University of Alberta

RESOURCE ALLOCATION FOR OFDMA-BASED MULTICAST WIRELESS SYSTEMS

by

Duy Trong Ngo

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science

in

Communications

Department of Electrical and Computer Engineering

Fall 2009
Edmonton, Alberta

Examining Committee

Dr. Chintha Tellambura, Electrical and Computer Engineering

Dr. Mike MacGregor, Computing Science

Dr. Hai Jiang, Electrical and Computer Engineering

To my parents and my sweetheart Thao,

With all my love.

Abstract

Regarding the problems of resource allocation in OFDMA-based wireless communication systems, much of the research effort mainly focuses on finding efficient power control and subcarrier assignment policies. With systems employing multicast transmission, the available schemes in literature are not always applicable. Moreover, the existing approaches are particularly inaccessible in practical systems in which there are a large number of OFDM subcarriers being utilized, as the required computational burden is prohibitively high. The ultimate goal of this research is therefore to propose affordable mechanisms to flexibly and effectively share out the available resources in multicast wireless systems deploying OFDMA technology. Specifically, we study the resource distribution problems in both conventional and cognitive radio network settings, formulating the design problems as mathematical optimization programs, and then offering the solution methods. Suboptimal and optimal schemes with high performance and yet of acceptable complexity are devised through the application of various mathematical optimization tools such as genetic algorithm and Lagrangian dual optimization. The novelties of the proposed approaches are confirmed, and their performances are verified by computer simulation with the presentation of numerical examples to support the findings.

Acknowledgements

My special thanks are given to my supervisor, Dr. Chintha Tellambura, for his enthusiasm, patience and kindness. Without his knowledge and constant supervision, this work would not have been successful. As well, I would like to thank Dr. Ha Nguyen from the University of Saskatchewan, Canada for his research collaboration and support over the years. Many thanks go to Dr. Tho Le-Ngoc from McGill University, Canada whose invaluable comments have helped improved the quality of my research results. I also wish to express my appreciation to the members of my thesis committee. Their comments and feedback have been very useful for refining this thesis.

I will surely remember all my labmates at the *i*CORE Wireless Communications Laboratory who, in one way or another, have made my two-year stay here a pleasurable and memorable one. I am definitely thankful to the Alberta Ingenuity Foundation and the Informatics Circle of Research Excellence, who have offered me financial support in the form of a full scholarship for graduate studies. Without that generosity, I would not be able to enjoy the advanced research in the province of Alberta.

My heartfelt gratitude goes out to my family for their endless encouragement and support. Lastly, it is my greatest happiness to meet and fall in love with Thao. She is always there, with her total understanding, her everlasting love and sacrifice. For that, I am deeply thankful.

Table of Contents

List of Tables

List of Figures

List of Symbols

$\arg$	argument
$\mathbf{dom} f$	domain of function f
$\lim$	limit
$\inf$	infimum
$\max$	maximum
$\min$	minimum
$[x]^+$	$\max(x, 0)$
$\mathcal{E}\{\cdot\}$	statistical expectation
$\forall$	for all
$\cap$	set intersection
$\cup$	set union
x	scalar x
$\mathbf{x}$	vector $\mathbf{x}$
$\mathbf{x}^T$	transpose of vector $\mathbf{x}$
$\mathbf{0}$	zero vector $[0, 0, \cdots, 0]$
$\mathbf{1}$	unit vector $[1, 1, \cdots, 1]$
$\mathcal{Z}_+$	set of all positive integer numbers
$\mathcal{R}$	set of all real numbers
$\mathcal{S} \backslash m$	set $\mathcal{S}$ without member m

Acronyms

AWGN Additive White Gaussian Noise

BER bit error rate

BS base station

CDMA Code Division Multiple Access

CSINR channel signal-to-interference-plus-noise ratio

CSNR channel signal-to-noise ratio

CSI channel state information

FDMA Frequency Division Multiple Access

FFT Fast Fourier Transform

GA Genetic Algorithm

i.i.d independent and identically distributed

IFFT Inverse Fast Fourier Transform

ISI inter-symbol interference

KKT Karush–Kuhn–Tucker

MDC Multiple Description Coding

NP-hard Non-deterministic Polynomial-time hard

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

pdf probability density function

PSD power spectral density

PSK Phase Shift Keying

QAM Quadrature Amplitude Modulation

QoS Quality-of-Service

s.t. subject to

SINR signal-to-interference-plus-noise ratio

SNR signal-to-noise ratio

TDMA Time Division Multiple Access

Chapter 1

Introduction

1.1 Background

In recent years, Orthogonal Frequency Division Multiplexing (OFDM) [1, 2], a multiplexing scheme utilized as a digital multi-carrier modulation technique, has become a popular advanced technology for wideband digital communication and also been considered a greatly promising candidate for the next generation networks. The basic idea of OFDM is to divide the transmitted bitstream into many different substreams and send these over a large number of closely-spaced orthogonal subchannels. Each subchannel can be represented by one subcarrier, and effectively one substream of data is transmitted through one subcarrier whereas individual subcarriers are modulated with a conventional modulation scheme such as Quadrature Amplitude Modulation (QAM) or Phase Shift Keying (PSK) at a low symbol rate. It should be noted that while the corresponding subchannel bandwidth is much less than the total system bandwidth, the total data rates achieved by OFDM maintain similar to conventional single-carrier modulation schemes subject to the same total bandwidth. The key advantage of employing OFDM over single-carrier schemes is its ability to easily adapt to severe channel conditions, especially in frequency-selective fading environments due to multipath propagation, without requiring complex equalization. This is because OFDM can be realized as transmitting many slowly-modulated narrowband signals rather than one rapidly-modulated wideband signal. The low symbol rate allows handling of time-spreading and eliminating inter-symbol interference (ISI) possible since it is now affordable to provide a guard interval between symbols. Furthermore, OFDM systems offer high spectral efficiency with an efficient implementation using Fast Fourier Transform (FFT). All those mentioned advantages have put OFDM forward as a commonly chosen scheme for wideband communication, whether wireless or over copper wires, used

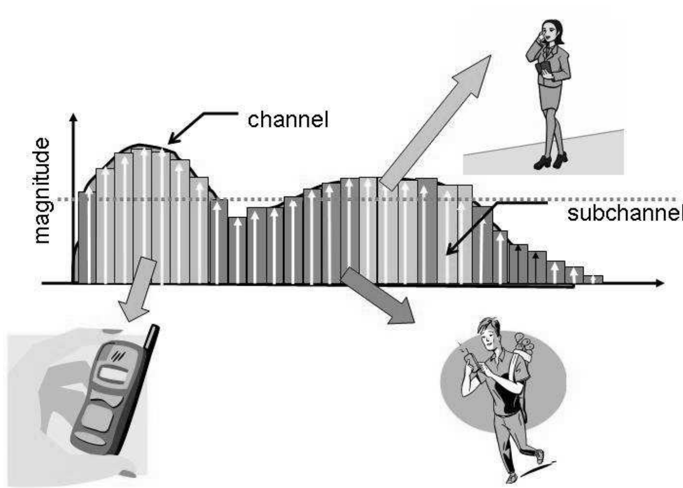


Figure 1.1: Orthogonal Frequency Division Multiple Access.

in various applications such as digital television and audio broadcasting, wireless networking and broadband internet access. The ideal structure of an OFDM system is described in detail in Appendix A.

Being employed for transmitting one bit stream over one communication channel using one sequence of OFDM symbols, OFDM in its primary form is considered a digital modulation technique rather than a multi-user channel access scheme. Nevertheless, it can be combined with multiple access using time, frequency or coding separation of the users. In Orthogonal Frequency Division Multiple Access (OFDMA), frequency-division multiple access is achieved by assigning different OFDM subchannels to different users, provided that each subchannel is allocated to at most one user at a time (see Fig. 1.1). In effect, differentiated Quality-of-Service (QoS) is supported by assigning various number of subcarriers to different users in a similar fashion as in Code Division Multiple Access (CDMA) technique, and hence complex packet scheduling or media access control can be avoided. Importantly, thanks to the independence in the fading channel states of different users, there is also an opportunity to take advantage of frequency selectivity and perform channel aware scheduling and resource allocation.

On the other hand, it is known that unicast is the sending of information to one single destination. As transmission with a unicast service is inherently point-to-point, it becomes necessary that the source replicates several identical data flows in order to transmit them to each of the receivers. Accordingly, bandwidth waste is generated. In contrast, broadcast is used to send the same content to all destinations indiscriminately. This mechanism also re-

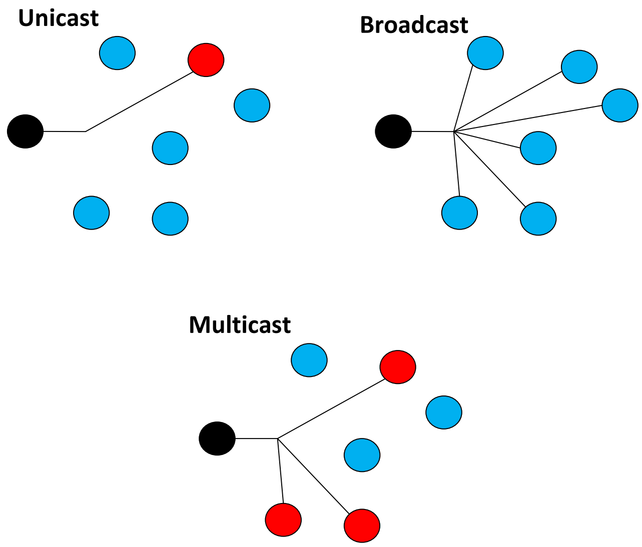


Figure 1.2: Transmission mechanisms.

sults in a waste of resource since it implies transporting the data to all the network stations, even if the number of receivers wishing to have that content is reduced. A more efficient way to transmit data is to provide a multicast service in which a single flow of data, originating from a given source, may be sent simultaneously to several interested receivers. With multicast, the source sends only one single copy of the data packets to a multicast group address. The network infrastructure replicates these packages in an intelligent way, directing the data according to the topology of receivers interested in that piece of information. When compared with the unicast and broadcast transmissions, multicast counterpart shows its clear advantages in numerous one-to-many and many-to-many applications such as real-time audio and video conferences, live concerts, distribution of software, news and market information, database updating, distance learning, distributed games, and so on. The three discussed mechanisms are illustrated in Fig. 1.2 and the benefits of employing multicast transmission are summarized in the following.

- Network performance is optimized through intelligent utilization of network resources to avoid unnecessary flow replication.
- Services run on multicast are scalable, easily dimensioned and thus allowing applications to be accessed by a large number of participants.
- Distributed applications are supported.

- Resource economy can be achieved through the reduction of network load, and subsequently network usage cost.

In wireless multicast, while all users within a multicast group receive the same rate from the base station (BS), the main issue arises from the mismatch data rates attainable by individual users of that group whose link conditions are typically asymmetric. If the BS transmits rate higher than the maximum rate that a user can handle, that user cannot decode *any* of the transmitted data at all. Therefore, a conventional approach is to transmit at the lowest rate of all the users within a group, which is determined by the user with the worst channel condition. This assures that the multicast services can be provided to all the subscribed users. On one hand, as all the multicast users within a group receive the same data rate from the BS, the total sum rate is scaled by the group size which is effectively the number of active users of that group. On the other hand, the lowest transmit rate typically decreases as the number of users increases because it is based on the least capable user. We, however, have established that as the number of users in a multi-carrier multicast system tends to infinity, the ergodic system capacity becomes independent of the group size but instead depends on the total number of subcarriers (see detailed derivation in Appendix B). This result confirms that the conventional multicast transmission scheme is, at least, both practical and beneficial, particularly with the use of multi-carrier transmission as in OFDM-based wireless networks.

1.2 Motivations and Research Aims

Two important resources in wireless communication are the available spectrum over which all the users signals may occupy, and the transmitted power budget. While more and more users desire to utilize the system, the actual system resources remain limited and thus making the resource allocation problem a very critical and challenging one. In literature, there are two key approaches to share out the available resources in wireless communication systems: i) fixed resource allocation, and ii) dynamic resource allocation. A fixed allocation scheme, such as Time Division Multiple Access (TDMA) or Frequency Division Multiple Access (FDMA), essentially assigns an independent dimension (time slots or frequency subchannels) to each individual user in a static manner. However, in a frequency selective fading environment there are time slots or subchannels left unused because these experience highly deep fade and therefore are not power efficient to carry any information bit.

As the allocation is fixed regardless of the current channel condition while the fundamental characteristic of wireless links is being varying, a preset resource distribution algorithm is certainly not optimal. On the contrary, a dynamic resource allocation method adaptively shares out the available dimensions to the users according to their respective channel conditions, and thus takes full advantage of channel diversity among users in different locations. This kind of diversity, commonly known as the "multiuser diversity", stems from independent pathloss and fading of different users. In a particular time slot or subchannel, although a certain user may be in deep fade, it is unlikely that all other users also experience bad channel conditions since fading parameters of different users are mutually independent. That time slot or subchannel is therefore not wasted as in the fixed allocation, but can be assigned to the users with good wireless links. By utilizing the available channel state information and also exploiting the multiuser diversity, adaptive resource distribution schemes can help to substantially enhance the system performance.

Recently, the problem of how to dynamically allocate the radio resources to improve the performance of OFDMA wireless systems has been the subject of intensive research. Broadly speaking, the solution methods obtainable in literature can be categorized into two classes – the margin adaptive and the rate adaptive. The former aims at minimizing the transmitted power under constraints on the individual user's data rate and/or bit error rate (BER) (see, for example, [3]), whereas the objective of the latter is to maximize the data rate of each user subject to transmitted power constraints and/or user's data rate (see, for instance, [4–11]). In literature, these problems have been studied for both downlink and uplink scenarios employing unicast or multicast transmission techniques, each of which has various system requirements, resulting in very different design formulations and solution methods. It is worth noticing that although much of research efforts have focused on the resource allocation for unicast OFDMA-based communication systems, many similar issues concerned with multicast settings remain open and, till now, no practical answer to these questions has been found. In these multiuser/multigroup scenarios, the problems involved the joint optimization of the subchannel and power variable sets are usually Non-deterministic Polynomial-time hard (NP-hard). As such, it is very likely that a simple application of the available solutions will lead to a prohibitively high computational complexity in most cases, whereas the optimal allocation solutions are always desired to be attained within a designated time due to quick variations of wireless channels. Indeed, this observation ascertains the impractical and inaccessible aspects of the existing approaches in multicast transmission situations.

Motivated by the shortcomings of the present resolutions, the aim of this thesis work is to provide accessible mechanisms to effectively distribute resources in a multicast wireless systems employing OFDMA. In particular, two design problems are investigated and their corresponding solutions are derived through the applications of Lagrangian duality theory, dual optimization method and genetic algorithm. First, in the context of a conventional multicast wireless system deploying OFDMA, we propose three novel efficient resource allocation schemes that balance the tradeoff between maximizing the total system throughput and ensuring a flexible and controllable spectrum sharing among different multicast groups. Specifically, a minimum number of subchannels is guaranteed to be designated to each group subject to a total power budget at the base station. Second, in a cognitive radio network setting, we devise a practically optimal resource allocation scheme which targets at maximizing the expected sum rate of all users in an OFDMA-based multicast secondary network, while satisfying the tolerable interference level induced to individual primary users. Remarkably, by defining a rate loss function as well as referring to a risk-return model, we in this design also take into account the primary user activities or the OFDM subchannel availability, an important issue which has not yet been paid adequate attention in literature. The proposed solutions to both examined problems are of great flexibility, affordable computational complexity and high performance so as to meet the vital design requirements in practical systems. Efficiency of the recommended approaches is verified by computer simulations with the presentation of numerical examples.

1.3 Thesis Outline

The rest of this thesis is organized as follows.

Chapter 2 provides necessary background knowledge on mathematical optimization. Special emphases are placed on the Lagrangian duality theory and genetic algorithm, which shall be utilized as the key tools of analysis throughout this work.

Chapter 3 presents three different efficient resource allocation algorithms for the conventional OFDMA-based multicast systems. By defining the "bandwidth control indices" which guarantee the minimum numbers of OFDM subcarriers to be assigned to each multicast group, the shares of available spectrum among individual groups can be flexibly and effectively controlled. The recommended solutions are proven to achieve high total throughput while their computational complexity is substantially reduced.

Chapter 4 proposes a resource allocation scheme for an OFDM-based multicast secondary network, subject to tolerable interference range introduced to primary users as well as the dynamics of primary users on the available radio spectrum. The proposed solution, obtained via a dual optimization framework, is claimed to achieve global optimality with fast computational time in practical systems wherein a large number of OFDM subcarriers is normally deployed.

Chapter 5 concludes the thesis, summarizing the findings and making recommendations for possible future work.

Chapter 2

Optimization Preliminaries

Optimization theory provides significant tools in the field of engineering, particularly in wireless communications. These tools are specially important for the analysis of theoretical as well as practical problems and their solutions. Chapter 2 therefore presents a foundation of optimization concepts and techniques used or referred to throughout this thesis. The main references for this chapter are [12–16].

Chapter 2 is organized as follows. Section 2.1 serves as an introduction to mathematical optimization theory. Section 2.2 presents the Lagrangian duality theory and dual optimization method. Section 2.3 discusses various concepts of Genetic Algorithm and its particular application to optimization problems as well. Finally, Section 2.4 summarizes the chapter.

2.1 Mathematical Optimization

Optimization problems occur in the mathematical modeling of a wide spectrum of applications; for instance, optimal allocation of scarce resources, scheduling, logistics, network routing, sequence alignment in genomics, portfolio optimization, engineering design optimization and so on. Most optimization problems of practical interest can be appropriately formulated as constrained optimization problems, the formulation of which is described in the following.

2.1.1 Optimization Problem Formulation

Mathematically, an optimization problem can be expressed in the form [16]

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq b_i \,;\, i = 1, \ldots, m \end{aligned} \tag{2.1}$$

where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{R}^n$ is the optimization variable, $f_0 : \mathcal{R}^n \to \mathcal{R}$ is the objective (or cost) function, and $f_i : \mathcal{R}^n \to \mathcal{R}$ (with $i = 1, \ldots, m$) are the constraint functions. Constants $b_1, \ldots, b_m$ are the bounds for the constraints. If there are no constraints, the problem is said to be unconstrained.

The set of points for which the objective and all constraint functions are defined, that is,

$$\mathcal{D} = \bigcup_{i=0}^{m} \mathbf{dom} f_i$$

is called the domain of the optimization problem (2.1). A point $\mathbf{x} \in \mathcal{D}$ is feasible if it satisfies the constraints $f_i(\mathbf{x}) \leq b_i,\ i = 1, \ldots, m$. The problem (2.1) is said to be feasible if there exists at least one feasible point. Otherwise, it is considered as an infeasible problem. The optimal value p^* of the problem (2.1) is denoted as p^* and is achieved at an optimal solution $\mathbf{x}^*$, that is, $p^* = f_0(\mathbf{x}^*)$. If that problem is infeasible, its optimal value is commonly denoted by $p^* = +\infty$.

The optimization problem (2.1) is an abstraction of the problem of finding the best possible choice of a vector in $\mathcal{R}^n$ from a set of candidate choices. While variable $\mathbf{x}$ represents the choice made, the constraints $f_i(x) \leq b_i$ represent specifications that limit the possible choices, and the objective value $f_0(\mathbf{x})$ represents the cost of choosing $\mathbf{x}$. An optimal solution $\mathbf{x}^*$ of (2.1) corresponds to a choice, from all the available choices satisfying the specifications, that has minimum cost.

2.1.2 Solving Optimization Problems

It can be argued that optimization problems are usually difficult to solve, even when the objective and constraint functions are either known to be or considered to be smooth. The solution by a single and all-purpose method is both cumbersome and inefficient, as it may involve some sort of compromise as well as involve highly complex computation. It is even possible that the use of a single and all-purpose method may give no resolution to the optimization problem at all. Because of all these difficulties, optimization problems are categorized as belonging to a particular class, where each class is defined by the properties of the objective and constraint functions of the problems belonging to that class. Solutions are then developed for each class of problems. Although finding the solutions for most optimization problems can be challenging, there are in fact some important exceptions. For a few problem classes (for example, least-squares problems, linear programs, or convex optimization), there are effective algorithms that can be employed to reliably solve problems belonging to those particular classes, even when those problems involve hundreds or thousands of variables and constraints.

Obviously, a global optimal solution for optimization problems is desirable. However, finding this global solution is far more difficult than discovering one or even many local optimal ones. Whereas an algorithm providing sub-optimality involves lower theoretical and computational complexity, resulting in many sub-optimal solutions for a particular problem, these outcomes do not guarantee a global optimal point being found [15, 17, 18]. According to [15, p. 109], a problem is said to be multi-extremal when it has many local minimizers with different objective functions values (so that a local minimizer may fail to be global). In fact, for some optimization problems, a local optimal solution is equivalent to failure, since global optimal point is a strict requirement for the correct solution.

Optimization problems can be categorized into two broad classes – constrained and non-constrained. Intuitively, problems with constraints are far more difficult to solve than unconstrained ones. Fortunately, there are many techniques available to remove restrictions, hence converting constrained problems into unconstrained ones. Lagrangian duality method, which will be discussed in the next section, is among the most efficient techniques available.

