on Communications, Control and Computing*, pp. 987–996, Sep. 1999.

[46] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit

beamforming," in *Proc. Handbook of Antennas in Wireless Communications*, L. C. Godara, Editor, CRC Press, 2001.

[47] J. Lofberg, "Yalmip: A toolbox for modelling and optimizationin MATLAB," in *Proc. IEEE Int. Symp. on Comp. Aided Control Sys. Design*, pp. 284–289, Taipei, Sep. 2004.

[48] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming." *Optimization Methods and Software*, Feb. 2007. Available [online]: http://www.stanford.edu/ boyd/cvx/V.1.0RC3.

[49] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, 1999. Special issue on Interior Point Methods (CD supplement with software).

[50] M. Schubert and H. Boche, "Iterative multiuser uplink and downlink beamforming under SINR constraints," *IEEE Trans. Signal Process.*, vol. 53, pp. 2324 – 2334, Jul. 2005.

[51] H. Boche and M. Schubert, "A general duality theory for uplink and downlink beamforming," in *Proc. IEEE Veh. Tech. Conf. (VTC) Fall*, Vancouver, Canada, Sep. 2002.

[52] H. Boche and M. Schubert, "Solution of the SINR downlink beamforming problem," in *Proc. 36th Conf. Inform. Sci. Syst. (CISS)*, Princeton, USA, Mar. 2002.

[53] H. Boche and M. Schubert, "A unifying theory for uplink and downlink multiuser beamforming," in *Proc. IEEE Intern. Zurich Seminar*, pp. 27–1–27–6, Zurich, Switzerland, Feb. 2002.

[54] D. Palomar and M. Lagunas, "Joint transmit-receive space-time equal-

ization in spatially correlated MIMO channels: a beamforming approach," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 730–743, Jun. 2003.

[55] D. Palomar, M. Lagunas, and J. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[56] D. Palomar, J. Cioffi, and M. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[57] A. Goldsmith, *Wireless Communications.* Cambridge University Press, 2005.

[58] G. Strang, *Linear Algebra and its Applications.* Fourth edition: Thomson Brooks/Cole, 2006.

[59] W. Yang and G. Xu, "Optimal downlink power assignment for smart antenna systems," in *Proc. IEEE ICASSP*, vol. 6, pp. 3337–3340, Seattle, WA, May 1998.

[60] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra.* SIAM, 2000.

[61] F. Rashid-Farrokhi, K. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1437–1450, Oct. 1998.

[62] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[63] P. Tseng, "Further results on approximating nonconvex quadratic optimization by semidefinite programming relaxation," *SIAM J. Optim.*, vol. 14, pp. 268–283, Jul. 2003.

[64] E. Karipidis, N. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, Mar. 2008.

[65] E. Karipidis, N. Sidiropoulos, and Z.-Q. Luo, "Far-field multicast beamforming for uniform linear antenna arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4916–4927, Oct. 2007.

[66] Y. J. Zhang and K. B. Letaief, "An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems," *IEEE Trans. Commun.*, vol. 13, no. 4, pp. 38 – 47, Aug. 2006.

[67] I. Koutsopoulos and L. Tassiulas, "Adaptive resource allocation in SDMA-based wireless broadband networks with OFDM signaling," in *Proc. IEEE INFOCOM*, vol. 3, pp. 1376–1385, 2002.

[68] H. Yang, "A road to future broadband wireless access: MIMO-OFDM-based air interface," *IEEE Commun. Mag.*, vol. 43, pp. 53 –60, Jan. 2005.

[69] A. Attar, O. Holland, M. R. Nakhai, and A. H. Aghvami, "Inteference-limited resource allocation for cognitive radio in orthogonal frequency division multiplexing networks," *IET J. Commun.*, vol. 2, no. 6, pp. 806–814, 2008.

[70] G. Bansal, J. Hossain, and V. K. Bhargava, "Adaptive power loading for OFDM-based cognitive radio system," in *Proc. IEEE Int. Conf. Commun.*, pp. 5137–5142, Glasgow, Scotland, Jun. 2007.

[71] S. Wang, "Efficient resource allocation algorithm for cognitive OFDM systems," *IEEE Commun. Lett.*, vol. 14, pp. 725 –727, Aug. 2010.

[72] Y. Zhang and C. Leung, "Resource allocation in an OFDM-based cog-

nitive radio system," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1928-1931, Jul. 2009.

