TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABREVIATIONS

OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
QoS	Quality of Service
SIR	Signal to Noise Ratio
SINR	Signal to Interference plus Noise Ratio
UE	User Equipment
BS	Base Station
RS	Relay Station
GP	Geometric Programming
KKT	Karush-Kuhn-Tucker
CSI	Channel State Information
IPM	Interior Point Method
LoS	Line of Sight
NLoS	none Line of Sight
LTE	Long-Term Evolution
WiMaX	Worldwide Interoperability for Microwave Access
5G	Fifth Generation wireless systems
CBRA	Conservative Binary Resource Allocation
OBRA	Optimal Binary Resource Allocation
C-RAN	Cloud Radio Access Networks
INR	Interference over Noise Ratio
FFR	Fractional Frequency Reuse
IM	Interference Management

DF	Decode-and-forward relaying
AF	Amplify-and-forward relaying
CF	Compress-and-forward relaying
DSTC	Distributed Space-Time Coding
MDP	Markov Decision Process
PHY	Physical layer, first layer in OSI network standard
MAC	Media Access Control, second layer in OSI network standard
GA	Genetic Algorithm
MCDH	Multi-cell Distributed Heuristic

LIST OF SYMBOLS

Φ	Set of base stations
R_c	Set of relay stations assigned to c-th base station
N	Set of available subcarriers
n_f	Number of available subcarriers
$\mathrm{K}_{c,m}$	Set of users assigned to m-th relay and c-th base station
T_s	Duration of one time slot
α	Time slot division parameter
$s_{c,m}^{u}(i,k)$	Allocation value of i-th subcarrier from u-th station to k-th user served by m-th relay and c-th base station
$p_{c,m}^{u}(i,k)$	Power allocation value for i-th subcarrier on u-th station for k-th user served by m-th relay and c-th base station
$Q_{c,m}^{u}(k)$	Length of the queue at u-th station for k-th user traffic
$r_{c,m}^{u}(k)$	Achievable transmission rate of u-th station for k-th user
σ_z^2	Variance of AWGN
$\sigma_{c,m}^{2}(i,k)$	Variance of interference received on i-th subcarrier to k-th user on RS-UE link
$\omega_{c,m}(k)$	Weight of k-th user
θ	Tolerable interference per subcarrier on RS-UE link
Δf	Bandwidth of one subcarrier
bps	bit per second
Mbps	Mega bit per second
ms	millisecond
Hz	Hertz

log	Logarithm in base 2
$\mathrm{CN}(\upsilon,\nu)$	Complex Gaussian random variable with variance ν and mean υ

INTRODUCTION

0.1 Motivation

Digital connectivity was once a nice to have, even a luxury. But now we have reached to the era of the Internet-dependent life style. This follows the revolutionary momentum of commercialized wireless communication systems in the 1990s, when the cellular phone started to gain its prominent role in people's day-to-day life. The thrive for commercial benefit speeded the evolution of cellular networks which has since shifted from a carrier of voice towards a primary means of data communications. In less than one decade, the rapid advancement of cell phones functionalities has made it the point of convergence for all forms of Internet-based gadgets. It did not take long to make the society ready for the introduction of the Internet of Things (IoT), which incarnated the vision of a life-style benefiting from automatized procedures performed by smart devices. From the technical point of view, IoT benefits from various-size networks of sensors and smart tools that use the Internet as an infrastructure to communicate and receive data services from a cloud of processing elements. According to a recent report on mobile traffic forecast by Cisco, the need for high data rate is expected to rise 10-fold by 2019 compared to 2014 (Cisco Systems Inc., 2017).The same study reports that sharing of Internet traffic sourced by smart devices will grow from 45% in 2016 to 75% in 2021.

This abrupt increase of Internet usage is only possible in the presence of fast, highly portable and interopratable means of communication. Wireless technology has been answering this substantial growing service demand with its rapid evolution through five technology generations. The fifth generation (5G) is the latest of all Internet-based mobile technology generations to date, which is promised to deliver higher capacity as well as higher density of mobile broadband users than its predecessors (Bangerter, Talwar, Arefi, & Stewart, 2014). The ongoing research for designing 5G-adaptable wireless systems is aimed to reduce communication latency as well as lowering power consumption, for better implementation of IoT.

Heterogeneous Networks (HetNets) is one of the methodologies provisioned for 5G technology. Aimed to extend accessibility, HetNet takes the advantage of a mixed system of various cell sizes to fulfill diverse functionalities (Bangerter et al., 2014). Figure 0.1 illustrates a schematic diagram of the HetNet concept that shows macro-, pico-, femto-cells and relays that have been employed in various indoor/outdoor locations to provide a uniform connection to the user equipment (UE).

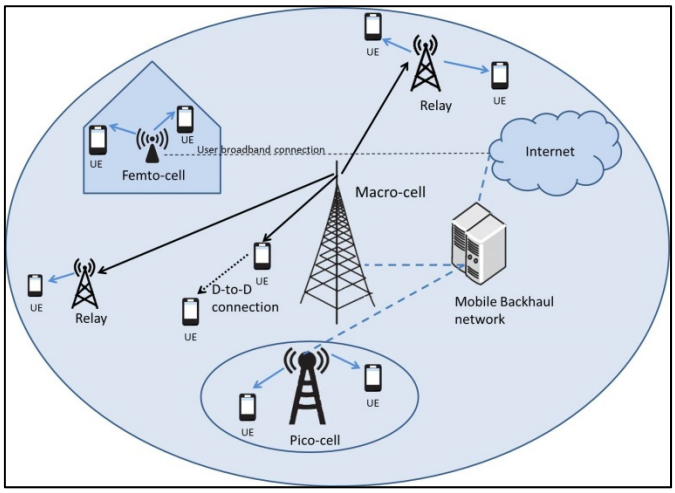

Figure 0.1 A HetNet system consisting of macro-, pico-, femto-cells and relay stations serving user equipment (UE) in various locations, arrows denote wireless links, dashed-lines denote backhaul links

As a reminder of the basics of wireless communication, a Cell is a geographical area in which a base station can cover user equipment for a specific wireless service. The coverage area of each type of base station is limited by choice of the carrier frequency and the limit on power consumption. The cell coverage deterioration factors are the blockage of wireless signals with numerous scattering objects, and the signal attenuation as a function of distance. Developing HetNet is particularly beneficial when the large size cell, namely macro-cell, becomes extensively populated by users and scattering objects. In this case, the users receiving weak signals in the macro-cell can be collaboratively served by low-cost and low-power base stations or relays.

Although HetNet remedies the need for coverage extension and improves the quality of experience for users, yet it is accompanied with numerous design challenges. These challenges have been categorized under cell association, user scheduling, resource allocation, load-balancing, and hand-over (M. Peng, Wang, Li, Xiang, & Lau, 2015). In addition, the combination of design limitation, choice of proper base station, and the versatility of users' service requirements bring a multi-dimensional problem which opens research field for devising novel and highly efficient solutions.

0.2 Problem statement

The transmission medium of wireless networks is a limited range in radio frequency spectrum. Therefore, it is essential to forge an efficient method for sharing this scarce resource among numerous subscribers. Based on the existing literature, the available wireless bandwidth can be accessed by multiple users or streams of traffic incorporating time-, frequency-, space- and code-division or hybrid techniques (Rappaport, 2001). Recently, orthogonal frequency division multiple access (OFDMA) has becomes the popular wireless access method because of its resilience in the presence of frequency selective fading and the ability for fully leveraging the multi-user diversity gain (Asadi & Mancuso, 2013). In particular, OFDMA divides the frequency-selective wideband channels into a number of orthogonal narrowband subcarriers. In the multi-user scenarios, the first step is to define an optimal scheduling mechanism which assigns OFDMA subcarriers to individual users with respect to their channel conditions. It is important for system designers to take into consideration that future mobile technologies are envisioned to be capable of delivering enhanced real-time video and multimedia such as augmented and virtual reality (Cisco Systems Inc., 2017). Therefore, one of the main missions of this research is to devise a scheduler that considers service requirements of real-time and delay-sensitive applications that is crucial for satisfying user experience in next generation wireless systems.

Based on Shannon's capacity theorem, by manipulating the transmission power of wireless signals, one can also tune the achievable data rate. Furthermore, applying OFDMA technique

enables the designers to adjust the transmission power for every single subcarrier that leads to better spectrum efficiency. Furthermore, reducing the energy consumption responds to the critical ecological concerns and helps to build a green communication system. Therefore, the control and optimization of transmission power as a system resource is a critical aspect of wireless design that is embodied in the current work.

As an element of HetNet systems, relay stations played the role of low-cost forwarding stations to increase the coverage of wireless services. In other words, the advantage of relay station becomes more tangible when heavy blockage and long distance travel deteriorate the signal strength for cell-edge users. Inclusion of relay stations shifts the system to the domain of multi-hop communication models. This will change the way the resources are distributed and how service quality is guaranteed, from the case of direct-link communication model. One of the design criterions is to prevent the relay station to become the bottleneck of a two-hop transmission system. Considering the buffered-relaying approach, it is essential to incorporates extra discussions on queue length control. This requires merging the PHY layer features to the MAC layer mechanisms and consequently building a cross layer approach. Therefore, one of the problems to be answered in this work is to define an innovative cross layer mechanism considering the supplementary complexities of a relay-aided communication model.

When it comes to a multi-cell communication system, the resource allocation problem is not limited to the condition of a single cell. In this situation, additional considerations are added to the problem such as combating the adverse effect of inter-cell interference. On the other hand, the signaling overhead and inter-cell communication latency incorporated into the multi-cell allocation mechanism, are limiting factors that reduce the scalability of the allocation method. Our objective is to design an efficient multi-cell allocation method that answers these challenges.

The multi-cell system raises an extra design challenge that corresponds to the extensive complexity of handling the aforementioned multi-dimensional problem. Therefore, it is

crucial to deploy a fast converging solution which imposes low computation cost and rapidly responses to the changes of wireless channels.

0.3 Contribution list of the thesis

Throughout this thesis, we consider a progressive approach to provide novel answers to the aforementioned design problems and challenges. This procedure starts with modeling a two-hop downlink transmission with the objective to enhance the throughput of cell-edge users by employing decode-and-forward relay stations. We study the queue length evolution at each hop and propose a rate control mechanism to stabilize the considered queues. The suggested novel allocation model aims to maximize user throughput with respect to the channel condition and the stability requirements. In order to solve the resulting combinatorial optimization problem we apply a time-shared approach and then convert the outcome to exclusive subcarrier allocation. In addition, we proposed a heuristic algorithm that is stabilizes the system and encounters significantly smaller execution time comparing to the optimization-based method.

Next, we enhance the aforementioned allocation method by considering a multi-cell system that accounts for more general and practical cellular networks. The OFDMA downlink subcarriers and the corresponding transmission power are to be allocated in a two-hop communication model with the objective to maximize system throughput. The multi-cell model embodies extra constraints for controlling the interference to the users of neighboring cells. We study the long-term stability requirements of the two-hop model and propose novel stability constraints. This is different from the stability constraints introduced in our single-cell model in the sense that it does not enquire a priori knowledge of the statistics of the arriving traffic. Since the defined model is a NP-complete problem, we defined a time-shared technique to achieve the closed-form optimal solution in polynomial time.

The proposed time-shared multi-cell approach is profitable when subcarriers can be shared by multiple users during one time-slot. However, time-sharing is not always beneficial

particularly when time synchronization is not possible in the network. For such a condition, we propose an innovative binary allocation mechanism, called OBRA, which exclusively assigns subcarriers to each user during the signaling period and optimizes the power allocation. We show that the complexity of calculating OBRA method is efficient and polynomial in time.

In an approach to follow green communication outlines, we propose an innovative power-conservative binary allocation method for the considered system. This model cuts off the unnecessary power allocation to the least value that can satisfy the stability and interference constraints. We show that the proposed conservative approach, although nonconvex, can reach the optimal solution in polynomial time. This is achieved by applying the theory of Geometric-programming and monomial approximation techniques. The simulation comparison shows that the conservative method remarkably reduces the power consumption.

As our final contribution, we enhance the proposed multi-cell allocation model to offer more scalability for larger number of cells and users. In this regard, we change the centralized allocation method to a novel distributed binary allocation mechanism while respecting the stability, power consumption and interference control criterions. With extensive simulation scenarios, we show that the distributed model is successful to maintain the system stability and improve the energy efficiency while offering a significant decrease in signaling overhead.

0.4 Thesis outline

The rest of the thesis is organized as follows. In CHAPTER 1, we review and analyze the existing literature in the considered domain. CHAPTER 2 introduces a traffic-aware resource allocation method in the downlink of relay-aided single-cell model with respect to stability criteria. This problem is then solved via three novel methods offering various performance advantages such as maximum throughput, exclusive subcarrier allocation and reduced computation burden. We extend the problem proposed in CHAPTER 2 to a multi-cell model

in CHAPTER 3. The extended model integrates transmission power into the allocation problem as well as introducing new queue-aware stability control and inter-cell interference constraints. In this regard, we offer varied solution methods to circumvent the complex allocation problem while achieving maximum throughput, exclusive subcarrier allocation, and energy and computation efficiency. In CHAPTER 4, a novel distributed scheme is proposed for the problem of resource allocation defined in CHAPTER 3 that offers high degree of scalability for crowded networks. The GP related analysis, KKT conditions and self-concordant formulation are instances of covered materials in the appendices.

CHAPTER 1

DESIGN METHODOLOGY AND LITERATURE REVIEW

The increasing demand for coverage extension and power gain, along with the need for decreasing implementation costs, raised the idea of developing Heterogeneous Networks (HetNet) (Ghosh et al., 2012). HetNet cellular systems, which have become an essential part in the evolution of fifth generation (5G) wireless systems, incorporate small cellular entities, such as picocells, femtocells, and relays, to provide a mix of low-power cells underlying the macrocell network. As a member of the HetNet family of solutions, relay implementations not only compensate for signal degradation due to propagation loss, strong shadowing and multipath fading, but also aim at gaining a high data rate and reliable communications (Y. L. Lee, Chuah, Loo, & Vinel, 2014). Furthermore, relaying can considerably save the transmit power since it enables communicating between nodes separated by a large distance. It is a particularly attractive method for low-power communications such as for small, battery-powered sensors (Cai, Yao, & Giannakis, 2005).

In the following, we provide a comprehensive overview of the state-of-the-art in this domain as well as a step-by-step guide for designing an efficient relay-assisted resource allocation mechanism. We also consider similar cooperative resource allocation methods in HetNet, e.g. in femto- or pico-cell, since they face similar set of design challenges for two-hop cooperative transmission.

1.1 Choosing a relaying technique

Relay-aided cooperative transmission, in which a transmitter and a relay cooperate to send a message to cell-edge users who cannot connect directly to the base station, is a popular scheme in wireless networks because of its spatial diversity gain against fading (Laneman, Tse, & Wornell, 2004; Sendonaris, Erkip, & Aazhang, 2003). The existing forwarding methods in relay-aided wireless networks are mainly known as decode-and-forward (DF) (Al-Tous & Barhumi, 2016; Ying Cui, Lau, & Yeh, 2015; Hausl, Işcan, & Rossetto, 2012;

Ng & Schober, 2011; J. Wang, Li, Zhong, Zhang, & Long, 2013), amplify-and-forward (AF) (Jia, Deng, Yang, & Zhu, 2015; Ng & Schober, 2010; Trinh & Pham, 2015), , compress-and-forward (Simoens, Muńoz-Medina, Vidal, & Coso, 2010), demodulate-and-forward (Annavajjala, Maaref, & Zhang, 2010), distributed space-time coding (DSTC) (Jing & Hassibi, 2006) and also different variations and combinations of these methods (Dai et al., 2014). AF offers simplicity and tractability since relays are only responsible for amplifying the signal and forwarding it to the destination. In contrast, CF and DF require more signal processing at relay stations. CF relays compress the received message and forward it to the destination as a new (or same) codeword. DF relays decode the received message and remove the noise then re-encode and forward the message. Demodulate-and-forward is an alternative to DF that reduces the power consumption at the receiver due to channel decoding at the relay. DSTC presents a different perspective by employing relay cooperation to simultaneously transmit different columns of a space-time coding matrix. Although all the relaying methods achieve diversity gain, DF is chosen in this work because it offers substantial performance gain by removing the noise and consequently improving the communication reliability.

Relays can communicate in half-duplex (Ng & Schober, 2011; Rashtchi, Gohary, & Yanikomeroglu, 2016; G. Zhang, Yang, Wu, Mei, & Zhao, 2015) or full-duplex (Jia et al., 2015; Khafagy, Ismail, Alouini, & Aissa, 2013; G. Liu, Yu, Ji, Leung, & Li, 2015) mode. In contrast to full-duplex, a relay cannot receive and transmit at the same time when functioning in the half-duplex mode. In cellular systems, half-duplex needs relay transmissions to be time-multiplexed while the simultaneous send/receive functions in Full-duplex require physical separation of send/receive antennas. Although its performance enhancement attracts more attention for future 5G implementation, deciding to employ full-duplex transmission is a trade-off between the extra cost of signal isolation and performance enhancement. In an approach to reduce the cost of deployment and technical complexity, half-duplex mode is considered as the relaying mode in the current thesis.

There are two categories of schemes to assign the carrier frequency on which the relay stations are operating; in-band and out-band relaying. In-band relaying means that relay uses the same carrier frequency for communication with base station and users. Employing out-band relaying requires dedicated frequency for transmission between relay and base station. The advantage of in-bound relaying for enhancing spectrum efficiency and achieving beamforming gain motivates us to consider it in this work.

1.2 Choosing a channel access method

In a multi-user system, the accessing method of the available network resources has an important role in improving the network performance. In single carrier wireless systems, multiple access strategies coordinates multiple users to access the available bandwidth through isolated time-slots (TDMA), or frequency bands (FDMA), or coding schemes (CDMA), or radiation pattern (SDMA) (Rappaport, 2001). In addition, in cooperative communication, in which an intermediate station assists the transmission to the destination, the aforementioned multiple access techniques need to be adjusted in regards to the problem of intra-cell interference. Many models have been introduced in cooperative communication to access the shared bandwidth based on the aforementioned methods. For instance TDMA is used in a cooperative system in (Cai et al., 2005; Host-Madsen, 2002; Laneman et al., 2004), CDMA in (Cai et al., 2005; Salem, Adinoyi, Yanikomeroglu, & Falconer, 2010; Sendonaris et al., 2003), and FDMA in (Z. Chen, Fan, Li, & Letaief, 2015; Liang, Wang, & Zhang, 2011). The introduction of multicarrier transmission offered higher system spectral efficiency than the single carrier models. Besides the advantages of multicarrier methods, the problem of cross-talk between subcarriers has been a design barrier which led to the idea of orthogonal signals.

Multicarrier orthogonality achieved the highest spectral efficiency with the introduction of OFDM technique. The robustness of OFDM against narrow-band co-channel interference and frequency-selective fading made it the prominent channel access technology for delivering ambitious performances in high speed wireless communication. OFDM provides

an excellent pairing for multi-antenna transmission, which is a promising part of 5G systems, by enabling the spatial interference to be dealt with at a subcarrier level (J. G. Andrews et al., 2014). From resource management point of view, the degree of freedom of OFDMA-based schemes is to allocate different portions of radio resources to different users in both the frequency and time domains (Salem, Adinoyi, Yanikomeroglu, et al., 2010). In particular, a high multiuser diversity gain can be achieved in OFDMA systems since the fading coefficients of different OFDM subcarriers are likely to be independent for different users. These benefits convinced us to employ OFDMA as the access method of the current research. In the following text we study the methodologies for resource allocation problems considering the multi-carrier and multi-user models which are the necessary bodies of advanced wireless networks.

1.3 Choosing the optimization objective for the resource allocation problem

In the context of this thesis, the process of assigning multiple subcarriers to multiple users in a specified timing scheme is called a resource allocation method. Because of the variation in design objectives, network topologies, configuration limitations, and diversity in the application requirements, the problem of resource allocation has been attracted a great number of researchers. In the framework of optimization problems, an objective or utility function quantifies an abstract concept and develops a tangible performance metric. In this section, we review the optimization objectives defined for resource allocation problems in the literature.

Throughput maximization, or maximizing the spectral efficiency has been one of the most popular optimization objectives in the relay-aided allocation problems (Al-Tous & Barhumi, 2016; Ho et al., 2015; Y. Liu & Chen, 2012; Ng & Schober, 2011; Phan, Le-Ngoc, Vorobyov, & Tellambura, 2009; Salem, Adinoyi, Yanikomeroglu, & Falconer, 2011; Son, Lee, Yi, & Chong, 2011). It is known that, selecting the best user for each OFDM subcarrier and adjusting the corresponding transmission power leads to maximum system spectral efficiency (Ng & Schober, 2011). A joint subcarrier and power allocation mechanism

introduced in (Ng & Schober, 2011) that maximizes a concave utility function that is defined as the minimum of instantaneous rate between base station-relay link (BS-RS) and relay-user link (RS-UE). Authors of (Son et al., 2011) maximized the sum of weighted rates in a distributed manner for each base station. Therein, to simplify the computation complexity, the subcarrier allocation is decoupled from power assignment. A joint subcarrier and transmission power allocation is proposed in (Al-Tous & Barhumi, 2016) with the objective to maximize the overall users weighted rates. Multiple approaches are then introduced considering both AF and DF relaying for high signal to noise ratio (SNR) channel condition, however these methods are not efficient for low SNR scenarios.

The throughput-optimal allocation model introduced in (Y. Liu & Chen, 2012) improves the spectrum efficiency by enabling the base station to transmit during relaying sub-frame on subcarriers unused by the relay. In a high SNR assumption, authors of (Phan et al., 2009) formulate the throughput maximization objective of the relaying power allocation problem to a utility function that is the product of SNR values. A sum-rate optimization problem defined in (Ho et al., 2015), solved the power allocation problem for the two-hop cooperative uplink transmission. Despite the advantage of decreasing the messaging overhead for single carrier method in (Ho et al., 2015), it can cause extensive computation and signaling load in multicarrier scenarios.

Power minimization is another design objective that can be applied in the relay-aided resource allocation problem. For instance, authors of (M. Chen, Serbetli, & Yener, 2005) defined a power allocation scheme in a DF relaying system with the objective to minimize the total transmit power while satisfying a target level of SNR at the destination. Among several power allocation methods for relaying systems that are deployed in (Phan et al., 2009), one notable approach consists of minimizing the maximum transmit power at the base station following the rational that energy constraints is less sever on relays.

Proportional fairness is a widely used resource allocation objective that compromises the network capacity and user fairness (Huang, Rong, Wang, Xue, & Schulz, 2007; Kelly,

Maulloo, & Tan, 1998; Z. Tang & Wei, 2009; Viswanath, Tse, & Laroia, 2002). For instance, the method introduced in (Viswanath et al., 2002) allocates a resource unit to a user targeting to maximize the ratio of its achievable rate to the exponentially weighted average rate.

The fact that a large number of papers have been dedicated to optimize the system throughput indicates the validity of this approach for optimal allocation of wireless resources. We thus chose throughput maximization as our optimization objective targeting to fill the gap for an efficient method in low SNR regime. Aside from the main objective, it is important to efficiently define the optimization constraints to confine a feasible solution in regard to specific performance criteria. For instance, optimization constraints can be defined for limiting the power consumption (Ho et al., 2015), co-channel interference (Ho et al., 2015; Lin, Tao, Stüber, & Liu, 2013; Ng & Schober, 2011; Venturino, Prasad, & Wang, 2009) or outage probability (Ho et al., 2015; Zarakovitis, Ni, Skordoulis, & Hadjinicolaou, 2012).

1.4 Cross-layer resource allocation

In previous section we pointed out the methods that only employ PHY layer features in the allocation problem. However, it has been shown that the joint optimization of resources allocation across PHY layer and MAC layer features, so called cross-layer methods, leads to significant throughput gain and efficiency. One of the main features of the MAC layer is the buffering of data in Queue-based data structures. Queue length has been employed in cross layer methods as a means of fairness, delay control, or stability. For instance, the introduced utility function for throughput maximization defined in (Salem, Adinoyi, Rahman, et al., 2010; Salem et al., 2011) is formulated as the product of rate and differential backlog. The idea of differential backlog that implies the difference of buffer length between the transmitter and the receiver, has been first introduced in (Neely, Modiano, & Rohrs, 2005) for a power allocation problem in wireless networks. It is shown in (Salem et al., 2011), that coupling the rate and buffer length in resource allocation problem leads to a higher fairness in throughput compared to traditional proportional fair scheduling methods. A utility fair rate

allocation in the downlink of cellular networks is considered in (Eryilmaz & Srikant, 2007). Therein, rates are allocated based on optimizing a queue-rate product utility function while the proposed admission control policy is in effect.

1.4.1 Delay optimal methods

Among various cross-layer approaches, delay optimal schemes focus on minimizing some functions of average delay using queuing features (Y. Cui, Lau, Wang, Huang, & Zhang, 2012). The method introduced in (Ying Cui et al., 2015) dynamically schedules the transmission in a two-hop relayed system targeted to minimize the average sum queue length over a finite horizon. Therein, the coupling of queue length information and quality of link at the source and relay forms an asymptotically optimal delay policy. Against the optimal closed-form expression in (Ying Cui et al., 2015), considering an on-off link model disqualifies this approach for multicarrier systems that offer higher diversity gain. A delay optimal multi-hop link selection mechanism is developed in (Tassiulas & Ephremides, 1994) using the technique of forward induction. The resulted method however is not applicable for random packet length.

Markov Decision Process (MDP) is a systematic approach for delay-aware resource allocation problems. In particular, MDP employs either the stochastic learning or the differential equation method to solve a problem formulated in Bellman equation. MDP is employed in (N. Wang & Gulliver, 2015) for optimizing the delay for a two-hop multi-relay network in a distributed and low-complexity mechanism. The delay-optimal power and subcarrier allocation problem of a single macrocell is defined in (J. Li, Peng, Cheng, Yu, & Wang, 2014) and modeled as a K-dimensional infinite horizon average reward MDP. To deal with the exponential state space, authors of (J. Li et al., 2014) employed reduced state Bellman equation, linear approximation and online stochastic learning. Delay constrained downlink cooperative transmission is modeled in (Lau, Zhang, & Cui, 2013) with discrete time MDP in which several approximation techniques are employed to find closed-form results. Despite the advantage of MDP for general delay regimes, dimensionality is a curse

since it is related to the exponential growth of the state space. Indeed the coupling of QSI with CSI and interference, makes a complicated stochastic model that usually requires a series of approximation techniques to solve it. Therefore, the final approximated results can suffer from some degree of imprecision.

Another approach for delay optimal mechanism in multi-hop networks is based on the asymptotic analysis of the large delay regime. With this perspective, reference (Neely, 2006) proposes a power-delay tradeoff to obtain near optimal dynamic control strategy for general queueing systems using Lyapunov optimization theory. In Lyapunov optimization theory, the formulation on average delay is solved by converting it to the Lyapunov drift-plus-utility function and finding the minimum value of this function.

QoS exponent is a widely used metric to analyze systems operating under statistical QoS constraints (Chang, 1994; Qiao & Gursoy, 2016; Qiao, Gursoy, & Velipasalar, 2013; J. Tang & Zhang, 2007; Wu & Negi, 2003). In particular, QoS exponent is the exponential decay rate of queue overflow probability when capacity of queue approaches infinity. Therefore, to lower down the probability of overflow, one can enforce large QoS exponent. Using the QoS exponent concept, the effective capacity as a Queue-aware performance metric, is introduced in (Wu & Negi, 2003). Effective capacity determines the maximum constant arrival rate that can be supported in a queue system by a given departure rate or service process, while satisfying statistical QoS constraints. Authors in (Qiao & Gursoy, 2016; Qiao et al., 2013) investigate the effective capacity of a two-hop wireless link subject to specific QoS exponents at each hop. In (J. Tang & Zhang, 2007) QoS exponent is employed to define a stringent QoS constraint in a throughput maximization problem for wireless relay-aided resource allocation.

1.4.2 System stabilizing methods

Queue stability is a cross-layer feature that can efficiently prevent congestion, buffer overflow and extensive delays. Stability has been integrated in resource allocation

approaches with different directions (Chang, 1994; Georgiadis, Neely, & Tassiulas, 2006; Neely et al., 2005; Qiao et al., 2013; Tassiulas & Ephremides, 1992; Zarakovitis et al., 2012). It is known that the queues are strongly stable if they have bounded time-average backlogs. One of the initial works in this domain is (Tassiulas & Ephremides, 1992) which shows that allocating resources to maximize the queue-length-weighted sum of rates is a stabilizing policy. Authors in (Qiao et al., 2013) introduce stability constraints that require the average transmission rate of source-relay link being smaller than the average transmission rate of relay-destination link. This policy is then utilized to define optimal time-slot division ratio in the downlink of relay-aided transmission system constraint on predefined QoS requirements. In (Zarakovitis et al., 2012), a cross-layer approach is proposed targeting to minimize the transmission power in OFDM subcarrier and power allocation problem. Using queueing theory, the authors of (Zarakovitis et al., 2012) proposed a constraint to guarantee the stability when the tolerable delay approaches to infinity. This cross-layer model, although elegant, does not consider multi-hop condition.

Lyapunov theory can also be an efficient method for defining stability condition of a system. Using Lyapunov theory, authors of (Georgiadis et al., 2006) employ Dynamic Backpressure to ensure the stability of the queues in the system if the traffic arrival rate falls within the stability region of the system. A similar approach is followed in (Neely et al., 2005) for adaptive power allocation aiming at maximizing the rate-backlog product and a Lyapunov drift analysis is utilized to define the sufficient condition of stability.

It can be seen that the existing delay optimal and stability mechanisms consider very large buffer capacity and are mostly good for large delay regimes; however they do not guarantee a satisfactory delay performance specifically in small delay regimes. The stability method introduced in the current thesis is designed to fill this gap observed in the literature.

1.5 Choosing a solver

Solving a comprehensive resource allocation problem that encompasses various constraints and performance metrics adjures optimization techniques to achieve efficient utilization of network resources. To prevent the co-channel interference in an OFDMA-based multiple access network, it is necessary to assign subcarriers exclusively to each user/traffic flow. However, the exclusive subcarrier assignment brings the allocation problem into the domain of binary programming that is NP-complete and infeasible to solve in polynomial time. This complex problem has received remarkable attention in the literature. In this section we review the existing works on the efficient linearization and approximation methods to solve the binary and non-convex allocation problem in a reasonable calculation time. An insightful research that compares various mathematical approaches to solve the problem in polynomial time is done in (Y. F. Liu & Dai, 2014) which studies the computation complexity of joint subcarrier and power allocation problem in different network scenarios.

The most popular approach to tackle the combinatorial problem is to relax the binary subcarrier allocation parameters to real values between zero and one, which implies the fraction of time a user can occupy a subcarrier (Ng & Schober, 2011; Wong, Cheng, Ben Letaief, & Murch, 1999; Zarakovitis et al., 2012). It has been shown that the relaxed problem results in a high precision estimation of the optimal binary allocation (Wong et al., 1999; Yu & Lui, 2006). The binary parameters in channel and power allocation problem in (Mao, Wang, & Lin, 2005) are first relaxed, and then a branch-and-bound based algorithm is proposed to solve the linearized problem. Therein, it is shown that the computation efficiency of the proposed method is remarkably improved compared to the exhaustive search scheme.

The other strategy to simplify the problem is to perform the channel allocation procedure separately from power allocation (Abdelnasser, Hossain, & Kim, 2014; Son et al., 2011; Stolyar & Viswanathan, 2009; Venturino et al., 2009; H. Zhang et al., 2014). In general, this strategy assigns each subcarrier to the user with the best channel gain for that subcarrier, and then a water filling-based method is employed to optimally allocate the power.

Lagrangian optimization technique is one of the most popular schemes to solve a complex constrained optimization problem with the help of an easier problem, the so called dual problem, in an enlarged feasible set (Y. Liu & Chen, 2012; Loodaricheh, Mallick, & Bhargava, 2014; Ng & Schober, 2011; Wunder, Zhou, Bakker, & Kaminski, 2008; Zarakovitis et al., 2012; H. Zhang et al., 2014). This approach simplifies the solution of nonconvex problems because the dual function forms a convex function and its complexity depends only on the number of constraints (Yu & Lui, 2006). For instance, by using the Lagrangian function of the subcarrier and power allocation problem in (Ng & Schober, 2011; H. Zhang et al., 2014), the dual problem is defined and then decomposed into subproblems to be solved in a semi-distributed manner. It is shown that the duality gap, that is the gap between the solution of dual problem and the original nonconvex problem, tends to zero as the number of frequency carriers grows very large (Yu & Lui, 2006). Accordingly, one can find various methods to solve the dual problem. For instance bisection search, as employed in the resource allocation problem in (Son et al., 2011), is a good method but its complexity grows exponentially with the number of constraints. Subgradient methods, as employed in relaying resource allocation problem in (Z. Tang & Wei, 2009) and femtocell spectrum sharing in (H. Zhang et al., 2014), offer the flexibility of being convergent for non-differentiable objective functions and enable distributed implementation (Boyd & Vandenberghe, 2004).

Hungarian algorithm is a primal dual strategy that solves combinatorial problems in polynomial time. The problem of OFDM subcarrier and power allocation is solved in (Basaran & Kurt, 2016) using Hungarian method targeting to minimize power consumption and outage probability. Hungarian method is employed in (Loodaricheh et al., 2014) to solve the energy-efficient power and subcarrier allocation problem in a cooperative cellular network. In (Salem et al., 2011) Hungarian is implemented to solve a cross layer fair resource allocation problem in OFDMA-based relying network.

Game theory is another approach to solve the multi-user two-hop radio resource allocation problems by modeling the interactions between active nodes (Bu, Yu, & Yanikomeroglu, 2015; Han, Peng, Zhao, & Wang, 2013). This approach relies on finding equilibrium state between individual agents which are trying to maximize their own selfish objectives.

Graph coloring is employed to address the interference control and channel allocation problem of HetNets in (Bhatia, Li, Luo, & Ramjee, 2006; Uygungelen, Auer, & Bharucha, 2011; Zaki & Fapojuwo, 2011; Zhou et al., 2016). It has been shown that graph coloring can be more computing efficient than other optimization methods depending on the special structure of the problem (Zaki & Fapojuwo, 2011).

Evolutionary algorithms, which are a group of heuristic methods that generate solution following the principle of natural evolution, can be deployed to obtain good sub-optimal solutions to the resource allocation problem in wireless networks (Cao, Zhang, Zeng, Chen, & Chai, 2014; Marshoud et al., 2016; Saad, Bettayeb, & Boubas, 2007; Saad & Muhaureq, 2008). For instance, Genetic algorithm (GA) is applied in (Marshoud et al., 2016; Saad et al., 2007) to solve the multidimensional allocation and routing problem in HetNet systems. The authors of (Marshoud et al., 2016) show that GA is advantageous over linear binary programming method (Estrada, Jarray, Otrok, & Dziong, 2014) in terms of speed of convergence. Particle Swarm Optimization is employed in (Saad & Muhaureq, 2008) to solve the joint problem of routing and resource allocation in multi-hop cellular systems. Therein, authors claim that Particle Swarm Optimization is easier to implement and more computationally efficient than GA. Using Quantum Particle Swarm technique, authors of (Cao et al., 2014) defined a relaying link selection method to solve the multi-objective problem for SNR maximization and power efficiency.

Geometric Programming (GP) is known as a convenient means of solving nonconvex problem while sustaining polynomial complexity. GP technique is employed in (Chiang, Tan, Palomar, O'Neill, & Julian, 2005) and provides a sub-optimal solution for the problem of power allocation in multiple-access systems. Authors of (Rashtchi et al., 2016; Rashtchi,

Gohary, & Yanikomeroglu, 2014) employed GP and monomial approximation technique to achieve optimal solutions for the nonconvex problems of joint resource allocation and routing in multi-hop multicarrier generic wireless systems. In (Phan et al., 2009) the power allocation problem in relaying networks is addressed by minimizing the total transmitted power while guaranteeing a minimum QoS level. Therein, authors showed that the global optimum is achievable by converting the nonlinear and nonconvex problems into a GP format. In (Simoens et al., 2010) GP is employed in the iterative coordinate descent method to solve the combinatorial and nonconvex power allocation problem. The advantage of GP that readily converts a nonconvex problem into a convex one and the availability of large-scale software solvers makes GP an appealing optimization model which is employed in this thesis.