2.2 Dual Optimization and Lagrangian Duality Theory

The optimization problem (2.1) can be written in standard form as described in [16]

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0 \text{ , } i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0 \text{ , } i = 1, \ldots, p \end{aligned} \tag{2.2}$$

with variable $\mathbf{x} \in \mathcal{R}^n$, inequality constraint functions $f_i : \mathcal{R}^n \to \mathcal{R}$ (where $i = 1, \ldots, m$), and equality constraint functions $h_i : \mathcal{R}^n \to \mathcal{R}$ (where $i = 1, \ldots, p$). The domain $\mathcal{D} = \bigcup_{i=0}^{m} \mathbf{dom} f_i \cup \bigcup_{i=1}^{p} \mathbf{dom} h_i$ of Problem (2.2) is assumed to be non-empty. The basic concept of Lagrangian duality is to take the constraints in (2.2) into account by augmenting the objective function with a weighted sum of the constraint functions. The Lagrangian associated with the problem (2.2) is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \mu_i h_i(\mathbf{x}), \tag{2.3}$$

where λ_i is the Lagrange multiplier associated with the *i*-th inequality constraint $f_i(\mathbf{x}) = 0$, and μ_i is the Lagrange multiplier associated with the *i*th equality constraint $h_i(\mathbf{x}) = 0$. The domain of L is then $\mathcal{D}_L = \mathcal{D} \times \mathcal{R}^n \times \mathcal{R}^p$.

The optimization variable $\mathbf{x}$ is called the primal variable and the vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are called the dual variables or Lagrange multiplier vectors associated with the problem (2.2). The original objective function $f_0(\mathbf{x})$ is termed the primal objective or primal function. The dual objective or dual function $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is defined as the minimum value of the Lagrangian over $\mathbf{x}$; that is, for $\boldsymbol{\lambda} \in \mathcal{R}^n$ and $\boldsymbol{\mu} \in \mathcal{R}^p$,

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{x \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}), \tag{2.4}$$

which is a concave function.

Lower Bound Property: The dual function $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is a lower bound on the optimal value p^* of the problem (2.2), that is,

$$\min_{\mathbf{x}} f_0(\mathbf{x}) \geq \max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \text{ , with } \boldsymbol{\lambda} \geq \mathbf{0}. \tag{2.5}$$

When attempting to solve the primal problem (2.2), one might consider finding the best lower bound of its optimal value p^*. From the Lower bound property (2.5), it is natural that the following optimization problem, called the Lagrange dual problem, is then examined

$$\begin{aligned} \max\nolimits_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \tag{2.6}$$

The difference between the original problem (2.2) and the dual problem (2.6) is called the duality gap. Weak duality holds if property (2.5) holds with strict inequality. Strong duality holds if the equality is satisfied. While weak duality always holds, this is generally not true for strong duality. However, there are conditions called the constraint qualifications that guarantee strong duality in the case where the primal problem is a convex optimization one [1]. As well, there also exists non-convex problems that have zero duality gap. In these instances, solving the primal problem is equivalent to solving the problem (2.6). As the Lagrange dual problem is always convex regardless of the convexity of primal problem, it can be solved very efficiently in practice. In certain cases, closed-form solutions

[1]A function $f : \mathcal{R}^n \to \mathcal{R}$ is convex if $\mathbf{dom} f$ is a convex set and if for all $x, y \in \mathbf{dom} f$, and θ with $0 \leq \theta \leq 1$, we have $f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$. A convex optimization problem is one of minimizing a convex function over a feasible set described by a set of inequalities involving convex functions and a set of linear equality constraints [16].

can be analytically obtained. However, in general iterative methods such as interior-point or cutting-plane are usually employed to solve the convex optimization problem (see, for example, [15, 17, 19]).

2.2.1 Example: Water-filling Solution via Dual Optimization

In solving engineering optimization problems, solutions involving water-filling structure are frequently obtained. The application of Lagrange duality theory in finding solutions for a typical information theory optimization problem is demonstrated in the water-filling example shown below.

Example 2.2.1 ([16], Example 5.2) *(Channel capacity maximization)*

Consider the following:

$$\begin{array}{ll} \max_{\mathbf{x}} & \sum_{i=1}^{n} \log(x_i + \alpha_i) \\ \text{s.t.} & \mathbf{x} \geq \mathbf{0}, \mathbf{1}^T \mathbf{x} = 1 \end{array} \tag{2.7}$$

where $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\alpha_i > 0$. This can be viewed as the problem of optimally allocating transmitted power, as represented by $\mathbf{x}$, to maximize the communication channel capacity as in the objective function. The constants α_i (with $i = 1, 2, \ldots, n$) represent the noise variances.

Lagrange multipliers $\boldsymbol{\lambda} \in \mathcal{R}^n$ and $\mu \in \mathcal{R}$ (for the inequality constraints $\mathbf{x} \geq \mathbf{0}$ and the equality constraint $\mathbf{1}^T \mathbf{x} = 1$ respectively) are now introduced. At optimality of $\mathbf{x}$, the Karush–Kuhn–Tucker (KKT) conditions give [16, p. 243]

$$\begin{gathered} \mathbf{x} \geq \mathbf{0}, \quad \mathbf{1}^T \mathbf{x} = 1, \quad \lambda \geq \mathbf{0}, \\ \lambda_i x_i = 0, \quad i = 1, \ldots, n \\ -\tfrac{1}{\alpha_i + x_i} - \lambda_i + \mu_i = 0, \quad i = 1, \ldots, n. \end{gathered}$$

The slack variable λ can then be eliminated, leaving

$$\begin{gathered} \mathbf{x} \geq \mathbf{0}, \quad \mathbf{1}^T \mathbf{x} = 1, \\ x_i(\mu - \tfrac{1}{\alpha_i + x_i}) = 0, \quad i = 1, \ldots, n \\ \mu \geq \tfrac{1}{\alpha_i + x_i}, \quad i = 1, \ldots, n. \end{gathered}$$

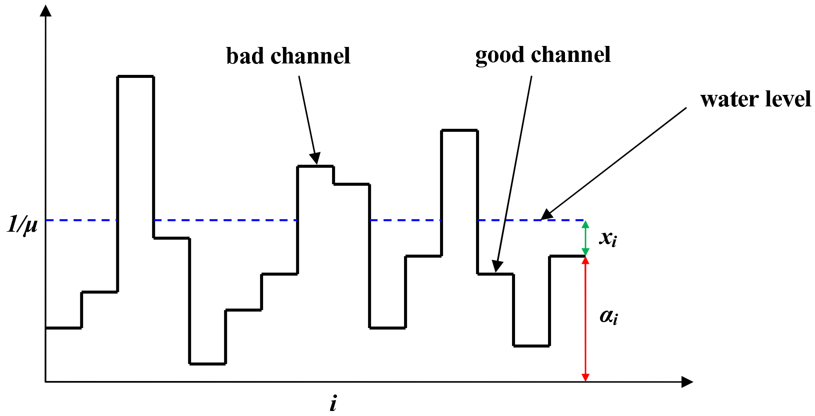


Figure 2.1: Water-filling structure.

Finally, this results in

$$x_i = \begin{cases} \frac{1}{\mu} - \alpha_i & \text{if } \mu < \frac{1}{\alpha_i}, \\ 0 & \text{if } \mu \geq \frac{1}{\alpha_i}. \end{cases} \tag{2.8}$$

That is, $x_i = \max(0, \frac{1}{\mu} - \alpha_i)$, where μ is determined from

$$\sum_{i=1}^{n} \max(0, \frac{1}{\mu} - \alpha_i) = 1. \tag{2.9}$$

In Figure 2.1, the water-filling structure (2.9) is illustrated. There are n patches. The height of each patch i is α_i. The region is flooded to a level $\frac{1}{\mu}$ which uses a total quantity of water equal to 1. The height of the water above each patch is the optimal value of x_i.

2.3 Genetic Algorithm for Optimization

2.3.1 Overview

Genetic Algorithm (GA) [12–14], categorized as global search heuristics, is a search technique used to find exact or approximate solutions to both constrained and unconstrained optimization problems. It is based on natural selection, the process driving biological evolution. In brief, the GA is implemented as a computer simulation in which a population of abstract representations (called *chromosomes* or the genotype of the genome) of candidate solutions (called *individuals* or creatures) to an optimization problem evolves toward better solutions. Specifically, the evolution usually starts from a population of randomly generated individuals and happens in generations. The GA repeatedly modifies a population of individuals through iterations, and, at each iteration, the algorithm randomly picks individuals from the current population to be parents which are then used to produce the *children* (or

Table 2.1: Comparison of Genetic Algorithm and classical algorithms.

Genetic Algorithm	**Classical Algorithms**
Generates a population of points at each iteration. The best point in the population approaches an optimal solution.	Generates a single point at each iteration. The sequence of points approaches an optimal solution.
Selects the next population by computation which uses random number generators.	Selects the next point in the sequence by a deterministic computation.

offsprings) for the next generation. Since the population evolves toward an optimum over successive generations, a sufficiently good solution to the optimization problem can finally be found. When compared to classical derivative-based optimization algorithms, the GA differs in two main aspects as summarized in Table 2.1.

The attractiveness of GAs comes from their simplicity and elegance as robust search algorithms as well as from their power to discover good solutions rapidly for difficult high-dimensional problems. This class of algorithms are particularly useful and efficient when

- The search space is large, complex or even poorly understood.
- It is difficult to encode to narrow the search space.
- No mathematical analysis is available.
- Traditional search methods fail.

For the above reasons, Genetic Algorithm has been employed in numerous applications in different fields such as machine learning, bioinformatics, economics, chemistry, manufacturing, mathematics, physics, and so on.

2.3.2 Basic Operations of Genetic Algorithm

A typical GA is presented in Table 2.2 [12, 14] and its detailed operations are explained in the rest of this section.

While there are many different implementations of the general GA for various problems, the key of success in the application of GA lies in an effective representation of the solution domain and also a meaningful fitness function to evaluate the solution.

- **Coding of Individuals:** A standard representation of the solution is as an array of bits. Arrays of other types and structures can be used in essentially the same way. The

Table 2.2: Outline of a basic Genetic Algorithm.

1. [**Start**] Generate a random population consisting K chromosomes (that is, the suitable solutions for the problem)
2. [**Fitness**] Evaluate the fitness $f(x)$ of each chromosome x in the population
3. [**New population**] Create a new population by repeating the following steps until the new population is complete
   a. [**Selection**] Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be chosen)
   b. [**Crossover**] With a certain crossover probability, cross over the parents to form a new offspring. If no crossover was performed, offspring is an exact copy of parents.
   c. [**Mutation**] With a certain mutation probability, mutate new offspring at a position in chromosome
   d. [**Accepting**] Place new offspring in a new population
4. [**Replace**] Use newly generated population for a further run of the algorithm
5. [**Test**] If an end condition is met, stop, and return the best solution in current population
6. [**Loop**] Go to step 2

main property that makes these genetic representations convenient is that their parts are easily aligned due to their fixed size, which in turn facilitates simple crossover operation. Variable length representations are also possible but implementation of crossover is more complex in this case.

- **Fitness Function:** Fitness function, a measurement of the quality of the represented solution, is defined over the genetic representation. In fact, the fitness function is always problem dependent. For certain instances, it is difficult or even impossible to define the fitness expression, in which case interactive genetic algorithms are to be utilized.

Once the genetic representation and the fitness function have been defined, GA proceeds to initialize a population of solutions randomly, then improve it through repetitive application of the following key rules. The reproduction operations of GA are also depicted in Fig. 2.2.

- **Selection Rule:** Selection rule determines how individuals are chosen for mating. If a selection method that picks only the best individual is used, the population will quickly converge to that individual. Therefore, it is important to design a selector that is not only biased toward better individuals but also able to pick some that are not quite as good (but hopefully have some good genetic material in them). Some of the more common rules include roulette wheel selection (that is, the likelihood of picking an individual is proportional to the individual's score), tournament selection (that is, a number of individuals are picked using roulette wheel selection, then the best of these is chosen for mating), and rank selection (that is, pick the best individual every time). Other selection rules such as stochastic remainder sampling, stochastic uniform sampling may also be effective.

- **Crossover Rule:** Crossover is used to combine two parents to form children for the next generation. Essentially, crossover enables the algorithm to extract the best genes from different individuals and recombine them into potentially superior offsprings. Some common rules include one-point or two-point, cut-and-splice and uniform crossover.

- **Mutation Rule:** Mutation applies random changes to individual parents to form children. The purpose of mutation in GA is to permit the algorithm to avoid local minima by preventing the population of individuals from becoming too similar to each other which, in turn, may slow down or even stop evolution. In this sense, mutation adds to the diversity of a population and thereby increases the likelihood that the algorithm will generate individuals with better fitness values.

- **Elite Children Rule:** If no operation is performed on a parent (likely to be one with the highest fitness value) and this individual is allowed to automatically survive to the next generation, then it is called the *elite child*.

The generational process is repeated until a terminating condition has been reached. Of which, some popular ones are:

- A solution that satisfies minimum criteria is found.

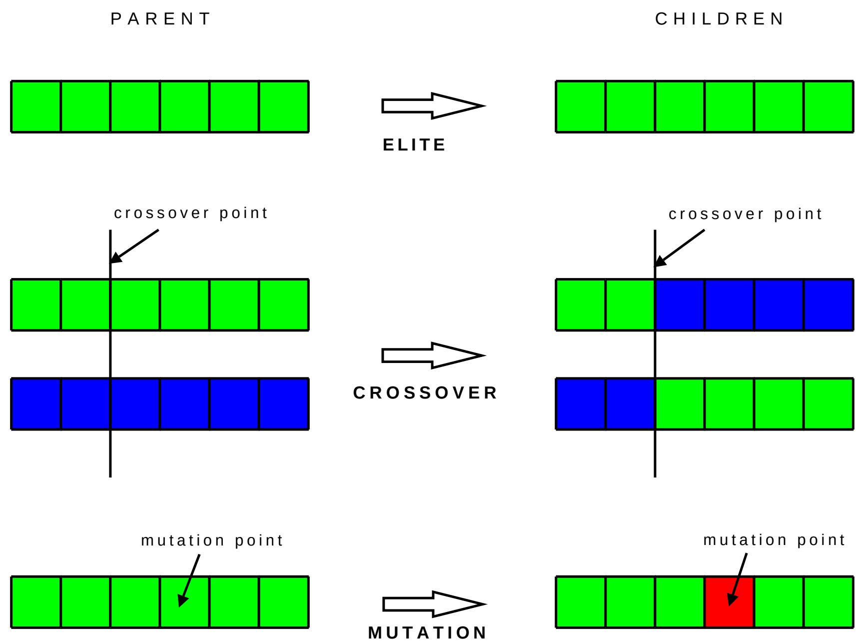


Figure 2.2: Reproduction process in Genetic Algorithm.

- Fixed number of generations has reached.
- Allocated budget (for instance, computation time) has been all spent.
- The highest ranking solution's fitness has reached a plateau such that successive iterations no longer produce better results.
- By manual inspection
- Combinations of the above

2.4 Concluding Remarks

In this chapter, we have presented some background knowledge on mathematical optimization, which shall be used as the main analytical tool throughout the thesis. Lagrangian dual method and genetic algorithm have been introduced with examples given to illustrate their operations.

Chapter 3

Efficient Resource Allocation for Conventional OFDMA-based Multicast Wireless Systems

It is apparent that radio spectrum is among the most precious resources in wireless communication. As the available spectrum is scarce and wireless channels are dynamic, resource allocation in multicast systems should be efficient, be able to cope with the channel variation, and more importantly, be flexible to adjust the level of priority or fairness in terms of accessible bandwidth provided to individual groups.

Chapter 3, which studies the efficient resource allocation algorithms for conventional OFDMA-based multicast systems with a controllable mechanism for spectrum sharing, is organized as follows[1]. Section 3.1 summarizes related study on the subject and also to highlight the original contributions of this research work. Section 3.2 formulates the OFDMA-based multicast resource allocation problem with spectrum-sharing constraints. Sections 3.3 and 3.4 propose the separate optimization and genetic-algorithm-based schemes, respectively. Section 3.5 analyzes computational complexity and also evaluates performance of the proposed solutions with the support of numerical results. Finally, Section 3.6 concludes the chapter.

[1]A version of this chapter has been accepted for presentation at the 2009 IEEE Radio and Wireless Symposium (RWS'09) held in San Diego, USA [20]. Further development of the results has been accepted for publication as a regular paper in the IEEE Transactions on Vehicular Technology [21].

3.1 Introduction

3.1.1 Background

For the uplink from mobile users to the base station (BS) of a conventional OFDMA wireless system, the study by [8] investigates the sum-rate maximization problem and derives the necessary optimality conditions, from which a near-optimal joint scheme of subcarrier and power allocation is proposed. Specifically, the subcarrier is assigned by a greedy algorithm while power is distributed in a water-filling manner. More generalized, reference [22] deals with the utility maximization problem in which three specific examples, namely throughput optimization, proportional fairness and max-min fairness, are examined. Regarding the issue of fairly distributing the available resources to different users, reference [11] devises a low-complexity algorithm to maximize the sum rate of an OFDMA uplink, constrained on both individual rate and transmitted powers. Here, the interpretation of fairness is to assure a minimum target rate to be met with a high probability for all users including those with bad link conditions. Specifically, the initial subcarrier allocation designates subcarriers to users whose rates are below the predefined target, and later the remaining subcarriers are allotted in such a way that the sum rate is enhanced. Once the subcarrier assignment process is finished, the optimal single-user power allocation is performed for each user by water-filling over the already determined subcarriers. As for the uplink of an OFDMA relay-assisted network, the work in [9] considers the subcarrier assignment problem wherein a certain notion of fairness is achieved by allowing a relay node is to use/aid only up to a maximum number of subcarriers/sources. The binary integer program formulated here is first transformed to a linear optimal distribution problem, which is then solved via graph theory.

For the downlink transmissions from the BS to the mobile users, the proposed algorithm in [3], through the application of the Lagrangian relaxation, solves the formulated multiuser OFDM margin adaptive problem by relaxing the subchannel assignment binary variables to take any real value between zero and one. In [5], it has been proven that the sum capacity is maximized only when each subchannel is assigned to the user with the best channel gain for that subcarrier, and power is distributed by the water-filling algorithm. On the other hand, [4] studies the fair max-min problem, in which all users are guaranteed to eventually achieve a similar rate through maximizing of the worst user's capacity. A suboptimal algorithm is also provided to alleviate the intensive computation required to find optimum of the resulting max-min problem. A more general fairness-aware scheme has been suggested by [6], where a set of proportional fairness constraints is imposed to ensure that each user can achieve its required data rate as in systems with QoS guarantees. Similar to [4], a low-complexity suboptimal algorithm that separates the subchannel allocation and the power distribution is proposed because it is too computationally complex to resolve the constrained non-linear fairness problem. Specifically, subchannel assignment is first accomplished by applying the algorithm of [4] to achieve coarse fairness. Then, fine fairness is attained by power allocation which targets at maximizing the sum capacity while maintaining proportional fairness via an iterative method such as Newton-Ralphson or quasi-Newton methods. Considering another form of fairness, namely fair bandwidth distribution among users, the study in [7] recommends an allocation of equal number of subcarriers to each single user. The total data rate can then be maximized via a greedy algorithm for a given transmitted power constraint and bit error rate (BER) requirements. In [10], the level of fairness achieved by [7] is further enhanced through swapping the subcarriers belonging to the user with the most number of loaded bits to the user with the least. Examining a fairness-aware dynamic resource allocation for the downlink of a multihop OFDMA system, the investigation in [23] formulates an optimization problem to maximize the system capacity while guaranteeing minimum resources for each user. Then, an efficient heuristic algorithm, which comprises of subchannel allocation, load balancing and power distribution steps based on the ideas of [4], is proposed. It should be noticed that the aforementioned approaches requires channel state information made available at the base station. To further avoid the extensive feedback of channel information from users to base station, the work in [24] suggests two constant complexity limited-feedback resource allocation algorithms for the downlink in OFDMA networks, which achieve near-optimal performance.

When multicast transmission [25, 26] is employed, the study in [27] proposes a low-complexity resource allocation scheme to improve the Shannon capacity of the downlink in an OFDM multicast wireless network. Specifically, each of the available subchannels is assigned to the group with the best channel and the most member users, under the assumption of equal transmitted power, followed by water-filling of the power. In [28], a low-complexity heuristic algorithm for suboptimally allocating resources of an OFDM multicast system is proposed to minimize the number of OFDM symbols that each individual user receives, thus resulting in a reduction of power consumed by the users. Considering the downlink transmission of an OFDM-based multicast wireless network with the assumption of multiple description coding (MDC), reference [29] solves the power control/bit loading problem for maximum throughput and proportional fairness. To avoid high complexity required to solve the resulting integer program, a two-step suboptimal algorithm is proposed. In particular, subcarriers are assigned assuming constant transmitted power being distributed to each subcarrier, then bits are loaded to the allocated subcarriers through the application of a modified Levin-Campello algorithm.

3.1.2 Research Contributions

Since the channel quality of every user in a multicast network may be very different, the attainable data rate of each multicast stream is usually restricted by the data rate of the least capable user. Furthermore, the number of users in a multicast group also has a direct impact on the aggregate data rate that can be achieved by that group. These critical factors lead to imbalanced opportunities in gaining access into the available system resources, such as bandwidth and power, of individual multicast groups. When the differences in pathloss and/or size among groups are large, it is likely that the typical adaptive resource allocation schemes, which try to maximize the system performance, will distribute most of the available bandwidth (and subsequently power) to the groups with high equivalent channel signal-to-noise-ratio (CSNR) and/or with larger user sets, for a significant portion of time. Consequently, the groups with worse channel conditions and/or with fewer member users may not be able to access to any available resources at all. As the system resources are valuable but scarce and maximizing total system throughput is not always the only design priority, the issue of fair resource utilization among multicast groups with diverse CSNR characteristics and with different group sizes becomes particularly important. The fair allocation of available resources in OFDMA-based systems has been discussed in different contexts for both unicast and multicast scenarios, including max-min fairness [4], proportional rate guarantee [6], minimum bandwidth assurance [23], equal bandwidth distribution [7], and proportional fairness [29]. However, none of these solutions accounts for a controllable sharing of the available radio spectrum to flexibly distribute the system resources in wireless multicast settings. Motivated by the works in [4, 23], this research work shall address the above-mentioned shortcoming of existing solutions.