[73] Y. Rahulamathavan, K. Cumanan, L. Musavian, and S. Lambotharan, "Optimal subcarrier and bit allocation techniques for cognitive radio networks using integer linear programming," in *Proc. IEEE Statistical Signal Processing Workshop*, pp. 293 – 296, Cardiff, UK, Aug. 2009.

[74] Y. Rahulamathavan, S. Lambotharan, C. Toker, and A. B. Gershman, "A suboptimal recursive optimization framework for adaptive resource allocation in spectrum sharing networks," *submitted to IET J. Signal Process. (provissionally accepted with revision)*, 2011.

[75] R. Zhang and Y. C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE J. Sel. Topics in Signal Process.*, vol. 2, pp. 88–102, Feb. 2008.

[76] S. Yiu, M. Vu, and V. Tarokh, "Interference reduction by beamforming in cognitive networks," in *Proc. IEEE GLOBECOM, New Orleans, LO*, pp. 1–6, Dec. 2008.

[77] L. Zhang, Y.-C. Liang, and Y. Xin, "Joint beamforming and power allocation for multiple access channels in cognitive radio networks," *IEEE J. Sel. Topics Commun.*, vol. 26, pp. 38 – 51, Jan. 2008.

[78] K. Phan, S. Vorobyov, N. Sidiropoulos, and C. Tellambura, "Spectrum sharing in wireless networks via QoS-aware secondary multicast beamforming," *IEEE Trans. Signal Process.*, vol. 57, pp. 2323 –2335, Jun. 2009.

[79] H. Islam, Y. C. Liang, and A. Hoang, "Joint power control and beamforming for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2415–2419, Jul. 2008.

[80] Y. Rahulamathavan, K. Cumanan, and S. Lambotharan, "Optimal resource allocation techniques for MIMO-OFDMA based cognitive radio networks using integer linear programming," in *Proc. IEEE 11th International Workshop on Signal Process. Advances Wireless Commun.*, pp. 1 –5, Jun. 2010.

[81] Y. Rahulamathavan, K. Cumanan, R. Krishna, and S. Lambotharan, "Adaptive subcarrier and bit allocation techniques for MIMO-OFDMA based uplink cognitive radio networks," in *Proc. First UK-India International Workshop on Cognitive Wireless Systems*, pp. 1 –5, Dec. 2009.

[82] L. Cao and H. Zheng, "Distributed spectrum allocation via local bargaining," in *Proc. IEEE SECON'05*, pp. 119–127, Santa Clara, California, USA, Sept. 2005.

[83] S. Haykin, *Communication System.* John Wiley and Sons, New york, 4 ed., 1994.

[84] K. B. Letaief and Y. J. Zhang, "Dynamic multiuser resource allocation and adaptation for wireless systems," *IEEE Trans. Wireless Commun.*, vol. 53, no. 1, pp. 107 – 116, Jan. 2005.

[85] Y. F. Chen and J. W. Chen, "A fast subcarrier, bit, and power allocation algorithm for multiuser OFDM-based systems," *IEEE Trans. Vehic. Tech.*, vol. 57, pp. 873 –881, Mar. 2008.

[86] G. Kulkarni, S. Adlakha, and M. Srivastava, "Subcarrier allocation and bit loading algorithms for OFDMA-based wireless networks," *IEEE Trans. Mob. Comp.*, vol. 4, pp. 652 – 662, Nov.-Dec. 2005.

[87] C. Mohanram and S. Bhashyam, "A sub-optimal joint subcarrier and power allocation algorithm for multiuser OFDM," *IEEE Commun. Lett.*, vol. 9, pp. 685 – 687, Aug. 2005.

[88] I. C. Wong, Z. Shen, B. L. Evans, and J. G. Andrews, "A low complexity algorithm for proportional resource allocation," in *Proc. IEEE Workshop Sig. Proces. Sys.*, 2004.

[89] C. Mohanram and S. Bhashyam, "A sub-optimal joint subcarrier power allocation algorithm for multiuser OFDM," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 685-687, Aug. 2005.

[90] K. Li and P. Kam, "On the error performance of 8-QAM involving a $\pi/2$-wedge-shaped decision region," in *Proc. Joint Conf. Fourth International Conference on Inf., Commun. and Sign. Process., and Fourth Pacific Rim Conference on Multimedia.*, pp. 884 – 887, Dec. 2003.