A class of resource allocation researches employ heuristic methods to solve the nonconvex and combinatorial optimization problem (H. Zhang et al., 2014; Y. J. Zhang & Ben Letaief, 2004). For instance, in (Y. J. Zhang & Ben Letaief, 2004) the channel allocation is based on a heuristic greedy-based approach targeting to maximize the system throughput. Those heuristic methods, although suboptimal, offer reduced computation complexity and their performance is remarkably close to that of optimal solution. With this perspective, in this thesis we introduce highly efficient heuristic algorithms as low complexity approaches to the proposed optimization methods to save on computation time and energy.

1.6 Challenges of multi-cell cooperation

1.6.1 Frequency planning

In multi-cell system design, frequency planning is the procedure to define the range of frequencies that each cell is allowed to use. Since the wireless channels are scarce resources, it is desired to reuse the available channels as much as possible. Relay deployment in multi-cell systems adds another degree of freedom to frequency planning since it can make inter-cell interference to users in neighbor cell. To elaborate this, we investigate three examples of

frequency planning in relay-aided cellular systems which are depicted in Figure 1.1. Note that in these examples a separate range of frequencies are dedicated to direct transmissions, which are in the central area of each cell and color coded in yellow. In subfigure (a) a unique range of frequencies is available for serving the relay-aided users in each adjacent cell, which implies the frequency reuse factor equals one. In this case, the spectral efficiency is high but inter-cell interference between adjacent cells can decrease the received signal quality for cell-edge users. In subfigure (b), the available bandwidth is divided to three frequency ranges, i.e. reuse factor equals 3, which completely prevents inter-cell interference. Considering that in case (b) the available channels are statically divided between adjacent cells, its spectral efficiency is one third of the case (a). In subfigure (c), reuse factor is three but each cell is also divided to three sectors using directional antennas. The latter case, so called Fractional Frequency Reuse (FFR), eliminates inter-cell interference with sectorization (Novlan, Ganti, Ghosh, & Andrews, 2012), however fixed division of available channels reduces the spectral efficiency to one third of the example (b).

This brief comparison shows the trade-off between higher spectral efficiency and lower inter-/intra-cell interference. Since the available wireless channels are even more limited in recent cellular and HetNet systems, achieving higher spectral efficiency becomes the main design goal of many researches as well as current thesis. In this regard, efficient interference management techniques are required to eliminate the adverse effect of co-channel interference on signal strength.

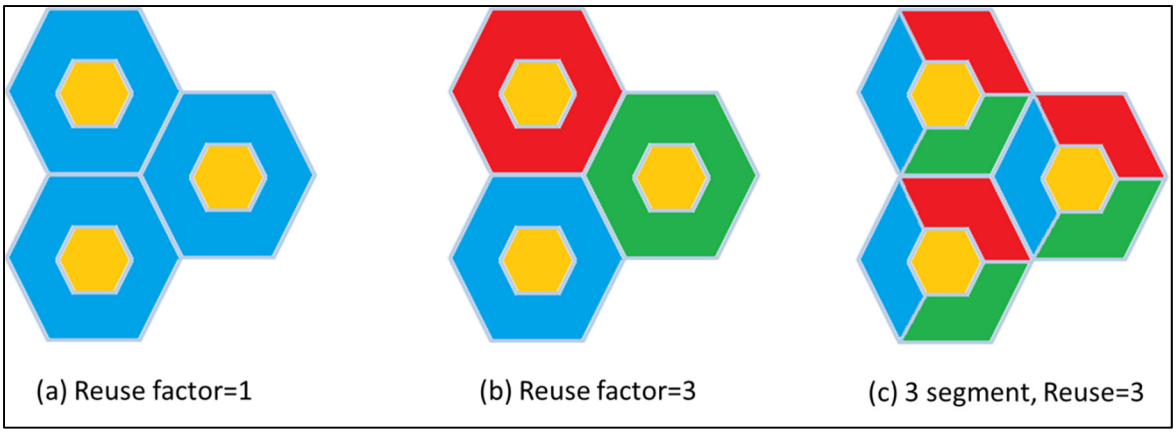

Figure 1.1 Three examples of frequency planning for relay-aided cellular systems, different colors imply different frequency range assigned

1.6.2 Inter-cell interference management

Universal frequency reuse, although it increases spectral efficiency, requires extra processing and messaging to control the co-channel interference for users in the neighbor cells. In the context of the inter-cell interference management, neighbor cells are required to exchange the channel state information (CSI) of their cell-edge users and to perform extra signal processing to adapt their allocation mechanism. In this cooperative scheme, the efficiency of the resource allocation increases by reducing the extra traffic load and the corresponding energy consumption due to the CSI feedback signaling. In this section, we review the existing body of research on the topic of interference handling and signaling overhead reduction.

The centralized approach in (Venturino et al., 2009) introduces reducing feedback methods which are deployed based on channel estimation and tone grouping. Although the centralized methods return a high precision result, they rely on signaling to the central station which is a scalability barrier for all centralized methods when the number of cooperative cells grows. In other words, when multi-cell system is highly populated and the number of cooperative cells grows largely, the signaling overhead and its communication delay to the central station becomes large, which reduces the overall performance of system. Therefore, it is more

reasonable to use centralized methods for less populated networks with a small number of cells.

Semi-distributed approaches are introduced in (Ng & Schober, 2011; H. Zhang et al., 2014) to reduce the feedback signaling that eliminates the dependence to a central station. Therein, the dual decomposition and subgradient techniques are employed which enables a collaborative resource allocation mechanism. The authors of (Ng & Schober, 2011) discuss the practicality of their method by showing that the delay imposed by multiple inter-cell information exchange and computation iterations per slot is smaller than the coherence time of a pedestrian user. A device-to-device relay-aided transmission model is considered in (Hasan & Hossain, 2015) that employs distributed resource allocation. In the latter work, multiple messages passing during one time slot increases the delay and decreases the practicality of the method for high velocity nodes and delay sensitive usages.

Distributed relay-aided methods introduced in (M. Chen et al., 2005) identify the optimum solution based on the partial knowledge of CSI of the cell-edge users to reduce the load of information exchange in each time-slot. In (Stolyar & Viswanathan, 2009), the inter-BS messaging frequency is reduced to one per n time-slots that condenses the complexity of updating power allocation to $1/n$. However, this method still needs several message passing routines during one time slot for the user scheduling procedure. To further reduce the signaling overhead, the potential interference is calculated in (Abdelnasser et al., 2014; Son et al., 2011) based on the worst case scenario in which a reference station is detected with highest channel gain on the link to the neighbor cell. In this manner, the neighbor relay (or base station) does not require the channel state information of all users and the information exchange is limited to the allocation parameters of the reference station. The signaling overhead reduction in reference-based mechanisms is further improved by grouping the inter-cell messages and quantizing the information, which are the approaches embodied in the proposed distributed method in the CHAPTER 4 of the current thesis.

1.7 Conclusion

In this chapter we presented the required steps to design an efficient relaying scheme in a cellular network. We reviewed the body of research over different aspects and compared the advantages and drawbacks of different methodologies. It can be seen that for every design step for an efficient relaying resource allocation scheme, there are various options and directions. This highlights the importance and necessity of a thorough literature review to deeply understand the problem and state of the art, as it was our target in writing this chapter. In the light of the existing research achievements we define our thesis in the following chapters targeting to fill the gaps and also benefit from successful approaches.

Because of the similarity between relaying resource allocation problem with other cooperative allocation methods in HetNets, we included those methods in our literature review. This expands the audience of this thesis to researchers who are willing to work on similar domains.

CHAPTER 2

CHANNEL- AND TRAFFIC-AWARE RESOURCE ALLOCATION CONSTRAINED TO SYSTEM STABILITY

In this chapter we propose a channel- and buffer-aware resource allocation method in the downlink of a relay assisted wireless network. The considered system employs DF relays to serve the cell-edge users. In order to eliminate the intra-cell interference, we assume that the transmissions of base station and relay stations are differed in time (Hausl et al., 2012). In this model, stability is a crucial criterion and a method needs to be implemented to prevent the relay stations from becoming bottlenecks. To fulfill this requirement, we model the buffers residing in the base station and relay stations, as systems of queues in tandem. Then, we introduce a set of stability constraints to the resource allocation problem. The ultimate goal in this chapter is to achieve the maximum system throughput by allocating the available subcarriers in the downlink of OFDMA system to relay-assisted users with regard to the stability constraints.

2.1 System model

The system model consists of a cell with one base station (BS), M relay stations (RS) and K total users. The BS is equipped with K buffers corresponding to each user. A schematic of the considered system is illustrated in Figure 2.1. Furthermore, we consider that each relay station has buffers dedicated to its users. We consider an OFDMA-based downlink with n_f subcarriers available for serving the relay-aided users. Such users are associated to the relays with strongest average channel gain. The users who do not suffer from dispersed and weak signals are assumed to be connected directly to the base station via separate resources. The resource partitioning issue is not studied in this work (interested researchers can read more on this subject in (de Moraes, Nisar, Gonzalez, & Seidel, 2012)). Continuous rate adaptation

scheme is considered, and it is assumed that the channel state information (CSI) of each link is known at the base station[1].

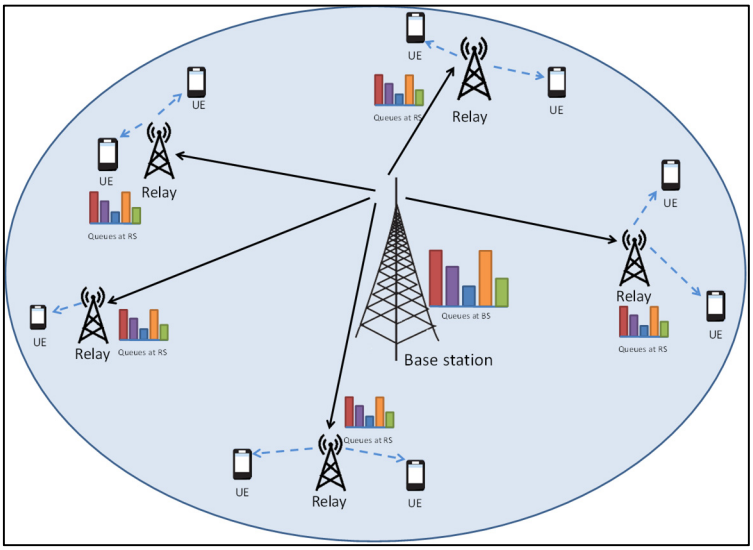

Figure 2.1 A cell with three relay stations serving the relay-assisted cell-edge users

To prevent intra-cell interference, the relay-assisted downlink transmission is proposed to be performed in two mini time-slots each equal to half of one time slot T_s. In particular, the base station transmits the data to the relay station in the first mini time-slot, using the entire available downlink sub-frame. Then, the relay station decodes the received signals, removes the noise and forwards the data to the user equipment (UE) in the second mini time-slot by using the entire available subcarriers n_f in the downlink sub-frame.

[1] Base station can acquire the exact or estimated value of channel state via signaling in previous time slot. Readers can find more details in (Y. Li, Cimini, & Sollenberger, 1998) since this subject is out of scope of the current work.

2.1.1 Channel capacity

In the OFDMA downlink with n_f subcarriers, the channel impulse response is assumed to be time-invariant during each time slot. The received OFDM symbol to the *m*-th relay $m \in \{1,2,..,M\}$ that is associated with the data stream of *k*-th user $k \in \{1,2,..,K\}$, sent from base station in the *i*-th subcarrier $i \in \{1,2,...,n_f\}$ at the *t*-th time slot, is given by

$$Y_m^{BS}(i,k,t) = \sqrt{P_m^{BS}(i,k,t) l_m^{BS}}\, H_m^{BS}(i) X(i,k) + Z_m^{BS}(i), \tag{2.1}$$

where $X(i,k)$ and $P_m^{BS}(i,k,t)$ represent respectively the transmitted symbol and power on the BS-RS link (i.e. the link from base station to the relay). l_m^{BS} is the path loss between base station and *m*-th relay. $Z_m^{BS}(i)$ is the Additive White Gaussian Noise (AWGN) at *m*-th relay using *i*-th subcarrier that is modeled by a zero mean complex Gaussian random variable and variance of σ_z^2. The shadowing/fading coefficient in the link between base station and *m*-th relay using *i*-th subcarrier is calculated by $H_m^{BS} = \gamma_m^{BS} \eta_m^{BS}$. γ_m^{BS} is the shadowing coefficient that is modeled with i.i.d random variables of lognormal distribution with zero mean and variance of $\sigma_{\gamma,BS}^2$. η_m^{BS} is the flat fading coefficient in LoS link and is modeled with i.i.d random variables of Rician distribution (Suraweera, Louie, Li, Karagiannidis, & Vucetic, 2009).

Since the base station and relays are usually placed in high positions, the propagation channel is expected to be in strong Line-of-Sight (LoS) with negligible blockage issues, therefore $H_m^{BS}(i)$ can be modeled by independent and identically distributed (i.i.d) random variables with a Rician distribution (Suraweera et al., 2009). With the help of directional antennas and applying interference cancellation techniques at the relay stations (Jeffrey G. Andrews, 2005), the inter-cell interference received at the relay station is negligible compared to the channel noise (Ng & Schober, 2011). Therefore, the received SNR at the *m*-th relay is given by

$$\Lambda_m^{BS}(i,k,t) = \frac{P_m^{BS}(i,k,t) l_m^{BS} \left| H_m^{BS}(i) \right|^2}{\sigma_z^2}$$

and the corresponding channel capacity on BS-RS link can be formulated as

$$C_m^{BS}(i,k,t) = \Delta f \log_2\left(1+\Lambda_m^{BS}\ (i,k,t)\right) \tag{2.2}$$

The message $X(i,k)$ sent by the BS, is then decoded at the *m*-th relay and forwarded to the *k*-th user. Hence, the received symbol at the *k*-th user from the *m*-th relay using the *i*-th subcarrier at the *t*-th time-slot, can be presented as

$$Y_m^{RS}(i,k,t) = \sqrt{P_m^{RS}(i,k,t)l_m(k)}H_m^{RS}(i,k)X(i,k) + Z^{RS}(i,k) \tag{2.3}$$

where variables $P_m^{RS}(i,k,t)$, $l_m(k)$, $H_m^{RS}(i,k)$ and $Z^{RS}(i,k)$ are defined with the same logic as the variables of BS-RS link; only the signaling direction is from relay station to user equipment (RS-UE link). Since users are randomly located and surrounded by numerous scatterers, the RS-UE link should be modeled as non-Line-of-Sight-(NLoS) link. Hence the shadowing/fading coefficient $H_m^{RS}(i,k)$ can be modeled by $H_m^{RS} = \gamma_m^{RS}\eta_m^{RS}$ where γ_m^{RS} is the shadowing coefficient that is modeled with i.i.d random variables of lognormal distribution with zero mean and variance of $\sigma_{\gamma,RS}^2$. η_m^{RS} is the flat fading coefficient in NLoS link and is modeled with i.i.d random variables of Rayleigh distribution (Suraweera et al., 2009). Therefore, the received SNR at the *k*-th user is given by

$$\Lambda_m^{RS}(i,k,t) = \frac{P_m^{RS}(i,k,t)l_m(k)\left|H_m^{RS}(i,k)\right|^2}{\sigma^2}$$

and the channel capacity on the RS-UE link can be formulated by

$$C_m^{RS}\left(i,k,t\right) = \Delta f \log_2\left(1+\Lambda_m^{RS}\left(i,k,t\right)\right) \tag{2.4}$$

We define $s^{BS}(i,k,t)$ as the binary allocation parameter at BS-RS link. In this regard, $s^{BS}(i,k,t)=1$ if the *i*-th subcarrier is allocated to transmit data of *k*-th user at the first mini time-slot, and $s^{BS}(i,k,t)=0$ otherwise. Therefore, user's achievable data transmission rate on the link between the base station and the *m*-th relay can be formulated as

$$r^{BS}(k,t) = \sum_{i=1}^{n_f} s^{BS}(i,k,t)C_m^{BS}(i,k,t) \tag{2.5}$$

We also define the binary allocation parameter at the second mini slot in RS-UE link, denoted by $s^{RS}(i,k,t)$. Consequently, the transmission rate associated to the link between the *m*-th relay and the *k*-th user is given by

$$r_m^{RS}(k,t) = \sum_{i=1}^{n_f} s^{RS}(i,k,t)\ C_m^{RS}(i,k,t) \tag{2.6}$$

2.1.2 Traffic-aware stability control

The two-hop transmission system considered in the current chapter is modeled as a network of queues in series (Bertsekas & Gallager, 1992). In this model, the stream of data associated to *k*-th user arrives at the *k*-th queue in the base station with a particular arrival bit rate $\lambda(k)$ during one time slot. The total capacity of a set of subchannels is considered as the service capacity of a queue. Therefore, the achievable transmission rate of the selected subchannels is the service rate of the queue. Therefore, the evolution of length of queues placed at the base station and associated to the *k*-th user $Q_{BS}(k,n)$, measuring in the number of bits, at the end of *t*-th time-slot (i.e. after two mini-slots), can be given by

$$Q_{BS}(k,t) = \left[Q_{BS}(k,t-1) - d_{BS}(k,t)\right]^+ + \lambda(k)T_s \tag{2.7}$$

where $d_{BS}(k,t)$ denotes the number of departed bits from the base station. Note that $[x]^+$ equals x if $x > 0$ otherwise it equals 0.

It is known that if the average arrival rate grows larger than the average service rate in a queue, the queue length and consequently the queuing delay will grow indefinitely which implies that the system becomes unstable (Bertsekas & Gallager, 1992). Therefore, one can prevent the unstable condition by enforcing

$$\lambda(k)T_s \le d_{BS}(k,t) \tag{2.8}$$

in the slot-by-slot resource allocation procedure. The departure rate is upper-bounded by the achievable rate on the BS-RS link, i.e. $d_{BS}(k,t) \le r^{BS}(k,t)\frac{T_s}{2}$, so Eq. (2.8) can be reformulated as

$$\lambda(k) \le \frac{1}{2} r^{BS}(k,t). \tag{2.9}$$

Similarly, the queue length growth at the *m*-th relay station at the end of *t*-th time-slot, is given by

$$Q_m(k,t) = \left[Q_m(k,t-1) - d_m(k,t) \right]^+ + a_m(k,t). \tag{2.10}$$

where $a_m(k,t)$ is the number of arrived bits to the *k*-th queue during the *t*-th time slot that is actually equal to the number of departed bits from the base station $d_{BS}(k,t)$. $d_m(k,t)$ in Eq. (2.10) denotes the number of departed bits from the relay station, which is practically upper-bounded to $r_m(k,t)T_s/2$. We define the stability requirement on relay's queue by $a_m(k,t) \le d_m(k,t)$, which can be reformulated as

$$d_{BS}(k,t) \le r_m^{RS}(k,t)\frac{T_s}{2}. \tag{2.11}$$

Using Eq. (2.8), one can rewrite Eq. (2.11) to

$$\lambda(k) \le r_m^{RS}(k,t)\frac{1}{2}. \tag{2.12}$$

Ultimately, Eq. (2.9) and Eq. (2.12) are the sufficient conditions for stability of queues in the relay-aided transmission model. These conditions are based on the assumption that the base station can predict the traffic arrival rate.

2.2 Defining an optimization model to solve the resource allocation problem

In this section, we define an optimisation model for subcarrier allocation problem in the downlink subframe using decode-and-forward relays; where the allocation decision is made centrally in the base station at the beginning of each time slot. The goal is to assign subcarriers

to each user in a way that maximizes the sum throughput of users based on instantaneous channel condition. The channel state varies from one time slot to another, so the optimization operation is performed once per time-slot and it repeats for the next time slot. Therefore, without loss of generality, we can eliminate the time-slot index (t) from the parameters in the remaining equations.

The instantaneous throughput of *k*-th user served by *m*-th relay is given in

$$\rho(k)=\min\left[\sum_{i=1}^{n_f} s^{BS}(i,k)\ C_m^{BS}(i,k), \sum_{i=1}^{n_f} s^{RS}(i,k)\ C_m^{RS}(i,k)\right], m\in\{1,2,...,M\}, k\in\{1,2,...,K\} \quad (2.13)$$

which is limited by the lowest value among the rate on BS-RS link Eq.(2.5) and RS-UE link Eq.(2.6).

It has been shown that by using an adaptive subcarrier allocation policy in favor of users with better channel condition, the throughput reduction is negligible when a fixed and uniformly distributed power allocation scheme is employed (Y. J. Zhang & Ben Letaief, 2004). Since the channel state information has already been fed back to the base station and considering the equal power allocation on all subcarriers, the channel capacities Eq. (2.2) and Eq. (2.4) can be pre-calculated and then be treated as constant values in the optimization procedure. So, the objective function that is given by

$$\max_{s^{BS},s^{RS}} \sum_{k=1}^{K} \omega(k)\rho(k),\ k\in\{1,2,...,K\} \quad (2.14)$$

becomes linear and the only optimization variables are subcarrier allocation parameters $s^{BS}(i,k)$ and $s^{RS}(i,k)$. $\omega(k)$ is a positive weight parameter that specifies scheduling priority and can enforce fairness policies to meet different users' requirements similar to approaches in (Song & Li, 2005).

To enforce the allocation mechanism to meet the required condition, the optimization constraints are defined in

$$\sum_{k=1}^{K} s^{BS}(i,k) \le 1,\ \sum_{k=1}^{K} s^{RS}(i,k) \le 1,\ \forall i \in \{1,2,...,n_f\} \tag{2.15}$$

$$2\lambda(k) \le r^{BS}(k), 2\lambda(k) \le r^{RS}(k),\ \forall k \in \{1,2,..,K\} \tag{2.16}$$

$$s^{BS}(i,k), s^{RS}(i,k) \in \{0,1\}, \forall k \in \{1,2,...,K\},\ \forall i \in \{1,2,...,n_f\} \tag{2.17}$$

where Eq. (2.15) implies the exclusive subcarrier assignment condition per BS-RS and RS-UE link. Moreover, the combination of conditions Eq.(2.15) and Eq.(2.17) ensure that every subcarrier is assigned exclusively to one user at each mini time-slot. The constraint Eq.(2.16) enforces the stability control requirement defined in Eq.(2.9) and Eq.(2.12). The proposed resource allocation method is fair in the sense that no user's packets are allowed to build up in the corresponding queues either in the base station or relay station (Viswanathan & Mukherjee, 2005).

We can transform the optimization problem Eq.(2.14) into its epigraph form by introducing the auxiliary parameter $x(k)$ (Boyd & Vandenberghe, 2004). In this regard, the objective function Eq.(2.14) and constraint Eq.(2.16) are respectively reformulated to

$$\max_{s^{BS}, s^{RS}, x(k)} \sum_{k=1}^{K} \omega(k) x(k) \tag{2.18}$$

$$2\lambda(k) \le x(k), \forall k \in \{1,2,..,K\} \tag{2.19}$$

The epigraph transformation requires two additional constraints, given by

$$x(k) \le \sum_{i=1}^{n_f} s^{BS}(i,k) C_m^{BS}(i,k) \tag{2.20}$$

$$x(k) \le \sum_{i=1}^{n_f} s^{RS}(i,k) C_m^{RS}(i,k) \tag{2.21}$$

,

which represent the hypograph of the original problem.

Since condition Eq.(2.17) enforces the allocation decision parameters to be binary variables, the total number of binary allocation parameters is equal to $2Kn_f$. This transforms the problem into a combinatorial optimization model, that is NP-Complete and is not possible to solve in polynomial time (Korte & Vygen, 2009). To resolve this issue, in the next section we propose a method by which we evaluate the solution of the allocation problem when the subcarriers are time-shared. Furthermore, we define a post processing mechanism, called Anti-relaxation approach that converts the resulting time-shared subcarrier allocation values into exclusive and binary parameters. In addition, a heuristic algorithm is proposed to further reduce the computation time of binary subcarrier allocation mechanism which is a critical issue especially for real-time applications.

2.3 Time-shared subcarrier allocation

It has been shown that if we relax the binary constraint to be the real values between zero and one, i.e.

$$s^{BS}(i,k), s^{RS}(i,k) \in [0,1], \tag{2.22}$$

then the solution of the relaxed problem yields an upper bound to the rates achieved by the binary problem. This upper bound converges to the optimal value if the number of subcarriers is sufficiently large (Yu & Lui, 2006). With this assumption, the real valued subcarrier allocation parameter $s^{BS}(i,k)$ defines the proportion of first mini time-slot that *i*-th subcarrier is assigned to *k*-th user on the BS-RS link. Similarly, $s^{RS}(i,k)$ defines the fraction of time that *i*-th subcarrier is assigned to *k*-th user on the RS-UE link. In other words, each subcarrier is assigned to each user in a fraction of one mini time slot.

Consequently, the problem Eq.(2.18) and its constraints Eq.(2.15), Eq. (2.19)-Eq.(2.21) are converted into a new problem given by

$$\max_{s^{BS},\, s^{RS},\, x(k)} \sum_{k=1}^{K} \omega(k)\, x(k) \quad (2.23)$$

,

subject to

$$\sum_{k=1}^{K} s^{BS}(i,k) \le 1,\ \sum_{k=1}^{K} s^{RS}(i,k) \le 1,\ \forall i \in \{1,2,...,n_f\}$$

,

$$2\lambda(k) \le x(k), \forall k \in \{1,2,..,K\}$$

,

$$x(k) \le \sum_{i=1}^{n_f} s^{BS}(i,k)\, C_m^{BS}(i,k)$$

,

$$x(k) \le \sum_{i=1}^{n_f} s^{RS}(i,k)\, C_m^{RS}(i,k)$$

,

$$s^{BS}(i,k), s^{RS}(i,k) \in [0,1], \forall k \in \{1,2,...,K\},\ \forall i \in \{1,2,...,n_f\}$$

.

Since the problem in Eq. (2.23) is now convex and linear, the optimal value can be calculated in polynomial time.

2.4 Exclusive subcarrier allocation

The proposed Time-shared approach results in an optimal solution when the subcarriers are allowed to be shared in one time-slot. However, the time-sharing assumption is not feasible for some network configurations when the time synchronization imposes unwanted communication overhead. In this section, we introduce two approaches for this situation.

2.4.1 Anti-relaxation approach (ARA)

To achieve a set of binary schedules, one can convert the continuous allocation parameters generated by the Time-shared formulation Eq.(2.23) into binary values. We call this procedure anti-relaxation approach (ARA), which constructs a set of potentially sub-optimal binary allocation parameters from the continuous ones resulted from performing the Time-shared method. In this regard, for each subcarrier $i \in \{1,2,...,n_f\}$, we set the largest allocation parameter amongst all users $k \in \{1,2,...,K\}$ to 1 and the rest to zero, i.e.

$$\begin{aligned} &\tilde{s}^u\left(i,\tilde{k}\right)=1, \quad \tilde{k}=\arg\max_k\left(s^u\left(i,k\right)\right) \\ &\tilde{s}^u\left(i,k\right)=0, \quad k\neq\tilde{k} \end{aligned} \tag{2.24}$$

where $u \in \{RS, BS\}$. After applying Eq.(2.24), the resulting allocation parameters satisfy the constraints Eq.(2.15), Eq.(2.17). In the numerical results section we will show that ARA method provides a sub-optimal solution on the sum rate gained by solving the Time-shared approach.

2.4.2 Two-Phase Heuristic Approach (TPHA)

We introduce a heuristic binary allocation method that invokes less computation complexity in comparison to ARA. The proposed heuristic algorithm is based on greedy methods and unlike general greedy algorithms that start from scratch; it computes an initial solution to speed up the computation process (Nemhauser & Wolsey, 1988).

As indicated by its name, the proposed mechanism is completed in two phases. The first phase is the initialization step where each subcarrier *i* is assigned to the user k^* with the best channel condition targeting to maximize system throughput. This policy can be formulated as

$$\begin{cases} s^{BS}\left(i,k^*\right)=s^{RS}\left(i,k^*\right)=1, \; k^*=\arg\max_k\left(\min\left(C_m^{BS}\left(i,k\right),C_m^{RS}\left(i,k\right)\right)\right) \\ s^{BS}\left(i,k^*\right)=s^{RS}\left(i,k^*\right)=0, \qquad \forall k\in\{1,2,..,K\}, k\neq k^* \end{cases}. \tag{2.25}$$

Next, we adjust the resulted binary allocation values in order to meet the stability conditions Eq.(2.9) and Eq.(2.12). Therefore, the second phase of the algorithm retains a part of bandwidth that has been allocated in phase one, so that any unsatisfied user $k^{\dagger}$ can meet stability conditions Eq.(2.9) and Eq.(2.12). Particularly, $k^{\dagger}$ is an unsatisfied user if its rate either on BS-RS or RS-UE link, cannot meet the stability conditions Eq.(2.9) and Eq.(2.12), i.e. $r^{u}(k^{*}) < 2\lambda(k^{*}), u \in \{BS, RS\}$. This method yields a constrained suboptimum solution with a significantly reduced complexity since the initial solution results from phase one and it is an unconstrained optimal Max-rate method.

Retaining the subcarriers which have been previously assigned in phase one and reallocating them to the unsatisfied users causes inevitable throughput reduction. Inspired by (Y. J. Zhang & Ben Letaief, 2004), we apply a cost function to minimize the degradation of system throughput. The cost calculation function $\Psi^{\mathrm{u}}\left(\mathrm{k}^{\dagger}, \mathrm{i}\right),\ \mathrm{u} \in \{\mathrm{BS}, \mathrm{RS}\}$ that is given in

$$\Psi^{u}\left(k^{\dagger}, i\right) = \frac{C_m^u(i,k^{*}) - C_m^u(i,k^{\dagger})}{C_m^u(i,k^{\dagger})}, \forall i \in \{1,2,...,n_f\}, s^{u}\left(i,k^{*}\right) = 1, u \in \{BS, RS\}, \tag{2.26}$$

estimates the cost of retaining any subcarrier from its initial owner k^{*} assigned in the phase one. The cost value is proportional to the decrease in system throughput and inversely proportional to the increase of unsatisfied user's data rate. Using Eq.(2.26), a vector of costs $\Phi^{u}(k^{\dagger}), u \in \{BS, RS\}$ is defined for each unsatisfied user $k^{\dagger}$. Then, the algorithm searches for the subcarrier $i^{\dagger}$ with the lowest cost in the cost vector of user $k^{\dagger}$. If it does not violate the minimum required rate, i.e. $r^{u}(k^{*}) - C_m^u(i,k^{*}) \geq 2\lambda(k^{*}), u \in \{BS, RS\}$, the subcarrier $i^{\dagger}$ will be retained from original owner k^{*} and reassigned to the unsatisfied user $k^{\dagger}$.

If the subcarrier reassignment violates the required rate of the original user, i.e. $r^{u}(k^{*}) - C_m^u(i,k^{*}) < 2\lambda(k^{*}), u \in \{BS, RS\}$, then the algorithm excludes that subcarrier and continues to search for another subcarrier with the lowest cost. The second phase is repeated for each unsatisfied user which in turn increases the number of satisfied users monotonically. This mechanism will continue until either all users meet the stability constraints or the available

subcarriers exhausted. In this way, the number of reallocation operations is finite and divergence is avoided.

One of the computation efficiency benefits of TPHA is that it only needs one constraint in each search stage. In addition, using the unconstrained optimal solution in the first phase as the initial allocation values, contributes to reducing the computation time. The number of required comparison operations in the first phase is $2Kn_f$. The number of required operations for cost calculation in the second phase is in the order of $2n_f(K-1)$. Furthermore, the maximum number of operations required for subcarrier swapping is ${n_f}^2$. Therefore, the computation complexity of the heuristic algorithm is in the order of $O\left(2Kn_f + {n_f}^2\right)$.

2.4.3 Computation complexity of the binary allocation mechanisms

To compare the computation efficiency of TPHA, we calculate the complexity of the ARA and two other binary allocation schemes; Exhaustive Search and Branch-and-Bound. The computation of ARA includes one-time execution of the Time-shared method plus rounding mechanism that need $2Kn_f$ comparison operations. Since the Time-shared method is convex, its computation complexity can be estimated by assuming an Interior Point Method (IPM) as a solver. We chose IPM because its complexity is readily measurable and is proportional to $z^{3/2}$, where z is the number of optimization constraints. The IPM complexity approximation requires that the log-barrier function, that is a function composed of the objective function and all of the constraints, be self-concordant (Nesterov, Nemirovskii, & Ye, 1994). Fortunately, Eq.(2.23) is a self-concordant problem[2]. Considering the number of constraints in problem Eq.(2.23), that is equal $4K + 4Kn_f + 2n_f$, the IPM complexity is by the order of

[2] More details on self-concordant property can be found in APPENDIX VI.

$O\left(\sqrt{\left(K+Kn_f+n_f\right)^3}\right)$. Therefore, the complexity of ARA is in the order of $O\left(2Kn_f+\sqrt{\left(K+Kn_f+n_f\right)^3}\right)$.

The exhaustive search method can find the global optimum solution with the complexity order of$\mathrm{O}\left(K^{2n_f}\right)$, where K^{2n_f} is the number of all possible allocation decisions. Branch-and-Bound is a popular method to find local optimal solution for problems that can be categorized under Integer Linear Programming (Sierksma, Dam, & Tijssen, 1996). Regarding to $2Kn_f$ integer variable in our problem, such a Branch-and-Bound scheme needs to solve 2^{2Kn_f} linear programming subproblems. The number of required iterations to solve one linear programming subproblem is approximately $2\left(x+y\right)$ where x is the number of variables, and y is the number of constraints. Furthermore, every iteration encompasses $\left(xy-y\right)$ multiplications, $\left(xy-y\right)$ summations, and $\left(x-y\right)$ comparison operations. For the case of our problem (14-17), we have $2\left(2\left(n_f+K+Kn_f\right)+2Kn_f\right)$ iterations considering $2Kn_f$ variables and $2\left(n_f+K+Kn_f\right)$ constraints. Therefore, LIP method can solve our problem by exponential complexity that is in the order of $O\left(2^{kn_f}\left(32kn_f\left(kn_f+K+n_f\right)^2\right)\right)$.

Table 2.1 Complexity comparison of binary subcarrier allocation schemes in a single cell model

Method	**Exhaustive search**	**Branch-and-Bound**	**TPHA**	**ARA**
complexity	$\mathrm{O}\left(K^{2n_f}\right)$	$\mathrm{O}\left(2^{kn_f}\left(32kn_f\left(kn_f+K+n_f\right)^2\right)\right)$	$\mathrm{O}\left(2Kn_f+{n_f}^2\right)$	$\mathrm{O}\left(2Kn_f+\sqrt{\left(K+Kn_f+n_f\right)^3}\right)$

Table 2.1 shows the complexity comparison of the suggested binary allocation methods, and it can be seen that TPHA and ARA have considerably lower complexities than LIP or

Exhaustive search. As a numerical example for comparing TPHA and ATA, we assume 128 subcarriers are available on each link and the number of users is 30. In this regard, using the formulation listed in Table 2.1, the computation complexity of TPHA and ARA can be approximated to 24000 and 260000, respectively. Accordingly, TPHA is more than 10 times computationally efficient than ARA. Regarding to the fact that the computation time is not always the desired performance metric for choosing an efficient allocation algorithm, therefore in the next section we compare the other performance qualities of the proposed methods.