We first provide a new formulation for the resource allocation problem in OFDMA-based multicast wireless systems that balances the tradeoff between maximizing the total throughput and ensuring a flexible and controllable spectrum sharing among different multicast groups. To this end, by introducing the “bandwidth control indices” which can be easily regulated, we impose constraints on the minimum numbers of subcarriers to be assigned to individual groups. The indices can be adjusted so that the formulated problem may be cast into the problem of sum rate maximization. More importantly, if a fair bandwidth sharing[2] among different groups with asymmetric links and diverse group sizes is desired, the minimum numbers of subcarriers can always be set to proper values which are determined from the respective channel conditions and sizes of individual groups. On the one hand, this prevents groups with good channels or with large user sets from greedily consuming all the available bandwidth. On the other hand, it guarantees that groups with poorer channel conditions or with smaller group sizes still have good opportunities to access the system resources.

We then propose three novel efficient schemes with low computational complexity to solve the formulated NP-hard design problem. In the first and second schemes, the allocation is accomplished via separate optimization of subcarriers and transmitted power where, specifically, subcarriers are assigned based on the assumption of uniform power allocation, followed by water-filling of the total power over the determined subcarrier assignment. In the third scheme, which is based on a modified genetic algorithm [12–14], each individual of the whole population corresponds to a subcarrier allocation, and whose fitness score is the system throughput computed on the basis of power water-filling procedure. It is shown that with proper adjustments of the minimum numbers of subcarriers assigned to individual groups, the proposed solutions provide a more flexibility in controlling the share of available radio spectrum given to each group and, at the same time, achieve a very high total sum rate. Complexity analysis of the proposed approaches is carried out, and their potentials are thoroughly verified via simulation with the illustration of numerical examples.

3.2 System Model and Problem Formulation

Consider a one-cell multicast wireless system employing OFDMA, in which one base station (BS) transmits G (downlink) traffic flows, each to one distinct multicast group, over M subcarriers. Assume that each user receives one traffic flow at a time, hence it belongs

[2] In this work, “fair bandwidth sharing” means that a certain multicast group deserves some portion of the total available bandwidth regardless of its link condition or group size. Also, the terms “bandwidth” and “radio spectrum” are used interchangeably.

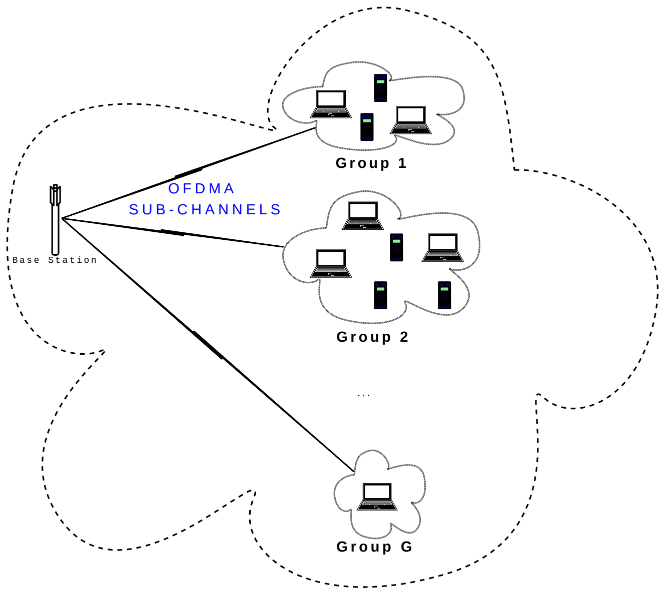


Figure 3.1: Downlink of an OFDMA multicast system.

to only one multicast group. Let K_g and $|K_g|$ $(g = 1, \cdots, G)$ denote the user set of group g and its cardinality, respectively. Since the g-th group is unicast if $|K_g| = 1$ whereas it is multicast if $|K_g| > 1$, the model is valid for both unicast and multicast settings. Clearly, all the users belong to the set $K = \bigcup_{g=1}^{G} K_g$, and $|K| = \sum_{g=1}^{G} |K_g|$ is the total number of users in the system. Let B denote the total system bandwidth and assume that each subcarrier has an equal bandwidth of $B_m = B_0 = B/M$. A generic system setup is depicted in Fig. 3.1.

In this work, the resource allocation is accomplished in a centralized manner at the BS, which has the perfect channel state information (CSI) of all the users in the systems via dedicated feedback channels. This is a typical assumption in literature [3, 5, 6]. The BS is then able to determine the maximum rate at which an individual user can reliably receive data, as well as the corresponding subcarrier over which the data shall be transmitted on. It is known that the maximum attainable rate of user $k \in K_g$ on subcarrier m is

$$r_{k,m} = \frac{B_0}{B} \log_2 \left(1 + \frac{|h_{k,m}|^2 P_m}{B_0 N_0} \right), \tag{3.1}$$

where $h_{k,m}$ represents the channel from the BS to user k on subcarrier m, P_m is the transmitted power allocated to subcarrier m, and N_0 is one-sided power spectral density of additive white Gaussian noise. It is further assumed that the channel conditions remain unchanged during the allocation period. This assumption is particularly valid for slowly-varying channels where the channel gains do not vary too significantly over time, for example, in high data rate systems and/or environments with reduced degrees of mobility.

An attractive feature of wireless multicast is that multicast data can be transmitted from the BS to multiple mobile users only through a single transmission. However, while all users within a multicast group receive the same rate from the BS, the main issue arises from the mismatch data rates attainable by individual users of that group whose link conditions are typically asymmetric. If the BS transmits rate higher than the maximum rate that a user can handle, that user cannot decode *any* of the transmitted data at all. Therefore, a conventional approach is to transmit at the lowest rate of all the users within a group, which is determined by the user with the worst channel condition [27]. This assures that the multicast services can be provided to *all* the subscribed users. On the one hand, as all the multicast users within a group receive the same data rate from the BS, the total sum rate is scaled by the group size which is effectively the number of active users of that group. On the other hand, the lowest transmit rate typically decreases as the number of users increases since it is based on the least capable user. Appendix B, however, establishes that as the number of users in a multi-carrier multicast system tends to infinity, the ergodic system capacity becomes independent of the group size but depends on the total number of subcarriers. This result confirms that the conventional multicast transmission scheme is indeed both practical and beneficial, particularly with the use of multi-carrier transmission as in OFDM-based wireless networks.

It is worth pointing out that other approaches such as exploiting the hierarchy in multicast data with the use of Multi Description Coding are possible (see, for instance, reference [29]). In this particular approach, a single data source is fragmented into several independent substreams (called descriptions) arranged in a hierarchy that provides progressive refinement. If only the first (base) description is received by the worst user, that user can decode the worst quality version. As more descriptions are received by more capable users, they can combine these descriptions to produce improved quality. However, such approaches are limited to multimedia (video and audio) applications whereas it may not be practical to perform partitioning of the data in other applications, for instance, file transfer. Therefore, MDC approaches are not pursued in this study where, instead, the conventional approach shall be followed to deal with a more general class of applications. Here, the BS is enforced to transmit at the lowest rate of all the users within a group, which is determined by the user with the smallest channel-to-noise ratio (CSNR). On subcarrier m, let

$$\beta_{g,m} = \min_{k \in K_g} \frac{|h_{k,m}|^2}{B_0 N_0} \tag{3.2}$$

be the equivalent CSNR of group g, then the maximum rate at which *all* users of group g are able to decode the transmitted data is

$$\check{r}_{g,m} = \frac{B_0}{B} \log_2(1 + \beta_{g,m} P_m). \tag{3.3}$$

As all users in a group receive the same rate, the aggregate data rate transmitted to group g on subcarrier m is thus

$$R_{g,m} = \sum_{k \in K_g} \check{r}_{g,m} = |K_g| \, \check{r}_{g,m}. \tag{3.4}$$

The goal of this work is to devise a subcarrier assignment and power allocation policy that maximizes the system sum rate of all multicast groups, while satisfying a constraint on the total transmitted power. Distinct from the existing works, here the important issue of providing a flexible mechanism to effectively govern the share of the accessible bandwidth among various multicast groups is also taken into account. One possible way to realize this idea is to guarantee a certain minimum number of subcarriers to be allocated to each group. Specifically, the design problem can be formulated as follows:

$$\max_{\{\rho_{g,m}, P_m\}} \quad \sum_{g=1}^{G} \sum_{m=1}^{M} \frac{|K_g|}{M} \rho_{g,m} \log_2(1 + \beta_{g,m} P_m) \tag{3.5}$$

$$\text{subject to:} \quad \sum_{m=1}^{M} P_m \leq P_{\text{tot}}, \tag{3.6}$$

$$P_m \geq 0, \ m = 1, \cdots, M, \tag{3.7}$$

$$\sum_{g=1}^{G} \rho_{g,m} = 1, \ m = 1, \cdots, M, \tag{3.8}$$

$$\rho_{g,m} \in \{0, 1\}, \tag{3.9}$$

$$\sum_{m=1}^{M} \rho_{g,m} \geq \alpha_g, \ g = 1, \cdots, G. \tag{3.10}$$

In this formulation, the binary variable $\rho_{g,m}$ represents the allocation of subcarrier m to group g. Constraints (3.6)-(3.7) express the power limitation at the BS, whereas constraints (3.8)-(3.9) ensure a disjoint subcarrier assignment in OFDMA systems in which one subcarrier can only be given to at most one group. Constraint (3.10) reflects the spectrum-sharing control of the design, where the "bandwidth control index" α_g is required to satisfy

$\alpha_g \in \mathcal{Z}_+$ and $\sum_{g=1}^{G} \alpha_g \leq M$. The value α_g manages the priority in terms of spectrum access opportunity provided to each multicast group. It varies from 0 to M and can be flexibly adjusted according to system design specifications. As α_g increases toward M, a higher priority is given to group g. In particular, if all α_g's approach 0, problem (3.5)–(3.10) becomes that of sum rate maximization. Moreover, as all α_g's approach $\lfloor \frac{M}{G} \rfloor$, the optimization formulation enforces (almost) a strict bandwidth fairness.

It should be pointed out that problem (3.5)–(3.10) is NP-hard. Therefore determining its optimal solution within a given time is very challenging. Performing a direct exhaustive search at the BS would obviously face a prohibitive computational burden where the optimal solutions must be obtained within a designated time period due to quick variations of wireless channels. Since such a solution method is too computationally expensive, it is impractical, particularly for systems with large number of subcarriers (which is often the case in practice). Suboptimal algorithms, which have a low complexity and yet provide good performance, are therefore preferable for cost-effective and delay-sensitive implementations. In the next sections, three efficient solutions to solve the formulated design problem (3.5)–(3.10) are proposed. The first two solutions are based on separate optimization of subcarriers and power, while the last one is obtained with a modified genetic algorithm.

3.3 Efficient Resource Allocation via Separate Optimization

Ideally, both subcarriers and power should be jointly allocated to achieve the global optimum of (3.5)–(3.10). However, this is highly complicated as the total number of variables becomes large. Instead of jointly optimizing $\{\rho_{g,m}\}$ and $\{P_m\}$, separate optimization over these two set of variables shall be performed. Although suboptimal, this approach enables significantly lower computational complexity since the number of variables in each separate optimization problem is reduced almost by half. Specifically, the subcarrier assignment problem is solved in the first phase by assuming a constant power allocation on subcarriers. In the second phase, the total power is distributed over the available subcarriers in a water-filling fashion.

3.3.1 Phase 1 – Subcarrier Allocation with Uniform Power Assumption

Under the assumption of equal-power distribution over the subcarriers, the data rate of the downlink traffic flow to multicast group g on subcarrier m in (4.7) becomes

$$R_{g,m} = \frac{|K_g|}{M} \log_2 \left(1 + \beta_{g,m} \frac{P_{\text{tot}}}{M}\right). \tag{3.11}$$

The proposed two-step subcarrier allocation is detailed in Algorithm 1. In Step 1, each subcarrier is assigned to the group who has the largest value of $R_{g,m}$ and who has not been given its required minimum number of subcarriers. Once a subcarrier is assigned, it will not be considered in all subsequent operations. Further, the group which has already been allocated its minimum number of subcarriers is discarded in all subsequent iterations. While the largest $R_{g,m}$ corresponds to the group with largest group size and/or the best link condition, the bandwidth constraint actually helps avoid the situation that subcarriers are all granted to the advantageous multicast groups. The procedure is repeated until all groups have been allocated their minimum numbers of subcarriers. In Step 2, the remaining subcarriers left from Step 1 are assigned to the group who has the largest value of $R_{g,m}$ in a sequential manner. Effectively, the allocation controls a certain level of bandwidth sharing as a result of Step 1, whereas the system throughput is further enhanced as a direct consequence of Step 2.

It can be easily seen that the lookup of subcarrier-group pair (g_m^*, m^*) in lines 8-11 of Algorithm 1 involves a two-dimensional search, which could be highly intensive for systems with large numbers of subcarriers and multicast groups. To alleviate this drawback, we now propose a reduced-complexity subcarrier assignment based on Algorithm 1. Different from Algorithm 1, the reduced-complexity approach performs the assignment on a per-subcarrier basis in Step 1 where randomization is carried out to pick a subcarrier for which all the eligible groups, that is, the ones that have not reached their minimum numbers of subcarriers, will compete. Since the assignment only requires a one-dimensional search for each subcarrier, its computational complexity is significantly lower. A full description of this algorithm is provided in Algorithm 2.

3.3.2 Phase 2 – Water-filling Power Allocation

Once subcarrier allocation is accomplished, all the values of $\rho_{g,m}$ are known. Hence, power allocation can be optimally completed on a per-subcarrier basis. The optimization

input : $\bar{R}_{G\times M} = \{R_{g,m}\}$ computed by (3.11),
$\mathcal{S} = \{1, 2, \cdots, M\}$, and $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_G] \in \mathcal{Z}_+^G$.
output: Subcarrier allocation $\{\rho_{g,m}\}$

1 **begin**
  `// Initialization`
2   $\rho_{g,m} \longleftarrow 0, \ \forall g, m$
3   $\Theta \longleftarrow [0, 0, \cdots, 0] \in \mathcal{Z}_+^G$
  `// Step 1`
4   **if** $\alpha_g == 0, \ \forall g = 1, \cdots, G$ **then**
5   go to Line 20
6   **else**
7   **repeat**
8   **find** pair $\{g_m^*, m^*\}$ that satisfy both conditions:
9   (a) $\bar{R}(g_m^*, m^*) \geq \bar{R}(g, m), \ \forall g, m,$
10   (b) $\Theta(g_m^*) \leq \alpha(g_m^*)$
11   **end**
12       $\rho_{g_m^*, m^*} \longleftarrow 1$
13       $\rho_{g \neq g_m^*, m^*} \longleftarrow 0$
14       $\Theta(g_m^*) \longleftarrow \Theta(g_m^*) + 1$
15       $\mathcal{S} \longleftarrow \mathcal{S} \backslash \{m^*\}$
16   delete $\bar{R}(:, m^*)$ from $\bar{R}$
17   **if** $\Theta(g_m^*) > \alpha(g_m^*)$ **then** delete $\bar{R}(g_m^*, :)$ from $\bar{R}$
18   **until** $\Theta_g \geq \alpha_g, \forall g = 1, \cdots, G$
19   **end**
  `// Step 2`
20   **foreach** $m \in \mathcal{S}$ **do**
21   find $g_m^* = \arg_g \max \bar{R}(g, m)$
22     $\rho_{g_m^*, m} \longleftarrow 1$
23     $\rho_{g \neq g_m^*, m} \longleftarrow 0$
24   **end**
25 **end**

Algorithm 1: Subcarrier Assignment with Bandwidth-sharing Control.

```
input : $\bar{R}_{G\times M} = \{R_{g,m}\}$ computed by (3.11),
        $\mathcal{S} = \{1, 2, \cdots, M\}$,
        $\mathcal{G} = \{1, 2, \cdots, G\}$, and
        $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_G] \in \mathcal{Z}_+^G$.
output: $\{\rho_{g,m}\}$
1  begin
       // Initialization
2      $\rho_{g,m} \longleftarrow 0,\ \forall g, m$
3      $\Theta \longleftarrow [0, 0, \cdots, 0] \in \mathcal{Z}_+^G$
       // Step 1
4      if $\alpha_g == 0,\ \forall g = 1, \cdots, G$ then
5          go to Line 17
6      else
7          repeat
8              randomly pick $m^*$ from $\mathcal{S}$
9              find $g_m^* = \arg\max_{g\in\mathcal{G}} \bar{R}(g, m^*)$
10             $\rho_{g_m^*, m^*} \longleftarrow 1$
11             $\rho_{g \neq g_m^*, m^*} \longleftarrow 0$
12             $\Theta(g_m^*) \longleftarrow \Theta(g_m^*) + 1$
13             $\mathcal{S} \longleftarrow \mathcal{S} \backslash \{m^*\}$
14             if $\Theta(g_m^*) > \alpha(g_m^*)$ then $\mathcal{G} = \mathcal{G} \backslash \{g_m^*\}$
15         until $\Theta_g \geq \alpha_g, \forall g = 1, \cdots, G$
16     end
       // Step 2
17     foreach $m \in \mathcal{S}$ do
18         find $g_m^* = \arg_g \max \bar{R}(g, m)$
19         $\rho_{g_m^*, m} \longleftarrow 1$
20         $\rho_{g \neq g_m^*, m} \longleftarrow 0$
21     end
22 end
```

Algorithm 2: Reduced-complexity Subcarrier Assignment.

problem (3.5)–(3.10) now becomes

$$\begin{aligned} \max_{P_m \geq 0,\, m=1,\ldots,M} \quad & \sum_{m=1}^{M} \frac{|K_{g_m^*}|}{M} \log_2(1 + \beta_{g_m^*,m} P_m) \qquad (3.12) \\ \text{subject to:} \quad & \sum_{m=1}^{M} P_m \leq P_{\text{tot}}, \end{aligned}$$

where each subcarrier m has been assigned to group g_m^*.

Clearly, (3.12) involves the maximization of a concave function over a linear set, thus it is a convex optimization problem. The closed-form solution can then be obtained by employing the Lagrange multiplier method. The Lagrangian of (3.12) can be expressed as follows:

$$\begin{aligned} \mathcal{L}(P_m, \mu) \quad = \quad & \sum_{m=1}^{M} \frac{|K_{g_m^*}|}{M} \log_2(1 + \beta_{g_m^*,m} P_m) \\ & -\mu \left(\sum_{m=1}^{M} P_m - P_{\text{tot}} \right), \qquad (3.13) \end{aligned}$$

where $\mu > 0$ is a Lagrange multiplier. The optimal power allocation can be derived from the Karush-Kuhn-Tucker (KKT) conditions to be

$$P_m \quad = \quad \max \left(\frac{|K_{g_m^*}|}{\mu M \log 2} - \frac{1}{\beta_{g_m^*,m}}, 0 \right). \qquad (3.14)$$

It can be observed that the solution in (3.14) has the form of water-filling, where μ can be easily found from the total power constraint $\sum_{m=1}^{M} P_m \leq P_{\text{tot}}$.

Combining Phase 1 and Phase 2 in the above results in two complete efficient resource allocation schemes, which shall be referred to as the Bandwidth Control – Separate Optimization (BC-SO) and Reduced-complexity Bandwidth Control – Separate Optimization (RCBC-SO), respectively. Though being simple, the allocation schemes devised in this section are suboptimal due to the separation of optimization variables in each allocation phase. In the next section, we propose another efficient scheme which utilizes the Genetic Algorithm to provide a global search for a jointly optimal subcarrier and power allocation.

3.4 Efficient Resource Allocation via Modified Genetic Algorithm

By its nature, a Genetic Algorithm (GA) does not begin its optimization process from a single point in the search space, but rather from an entire set of individuals, which form the initial population. Hence, GA may be invoked in robust global search and optimization procedures that do not require the knowledge of the objective function's derivatives or any gradient-related information concerning the search space. It is therefore particularly suitable for optimization problems which are not well suited for standard optimization algorithms, including problems whose objective function is discontinuous, non-differentiable, stochastic, or highly nonlinear[3]. Regarding the NP-hard design problem (3.5)–(3.10), the objective function involves both continuous and discrete variables and thus represents a class of problems for which GA can be efficiently applied.

The proposed efficient scheme, which shall be referred to as <u>B</u>andwidth <u>C</u>ontrol – <u>G</u>enetic <u>A</u>lgorithm (BC-GA), follows the general procedure of a GA together with the following features to specifically solve the design problem under investigation (3.5)–(3.10).

Coding of Individuals

Each individual of the population corresponds to a subcarrier allocation. It is coded as a vector of length M whose indices represent the subcarriers, and the value of each vector entry is an integer in the range $[1, G]$ representing the group that has been assigned the subcarrier corresponding to that entry. For instance, the m-th entry of an individual has value of g implies that subcarrier m is designated to multicast group g. Fig. 3.2 depicts the coding of individuals and the entire population in one generation.

Initial Population

The initial population of size N_p can be randomly generated, with high-quality individuals possibly being fed into the population. A fine individual could be either a good subcarrier allocation generated by appropriate randomization, or the suboptimal solutions derived via the proposed BC-SO and RCBC-SO schemes in Section 3.3. With a well-chosen starting population, the time required for BC-GA to reach an optimum solution would be substantially reduced.

Fitness Function

For each individual, its fitness value is the corresponding total sum rate. To compute this value, first the bandwidth-control constraint in (3.10) is checked against each individual (that is, each subcarrier allocation). If the constraint is unsatisfied, the individual

[3]See, for instance, the adaptive resource allocation and call admission control problem in [30], or the multiple-antenna OFDM multiuser detection problem in [31].