[91] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," in *Proc. Wireless Personal Communications*, vol. 6, pp. 311–335, Mar. 1998.

[92] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, pp. 461 – 471, Feb. 2004.

[93] E. Matskani, N. D. Sidiropoulos, Z. Q. Luo, and L. Tassiulas, "Convex approximation techniques for joint multiuser downlink beamforming and admission control," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2682–2693, Jul. 2008.

[94] J. Li and P. Stoica, *Robust Adaptive Beamforming.* United Kingdom: Wiley Series in Telecommunications and Signal Processing, 2006.

[95] K. Hamdi, W. Zhang, and K. B. Letaief, "Joint beamforming and scheduling in cognitive radio networks," in *Proc. IEEE GLOBECOM*, pp. 2977–2981, Washington, DC, Nov. 2007.

[96] K. Cumanan, L. Musavian, S. Lambotharan, and A.B.Gershman, "SINR balancing technique for downlink beamforming in cognitive radio networks," *IEEE Signal Process. Lett.*, vol. 17, pp. 133 –136, Feb. 2010.

[97] K. Cumanan, R. Krishna, Z. Xiong, and S. Lambotharan, "SINR balancing technique and its comparison to semidefinite programming based QoS provision for cognitive radios," in *Proc. IEEE 69th VTC Spring*, pp. 1 –5, Apr. 2009.

[98] L. Zhang, R. Zhang, Y. C. Liang, Y. Xin, and H. V. Poor, "On Gaussian MIMO BC-MAC duality with multiple transmit covariance constraints," *submitted to IEEE Trans. Inform. Theory.*, Available online at http://arxiv.org/PS cache/arxiv/pdf/0809/0809.4101v1.pdf.

[99] A. Fehske, F. Richter, and G. Fettweis, "SINR balancing for the multiuser downlink under general power constraints," in *Proc. IEEE GLOBECOM 2008*, pp. 1 –6, Dec. 2008.

[100] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *IEEE Trans. Inform. Theory*, vol. 55, pp. 3945 –3958, Sep. 2009.

[101] A. Ghasemi and E. S. Sousa, "Fundamental limits of spectrum-sharing in fading environments," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 649 –658, Feb. 2007.

[102] L. Zhang, Y. Xin, and Y. C. Liang, "Weighted sum rate optimization for cognitive radio MIMO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 2950 –2959, Jun. 2009.

[103] L. Zhang, Y. Xin, Y. C. Liang, and H. V. Poor, "Cognitive multiple access channels: optimal power allocation for weighted sum rate maximization," *IEEE Trans. Commun.*, vol. 57, pp. 2754–2762, Sep. 2009.

[104] G. Zheng, K.-K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, pp. 4871 –4881, Dec. 2009.

[105] S. Boyd, L. Xiao, and A. Mutapici, *Subgradient methods.* Available online at http://mit.edu/6.976/www/notes/subgrad-method.pdf, Oct. 2003.

[106] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, pp. 1310 –1322, Jul. 2006.

[107] V. Y. Pan and Z. Q. Chen, "The complexity of the matrix eigenproblem," in *Proc. 31st Ann. ACM Symp. on Theory of Computing*, ACM Press, New York, 1999.

[108] N. J. S. Vishwanath and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[109] E. Jorswieck and H. Boche, "Rate balancing for the multi-antenna gaussian broadcast channel," in *IEEE Seventh International Symposium on Spread Spectrum Techniques and Applications*, 2002.

[110] T. Sartenaer, L. Vandendorpe, and J. Louveaux, "Balanced capacity of wireline multiuser channels," *IEEE Trans. Commun.,*, vol. 53, pp. 2029 – 2042, Dec. 2005.

[111] J. Lee and N. Jindal, "Symmetric capacity of MIMO downlink channels," in *Proc. IEEE International Symposium on Inform. Theory (ISIT)*, pp. 1031 –1035, Seattle, WA, Jul. 2006.

[112] P. Tejera, W. Utschick, J. Nossek, and G. Bauch, "Rate balancing in multiuser MIMO OFDM systems," *IEEE Trans. Commun.*, vol. 57, pp. 1370 –1380, May 2009.

[113] D. Tse and S. Hanly, “Multiaccess fading channels-part I: polymatriod structure, optimal resource allocation and throughput capacities,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 2796–2815, Nov. 1998.