2.5 Numerical results and discussion

We defined a network consisting of a single cell and three relay stations fixed at a distance of 15 Km from the base station. The relay-aided users are randomly placed in a distance of 220 m from their corresponding relays, as depicted in Figure 2.1. There are 64 subcarriers available in the downlink subframe. Each subcarrier has 15 KHz bandwidth and the carrier center frequency is 2.5 GHz. The BS-RS and RS-UE links both follow the suggested 3GPP path loss model, i.e urban macrocell in LoS and urban microcell in NLoS, respectively (*Universal Mobile Telecommunications System (UMTS); Spacial channel model for Multiple Input Multiple Output (MIMO) simulations*, n.d.). Each RS-UE link, that is NLoS, suffers multipath Rayleigh fading that is modeled with i.i.d random variables of complex Gaussian distribution with zero mean and unit variance, i.e. $\mathcal{CN}(0,1)$. The LoS BS-RS link experiences Rician fading with Rician factor $\kappa = 6dB$ that can be modeled by i.i.d random variable of complex Gaussian distribution where $\mathcal{CN}\left(\sqrt{\kappa/1+\kappa}, 1/1+\kappa\right)$. The shadowing coefficients are i.i.d random variable modeled by complex Gaussian distribution with zero mean and variance of $\sigma^2_{\gamma,BS} = 3dB$ and $\sigma^2_{\gamma,RS} = 8dB$, respectively for BS-RS and RS-UE links. The noise power density equals -174 dBm/Hz. The total transmission power of BS and RS equals 20W and 10W, respectively. The transmit power per subcarrier in BS and RS is fixed to the ratio of total power to the number of subcarriers. The length of a time slot is 5 ms. The entering data traffic at the base station can be modeled by a bursty On/Off traffic model (Lakani, Gagnon,

& Groleau, 2015). However, the length of On-time (normally 2 s) is far larger than the time slot length (5 ms), which resembles a constant bit rate (CBR) traffic during one time slot. For this reason, we consider CBR traffic arrival rate in the current work. To avoid complexity, all incoming traffic streams are assumed to have identical mean arrival rate and the users' weight parameters are also amounted to unity i.e. $\omega(k)=1, k\in\{1,2,...,K\}$. We apply MOSEK solver (MOSEK Apps., 2014) in CVX package (M. Grant & S. Boyd, 2014) for solving the Time-shared optimization problem.

Because the channel samples are drawn from probabilistic distributions, we demonstrated the numerical results by averaging the performance values over 100 independent realization of the simulated system. During various simulation examinations, we verified that 100 realization is enough to prevent fluctuations in the performance trends.

2.5.1 System stability

In this section we investigate the capability of stabilizing system queues among the proposed methods. For presentation purposes, the length of queues at the base station is summed with those of relay stations, and the queue lengths are recorded during 200 consecutive time-slots. The results presented in this section are the arithmetic average queue lengths of 15 users served by each of three relay stations. For comparison purposes, we first present the average queue length in Figure 2.2 when the arrival rate is low (i.e. 20Kbps for the considered system configuration) so that all methods can stabilize the system. The Time-shared and ARA methods return almost steady queue length around $2\lambda T_s = 200$ bits. It can be seen that TPHA is abruptly fluctuating, although it is almost upper bounded to 1000 bits.

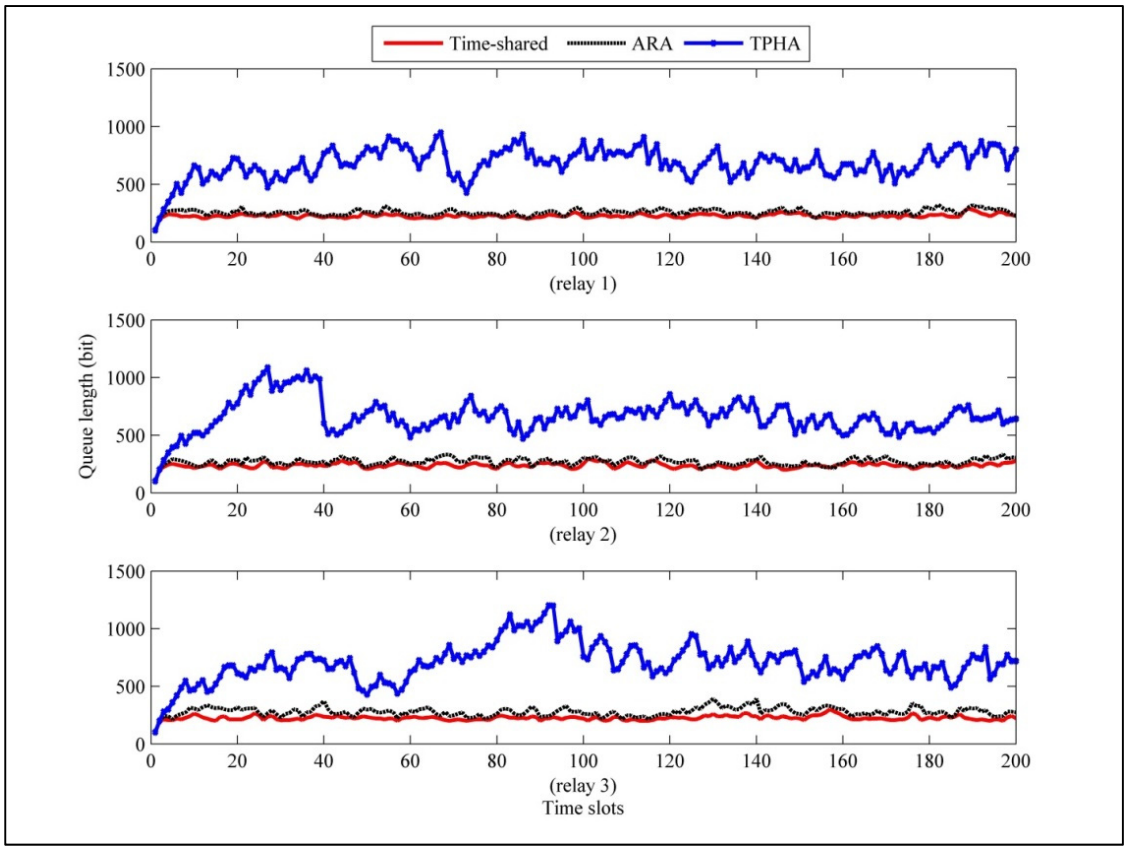

Figure 2.2 Average queue length per relay, subscripted by (relay 1), (relay 2), (relay 3) the arrival rate 20 Kbps, the number of users K=45

In Figure 2.3, we present the results of worst case scenario i.e. the largest arrival rate at which the Time-shared method can stabilize the system (45Kbps). It can be seen in Figure 2.3 that the queue length resulted from Time-shared and ARA methods is upper bounded to 1000 bits. On the other hand, TPHA tries to control user queues in relay 1 and 2; however due to its heuristic nature it fails to guarantee stability for all of the users as can be seen for users of relay 3. Overall, the Time-shared, ARA, and TPHA methods are successful to stabilize the queues in low traffic load conditions; however the time-shared and ARA methods are the more reliable stabilizing approaches in heavy traffic scenarios.

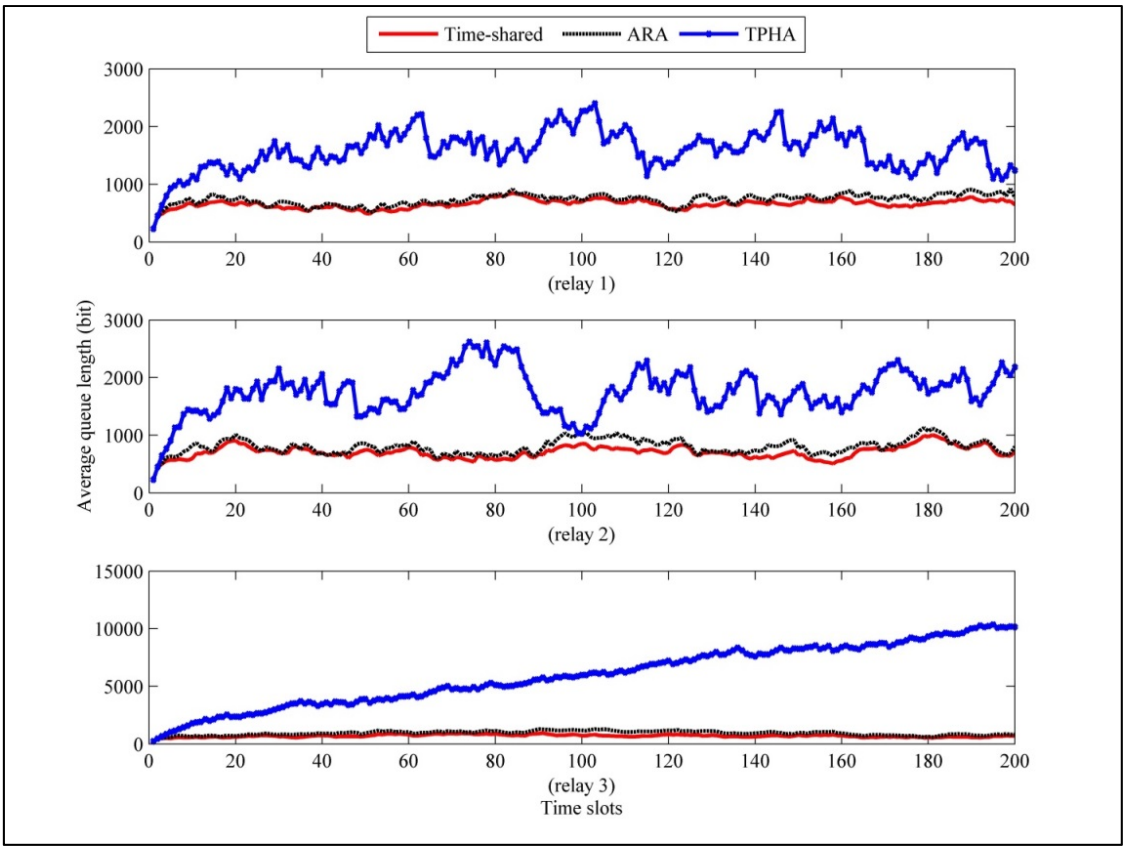

Figure 2.3 Average queue length per relay, subscripted by (relay 1), (relay 2), (relay 3), the arrival rate 45 Kbps, number of users K=45

2.5.2 System throughput

In this section we compare the achieved sum system throughputs, calculated by Eq.(2.14) upon changing the arrival rate and various number of users in the system. In Figure 2.4, we illustrate the system throughput with small number of users in the system, i.e. K=12. It is expected and observed that the illustrated throughputs in Figure 2.4 are constant and independent from the variation of arrival rate. It can be seen that the two binary allocation methods, ARA and TPHA, are respectively the lower-bound and the upper-bound to the results of Time-shared method. The aforementioned bounds are in ±0.3% proximity of the achievable throughput of the Time-shared method. In particular, when the system is lightly loaded, the suggested binary allocation methods are good estimates of the optimal values.

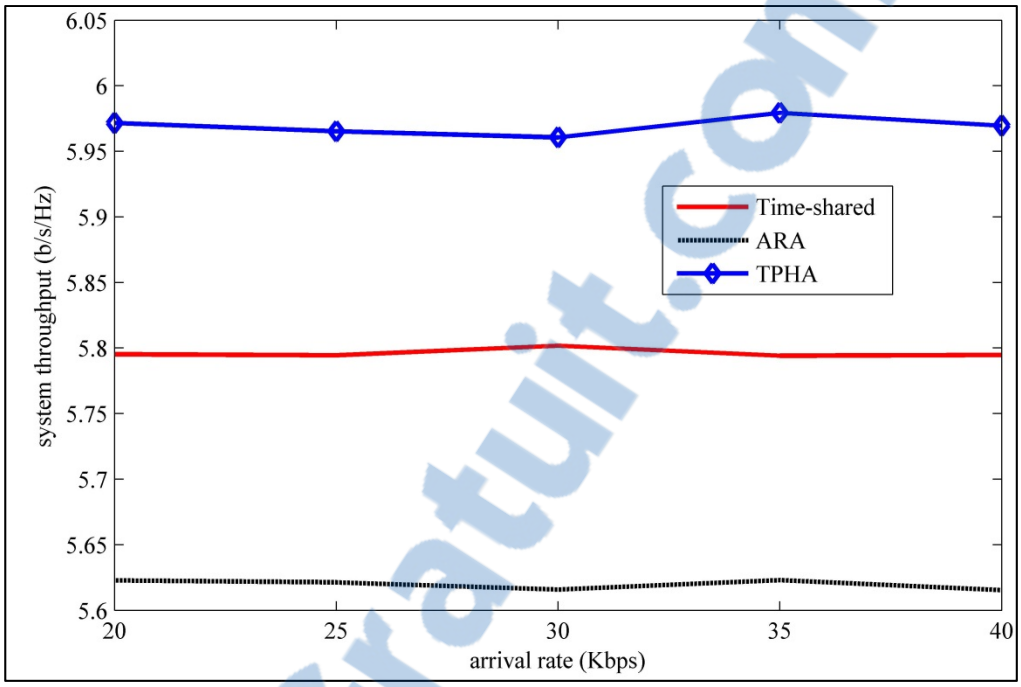

Figure 2.4 Achievable system throughput , applying Time-shared, ARA and TPHA methods, the number of total users 12

We also examine the performance of the proposed methods in a heavy loaded scenario. In Figure 2.5, we illustrate the system throughput achieved by applying Time-shared, ARA and TPHA methods when number of users is K=45. It can be seen that the proximity of two binary methods to the Time-shared method are noticeably different from the lightly loaded scenario in Figure 2.4. In particular, the throughput of TPHA degrades to 50% less than the Time-shared results. The gap between the achieved throughput of ARA and Time-shared is also increased to 18%. Overall, it can be noted that TPHA loses its efficiency when the traffic and number of users grow large. We have shown that the computation complexity of TPHA is considerably smaller than ARA, however the results presented in Figure 2.5 points out the advantage of ARA in heavy traffic scenario rather than TPHA.

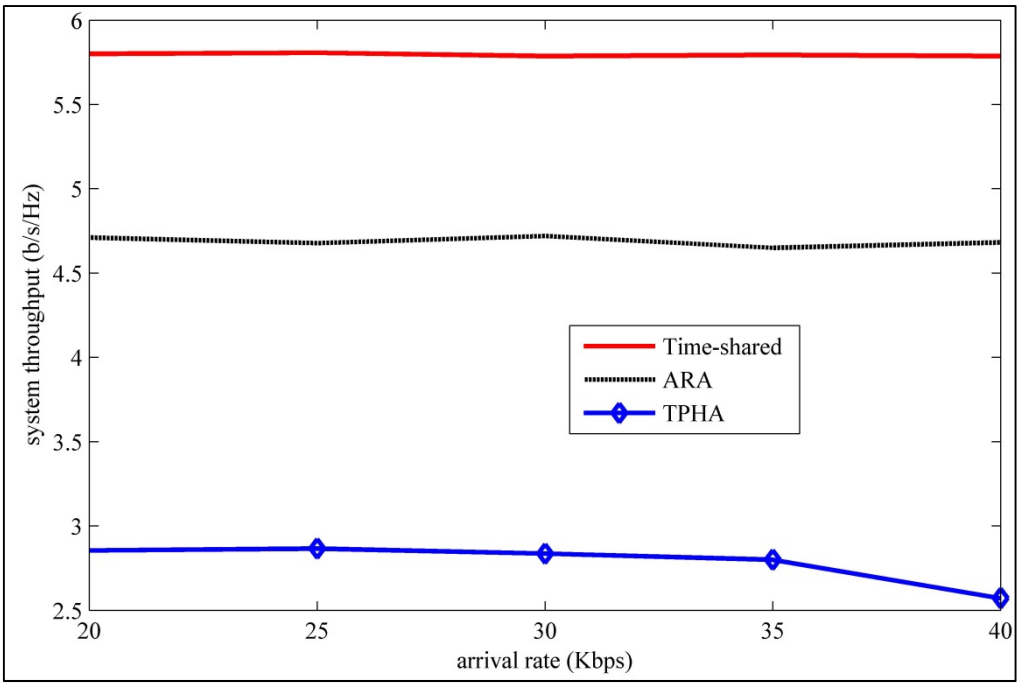

Figure 2.5 Achievable system throughput applying Time-shared, ARA and TPHA methods, total users 45

2.6 Conclusion

In this chapter, we defined the problem of downlink resource allocation to maximize the throughput of relay-aided users. We inspected the stability requirement of queues in tandem which refers to the buffers placed in the base station and relay stations. Then, we proposed stability constraints to integrate with the resource allocation problem. This mechanism controls the subcarrier allocation in regard to the traffic arrival rates to prevent instability in the corresponding queues. To solve the resource allocation problem, we employed an optimal method called Time-shared, that allows users to share subcarriers in BS-RS or RS-UE link, in a proportion of one mini-time-slot. For the case that network design limitations prohibit Time-shared approach, we proposed two heuristic methods, called ARA and TPHA that allocate subcarriers exclusively to each user. The ARA method is a post-processing procedure on the results of Time-shared method. TPHA is a greedy-based heuristic method with lower computational complexity which can be advantageous for delay sensitive functionalities. Using simulation scenarios, it is shown that the proposed stability constraints

can efficiently control the queue length in various network scenarios using Time-shared method. As a low complexity approach, TPHA succeeds stabilizing the system and achieving close to optimal throughput when the considered network is not heavily loaded. It is also shown that, if the computation complexity is not a concern, ARA method can exclusively allocate subcarriers and gain close to optimal results in various network scenarios.

In this chapter we defined stability constraints with respect to traffic arrival rates. In the following chapters we introduce resource allocation mechanisms for the case that the information about arrival rates is not known.

CHAPTER 3

CHANNEL- AND QUEUE-AWARE RESOURCE ALLOCATION IN A MULTI-CELL RELAY-AIDED SYSTEM CONSTRAINED TO STABILITY AND INTERFERENCE CONTROL

This chapter expands the topic of resource allocation in the downlink of an OFDMA decode-and-forward relaying, considering a multi-cell system. To optimize the transmission power, we include the inter-cell interference control and power allocation parameters in the optimization problem. In addition, we introduce a new stability analysis that does not require a priori knowledge of the arrival process. To further control the queuing delay and enhance system throughput, we propose an adjustable time-slot division mechanism.

The allocation objective is to maximize the overall throughput of cell-edge users served by relay stations. We circumvent the combinatorial nature of the introduced problem by a time-shared approach. The need for exclusive subcarrier allocation is then satisfied by a new optimal binary allocation solution whose computation complexity is polynomial in time and comparable to that of the Time-shared method. In addition, we introduce a novel binary allocation approach to reduce the system power consumption while respecting the stability and interference constraints. Geometric Programming and monomial approximation technique are employed to offer a computationally efficient solution to the non-convex problem.

3.1 System model

We assume a cellular system that includes a set of cells/base stations denoted by Φ, and a set of relays pertaining to the *c*-th base station denoted by R_c. The design goal is to assign a set of available subcarriers denoted by N (where its size equals $|\mathrm{N}| = n_f$) to the set of users $\mathrm{K}_{c,m}$ served by *m*-th relay in *c*-th cell. The aforementioned users are placed at the edge of cells and it is impossible for them to connect to the base station due to heavy blockage and long distance transmission. The users who do not suffer from dispersed and weak signals are

assumed to be connected directly to the base station via separate resources. In this regard, we assume that the resource allocation for the users close to the base station is done via independent procedures[3] .

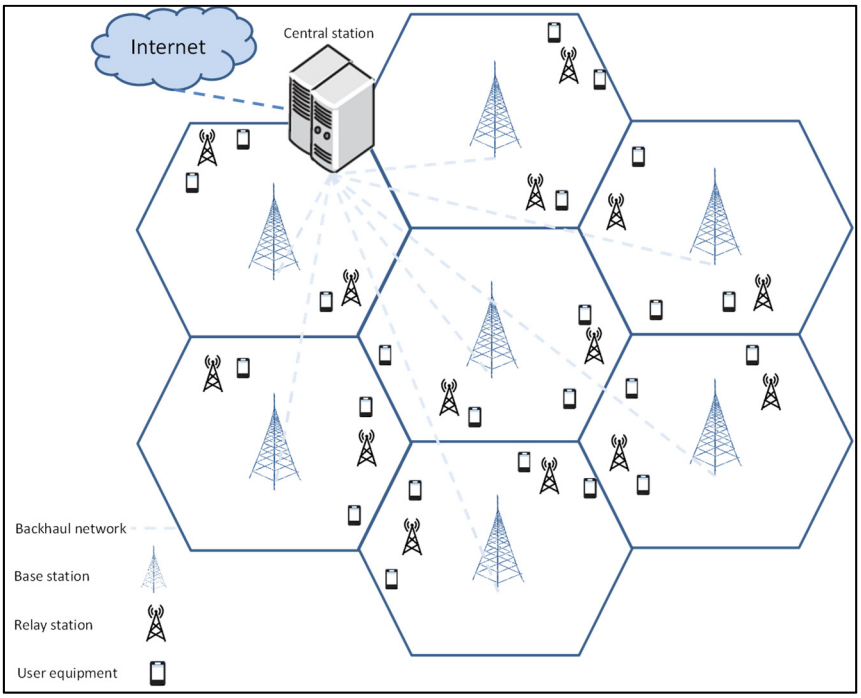

Figure 3.1 Schematic of a multicell relaying system controlled by a central station

We consider that each relay is served by only one base station, and each cell-edge user is served by only one relay. As a relay selection strategy, it is assumed that the users are connected to the relays with the strongest average channel. Base stations are connected to a central unit, similar to the schematic in Figure 3.1, with optical fiber backhaul links, to which the channel state information (CSI) of all links is fed back. This central unit performs the allocation procedure for the entire system. To benefit from the spectral efficiency advantage

[3] We note that a joint resource allocation for non-relay assisted and relay assisted users would lead to a better system performance but the increased computational complexity can degrade the system efficiency for practical scenarios.

of a full frequency reusing scenario, we assume that the set of n_f subcarriers are shared for transmission on both BS-RS and RS-UE links.

3.1.1 Adjustable time-slot partitioning

We consider that the downlink subframe is time-slotted and the length of one slot equals T_s. In order to prevent the intra-cell interference, the transmissions of base station and relay stations are separated into two mini time-slots. In particular, the BS-RS transmission is assumed to take place at the first mini time-slot of the downlink subframe and the RS-UE transmission at the second mini time-slot. Considering that the signal to interference plus noise ratio (SINR) on the Line-of-Sight type BS-RS link is different from the SINR of the Non-Line-of-Sight RS-UE link, an adjustable time-slot partitioning is beneficial in order to compromise the capacity difference. Therefore, we consider an adjustable time-slot division parameter $\alpha \in (0,1)$, so that αT_s and $(1-\alpha)T_s$ define the length of a mini time-slot assigned for the BS-RS and RS-UE transmissions, respectively.

3.1.2 Channel capacity

The channel state information from relay and user equipment (UE) is assumed to be fed back to the base station via control information, and therefore, the *c*-th base station can calculate the instantaneous capacity $C_{c,m}^{BS}(i,k,t)$ on the BS-RS link to its *m*-th relay destined to the *k*-th user, in the *i*-th subcarrier and during the *t*-th time slot, that is given by

$$C_{c,m}^{BS}(i,k,t) = \Delta f \log_2\left(1+\Lambda_{c,m}^{BS}(i,k,t)\right).$$

The variable Δf denotes the bandwidth of each subcarrier, that is equal to the total bandwidth divided by the number of subcarriers n_f. Due to multi-cell assumption, we need to reformulate the SINR received at relay station that is given by

$$\Lambda_{c,m}^{BS}(i,k,t)=\frac{P_{c,m}^{BS}(i,k,t)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2+\sigma_c^2(i,k,t)}.$$

The parameters in the above SINR formulation is similar to the one defined in CHAPTER 2, i.e. $\Lambda_m^{BS}(i,k,t)$, except the index c which denotes the serving BS and the inter-cell interference variance, i.e. $\sigma_c^2(i,k,t)$, in the denominator. The latter parameter corresponds to the potential interference imposed by neighbor BSs. We assume that because of the large cell dimension, the use of directional antennas and applying interference cancellation techniques at relay stations (Jeffrey G. Andrews, 2005), the inter-cell interference received by the relay station becomes negligible compared to the channel noise i.e. $\sigma_c^2(i,k,t)\ll\sigma_z^2$ (Ng & Schober, 2011). Therefore, in the remainder of this text we consider the approximated SNR at the *m*-th relay, given by

$$\Lambda_{c,m}^{BS}(i,k,t)\approx\frac{P_{c,m}^{BS}(i,k,t)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}.$$

The instantaneous channel capacity of RS-UE link is also similar to the Eq.(2.4), except for the extra index *c* pertaining to the serving base station and it is given by

$$C_{c,m}^{RS}\left(i,k,t\right)=\Delta f\log_2\left(1+\Lambda_{c,m}^{RS}\left(i,k,t\right)\right).$$

Due to multi-cell assumption, we need to reformulate the SINR received at *k*-th user $\Lambda_{c,m}^{RS}(i,k,t)$ as follows

$$\Lambda_{c,m}^{RS}\left(i,k,t\right)=\frac{p_{c,m}^{RS}(i,k,t)l_{c,m}^{RS}(k)\left|H_{c,m}^{RS}(i,k)\right|^2}{\sigma_z^2+\sigma_{c,m}^2(i,k,t)}. \tag{3.1}$$

In the latter SINR formulation, we take the co-channel interference from the relays in the neighboring cells into consideration and its variance is given by

$$\sigma_{c,m}^2(i,k,t)=\sum_{c'\in\Phi,c'\neq c}\sum_{m'\in R_{c'}}\sum_{k'\in K_{c',m'},k'\neq k}s_{c',m'}^{RS}(i,k',t)p_{c',m'}^{RS}(i,k',t)\left(l_{c',m'}^{RS}(k)\left|H_{c',m'}^{RS}(i,k)\right|^2\right).$$

where $s_{c',m'}^{RS}(i,k',t)$ is the binary subcarrier allocation indicator at the second mini-slot associated with the RS-UE link. Setting $s_{c',m'}^{RS}(i,k',t)=1$ implies that the *i*-th subcarrier is

assigned to transmit data destined to the k'-th user. $l^{RS}_{c',m'}(k)$ and $H^{RS}_{c',m'}(i,k)$ denote respectively the path loss and fading/shadowing coefficient of the i-th subcarrier on the interfering link between the k-th user and m'-th relay pertaining to the c'-th neighbor cell.
Similar to Eq.(2.5) and Eq.(2.6), we define the data transmission rate (bit per second), respectively in BS-RS link

$$r^{BS}_{c,m}(k,t) = \sum_{i \in N} s^{BS}_{c,m}(i,k,t) C^{BS}_{c,m}(i,k,t)$$

and RS-UE link

$$r^{RS}_{c,m}(k,t) = \sum_{i \in N} s^{RS}_{c,m}(i,k,t) C^{RS}_{c,m}(i,k,t) .$$

Similarly, $s^{BS}_{c,m}(i,k,t)$ and $s^{RS}_{c,m}(i,k,t)$ define the binary allocation parameters of the BS-RS link and RS-UE link, respectively.

3.1.3 Queue-aware stability control

It is worth to mention that, we assume an admission control policy has been performed at a higher level in order to explore the maximum admissible traffic, which accordingly rejects the excessive data flow (Tassiulas & Ephremides, 1992). Unlike the stability strategy defined in CHAPTER 1, we use the fundamental definition of stability in (Georgiadis et al., 2006). It is known that, a queueing system is stable if all of the individual queues in the system have a bounded time average backlog. This is mathematically defined as

$$\lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} E\left[Q(k,t)\right] < \infty, \forall k \in K$$

where $E[.]$ denotes the expected value.

To calculate the expected value of the queue length, we formulate the queue length evolution conditioned on the allocated service rate. In this regard, the queue length evolution at the base station associated with the k-th user data flow, during one time-slot, is given by

$$Q_{c,m}^{BS}(k,t+1)=\begin{cases}Q_{c,m}^{BS}(k,t)-r_{c,m}^{BS}(k,t)\alpha T_s+a_{BS}(k,t), & \text{if } Q_{c,m}^{BS}(k,t)>r_{c,m}^{BS}(k,t)\alpha T_s\\ a_{BS}(k,t), & \text{if } Q_{c,m}^{BS}(k,t)\le r_{c,m}^{BS}(k,t)\alpha T_s\end{cases} \quad (3.2)$$

where $a_{BS}(k,t)$ indicates the number of arrived bits during the *t*-th time slot. Using (3.2) the long-term average of the expected value of the queue length can be formulated as

$$\begin{aligned}&\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau}E\left[Q_{c,m}^{BS}(k,t+1)\right] \quad (3.3)\\ &=\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau}\left(\left(Q_{c,m}^{BS}(k,t)-r_{c,m}^{BS}(k,t)\alpha T_s+a_{BS}(k,t)\right)\left(1-P_{e1}(k,t)\right)\right)+a_{BS}(k,t)P_{e1}(k,t)\end{aligned}$$

where $P_{e1}(k,t)$ equals $Pr\left[Q_{c,m}^{BS}(k,t)\le r_{c,m}^{BS}(k,t)\alpha T_s\right]$ which implies the probability that the queue will be emptied by using the assigned departure rate in the *t*-th time slot. If $P_{e1}(k,t)=1$, the updated queue backlog $Q_{c,m}^{BS}(k,t+1)$ would be limited to the arrived bits during the last time slot. In this respect, in *Proposition 1*, we introduce a stability condition that does not require a priori knowledge of stationary probability of channel state or arrival process.

Proposition 1. Assigning the BS-RS rate in each allocation time slot such that

$$\frac{1}{\alpha T_s}Q_{c,m}^{BS}(k,t)\le r_{c,m}^{BS}(k,t), \quad (3.4)$$

is sufficient to stabilize the *k*-th queue in the base station.

Proof. We assume the condition in Eq.(3.4) is enforced for each time slot, i.e. $P_{e1}(k,t)=1$. Therefore Eq.(3.3) can be rewritten as

$$\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau}E\left[Q_{c,m}^{BS}(k,t+1)\right]=\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau}\left(a_{BS}(k,t)\right).$$

The number of arrived bits in each time slot, from the source to the base station $a_{BS}(k,t)$ equals the data arrival rate $\lambda(k,t)$ multiplied by the duration of time slot T_s. Considering the

fact that every traffic flow, destined to the *k*-th user, has a definitive average arrival rate $\overline{\lambda}(k)$ [4], the long-term average queue length is finite and equal to

$$\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau}E\left[Q_{c,m}^{BS}(k,t+1)\right]=\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau}T_s\lambda(k,t)=T_s\overline{\lambda}(k) \tag{3.5}$$

which satisfies the sufficient condition for stability of queues at the base station. □

Remark 1. The Little's law defines that the average delay of a queue equals the ratio of the average queue length to the average arrival rate (Leon-Garcia, 2008). By satisfying conditions defined in Eq.(3.4) and Eq.(3.5) and considering Little's law, the average delay in the BS queues is bounded to T_s.

Next, we apply a similar method for analyzing the queues placed in the relay stations conditioned on the allocated service rate $r_{c,m}^{RS}(k,t)$. The queue length of *m*-th relay at the beginning of $(t+1)$-th time slot is given by

$$Q_{c,m}^{RS}(k,t+1)=\begin{cases}Q_{c,m}^{RS}(k,t)-r_{c,m}^{RS}(k,t)(1-\alpha)T_s+\min\left(Q_{c,m}^{BS}(k,t),r_{c,m}^{BS}(k,t)\alpha T_s\right), & \text{if } Q_{c,m}^{RS}(k,t)>r_{c,m}^{RS}(k,t)(1-\alpha)T_s\\ min\left(Q_{c,m}^{BS}(k,t),r_{c,m}^{BS}(k,t)\alpha T_s\right), & \text{if } Q_{c,m}^{RS}(k,t)\le r_{c,m}^{RS}(k,t)(1-\alpha)T_s\end{cases}$$

where $min\left(Q_{c,m}^{BS}(k,t),r_{c,m}^{BS}(k,t)T_s\alpha\right)$ is the actual number of bits already departed from the base station and destined to the *k*-th queue at the relay station. Accordingly, we can write the long-term average of the expected value of the queue length at the *m*-th relay as

$$\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau}E\left[Q_{c,m}^{RS}(k,t+1)\right]=\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau}\Big(\left(Q_{c,m}^{RS}(k,t)-r_{c,m}^{RS}(k,t)(1-\alpha)T_s+min\left(Q_{c,m}^{BS}(k,t),r_{c,m}^{BS}(k,t)T_s\alpha\right)\right)\left(1-P_{e2}(k,t)\right)+$$
$$min\left(Q_{c,m}^{BS}(k,t),r_{c,m}^{BS}(k,t)T_s\alpha\right)P_{e2}(k,t)\Big)$$

where the probability term $P_{e2}(k,t)$ implies $Pr\left[Q_{c,m}^{RS}(k,t)\le r_{c,m}^{RS}(k,t)(1-\alpha)T_s\right]$. In *Proposition 2*, we define the stability condition for queues resided in the relay station.

[4] In the case that the traffic is received from the source generating application, the arrival rate is bounded by the capacity of data generation rate. For further readings you can find several statistically modeled traffic types in (Lakani, Gagnon, & Groleau, 2015). In the case that data is forwarded via another station, the data rate is bounded by the capacity of the forwarding link.

Proposition 2. Assigning the rate in RS-UE link in each allocation time slot such that

$$\frac{1}{(1-\alpha)T_s} Q_{c,m}^{RS}(k,t) \le r_{c,m}^{RS}(k,t) \quad (3.6)$$

is sufficient to stabilize the *k*-th queue in the relay station.

Proof. Assume that Eq.(3.6) is enforced at each time-slot ; it means that $P_{e2} = 1$. So the long-term average of queue length at *m*-th relay is given by $\lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=0}^{\tau} min\left(Q_{c,m}^{BS}(k,t), r_{c,m}^{BS}(k,t) T_s \alpha\right)$. Next we study the various outcomes of the minimum term in the latter formulation. First consider case (a) in which we assume that $min\left(Q_{c,m}^{BS}(k,t), r_{c,m}^{BS}(k,t) T_s \alpha\right) = Q_{c,m}^{BS}(k,t)$. It can be readily seen that case (a) satisfies Eq.(3.4), which leads to the conclusion that the average queue length of the *m*-th relay is bounded to the average traffic arrival rate at the base station. This is given by

$$\lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=0}^{\tau} E\left[Q_{c,m}^{RS}(k,t+1)\right] = T_s \bar{\lambda}(k).$$

Now consider case (b), that is $min\left(Q_{c,m}^{BS}(k,t), r_{c,m}^{BS}(k,t) T_s \alpha\right) = r_{c,m}^{BS}(k,t) T_s \alpha$. In this case, the long-term average queue length is also bounded, because the achievable rate of each link is bounded to a set of finite rate levels, i.e. $\left\{r^{-BS}\right\}$, due to the practical power constraints and limited code design. Therefore, we can conclude that the condition in Eq.(3.6) is sufficient for the stability of queues at the relays. □

Remark 2. By satisfying condition Eq.(3.6) and using Little's law, the queuing delay in the relay stations can be calculated as given in

$$\frac{E\left[Q_{c,m}^{RS}(k)\right]}{\mathrm{E}\left[min\left(Q_{c,m}^{BS}(k,t), r_{c,m}^{BS}(k,t) T_s \alpha\right)\right] / T_s \alpha} = T_s \alpha$$

.

With the suggested stability conditions Eq.(3.4) and Eq.(3.6) , not only the queue length of a data flow destined to the *k*-th user is bounded in each time-slot, but also the aggregated long-

term and short-term queuing delay from BS to UE, is bounded to $T_s(1+\alpha)$ with respect to *Remark 1* and *Remark 2*. This shows the importance of finding an optimal value for α that decreases the delay even more than fixed ratio time-slotting.