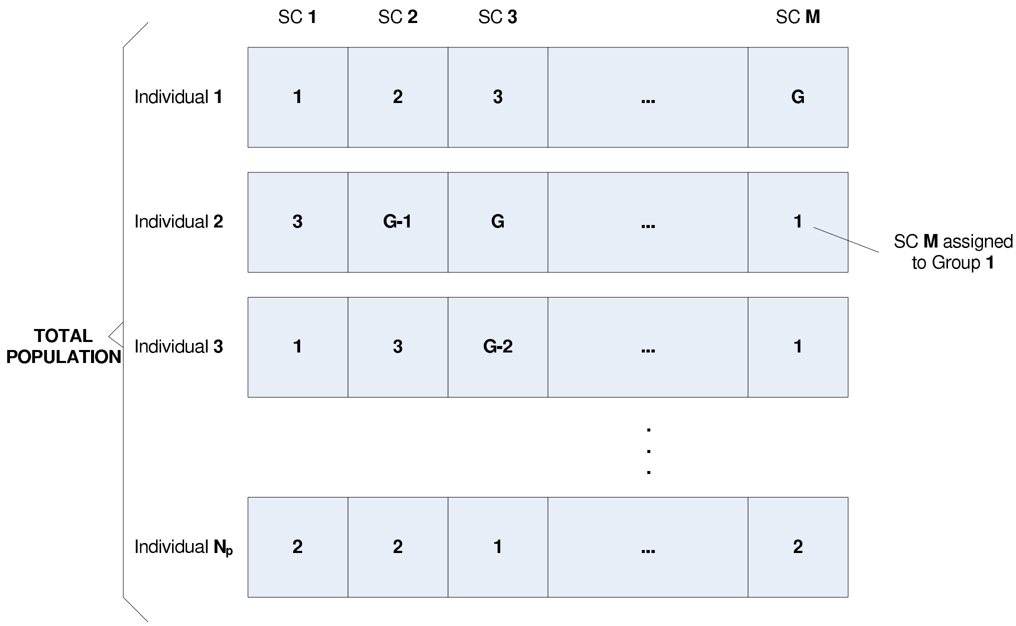


Figure 3.2: Coding of individuals and total population in one generation.

will be given fitness value of $-\infty$. Otherwise, by performing the water-filling of power over the known subcarrier assignment as described in Section 3.3, the fitness score of this subcarrier-power allocation can be computed. Since the objective is to maximize the system throughput, individuals with higher fitness values (that is, higher sum rates) are preferable in the proposed solution.

Producing Next Generation

To produce the next generation, the following rules apply and their operations are also illustrated in Fig. 3.3.

(i) **Elite Children Rule**: Elite children are the individuals in the current generation with the best fitness values. These individuals automatically survive to the next generation. We propose the number of elite children N_e in our genetic algorithm to be fewer than 5 as setting this number to a high value causes the fittest individuals to dominate the population, which in turn may lead to a less effective search.

(ii) **Crossover Rule**: Crossover enables the algorithm to extract the best genes from different individuals and recombine them into potentially superior children. In our

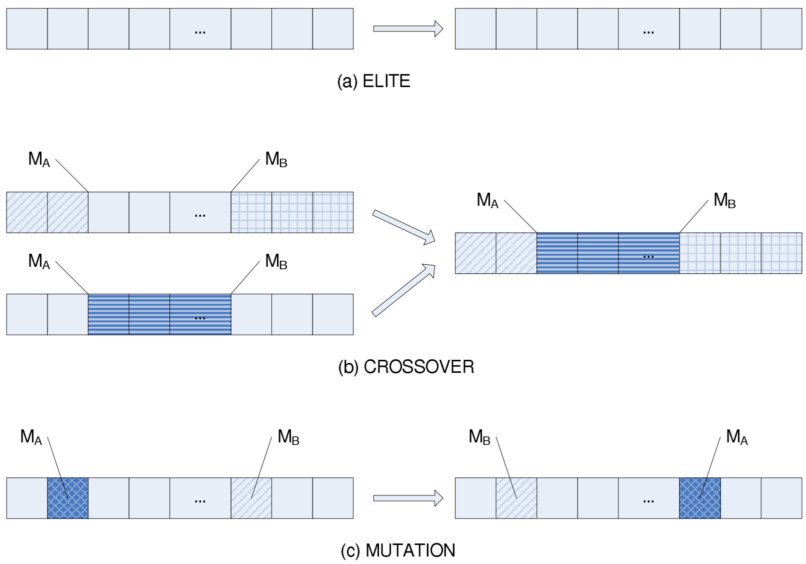


Figure 3.3: Operations to create new generations.

proposed scheme, we apply a two-point crossover rule which selects two unequal points M_A and M_B at random ($1 \leq M_A, M_B \leq M$). The child has the vector entries (genes) of the first parent at the locations before M_A and after M_B, and the vector entries (genes) of the second parent after M_A and before M_B.

(iii) **Mutation Rule**: The mutation process adds to the diversity of a population and hence increases the likelihood that the algorithm will generate individuals whose fitness values are better. Here, we propose a swapping of two randomly selected entries in a single parent to produce a new child.

Stopping Criteria

The proposed GA is terminated when at least one of the following conditions is met:

(i) A maximum number of generations $L_{\max}$ is exceeded.

(ii) The number of generations, over which a cumulative change in fitness function value is less than a tolerance value ϵ, exceeds $L_{\lim}$.

3.5 Complexity Analysis and Performance Evaluation

3.5.1 Complexity Analysis

Regarding the resource allocation problem (3.5)–(3.10), an optimal search can be accomplished via exhaustive comparison of all G^M possible subcarrier assignments, each of which requires a total of M runs of power water-filling to compute the achieved throughput. As a result, the direct search has an exponential complexity of $\mathcal{O}(G^M M)$. On the other hand, after obtaining the matrix $\bar{R}$ via GM operations, both the BC-SO and RCBC-SO schemes only need to perform either a one- or two-dimensional search to find the eligible group corresponding to individual subcarriers. Once an optimal solution of subcarrier assignment has been found, the actual number of water-filling executions in these cases is simply M. Assuming that a search through a one-dimensional (sorted) list with K elements is of $\mathcal{O}(K \log K)$ complexity, then the total number of operations required by the BC-SO approach is indeed $GM + GM^2 \log(GM) + M$ whereas that by the RCBC-SO design is only $GM + GM \log G + M$. It is worth pointing out that the considerably lower complexity of the two proposed algorithms is mainly attributed to the separation of $\{\rho_{g,m}\}$ and $\{P_m\}$ variable sets with the assumption of uniform power distribution, as previously discussed in Section 3.3.1.

In contrast, complexity of BC-GA scheme depends on the maximum number of generations $L_{\max}$ required to be produced before the algorithm terminates, as well as on the size N_p of each generated population. Within a population, water-filling of power is completed for each individual in the computation of fitness scores, followed by a one-dimensional search to select the most fitted individuals. It should be noted that the efficiency of a genetic-algorithm-based approach also depends on other factors (such as the choice of initial population, the rules to produce new generations and the tolerance allowable for cumulative changes in fitness scores), which can be difficult to explicitly quantify. Excluding these factors, the total complexity of the BC-GA scheme can be shown to be $\mathcal{O}\left(L_{\max} N_p (M + \log N_p)\right)$.

All of the above-mentioned analyses are summarized in Table 3.1. Compared with the optimal exhaustive search, the three proposed methods clearly require far less computational effort. However, this benefit comes at the cost of sacrificing attainable system throughput as the devised schemes, by their nature, are suboptimal. In selecting suitable algorithms for different applications, it is therefore critical to balance the contradicting re-

Table 3.1: Complexity analysis.

Algorithm	**Number of Operations**	**Order of Complexity**
Optimal search	$G^M M$	$\mathcal{O}(G^M M)$
BC-SO scheme	$GM + GM^2 \log(GM) + M$	$\mathcal{O}(GM^2 \log(GM))$
RCBC-SO scheme	$GM + GM \log G + M$	$\mathcal{O}(GM \log G)$
BC-GA scheme	$L_{\max} N_p(M + \log N_p)$	$\mathcal{O}\left(L_{\max} N_p(M + \log N_p)\right)$

Table 3.2: Parameters for the BC-GA scheme.

Parameter	N_p	N_e	$L_{\max}$	$L_{\lim}$	ϵ
Value	32	2	60	20	10^{-6}

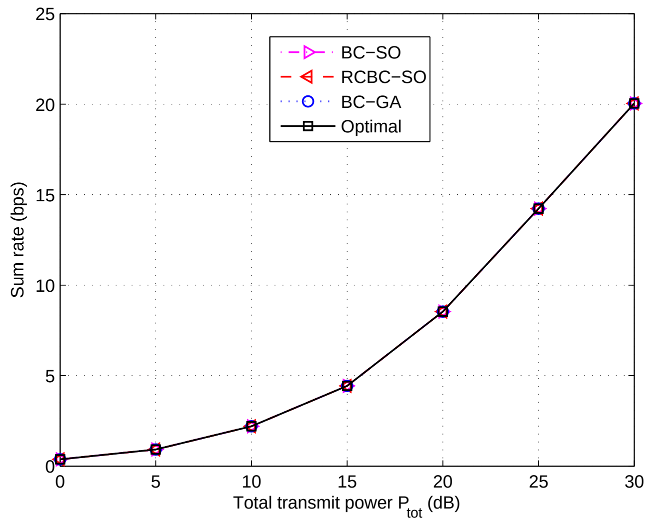
quirements of reducing computational burden and achieving the highest possible sum rates. In what follows, some numerical examples are provided to evaluate the performance of the proposed designs in various scenarios.

3.5.2 Numerical Examples

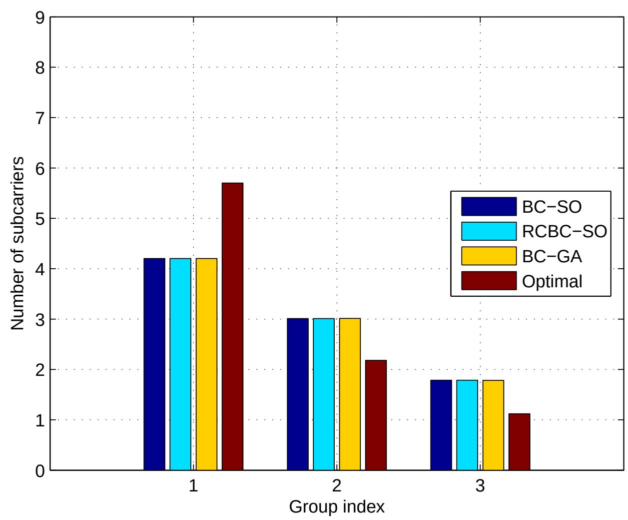
Considered is an OFDMA system with $M = 9$ subcarriers in which the BS communicates with $G = 3$ multicast groups, each has equal $|K_1| = |K_2| = |K_3| = 4$ users. Assume that K_1 is located closer to the BS thus causes a pathloss advantage of 1.5dB to K_2, and of 3dB to K_3. To have a meaningful interpretation of the results, 100 sets of independent channel coefficients $\{h_{k,m}\}$ are randomly generated according to the Rayleigh distribution in each simulation study. The equivalent CSNR of group K_g on subcarrier m is computed as $\beta_{g,m} = \min_{k \in K_g} |h_{k,m}|^2$. The final results are then averaged for plotting. For simplicity, the average channel gain, the noise power in each subcarrier, and the individual subcarrier bandwidth are all normalized to 1.

We shall now demonstrate three illustrative examples wherein the values of bandwidth-control indices α_g $(g = 1, 2, 3)$ are properly adjusted to either provide throughput maximization, or offer a fair spectrum sharing by guaranteeing certain portions of the total available bandwidth to be designated to individual multicast groups. Performance of the proposed solutions (namely BC-SO, RCBC-SO and BC-GA) are to be compared against one another and also with that of optimal exhaustive search. In all the examples presented here, the parameters used for the BC-GA scheme are listed in Table 3.2.

First, notice that by not guaranteeing any minimum numbers of subcarriers to be allo-



(a) Achieved throughput



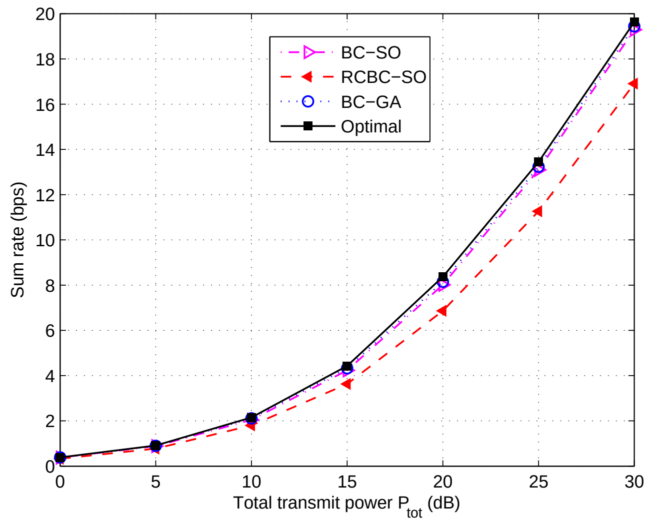
(b) Distribution of total bandwidth (averaged)

Figure 3.4: Performance in the SRM example.

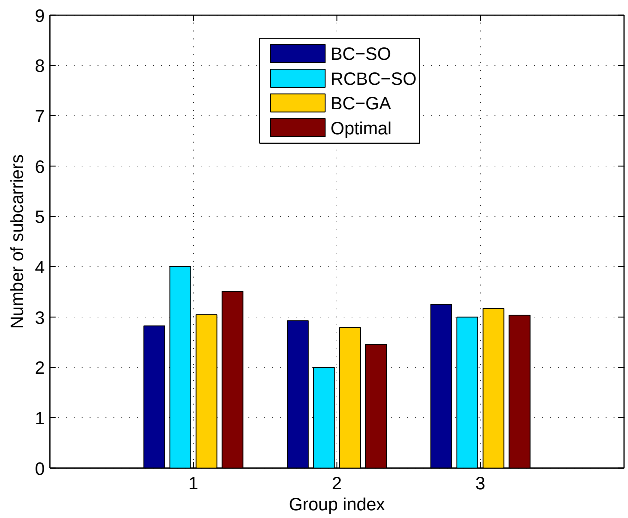
cated to individual multicast groups, that is, by setting $\alpha_1 = \alpha_2 = \alpha_3 = 0$, the formulation (3.5)–(3.10) actually becomes the problem of throughput maximization. In this case, there is a free competition among K_1, K_2 and K_3, and the group who contributes the most to the total sum rate will finally secure the available system resources for its own usage. We will refer to this example as Sum Rate Maximization (SRM), where it is clear from Fig. 3.4a that all the proposed algorithms approach optimality. This numerical result, in particular, verifies that an equal transmit power allocation hardly decreases the data throughput of an OFDMA-based system since each subchannel is only given to a user whose channel gain is good in that subchannel. Further, because Step 1 of both Algorithms 1 and 2 is omitted in the SRM example, the BC-SO and RCBC-SO schemes reduce to a throughput maximization algorithm and hence perform identically. In terms of bandwidth sharing, the proposed algorithms allocate more subcarriers to the group with better link conditions in this case, as can be clearly seen in Fig. 3.4b. It should also be pointed out that although the optimal search assigns more subcarriers to the advantageous groups, these subcarriers might have been distributed zero power by the water-filling procedure, resulting in no improvement in the attained throughput at all.

Since groups K_2 and K_3 are located farther away from the BS, their effective equivalent channel gains (including long-term pathloss and short-term fading) are potentially smaller than that of group K_1. The former groups are therefore likely to be in disadvantageous position, having fewer chances to gain access into the available radio spectrum. As such, in this second example we impose $\alpha_1 = 1,\ \alpha_2 = 2,\ \alpha_3 = 3$ to ensure a fairer allocation in terms of bandwidth to the disadvantaged (inferior) groups K_2 and K_3. The remaining 3 subcarriers are then open for competition among the three groups. We will refer to this example as Inferior-fair Bandwidth Allocation (IBA). From Fig. 3.5a, it can be seen that the sum rates achieved by the BC-SO and BC-GA schemes are only 5% away from optimality whereas the simple RCBC-SO design attains even more than 82% of the optimal throughput. In addition, the total bandwidth has been shared more fairly among the multicast groups, as can be seen in Fig. 3.5b. Note that the values of α_g in this example are chosen for an illustrative purpose only and they are completely adjustable at the discretion of the system designer. If the channel condition of the worst user in group K_2 or K_3 remains unfavorable for a relatively long period of time, it becomes necessary to readjust the value of α_g to avoid an unacceptable sacrifice in the system throughput (one, for instance, may opt to increase α_1 and decrease α_2, α_3).

Even more strictly, a totally fair bandwidth allocation for all three multicast groups can be enforced by setting $\alpha_1 = \alpha_2 = \alpha_3 = 9/3 = 3$, in which case each group will be given exactly a third of the accessible bandwidth regardless of its respective channel state. This example shall be referred to as Equal Bandwidth Allocation (EBA). Fig. 3.6a illustrates that the sum rates obtained by the proposed solutions are very close to that offered by the optimal search, with both the BC-SO and the BC-GA algorithms achieving more than 97%



(a) Achieved throughput

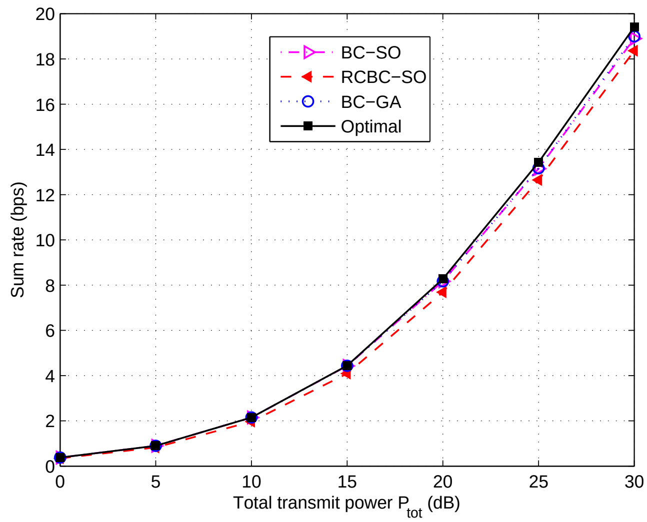


(b) Distribution of total bandwidth (averaged)

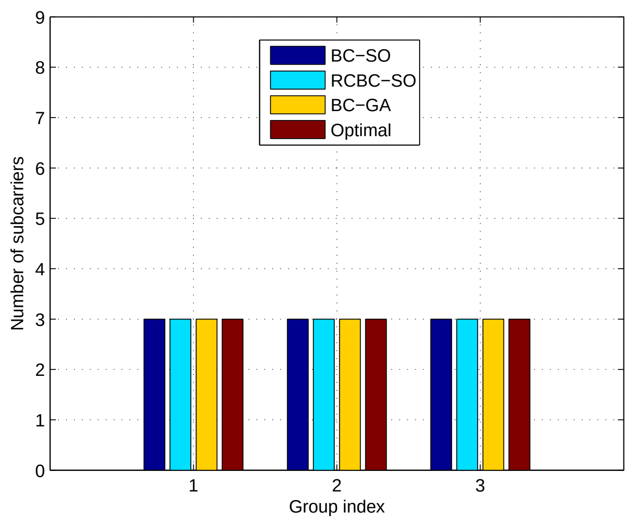
Figure 3.5: Performance in the IBA example.

of the optimal throughput and that for the RCBC-SO solution being above 91%. Regarding the distribution of available bandwidth, Fig. 3.6b verifies that subcarriers have been shared equally among individual multicast groups by all the schemes under investigation in this third example.

The above numerical results have clearly confirmed that by properly adjusting the minimum numbers of subcarriers allocated to individual multicast groups, the design formula-



(a) Achieved throughput



(b) Distribution of total bandwidth (averaged)

Figure 3.6: Performance in the EBA example.

tion and proposed schemes offer a more flexibility in controlling the share of available radio spectrum given to each group and, at the same time, still achieve a high system throughput. In particular, the BC-GA algorithm, with an appropriate choice of parameters, always offers the highest attainable data rate among the three proposals. This is expected since the BC-GA scheme performs a robust global search for jointly optimal solution of subcarriers and power, as opposed to separate optimization of those two variable sets in the BC-SO and

RCBC-SO solutions. Moreover, the RCBC-SO design experiences the lowest throughput among the three at the benefit of having a significantly lower computational complexity.

3.6 Concluding Remarks

This chapter proposed three efficient low-complexity resource allocation schemes for OFDMA-based multicast wireless systems. The novelty in the proposed schemes is that the issue of controllable and flexible distribution of the available radio spectrum among multicast groups was explicitly taken into account. In the separate optimization schemes, the subcarrier allocation ensures minimum numbers of subcarriers to be assigned to individual groups according to their respective channel gains and group sizes, while power is allocated in a water-filling fashion. With the scheme based on the modified genetic algorithm, the jointly optimal subcarrier-power allocation is iteratively evolved through a global search while satisfying the imposed bandwidth constraints among different multicast groups. Numerical examples showed that the proposed designs can be utilized to attain a high total sum rate while at the same time distributing the available bandwidth more flexibly and fairly among multicast groups. The computational complexity of our proposed approaches has been analyzed, and their benefits have also been confirmed by numerical examples.