The stability conditions proposed is deployed in a slot-per-slot allocation mechanism. This is advantageous for certain real-time wireless multimedia or video streaming applications, where a packet can be discarded in each time window of transmission, if it has been received later than an expected delay. In contrast with the stability analysis provided in (Jamali, Zlatanov, & Schober, 2015; Neely et al., 2005; Tassiulas & Ephremides, 1992), our analysis of stability neither requires the complex Markovian modeling on channel states and/or arrival processes, nor requires a priori knowledge of the statistics of the arrival traffic.

3.2 Defining the optimization problem

Similar to the approach in 2.3, in this section we define an optimization problem to allocate OFDMA subcarriers on BS-RS and RS-UE links. We extend the allocation problem by including the transmission power per subcarrier, denoted by $p_{c,m}^{BS}(i,k)$ and $p_{c,m}^{RS}(i,k)$, in order to optimize energy consumption of the base station and relay stations, respectively. In addition, the multi-cell model requires considering inter-cell interference, which also can be handled by optimal power and subcarrier allocation scheme. To this end, the optimization objective is to maximize the throughput of users in all cells for each time slot. We define the users instantaneous throughput $\Pi_{c,m}(k,t)$ by the minimum of the transmitted bits in each of two links as follows:

$$\Pi_{c,m}(k,t) = \min\left\{T_s \alpha r_{c,m}^{BS}(k,t), T_s(1-\alpha) r_{c,m}^{RS}(k,t)\right\}.$$

Similar to CHAPTER 2, we assume that the channel state is invariant during one time slot, and therefore we can solve the channel-aware allocation problem once per time slot. Therefore, without loss of generality and for the sake of simplicity, we eliminate time slot

index *t* from the remaining equations. Consequently, the optimization problem can be defined as

$$\max_{\{s^{BS}\},\{p^{BS}\},\{s^{RS}\},\{p^{RS}\}} \sum_{c\in\Phi}\sum_{m\in \mathrm{R}_c}\sum_{k\in \mathrm{K}_{c,m}} \omega_{c,m}(k)\Pi_{c,m}(k) \quad , \tag{3.7}$$

where $\omega_{c,m}(k)$ denotes a positive weight parameter that can reflect the users scheduling priority and can also enforce fairness policies in order to meet the QoS requirements of different users.

The total transmit power of the base station and each relay station are limited to predefined maximum values $p_{BS}^{\max}$ and $p_{RS}^{\max}$, respectively. That leads to the set of power constraints that are denoted as:

$$p_{c,m}^{BS}(i,k)\geq 0,\ p_{c,m}^{RS}(i,k)\geq 0, \qquad \forall i\in N,\forall c\in\Phi,\forall m\in \mathrm{R}_c,\forall \mathrm{k}\in \mathrm{K}_{c,m}, \tag{3.8}$$

$$\sum_{m\in \mathrm{R}_c}\sum_{k\in \mathrm{K}_{c,m}}\sum_{i\in N} s_{c,m}^{BS}(i,k)p_{c,m}^{BS}(i,k)\leq p_{BS}^{max}, \qquad \forall c\in\Phi \quad , \tag{3.9}$$

$$\sum_{k\in \mathrm{K}_{c,m}}\sum_{i\in N} s_{c,m}^{RS}(i,k)\,p_{c,m}^{RS}(i,k)\leq p_{RS}^{max}, \qquad \forall c\in\Phi, m\in \mathrm{R}_c \quad . \tag{3.10}$$

The following three constraints are also necessary in order to assign subcarriers exclusively for one user per mini time-slot.

$$\sum_{m\in \mathrm{R}_c}\sum_{k\in \mathrm{K}_{c,m}} s_{c,m}^{BS}(i,k)\leq 1, \qquad \forall i\in N,c\in\Phi \quad , \tag{3.11}$$

$$\sum_{m\in \mathrm{R}_c}\sum_{k\in \mathrm{K}_{c,m}} s_{c,m}^{RS}(i,k)\leq 1, \qquad \forall i\in N,c\in\Phi \quad , \tag{3.12}$$

$$s_{c,m}^{BS}(i,k)\in\{0,1\}, s_{c,m}^{RS}(i,k)\in\{0,1\}, \qquad \forall i\in N,c\in\Phi,m\in \mathrm{R}_c,k\in \mathrm{K}_{c,m}. \tag{3.13}$$

To ensure the queue stability conditions (Eq.(3.5), Eq.(3.6)) are enforced, we assert the following two constraints

$$\frac{Q_{c,m}^{BS}(k)}{\alpha T_s} \le r_{c,m}^{BS}(k), \qquad \forall m \in \mathrm{R}_c, c \in \Phi, k \in \mathrm{K}_{c,m} \tag{3.14}$$

,

$$\frac{Q_{c,m}^{RS}(k)}{(1-\alpha) T_s} \le r_{c,m}^{RS}(k), \qquad \forall m \in \mathrm{R}_c, c \in \Phi, k \in \mathrm{K}_{c,m} \tag{3.15}$$

.

The minimum term in the objective function formulation can be handled by introducing an auxiliary variable $\rho_{c,m}(k) = \Pi_{c,m}(k) / T_s$, which transforms the problem into its epigraph form, given by

$$\max_{\{s^{BS}\},\{p^{BS}\},\{s^{RS}\},\{p^{RS}\},\{\rho\}} T_s \sum_{c \in \Phi} \sum_{m \in \mathrm{R}_c} \sum_{k \in \mathrm{K}_{c,m}} \omega_{c,m}(k) \rho_{c,m}(k) \tag{3.16}$$

,

subject to Eq.(3.8)-Eq.(3.15), along with two additional constraints given by

$$\rho_{c,m}(k) \le \alpha r_{c,m}^{BS}(k), \qquad \forall m \in \mathrm{R}_c, c \in \Phi, k \in \mathrm{K}_{c,m}, \tag{3.17}$$

$$\rho_{c,m}(k) \le (1-\alpha) r_{c,m}^{RS}(k), \qquad \forall m \in \mathrm{R}_c, c \in \Phi, k \in \mathrm{K}_{c,m}. \tag{3.18}$$

The new objective function Eq.(3.16) is convex but due to binary constraint Eq.(2.13) the problem becomes combinatorial and nonconvex, hence it is not feasible to find the global optimum solution in polynomial time. In the following sections, we study the possible methods that can return the computationally efficient optimal solution.

3.3 The time-shared solution in multi-cell network

In solving the problem, the BS-RS and RS-UE links are assumed under low SNR profiles; hence we cannot simplify the problem by approximating the capacity formula $\Delta f \log_2\left(1 + \Lambda_{c,m}(i,k,t)\right)$ to $\Delta f \log_2\left(\Lambda_{c,m}(i,k,t)\right)$. The proposed problem can be optimally solved using techniques like bisection search (Cendrillon, Yu, Moonen, Verlinden, & Bostoen, 2006) or ellipsoid/subgradient methods (Yu & Lui, 2006), however their exponential computation complexity makes it inefficient for practical multi-cell systems. In the following text, we apply some techniques to convert the problem Eq.(3.16) into a convex form, so that the solution can be attained in polynomial time.

First, similar to the approach in CHAPTER 2, we relax the binary constraint to be the real values between zero and one, i.e. $s_{c,m}^{BS}(i,k), s_{c,m}^{RS}(i,k) \in [0,1]$. The relaxed subcarrier allocation can be drawn as time sharing parameters that assign the *i*-th subcarrier to the *k*-th user during a proportion of the assigned time-slot.

The cause of non-convexity in Eq.(3.15) and Eq.(3.18) is the received interference $\sigma_{c,m}^2(i,k,t)$ embodied in the rate formula in RS-UE link. To circumvent this, we can replace the received interference by a constant value θ denoting the maximum tolerable interference per subcarrier, imposed by neighboring relays in the adjacent cells. Therefore, the new SINR formula is given by

$$\Lambda_{c,m}^{RS}(i,k) = \frac{p_{c,m}^{RS}(i,k) l_{c,m}^{RS}(k) \left| H_{c,m}^{RS}(i,k) \right|^2}{\sigma_z^2 + \theta}, \tag{3.19}$$

which replaces Eq.(3.1) in the rate formulation $r_{c,m}^{RS}(k)$. By tuning θ, the scheduler can control the level of tolerable interference in order to adjust system performance, which is further examined in the numerical results section. By replacing the actual received interference $\sigma_{c,m}^2(i,k,t)$ by tolerance value θ, an extra inequality condition $\sigma_{c,m}^2(i,k) \le \theta$ is required to be added to the set of optimization constraints.

The other step to combat non-convexity is to apply a variable mapping, given by

$$\hat{p}_{c,m}^{BS}(i,k) = s_{c,m}^{BS}(i,k) p_{c,m}^{BS}(i,k),$$

$$\hat{p}_{c,m}^{RS}(i,k) = s_{c,m}^{RS}(i,k) p_{c,m}^{RS}(i,k).$$

We call the new parameters $\left\{\hat{p}_{c,m}^{BS}(i,k)\right\}$ and $\left\{\hat{p}_{c,m}^{RS}(i,k)\right\}$ as the actual transmit power per subcarrier for *k*-th dataflow at the base station and at the *m*-th relay station, respectively. The new variables implicitly enforce zero power allocation when its corresponding subcarrier allocation parameter is zero. Consequently, the rate formulas need to be rewritten as follows

$$\hat{r}_{c,m}^{BS}(k) = r_{c,m}^{BS}(k)\Big|_{p_{c,m}^{BS}(i,k)=\hat{p}_{c,m}^{BS}(i,k)s_{c,m}^{BS}(i,k)}, \tag{3.20}$$

$$\hat{r}_{c,m}^{RS}(k) = r_{c,m}^{RS}(k)\Big|_{p_{c,m}^{RS}(i,k)=\hat{p}_{c,m}^{RS}(i,k)s_{c,m}^{RS}(i,k)}. \tag{3.21}$$

With the help of convexity property of *the perspective functions* (Boyd & Vandenberghe, 2004), one can readily show that the converted rate function in the BS-RS link $\hat{r}_{c,m}^{BS}(k)$ is concave in $\left(\hat{p}_{c,m}^{BS}(i,k), s_{c,m}^{BS}(i,k)\right)$and similarly, the rate function in RS-UE link $\hat{r}_{c,m}^{RS}(k)$ is concave in $\left(\hat{p}_{c,m}^{RS}(i,k), s_{c,m}^{RS}(i,k)\right)$, see *APPENDIX I*. With the latter power mapping operations, the resulted optimization problem is reformulated as

$$\max_{\left\{s^{BS}\right\},\left\{\hat{p}^{BS}\right\},\left\{s^{RS}\right\},\left\{\hat{p}^{RS}\right\}\{\rho\}} T_s \sum_{c\in\Phi}\sum_{m\in \mathrm{R}_c}\sum_{k\in \mathrm{K}_{c,m}} \omega_{c,m}(k)\rho_{c,m}(k) \tag{3.22}$$

subject to Eq.(3.11) and Eq.(3.12), along with

$$\hat{p}_{c,m}^{BS}(i,k) \geq 0, \quad \hat{p}_{c,m}^{RS}(i,k) \geq 0, \qquad \forall i \in N, k \in \mathrm{K}_{c,m},$$

$$s_{c,m}^{BS}(i,k), s_{c,m}^{RS}(i,k) \in [0,1], \qquad \forall c \in \Phi, i \in N, m \in \mathrm{R}_c, k \in \mathrm{K}_{c,m},$$

$$\sum_{m\in \mathrm{R}_c}\sum_{k\in \mathrm{K}_{c,m}}\sum_{i\in N} \hat{p}_{c,m}^{BS}(i,k) \leq p_{BS}^{max}, \qquad \forall c \in \Phi,$$

$$\sum_{k\in \mathrm{K}_{c,m}}\sum_{i\in N} \hat{p}_{c,m}^{RS}(i,k) \leq p_{RS}^{max}, \qquad \forall c \in \Phi, m \in \mathrm{R}_c,$$

$$\rho_{c,m}(k) \leq \alpha \hat{r}_{c,m}^{BS}(k), \qquad \forall c \in \Phi, m \in \mathrm{R}_c, k \in \mathrm{K}_{c,m},$$

$$\rho_{c,m}(k) \leq (1-\alpha)\hat{r}_{c,m}^{RS}(k), \qquad \forall c \in \Phi, m \in \mathrm{R}_c, k \in \mathrm{K}_{c,m},$$

$$\hat{\sigma}_{c,m}^2(i,k) \leq \theta, \qquad \forall i \in N, c \in \Phi, m \in \mathrm{R}_c, k \in \mathrm{K}_{c,m},$$

$$\frac{Q_{c,m}^{BS}(k)}{\alpha T_s} \le \hat{r}_{c,m}^{BS}(k), \qquad \forall c \in \Phi, m \in \mathrm{R}_c, k \in \mathrm{K}_{c,m},$$

$$\frac{Q_{c,m}^{RS}(k)}{(1-\alpha) T_s} \le \hat{r}_{c,m}^{RS}(k), \qquad \forall c \in \Phi, m \in \mathrm{R}_c, k \in \mathrm{K}_{c,m},$$

where the variance of the received interference at the *i*-th subcarrier is rewritten as

$$\hat{\sigma}_{c,m}^2(i,k) = \sigma_{c,m}^2(i,k)\Big|_{p_{c,m}^{RS}(i,k)=\hat{p}_{c,m}^{RS}(i,k)s_{c,m}^{RS}(i,k)}.$$

The resulting optimization problem Eq.(3.12) is now convex, along with all of the corresponding constraints, and the optimal value can be found efficiently in polynomial calculation time. In the remaining of this chapter, we refer to this scheme as the Time-shared method.

It is worth to point out the infeasible situation that occurs when the constraints are mutually inconsistent due to a case of practical network settings[5]. Relaxing the constraints or tuning the problem boundaries are examples of possible treatments for the infeasible situation. Adjusting the modulation and coding schemes or applying a congestion control mechanism can also be considered as the other counteracting methods. An admission control policy can also be useful in order to prevent infeasible situation. One efficient admission policy can be designed to filter out the traffic flows which are not within the system's ergodic capacity region. It can be shown that this admission control policy satisfies the necessary condition for system stability (see *APPENDIX II*).

The defined Time-shared approach results in the optimal solution when the subcarriers are allowed to be shared in one time-slot. However, the time-sharing assumption is not feasible for some network configurations when the time-synchronization imposes impractical

[5] For interested readers, there are feasibility check methods that define measures for infeasibility or violation of constraints. These methods can be used before actually performing the allocation procedure. The advantage of these approaches is that they identify which constraints should be relaxed to achieve a feasible problem or to find a potentially useful nearly feasible point (Boyd & Vandenberghe, 2004).

communication overhead. In the following section, we introduce an optimal solution for problem Eq.(3.16) when the binary subcarriers allocation is requested.

3.4 The Optimized Binary Resource Allocation (OBRA) approach

In order to have exclusive subcarrier allocation, we suggest an optimal binary resource allocation (OBRA) scheme that respects the power limit, interference and queue length constraints as stated in problem Eq.(3.16). In this scheme, we import the continuous allocation parameters $s^{*u}_{c,m}(i,k), u \in \{BS, \mathrm{R}S\}$ resulted from solving the Time-shared problem and round them to binary values. During this rounding procedure, for each subcarrier, the largest time-sharing allocation parameter among all users in one cell is set to 1, and that of the rest of users is set to zero, as stated in

$$\begin{aligned} &\tilde{s}^{u}_{c,m}(i,k^*) = 1, && k^* = \arg\max_{k} s^{*u}_{c,m}(i,k^*), \\ &\tilde{s}^{u}_{c,m}(i,k) = 0, && k \neq k^* \end{aligned} \tag{3.23}$$

where $u \in \{RS, BS\}$. The resulting subcarrier allocation parameters $\tilde{s}^{u}_{c,m}(i,k)$ respect the set of constraints Eq.(3.11) to Eq.(3.13).

Next step is to perform a power allocation optimization procedure in which the subcarrier allocation parameters $\tilde{s}^{u}_{c,m}(i,k)$ are fixed and binary. However, note that the rounding procedure Eq.(3.23) may lead to a set of subcarrier allocation parameters that could violate constraints Eq.(3.14) and Eq.(3.15). Therefore, we propose a subcarrier refining mechanism described in Algorithm 3.1 that adjusts the binary subcarrier allocation parameters in regard to stability constraints, in advance of running the power allocation optimization problem.

In Algorithm 3.1, we use the power allocation values $p^{*u}_{c,m}(i,k), u \in \{BS, RS\}$ resulted from Time-shared solution. However, we reverse the mapping function to retain the original power allocation values as $\tilde{p}^{u}_{c,m}(i,k) = p^{*u}_{c,m}(i,k) / s^{*u}_{c,m}(i,k)$. Consequently, the SINR formulas need

to be redefined as $\tilde{\Lambda}^{u}_{c,m}(i,k) = \Lambda^{u}_{c,m}(i,k)\Big|_{p^{u}_{c,m}=\tilde{p}^{u}_{c,m}}$, $u \in \{BS, RS\}$. In Algorithm 3.1 we employ the idea introduced in TPHA in CHAPTER 2, to find the unsatisfied users (lines 2 to7) and to swap subcarriers to meet the stability constraints Eq.(3.14) and Eq.(3.15) (lines 8 to 20). To prevent repetition, we only point out the function that calculates the subcarrier exchange cost vector in the SWAPPING procedure (lines 8 to 20), that is given by

$$\Psi^{k^{\dagger}} = \left\{ \psi_{i,k} \left| \psi_{i,k} = \frac{\tilde{\Lambda}^{u}_{c,m}(i,k) - \tilde{\Lambda}^{u}_{c,m}(i,k^{\dagger})}{\tilde{\Lambda}^{u}_{c,m}(i,k^{\dagger})},\ \tilde{s}^{u}_{c,m}(i,k) = 1, \forall i \in N, \forall k \in \mathrm{K}_c \right. \right\}.$$

We note that, similar to TPHA, the subcarrier i^* can be exchanged if it does not cause the donor user k^* to violate the stability constraint. For instance, on BS-RS link, the subcarrier i^* will be retained from k^* user, if the following condition is met:

$$\Delta f \sum_{i \neq i^*}^{n_f} \tilde{s}^{BS}_{c,m}(i,k^*) \log_2\left(1 + \tilde{\Lambda}^{BS}_{c,m}(i,k^*)\right) > \frac{Q^{BS}_{c,m}(k^*)}{\alpha T_s}.$$

Finally, the swapping procedure terminates if all users become satisfied or no more swapping would be possible because of resource exhaustion.

Now, given the fixed binary subcarrier allocation values $\left\{\tilde{s}^{u}_{c,m}(i,k), u \in \{BS, RS\}\right\}$ that are resulted from Algorithm 3.1, a power allocation optimization is essential to satisfy the rest of constraints defined in original problem.

Algorithm 3.1 Refining the binary subcarrier allocations

for all $k \in K_{c,m}, \forall m \in R_c, \forall c \in \Phi$ **do**
Find an unsatisfied user $k^{\dagger}$:
  if
$\Delta f \sum_{i}^{n_f} \tilde{s}_{c,m}^{BS}(i,k^{\dagger}) \log_2\left(1+\frac{\tilde{p}_{c,m}^{BS}(i,k^{\dagger}) l_{c,m}^{BS} \left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right) < \frac{Q_{c,m}^{BS}(k^{\dagger})}{\alpha T_s}$
    Call SWAPPING $(k^{\dagger}, BS)$
  else if
$\Delta f \sum_{i}^{n_f} \tilde{s}_{c,m}^{RS}(i,k^{\dagger}) \log_2\left(1+\frac{\tilde{p}_{c,m}^{RS}(i,k^{\dagger}) l_{c,m}^{RS}(k^{\dagger}) \left|H_{c,m}^{RS}(i,k^{\dagger})\right|^2}{\sigma_z^2+\theta}\right)$
$< \frac{Q_{c,m}^{RS}(k^{\dagger})}{(1-\alpha)T_s}$
    Call SWAPPING $(k^{\dagger}, RS)$
  end if
end for
procedure SWAPPING$(k^{\dagger}, u)$
  Calculate the subcarrier exchange cost vector $\Psi^{k^{\dagger}}$.
  Find the donor user k^* and candidate subcarrier $i^*$for subcarrier swapping:
  while $\exists k^* \in K$ Such that $\tilde{s}_{c,m}^{u}(i^*,k^*)=1$ & $\psi_{i^*,k^*} = min\left(\Psi^{k^{\dagger}}\right)$ **do**
    if reallocation of subcarrier i^* does not violate the stability requirement of donner user k^* **then**
      $\tilde{s}_{c,m}^{u}(i^*,k^*)=0$ & $\tilde{s}_{c,m}^{u}(i^*,k^{\dagger})=1$
    else
      update the cost entry $\psi_{i^*,k^{\dagger}} = \infty$,
      continue While
    end if
    if $k^{\dagger}$ is still unsatisfied **then**
      Continue While
    end if
  end while
end procedure

Therefore, the optimization objective of OBRA model is given by

$$\max_{\{p^{BS}\},\{p^{RS}\},\{\rho\}} T_s \sum_{c\in\Phi}\sum_{m\in R_c}\sum_{k\in K_{c,m}} \omega_{c,m}(k)\rho_{c,m}(k), \tag{3.24}$$

subject to (3.8), along with

$$\sum_{m\in R_c}\sum_{k\in K_{c,m}}\sum_{i\in N} \tilde{s}_{c,m}^{BS}(i,k)\, p_{c,m}^{BS}(i,k) \le p_{BS}^{max}, \qquad \forall c\in\Phi, \tag{3.25}$$

$$\sum_{k\in \mathrm{K}_{c,m}} \sum_{i\in N} \tilde{s}_{c,m}^{RS}(i,k)\, p_{c,m}^{RS}(i,k) \le p_{RS}^{max}, \qquad \forall c\in \Phi, m\in \mathrm{R}_c \tag{3.26}$$

,

$$\rho_{c,m}(k) \le \alpha \tilde{r}_{c,m}^{BS}(k), \qquad \forall c\in \Phi, m\in \mathrm{R}_c, k\in \mathrm{K}_{c,m} \tag{3.27}$$

,

$$\rho_{c,m}(k) \le (1-\alpha)\tilde{r}_{c,m}^{RS}(k), \qquad \forall c\in \Phi, m\in \mathrm{R}_c, k\in \mathrm{K}_{c,m} \tag{3.28}$$

,

$$\tilde{\sigma}_{c,m}^{2}(i,k) \le \theta, \qquad \forall i\in N, c\in \Phi, m\in \mathrm{R}_c, k\in \mathrm{K}_{c,m} \tag{3.29}$$

,

$$\frac{Q_{c,m}^{BS}(k)}{\alpha T_s} \le \tilde{r}_{c,m}^{BS}(k), \qquad \forall c\in \Phi, m\in \mathrm{R}_c, k\in \mathrm{K}_{c,m} \tag{3.30}$$

,

$$\frac{Q_{c,m}^{RS}(k)}{(1-\alpha)T_s} \le \tilde{r}_{c,m}^{RS}(k), \qquad \forall c\in \Phi, m\in \mathrm{R}_c, k\in \mathrm{K}_{c,m} \tag{3.31}$$

,

where $\tilde{r}_{c,m}^{BS}(i,k) = r_{c,m}^{BS}(i,k)\big|_{\tilde{s}_{c,m}^{BS}}$, $\tilde{r}_{c,m}^{RS}(i,k) = r_{c,m}^{RS}(i,k)\big|_{\tilde{s}_{c,m}^{RS}}$ and $\tilde{\sigma}_{c,m}^{2}(i,k) = \sigma_{c,m}^{2}(i,k)\big|_{\tilde{s}_{c,m}^{RS}}$. Since the problem Eq.(3.24) is convex, along with all of the constraints; we can achieve the closed form optimal solution in polynomial time. Note that the binary conversion operation defined in Eq.(3.23) and swapping the subcarrier allocation parameters actually extend the feasible rate region, and consequently, the sum rate resulted from the OBRA method brings an upper bound to the Time-shared method.

3.5 The Conservative Binary Resource Allocation (CBRA) approach

Following the trend of green communication, we are interested to find a stabilizing resource allocation method that is energy efficient. In particular, our goal is to cut off extra subcarriers that are not necessary to satisfy the stability constraints. In this regard, we

propose to tighten the stability constraints from inequality into equality constraints, as follows

$$r_{c,m}^{BS}(k) = \frac{Q_{c,m}^{BS}(k)}{\alpha T_s}, \qquad ? \quad \forall c \in \Phi, m \in \mathrm{R}_c, k \in \mathrm{K}_{c,m}, \tag{3.32}$$

$$r_{c,m}^{RS}(k) = \frac{Q_{c,m}^{RS}(k)}{(1-\alpha) T_s}, \qquad \forall c \in \Phi, m \in \mathrm{R}_c, k \in \mathrm{K}_{c,m} \tag{3.33}$$

.

This change of constraints contributes to minimizing the power consumption up to the point that it does not violate the stability criteria. It also eliminates waste of bandwidth by preventing service assignment to the empty queues. The proposed conservative method is formatted similar to OBRA. In particular, it first inputs the subcarrier allocation values obtained by solving the Time-shared method. Second, it rounds them to binary values using Eq.(3.23). Third, it applies Algorithm 3.1 to achieve a set of fixed binary subcarrier values. Next step is using the fixed subcarrier allocation values to form a power allocation optimization problem. However, by replacing the inequality stability constraints with equality condition, the problem turns into nonconvex format and therefore difficult to solve in polynomial time.

Among various approaches, Geometric Programming (GP) provides a more efficient method to solve nonconvex problems. A brief review of the acceptable formats of Geometric programming can be found in the *APPENDIX III*. The advantage of GP method is that it can be readily converted to the convex format using a standard logarithmic transformation. In order to transfer our model into the GP compliant format, we first define a logarithmic mapping $\rho_c(k) = \log_2 x_c(k)$ .The advantage of this one-to-one mapping is that it simply enables us to retrieve the original variables. This parameter mapping is then followed by a reformulation of the objective function and also the constraints in logarithmic format (see *APPENDIX IV*). The resulting product form of the optimization problem is then given by

$$\max_{\{x\}\{p^{BS}\}\{p^{RS}\}} \prod_{c\in\Phi} \prod_{m\in \mathrm{R}_c} \prod_{k\in \mathrm{K}_{c,m}} x_{c,m}(k)^{\omega_{c,m}(k)} \tag{3.34}$$

subject to Eq.(3.8), Eq.(3.25), Eq.(3.26), Eq.(3.29), along with

$$\prod_{i\in N}\left(1+\frac{p_{c,m}^{BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(i,k)}=2^{\left(\frac{Q^{BS}(k)}{\Delta f\alpha T_s}\right)},\qquad ?\quad \forall c\in\Phi,m\in \mathrm{R}_c,k\in \mathrm{K}_{c,m}, \tag{3.35}$$

$$\prod_{i\in N}\left(1+\frac{p_{c,m}^{RS}(i,k)l_{c,m}^{RS}(k)\left|H_{c,m}^{BS}(i,k)\right|^2}{\sigma_z^2+\theta}\right)^{\tilde{s}_{c,m}^{RS}(i,k)}=2^{\left(\frac{Q^{RS}(k)}{\Delta f(1-\alpha)T_s}\right)},\qquad \forall c\in\Phi,m\in \mathrm{R}_c,k\in \mathrm{K}_{c,m} \tag{3.36}$$

$$\prod_{i\in N}\left(1+\frac{p_{c,m}^{BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(i,k)}\geq x_{c,m}(k)^{1/\alpha\Delta f},\qquad ?\quad \forall c\in\Phi,m\in \mathrm{R}_c,k\in \mathrm{K}_{c,m}, \tag{3.37}$$

$$\prod_{i\in N}\left(1+\frac{p_{c,m}^{RS}(i,k)l_{c,m}^{RS}(k)\left|H_{c,m}^{BS}(i,k)\right|^2}{\sigma_z^2+\theta}\right)^{\tilde{s}_{c,m}^{RS}(i,k)}\geq x_{c,m}(k)^{1/(1-\alpha)\Delta f},\qquad \forall c\in\Phi,m\in \mathrm{R}_c,k\in \mathrm{K}_{c,m} \tag{3.38}$$

The stability constraints Eq.(3.32) and Eq.(3.33) are reformulated into product forms Eq.(3.35) and Eq.(3.36), so that it can be simply converted into monomial format, which is the standard format for equality constraints in GP. Similarly, the epigraph constraints Eq.(3.17) and Eq.(3.18) are reformulated into Eq.(3.37) and Eq.(3.38) respectively, using the aforementioned logarithmic mapping.

Unfortunately, the latter product form constraints still do not comply with the GP format. More specifically, the equality constraints Eq.(3.35) and Eq.(3.36) do not respect the monomial format and the inequality constraints Eq.(3.37) and Eq.(3.38) cannot be considered as posynomial functions. However, the solution of problem in Eq.(3.34) can be closely estimated using *single condensation method* (Chiang et al., 2005). The idea of single condensation method is to solve an iterative series of GPs, which are approximates of the original problem Eq.(3.34). To do this, the monomial approximation technique can help to replace the constraints Eq.(3.35) to Eq.(3.38) with their monomial estimation (see *APPENDIX V*). Ultimately, the corresponding monomial approximated constraints are respectively given by

$$\prod_{i\in N}\left(1+\frac{p_{c,m}^{(0)BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(i,k)}\prod_{i\in N}\left(\frac{p_{c,m}^{BS}(i,k)}{p_{c,m}^{(0)BS}(i,k)}\right)^{\beta_{c,m}^{BS}(i,k)}=2^{\left(\frac{Q_{c,m}^{BS}(k)}{\alpha T_s\Delta f}\right)}, \tag{3.39}$$

$$\prod_{i\in N}\left(1+\frac{p_{c,m}^{(0)RS}(i,k)l_{c,m}^{RS}(k)\left|H_{c,m}^{RS}(i,k)\right|^2}{\sigma_z^2+\theta}\right)^{\tilde{s}_{c,m}^{RS}(i,k)}\prod_{i\in N}\left(\frac{p_{c,m}^{RS}(i,k)}{p_{c,m}^{(0)RS}(i,k)}\right)^{\beta_{c,m}^{RS}(i,k)}=2^{\left(\frac{Q_{c,m}^{RS}(k)}{(1-\alpha)T_s\Delta f}\right)}, \tag{3.40}$$

$$\prod_{i\in N}\left(1+\frac{p_{c,m}^{(0)BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(i,k)}\prod_{i\in N}\left(\frac{p_{c,m}^{BS}(i,k)}{p_{c,m}^{(0)BS}(i,k)}\right)^{\beta_{c,m}^{BS}(i,k)}\geq x_{c,m}(k)^{1/\alpha\Delta f}, \tag{3.41}$$

$$\prod_{i\in N}\left(1+\frac{p_{c,m}^{(0)RS}(i,k)l_{c,m}^{RS}(k)\left|H_{c,m}^{RS}(i,k)\right|^2}{\sigma_z^2+\theta}\right)^{\tilde{s}_{c,m}^{RS}(i,k)}\prod_{i\in N}\left(\frac{p_{c,m}^{RS}(i,k)}{p_{c,m}^{(0)RS}(i,k)}\right)^{\beta_{c,m}^{RS}(i,k)}\geq x_{c,m}(k)^{1/(1-\alpha)\Delta f}. \tag{3.42}$$

where the exponents $\beta_{c,m}^{BS}(i,k)$ and $\beta_{c,m}^{RS}(i,k)$ are respectively defined as

$$\beta_{c,m}^{BS}(i,k)=\frac{\tilde{s}_{c,m}^{BS}(i,k)\,p_{c,m}^{(0)BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2+p_{c,m}^{(0)BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2},$$

$$\beta_{c,m}^{RS}(i,k)=\frac{\tilde{s}_{c,m}^{RS}(i,k)\,p_{c,m}^{(0)RS}(i,k)l_{c,m}^{RS}(k)\left|H_{c,m}^{RS}(i,k)\right|^2}{\sigma_z^2+\theta+p_{c,m}^{(0)RS}(i,k)l_{c,m}^{RS}(k)\left|H_{c,m}^{RS}(i,k)\right|^2}.$$

Now, the problem in Eq.(3.34) and the corresponding constraints Eq.(3.8), Eq.(3.25), Eq.(3.26), Eq.(3.29), Eq.(3.39), Eq.(3.40), Eq.(3.41),and Eq.(3.42) form a GP compatible problem to be solved in each iteration. Based on *the single condensation method*, each iteration needs to be fed by initial power values; $p_{c,m}^{(0)BS}(i,k)$, $p_{c,m}^{(0)RS}(i,k)$ and the solution of each iteration can be used as the initial point for the next iteration and so on. The initial values of the very first iteration can be derived from solving the Time-shared problem Eq.(3.22) and projecting them onto the feasible region, i.e. $p_{c,m}^{(0)BS}(i,k)=\hat{p}_{c,m}^{BS}(i,k)/\tilde{s}_{c,m}^{BS}(i,k)$

and $p_{c,m}^{(0)RS}(i,k) = \hat{p}_{c,m}^{RS}(i,k) / \tilde{s}_{c,m}^{RS}(i,k)$. The iterations generated by this sequential algorithm converge to a solution of KKT system corresponding to the original problem upon satisfying specific conditions (see APPENDIX VI) outlined in. (Chiang et al., 2005; Marks & Wright, 1978) Since the problem Eq.(3.34) is not convex, the KKT conditions are only necessary for optimality, and Eq.(3.34) returns a lower bound on the optimal achievable system throughput.

3.6 Comparison of computation complexity

We examine the computational complexity of solving the suggested resource allocation problem when applying the Time-shared, OBRA and CBRA methods. To have a basic comprehension, we note that the exact optimal binary allocation can be achieved by using the exhaustive-search method, in which all possible allocation profiles need to be checked. In our considered multi-cell system consisting of φ cells, the number of possible allocation profiles will increase exponentially to $\varphi^{K^{2n_f}}$. Each of the subcarrier allocation profiles are subsequently employed to solve a convex power allocation problem. This power allocation problem is similar to the Time-shared problem, with the exception that the subcarrier allocation parameters $\left\{s_{c,m}^{BS}(i,k), s_{c,m}^{RS}(i,k)\right\}$ are not involved in the optimization procedure and are treated as constants, and correspondingly the constraints Eq.(3.11) to Eq.(3.13) are eliminated.

Similarly, one can calculate the computation complexity of solving the binary allocation problem using Branch-and-Bound mechanism, described in CHAPTER 2, which sequentially solves $2^{4\varphi K_c n_f}$ convex subproblems. We illustrate the results of complexity estimation in Table 3.1.

The Time-shared and OBRA problems are convex, and therefore, the global optimum solution can be found in polynomial time using an efficient solver. However in the case of CBRA problem, we can find a close approximation of the optimal solution using single

condensation method. For the sake of comparison, we consider highly efficient interior-point methods (IPM) to solve our introduced convex problems since their computation complexity is measurable and is proportional to the number of constraints and optimization variables. The underlying discipline of IPM is based on Newton's method along a central path. The complexity of each Newton step grows with the cube of the number of constraints. The number of Newton's steps is bounded by $\sqrt{Z}$, where Z indicates the complexity of one Newton step. This approximation requires that the log-barrier function, which is composed of the objective function and all of the constraints, be self-concordant (Nesterov et al., 1994). Although the log-barrier functions of the problems described in the Time-share, OBRA and CBRA methods are not self-concordant, without loss of generality, we can add extra constraints that do not change the problem but convert it to the self-concordant format, see *APPENDIX VII*.