Chapter 4

Optimal Resource Allocation for OFDMA-based Cognitive Radio Multicast Networks

In November 2002, the Federal Communications Commission (FCC) published a report on the current management of the precious radio spectrum resource in the United States. One of the main findings stated in the report is [32]

> "In many bands, spectrum access is a more significant problem than physical scarcity of spectrum, in large part due to legacy command-and-control regulation that limits the ability of potential spectrum users to obtain such access."

Simply put, it has been confirmed that much of the licensed spectrum lies idle at any given time and location, and that the spectrum shortage results from the spectrum management policy rather than the physical scarcity of the usable frequencies. Spectrum utilization can thus be significantly improved by allowing secondary users to access spectrum holes unoccupied by the primary users at given locations and times. Cognitive radio [33, 34] has been identified as an efficient technology to promote this idea by exploiting the existence of the spectrum portions unoccupied by the primary (or licensed) users. Potentially, while the primary users have priority access to the spectrum, the secondary (or unlicensed or cognitive) users have restricted access, subject to a constrained degradation on the primary users' performance [35]. In spectrum sharing environments, the key design challenges of a cognitive radio network are therefore to guarantee a protection of the primary users from excessive interference induced by the secondary users as well as to meet some Quality-of-Service (QoS) requirements for the latter [36, 37].

Chapter 4, which examines the design of a dynamic resource allocation algorithm for OFDMA-based multicast cognitive radio networks with primary user activity (or subchannel availability) consideration, is organized as follows[1]. Section 4.1 reviews related works on the similar subject in literature and also highlights the original research contributions of this study. Section 4.2 presents the system model under consideration. Also formulated in this section is the resource allocation problem for OFDMA-based multicast secondary networks where primary user activities are taken into account. Section 4.3 introduces the dual optimization method, an effective approach to deal with a large class of multi-carrier resource allocation problems. In Section 4.4, an iterative scheme, derived from the dual framework to resolve the design under investigation, is proposed. Section 4.5 provides numerical examples to verify performance of the devised solution. Finally, Section 4.6 concludes the chapter with several remarks.

4.1 Introduction

4.1.1 Spectrum Pooling Approach for Opportunistic Spectrum Access

Spectrum pooling is an opportunistic spectrum access approach that enables public access to the already licensed frequency bands [40, 41]. The basic idea is to merge spectral ranges from different spectrum owners (for example, military, trunked radios) into a common pool, from which the secondary users may temporarily rent spectral resources during idle periods of licensed users. In effect, the licensed system does *not* need to be changed while the secondary users access unused resources. Among many possible technologies for unlicensed users' transmission in spectrum-pooling radio systems, orthogonal frequency division multiplexing (OFDM) has already been widely recognized as a highly promising candidate, mainly due to its great flexibility in dynamically allocating the unused spectrum among secondary users as well as its ability to monitor the spectral activities of licensed users at no extra cost [42]. However, it has been shown that employing OFDM also affects the performance of a cognitive radio network, for instance, causing mutual interference between the primary and secondary users due to the non-orthogonality of respective transmitted signals [43, 44].

[1]A version of this chapter has been presented at the 2009 IEEE Wireless Communications and Networking Conference (WCNC'09) held in Budapest, Hungary [38]. Further development of the results has been submitted for publication as a regular paper in the IEEE Transactions on Vehicular Technology [39].

4.1.2 Resource Allocation in OFDM-based Cognitive Radio Systems

Resource allocation for OFDM-based cognitive radio networks has been examined in [45], where an optimal scheme, derived via Lagrangian formulation, is proposed to maximize the downlink capacity of a single cognitive user while guaranteeing the interference to the primary user being below a specified threshold. The work of [46] extends [45] to multiuser scenarios, in which discrete sum rate of the secondary network is maximized constrained on the interference to the primary user bands and also on the total transmitted power. Subject to the per-subchannel power constraints (due to primary users interference limits), the study in [47] proposes a partitioned iterative water-filling algorithm that enhances the capacity of an OFDM cognitive radio system. Further, the issue of downlink channel assignment and power control for FDMA-based cognitive networks has been addressed in [48], wherein a set of base stations make opportunistic spectrum access to serve the fixed-location wireless users within their cells. To maximize the total number of active users that can be supported while guaranteeing the minimum signal-to-interference-plus-noise ratio (SINR) requirements of secondary users and also protecting the primary users, suboptimal schemes are suggested for the formulated mixed-integer program. Considering networks with the coexistence of multiple primary and secondary links through OFDMA-based air-interface, reference [49] utilizes the dual framework from [50] to provide centralized and distributed algorithms for improving the total achievable sum rate of secondary networks subject to interference temperature constraints specified at primary users' receivers.

While the previous related studies implicitly assume that the designated spectrum for secondary usage is fixed and always available, the work of [51] investigates another important aspect of subchannel availability or primary user activity in an OFDM cognitive radio system. Here, cognitive radio can be realized as a risky environment where the licensed users may, at any time, come back and take up the frequency bands currently available for secondary access. In such scenarios, the power already invested by unlicensed users in those bands becomes wasted. By referring to a risk-return model and upon defining a general rate-loss function which gives a decrease in total throughput whenever primary users reoccupy the temporarily accessible subchannels, a problem of optimally allocating power for a single cognitive user is formulated incorporating the reliability or the availability of OFDM subchannels. For the special case of linear rate loss, multi-level water-filling solution for the resulting convex program has been derived in [51], but other types of rate-loss functions are not yet investigated.

4.1.3 Research Contributions

Different from all the aforementioned works which only consider unicast transmission, this research studies resource allocation in a secondary OFDMA-based multicast network where the patterns of primary user activities on the available radio spectrum are dynamic. As an efficient means of transmitting the same content to multiple receivers while minimizing the network resource usage, multicasting [25, 26] is clearly an attractive transmission technique for secondary networks who only have a limited access to the available spectrum. However, in the multi-group/multi-user settings, the problem of joint subcarrier assignment and power distribution usually turns out to be non-convex, making the solution derived in [51] no longer applicable. As well, performing an direct exhaustive search to find the global optimal solutions is certainly impractical in these cases as computational complexity of such approach is prohibitively demanding. Motivated by the shortcomings of the existing designs, we propose in this work a dual optimization scheme to efficiently solve the challenging resource allocation in a cognitive OFDMA network consisting of multiple multicast groups.

Adopting a similar risk-return model of [51] to account for the primary user activities, our proposed subcarrier assignment and power allocation solution targets at maximizing the expected sum rate of all secondary users in an OFDMA-based cognitive radio multicast network, while satisfying the tolerable interference level induced to individual licensed users. Specifically, the original non-convex optimization problem is solved effectively in the dual domain with global optimum obtained in the limit as the number of subcarriers goes to infinity. More significantly, it is shown that the proposed approach has only linear complexity in the total number of subcarriers, resulting in a huge reduction in computational burden. These features are certainly attractive for practical OFDMA-based systems that deploy a large number of subcarriers. Further, the dual approach presented here is valid for both unicast and multicast scenarios, and is applicable for a wide range of rate-loss functions among which linear being a special case. As well, the mutual interference between secondary and primary networks, which is an important factor, is explicitly quantified. The effects of adjacent subcarrier nulling technique [43], used to decrease the mutual interference, on the proposed design are also carefully analyzed.

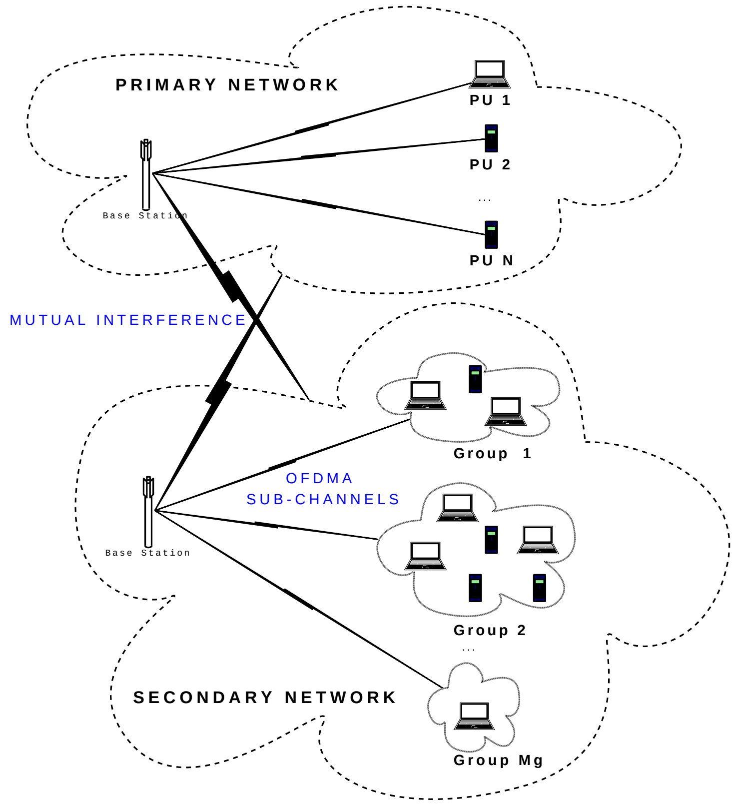
4.2 System Model and Problem Formulation

4.2.1 System Model

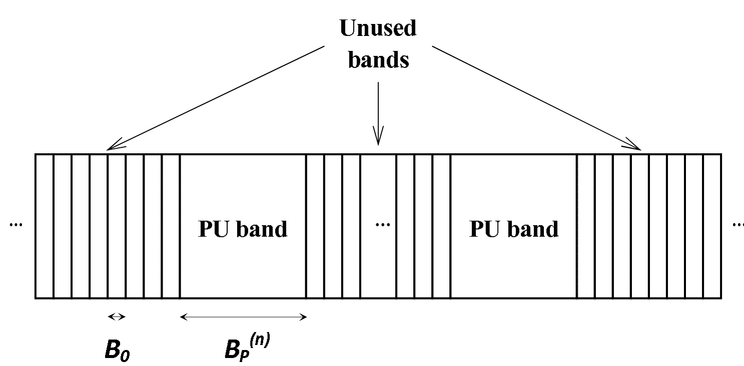
We consider a primary base station (BS) that transmits (not necessarily OFDM) signals to its N primary users, each of which occupies a predetermined frequency band $B_P^{(n)}$ ($n = 1, \cdots, N$) in the available spectrum. To implement efficient opportunistic spectrum access, a secondary BS is permitted to employ K OFDM subcarriers to transmit G downlink traffic flows, each to one distinct multicast group consisting of secondary users, over the temporarily unused/available frequency bands. Information regarding the availability of these bands is made known at the secondary BS either by means of signalling from the primary BS, or as the result of spectrum sensing performed by the secondary BS itself. Notice that since licensed users have priority access to the radio spectrum, the unused frequency bands need, at any time, to be handed back to the primary network upon request. Therefore, depending on the activity of primary users, there is a chance that the temporarily unused spectrum becomes reoccupied.

Assume that each secondary user receives one traffic flow at a time, and hence it belongs to only one multicast group. Let M_g and $|M_g|$ ($g = 1, \cdots, G$) denote the user set of group g and its cardinality, respectively. The g-th group is unicast if $|M_g| = 1$, whereas it is multicast if $|M_g| > 1$. Thus, the system framework presented here is applicable to the both unicast and multicast transmissions. Clearly, all the secondary users belong to the set $M = \bigcup_{g=1}^{G} M_g$, and $|M| = \sum_{g=1}^{G} |M_g|$ is the total number of users in the cognitive multicast network. Let B denote the total bandwidth available for secondary usage, and also assume that each subchannel has an equal bandwidth of $B_s = B/K$. The system setup is depicted in Fig. 4.1a with the distribution of accessible spectrum shown in Fig. 4.1b.

As the consequence of having two coexisting networks, the OFDM signals from secondary BS, which are intended for its own serviced users, might interfere the reception at the primary users' receivers. Upon defining $\check{P}_{m,k}$ as the power spent for transmitting to secondary user m in group g on subcarrier k and denoting T_s the OFDM symbol duration, the power spectral density (PSD) of the subcarrier-k signal can be modeled as $\Phi_k(f) = \check{P}_{m,k} T_s \left(\frac{\sin \pi f T_s}{\pi f T_s}\right)^2, \ m \in M_g$. Then, the interference caused by this signal



(a) Coexistence of primary and secondary networks in cognitive radio.



(b) Distribution of available spectrum to primary and secondary users.

Figure 4.1: System model.

onto primary user n is given as [43]

$$\begin{aligned} I_k^{(n)} &= |g_{SP}^{(n)}|^2 \int_{d_k^{(n)}-B_P^{(n)}/2}^{d_k^{(n)}+B_P^{(n)}/2} \Phi_k(f)\,\mathrm{d}f \\ &= \check{P}_{m,k} \left\{ |g_{SP}^{(n)}|^2 T_s \int_{d_k^{(n)}-B_P^{(n)}/2}^{d_k^{(n)}+B_P^{(n)}/2} \left(\frac{\sin \pi f T_s}{\pi f T_s} \right)^2 \mathrm{d}f \right\} \\ &= \check{P}_{m,k} \check{I}_k^{(n)}, \end{aligned} \tag{4.1}$$

where $d_k^{(n)} = |f_k - f_n|$ represents the spectral distance between subcarrier k and the center frequency of primary user n, and $g_{SP}^{(n)}$ denotes the channel from secondary BS to primary user n. Clearly, interference $I_k^{(n)}$ depends on both $d_k^{(n)}$ and $\check{P}_{m,k}$.

In addition, the coexistence of primary users and multicast groups of secondary users may cause interference induced by the signals from primary BS, which are destined to primary users, onto secondary users' frequency bands. Let $g_{m,k}^{PS}$ be the channel from primary BS to secondary user m in group g on subcarrier k, and $\Phi_{PU}^{(n)}(\mathrm{e}^{jw})$ be the PSD of the signal transmitted from primary BS to primary user n. Then, the interference power caused by this signal onto secondary user $m \in M_g$ on subcarrier k can be computed as [43]

$$J_{m,k}^{(n)}(d_k^{(n)}) \quad = \quad |g_{m,k}^{PS}|^2 \int_{d_k^{(n)}-B_s/2}^{d_k^{(n)}+B_s/2} \mathcal{E}\{I_K(w)\}\,\mathrm{d}w, \tag{4.2}$$

where $\mathcal{E}\{I_K(w)\} = \frac{1}{2\pi K}\int_{-\pi}^{\pi} \Phi_{PU}^{(n)}(\mathrm{e}^{jw}) \left(\frac{\sin(w-\phi)K/2}{\sin(w-\phi)/2}\right)^2 \mathrm{d}\phi$ is the PSD of primary user n's signal after K-FFT processing.

In this work, the resource allocation of secondary network is accomplished in a centralized manner with perfect channel state information of all primary and secondary users in the system being assumed (for example, via training and feedbacks from the users through dedicated channels). We further assume that the channel conditions remain unchanged during the allocation period. Hence, this model is particularly valid for slowly-varying channels where the channel gains do not vary too significantly over time, such as in high data rate systems and/or environments with reduced degrees of mobility [3,5,6]. With the perfect link information available, it is therefore possible to determine the maximum rate at which an individual secondary user can reliably receive data, as well as the corresponding subcarrier over which the data shall be transmitted on. The channel signal-to-interference-plus-noise ratio (CSINR) of secondary user $m \in M_g$ on subcarrier k can be shown to be

$$\alpha_{m,k} = \frac{|h_{m,k}^{SS}|^2}{\Gamma(N_0 B_s + \sum_{n=1}^{N} J_{m,k}^{(n)})}, \tag{4.3}$$

where $h_{m,k}^{SS}$ is the corresponding channel coefficient and N_0 is the one-sided PSD of additive white Gaussian noise (AWGN). The parameter Γ represents the signal-to-noise ratio (SNR) gap to the capacity limit, which is a function of the desired Bit-Error-Rate (BER), coding gain and noise margin [52]. The maximum attainable rate of secondary user $m \in M_g$ on subcarrier k is then

$$r_{m,k} = \frac{B_s}{B} \log_2(1 + \alpha_{m,k}\check{P}_{m,k}). \tag{4.4}$$

Similar to the study in Chapter 3, the conventional multicast transmission approach is also followed here by enforcing the secondary BS to transmit at the lowest rate of all the users within a group, which is determined by the user with the smallest CSINR [27]. This assures that the multicast services can be provided to *all* the subscribed users. Let

$$\gamma_{g,k} \quad = \quad \min_{m \in M_g} \alpha_{m,k} \tag{4.5}$$

be the equivalent CSINR of group g on subcarrier k. Then, the maximum rate at which all secondary users of group g are able to decode the data transmitted on that same subcarrier is

$$\check{r}_{g,k} = \frac{B_s}{B} \log_2(1 + \gamma_{g,k} P_{g,k}), \tag{4.6}$$

where $P_{g,k}$ denotes power allocated to group g on subcarrier k. Since all the secondary users in a group receive the same rate, the aggregate rate transmitted to group g on subcarrier k is scaled by the group size as

$$\check{R}_{g,k} = \sum_{m \in M_g} \check{r}_{g,k} = |M_g| \, \check{r}_{g,k}. \tag{4.7}$$

4.2.2 Problem Formulation with Primary User Activity Consideration

In a cognitive radio environment, there is likely a delay from the moment that a channel is made available for secondary usage to the time that the secondary network is fully aware of that accessibility. The time delay could be due to, for example, the efficiency of spectrum sensing algorithms performed by the cognitive network. This effect is of particular concern if the patterns of spectrum usage by primary users are greatly dynamic, for instance, frequent occurrences of releasing and reoccupying certain bands. Consequently, secondary BS may carry out the resource allocation at the current time frame t with the available information (regarding, for example, locations of spectrum holes, link conditions, interference, etc.) valid at time $t - \Delta t$. During the time delay $\Delta t > 0$, it is possible that primary users may have come back and taken up the subchannels that were once available at time $t - \Delta t$. As this is the case, performance of current resource allocation for secondary network can be severely affected.

To account for the primary user activities (or equivalently, the availability of OFDM subchannels), we refer to the risk-return model in which power allocated to a frequency band is considered an investment in that band [51]. In this model, cognitive radio environment can be thought as a risky environment where the primary users may return to take up the available band at any time. In such cases, the secondary users power investment in that band becomes wasted, representing a loss in the data rate achieved by the secondary users, possibly due to, for instance, other better allocation schemes that could have been utilized, or an increase in the amount of interference caused to primary users when the unused bands are reoccupied. In order to model this loss, we define a rate loss $L(P)$ which is a function of the power invested by cognitive network. Strictly, $L(P)$ is required to satisfy the following two conditions:

(i) $L(P) > 0$ for $P > 0$, and

(ii) $L(P) = 0$ for $P = 0$.

Given the probability ϕ_k that the subchannel k is taken up by the primary users in the current time frame, the expected rate-loss can be written as

$$\mathcal{E}\{\Delta R_{g,k}\} \quad = \quad \phi_k L(P_{g,k}). \tag{4.8}$$

Then, the expected rate transmitted from secondary BS to group g on subcarrier k becomes

$$\begin{aligned} \mathcal{E}\{R_{g,k}\} &= \check{R}_{g,k} - \mathcal{E}\{\Delta R_{g,k}\} \\ &= \frac{|M_g|}{K} \log_2(1 + \gamma_{g,k} P_{g,k}) - \phi_k L(P_{g,k}). \end{aligned} \tag{4.9}$$

The goal of this study is to devise a subcarrier assignment and power allocation policy that maximizes the expected sum rate of all multicast groups of secondary users, while satisfying constraints on the tolerable interference level of each individual primary user. Specifically, the design problem can be formulated as follows:

$$\max_{\{P_{g,k}\}} \quad \sum_{g=1}^{G} \sum_{k=1}^{K} \frac{w_g |M_g|}{K} \log_2(1 + \gamma_{g,k} P_{g,k}) - \phi_k L(P_{g,k}) \tag{4.10}$$

$$\text{s.t.} \quad \sum_{g=1}^{G} \sum_{k=1}^{K} P_{g,k} \check{I}_k^{(n)} \leq I_{\text{th}}^{(n)};\ n = 1, \cdots, N, \tag{4.11}$$

$$P_{g,k} \geq 0;\ g = 1, \cdots, G,\ k = 1, \cdots, K, \tag{4.12}$$

$$P_{g,k} P_{g',k} = 0;\ \forall g' \neq g. \tag{4.13}$$

In this formulation, weight $w_g \geq 0$ reflects the priority designated to group g and is obliged to satisfy $\sum_{g=1}^{G} w_g = 1$. Constraint (4.11) expresses the tolerable interference level at the receiver of primary user n, with $I_{\text{th}}^{(n)}$ representing the interference temperature threshold.

Constraints (4.12)-(4.13) enforce a disjoint subchannel assignment in OFDMA systems, that is, one subcarrier is permitted to be assigned to at most one group at a time [3]. It is note-worthy that the optimization problem (4.10)–(4.13) is NP-hard since it requires the allocation of an optimal set of subcarriers to each multicast group of secondary users. Complexity needed to directly solve this combinatorial problem increases, at least, exponentially with the number of subcarriers K. Such prohibitively high computational effort is required even for a simplified case as discussed in Appendix C. Moreover, the multiple constraints in (4.11) make it even more challenging to derive an analytical solution for problem (4.10)–(4.13).

In the following sections, we will first introduce the dual optimization method for non-convex multicarrier resource allocation. Then, we will show how the optimization problem (4.10)–(4.13) can be effectively resolved in the dual domain with virtually zero duality gap, and thereby global optimal solutions can be obtained in the limit as the number of subcarriers goes to infinity. Further, we establish that the complexity of the proposed dual scheme is only linear in the total number of subcarriers, and thus representing a significant reduction in computational burden at the BS, where it is desirable to find the optimal solutions rapidly to mitigate the fluctuations of wireless channels.