Table 3.1 Complexity comparison of the resource allocation methods in a single cell model

Scheme	Estimated complexity
Time-shared	$O\left(\sqrt{\left(\varphi K_c\left(5+7n_f\right)+\varphi\left(1+M_c+2n_f\right)\right)^3}\right)$
OBRA	Time-shared+ $O\left(\sqrt{\left(\varphi K_c\left(5+3n_f\right)+\varphi\left(1+M_c\right)\right)^3}\right)$
CBRA	Time-shared+ $O\left(X\sqrt{\left(\varphi K_c\left(5+3n_f\right)+\varphi\left(1+M_c\right)\right)^3}\right)$
Branch-and-Bound	$O\left(2^{4\varphi K_c n_f}\sqrt{\left(\varphi K_c\left(5+3n_f\right)+\varphi\left(1+M_c\right)\right)^3}\right)$
Exhaustive search	$O\left(\varphi^{K^{2n_f}}\sqrt{\left(\varphi K_c\left(5+3n_f\right)+\varphi\left(1+M_c\right)\right)^3}\right)$

Note that the computation complexity of OBRA method includes the execution of the Time-shared method followed by Algorithm 3.1 and the power allocation problem Eq.(3.24). This is the same for the case of CBRA method however, this time Algorithm 3.1 is followed by an iterative procedure of GP optimization problems. The number of required iterations X for the GP problem to converge never exceeds 100, which is independent from the number of users or subcarriers, based on the simulation observations herein and in (Chiang et al., 2005;

Rashtchi et al., 2014). Note that the computation complexity of Algorithm 3.1 is not taken into consideration because it is significantly smaller than the complexity of the optimization problems.

As presented in Table 3.1, the complexity of the Time-shared, OBRA and CBRA approaches are comparable and polynomial in number of subcarriers n_f, cells φ, relays M_c and users in each cell K_c, which is rather efficient than using exhaustive search or Branch-and-Bound methods.

3.7 Numerical results and discussion

In order to evaluate the performance of the proposed methods, we consider a multi-cell scenario similar to the one depicted in Figure 3.1. In this section, we describe the system level model and the corresponding default value of the configuration parameters. The considered multi-cell system consists of seven cells, each serving three relay stations. The LoS link between the base station and its corresponding relays is modeled similar to CHAPTER 2, i.e. it experiences lognormal shadowing, Rician fading with Rician factor 6 dB. The distance between the base station and each relay is 20 Km. A relay serves cell-edge UEs, which are slowly moving in a radius of 100 m from the connected relay station. The maximum Doppler frequency equals 15 Hz. The RS-UE link is NLoS, and suffers multipath Rayleigh fading and lognormal shadowing, which can be modeled with i.i.d random variables as described in CHAPTER 2. The BS-RS and RS-UE links both follow the suggested 3GPP path loss model, i.e urban macrocell in LoS and urban microcell in NLoS, respectively (*Universal Mobile Telecommunications System (UMTS); Spacial channel model for Multiple Input Multiple Output (MIMO) simulations*, n.d.). In order to model the interfering signal received from neighboring relay stations to the cell-edge users, we use the NLoS path loss in the distance 220 m and the multipath Rayleigh fading and lognormal shadowing, similar to local RS-UE links. Twelve subcarriers of 15 kHz bandwidth are available in the downlink subframe, and the carrier center frequency is 2.5 GHz. The noise

power density equals -174 dBm/Hz. The default value of maximum transmit power at the BS and RS are equally set to $P_{BS}^{max} = P_{RS}^{max} = 20W$.

Without loss of generality, all incoming streams are assumed to have identical mean arrival rate and the users' weight parameters are also amounted to unity i.e. $\omega_{c,m}(k) = 1, k \in \{1,2,...,K\}$. The duration of one time slot is 5 ms, and the proper default time slot division value is determined by offline simulation resulted to $\alpha = 0.6$, which will be discussed later in this section. The entering data traffic at the base station can be modeled by a bursty On/Off traffic model (Lakani et al., 2015). However, the length of On-time (normally 2 s) is far larger than the time slot length (5 ms), which resembles a constant bit rate (CBR) traffic during one time slot. For this reason, we consider CBR traffic arrival rate in the current work and its default value is 5 Kbps. We normalize the interference tolerance parameter over the value of AWGN, denoted by $INR = \theta / \sigma_z^2$, for the sake of comparability. Its proper default value is set to INR=100 using an offline simulation, as we discuss it later in this section.

We compare our suggested methods to two other allocation schemes, which we refer to as benchmark-1 (Y. J. Zhang & Ben Letaief, 2004) and benchmark-2 (Ng & Schober, 2011). Particularly, the method defined in benchmark-1, distributes the maximum transmission power evenly on all subcarriers, and then performs a heuristic-based binary subcarrier allocation problem in order to maximize the system throughput while the allocated rate is greater than or equal to a particular minimum rate requirement. The binary allocation feature makes benchmark-1 a proper choice of performance comparison to our proposed OBRA and CBRA methods. Benchmark-2 considers the throughput maximization in an interference limited multi-cell and multi-user scenario that is constrained on a minimum data rate. Benchmark-2 allocates transmission power and OFDMA subcarriers in a flexible two mini-slot scenario, and also benefits from relaxing the binary subcarrier allocation into time-shared values. This makes benchmark-2 a good choice for performance comparison to our proposed Time-shared model. We assume the minimum data rate requirement for both benchmark methods equal to 0.01 b/s/Hz. We apply MOSEK (MOSEK Apps., 2014) solver in CVX

package (M. Grant & S. Boyd, 2014) for solving the optimization problems described in Time-shared, OBRA, CBRA and Benchmark-2 methods.

3.7.1 System throughput

For the first evaluation scenario, we examine the system throughput calculated by Eq.(3.7) resulted by applying the Time-shared, OBRA, CBRA in comparison to the performance of benchmark-1 and benchmark-2 schemes. The system throughput is calculated for different values of maximum power limit, which is assumed to be equal for the base stations and relays, i.e. $p_{BS}^{max} = p_{RS}^{max}$. For each power limit value, we average the system throughput per bandwidth (b/s/Hz) over 100 independent realizations of the simulated system and illustrate it in Figure 3.2.

As it can be seen in Figure 3.2, the OBRA and CBRA methods return an upper bound and a lower bound, on the throughput of the Time-shared method, respectively. This result has been expected regarding to the extended feasible regions of OBRA and tightened rate constraints in CBRA, comparing to the Time-shared method. Note that CBRA returns a fixed throughput due to its restricted rate policy, while the throughputs of other methods are rising by power increase.

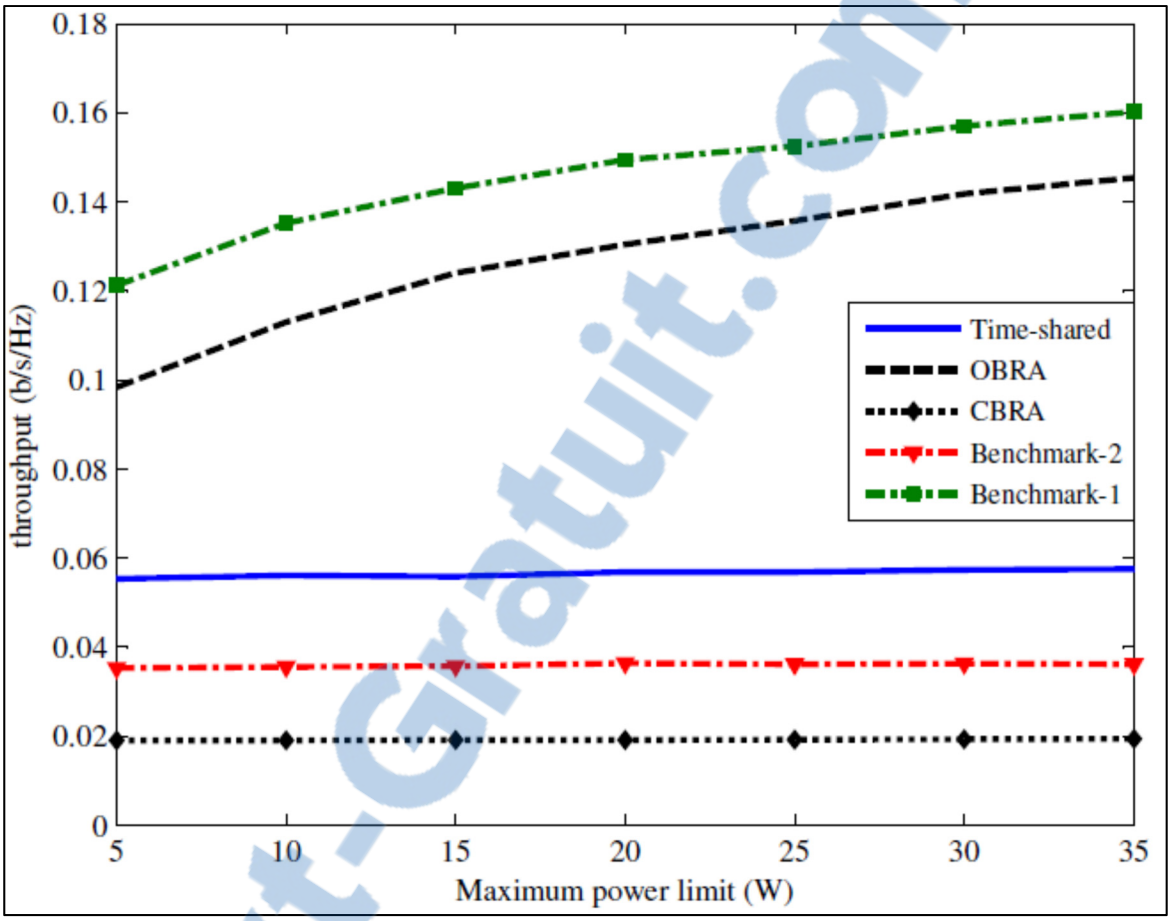

Figure 3.2 System throughput comparison by applying the Time-shared, OBRA, benchmark-1 and benchmark-2 methods

The throughput of Benchmark-2 is in close proximity to that of the Time-shared model. However, the Time-shared method gains 30% larger throughput due to its flexible rate constraints compared to Benchmark-2, which considers fixed data rate requirement for both links. Benchmark-1 gains the highest throughput but it is interesting to further study the cost that is imposed to the system to achieve this throughput gain by the next experiment on power consumption.

3.7.2 Energy efficiency

To measure the performance of the introduced methods in terms of power consumption, we calculate the sum of transmitted power from all the base stations, using

$\sum_{c\in\Phi}\sum_{m\in R_c}\sum_{k\in K_{c,m}}\sum_{i\in N} s_{c,m}^{BS}(i,k)\,p_{c,m}^{BS}(i,k)$ added to the transmitted power from relay station, using $\sum_{c\in\Phi}\sum_{m\in R_c}\sum_{k\in K_{c,m}}\sum_{i\in N} s_{c,m}^{RS}(i,k)\,p_{c,m}^{RS}(i,k)$. Then, we illustrate the energy efficiency of the compared methods by dividing the sum system throughput Eq.(3.7) over the calculated sum system power consumption.

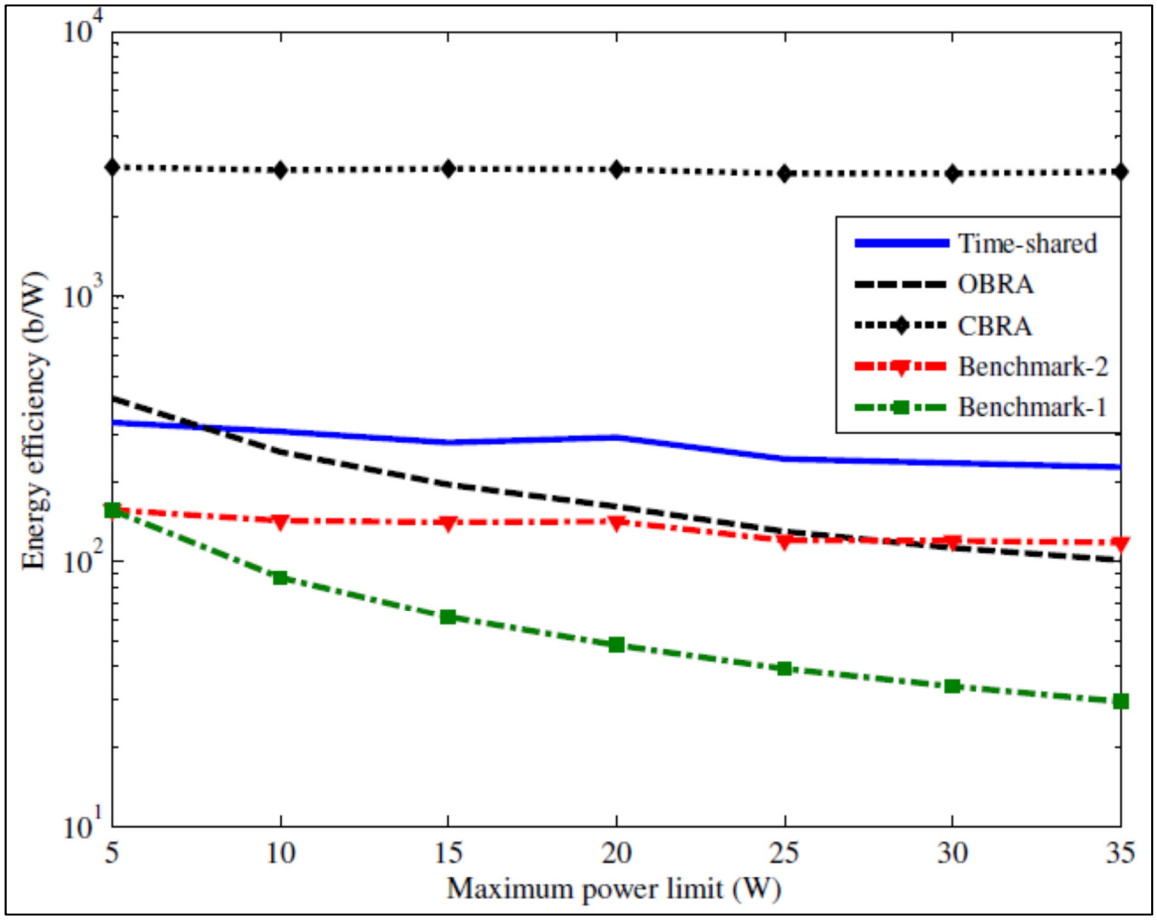

Figure 3.3 System energy efficiency applying the Time-shared, OBRA, CBRA, benchmark-1 and benchmark-2 methods

In Figure 3.3 we present the achieved energy efficiency over various power limitation scenarios. It can be seen that CBRA, OBRA and Time-shared methods are more energy efficient than two benchmark schemes. Furthermore, CBRA is the most energy efficient method. Particularly, when power limit is 35 W, CBRA achieves 92%, 96%, and 99% power

improvement comparing to Time-shared, Benchmark-2 and Benchmark-1, respectively. It can be seen that, the reason Benchmark-1 gains the highest throughput is due to its large power consumption, which makes this method the least power-efficient one.

3.7.3 Impact of varying channel state on throughput

For the next performance evaluation scenario, we examine the optimal system throughput when varying the channel condition. We define three distance profiles denoted by $DP\#(d1,d2)$, where $d1$ and $d2$ express respectively the BS-RS and RS-UE distances in meters. These distance profiles are defined as $DP1=(15000,100)$, $\mathrm{DP2}=(7500,100)$, $DP3=(15000,500)$. We depict the cumulative distribution function of the system throughput in Figure 3.4 as a result of applying the time-shared method averaged over 100 independent realizations.

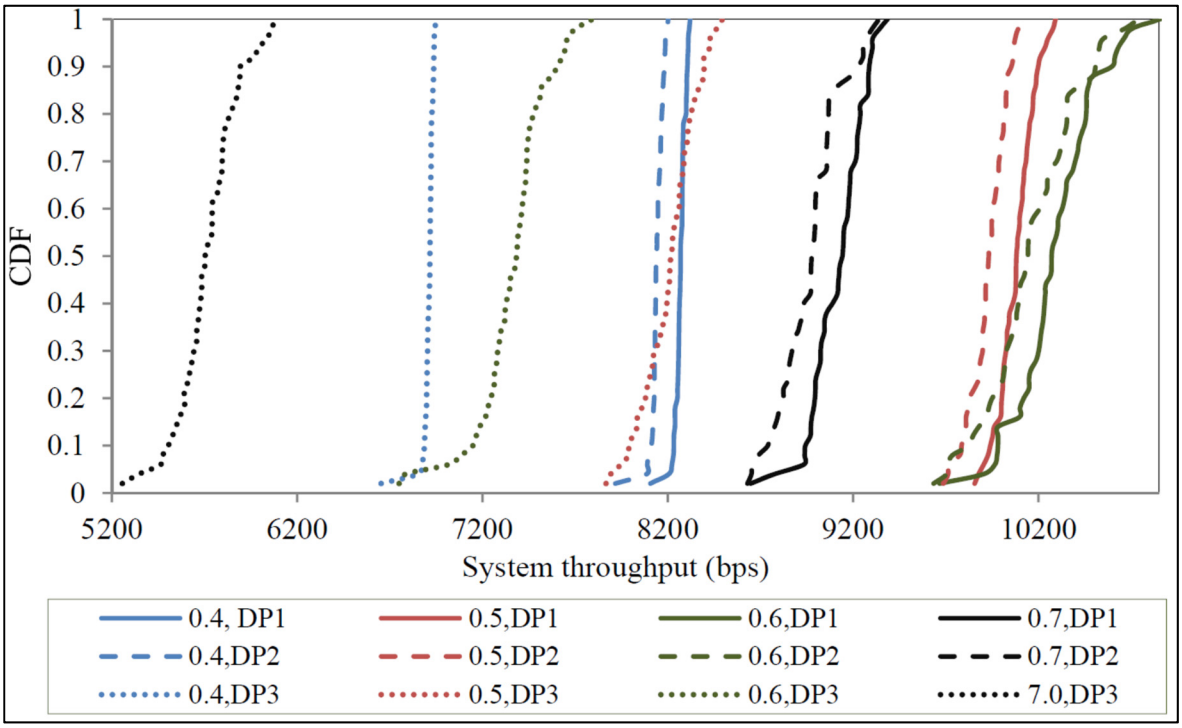

Figure 3.4 Cumulative distribution function of system throughput applying the multi-cell Time-shared method compared for different time-slot division parameters and distance profiles {DP1,DP2,DP3}, denoted by $\{\alpha, DP\#\}$

It is observed that higher system throughput can be achieved when the distance between the RS and UE is short (100 m). Considering the system throughput formulation Eq.(3.7) that is calculated by the minimum rate achieved in each of the two links, the latter observation in Figure 3.4 highlights the fact that the system throughput is rather sensitive to the variation of the channel gain at RS-UE link which has a lower SINR than the BS-RS link. Therefore, when the RS-UE distance increases, as in DP3, the system throughput drastically decreases.

3.7.4 Impact of adaptive time-slot division on throughput

The results depicted in Figure 3.4 present another performance metric that affects the system throughput that is the time-slot division parameter α. It can be seen that when the relay station has a larger portion of time-slot for transmission than the base station, e.g. $\alpha = 0.4$, it does not increase the RS-UE data rate and similarly the system throughput. However, shrinking the relay's portion of time-slot increases the data rate on RS-UE link as it can be seen when α equal to 0.6 or 0.5.

In addition, when the RS-UE distance reaches 500 meters as in DP3, the RS-UE channel condition drastically degrades to the point that shrinking the relay's share of time-slot more than 50%, e.g. $\alpha = \{0.6, 0.7\}$, does not contribute to increasing the rate any further. Therefore, it is worth to note the limit of shrinking the relay's portion of time slot in order to gain the desired performance. For instance, the equilibrium of maximum throughput has been achieved by $\alpha = 0.6$ when RS-UE distance is 100 m, while setting $\alpha = 0.7$ leaves the relay stations struggling to acquire appropriate rate. This observation highlights the importance of adjustable time-slot partitioning, which challenges the efficiency of the resource allocation mechanisms that are based on fixed time-slot partitioning. For the simulation trials of the current work, we applied the resulted optimum time-slot ratio $\alpha = 0.6$ which we acquired offline. However, the idea of adding the time-slot partitioning value to the set of optimization parameters can represent an opportunity to extend the current work in the future.

3.7.5 Impact of varying channel state and time-slot division on power consumption

The power consumption that is summed over all base stations when applying the Time-shared method is depicted in Figure 3.5-(a), targeting to compare the effect of adjusting time-slot ratio $\alpha \in [0.1, 0.8]$ and different distance profiles. It can be seen that power consumption is rather sensitive to the channel condition than the time-slot division ratio. By improving the channel condition on BS-RS link from DP1 to DP2, we achieve 84% decrease of power consumption at the base stations. However, this change does not variate the system throughput more than 0.5%, as we have seen in Figure 3.4.

The sum power consumption of relay stations is depicted in Figure 3.5-(b). It can be noted that when the RS-UE channel gain degrades from DP1 to DP3 due to distance increase, the relay stations consume 51% more power. Interestingly, this channel gain degradation in RS-UE link, i.e. from DP1 to DP3, contributes to 24% less power consumption at the base stations as can be seen in Figure 3.5-(a). This observation shows an interesting correlation between the rate of the BS-RS link and that of RS-UE link in our optimization problem. Recall that the optimization problem tries to maximize the minimum rate of two links, hence when the RS-UE link condition is good, i.e. DP1; it is highly probable that the minimum rate equals the rate at the BS-RS link. We examined this claim via simulations and it shows that there is 35% chance that the minimum rate equals to the rate at the BS-RS link, when distance profile is DP1, and this chance is 0% for distance profile DP3. So when the minimum rate is equal to the rate at the BS-RS link, the system tries to converge to the values where BS-RS rates are maximized, which leeds to more power consumption at the base stations.

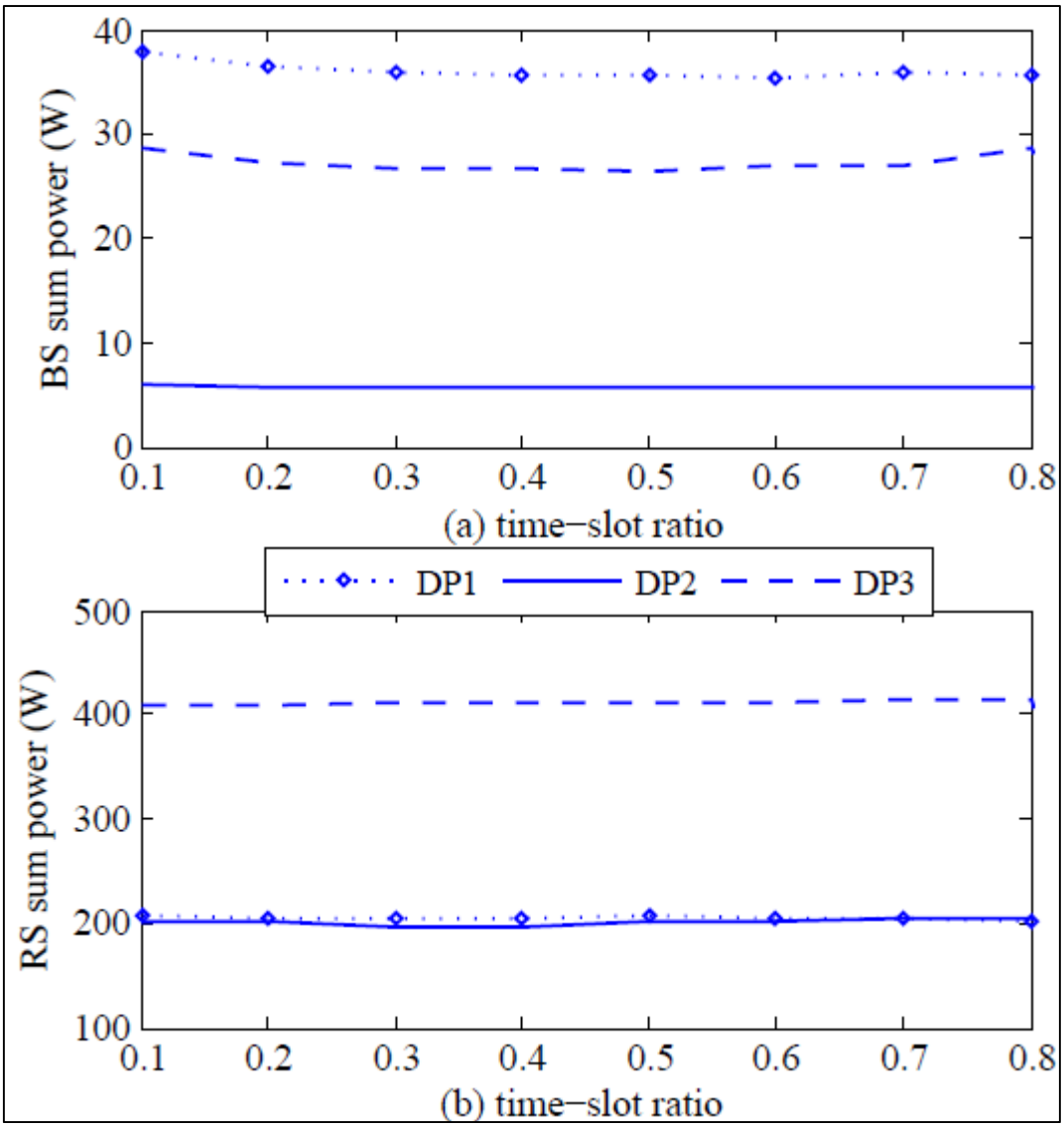

Figure 3.5 Sum power consumption of the base stations (a) and relay stations (b), applying the Time-shared method in multi-cell network, when varying distance profiles {DP1,DP2,DP3} and time-slot division ratios

3.7.6 System Stability

We examine the stability feature of the proposed method by depicting the length of the queues placed at the base stations in Figure 3.6 and at relay stations in Figure 3.7, during 200 consecutive time slots. The depicted results correspond to the queue length assigned to accommodate the traffic flow of 3 users placed in the central cell in the 7-cell model depicted

in Figure 3.1. In this experiment, the arrival rate is 70 Kbps that is the largest arrival rate under which the proposed Time-shared method could stabilize the queues. The result of proposed Time-shared scheme is compared to Benchmark-1 and Benchmark-2 methods. We cast aside the queue length results from applying OBRA and CBRA methods for the sake of readability, since they show analogous values to the Time-shared results.

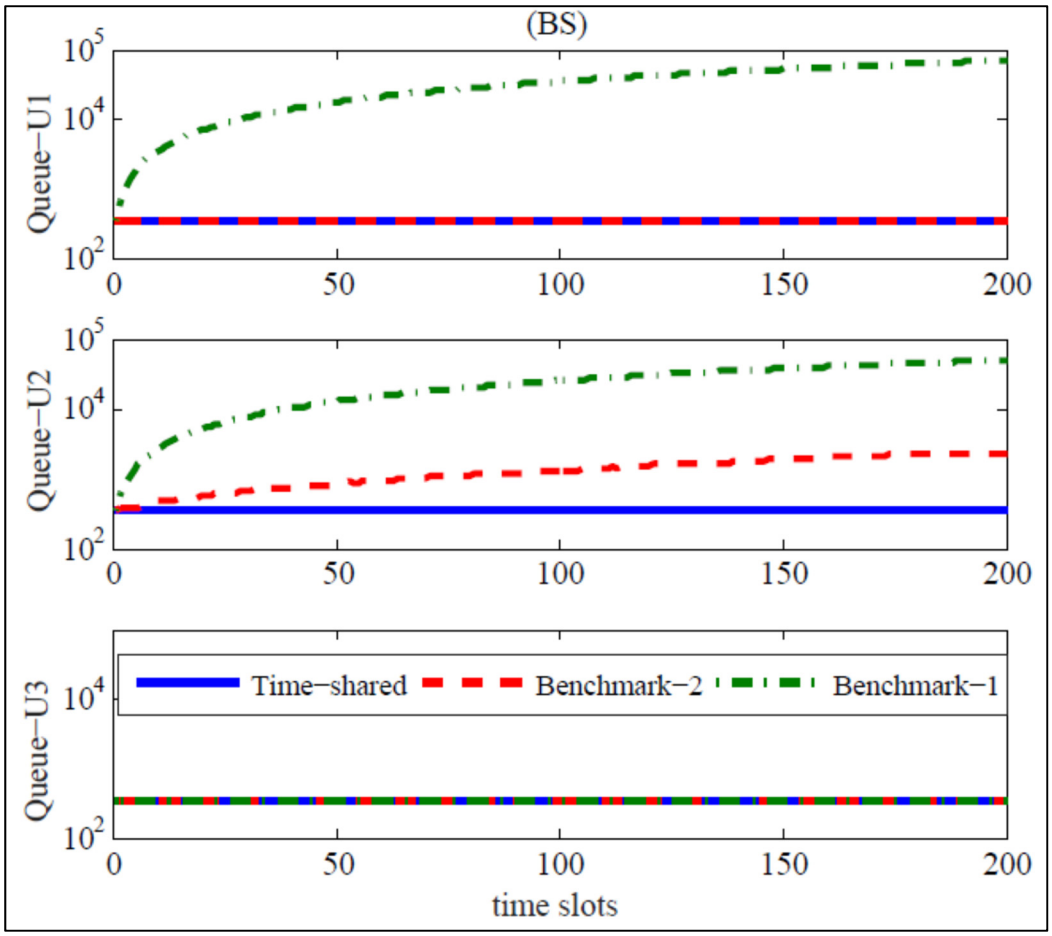

Figure 3.6 Queue length growth at the base station, applying the Time-shared and two benchmark schemes in the multi-cell network, the queue lengths (shown in bits at the vertical axes) correspond three users (U1,U2,U3) in the central cell

In this experiment, we observe that the Time-shared method can stabilize the queue lengths at both the base station and relay stations and the length is close to the quantity of arrived bits during one time-slot, i.e. 350 bits. However it can be seen that, without the stability constraints enforcement, as in the Benchmark-1 and Benchmark-2 methods, the allocation

mechanism fails to stabilize the system and therefore the queues length grow exponentially. In a closer look, the Benchmark-1 fails to stabilize the queue length assigned to user 1 and user 2 at the base station. At the relay stations, Benchmark-1 keeps user 2 and user 3 queues under control at the cost of depriving user 1 from any rate. Contrary to Benchamrk-1, Benchmark-2 does not deprive users from bandwidth; however it causes the queue length to grow exponentially.

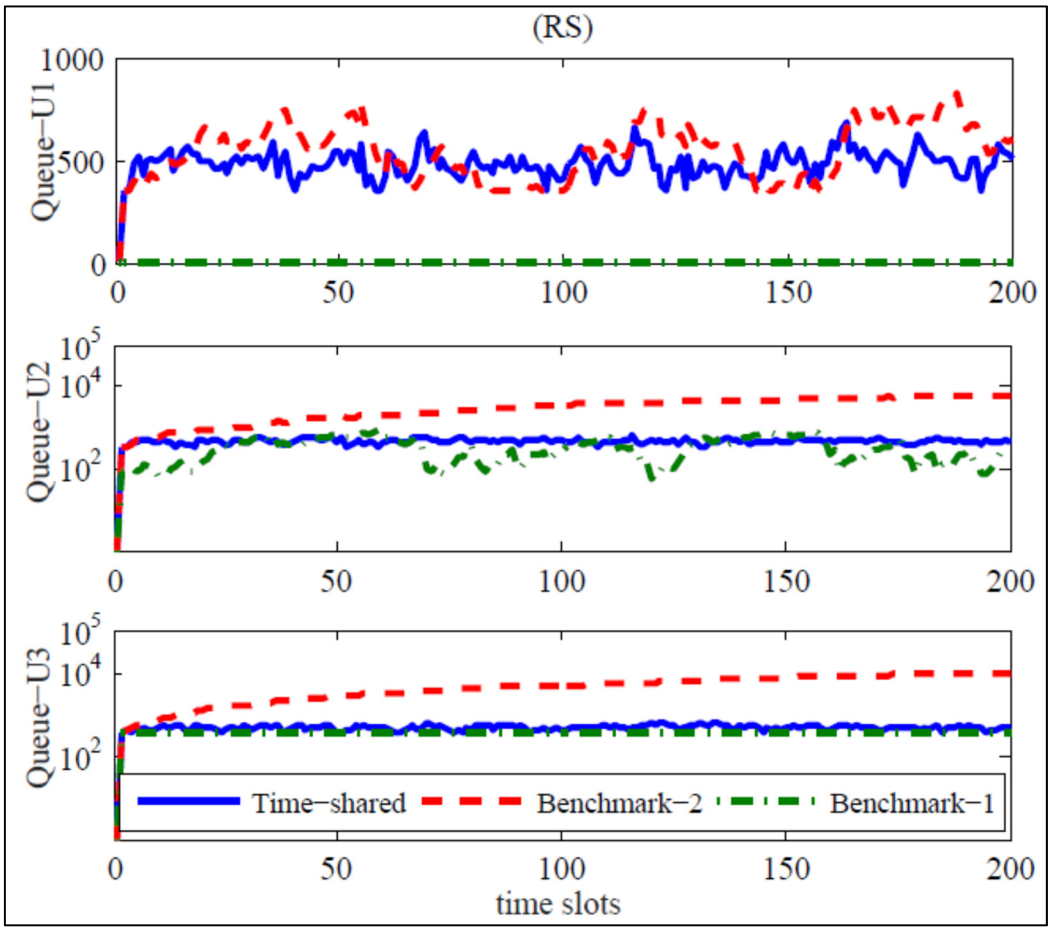

Figure 3.7 Queue length growth at relay stations applying the Time-shared and two benchmark schemes, the queue lengths (shown in bits at the vertical axes) correspond three users (U1,U2,U3) in the central cell

3.7.7 Offline tuning of a proper interference tolerance value

The interference tolerance value is a tuning parameter that can be adjusted with different system configurations to keep the performance in a desired level. In order to find the most proper value of interference tolerance, in Figure 3.8 we show the empirical probability that the received interference, i.e. $\sigma^2_{c,m}(i,k), \forall i \in N, m \in \mathrm{R}_c, c \in \Phi$, is smaller than the interference tolerance θ. This trend is resulted from running the Time-shared method with the default configuration values.

Considering the range $INR \in \{0,1,10,50,100,150,200\}$, the aforementioned probability becomes equal to one when $INR \geq 100$. Since INR > 100 can cause waste of transmission power, we set $INR = 100$ as the proper default value for all the simulation scenarios in this chapter. Performing an offline interference investigation is necessary to find the proper interference tolerance before the actual online allocation mechanism. This offline procedure can be provisioned since we assumed that UEs and relays feedback their channel state information to the base station. This experiment opens an opportunity to future studies in order to engage the calculation of the optimal interference tolerance value as a part of the online resource allocation procedure.

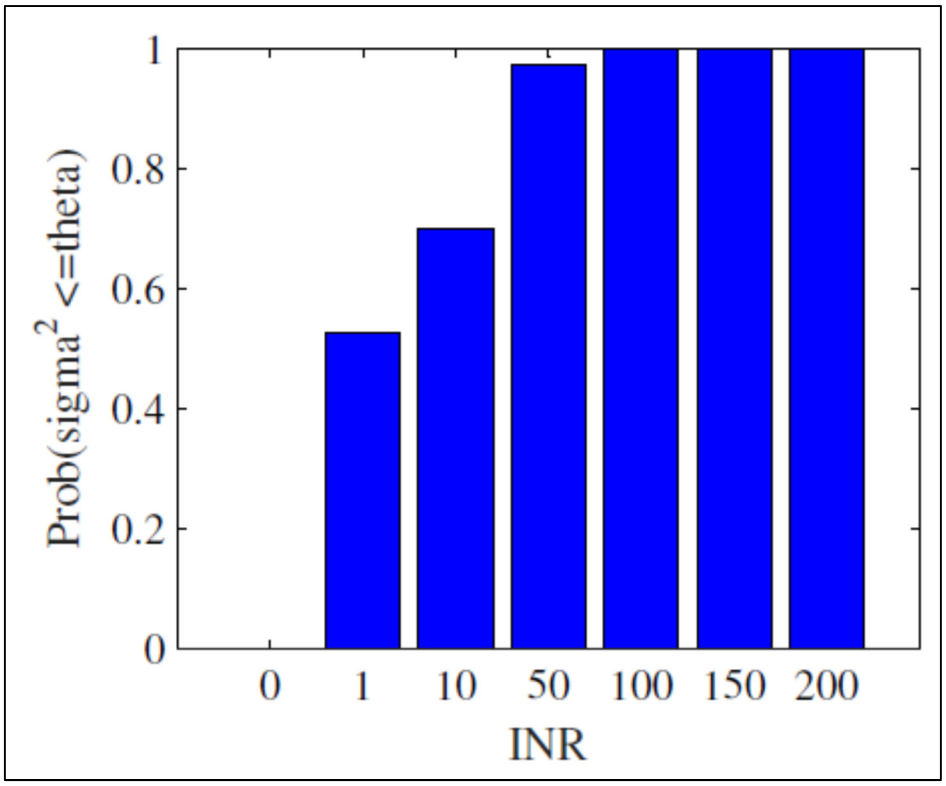

Figure 3.8 Empirical probability of the received interference on RS-UE link being smaller than the interference tolerance, applying the Time-shared method over various values of Interference over Noise Ratio (INR)

3.8 Conclusion

In this chapter, we combine the problem of OFDMA subcarrier allocation with optimal power assignment in a multi-cell decode-and-forward relaying network. One of the innovative aspects of this work is to consider several performance metrics such as stability an interference control in one inclusive resource allocation problem. We propose a set of queue length constraints in order to meet the stability requirement for the buffers placed in the base station and relay stations. The advantage of the stability constraints to the model described in is that it does not require a priori knowledge of the arrival process statistics. The inter-cell interference is also incorporated into the allocation problem, assuming the relay stations to be the effective sources of inter-cell interference for the cell-edge users located in neighbor cells.

Since the defined problem of OFDMA subcarrier and power allocation is combinatorial and nonconvex, we proposed the Time-shared approach in order to achieve a closed form optimal solution. When system design limitations do not allow the subcarriers to be time-shared, we propose an optimal binary allocation scheme, namely OBRA. Despite the other binary allocation mechanisms, we show that the computation complexity of OBRA is polynomial in time. In addition, we propose a power conservative approach, namely CBRA, which can prevent waste of bandwidth, and consequently decrease the system power consumption. With the help of Geometric Programing technique and monomial approximations, we convert the nonconvex CBRA problem to a convex form that can be solved in polynomial time.

Simulation results demonstrate the success of the proposed methods toward stabilizing the queues. It has been observed that, among the proposed allocation models, OBRA achieves larger system throughput and CBRA offers the best energy efficiency. We also discuss that the dynamic tuning of the time-slot division and interference tolerance parameters play an important role for achieving the desired performance. Therefore, as future extensions of this work, one can extend the optimization model by including the time-slot division and interference tolerance parameters into the optimization variables.

CHAPTER 4

LOW OVERHEAD AND DISTRIBUTED SUBCARRIER AND POWER ALLOCATION IN RELAY-AIDED MULTI-CELL NETWORK WITH STABILITY CONSTRAINTS

Choosing between a centralized and a distributed approach is a controversial subject of study in radio resource allocation problems since each approach offers different performance gain. The centralized schemes tend to gather all the information and the feedback required in order to perform global resource allocation likewise the methods introduced in previous chapters. On the other hand, the centralized methods involve a burdensome computation complexity in the presence of very populated networks. In other words, as more users and coordinated cells are managed, the required feedback load for CSI and QSI increases even more since it is needed at the central decision making station. This also increases the size of parameters in the optimization problem that requires larger processing time and calculation power. This degrades the practicality and scalability of centralized methods in terms of energy efficiency and latency in the presence of large number of collaborative cells. This motivates us to design a queue-aware distributed method that can speed-up the computation in several local units which also imposes less signaling overhead than the centralized one.

4.1 System model

4.1.1 Channel capacity

We adopt the system model defined in 3.1.2, however in this chapter, we consider a more general condition for channel capacity allocation. In particular, we assume that the RS does not benefit from interference cancelation techniques, and consequently we need to control the inter-cell interference each BS imposes to the relays in neighbor cells, as depicted in Figure 4.1.

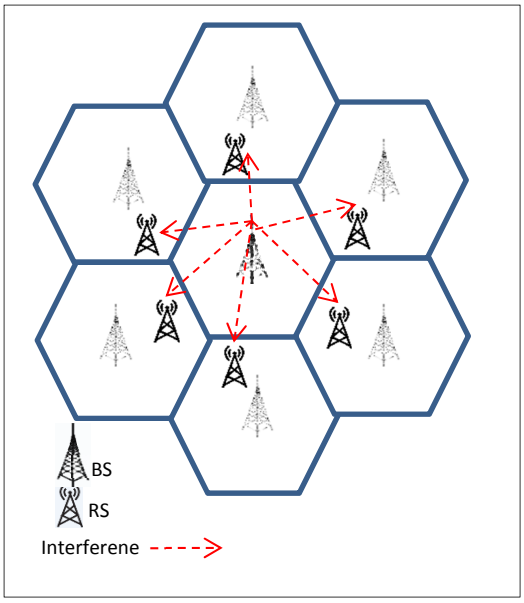

Figure 4.1 Interfering links from BS to neighboring relays

In this regard, the received SINR at the relay station is reformulated and is given by

$$\Lambda_{c,m}^{BS}(i,k,t)=\frac{P_{c,m}^{BS}(i,k,t)h_{c,m}^{BS}(i)}{\sigma_z^2+\sum_{\substack{c'\in\Phi,\\ c'\neq c}}\sum_{m'\in R_{c'}}\sum_{k'\in K_{c',m'}} s_{c',m'}^{BS}(i,k',t)\,p_{c',m'}^{BS}(i,k',t)h_{c'}^{c,m}(i)} \tag{4.1}$$

where $h_{c,m}^{BS}(i)$ is the channel gain of the direct BS-RS link on the i-th subcarrier and $h_{c'}^{c,m}(i)$ is the channel gain of the interfering link between neighbor BS (denoted by c') and the *m*-th relay of local cell (indexed by *c*). In case of channel capacity on RS-UE link, the received SINR at UE is similar to that of CHAPTER 3, which is given by

$$\Lambda_{c,m}^{RS}(i,k,t)=\frac{p_{c,m}^{RS}(i,k,t)h_{c,m}^{RS}(i,k,t)}{\sigma_z^2+\sum_{\substack{c'\in\Phi\\ c'\neq c}}\sum_{m'\in R_{c'}}\sum_{k'\in K_{c',m'}} s_{c',m'}^{RS}(i,k',t)\,p_{c',m'}^{RS}(i,k',t)h_{c',m'}^{c,m}(i,k,t)} \tag{4.2}$$

where $h_{c,m}^{RS}(i,k)$ is the channel gain on the direct RS-UE link on the *i*-th subcarrier, and $h_{c',m'}^{c,m}(i,k)$ is the gain of the interfering link between the *k*-th user in the local cell and the *m'*-th relay in the *c'*-th neighbor cell. The channel gain in BS-RS and RS-UE links takes into account the path loss, log-normal shadowing and fast fading coefficients. Note that, the fading coefficient in BS-RS link and RS-UE link follows Rician and Rayleigh distribution,

respectively, due to LoS and NLoS nature of those links. Similarly to CHAPTER 2 and CHAPTER 3, $\left\{s_{c,m}^{BS}(i,k,t), \mathrm{p}_{c,m}^{BS}(i,k,t)\right\}$ and $\left\{s_{c,m}^{RS}(i,k,t), p_{c,m}^{RS}(i,k,t)\right\}$ define the binary subcarrier and power allocation parameters at the BS-RS link and RS-UE link, respectively.

4.1.2 Queue-aware stability control

We employ the queue stability constraints and the time-slot division ratio policy, i.e. $\alpha \in (0,1)$, defined in CHAPTER 3, without any change. However, in this chapter, we highlight the fact that the stability constraints corresponding to the BS-RS link, i.e. $\frac{1}{\alpha T_s} Q_{c,m}^{BS}(k,t) \le r_{c,m}^{BS}(k,t)$, and the one in the RS-UE link, i.e. $\frac{1}{(1-\alpha)T_s} Q_{c,m}^{RS}(k,t) \le r_{c,m}^{RS}(k,t)$ can be satisfied independently from one another. This fact is one of the building blocks for developing a distributed model where the resource allocation of RS-UE link is performed independently from that of BS-RS link. As mentioned in CHAPTER 3, the proposed constraints are sufficient conditions for stability however, one can enforce the necessary condition of stability by applying an admission control policy for the incoming traffic similar to the one suggested in APPENDIX II.

4.2 Distributed subcarrier and power allocation

Regarding the fact that the stability criteria on the BS-RS link is independent from that of RS-UE link, one can divide the time-shared optimization problem in CHAPTER 3 into two resource allocation problems. Recall that by maintaining the universal frequency planning, multiple relay stations in one cell can access the same set of subcarriers during the second time-slot. Therefore, the resource allocation on RS-UE link needs to be done in a central unit in each cell, such as BS, where the CSIs of all channels on RS-UE link are known. In addition, a collaborative mechanism is required to control the inter-cell interference. However, this method requires full CSI and QSI feedback from all links to all base stations, which produces an extensive load of signaling overhead. Inspired by the interference control

method in (Son et al., 2011), we can decrease the inter-cell signaling load to the information of only one station per subcarrier. This particular station is called *reference-station* who belongs to the neighbor cell and has the largest channel gain on the interfering link. This idea originated from the worst case scenario strategy implying that by considering the worst interference case the service quality of users in better channel condition is ensured.

To deploy this idea, for each BS, identified by *c*, we define the set of neighbor BSs denoted by Φ_c^{ne}. The reference-station for *c*-th base station is the $\hat{m}$-th relay in the neighbor cell ($\hat{c} \in \Phi_c^{ne}$) that has the largest channel gain ($h_c^{\hat{c},\hat{m}}(i)$) on the *i*-th subcarrier on the interfering link. This can be formulated as

$$\hat{m} = \underset{\substack{\hat{c}\in\Phi_c^{ne} \\ \hat{m}\in \mathrm{R}_{\hat{c}} \\ \hat{k}\in \mathrm{K}_{\hat{c},\hat{m}}}}{\arg\max} \left(s_{c,m}^{BS}(i,k,t)\, s_{\hat{c},\hat{m}}^{BS}\left(i,\hat{k},t\right) h_c^{\hat{c},\hat{m}}(i) \right) \tag{4.3}$$

In Eq.(4.3), $\hat{k}$ is the user connected to the reference-relay ($\hat{m}$) on the *i*-th subcarrier.

Similarly, the reference-station for a relay is the $\hat{k}$-th UE connected to one of the neighbor relays ($\hat{m} \in \mathrm{R}_{c,m}^{ne}$) in the neighbor cell and has the largest channel gain ($h_{c,m}^{\hat{c},\hat{m}}\left(i,\hat{k},t\right)$, $\hat{c} \in \Phi_c^{ne}, \hat{k} \in \mathrm{K}_{\hat{c},\hat{m}}$) on the i-th subcarrier on the interfering link. This can be formulated as

$$\hat{k} = \underset{\substack{\hat{c}\in\Phi_c^{ne} \\ \hat{m}\in \mathrm{R}_{c,m}^{ne} \\ \hat{k}\in \mathrm{K}_{\hat{c},\hat{m}}}}{\arg\max} \left(s_{c,m}^{RS}(i,k,t)\, s_{\hat{c},\hat{m}}^{RS}\left(i,\hat{k},t\right) h_{c,m}^{\hat{c},\hat{m}}\left(i,\hat{k},t\right) \right) \tag{4.4}$$

The formula indicates that for finding the reference-user, the CSI on the interference link between users in the neighbor cell and the local relay station should be known at the central decision making station, i.e. the BS. Assuming that each relay preserves and updates the set of neighbor relays $\mathrm{R}_{c,m}^{ne}$, we can delegate the task of finding the reference-user to be done on each relay station. This provides the rudiments of a distributed scheme that assigns a part of the resource allocation problem to relay stations. Particularly, we introduce a Multi-Cell

Distributed Heuristic (MCDH) method where the subcarrier allocation task is completed in the base stations while BS and RSs individually allocate the transmission power based on the knowledge of their reference-stations, as summarized in Figure 4.2.

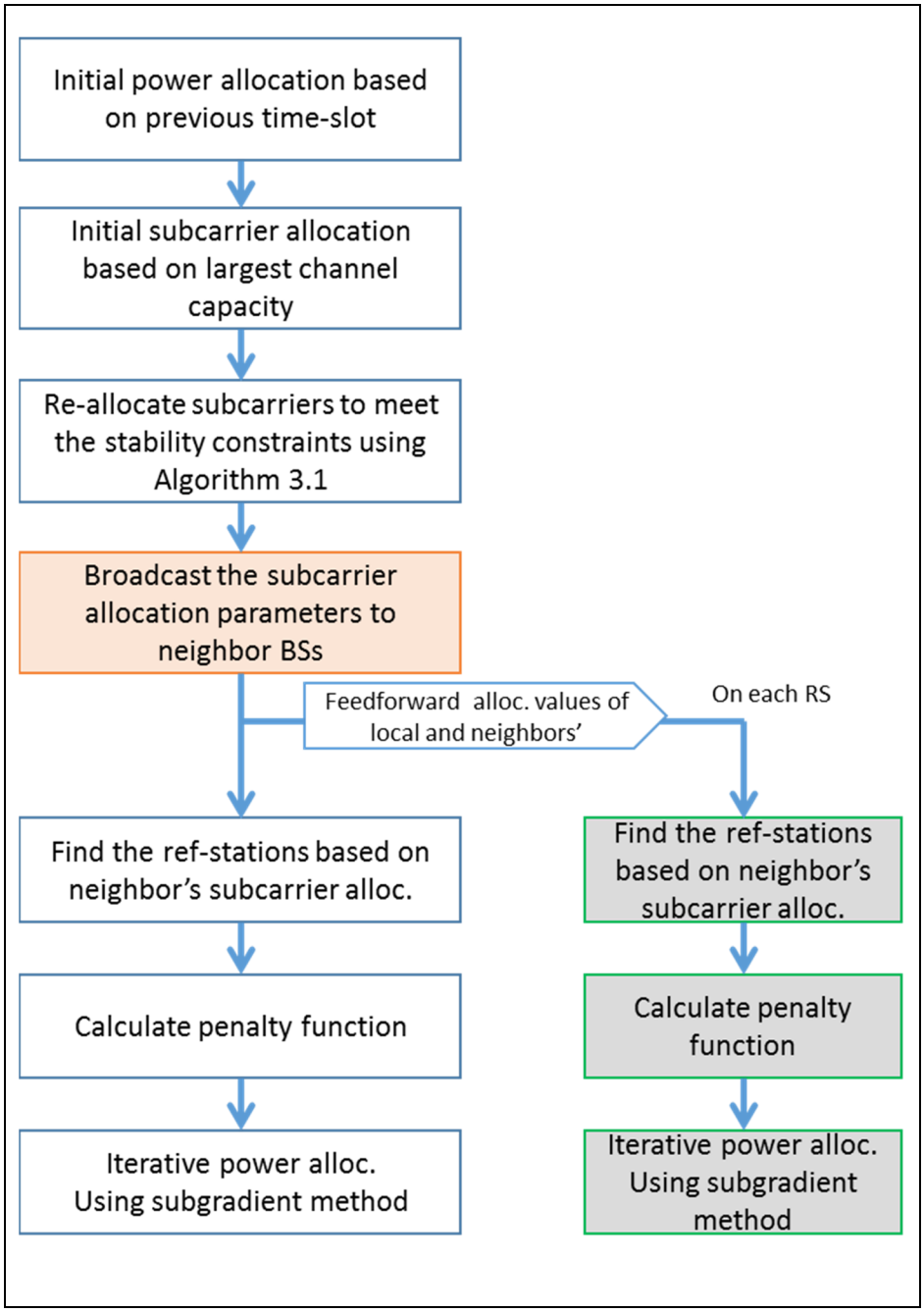

Figure 4.2 Flowchart of Multi-Cell Distributed Heuristic (MCDH) method

The first step in MCDH is the subcarrier allocation which is rooted from the phase one of TPHA method defined in CHAPTER 2. As a reminder, this procedure requires an initial power allocation to compute the channel capacity. Note that, the initial power allocation is the starting point for our allocation method and can significantly affect the final solution. It has been shown that the allocation algorithm that uses the initial power allocation based on previous time-slot outperforms the policy that uses uniform power allocation as initial power values (Son et al., 2011). In this regard, we consider the values of transmission power of the previous time-slot, i.e. $p_{c,m}^{BS}(i,k,t) = p_{c,m}^{BS}(i,k,t-1), p_{c,m}^{RS}(i,k,t) = p_{c,m}^{BS}(i,k,t-1)$, to initialize our heuristic algorithm and calculate the channel capacity. Note that for calculating the channel capacity, we assume that the channel gain, noise and the received interference of each subcarrier on both BS-RS links and RS-UE links in a cell are known at BS. It is clear that the capacity of a two-hop link is limited to the capacity of the link with lowest channel gain. Therefore, we can conduct the subcarrier allocation based on the largest minimum channel capacity of BS-RS and RS-UE links, that can be formulated as

$$\begin{cases} s_{c,m}^{BS}\left(i,k^*,t\right) = s_{c,m}^{RS}\left(i,k^*,t\right) = 1,\ k^* = \arg\max\limits_{k}\left(\min\left(C_{c,m}^{BS}\left(i,k,t\right), C_{c,m}^{RS}\left(i,k,t\right)\right)\right) \\ s_{c,m}^{BS}\left(i,k^*,t\right) = s_{c,m}^{RS}\left(i,k^*,t\right) = 0, \forall k \in \left\{1,2,..,K\right\}, k \neq k^* \end{cases}.$$

Next, the subcarrier allocation values have to be refined to meet the stability constraints. For this goal, we apply Algorithm 3.1 introduced in CHAPTER 3. Note that, the SINR formulas in BS-RS and RS-UE rate formulations in Algorithm 3.1 are updated with Eq.(4.1) and Eq.(4.2), respectively. Afterwards, each BS broadcasts the subcarrier allocation parameters corresponding to BS-RS and RS-UE links to its neighbor cells. The subsequent procedures are conducted in fully distributed fashion on each BS and RS independently.

With the subcarrier allocation information received from neighbor cells, each BS is capable of finding the reference-station using Eq.(4.3). In addition, the base station is responsible for forwarding the subcarrier allocation information of neighbor cells to its local relays. Consequently, relays are also capable of finding their reference-stations using Eq.(4.4).

Next step is to conduct a distributed power allocation mechanism. We start by defining the individual optimization problems that assign the transmission power of BS and RSs independently. The first problem corresponds to the power allocation in the BS-RS link on the first mini time-slot. The utility function is defined to maximize the sum rates on BS-RS link that is added to the rate-per-subcarrier corresponding to all reference-stations. This is given by

$$f_{BS}\left(p^{BS}\right)=\sum_{m\in \mathrm{R}_c}\sum_{k\in \mathrm{K}_{c,m}}\sum\nolimits_{i\in \mathrm{N}}\omega_{c,m}(k)s_{c,m}^{BS}\left(i,k\right)\log\left(1+\frac{p_{c,m}^{BS}\left(i,k\right)h_{c,m}^{BS}\left(i\right)}{\sigma^2+\sum_{c'\in\Phi_c^{ne}}\sum_{m'\in \mathrm{R}_{c'}}\sum_{k'\in \mathrm{K}_{c',m'}}s_{c',m'}^{BS}\left(i,k'\right)p_{c',m'}^{BS}\left(i,k'\right)h_{c'}^{c,m}\left(i\right)}\right) \tag{4.5}$$

$$+\sum_{\substack{i\in \mathrm{N}\\ \hat{c}\in\Phi_c^{ne}\\ \hat{m}\in \mathrm{R}_{\hat{c}}\\ \hat{k}\in \mathrm{K}_{\hat{c},\hat{m}}}}\log\left(1+\frac{p_{\hat{c},\hat{m}}^{BS}\left(i,\hat{\mathrm{k}}\right)h_{\hat{c},\hat{m}}^{BS}\left(i\right)}{\sigma^2+\sum_{c'\in\Phi_c^{ne}}\sum_{m'\in \mathrm{R}_{c'}}\sum_{k'\in \mathrm{K}_{c',m'}}s_{c',m'}^{BS}\left(i,k'\right)p_{c',m'}^{BS}\left(i,k'\right)h_{c'}^{\hat{c},\hat{m}}\left(i\right)}\right)$$

For the sake of simplicity, we eliminated the time-slot index t from Eq.(4.5) and the remaining formulations since the MCDH procedure is reoccurring per time-slot. Ultimately, the optimization problem can be defined as

$$\max_{\left\{p^{BS}\right\}} f_{BS}\left(p^{BS}\right) \tag{4.6}$$

subject to the set of constraints as defined in CHAPTER 3, given as

$$p_{c,m}^{BS}(i,k)\geq 0, \qquad \forall i\in N,\forall m\in \mathrm{R}_c,\forall \mathrm{k}\in \mathrm{K}_{c,m}, \tag{4.7}$$

$$\sum_{m\in \mathrm{R}_c}\sum_{k\in \mathrm{K}_{c,m}}\sum_{i\in N}s_{c,m}^{BS}(i,k)p_{c,m}^{BS}\left(i,k\right)\leq p_{BS}^{max}, \tag{4.8}$$

$$\frac{Q_{c,m}^{BS}(k)}{\alpha T_s\Delta f}\leq\sum\nolimits_{i\in \mathrm{N}}s_{c,m}^{BS}\left(i,k\right)\log\left(1+\frac{p_{c,m}^{BS}\left(i,k\right)h_{c,m}^{BS}\left(i\right)}{\sigma^2+\sum_{c'\in\Phi_c^{ne}}\sum_{m'\in \mathrm{R}_{c'}}\sum_{k'\in \mathrm{K}_{c',m'}}s_{c',m'}^{BS}\left(i,k'\right)p_{c',m'}^{BS}\left(i,k'\right)h_{c'}^{c,m}\left(i\right)}\right), \qquad \forall m\in \mathrm{R}_c,k\in \mathrm{K}_{c,m} \tag{4.9}$$

The second problem corresponds to the optimization of the transmission power per subcarriers on RS-UE link that is calculated by each relay station in the second mini time-

slot. The utility function of the RS-UE link can be defined in the same way as the BS-RS link, and is given by

$$f_{RS}\left(p^{RS}\right)=\sum_{k\in \mathrm{K}_{c,m}}\sum\nolimits_{i\in \mathrm{N}}\omega_{c,m}(k)s_{c,m}^{RS}\left(i,k\right)\log\left(1+\frac{p_{c,m}^{RS}\left(i,k\right)h_{c,m}^{BS}\left(i,k\right)}{\sigma^{2}+\sum\limits_{\substack{m'\in \mathrm{R}_{c,m}^{ne}\\ m'\in \mathrm{R}_{c'}\\ c'\in\Phi_{c}^{ne}}}\sum\limits_{k'\in \mathrm{K}_{c',m'}}s_{c',m'}^{RS}\left(i,k'\right)p_{c',m'}^{RS}\left(i,k'\right)h_{c',m'}^{c,m}\left(i,k\right)}\right) \tag{4.10}$$

$$+\sum_{\substack{i\in \mathrm{N}\\ \hat{c}\in\Phi_{c}^{ne}\\ \hat{m}\in \mathrm{R}_{c,m}^{ne}\\ \hat{k}\in \mathrm{K}_{\hat{c},\hat{m}}}}\log\left(1+\frac{p_{\hat{c},\hat{m}}^{RS}\left(i,\hat{\mathrm{k}}\right)h_{\hat{c},\hat{m}}^{BS}\left(i,\hat{\mathrm{k}}\right)}{\sigma^{2}+\sum\limits_{\substack{m'\in \mathrm{R}_{c,m}^{ne}\\ m'\in \mathrm{R}_{c'}\\ c'\in\Phi_{c}^{ne}}}\sum\limits_{k'\in \mathrm{K}_{c',m'}}s_{c',m'}^{RS}\left(i,k'\right)p_{c',m'}^{RS}\left(i,k'\right)h_{c',m'}^{\hat{c},\hat{m}}\left(i,\hat{\mathrm{k}}\right)}\right)$$

Note that in (4.10), each relay has a list of neighbor relays, i.e. $\left\{m'\middle|m'\in \mathrm{R}_{c,m}^{ne},m'\in \mathrm{R}_{c'},c'\in\Phi_{c}^{ne}\right\}$, and requires only considering the interference and reference-user among those stations. This helps to reduce the problem size and computation burden on relay stations. Therefore the maximization problem is defined by

$$\max_{\left\{p^{RS}\right\}} f_{RS}\left(p^{RS}\right) \tag{4.11}$$

subject to the set of constraints, similarly defined in CHAPTER 3 and are given by

$$p_{c,m}^{RS}(i,k)\geq 0,\qquad \forall i\in N,\forall m\in \mathrm{R}_{c},\forall \mathrm{k}\in \mathrm{K}_{c,m}, \tag{4.12}$$

$$\sum_{m\in \mathrm{R}_{c}}\sum_{k\in \mathrm{K}_{c,m}}\sum_{i\in N}s_{c,m}^{RS}(i,k)p_{c,m}^{RS}\left(i,k\right)\leq p_{\max}^{RS}, \tag{4.13}$$

$$\frac{Q_{c,m}^{RS}(k)}{(1-\alpha)T_{s}\Delta f}\leq\sum\nolimits_{i\in \mathrm{N}}s_{c,m}^{RS}\left(i,k\right)\log\left(1+\frac{p_{c,m}^{RS}\left(i,k\right)h_{c,m}^{RS}\left(i,k\right)}{\sigma^{2}+\sum\limits_{\substack{m'\in \mathrm{R}_{c,m}^{ne}\\ m'\in \mathrm{R}_{c'}\\ c'\in\Phi_{c}^{ne}}}\sum\limits_{k'\in \mathrm{K}_{c',m'}}s_{c',m'}^{RS}\left(i,k'\right)p_{c',m'}^{RS}\left(i,k'\right)h_{c',m'}^{c,m}\left(i,k\right)}\right),\qquad \forall k\in \mathrm{K}_{m}. \tag{4.14}$$

For a given set of subcarrier allocation values $\{s^{BS}, s^{RS}\}$ and the information corresponding to reference-stations, the optimal transmission power can be calculated by solving the dual problem. The dual problem at the BS is given by

$$\mathrm{L}_c(\lambda,\beta,\mathrm{p}) = f_{BS}(p) - \lambda_c\left[\sum_{i\in\mathrm{N}} p_{c,m}^{BS}(i,k)\, s_{c,m}^{BS}(i,k) - P_{\max}^{BS}\right]$$
$$-\sum_{k}\beta_{c,\mathrm{m},k}\left[\frac{Q_{c,m}^{BS}(i,k)}{\alpha T_s \Delta f} - \sum_{i\in\mathrm{N}} s_{c,m}^{BS}(i,k)\log\left(1+\frac{p_{c,m}^{BS}(i,k)\, h_{c,m}^{BS}(i)}{\sigma_z^2 + \sum_{c'\in\Phi_c^{ne}}\sum_{m'\in\mathrm{R}_{c'}}\sum_{k'\in\mathrm{K}_{c',m'}} s_{c',m'}^{BS}(i,k')\, p_{c',m'}^{BS}(i,k')\, h_{c'}^{c,m}(i)}\right)\right]$$

where λ_c and $\beta_{c,m,k}$ are non-negative Lagrangian multipliers related to constraints Eq.(4.8) and Eq.(4.9), respectively. Similarly, the dual problem to be solved at each relay station is specified as

$$\mathrm{L}_m(\lambda,\beta,\mathrm{p}) = f_{RS}(p) - \lambda_m\left[\sum_{i\in\mathrm{N}} p_{c,m}^{RS}(i,k)\, s_{c,m}^{RS}(i,k) - P_{\max}^{RS}\right]$$
$$-\sum_{k}\beta_{\mathrm{m},k}\left[\frac{Q_{c,m}^{RS}(i,k)}{(1-\alpha)T_s\Delta f} - \sum_{i\in\mathrm{N}} s_{c,m}^{RS}(i,k)\log\left(1+\frac{p_{c,m}^{RS}(i,k)\, h_{c,m}^{RS}(i,k)}{\sigma_z^2 + \sum_{\substack{m'\in\mathrm{R}_{c,m}^{ne}\\ m'\in\mathrm{R}_{c'}\\ c'\in\Phi_c^{ne}}}\sum_{k'\in\mathrm{K}_{c',m'}} s_{c',m'}^{RS}(i,k')\, p_{c',m'}^{RS}(i,k')\, h_{c',\mathrm{m}'}^{c,m}(i,k)}\right)\right]$$

where λ_m and $\beta_{m,k}$ are non-negative Lagrangian multipliers related to constraints Eq.(4.13) and Eq.(4.14), respectively. To satisfy KKT conditions (Boyd & Vandenberghe, 2004), we set the derivative of dual functions to zero, i.e. $\frac{\partial \mathrm{L}_c}{\partial p_{c,m}^{BS}(i,k)} = 0$ and $\frac{\partial \mathrm{L}_m}{\partial p_{c,m}^{RS}(i,k)} = 0$.

After some manipulations, the optimal transmission power on BS-RS and RS-UE link are respectively written as

$$p_{c,m}^{*BS}(i,k) = \left[\frac{\omega(i,k)+\beta_{c,m,k}}{\lambda_c \ln 2 + \Psi_{c,m}^{BS}(i,\hat{k})} - \frac{\sigma_z^2 + \sum_{c'\in\Phi_c^{ne}}\sum_{m'\in\mathrm{R}_{c'}}\sum_{k'\in\mathrm{K}_{c',m'}} s_{c',m'}^{BS}(i,k')\, p_{c',m'}^{BS}(i,k')\, h_{c'}^{c,m}(i)}{h_{c,m}^{BS}(i)}\right]^{+}, \quad (4.15)$$

$$p_{c,m}^{*RS}(i,k)=\left[\frac{\omega(i,k)+\beta_{m,k}}{\lambda_m \ln 2+\Psi_{c,m}^{RS}\left(i,\hat{k}\right)}-\frac{\sigma_z^2+\sum_{\substack{m'\in R_{c,m}^{ne}\\ m'\in R_{c'}\\ c'\in\Phi_c^{ne}}}\sum_{k'\in K_{c',m'}} s_{c',m'}^{RS}(i,k')\,p_{c',m'}^{RS}(i,k')\,h_{c',m'}^{c,m}(i,k)}{h_{c,m}^{RS}(i,k)}\right]^{+}. \tag{4.16}$$

By applying $[x]^{+}$ (which equals x, if $x>0$, otherwise 0), we ensure the optimal power value is non-negative. $\Psi_{c,m}^{BS}\left(i,\hat{k}\right), \Psi_{c,m}^{RS}\left(i,\hat{k}\right)$ are called penalty functions formulated as

$$\Psi_{c,m}^{BS}\left(i,\hat{k}\right)=\frac{h_c^{\hat{c},m}(i)\,\Lambda_{\hat{c},m}^{BS}\left(i,\hat{k}\right)}{\mathrm{I}_{\hat{c},m}^{BS}\left(i,\hat{k}\right)\left(\Lambda_{\hat{c},m}^{BS}\left(i,\hat{k}\right)+1\right)},$$

$$\Psi_{c,m}^{RS}\left(i,\hat{k}\right)=\frac{h_{c,m}^{\hat{c},m}\left(i,\hat{k}\right)\Lambda_{c,m}^{RS}\left(i,\hat{k}\right)}{\mathrm{I}_{\hat{c},m}^{RS}\left(i,\hat{k}\right)\left(\Lambda_{c,m}^{RS}\left(i,\hat{k}\right)+1\right)}.$$

The received interference plus noise on the *i*-th subcarrier corresponding to the reference-station of BS-RS and RS-UE links, respectively are given by

$$\mathrm{I}_{\hat{c},m}^{BS}\left(i,\hat{k}\right)=\sigma_z^2+\sum_{c'\in\Phi_c^{ne}}\sum_{m'\in R_{c'}}\sum_{k'\in K_{c',m'}} s_{c',m'}^{BS}(i,k')\,p_{c',m'}^{BS}(i,k')\,h_{c'}^{\hat{c},m}(i),$$

$$\mathrm{I}_{\hat{c},m}^{RS}\left(i,\hat{k}\right)=\sigma_z^2+\sum_{\substack{m'\in R_{\hat{c},m}^{ne}\\ m'\in R_{c'}\\ c'\in\Phi_c^{ne}}}\sum_{k'\in K_{c',m'}} s_{c',m'}^{RS}(i,k')\,p_{c',m'}^{RS}(i,k')\,h_{c',m'}^{\hat{c},m}\left(i,\hat{k}\right).$$

It can be seen that the formulations in Eq.(4.15) and Eq.(4.16) illustrate a water-filing power allocation problem. In particular, a larger value of penalty function results in lower power level to reduce the interference impact on the reference-station. The complementary slackness conditions of the optimization problem on BS-RS link are defined as follows

$$\lambda_c\left[\sum_{i\in N} p_{c,m}^{*BS}(i,k)\,s_{c,m}^{BS}(i,k)-P_{\max}^{BS}\right]=0, \tag{4.17}$$

$$\beta_{c,m,k}\left[\frac{Q_{c,m}^{BS}(i,k)}{\alpha T_s \Delta f}-\sum_{i\in \mathrm{N}} s_{c,m}^{BS}(i,k)\log\left(1+\frac{p_{c,m}^{*BS}(i,k)h_{c,m}^{BS}(i)}{\sigma_z^2+\sum_{c'\in\Phi_c^{ne}}\sum_{m'\in \mathrm{R}_{c'}}\sum_{k'\in \mathrm{K}_{c',m'}} s_{c',m'}^{BS}(i,k')\,p_{c',m'}^{BS}(i,k')h_{c'}^{c,m}(i)}\right)\right]=0, \tag{4.18}$$

and the ones corresponding to RS-UE link are given by

$$\lambda_m\left[\sum_{i\in \mathrm{N}} p_{c,m}^{*RS}(i,k)\,s_{c,m}^{RS}(i,k)-P_{\max}^{RS}\right]=0, \tag{4.19}$$

$$\beta_{m,k}\left[\frac{Q_{c,m}^{RS}(i,k)}{(1-\alpha)T_s \Delta f}-\sum_{i\in \mathrm{N}} s_{c,m}^{RS}(i,k)\log\left(1+\frac{p_{c,m}^{*RS}(i,k)h_{c,m}^{RS}(i,k)}{\sigma_z^2+\sum_{\substack{m'\in \mathrm{R}_{c,m}^{ne}\\ m'\in \mathrm{R}_{c'}\\ c'\in\Phi_c^{ne}}}\sum_{k'\in \mathrm{K}_{c',m'}} s_{c',m'}^{RS}(i,k')\,p_{c',m'}^{RS}(i,k')h_{c',\mathrm{m}'}^{c,m}(i,k)}\right)\right]=0. \tag{4.20}$$

The defined slackness conditions in Eq.(4.17) and Eq.(4.18) (similarly Eq.(4.19) and Eq.(4.20) for RS problem) are required to be satisfied in the optimal solution. Note that the defined optimization problems Eq.(4.6) and Eq.(4.11) are not convex, thus the slackness conditions are not sufficient for the convergence to the global optimum (Boyd & Vandenberghe, 2004). On the other hand, it is shown that the duality gap tends to zero with very large number of subcarriers (Yu & Lui, 2006). In the following, we define an efficient iterative algorithm that rapidly converges to a local optimum with a close approximation to the global optimum solution.