4.3 Dual Optimization of Non-convex Multicarrier Resource Allocation

Consider the problem of optimally allocating resources in a multicarrier system with M users and K subcarriers. The objective and constraints of the optimization consist of a number of individual functions, each corresponding to one of the K subcarriers, and can be expressed as

$$\begin{aligned} \max_{\{\mathbf{x}_k\}\in\mathcal{R}^M} \quad & \sum_{k=1}^{K} f_k(\mathbf{x}_k) \\ \text{s.t.} \quad & \sum_{k=1}^{K} \mathbf{h}_k(\mathbf{x}_k) \leq \mathbf{P}, \end{aligned} \tag{4.14}$$

where $f_k(\cdot)$ are $\mathcal{R}^M \to \mathcal{R}$ (not necessarily concave) functions, $\mathbf{h}_k(\cdot)$ are $\mathcal{R}^M \to \mathcal{R}^N$ (not necessarily convex) functions, and constant $\mathbf{P}$ denotes the N-vector of constraints.

The idea of dual optimization is to solve (4.14) by first forming its Lagrangian dual, which is defined as [16]

$$\mathcal{L}(\{\mathbf{x}_k\}, \boldsymbol{\lambda}) \;=\; \sum_{k=1}^{K} f_k(\mathbf{x}_k) - \boldsymbol{\lambda}^T \left(\sum_{k=1}^{K} \mathbf{h}_k(\mathbf{x}_k) - \mathbf{P} \right), \tag{4.15}$$

where $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_N]^T \succeq \mathbf{0}$ is a vector of Lagrange dual variables.

Then, upon defining the dual objective as $D(\boldsymbol{\lambda}) = \max_{\{\mathbf{x}_k\}} \mathcal{L}(\{\mathbf{x}_k\}, \boldsymbol{\lambda})$, the dual optimization problem of (4.14) becomes

$$\begin{aligned} \min_{\boldsymbol{\lambda}} \quad & D(\boldsymbol{\lambda}), \\ \text{s.t.} \quad & \boldsymbol{\lambda} \succeq \mathbf{0} \end{aligned} \tag{4.16}$$

The Lagrange dual problem is a convex optimization problem, which can be solved very efficiently in practice. This is the case whether or not the primal problem is convex. Nevertheless, solving a dual problem is not always equivalent to solving the primal one. Let f^* and D^* denote the primal and the dual optimal values, respectively. Then the difference $d = D^* - f^*$ is defined as the optimal duality gap. It has been proven from duality theory that $d \geq 0$ always holds. In particular, when $f_k(\mathbf{x}_k)$'s are concave and $\mathbf{h}_k(\mathbf{x}_k)$'s are convex (that is, problem (4.14) is convex), strong duality is guaranteed, which implies $d = 0$. In such cases, the primal and dual problems have the same optimal value and thus the globally optimal solution can be derived in the dual domain via Lagrangian decomposition. However, this gap in general is not always zero and the optimal solution of the dual problem only gives the best upper bound on that of the primal.

Interestingly enough, it has been shown in [50, 53] that even if the multicarrier optimization problem (4.14) is non-convex, the duality gap is zero if either of the following conditions is met.

Condition 1 $\mathbf{x}_k^*(\boldsymbol{\lambda}) = \arg\max_{\mathbf{x}_k} \mathcal{L}(\{\mathbf{x}_k\}, \boldsymbol{\lambda})$ *is continuous at optimal* $\boldsymbol{\lambda}^*$.

Condition 2 *The optimal value of* $\sum_{k=1}^{K} f_k(\mathbf{x}_k)$ *is concave in* $\mathbf{P}$.

It turns out that for the non-convex problem with a general form as in (4.14), Condition 2 in the above, called the frequency-sharing condition [50], is always satisfied in the limit as the number of subcarriers K approaches infinity. A more general theoretical justification of this observation can be found in [18, Sec. 5.1.6]. Significantly, this important result allows the original challenging non-convex problem to be efficiently solved in the dual domain with a virtually negligible duality gap for a realistically large number of subcarriers.

4.4 Practically Optimal Subcarrier and Power Allocation via Duality

It can be observed that the particular structure of problem (4.10)–(4.13) satisfies the frequency-sharing condition, and hence its global optimum can be obtained in the dual domain by an iterative method, at a significantly reduced computational complexity [50]. In brief, for a fixed Lagrange dual variable set, it is possible to first decompose (4.10)–(4.13) by Lagrangian into several unconstrained per-tone power allocation subproblems, each of which can be solved by water-filling or exhaustive search. Once the optimal distribution of powers is found for all subcarriers, the Lagrange dual variables are updated by a subgradient-based or an ellipsoid method. The procedure is repeated until convergence, and the optimal solution of subcarrier and power allocation obtained in the dual domain becomes that of the primal problem (4.10)–(4.13) as the number of subcarriers tends to be large.

4.4.1 Resource Allocation for Linear Loss Function

For the purpose of demonstration, let us for now consider a linear rate-loss function $L(P_{g,k}) = C \cdot P_{g,k}$ where C is the normalized average cost per unit power for the secondary users to utilize the resource. Note that the intention to choose a linear function is to simplify the analysis while giving a better understanding of the proposed approach. The dual method presented here is as well applicable for other types of rate-loss function, which shall be discussed in a later section. With this linear loss, the objective (4.10) becomes

$$\sum_{g=1}^{G}\sum_{k=1}^{K}\frac{w_g|M_g|}{K}\log_2(1+\gamma_{g,k}P_{g,k}) - C\phi_k P_{g,k}. \tag{4.17}$$

The exclusive channel assignment constraint (4.12)-(4.13) can be expressed as $P_{g,k} \in \mathcal{S}$, where domain $\mathcal{S}$ is defined as $\mathcal{S} = \{P_{g,k} \geq 0;\ g = 1, \cdots, G,\ k = 1, \cdots, K \,|\, P_{g,k}P_{g',k} = 0,\ \forall g' \neq g\}$. Over the domain $\mathcal{S}$, the Lagrangian of problem (4.10)-(4.11) is given as

$$\begin{aligned}\mathcal{L}(\{P_{g,k}\}, \boldsymbol{\lambda}) \quad = \quad & \sum_{g=1}^{G}\sum_{k=1}^{K}\frac{w_g|M_g|}{K}\log_2(1+\gamma_{g,k}P_{g,k}) - C\phi_k P_{g,k} \\ & - \sum_{n=1}^{N}\lambda_n\left(\sum_{g=1}^{G}\sum_{k=1}^{K}P_{g,k}\check{I}_k^{(n)} - I_{\text{th}}^{(n)}\right),\end{aligned} \tag{4.18}$$

where $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_N]^T \succeq \mathbf{0}$ is the vector of dual variables.

Now, thanks to the disjoint subchannel constraint in OFDMA-based systems, it is possible to decompose the Lagrange dual function of (4.18) into K independent optimization problems, one for each subcarrier k, as follows:

$$\begin{aligned} D(\boldsymbol{\lambda}) &= \max_{\{P_{g,k}\}\in\mathcal{S}} \mathcal{L}(\{P_{g,k}\},\boldsymbol{\lambda}) \\ &= \sum_{k=1}^{K} D_k(\boldsymbol{\lambda}) + \sum_{n=1}^{N} \lambda_n I_{\text{th}}^{(n)}, \end{aligned} \tag{4.19}$$

where the per-subcarrier problem is

$$\begin{aligned} D_k(\boldsymbol{\lambda}) = \max_{\{P_{g,k}\}\in\mathcal{S}} \sum_{g=1}^{G} \Bigg\{ & \frac{w_g|M_g|}{K}\log_2(1+\gamma_{g,k}P_{g,k}) \\ & - \left(\sum_{n=1}^{N}\lambda_n \check{I}_k^{(n)} + C\phi_k\right) P_{g,k} \Bigg\}, \end{aligned} \tag{4.20}$$

for $k = 1, \cdots, K$.

For each subcarrier k, there is at most one $P_{g,k} > 0$ for all $g = 1, \cdots, G$. Therefore, the optimal group assignment for subcarrier k can be found by first deriving G optimal power allocations, one for each of the total G groups, and then selecting the one that maximizes $D_k(\boldsymbol{\lambda})$. Assume that multicast group g is active on subcarrier k. For a fixed $\boldsymbol{\lambda}$, the objective of the maximization in (4.20) is a concave function of $P_{g,k}$. Hence, from the Karush-Kuhn-Tucker conditions [16], the optimal power allocation can be devised as

$$P_{g,k}^* = \left(\frac{1}{\gamma_{0,k}} - \frac{1}{\gamma_{g,k}}\right)^+, \tag{4.21}$$

where $x^+ = \max(x, 0)$. Apparently, this is a form of water-filling where the water level is

$$\gamma_{0,k} = \frac{K(C\phi_k + \sum_{n=1}^{N}\lambda_n \check{I}_k^{(n)})\log 2}{w_g|M_g|}. \tag{4.22}$$

Then, by searching over all G possible group assignments for subcarrier k, the optimal value of (4.20) is actually

$$\begin{aligned} D_k^*(\boldsymbol{\lambda}) = \max_g \Bigg\{ & \frac{w_g|M_g|}{K}\log_2\left[1+\gamma_{g,k}\left(\frac{1}{\gamma_{0,k}} - \frac{1}{\gamma_{g,k}}\right)^+\right] \\ & - \left(\sum_{n=1}^{N}\lambda_n \check{I}_k^{(n)} - I_{\text{th}}^{(n)}\right)\left(\frac{1}{\gamma_{0,k}} - \frac{1}{\gamma_{g,k}}\right)^+ \Bigg\}, \end{aligned} \tag{4.23}$$

for $k = 1, \cdots, K$. This is achieved when the power allocation on subcarrier k is $P_{g^*,k} = P_{g^*,k}^*$ and $P_{g,k} = 0$ for all $g \neq g^*$, where g^* represents the group being allocated the

Table 4.1: Subcarrier and power allocation by dual optimization method

THE PROPOSED DUAL SCHEME	
1:	**Initialize** $(\lambda_1, \ldots, \lambda_N)$
2:	**Repeat**
3:	for $k = 1, \ldots, K$
4:	compute $P^*_{g,k}$ for all groups $g = 1, \ldots, G$ by (4.21)
5:	pick group g^* that gives minimum of $D_k(\boldsymbol{\lambda})$ as in (4.23)
6:	assign subcarrier k to group g^*
7:	set $P_{g^*,k} := P^*_{g^*,k}$ and $P_{g,k} := 0,\ \forall g \neq g^*$
8:	end
9:	update $(\lambda_1, \ldots, \lambda_N)$ according to (4.24)
10:	**Until** convergence of $\boldsymbol{\lambda}$

subcarrier k. From (4.21) and (4.22), it is worth mentioning that the allocation depends not only on the CSINR and the number of group users, but also on the availability of subchannel k as represented by ϕ_k.

Once (4.23) is solved for all subcarriers ($k = 1, \cdots, K$), the overall Lagrange dual function $D(\boldsymbol{\lambda})$ in (4.19) can be evaluated for the fixed $\boldsymbol{\lambda}$. Finally, it remains to find $\boldsymbol{\lambda}^* \succeq \mathbf{0}$ that minimizes $D(\boldsymbol{\lambda})$. This can be efficiently done by iteratively updating $\boldsymbol{\lambda}$ utilizing a subgradient-based or an ellipsoid method until convergence of $\boldsymbol{\lambda}$. At that point, the sum interference, induced by transmission from secondary BS to all of its multicast groups, to each primary user's frequency band also converges, and (positive) optimal powers have been distributed to the eligible multicast groups.

The update of $\boldsymbol{\lambda}$ may be performed as follows:

$$\lambda_n^{(t+1)} = \left(\lambda_n^{(t)} - \delta^{(t)}\left(I_{\text{th}}^{(n)} - \sum_{g=1}^{G}\sum_{k=1}^{K} P_{g,k}\check{I}_k^{(n)}\right)\right)^+, \tag{4.24}$$

for $n = 1, \cdots, N$, where $\delta^{(t)} > 0$ is a sequence of scalar step sizes. This subgradient update is guaranteed to converge to the optimal $\boldsymbol{\lambda}^*$ as long as $\delta^{(t)}$ is chosen to be sufficiently small. A popular choice is that $\delta^{(t)}$ is square summable but not absolute summable, for instance, $\delta^{(t)} = \frac{\beta}{t}$, for some constant $\beta > 0$.

The overall proposed dual scheme is summarized in Table 4.1. Note that the search for the best pairs of subcarriers and multicast groups (in Lines 3–8) has the form of a Greedy Algorithm [54]. It is also important to stress out that as the number of subcarriers goes to infinity, the gap between primal and dual optimal solutions vanishes quickly to zero. In practice, OFDM systems employ very large number of subcarriers (for example, as many as 4096), thus the optimal solution, obtained in the dual domain by the proposed scheme, virtually becomes a global optimum of the primal problem (4.10)–(4.13), with a negligible duality gap. Evidently, this demonstrates the practical optimality achieved by our proposed design.

4.4.2 Complexity Analysis

For a fixed $\boldsymbol{\lambda}$, solving problem (4.19) requires $\mathcal{O}(KG)$ executions. An ellipsoid method used to update $\boldsymbol{\lambda}$ converges in $\mathcal{O}(N^2)$ iterations. Therefore, the total complexity of the proposed dual scheme is $\mathcal{O}(KGN^2)$, which is only *linear* in the number of subcarriers. Since the number of subcarriers is often large in practical scenarios, a huge reduction in computational burden is expected from the proposed dual scheme as compared to, at least, $\mathcal{O}(KG^K)$ operations (for a simple case with only $N = 1$ primary user) by an optimal exhaustive search. Certainly, this is a highly desirable feature of adaptive algorithms designed for wireless communication systems where resolutions often need to be found within a very short time due to the dynamics of wireless channels.

4.4.3 Application of Dual Method to Other Types of Rate-loss Function

As seen before, the choice of a linear rate-loss function makes the problem more straightforward to analyze and a water-filling solution of power can be obtained as in (4.21). However, other rate-loss functions such as those illustrated in Fig. 4.2 could also be possible, and the proposed dual scheme is completely applicable in such cases.

If $L(P_{g,k})$ is concave and differentiable, a closed-form solution for the per-subcarrier problem can be obtained. For instance, let $L(P_{g,k}) = C \cdot \log(P_{g,k})$ then the objective of the maximization in the per-subcarrier problem is also a concave function of $P_{g,k}$. Solving a quadratic equation will yield an analytical solution for the optimal $P^*_{g,k}$. On the other hand, if the rate-loss function is not concave (for instance, exponential $L(P_{g,k}) = C \cdot (e^{P_{g,k}} - 1)$), an exhaustive search is normally required to determine the solution of the per-subcarrier problem. Nevertheless, since this problem is unconstrained, it is easier to handle than the original one. Indeed, any other form of the rate-loss function only leads to a difference in the resolution of per-subcarrier problems, whereas all other steps in the proposed dual scheme still remain unchanged.

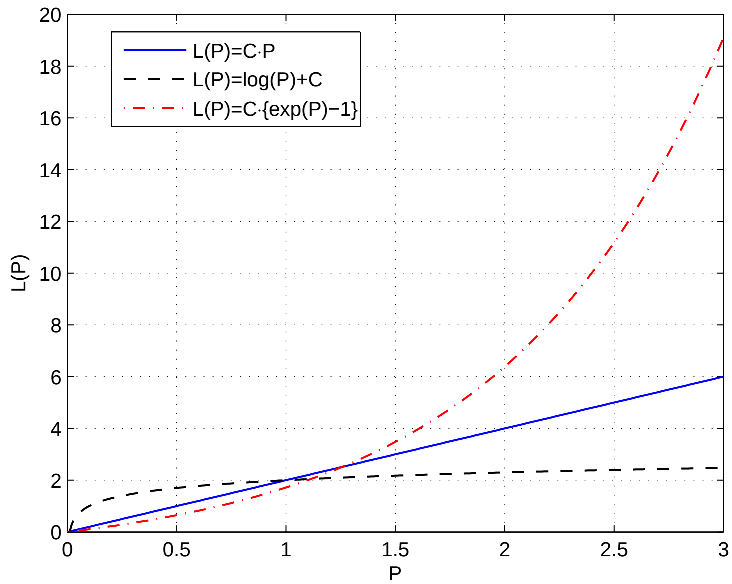


Figure 4.2: Some common rate-loss functions.

4.4.4 Effects of Adjacent Subcarrier Nulling Technique

The study in [43] proposes the method of dynamically deactivating subcarriers as a countermeasure to reduce the amount of interference from secondary to primary bands. Essentially, the suggested approach provides flexible guard bands between primary and secondary users by nulling subcarriers adjacent to the primary users' bands. However, this benefit comes at the cost of sacrificing bandwidth and, consequently, throughput of secondary users. It is therefore critical to balance the contradicting requirements of reducing interference and achieving the highest possible throughput of secondary users.

The computational complexity needed to find the optimal solutions in the proposed dual scheme can be further reduced by applying the adjacent subcarrier nulling method since the number of subcarriers K decreases. As well, more power can be distributed into the far-away subcarriers for a given interference threshold $I_{\text{th}}^{(n)}$. However, nulling adjacent subcarriers also reduces the available degree of freedom, which is the number of available subcarriers for possible transmission from the secondary BS to its own users, and in turn leads to a decrease in the throughput achieved by all cognitive multicast groups. For this reason, only the deactivation of a few adjacent subcarriers on each side of primary users' bands are normally recommended.

4.5 Numerical Examples

4.5.1 Simulation Setup and Assumptions

We consider a wireless system in which the primary BS communicates with N primary users. The primary-user frequency bands are predetermined in the available spectrum. All the primary user signals are assumed to be elliptically filtered white noise with equal amplitude $P_{PU} = 1$. To exploit temporary unused spectrum holes, a secondary BS is also allowed to simultaneously transmit to $G = 2$ multicast groups, each respectively consists of $|M_1| = 5, |M_2| = 3$ secondary users. The number of OFDM subcarriers used by the secondary BS is K, and the unused frequency bands are located on the sides of the already occupied bands. Moreover, the probability that primary users reoccupy unused subchannel k is assumed to be ϕ for all k.

In the computation of attainable sum rates, 100 sets of independent channel coefficients $\{g_{SP}^{(n)}\}$, $\{g_{m,k}^{PS}\}$ and $\{h_{m,k}^{SS}\}$ are randomly generated according to the Rayleigh distribution. The average channel gains, the noise power of each subcarrier, the OFDM symbol duration, and the individual subcarrier bandwidth are all normalized to 1. It is further assumed that perfect coding is employed, which means that $\Gamma = 1$. Since all the spectral distances $d_k^{(n)}$ can be determined, it is possible to compute the interferences $\check{I}_k^{(n)}$, $J_{m,k}^{(n)}$, and the CSINR of individual secondary user $\alpha_{m,k}$. Then, the equivalent CSINR of group M_g on subcarrier k is simply $\gamma_{g,k} = \min_{m \in M_g} \alpha_{m,k}$. In addition, both groups are assumed to have equal priority $w_1 = w_2 = 0.5$. For brevity, the numerical examples are only performed for the case of linear rate-loss function $L(P_{g,k}) = C \cdot P_{g,k}$.

4.5.2 Simulation Results

In order to confirm the practical optimality achieved by the proposed dual scheme, we first study a simple case with $N = 1$ primary user, $K = 8$ OFDM subcarriers and $L(P_{g,k}) = 0$. Fig. 4.3 plots the actual achieved throughput by both the dual optimization and the direct exhaustive search. As can be seen, the two rate curves are almost indistinguishable. The very small gap between the primal optimum obtained by exhaustive search and the dual optimum is also shown in Table 4.2. Note that the duality gap is already insignificant even with only 8 active subcarriers. Furthermore, we examine the case of multiple primary users $N = 2$, a larger number of OFDM subcarriers $K = 36$ and a positive rate loss. Since it is too complex to carry out an exhaustive search to find the optimal so-

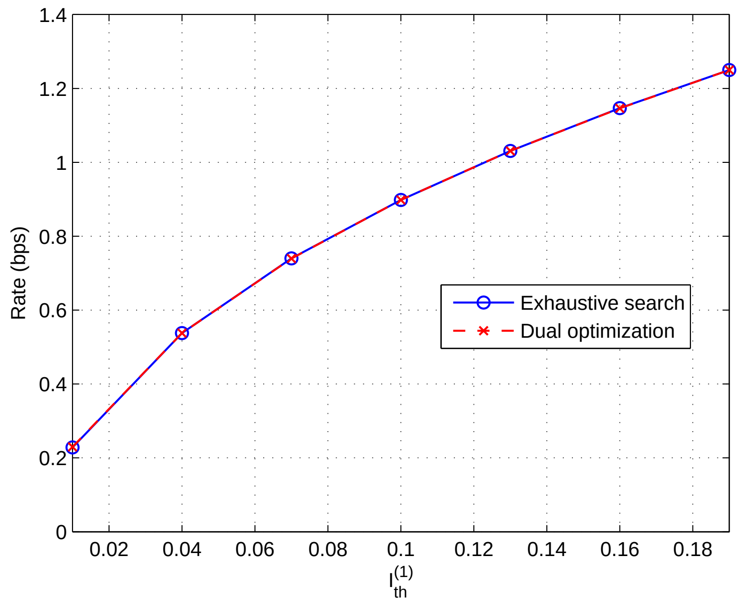


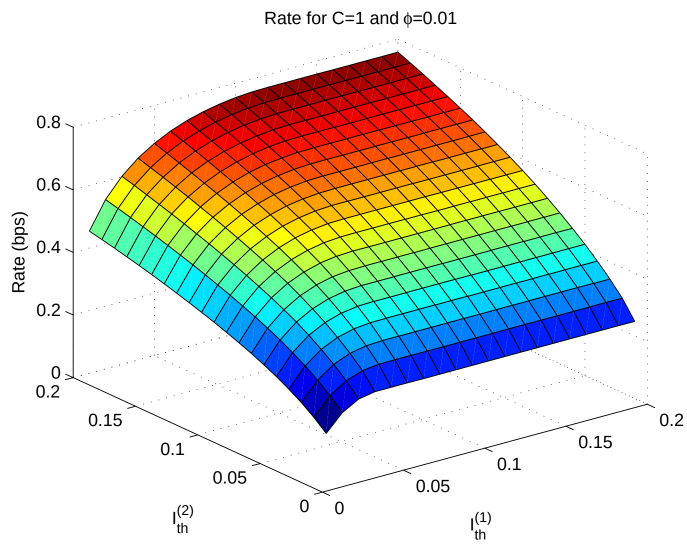
Figure 4.3: Comparison of dual optimization and exhaustive search methods for $N = 1, K = 8, L(P_{g,k}) = 0$.