4.2.1 Iterative power allocation mechanism

Starting from the initial power allocation set mentioned earlier in this chapter, we first define the subgradient upgrade formula for BS power allocation that is given by the following update rules

$$\lambda_c[n+1]=\left[\lambda_c[n]-\upsilon_\lambda[n]\left(P_{\max}^{BS}-\sum_{i\in \mathrm{N}} p_{c,m}^{*BS}(i,k)\,s_{c,m}^{BS}(i,k)\right)\right]^{+}, \tag{4.21}$$

$$\beta_{c,m,k}[n+1]=\left[\beta_{c,m,k}[n]-\upsilon_{\beta}[n]\left[\sum_{i\in \mathrm{N}} s_{c,m}^{BS}(i,k)\log\left(1+\frac{p_{c,m}^{*BS}(i,k)h_{c,m}^{BS}(i)}{\sigma_z^2+\sum_{c'\in\Phi_c^{ne}}\sum_{m'\in \mathrm{R}_{c'}}\sum_{k'\in \mathrm{K}_{c',m'}} s_{c',m'}^{BS}(i,k')\,p_{c',m'}^{BS}(i,k')h_{c'}^{c,m}(i)}\right)-\frac{Q_{c,m}^{BS}(i,k)}{\alpha T_s\Delta f}\right]\right]. \quad (4.22)$$

Likewise, the update rule for the power allocation problem in the RS is given by

$$\lambda_m[n+1]=\left[\lambda_m[n]-\upsilon_{\lambda}[n]\left(P_{\max}^{RS}-\sum_{i\in \mathrm{N}} p_{c,m}^{*RS}(i,k)\,s_{c,m}^{RS}(i,k)\right)\right]^{+}, \quad (4.23)$$

$$\beta_{m,k}[n+1]=\left[\beta_{m,k}[n]-\upsilon_{\beta}[n]\left[\sum_{i\in \mathrm{N}} s_{c,m}^{RS}(i,k)\log\left(1+\frac{p_{c,m}^{*RS}(i,k)h_{c,m}^{RS}(i,k)}{\sigma_z^2+\sum_{\substack{m'\in \mathrm{R}_{c,m}^{ne}\\ m'\in \mathrm{R}_{c'}\\ c'\in\Phi_c^{ne}}}\sum_{k'\in \mathrm{K}_{c',m'}} s_{c',m'}^{RS}(i,k')\,p_{c',m'}^{RS}(i,k')h_{c',\mathrm{m}}^{c,m}(i,k)}\right)-\frac{Q_{c,m}^{RS}(i,k)}{(1-\alpha)T_s\Delta f}\right]\right]. \quad (4.24)$$

In Eq.(4.21) and Eq.(4.22) (likewise Eq.(4.23) and Eq.(4.24)) the step size vectors at *n+1* iteration is denoted by $\upsilon_{\lambda}[n+1]$ and $\upsilon_{\beta}[n+1]$, respectively. It is known that the iterative algorithm converges to the optimal solution if the chosen step sizes satisfy the infinite travel condition (Boyd, Xiao, & Mutapcic, 2003), i.e.

$$\sum_{n=1}^{\infty}\upsilon_u[n]=\infty, \mathrm{u}\in\{\lambda,\beta\}.$$

Therefore, we define the step size as $\upsilon_u[n]=\frac{b}{\sqrt{n}}, u\in\{\beta,\lambda\}$ where b is a constant. This is known as diminishing step size rule (Boyd et al., 2003) and it has been shown that this rule has a better convergence speed and precision than other rules employed in spectrum optimization problems (Yu & Lui, 2006).

Algorithm 4.1 Iterative Power allocation

1: Initialize $p_{c,m}^{u}, u \in BS, RS$, Lagrangian multipliers λ, β , maximum number of iteration n_{max} and set $n = 1$
2: **while** (4.17) and (4.18) for BS-RS allocation are not satisfied ((4.19) and (4.20) for RS-UE allocation) **do**
3: Calculate power using (4.15) ((4.16) for RS-UE allocation)
4: update Lagrangians using (4.21),(4.22)((4.23),(4.24) for RS-UE allocation)
5: $n = n + 1$
6: update step size
7: **end while**

The iterative algorithm continuous updating λ, β and $p_{c,m}^{*u}(i,k), u \in \{BS, RS\}$ respectively, using the defined updates rules, until Eq.(4.17) and Eq.(4.18) (similarly Eq.(4.19) and Eq.(4.20) on RS equations) become satisfied. The iterative algorithm, which runs independently on all BS and RSs, is summarized in Algorithm 4.1. Based on our observation in numerical simulations, the iterative power allocation model running in BS and RS converges while the average number of iterations equals 3. This fast convergence highlights the practicality of this algorithm for designing larger networks.

4.3 Further reducing the signaling overhead

To efficiently reduce the messaging overhead, we first identify the messaging phases deployed in MCDH method. Initially, the relays are required to feedback the CSI received from UEs in addition to their own CSIs and QSIs to the base station. Note that, each BS and RS is required to maintain a data structure to save and update the CSIs of neighbors as well as CSIs and QSIs of local stations. Depending on the mobility of each station, the frequency of sending those feedback messages can be reduced by averaging the channel capacity and sending the results in larger periods than one time-slot. This idea is applicable for sending the averaged CSI of the direct links as well as the averaged CSI of the interfering links. For the

case of QSI, the size and feedback frequency can be reduced by reporting only the changes of queue length and reporting once it has been changed[6].

The next signaling phase is related to the broadcast of subcarrier allocation parameters by base stations and is called feedforward messaging. One way to reduce the volume of data in this phase is to only send indexes of scheduled users. Moreover, the broadcast messages can be securely transmitted using the backhaul network that connects all the base stations that is mostly fiber optic-based and incurs extremely low latency. This event can coincide with sending the CSI data to neighbor base stations that is required for identifying the reference-user and calculating the penalty function.

It is important to elaborate the details of the required CSI data from neighbor cells. The first part of this data is the channel gain of the interfering link on each subcarrier that is necessary for finding the reference-station. The second part is a data tuple related to the reference-station of each subcarrier i.e. $\left[\Lambda^{u}_{\hat{c},m}\left(i,\hat{k}\right),\mathrm{I}^{u}_{\hat{c},m}\left(i,\hat{k}\right)\right], u\in\{BS,RS\}$, which is required for calculating the penalty function. Although the volume of the required CSI from neighbor cells is noticeable, we can limit it by sending the information corresponding to only the scheduled users $\left\{k^{*}\middle| s^{u}_{c,m}\left(i,k^{*}\right)=1, u\in\{BS,RS\}\right\}$.

For relay stations to be informed via feedforward messages, the base station can integrate several data in one multi-purpose message and send it via downlink control signals to the relay stations. To achieve this goal, the multi-purpose message can include the subcarrier allocation decision of the local RS-UE links and those of neighbor cells in addition to CSI data of the scheduled users from neighbor relays. Table 4.1 presents the calculated messaging

[6] It is worth to mention that, for maintaining an acceptable performance level, it is crucial to dynamically change the feedback period which becomes challenging when mobile nodes are moving with high speed. This issue is out of scope of the current text but interested readers can find more information in (N. Lee & Heath, 2014).

Table 4.1 Comparison of messaging load of MCDH and the centralized method

Message content	MCDH method	Centralized method
direct BS-RS: CSI per subcarr.	$\lvert\mathcal{R}_c\rvert$ †	$\lvert\mathcal{R}_c\rvert\lvert\Phi\rvert$
direct RS-UE: CSI per subcarr.	$2\lvert\mathcal{K}_c\rvert$ †	$2\lvert\mathcal{K}_c\rvert\lvert\Phi\rvert$
interfering BS-RS: CSI per subcarr.	Combined data per scheduled user: $\lvert\Phi_c^{ne}\rvert$ †	$\lvert\mathcal{R}_c\rvert\lvert\Phi(\Phi-1)\rvert$
interfering RS-UE: CSI per subcarr.	Combined data per scheduled user: $2\lvert\Phi_c^{ne}\rvert$ †	$\lvert\mathcal{K}_c\rvert\lvert\mathcal{R}_c\rvert\lvert\Phi(\Phi-1)\rvert$
QSI at BS	_	$\lvert\mathcal{K}_c\rvert\lvert\Phi\rvert$
QSI at RS	$\lvert\mathcal{K}_c\rvert$ †	$2\lvert\mathcal{K}_c\rvert\lvert\Phi\rvert$
BS-RS Subcarr. alloc. (per subcarrier)	Broadcast to neighbors:$\lvert\Phi_c^{ne}\rvert$	$\lvert\mathcal{K}_c\rvert\lvert\Phi\rvert$
RS-UE Subcarr. alloc. (per subcarrier)	• to local relays:$\lvert\mathcal{K}_c\rvert+\lvert\Phi_c^{ne}\rvert$ • Broadcast to neighbors:$\lvert\Phi_c^{ne}\rvert$	$2\lvert\mathcal{K}_c\rvert\lvert\Phi\rvert$
BS-RS Power alloc. per subcarrier	_	$\lvert\mathcal{K}_c\rvert\lvert\Phi\rvert$
RS-UE Power aloc. per subcarrier	_	$2\lvert\mathcal{K}_c\rvert\lvert\Phi\rvert$

(†) indicates that the feedback period can be larger than one time-slot. $\lvert\Phi\rvert$is the size of network, $\lvert\Phi_c^{ne}\rvert$ denotes the number of neighbor BSs. The broadcast of subcarrier allocation decision is limited to the index of scheduled users. Combined data on the interfering link implies the combination of channel gain, SINR and received interference plus noise. A numeric multiplier in front of some formulations resulted from the number of hops over which the feedback message is traveling.

load required for MCDH method and compares it with the centralized scheme defined in CHAPTER 3.

For the sake of comparison, we present the signaling overhead of the proposed centralized and distributed methods with a semi-distributed relaying allocation scheme (Ng & Schober, 2011). We add up the messages mentioned in rows of Table 4.1 and show the average messaging load per cell in Figure 4.3. For a numerical instance, consider the unit of data is one byte for a system with 3 relays per cell and 32 subcarriers, thus in MCDH method the required CSI corresponding to the direct BS-RS links is 96 bytes and $3*7*32=672$ bytes in the centralized method.

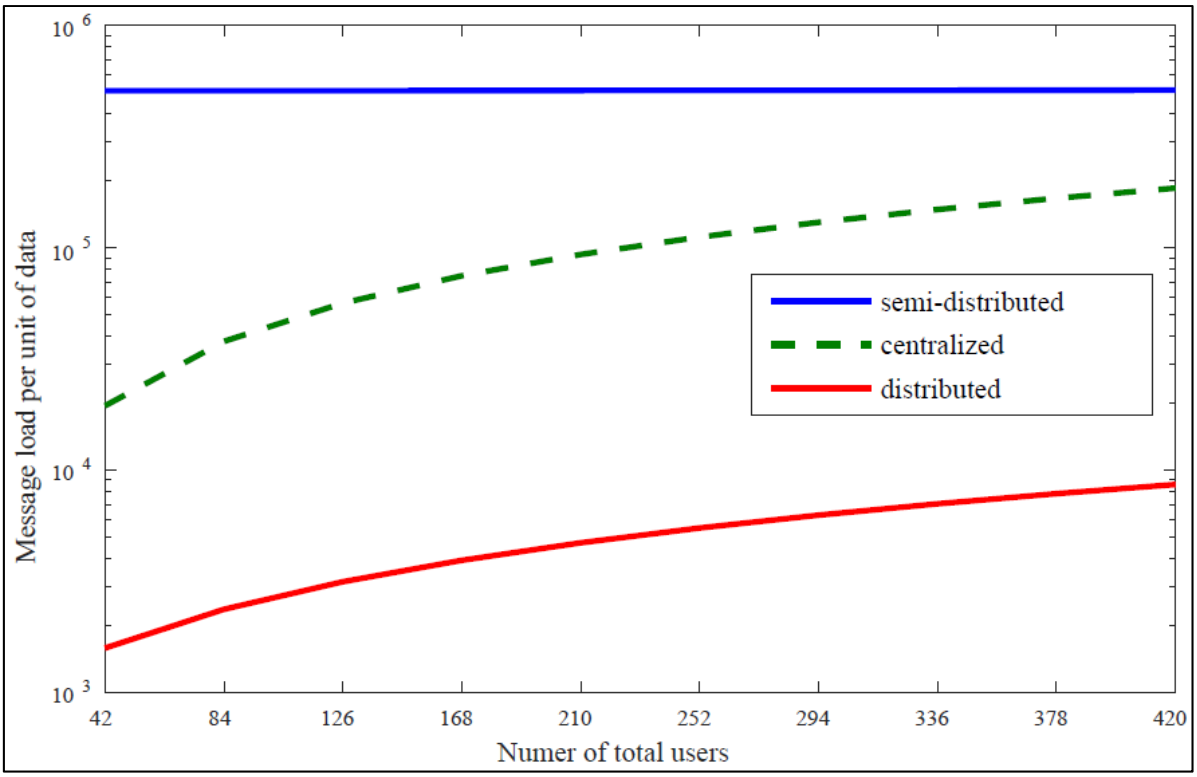

Figure 4.3 Message load in general unit of data (e.g. one byte) versus the number of total users in the system, applying the centralized, distributed and semi-distributed methods

It can be seen in Figure 4.3 that the proposed distributed method can significantly reduce the messaging overhead in comparison to the centralized and semi-distributed methods. In particular, in the semi-distributed method the cooperative cells are required to iteratively transfer allocation values till reaching to the optimal value. Authors of (Ng & Schober, 2011) reported that, the number of iterations required to reach to 90% of optimal value is 10. The reason that the messaging load of the semi-distributed method is almost steady is that it is rather dependent on the number of iterations and subcarriers than the number of users in the system.

4.4 Numerical results and performance evaluation

The system configuration for numerical analysis is similar to CHAPTER 3 except that we decrease the cell dimension in order to analyze the impact of interference between base

station and neighbor relays. In this regard, the distance between BS to the local RS and to the neighbor RS is 1000 m and 1320 m, respectively. The distance on direct RS-UE link and interfering RS-UE link is 100 m and 220 m, respectively. The channel coefficients on the link between BS and neighbor relays are modeled with Rician fading, lognormal shadowing and path loss for urban macrocell model, identical to that of direct BS-RS link defined in CHAPTER 3. The system consists of 7 BSs, the number of users per relay is 4 and the number of relays per cell is 3. For comparison purpose, we illustrate the performance evaluation of the centralized method defined in CHAPTER 3 named the Time-shared model. Final note is that the number of available OFDMA subcarriers per link is 16, the bandwidth of each subcarrier is 15 kHz, the arrival rate is 5 kbps, and the default time-slot division ratio is 0.5.

Because the channel samples are drawn from probabilistic distributions, we demonstrated the numerical results by averaging the performance values over 100 independent realization of the simulated system.

4.4.1 System throughput

The first important factor is to check the achieved throughput that is the sum of minimum rates on BS-RS and RS-UE links in the system. In Figure 4.4 we present the average throughput per subcarrier achieved via MCDH and the centralized time-shared method of CHAPTER 3. In this experiment we equally increase the maximum power limit in BS and RS (i.e. $P_{BS}^{\max} = P_{RS}^{\max}$), which accordingly increases the received interference on the neighbor BSs and RSs.

The first observation of Figure 4.4 indicates that the worst case interference consideration in the calculation of MCDH scheme results in lower throughput than the centralized method. The second observation is that the growth of power limit does not contribute to significant increase of throughput in either centralized or distributed methods. This is expected because

as much the power increases, the interference also increments which prevent the data rate to be improved.

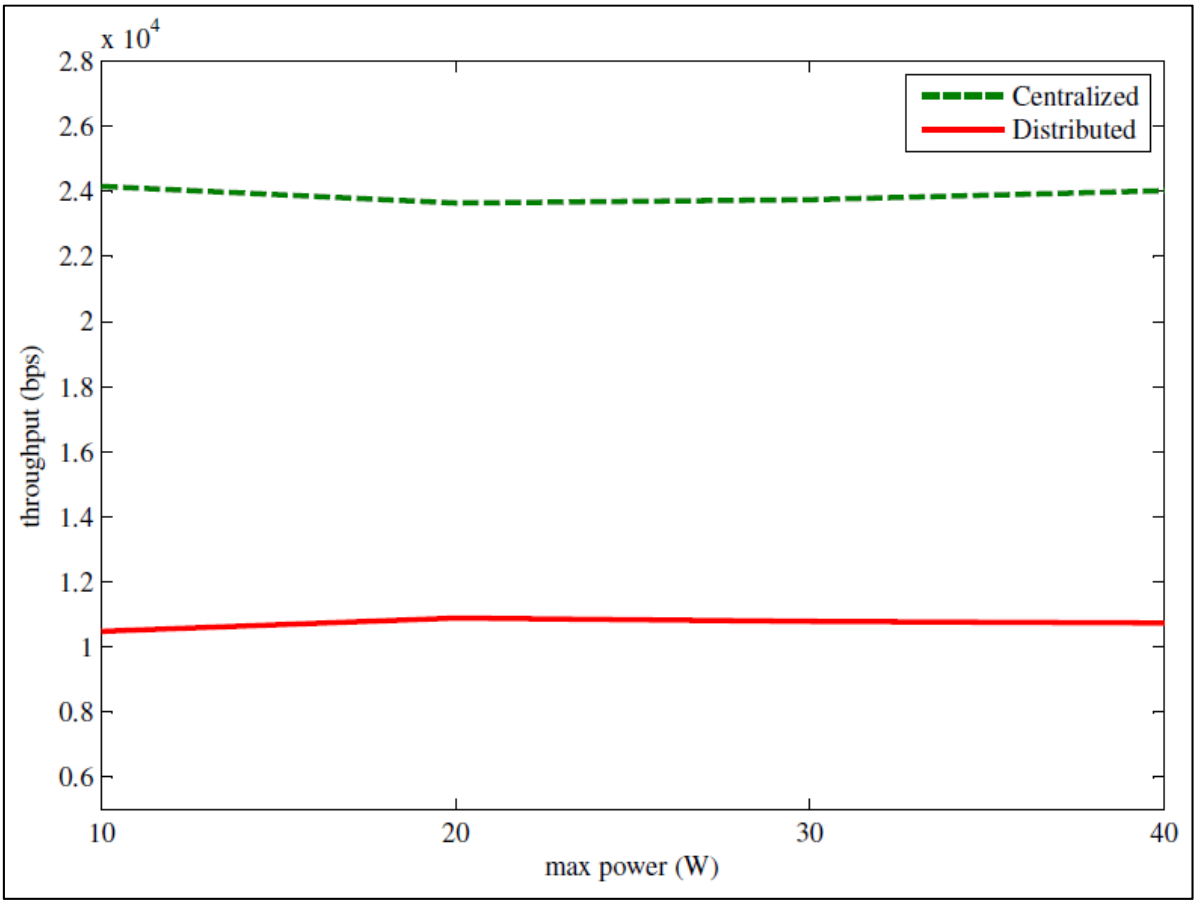

Figure 4.4 Average system throughput per subcarrier, applying MCDH method and the centralized time-shared method

4.4.2 Energy efficiency

We apply the same scheme defined in 3.7.2 to calculate the energy efficiency of the proposed distributed method and compare it with the centralized time-shared scheme, that is illustrated in Figure 4.5. It can be seen that the MCDH approach significantly improves the energy efficiency in the system comparing to the centralized time-shared method when the power limit ranges from 5 W to 25 W. This is due the fact that the reference-station technique in

MCDH, adapts the power allocation values with better estimation of interference condition than the fixed interference temperature applied in centralized method.

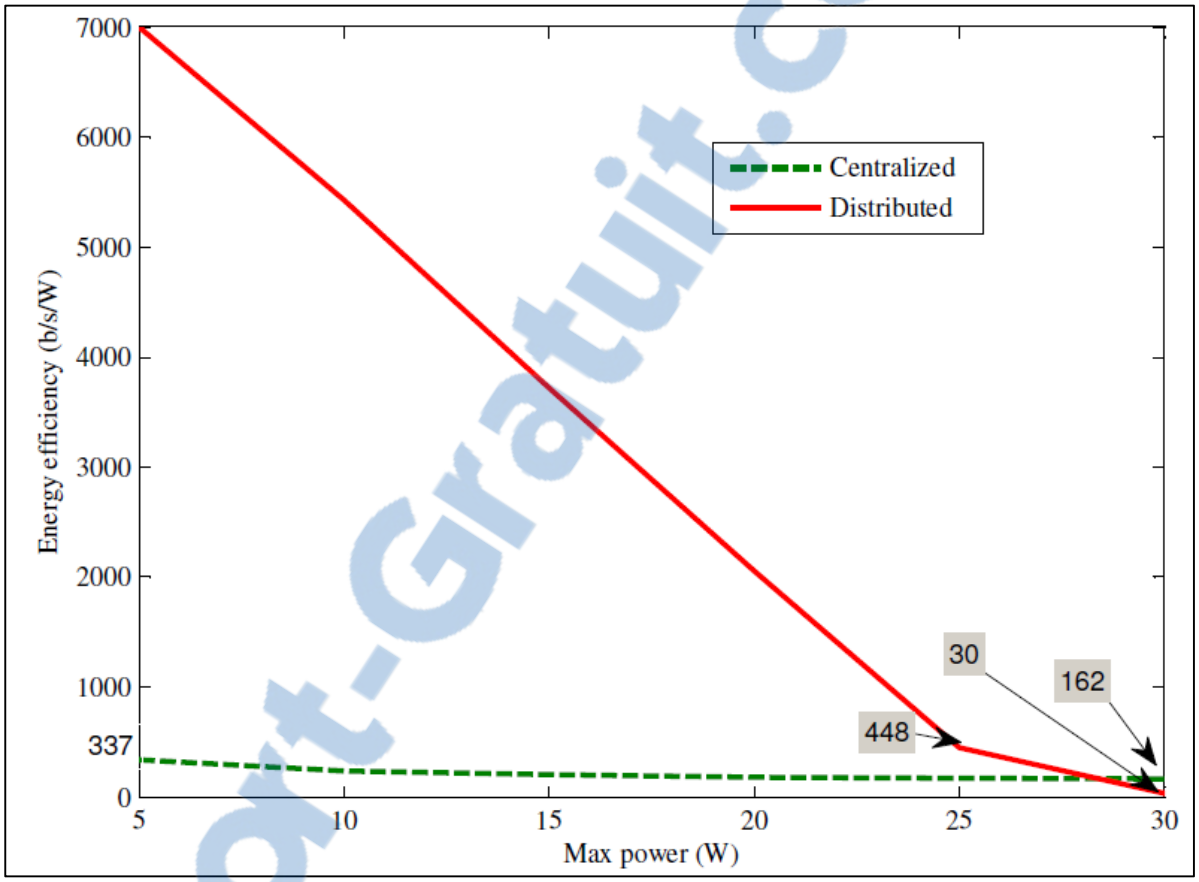

Figure 4.5 Energy consumption per bit, i.e. energy efficiency, applying the MCDH and centralized time-shared method

As it has been observed previously in numerical results of CHAPTER 3, the increase of power limit degrades the system energy efficiency. This degradation of energy efficiency is more abrupt in MCDH method than in the centralized one. In addition, it can be seen that the power limit of 30 W is a turning point for the distributed method and it cannot offer better energy efficiency for larger power limits comparing to the centralized method. This is expected because the worst-case interference becomes impractical when interference from reference-station is significantly larger than the remaining interfering stations. It can be

concluded that the efficiency of the MCDH method is better when applied in low power scenarios.

4.4.3 System stability

We examine the stability feature of the proposed method by calculating the length of the queues for MCDH and the centralized method. In Figure 4.6 we depict the results corresponding to the average of sum queue lengths in BS and RSs in the central cell, as depicted in Figure 4.1, during 200 consecutive time-slots.

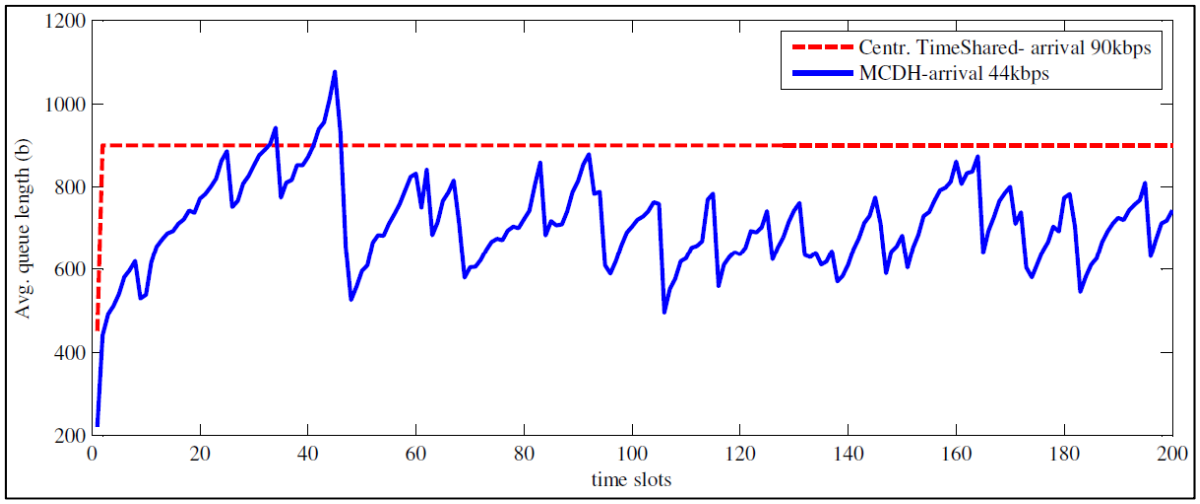

Figure 4.6 Evolution of averaged queue length of users in the central cell during 200 consecutive time-slots, applying MCDH and centralized methods

Considering the defined system configuration, the largest arrival rate under which the proposed distributed method could stabilize the queues is 44 kbps. However, the centralized method can stabilize the system up to the arrival rate of 90Kbps. Therefore, MCDH has less robustness for higher data rates traffic than the centralized method. This reveals the fact that using MCDH method reduces the stability region compared to the centralized method.

4.5 Conclusion

In this chapter, we discussed a new idea to enhance the resource allocation model proposed in CHAPTER 3 in order to improve its scalability and decrease the CSI and QSI feedback overhead. To achieve this goal we proposed a multi-cell distributed heuristic method in which the resource allocation task is run partially on BSs and RSs. In particular, the first part of allocation process corresponds to assigning subcarriers to both BS-RS and RS-UE links that is performed by each base station. The second part corresponds to the optimum power allocation that is performed independently on BS and RSs in each cell.

In addition, we generalized the interference model previously defined in CHAPTER 3 by including the received interference on the relay stations in the BS-RS link model. One of the enhancements of the MCDH is that it limits the inter-cell messaging overhead to the neighbor cells which are considered as the potential sources of interference. Moreover, the interference control is deployed based on the technique of reference-station, which is a station that receives the most significant interference on each subcarrier. This idea is based on abstracting the interference model with the worst case scenario which substantially reduces the computation and signaling overhead. To further reduce the feedback overhead, we also defined methods to decrease the size and frequency of sending feedback messages.

The numerical results show a significant improvement in terms of power consumption thanks to the efficient interference control mechanism in MCDH. It is also observed that the MCDH method can sustain the system stability although the stability range of MCDH is noticeably smaller than the centralized method.

The other benefit of MCDH method is its high convergence speed which is a strong motivation to make it a good replacement for centralized allocation schemes for highly populated and large networks. In essence, the MCDH method offers low signaling overhead, high power-efficiency and stabilizing the queues at the cost of imposing a light processing overhead at each relay station to calculate the local iterative power allocation. The MCDH

method can also be an effective allocation method for other forms of cooperative transmission systems such as small cells, where the BS of small cell can handle the light processing requirement likewise the role of relay station in MCDH method.

CONCLUSION AND RECOMMENDATIONS

In this final chapter, we first summarize the results and highlights of this thesis. Next, we introduce ideas for future perspectives on this work and related researches.

Thesis results summary

In the current work we studied the problem of allocating OFDM subcarriers in the time-slotted downlink subframe of a DF relaying cellular network. Our review on the existing body of research revealed that the joint problem of system stability, interference control and resource allocation in multi-hop cooperative communication models has not received sufficient attention. With this motivation, we initialized our research with a simplified single cell model and defined a subcarrier allocation method that satisfies the necessary conditions of stability. This stability scheme is advantageous in cases that the source node, i.e. BS, is aware of the statistics of the arriving traffic. To solve the combinatorial optimization model of the single-cell subcarrier allocation problem, we suggested three approaches. The first variation is based on the fact that the subcarriers on each link can be shared during each time-slot. Although this assumption resolves the combinatorial problem and results in the optimal solution, it imposes extra synchronization complexity which is not practical for general system models. To fill this gap, the second approach, called ARA, is introduced and it is shown to achieve close to optimal solution. The final scheme, called TPHA, is defined to provide a computationally efficient resolution for the problem of exclusive subcarrier allocation with stability consideration.

Next, we extended the system model to a multi-cell network that also integrates the interference control subject to the resource allocation problem. To improve the energy efficiency, we also integrated the transmission power into the problem. Moreover, a tunable time-slot division parameter is applied to the model to employ dynamic adaptation with the channel state variation of the two-hop communication link. A novel queue-aware stability policy is introduced, which is mostly advantageous for scenarios that the traffic arrival rate cannot be recognized in the source. To solve the defined combinatorial and non-convex

optimization problem, we proposed three different approaches. The first approach solves the problem with time-shared assumption. The time-shared method shows high resiliency to stabilize the system even in highly loaded networks. In case that system design limitations do not allow the subcarriers to be time-shared, we proposed an optimal binary allocation scheme, namely OBRA. Despite the other optimal binary allocation mechanisms, we showed that the computation complexity of OBRA is polynomial in time. For the third approach, we introduced a power conservative method, namely CBRA, which prevents the waste of bandwidth, and consequently decreases the system power consumption significantly.

CSI acquisition is a challenging issue for cooperative transmission schemes. Most resource allocation regimes require large amount of feedback that also incurs severe energy loss due to CSI acquisition. For the last enhancement step, we proposed a novel distributed method, called MCDH, which reduces the feedback overhead and also improves the scalability of allocation problem for networks with large number of cells and users. The high scalability feature relies on the fact that the distributed method enables each cell to solve the resource allocation problem independently. This way, MCDH method also prevents large signaling load and latency that is unavoidable in centralized methods. We employed an effective neighbor abstraction technique that only accounts the worst interfering link in the neighbor cells. This, MCDH succeeded to dynamically adapt the system model with changes of channel condition in neighbor cells with less amount of CSI feedback that leads to better energy efficiency. Although in MCDH the relay stations are required to solve a part of resource allocation problem, it does not reduce the system efficiency thanks to the fast convergence speed and small problem size of the proposed power allocation algorithm.

Recommendations for future extensions

In this section we propose a list of possible extensions to the current work that opens new perspectives for future research.

1- To reduce the cross layer data processing, a joint model of admission control and subcarrier allocation method constrained to system stability and interference management can be a beneficial extension to the current work;

2- In a two hop-transmission model defined in the current work, the optimal value of a problem defined in $\max\left\{\min\left\{\alpha r_{c,m}^{BS}(k),(1-\alpha)r_{c,m}^{RS}(k)\right\}\right\}$ is attainable when $\alpha r_{c,m}^{BS}(k)=(1-\alpha)r_{c,m}^{RS}(k)$. Therefore, one can extend the optimization problem defined in this work by adding the time-slot division variable to set the optimization parameters and an extra constraint as $(1-\alpha)r_{c,m}^{RS}(k)-\alpha r_{c,m}^{BS}(k)\le\delta$ where δ is a very small positive constant. Although this adjustment adds a nonlinear term to the problem, it is beneficial since it results optimal resource allocation in optimum time-slot division ratio;
3- For a centralized scheme, it can be beneficial to jointly optimize the interference tolerance value along with the allocation of resources. This could improve the accuracy of allocation decisions and reduces power consumption due to inaccurate channel estimation;
4- Integrating the resource allocation model with other criteria such as routing or handover can also be beneficial to decrease the cross layer processing. Since such integrations cause extremely complex and large size problems, the critical gain to achieve is to find a solution that is efficient in computation time and energy and requires low communication overhead. This is a necessary satisfaction factor the future network applications which have high expectations on delay and response time;
5- To improve the dynamical energy efficiency, energy harvesting is the recent promising solution (M. Peng et al., 2015). As a future perspective, one can employ the received interference as a source of energy harvesting for low power nodes;
6- Cloud computing has been recognized as a solution for centralized large-scale cooperative signal processing (Mugen Peng, I, Tan, & Huang, 2014). Benefiting from cloud radio access networks (C-RANs) can be a future direction to alleviate the overload of backhaul and computation complexity for centralized allocation schemes;
7- Due to the similarities in configurations, the methods introduced in this thesis can be readily customized for other forms of HetNet models that are based on cooperative transmissions.

APPENDIX I

CONVEX PROPERTTY OF THE PERSPECTIVE FUNCTION

Recalling the perspective function and its convexity property, let $g(v,t)$ be the perspective function of $f(v)$ so that $g(v,t) \cong tf\left(v/t\right)$. Then $f(v)$ is concave in *(v)* if and only if $g(v,t)$ is concave in (v,t) (Boyd & Vandenberghe, 2004). Now let assume the functions

$$g_1\left(\hat{p}_{c,m}^{BS}(i,k), s_{c,m}^{BS}(i,k)\right) = s_{c,m}^{BS}(i,k)\log_2\left(\frac{\hat{p}_{c,m}^{BS}(i,k) l_{c,m}^{BS}(k)\left|H_{c,m}^{BS}(k)\right|^2}{s_{c,m}^{BS}(i,k)\sigma^2}\right),$$

$$g_2\left(\hat{p}_{c,m}^{RS}(i,k), s_{c,m}^{RS}(i,k)\right) = s_{c,m}^{RS}(i,k)\log_2\left(\frac{\hat{p}_{c,m}^{RS}(i,k) l_{c,m}^{RS}(k)\left|H_{c,m}^{RS}(i,k)\right|^2}{s_{c,m}^{RS}(i,k)\left(\sigma^2+\theta\right)}\right)$$

are respectively the perspective functions of

$$f_1\left(\hat{p}_{c,m}^{BS}(i,k)\right) = \log_2\left(\frac{\hat{p}_{c,m}^{BS}(i,k) l_{c,m}^{BS}(k)\left|H_{c,m}^{BS}(k)\right|^2}{\sigma^2}\right),$$

$$f_2\left(\hat{p}_{c,m}^{RS}(i,k)\right) = \log_2\left(\frac{\hat{p}_{c,m}^{RS}(i,k) l_{c,m}^{RS}(k)\left|H_{c,m}^{RS}(i,k)\right|^2}{\left(\sigma^2+\theta\right)}\right)$$

It can be readily shown that the capacity functions f_1 and f_2 are concave in $\hat{p}_{c,m}^{BS}(i,k)$ and in $\hat{p}_{c,m}^{RS}(i,k)$ respectively. Then it is concluded that their perspective functions g_1 and g_2, are also concave in $\left(\hat{p}_{c,m}^{BS}(i,k), s_{c,m}^{BS}(i,k)\right)$ and $\left(\hat{p}_{c,m}^{RS}(i,k), s_{c,m}^{RS}(i,k)\right)$ respectively. Therefore the converted rate functions $\hat{r}_{c,m}^{BS}(k)$ Eq.(3.20) and $\hat{r}_{c,m}^{RS}(k)$ Eq.(3.21), which equals the sum of the concave functions g_1 and g_2, are also concave in $\left(\hat{p}_{c,m}^{BS}(i,k), s_{c,m}^{BS}(i,k)\right)$ and $\left(\hat{p}_{c,m}^{RS}(i,k), s_{c,m}^{RS}(i,k)\right)$, respectively.