Table 4.2: Average duality gap for $N = 1, K = 8, L(P_{g,k}) = 0$

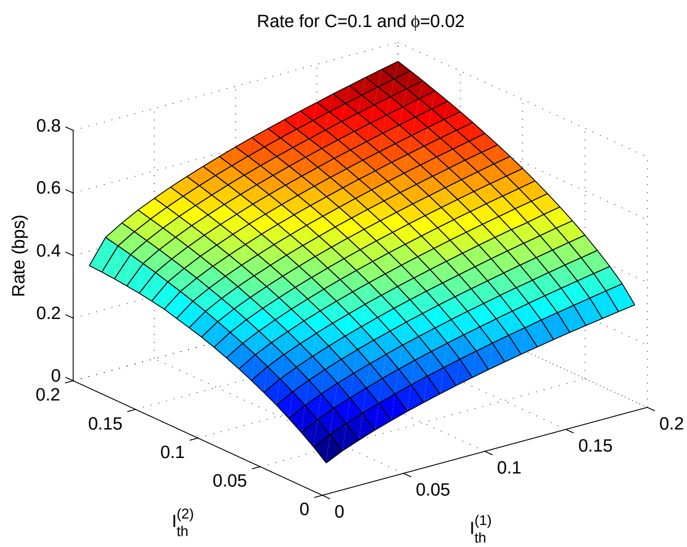
Interf. limit I_{th}	0.0100	0.0400	0.0700	0.1000	0.1300	0.1600	0.1900
Abs. gap $\lvert D^* - f^* \rvert$ $(\times 10^{-3})$	1.1831	0.0250	0.2934	0.0009	0.0003	0.3043	0.0004
Rel. gap $\lvert \frac{D^* - f^*}{f^*} \rvert$ $(\times 10^{-3})$	5.1783	0.0465	0.3966	0.0010	0.0003	0.2654	0.0004

lutions of (4.10)–(4.13) in this case, we instead verify that Condition 2 stated in Section 4.3 is met. For a randomly chosen instance of channels and with different values of C and ϕ, Fig. 4.4 demonstrates that the total expected rate-sum at optimality is indeed a concave function of $\bar{\mathbf{I}}_{\text{th}} = [I_{\text{th}}^{(1)}, I_{\text{th}}^{(2)}]$. It is expected that the concavity of optimal throughput with respect to $\bar{\mathbf{I}}_{\text{th}}$ becomes even more visible as the number of subcarriers K is much larger than 36. From Condition 2, this observation implies a negligible duality gap and, again, indicates that the solution obtained by the proposed dual method is virtually the primal global optimum. Therefore, in what follows, all the simulation results are presented for this more practical scenario with $N = 2$ primary users and $K = 36$ OFDM subcarriers.

Assuming $C = 0.1$ and $\phi = 0.02$, Figs. 4.5a and 4.5b illustrate the actual convergence process of the sum interference introduced to each primary user and that of the achieved data rate for a snapshot of channel at interference threshold $I_{\text{th}}^{(1)} = I_{\text{th}}^{(2)} = I_{\text{th}} = 0.1$, respectively. As can be seen, the dual optimization scheme converges very fast (after only

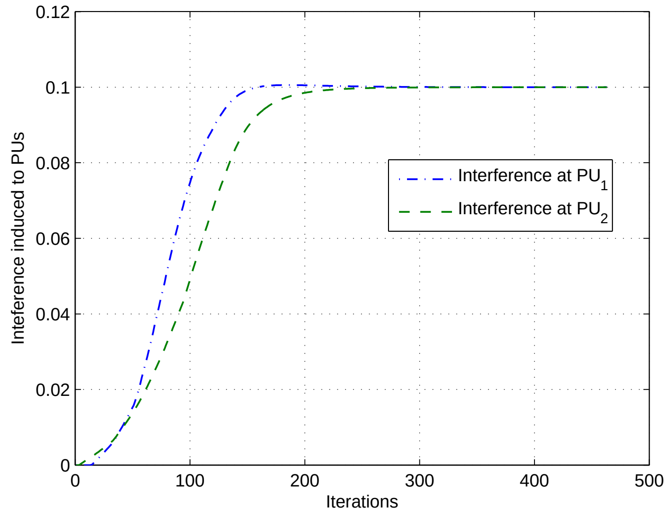


(a)

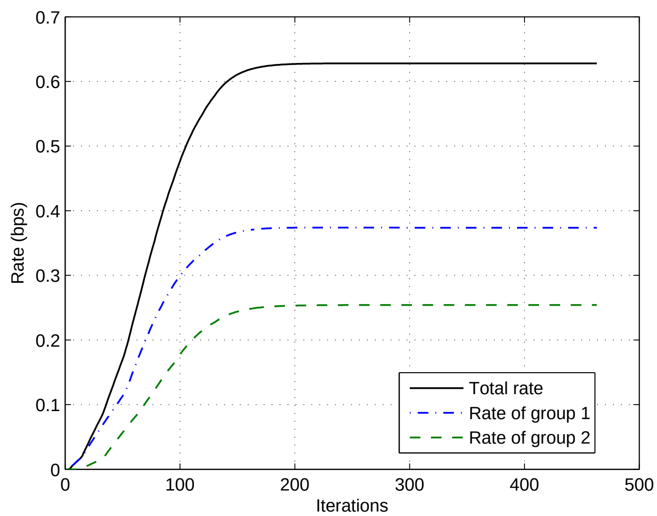


(b)

Figure 4.4: Concavity of the optimal throughput for $N = 2, K = 36, L(P_{g,k}) > 0$.



(a) Sum interference introduced to primary users (PU).



(b) Data rates.

Figure 4.5: Convergence process of the proposed dual scheme.

few hundred of iterations) and the constraints on interference limits are met with equality. Together with the above-confirmed practical optimality, this quick-converging feature certainly makes the proposed design highly attractive for practical wireless applications.

To clearly evaluate the effect of adjacent subcarrier nulling technique, Fig. 4.6 plots the total system throughput (assuming no rate loss) as well as individual group rates obtained by the proposed scheme, with and without nulling adjacent subcarriers on each side

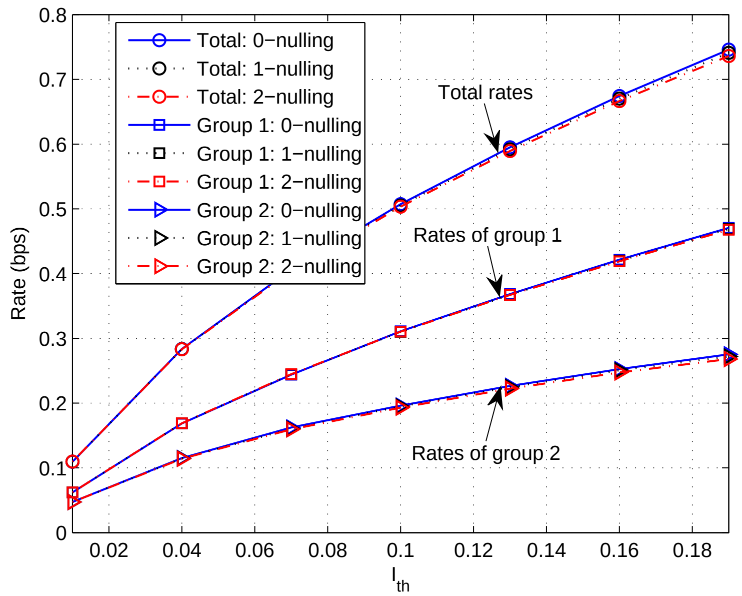


Figure 4.6: Effect of adjacent subcarrier nulling (with zero rate-loss assumed).

of the primary users' bands. Note that only deactivating 1 and 2 adjacent subcarriers are considered in the simulations. It is clear that the multicast group with more user members (that is, group M_1) achieves a higher rate as it is allocated more subchannels in the unused spectrum. Moreover, although the attainable rates in the $1-$nulling and $2-$nulling cases actually decrease as the adjacent subcarriers are assigned zero power even when their respective channel conditions are very good, the degradation is minor. This is possibly due to the compensation provided by the so-called multiuser diversity, and to the fact that the equivalent CSINR of each multicast group is determined by user with the worst channel condition.

Finally, the consequences of varying different rate-loss parameters on the system throughput are assessed. Figs. 4.7, 4.8 and 4.9 exhibit the behaviors of achieved data rate for several selected values of C, ϕ and I_{th}. It is apparent that increasing rate-loss (that is, by increasing C and/or ϕ) results in a decrease in the attained throughput. In particular, if the OFDM subchannels are too busy or secondary access of the available resources is too costly, the achievable data rates of cognitive radio network may even approach zero. Clearly, the performance of resource management algorithms designed for cognitive radio network depends very much on the activities of primary users on the spectrum available for secondary access, which are quantified by a loss in the achieved sum rates.

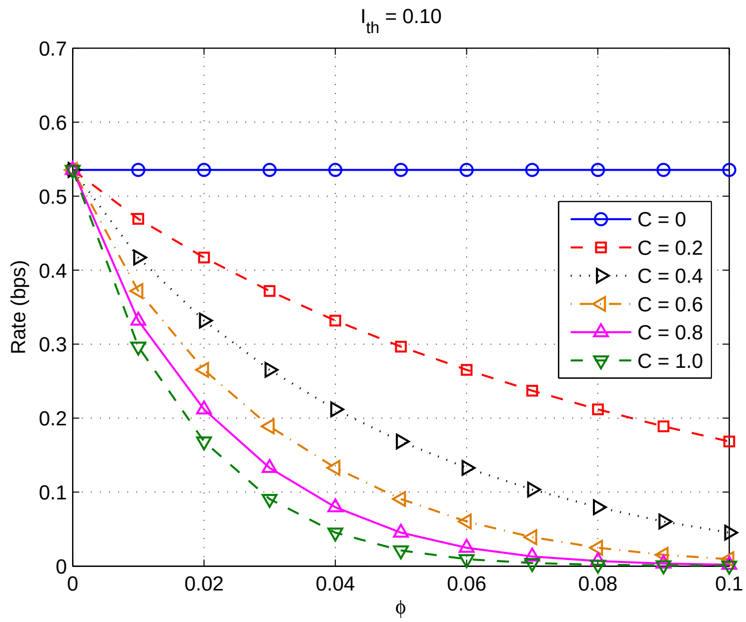


Figure 4.7: Attained system throughput for a fixed $I_{\mathrm{th}} = 0.10$.

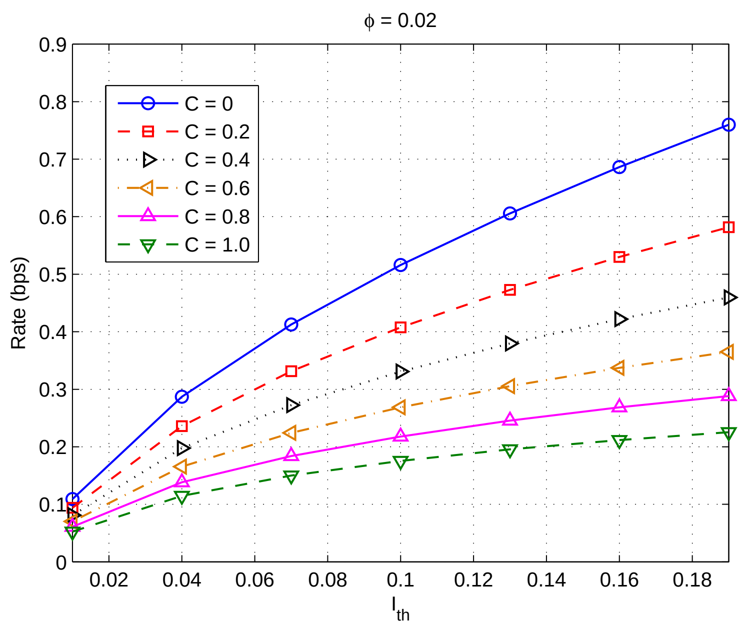


Figure 4.8: Attained system throughput for a fixed $\phi = 0.02$.

4.6 Concluding Remarks

In this chapter, we have proposed a dual scheme for the allocation of subcarriers and power to maximize the expected throughput of a secondary network employing OFDMA,

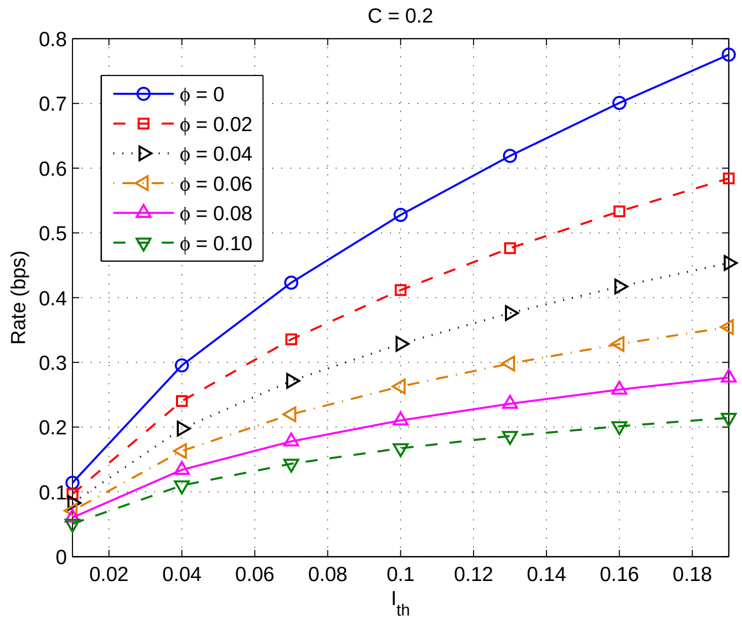


Figure 4.9: Attained system throughput for a fixed $C = 0.2$.

subject to tolerable interference at the primary users in cognitive radio settings. The solution also takes into account the subchannel availability or the primary users' activities by incorporating a rate-loss function in the design. Global optimality can be achieved by the devised scheme for a realistically large number of OFDM subcarriers. Further, the proposed dual optimization method can handle both unicast and multicast transmissions, and its complexity is only linear in the number of subcarriers. The effects of nulling adjacent subcarriers on the proposed design have also been investigated. Numerical results confirms the potential benefits of our proposed approach.

Chapter 5

Conclusion

5.1 Summary

Much of the existing work, relating to the problems of resource allocation in OFDMA-based wireless systems, focuses on finding efficient subcarrier assignment and power control schemes. In systems employing multicast transmission, the present solutions in literature are not always applied. Further, when the number of OFDM subcarriers utilized is practically large, the existing approaches are, as best, inaccessible since they involve prohibitively high computational effort. The aim of this thesis is therefore to provide effective and affordable mechanisms to share out the available resources in multicast wireless systems deploying OFDMA technology. Through the application of various mathematical optimization techniques such as genetic algorithm and Lagrangian dual optimization, sub-optimal and optimal solutions with high performance and yet of acceptable complexity have been presented for different system models. Specifically, we have studied the resource management problems in two contexts, formally described them as mathematical optimization programs, and subsequently provided the solution methods. The novelties of the proposed designs have been confirmed and their performance have been verified by simulation with the illustration of numerical examples.

In the first design, we aim at maximizing the total sum rates of a conventional OFDMA-based multicast network while ensuring a flexible and effective control of the spectrum shares among individual multicast groups. Three novel schemes, which are shown to offer high system throughput with significantly reduced computational complexity, have been devised. The first two solutions are based on separate optimization of subcarriers and power, while the last one is obtained with a modified genetic algorithm. In the separate optimization schemes, the subcarrier allocation ensures minimum numbers of subcarriers to be assigned to individual groups according to their respective channel gains and group sizes, while power is distributed in a water-filling fashion. With the scheme based on the modified genetic algorithm, the jointly optimal subcarrier-power allocation is iteratively evolved through a global search while satisfying the imposed bandwidth constraints among different multicast groups. The proposed approaches, whose complexity is totally affordable, are particularly relevant for cost-effective and delay-sensitive wireless applications which require the resource allocation to be completed within a short time due to the dynamics of wireless channels.

In the second design, we consider the aspect of primary user activity or subchannel availability in optimally managing the resources of an OFDMA-based cognitive radio multicast network. For this purpose, a risk-return model is presented and a general rate loss function, which gives the rate loss whenever primary users reoccupy the temporarily available subchannels, is defined. Taking the maximization of the expected sum rate of secondary multicast groups as the design objective, a practically optimal subcarrier and power allocation scheme is proposed under constraints on the tolerable interference thresholds at individual primary user's frequency bands and also on the dynamics of primary users in the accessible radio spectrum. Specifically, the original challenging non-convex optimization problem is solved effectively via a dual optimization framework, and as the number of subcarriers becomes realistically large, the duality gap between primal and dual optimal solutions turns out to be negligible. Even more attractively, complexity of the derived scheme grows only linearly in the number of subcarriers, representing a huge reduction in computational burden. Accordingly, it can be claimed that the proposed solution is capable of achieving the global optimality with a fast computational time in practical systems in which a large number of OFDM subcarriers is normally deployed.

5.2 Future Work

Throughout this work, we have assumed that various parameters such as the number of users in individual multicast groups, the channel gains and the interference thresholds are fixed. In practical systems, however, the number of active users may be highly dynamic due to users leaving and joining the systems. Further, the link quality and the interference levels may vary quickly, especially in environments involved a high degree of mobility. Under those circumstances, the solutions, once computed for the fixed scenarios, may no longer be valid. A more appropriate and practical approach could involve the application of stochastic optimization in analyzing such cases [55, 56].

Moreover, it is note-worthy that the power control and subcarrier assignment schemes presented in this work are centralized, where the base station dictates and assigns subcarrier as well as transmitted power levels to its users based on, for instance, their channel qualities. The resource allocation can also be accomplished in a distributed manner, in which users compete for the subchannels and update their powers autonomously, independent of the base station, based on the perceived service quality. In such distributed settings, the wireless users can be realized as selfish agents or players who try to maximize their utilities (for example, the corresponding throughput). Hence, game theory could be a more powerful tool to study such scenarios [57, 58].

Bibliography

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

[2] L. J. Cimini and N. R. Sollenberger, "OFDM with diversity and coding for advanced cellular internet services," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 1, Nov. 1997, pp. 305–309.

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

[4] W. Rhee and J. Cioffi, "Increasing in capacity of multiuser OFDM systems using dynamic subchannel allocation," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, vol. 2, May 2000, pp. 1085–1089.

[5] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.

[6] 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.

[7] Y. Otani, S. Ohno, K. Teo, and T. Hinamoto, "Subcarrier allocation for multi-user OFDM system," in *Asia-Pacific Conf. on Commun.*, Oct. 2005, pp. 1073–1077.

[8] K. Kim, Y. Han, and S.-L. Kim, "Joint subcarrier and power allocation in uplink OFDMA systems," *IEEE Commun. Lett.*, vol. 9, no. 6, pp. 526–528, Jun. 2005.

[9] G. Li and H. Liu, "Resource allocation for OFDMA relay networks with fairness constraints," *IEEE J. Select. Areas Commun.*, vol. 24, no. 11, pp. 2061–2069, Nov. 2006.

[10] Z. Mingyu, L. Lihua, and Z. Minghua, "A fair and effective adaptive resource allocation method for multi-user OFDM systems," in *IEEE Int. Symposium on Commun. and Infor. Technologies*, Oct. 2007, pp. 95–100.

[11] L. Gao and S. Cui, "Efficient subcarrier, power, and rate allocation with fairness consideration for OFDMA uplink," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1507–1511, May 2008.

[12] J. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, Michigan: University of Michigan Press, 1975.

[13] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, Massachusetts: Addison-Wesley, 1989.

[14] M. Mitchell, *An Introduction to Genetic Algorithm*. Cambridge, Massachusetts: MIT Press, 1996.

[15] H. Tuy, *Convex Analysis and Global Optimization*. The Netherlands: Kluwer Academic Publishers, 1998.

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

[17] R. Horst, P. Pardalos, and N. Thoai, *Introduction to Global Optimization*. The Netherlands: Kluwer Academic Publishers, 1995.

[18] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Boston: Athena Scientific, 1999.

[19] D. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[20] D. T. Ngo, C. Tellambura, and H. H. Nguyen, "Efficient resource allocation for OFDMA multicast systems with fairness consideration," in *Proc. IEEE Radio and Wireless Symposium (RWS)*, San Diego, USA, Jan. 2009, pp. 392–395.

[21] ——, "Efficient resource allocation for OFDMA multicast wireless systems with spectrum-sharing control," *IEEE Trans. Veh. Technol.*, 2009, to appear.

[22] C. Ng and C. W. Sung, "Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1667–1675, May 2008.

[23] C. Bae and D.-H. Cho, "Fairness-aware adaptive resource allocation scheme in multihop OFDMA systems," *IEEE Commun. Lett.*, vol. 11, no. 2, pp. 134–136, Feb. 2007.

[24] R. Agarwal, V. Majjigi, Z. Han, R. Vannithamby, and J. Cioffi, "Low complexity resource allocation with opportunistic feedback over downlink OFDMA networks," *IEEE J. Select. Areas Commun.*, vol. 26, no. 8, pp. 1462–1472, Oct. 2008.

[25] U. Varshney, "Multicast over wireless networks," *Commun. of the ACM*, vol. 45, pp. 31–37, Dec. 2002.

[26] H. Won, H. Cai, D. Eun, K. Guo, R. Netravali, I. Rhee, and K. Sabnani, "Multicast scheduling in cellular data networks," in *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, 2007, pp. 1172–1180.

[27] J. Liu, W. Chen, Z. Cao, and K. B. Letaief, "Dynamic power and sub-carrier allocation for OFDMA-based wireless multicast systems," in *Proc. IEEE Int. Conf. Communications (ICC)*, May 2008, pp. 2607–2611.

[28] J. Kim, T. Kwon, and D.-H. Cho, "Resource allocation scheme for minimizing power consumption in OFDM multicast systems," *IEEE Commun. Lett.*, vol. 11, no. 6, pp. 486–488, Jun. 2007.

[29] C. Suh and J. Mo, "Resource allocation for multicast services in multicarrier wireless communications," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 27–31, Jan. 2008.

[30] M. R. Sherif, I. W. Habib, M. Nagshineh, and P. Kermani, "Adaptive allocation of resources and call admission control for wireless ATM using genetic algorithms," *IEEE J. Select. Areas Commun.*, vol. 18, no. 2, pp. 268–282, Feb. 2000.

[31] M. Alias, S. Chen, and L. Hanzo, "Multiple-antenna-aided OFDM employing genetic-algorithm-assisted minimum bit error rate multiuser detection," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1713–1721, Sep. 2005.

[32] FCC Spectrum Policy Task Force, "Report of the spectrum efficiency working group," Federal Communications Commission, Tech. Rep. ET Docket No. 02-135, Nov. 2002.

[33] J. Mitola and G. Q. Maguire, “Cognitive radio: Making software radios more personal,” *IEEE Personal Commun. Mag.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

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

[35] Q. Zhao and B. M. Sadler, “A survey of dynamic spectrum access,” *IEEE Signal Processing Mag.*, pp. 79–89, May 2007.

[36] T. Clancy, “Formalizing the interference temperature model,” *Wirel. Commun. Mob. Comput. (UK)*, vol. 7, no. 9, pp. 1077–1086, 2007.

[37] Y. Xing, C. N. Mathur, M. Haleem, R. Chandramouli, and K. Subbalakshmi, “Dynamic spectrum access with QoS and interference temperature constraints,” *IEEE Trans. Mobile Computing*, vol. 6, no. 4, pp. 423–433, Apr. 2007.

[38] D. T. Ngo, C. Tellambura, and H. H. Nguyen, “Resource allocation for OFDM-based cognitive radio multicast networks,” in *Proc. IEEE Wireless Commun. and Networking Conf.*, Budapest, Hungary, Apr. 2009, pp. 1–6.

[39] ——, “Resource allocation for OFDMA-based cognitive radio multicast networks with primary user activity consideration,” *IEEE Trans. Veh. Technol.*, May 2009, submitted.

[40] T. Weiss and F. Jondral, “Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency,” *IEEE Commun. Mag.*, vol. 42, no. 3, pp. S8–14, Mar. 2004.

[41] U. Berthold, F. Jondral, S. Brandes, and M. Schnell, “OFDM-based overlay systems: A promising approach for enhancing spectral efficiency [Topics in radio communications],” *IEEE Commun. Mag.*, vol. 45, no. 12, pp. 52–58, Dec. 2007.

[42] S. Chuprun, J. Kleider, and C. Bergstrom, “Emerging software defined radio architectures supporting wireless high data rate OFDM,” in *Proc. IEEE Radio and Wireless Conf. (RAWCON)*, 1999, pp. 117–120.

[43] T. Weiss, J. Hillenbrand, A. Krohn, and F. Jondral, “Mutual interference in OFDM-based spectrum pooling systems,” in *Proc. IEEE Vehicular Technology Conf. (VTC)*, vol. 4, May 2004, pp. 1873–1877.

[44] B. Farhang-Boroujeny and R. Kempter, “Multicarrier communication techniques for spectrum sensing and communication in cognitive radios,” *IEEE Commun. Mag.*, pp. 80–85, Apr. 2008.

[45] G. Bansal, M. Hossain, and V. Bhargava, “Optimal and suboptimal power allocation schemes for OFDM-based cognitive radio systems,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4710–4718, Nov. 2008.

[46] T. Qin and C. Leung, “Fair adaptive resource allocation for multiuser OFDM cognitive radio systems,” in *Proc. Second Int. Conf. Communications and Networking in China (CHINACOM)*, Aug. 2007, pp. 115–119.

[47] P. Wang, M. Zhao, L. Xiao, S. Zhou, and J. Wang, “Power allocation in OFDM-based cognitive radio systems,” in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2007, pp. 4061–4065.

[48] A. T. Hoang and Y.-C. Liang, “Downlink channel assignment and power control for cognitive networks,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3106–3117, Aug. 2008.

[49] P. Cheng, Z. Zhang, H.-H. Chen, and P. Qiu, "Optimal distributed joint frequency, rate and power allocation in cognitive OFDMA systems," *IET Communications*, vol. 2, no. 6, pp. 815–826, July 2008.

[50] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

[51] Z. Hasan, E. Hossain, C. Despins, and V. K. Bhargava, "Power allocation for cognitive radios based on primary user activity in an OFDM system," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2008, pp. 1–6.

[52] A. J. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.

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

[54] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Boston: MIT Press, Sep. 2001.

[55] N. Ansari, A. Arulambalam, and S. Balasekar, "Traffic management of a satellite communication network using stochastic optimization," *IEEE Trans. Neural Networks*, vol. 7, no. 3, pp. 732–744, May 1996.

[56] M. van der Schaar, D. Turaga, and R. Sood, "Stochastic optimization for content sharing in P2P systems," *IEEE Trans. Multimedia*, vol. 10, no. 1, pp. 132–144, Jan. 2008.

[57] D. Niyato and E. Hossain, "A noncooperative game-theoretic framework for radio resource management in 4G heterogeneous wireless access networks," *IEEE Trans. Mobile Computing*, vol. 7, no. 3, pp. 332–345, Mar. 2008.

[58] ——, "A game theoretic analysis of service competition and pricing in heterogeneous wireless access networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5150–5155, Dec. 2008.

[59] C. Suh and J. Mo, "Resource allocation for multicast services in multicarrier wireless communications," in *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, April 2006, pp. 1–12.

[60] L. M. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, "Multiuser transmit optimization for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 922–930, Jun. 2004.

Appendix A

Ideal Structure of an OFDM System

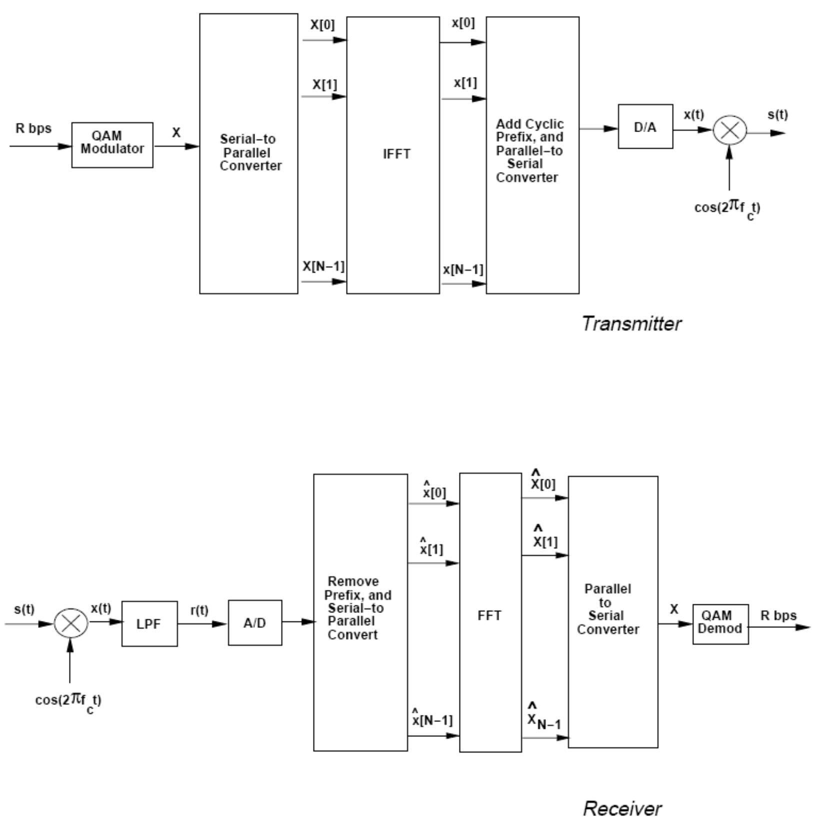


Figure A.1: Structure of an OFDM system. Taken from [1, p.387].

Block diagram of an ideal OFDM transceiver is depicted in Fig. A.1. On the transmitter side, the input bit stream is first modulated by a Quadrature Amplitude Modulator (QAM) to create a symbol stream $\mathbf{X}$ of N complex symbols. A serial-to-parallel converter is used to split this stream of symbols into N parallel QAM symbols $\mathbf{X}[0], \mathbf{X}[1], \ldots, \mathbf{X}[N-1]$ to be transmitted over each of the N sub-carriers. These frequency components are converted into time samples by an Inverse-FFT (IFFT), yielding an OFDM symbol which consists of N elements:

$$x[n] = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \mathbf{X}[i] e^{j2\pi ni/N}, \quad 0 \leq n \leq N-1. \tag{A.1}$$

More generally, the length-N vector $\mathbf{X}$ can be written as

$$\begin{aligned} \mathbf{X} \;\; = \;\; & [\mathbf{X}[0], \mathbf{X}[1], \ldots, \mathbf{X}[(N_{\text{act}}-1)/2], 0, 0, \ldots \\ & 0, 0, \mathbf{X}[(N_{\text{act}}-1)/2+1], \ldots, \mathbf{X}[N_{\text{act}}-1]], \end{aligned} \tag{A.2}$$

which represents one OFDM symbol, where the number of active sub-carriers conveying information is N_{act} (that is, there are N_{act} QAM symbols), while the other carriers are set to zero to avoid spectral overlapping. This can also be thought of as a form of up-sampling, as the rate at the output of the IFFT will be increased. The cyclic prefix is then added as a guard interval to the OFDM symbol. The role of cyclic prefix is to eliminate inter-symbol interference between data blocks as it ensures the channel output is a circular convolution. The parallel-to-serial converter is employed to reorder the time samples before passing them through a digital-to-analog (D/A) converter to obtain the baseband OFDM signal $x(t)$. This signal is then upconverted to frequency f_c and sent to the channel.

The transmitted signal is filtered by channel impulse response, corrupted by noise and the signal $s(t)$ is received at the receiver front end. Here, the received signal is first down-converted to baseband before being passed through a low-pass filter to remove all the high-frequency components. The resulting continuous-time signal $r(t)$ is then converted to a digital signal by an analog-to-digital (A/D) converter. The cyclic prefix of this digital signal is removed, and the remaining serial time samples are then converted into N parallel symbols $\hat{x}[0], \hat{x}[1], \ldots, \hat{x}[N-1]$. An FFT, whose outputs $\hat{\mathbf{X}}[0], \hat{\mathbf{X}}[1], \ldots, \hat{\mathbf{X}}[N-1]$ will be parallel-to-serial converted, is used to obtain scaled versions of the original symbols. Finally, the symbol stream $\mathbf{X}$ is demodulated by a QAM demodulator to recover the original data. In the case when up-sampling is considered, to avoid spectral overlap, the output of the N-point FFT, will convey N_{act} sub-carriers of data, while the other carriers will be equal to zero plus the additive channel noise. The samples containing additive channel noise can be removed before parallel-to-serial conversion. This operation is a form of down-sampling, as it is the inverse of the up-sampling process (by zero padding) at the transmitter.

Appendix B

Capacity of Multicarrier Multicast Systems

Consider a conventional multicast transmission from BS to a group of K active users over M OFDM subcarriers[1]. Upon defining $X_k^{(m)}$ $(m = 1, \cdots, M;\ k = 1, \cdots, K)$ the random variable representing the signal-to-noise ratio (SNR) of user k on subcarrier m and denoting $X_{(1)}^{(m)} = \min\{X_1^{(m)}, X_2^{(m)}, \cdots, X_K^{(m)}\}$ the group equivalent SNR on that same subcarrier, the multicast transmission rate at which BS transmits to all the K users on subcarrier m can be written as

$$R_{\mathrm{MC}}^{(m)} = \log_2\left(1 + X_{(1)}^{(m)}\right). \tag{B.1}$$

The system multicast capacity over all M subcarriers is then

$$C_{\mathrm{MC}} = \sum_{m=1}^{M} K \cdot R_{\mathrm{MC}}^{(m)}. \tag{B.2}$$

The ergodic capacity for multicast service now becomes

$$\mathcal{E}\{C_{\mathrm{MC}}\} = \sum_{m=1}^{M} \mathcal{E}\Big\{K \cdot \log_2\left(1 + X_{(1)}^{(m)}\right)\Big\}, \tag{B.3}$$

For Rayleigh fading channels, we have the following result.

Proposition B.0.1 *Assume that $X_k^{(m)}, k = 1, \cdots, K$ are i.i.d exponential random variables with parameter $\beta^{(m)}$, the ergodic capacity defined in (B.3) only depends on $\beta^{(m)}$ in the limit as $K \to \infty$. If we further assume that $\beta^{(1)} = \beta^{(2)} = \cdots = \beta^{(M)}$, the ergodic capacity increases linearly with M in the limit as $K \to \infty$.*

[1] A version of this appendix has been presented in [21].

Proof B.0.1 *The proof is based on [59]. From the pdf of $X_k^{(m)}$ which is*

$$f(x) = \frac{1}{\beta^{(m)}} e^{-\frac{x}{\beta^{(m)}}}, \tag{B.4}$$

the pdf of $X_{(1)}^{(m)}$ can be derived via order statistics as

$$f_{(1)}(x) = \frac{K}{\beta^{(m)}} e^{-\frac{Kx}{\beta^{(m)}}}. \tag{B.5}$$

The ergodic capacity multi-carrier multicast system can be expressed as

$$\begin{aligned}
\mathcal{E}\{C_{\mathrm{MC}}\} &= \mathcal{E}\left\{ \sum_{m}^{M} C_{\mathrm{MC}}^{(m)} \right\} \\
&= \sum_{m=1}^{M} \mathcal{E}\left\{ K \cdot \log_2 \left(1 + X_{(1)}^{(m)} \right) \right\} \\
&= \sum_{m=1}^{M} K \cdot \int_0^{\infty} \log_2 (1+x) \, f_{(1)}(x) dx \\
&= \sum_{m=1}^{M} \frac{K^2}{\beta^{(m)}} \cdot \int_0^{\infty} \log_2 (1+x) \, e^{-\frac{Kx}{\beta^{(m)}}} dx
\end{aligned} \tag{B.6}$$

Applying the result in [59] and by the change of variable $u = t - \frac{K}{\beta^{(m)}}$, (B.6) simplifies to

$$\mathcal{E}\{C_{\mathrm{MC}}\} = \log_2 e \cdot \sum_{m=1}^{M} \beta^{(m)} \int_0^{\infty} \frac{e^{-u}}{1 + \frac{\beta^{(m)} u}{K}} du. \tag{B.7}$$

As $K \to \infty$, (B.7) becomes

$$\lim_{K \to \infty} \mathcal{E}\{C_{\mathrm{MC}}\} = \log_2 e \cdot \sum_{m}^{M} \beta^{(m)}. \tag{B.8}$$

If we further assume that $\beta^{(m)} = \bar{\beta}$, $\forall m$, then (B.8) evaluates to

$$\lim_{K \to \infty} \mathcal{E}\{C_{\mathrm{MC}}\} = \log_2 e \cdot M \cdot \bar{\beta}. \tag{B.9}$$

This completes the proof.

For Ricean fading channels, analyzing the ergodic multicast capacity is challenging since the Ricean distribution involves the modified Bessel function. Instead, we claim that a similar result applies for the case of Ricean fading and verify it with simulation results in the following.

Assuming that the average SNR is normalized to 1, Figs. B.1 and B.2 demonstrate the dependence of multicast ergodic capacity on the group sizes and on the number of

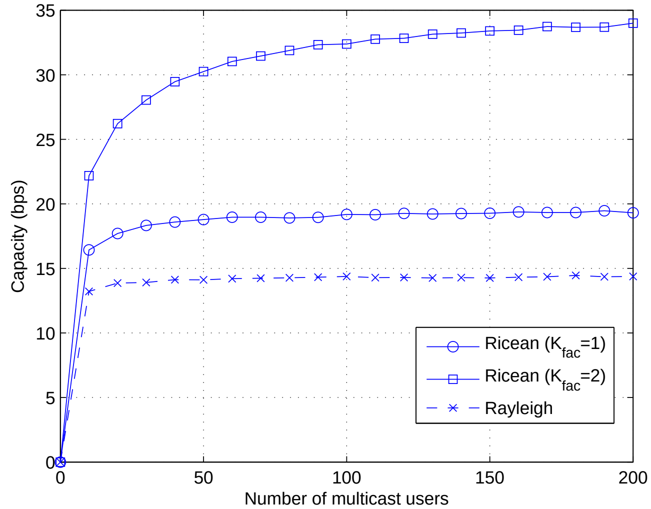


Figure B.1: Ergodic multicast capacity as a functions of group size, assuming all the users have an identical K_{factor}.

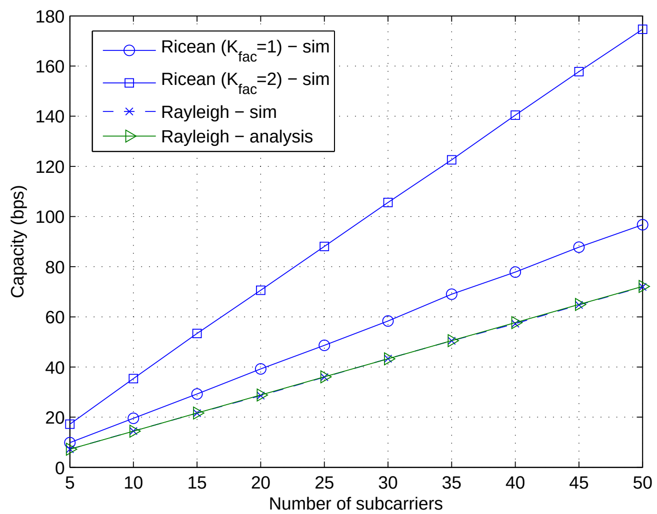


Figure B.2: Ergodic multicast capacity as a function of the number of subcarriers, assuming all the users have an identical K_{factor}.

subcarriers, respectively. As the group size K increases, the capacity becomes saturated and independent of K for multicast systems employing $M = 10$ subcarriers. However, the capacity of a multicast system with $K = 100$ users employing conventional transmission increases linearly with the number of subcarriers.

Appendix C

Multi-level Water-filling for Non-convex Multicarrier Resource Allocation

In this appendix, we shall establish that although it is possible to directly derive an optimal solution for the design problem (4.10)–(4.13) in the primal domain, the complexity of such approach is actually exponential in the number of OFDM subcarriers. For simplicity, let us consider the case of $N = 1$ primary user and zero rate-loss $L(P_{g,k}) = 0$. The optimization problem can now be reduced to

$$\max_{\{P_{g,k}\}} \quad \sum_{g=1}^{G}\sum_{k=1}^{K} \frac{w_g|M_g|}{K}\log_2(1+\gamma_{g,k}P_{g,k}) \tag{C.1}$$

$$\text{s.t.} \quad \sum_{g=1}^{G}\sum_{k=1}^{K} P_{g,k}\check{I}_k \le I_{\text{th}}^{(1)}, \tag{C.2}$$

$$P_{g,k} \ge 0;\ g = 1,\cdots,G,\ k = 1,\cdots,K, \tag{C.3}$$

$$P_{g,k}P_{g',k} = 0;\ \forall g' \neq g. \tag{C.4}$$

Let S_g denote the set of subcarriers allocated to tone group g. For any fixed channel assignment S_g, problem (C.1)–(C.4) is convex and thus its optimal solution can be determined from the KKT conditions as follows.

$$P_{g,k}^{*} = \left(\frac{w_g|M_g|}{K\check{I}_k\log 2}\cdot\mu - \frac{1}{\gamma_{g,k}}\right)^{+}, \tag{C.5}$$

$$\mu = \frac{I_{\text{th}}^{(1)} + \sum_{g=1}^{G}\sum_{k\in\{S_g:P_{g,k}>0\}}\frac{\check{I}_k}{\gamma_{g,k}}}{\sum_{g=1}^{G}\frac{w_g|M_g||S_g|}{K\log 2}}. \tag{C.6}$$

Clearly, this is a form of multi-level waterfilling wherein the number of used OFDM subchannels needs to be optimized until all powers are positive [60]. As finding optimal subchannel assignment among G groups of secondary users requires G^K searches, the overall optimization requires $\mathcal{O}(KG^K)$ operations which is exponentially complex.

Also notice that the analytical solution in the above derivation is possible due to the many simplified assumptions. In the presence of a positive rate loss function and multiple primary users, the optimal search in the primal domain would be far more complicated. This emphasizes the need of having more suitable approaches to efficiently solve the design problem (4.10)–(4.13).