APPENDIX II

ADMISSION CONTROL POLICY

To prevent the optimization problem from infeasible situation, one solution is to employ an admission control policy. To support the proposed stability constraints defined in CHAPTER 2 and CHAPTER 3, which are the sufficient conditions of stability, in this section we explain an admission control policy that only accepts users whose traffic arrival rate is within the stability region of the considered model. The stability region of a policy is defined as the set of all possible arrival rate vectors for which the system is stable under the considered allocation policy (Song, Li, & Cimini, 2009). It is shown in (Song et al., 2009) that the necessary condition for stability is that the vector of stationary and ergodic arrival data rates lies in the interior of the ergodic capacity region. To calculate the ergodic capacity region of our model, we first find the instantaneous capacity region that is the set of all achievable data rates in the current systems state H (Neely et al., 2005). The instantaneous capacity region of the two-hop transmission system considered in the current work can be defined as

$$\mathrm{C}(\mathrm{H}) = \min\{\mathbf{R}_{BS-RS}, \mathbf{R}_{RS-UE}\}$$

where $\mathbf{R}$ is the vector of achievable rates using the allocation method defined in CHAPTER 2 or CHAPTER 3. The ergodic capacity region $\overline{\mathrm{C}}$ under the allocation constraints defined in CHAPTER 2 or CHAPTER 3, can be calculated by averaging the capacity regions achieved during a time window of previous allocation procedures. Finally, the responsibility of the admission control policy is to accept users whose traffic arrival rates is within the ergodic capacity region. Note that this admission policy is only defined for the sake of completeness and it is not a part of resource allocation procedure. Therefore, it can be performed as a pre-processing procedure prior to the resource allocation operations.

APPENDIX III

REVIEW ON GEOMETRIC PROGRAMMING

In order to define the standard format of a Geometric Programming (GP) model, we need to review the definition of monomial and posynomial functions. Assume the optimization variable $\boldsymbol{z} \in \mathbb{R}$ is a vector of positive elements. A monomial function in $\boldsymbol{z}$ is of the form $b_0 \prod_i z_i^{\alpha_i}$, and a posynomial in $\boldsymbol{z}$ is of the form $\sum_k b_k \prod_i z_{k,i}^{\alpha_{k,i}}, i = 1,\ldots,n, k = 0,1,\ldots,K$ where $b_k > 0$, α_i, and $\alpha_{k,i}$ are arbitrary constants. A standard GP problem objective function is given by

$$\min_{\mathbf{z}} g_0(\mathbf{z}) \text{ or } \max_{\mathbf{z}} h_0(\mathbf{z})$$

subject to

$$g_j(\mathbf{z}) \le 1, \qquad j = 1,\ldots,m$$

$$h_j(\mathbf{z}) = 1, \qquad j = 1,\ldots,l$$

where $\{g_j\}$ and $\{h_j\}$ are posynomial and monomial, respectively (Boyd & Vandenberghe, 2004; Chiang et al., 2005). Note that the monomials are closed under division and multiplication, while posynomials are closed under addition, multiplication, and positive scaling.

APPENDIX IV

CONVERSION OF OBJECTIVE FUNCTION OF CBRA PROBLEM TO GP-COMPATIBLE FORMAT

By mapping $\rho_{c,m}(k)$ to $\log_2 x_{c,m}(k)$, the objective function is reformulated to $T_s \sum_c \sum_m \sum_k \log\left(x_{c,m}(k)\right)^{\omega_{c,m}(k)}$, which is a monotone increasing function in $x_{c,m}(k)$ assuming a constant positive set of weights $\{\omega\}$. One can rewrite the latter objective function format into product form $G: T_s \log \prod_c \prod_m \prod_k x_{c,m}(k)^{\omega_{c,m}(k)}$. For the sake of simplicity and without loss of generality, we eliminate the logarithmic objective function and replace it with $G_0: \prod_c \prod_m \prod_k x_{c,m}(k)^{\omega_{c,m}(k)}$, which is also monotone increasing in $x_{c,m}(k)$. Ultimately, G_0 is a posynomial that respects the GP format and we can use it as the objective function in CBRA method.

APPENDIX V

MONOMIAL APPROXIMATION EMPLOYED IN CBRA PROBLEM

In order to acquire the monomial approximation of a differentiable function $f(\mathbf{z})$, $\boldsymbol{z} \in \mathbb{R}$ near a feasible point $\mathbf{z}^{(0)}$, we need to define the first order Taylor approximation of the function in the logarithmic domain (Marks & Wright, 1978), which is given by

$$\hat{f}(\mathbf{z}) \cong f(\mathbf{z}^{(0)}) \prod_{i=1}^{n} \left(\frac{z_i}{z_i^{(0)}} \right)^{\beta_i}$$

where $\beta_i = \frac{z_i^{(0)}}{f(\mathbf{z}^{(0)})} . \frac{\partial f(\mathbf{z})}{\partial z_i}\Big|_{\mathbf{z}=\mathbf{z}^{(0)}}$ .This approximation can be used to convert the constraint into monomial format in order to comply with the necessary monomial property of the equality constraints in GP format. We assume a function of power vector $\mathbf{p} = \left\{ p(i,k) \mid i = 1,\ldots,n_f, k \in K \right\}$, which is given by

$$f(\mathbf{p}) = \prod_{i=1}^{n_f} \left(1 + \frac{p(i,k) l \left| H(i) \right|^2}{x_i} \right)^{\tilde{s}(i,k)}$$

where x_i is a constant equal to noise. The monomial approximation of $f(\mathbf{p})$ is then calculated by

$$\hat{f}(\mathbf{z}) \cong f\left(\mathbf{p}^{(0)}\right) \prod_{i=1}^{n_f} \left(\frac{p(i,k)}{p^{(0)}(i,k)} \right)^{\beta_i}$$

where

$$\beta_i = \frac{p^{(0)}(i,k)}{f(\mathbf{p}^{(0)})} . \frac{\partial f(\mathbf{p})}{\partial p(i,k)}\Big|_{\mathbf{p}=\mathbf{p}^{(0)}} = \frac{\tilde{s}(i,k)\, p^{(0)}(i,k) l \left| H(i) \right|^2}{x_i + p^{(0)}(i,k) l \left| H(i) \right|^2},$$

$$f\left(\mathbf{p}^{(0)}\right) = \prod_{i=1}^{n_f} \left(1 + \frac{p^{(0)}(i,k) l \left| H(i) \right|^2}{x_i} \right)^{\tilde{s}(i,k)} .$$

APPENDIX VI

CONVERGENCE CONDITIONS OF THE ITERATIVE GP PROBLEM

IN CBRA METHOD

The sequence of iterative GP problems will converge to the solution of the corresponding KKT system of CBRA problem when the three conditions, outlined in (Chiang et al., 2005; Marks & Wright, 1978) are satisfied on the monomial approximated function. First consider the function $f_k^{BS}(\mathbf{p})$ given by

$$f_k^{BS}(\mathbf{p}) = \prod_{i=1}^{n_f} \left(1 + \frac{p_{c,m}^{BS}(i,k) l_{c,m}^{BS} \left| H_{c,m}^{BS}(i) \right|^2}{\sigma_z^2} \right)^{\tilde{s}_{c,m}^{BS}(i,k)},$$

that is approximated by

$$\hat{f}_k^{BS}(\mathbf{p}) = \prod_{i \in N} \left(1 + \frac{p_{c,m}^{(0)BS}(i,k) l_{c,m}^{BS} \left| H_{c,m}^{BS}(i) \right|^2}{\sigma_z^2} \right)^{\tilde{s}_{c,m}^{BS}(i,k)} \prod_{i \in N} \left(\frac{p_{c,m}^{BS}(i,k)}{p_{c,m}^{(0)BS}(i,k)} \right)^{\beta_{c,m}^{BS}(i,k)}$$

where $\beta_{c,m}^{BS}(i,k) = \dfrac{\tilde{s}_{c,m}^{BS}(i,k) p_{c,m}^{(0)BS}(i,k) l_{c,m}^{BS} \left| H_{c,m}^{BS}(i) \right|^2}{\sigma_z^2 + p_{c,m}^{(0)BS}(i,k) l_{c,m}^{BS} \left| H_{c,m}^{BS}(i) \right|^2}$. In this case, the three conditions are as follows (the same set of three conditions can be considered and approved as corresponding to the constraints involving relay power $p_{c,m}^{RS}(i,k)$, which are not repeated here.) :

1- $$1 - \left(\prod_{i=1}^{n_f} \left(1 + \frac{p_{c,m}^{(0)BS}(i,k) l_{c,m}^{BS} \left| H_{c,m}^{BS}(i) \right|^2}{\sigma_z^2} \right)^{\tilde{s}_{c,m}^{BS}(i,k)} \right) \left(\prod_{i=1}^{n_f} \left(\frac{p_{c,m}^{BS}(i,k)}{p_{c,m}^{(0)BS}(i,k)} \right)^{\beta_{c,m}^{BS}(i,k)} \right) \le \prod_{i=1}^{n_f} \left(1 + \frac{p_{c,m}^{BS}(i,k) l_{c,m}^{BS} \left| H_{c,m}^{BS}(i) \right|^2}{\sigma_z^2} \right)^{\tilde{s}_{c,m}^{BS}(i,k)}$$

2- $$\prod_{i=1}^{n_f} \left(1 + \frac{p_{c,m}^{BS}(i,k) l_{c,m}^{BS} \left| H_{c,m}^{BS}(i) \right|^2}{\sigma_z^2} \right)^{\tilde{s}_{c,m}^{BS}(i,k)} \Bigg|_{p_{c,m}^{BS}(i,k) = p_{c,m}^{(0)BS}(i,k)} = \left(\prod_{i=1}^{n_f} \left(1 + \frac{p_{c,m}^{(0)BS}(i,k) l_{c,m}^{BS} \left| H_{c,m}^{BS}(i) \right|^2}{\sigma_z^2} \right)^{\tilde{s}_{c,m}^{BS}(i,k)} \right) \left(\prod_{i=1}^{n_f} \left(\frac{p_{c,m}^{BS}(i,k)}{p_{c,m}^{(0)BS}(i,k)} \right)^{\beta_{c,m}^{BS}(i,k)} \right) \Bigg|_{p_{c,m}^{BS}(i,k) = p_{c,m}^{(0)BS}}$$

$$3\text{-}\frac{\partial}{\partial p_{c,m}^{BS}(i,k)}\prod_{i=1}^{n_f}\left(1+\frac{p_{c,m}^{BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(i,k)}\Bigg|_{p_{c,m}^{BS}(i,k)=p_{c,m}^{BS}(i,k)^*}$$

$$=\frac{\partial}{\partial p_{c,m}^{BS}(i,k)}\left(\left(\prod_{i=1}^{n_f}\left(1+\frac{p_{c,m}^{(0)BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(i,k)}\right)\left(\prod_{i=1}^{n_f}\left(\frac{p_{c,m}^{BS}(i,k)}{p_{c,m}^{(0)BS}(i,k)}\right)^{\beta_{c,m}^{BS}(i,k)}\right)\right)\Bigg|_{p_{c,m}^{BS}(i,k)=p_{c,m}^{BS}(i,k)^*}$$

where $p_{c,m}^{BS}(i,k)^*$ is assumed to be the power allocation at convergence.

The first condition implies that any solution to the approximated problem will be a feasible point of the original CBRA problem. This condition is approvable according to *Lemma 1* in (Chiang et al., 2005).

The second condition ensures the successive increase of the objective function of CBRA problem at each iteration. This condition is satisfied, considering the fact that the monomial approximation is actually the Tayler expansion around $p_{c,m}^{(0)BS}(i,k)$.

The third condition guarantees that the solution of the iterative algorithm after convergence satisfies the KKT conditions corresponding to CBRA problem. This can be proven by first considering the partial differential of the original function, which is given by

$$\frac{\partial}{\partial p_{c,m}^{BS}(i,k)}\prod_{i=1}^{n_f}\left(1+\frac{p_{c,m}^{BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(i,k)}\Bigg|_{p_{c,m}^{BS}(i,k)=p_{c,m}^{BS}(i,k)^*}=$$

$$\frac{\tilde{s}_{c,m}^{BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma^2+p_{c,m}^{BS*}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}\left(1+\frac{p_{c,m}^{BS*}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(i,k)}\prod_{j=1,j\neq i}^{n_f}\left(1+\frac{p_{c,m}^{BS}(j,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(j)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(j,k)}$$

$$=\begin{cases}0\ ,if\ \ \tilde{s}_{c,m}^{BS}(i,k)=0\ ,\forall j\neq i;\tilde{s}_{c,m}^{BS}(j,k)=1\\ \dfrac{l_{c,m}^{BS}\left|H_{c,m}^{BS}(j)\right|^2}{\sigma_z^2},if\ \ \tilde{s}_{c,m}^{BS}(i,k)=1,\forall j\neq i;\tilde{s}_{c,m}^{BS}(j,k)=0\end{cases}.$$

Similar result can be achieved by the partial differential of the approximated function that is given by

$$\frac{\partial}{\partial p_{c,m}^{BS}(i,k)}\left(\prod_{i=1}^{n_f}\left(1+\frac{p_{c,m}^{(0)BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(i,k)}\right)\left(\prod_{i=1}^{n_f}\left(\frac{p_{c,m}^{BS}(i,k)}{p_{c,m}^{(0)BS}(i,k)}\right)^{\beta_{c,m}^{BS}(i,k)}\right)\Bigg|_{p_{c,m}^{BS}(i,k)=p_{c,m}^{BS*}(i,k)}$$

$$=\left(1+\frac{p_{c,m}^{(d-1)BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(i,k)}\prod_{j=1,j\neq i}^{n_f}\left(1+\frac{p_{c,m}^{BS}(j,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(j)\right|^2}{\sigma_z^2}\right)^{\tilde{s}_{c,m}^{BS}(j,k)}\frac{\tilde{s}_{c,m}^{BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{\sigma^2+p_{c,m}^{(d-1)BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}\frac{p_{c,m}^{(d-1)BS}(i,k)}{p_{c,m}^{BS*}(i,k)}\left(\frac{p_{c,m}^{BS*}(i,k)}{p_{c,m}^{(d-1)BS}(i,k)}\right)^{\beta_{c,m}^{BS}}\prod_{j=1,j\neq i}^{n_f}\left(\frac{p_{c,m}^{BS}(j,k)}{p_{c,m}^{(d-1)BS}(j,k)}\right)^{\beta_{c,m}^{BS}}$$

$$=\begin{cases}0\ ,if\ \ \tilde{s}_{c,m}^{BS}(i,k)=0\ ,\forall j\neq i;\tilde{s}_{c,m}^{BS}(j,k)=1\\ \dfrac{l_{c,m}^{BS}\left|H_{c,m}^{BS}(j)\right|^2}{\sigma_z^2},if\ \ \tilde{s}_{c,m}^{BS}(i,k)=1,\forall j\neq i;\tilde{s}_{c,m}^{BS}(j,k)=0\end{cases}$$

This result is achieved using the fact that the optimization parameters at the convergence are equal to those at the very last iteration (i.e. $p_{c,m}^{(d-1)BS}(i,k)=p_{c,m}^{(d)BS}(i,k)$ where the d-th iteration is assumed to be the convergence state).

APPENDIX VII

THE SELF-CONCORDANT PROPERTY

Based on the definition in (Marks & Wright, 1978), a function $f:\mathrm{R}^n \to \mathrm{R}$ is called self-concordant if for all $x, v \in \mathrm{R}^n$, $t \in \mathrm{R}$ where $x+tv$ is in the domain of f, we have $\left|\frac{\partial^3}{\partial t^3} f(x+tv)\right| \le 2\frac{\partial^2}{\partial t^2} f(x+tv)^{3/2}$.

The log-barrier function Ω of multi-cell time-shared problem defined in CHAPTER 3, can be constructed by superimposing the sum of the logarithm of the constraints onto the *t*-scaled objective function. This is given by

$$\Omega = -t\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}} \omega_{c,m}(k)\rho_{c,m}(k) - \sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}} \log\left(\hat{p}^{BS}_{c,m}(i,k)\right) - \sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}} \log\left(\hat{p}^{RS}_{c,m}(i,k)\right) -$$

$$\sum_{c\in\Phi}\sum_{i\in\mathrm{N}} \log\left(1-\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}} s^{BS}_{c,m}(i,k)\right) - \sum_{c\in\Phi}\sum_{i\in\mathrm{N}} \log\left(1-\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}} s^{RS}_{c,m}(i,k)\right) -$$

$$\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}} \log\left(s^{BS}_{c,m}(i,k)\right) - \sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}} \log\left(s^{RS}_{c,m}(i,k)\right) -$$

$$\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}} \log\left(1-s^{BS}_{c,m}(i,k)\right) - \sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}} \log\left(1-s^{RS}_{c,m}(i,k)\right) -$$

$$\sum_{c\in\Phi}\log\left(P^{\max}_{BS} - \sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}} \hat{p}^{BS}_{c,m}(i,k)\right) - \sum_{c\in\Phi}\sum_{m\in\Re}\log\left(P^{\max}_{RS} - \sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}} \hat{p}^{RS}_{c,m}(i,k)\right) - \sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}} \log\left(\theta-\hat{\sigma}^2_{c,m}(i,k)\right) -$$

$$\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}} \log\left(\Delta f\sum_{i\in\mathrm{N}}\left(s^{BS}_{c,m}(i,k)\log_2\left(1+\frac{\hat{p}^{BS}_{c,m}(i,k)l^{BS}_{c,m}\left|H^{BS}_{c,m}(i)\right|^2}{s^{BS}_{c,m}(i,k)\sigma^2_z(i,k)}\right)\right) - \frac{1}{\alpha T_s}Q^{BS}_{c,m}(k)\right) -$$

$$\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}} \log\left(\Delta f\sum_{i\in\mathrm{N}}\left(s^{RS}_{c,m}(i,k)\log_2\left(1+\frac{\hat{p}^{RS}_{c,m}(i,k)l^{RS}_{c,m}(k)\left|H^{RS}_{c,m}(i,k)\right|^2}{s^{RS}_{c,m}(i,k)\left(\theta+\hat{\sigma}^2_{c,m}(i,k)\right)}\right)\right) - \frac{1}{(1-\alpha)T_s}Q^{RS}_{c,m}(k)\right) -$$

$$\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}} \log\left(\alpha\Delta f\sum_{i\in\mathrm{N}}\left(s^{BS}_{c,m}(i,k)\log_2\left(1+\frac{\hat{p}^{BS}_{c,m}(i,k)l^{BS}_{c,m}\left|H^{BS}_{c,m}(i)\right|^2}{s^{BS}_{c,m}(i,k)\sigma^2_z(i,k)}\right)\right) - \rho_{c,m}(k)\right) -$$

$$\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\log\left((1-\alpha)\Delta f\sum_{i\in\mathrm{N}}\left(s_{c,m}^{RS}(i,k)\log_2\left(1+\frac{\hat{p}_{c,m}^{RS}(i,k)l_{c,m}^{RS}(k)\left|H_{c,m}^{RS}(i,k)\right|^2}{s_{c,m}^{RS}(i,k)\left(\theta+\hat{\sigma}_{c,m}^2(i,k)\right)}\right)\right)-\rho_{c,m}(k)\right).$$

It can be seen that Ω is self-concordant except for the last two terms. It has been shown that for a convex function $f:\mathrm{R}_{++}{}^n\to\mathrm{R}$,where $\left|\frac{\partial^3}{\partial x^3}f(x)\right|\le\frac{3\partial^2}{x\partial x^2}f(x)$ for all x in the domain of f, the function $g(x)=-\log\left(-f(x)\right)-\log(x)$ is self-concordant on $\{x\mid x>0, f(x)<0\}$ (Marks & Wright, 1978). By applying this theory, our matching $f(x)$ functions can be written as

$$f_1=-\left(\alpha\Delta f\sum_{i\in\mathrm{N}}\left(s_{c,m}^{BS}(i,k)\log_2\left(1+\frac{\hat{p}_{c,m}^{BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{s_{c,m}^{BS}(i,k)\sigma_z^2(i,k)}\right)\right)+\rho_{c,m}(k)\right)$$

and

$$f_2=-\left((1-\alpha)\Delta f\sum_{i\in\mathrm{N}}\left(s_{c,m}^{RS}(i,k)\log_2\left(1+\frac{\hat{p}_{c,m}^{RS}(i,k)l_{c,m}^{RS}(k)\left|H_{c,m}^{RS}(i,k)\right|^2}{s_{c,m}^{RS}(i,k)\left(\hat{\sigma}_{c,m}^2(i,k)+\theta\right)}\right)\right)+\rho_{c,m}(k)\right).$$

We add an auxiliary constraint $\rho_{c,m}^2(k)\ge 0$,$\forall m\in\mathrm{R}_c,\forall c\in\Phi,\forall\mathrm{k}\in\mathrm{I}_{c,m}$, which does not affect the final solution or the actual feasible set. This auxiliary constraint, along with two original constraints, i.e. $\hat{p}_{c,m}^{BS}(i,k),\hat{p}_{c,m}^{RS}(i,k)\ge 0$ and $s_{c,m}^{BS}(i,k),s_{c,m}^{RS}(i,k)\ge 0$, help to reformulate the last two terms into the self-concordant format given by

$$g_1=-\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\log\left(\alpha\Delta f\sum_{i\in\mathrm{N}}\left(s_{c,m}^{BS}(i,k)\log_2\left(1+\frac{\hat{p}_{c,m}^{BS}(i,k)l_{c,m}^{BS}\left|H_{c,m}^{BS}(i)\right|^2}{s_{c,m}^{BS}(i,k)\sigma_z^2(i,k)}\right)\right)-\rho_{c,m}(k)\right)-$$
$$\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}}\log\left(s_{c,m}^{BS}(i,k)\right)-\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in\mathrm{N}}\log\left(\hat{p}_{c,m}^{BS}(i,k)\right)-\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\log\left(\rho_{c,m}(k)\right)$$

and

$$g_2 = -\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}} \log\left((1-\alpha)\Delta f \sum_{i\in \mathrm{N}} \left(s_{c,m}^{RS}(i,k)\log_2\left(1+\frac{\hat{p}_{c,m}^{RS}(i,k) l_{c,m}^{RS}(k)\left|H_{c,m}^{RS}(i,k)\right|^2}{s_{c,m}^{RS}(i,k)\left(\hat{\sigma}_{c,m}^2(i,k)+\theta\right)}\right)\right) - \rho_{c,m}(k)\right) -$$

$$\sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in \mathrm{N}} \log\left(s_{c,m}^{RS}(i,k)\right) - \sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}}\sum_{i\in \mathrm{N}} \log\left(\hat{p}_{c,m}^{RS}(i,k)\right) - \sum_{c\in\Phi}\sum_{m\in\Re_c}\sum_{k\in\Im_{c,m}} \log\left(\rho_{c,m}(k)\right) .$$

The Time-shared problem, that is defined in CHAPTER 3, now possesses the self-concordance property, and therefore, the number of Newton steps required to solve the problem is bounded to $\sqrt{x}$, where x is the complexity of one Newton step. It is known that the complexity of each Newton step is equal to the cube of the number of inequality constraints. Altogether, the computational complexity of Time-shared method can be estimated by $O\left(\sqrt{\left(\varphi \mathrm{K}\left(5+7\mathrm{N}\right)+\varphi\left(1+\mathrm{M}+2\mathrm{N}\right)\right)^3}\right)$.

One can follow the same procedure to show the self-concordance of OBRA and CBRA, which is dismissed here to prevent repetition. Note that the same auxiliary constraint $\rho_{c,m}^2(k) \geq 0$, $\forall m \in \mathrm{R}_c, \forall c \in \Phi, \forall \mathrm{k} \in \mathrm{I}_{c,m}$ has been added to OBRA and CBRA problem sets as well. In results, the number of constraints in of OBRA and CBRA problems equals $\varphi \mathrm{K}\left(5+3\mathrm{N}\right)+\varphi\left(1+\mathrm{M}\right)$.

APPENDIX VIII

LIST OF PUBLICATIONS

Published journal article

(Lakani & Gagnon, 2017)

Conference papers

(Lakani & Gagnon, 2016)

(Lakani et al., 2015)

Poster presentation

(Lakani, S., 2017)

BIBLIOGRAPHY

Abdelnasser, A., Hossain, E., & Kim, D. I. (2014). Clustering and Resource Allocation for Dense Femtocells in a Two-Tier Cellular OFDMA Network. *IEEE Transactions on Wireless Communications*, *13*(3), 1628–1641. DOI: 10.1109/TW.2014.011614.131163

Al-Tous, H., & Barhumi, I. (2016). Resource Allocation for Multiple-Sources Single-Relay Cooperative Communication OFDMA Systems. *IEEE Transactions on Mobile Computing*, *15*(4), 964–981. DOI: 10.1109/TMC.2015.2434813

Andrews, J. G., Buzzi, S., Choi, W., Hanly, S. V., Lozano, A., Soong, A. C. K., & Zhang, J. C. (2014). What Will 5G Be? *IEEE Journal on Selected Areas in Communications*, *32*(6), 1065–1082. DOI: 10.1109/JSAC.2014.2328098

Andrews, Jeffrey G. (2005). Interference cancellation for cellular systems: a contemporary overview. *IEEE Wireless Communications*, *12*(2), 19–29.

Annavajjala, R., Maaref, A., & Zhang, J. (2010). Demodulate-and-Forward Relaying with Higher Order Modulations: Impact of Channel State Uncertainty. In *2010 IEEE International Conference on Communications* (pp. 1–5). DOI: 10.1109/ICC.2010.5502353

Asadi, A., & Mancuso, V. (2013). A survey on opportunistic scheduling in wireless communications. *Communications Surveys & Tutorials, IEEE*, *15*(4), 1671–1688.

Bangerter, B., Talwar, S., Arefi, R., & Stewart, K. (2014). Networks and devices for the 5G era. *IEEE Communications Magazine*, *52*(2), 90–96.

Basaran, S. T., & Kurt, G. K. (2016). Joint Subcarrier and Power Allocation in OFDMA Systems for Outage Minimization. *IEEE Communications Letters*, *20*(10), 2007–2010. DOI: 10.1109/LCOMM.2016.2586038

Bertsekas, D., & Gallager, R. (1992). Delay models in data networks. In *Data Networks (2Nd Ed.)* (pp. 149–270). Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Bhatia, R., Li, L., Luo, H., & Ramjee, R. (2006). ICAM: integrated cellular and ad hoc multicast. *IEEE Transactions on Mobile Computing*, *5*(8), 1004–1015. DOI: 10.1109/TMC.2006.116

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Boyd, S., Xiao, L., & Mutapcic, A. (2003). Subgradient methods. *Lecture Notes of EE392o, Stanford University, Autumn Quarter*, *2004*. Retrieved from http://web.mit.edu/6.976/www/notes/subgrad_method.pdf

Bu, S., Yu, F. R., & Yanikomeroglu, H. (2015). Interference-Aware Energy-Efficient Resource Allocation for OFDMA-Based Heterogeneous Networks With Incomplete Channel State Information. *IEEE Transactions on Vehicular Technology*, *64*(3), 1036–1050. DOI: 10.1109/TVT.2014.2325823

Cai, X., Yao, Y., & Giannakis, G. B. (2005). Achievable rates in low-power relay links over fading channels. *IEEE Transactions on Communications*, *53*(1), 184–194. DOI: 10.1109/TCOMM.2004.840634

Cao, J., Zhang, T., Zeng, Z., Chen, Y., & Chai, K. K. (2014). Multi-relay selection schemes based on evolutionary algorithm in cooperative relay networks. *International Journal of Communication Systems*, *27*(4), 571–591. DOI: 10.1002/dac.2710

Cendrillon, R., Yu, W., Moonen, M., Verlinden, J., & Bostoen, T. (2006). Optimal multiuser spectrum balancing for digital subscriber lines. *Communications, IEEE Transactions on*, *54*(5), 922–933.

Chang, C.-S. (1994). Stability, queue length, and delay of deterministic and stochastic queueing networks. *IEEE Transactions on Automatic Control*, *39*(5), 913–931. DOI: 10.1109/9.284868

Chen, M., Serbetli, S., & Yener, A. (2005). Distributed power allocation for parallel relay networks. In *GLOBECOM '05. IEEE Global Telecommunications Conference.* (Vol. 3, p. 5 pp.-). DOI: 10.1109/GLOCOM.2005.1577839

Chen, Z., Fan, P., Li, T., & Letaief, K. B. (2015). On the cooperation gain in 5g heterogeneous networking systems. In *2015 IEEE International Symposium on Information Theory (ISIT)* (pp. 1014–1018). DOI: 10.1109/ISIT.2015.7282608

Chiang, M., Tan, C. W., Palomar, D. P., O'Neill, D., & Julian, D. (2005). Power control by geometric programming. *Resource Allocation in Next Generation Wireless Networks*, *5*, 289–313.

Cisco Systems Inc. (2017, February). Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021. Retrieved from http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf

Cui, Y., Lau, V. K. N., Wang, R., Huang, H., & Zhang, S. (2012). A Survey on Delay-Aware Resource Control for Wireless Systems;Large Deviation Theory, Stochastic Lyapunov Drift, and Distributed Stochastic Learning. *IEEE Transactions on Information Theory*, *58*(3), 1677–1701. DOI: 10.1109/TIT.2011.2178150

Cui, Ying, Lau, V. K., & Yeh, E. (2015). Delay optimal buffered decode-and-forward for two-hop networks with random link connectivity. *IEEE Transactions on Information Theory*, *61*(1), 404–425.

Dai, M., Wang, P., Zhang, S., Chen, B., Wang, H., Lin, X., & Sun, C. (2014). Survey on cooperative strategies for wireless relay channels. *Transactions on Emerging Telecommunications Technologies*, *25*(9), 926–942. DOI: 10.1002/ett.2789

de Moraes, T. M., Nisar, M. D., Gonzalez, A. A., & Seidel, E. (2012). Resource allocation in relay enhanced LTE-Advanced networks. *EURASIP Journal on Wireless Communications and Networking*, *2012*(1), 364.

Eryilmaz, A., & Srikant, R. (2007). Fair Resource Allocation in Wireless Networks Using Queue-length-based Scheduling and Congestion Control. *IEEE/ACM Trans. Netw.*, *15*(6), 1333–1344. DOI: 10.1109/TNET.2007.897944

Estrada, R., Jarray, A., Otrok, H., & Dziong, Z. (2014). Base station selection and resource allocation in macro–femtocell networks under noisy scenario. *Wireless Networks*, *20*(1), 115–131. DOI: 10.1007/s11276-013-0594-9

Georgiadis, L., Neely, M. J., & Tassiulas, L. (2006). *Resource allocation and cross-layer control in wireless networks*. Now Publishers Inc.

Ghosh, A., Mangalvedhe, N., Ratasuk, R., Mondal, B., Cudak, M., Visotsky, E., … others. (2012). Heterogeneous cellular networks: From theory to practice. *IEEE Communications Magazine*, *50*(6). Retrieved from http://IEEExplore.IEEE.org/abstract/document/6211486/

Han, B., Peng, M., Zhao, Z., & Wang, W. (2013). A Multidimensional Resource-Allocation Optimization Algorithm for the Network-Coding-Based Multiple-Access Relay Channels in OFDM Systems. *IEEE Transactions on Vehicular Technology*, *62*(8), 4069–4078. DOI: 10.1109/TVT.2013.2251025

Hasan, M., & Hossain, E. (2015). Distributed Resource Allocation for Relay-Aided Device-to-Device Communication Under Channel Uncertainties: A Stable Matching Approach. *IEEE Transactions on Communications*, *63*(10), 3882–3897. DOI: 10.1109/TCOMM.2015.2466238

Hausl, C., Işcan, O., & Rossetto, F. (2012). Resource allocation for asymmetric multi-way relay communication over orthogonal channels. *EURASIP Journal on Wireless Communications and Networking*, *2012*(1), 1–12.

Ho, T. M., Tran, N. H., Do, C. T., Kazmi, S. M. A., Huh, E.-N., & Hong, C. S. (2015). Power Control for Interference Management and QoS Guarantee in Heterogeneous Networks. *IEEE Communications Letters*, *19*(8), 1402–1405. DOI: 10.1109/LCOMM.2015.2444844

Host-Madsen, A. (2002). On the capacity of wireless relaying. In *Proceedings IEEE 56th Vehicular Technology Conference* (Vol. 3, pp. 1333–1337 vol.3). DOI: 10.1109/VETECF.2002.1040432

Huang, L., Rong, M., Wang, L., Xue, Y., & Schulz, E. (2007). Resource Scheduling for OFDMA/TDD Based Relay Enhanced Cellular Networks. In *2007 IEEE Wireless Communications and Networking Conference* (pp. 1544–1548). DOI: 10.1109/WCNC.2007.291

Jamali, V., Zlatanov, N., & Schober, R. (2015). Bidirectional Buffer-Aided Relay Networks With Fixed Rate Transmission-Part II: Delay-Constrained Case. *IEEE Transactions on Wireless Communications*, *14*(3), 1339–1355. DOI: 10.1109/TWC.2014.2365806

Jia, X., Deng, P., Yang, L., & Zhu, H. (2015). Spectrum and Energy Efficiencies for Multiuser Pairs Massive MIMO Systems With Full-Duplex Amplify-and-Forward Relay. *IEEE Access*, *3*, 1907–1918. DOI: 10.1109/ACCESS.2015.2486039

Jing, Y., & Hassibi, B. (2006). Distributed space-time coding in wireless relay networks. *IEEE Transactions on Wireless Communications*, *5*(12). Retrieved from http://IEEExplore.IEEE.org/abstract/document/4027587/

Kelly, F. P., Maulloo, A. K., & Tan, D. K. H. (1998). Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, *49*(3), 237–252. DOI: 10.1057/palgrave.jors.2600523

Khafagy, M., Ismail, A., Alouini, M.-S., & Aissa, S. (2013). On the outage performance of full-duplex selective decode-and-forward relaying. *IEEE Communications Letters*, *17*(6), 1180–1183.

Korte, B., & Vygen, J. (2009). *Combinatorial optimization: theory and algorithms*. Springer Science & Business Media.

Lakani, S. (2017, May). *Stability and Interference Control in Relay-aided Resource Allocation in Cellular Networks*. Presented at the Poster session for the Chair of Wireless Emergency and Tactical Communication, Ultra Electronics Inc., Montreal, Canada.

Lakani, S., & Gagnon, F. (2016). Resource allocation for relay-aided OFDMA networks with constraints on queue stability (pp. 1–7). Presented at the 2016 International Conference on Computing, Networking and Communications (ICNC), Kauai, HI, USA: IEEE. DOI: 10.1109/ICCNC.2016.7440662

Lakani, S., & Gagnon, F. (2017). Optimal design and energy efficient binary resource allocation of interference-limited cellular relay-aided systems with consideration of queue stability. *IEEE Access*.

Lakani, S., Gagnon, F., & Groleau, R. (2015). An IEEE 802.16 MAC Layer Downlink Scheduler Implemented in NS3 to Improve the Performance of Real-Time and Non-Real-Time Traffic Transmission. In *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st* (pp. 1–6). IEEE.

Laneman, J. N., Tse, D. N. C., & Wornell, G. W. (2004). Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transactions on Information Theory*, *50*(12), 3062–3080. DOI: 10.1109/TIT.2004.838089

Lau, V. K. N., Zhang, F., & Cui, Y. (2013). Low Complexity Delay-Constrained Beamforming for Multi-User MIMO Systems With Imperfect CSIT. *IEEE Transactions on Signal Processing*, *61*(16), 4090–4099. DOI: 10.1109/TSP.2013.2264058

Lee, N., & Heath, R. W. (2014). Space-Time Interference Alignment and Degree-of-Freedom Regions for the MISO Broadcast Channel With Periodic CSI Feedback. *IEEE Transactions on Information Theory*, *60*(1), 515–528. DOI: 10.1109/TIT.2013.2285216

Lee, Y. L., Chuah, T. C., Loo, J., & Vinel, A. (2014). Recent Advances in Radio Resource Management for Heterogeneous LTE/LTE-A Networks. *IEEE Communications Surveys Tutorials*, *16*(4), 2142–2180. DOI: 10.1109/COMST.2014.2326303

Leon-Garcia, A. (2008). *Probability, statistics, and random processes for electrical engineering*. Pearson/Prentice Hall 3rd ed. Upper Saddle River, NJ.

Li, J., Peng, M., Cheng, A., Yu, Y., & Wang, C. (2014). Resource Allocation Optimization for Delay-Sensitive Traffic in Fronthaul Constrained Cloud Radio Access Networks. *IEEE Systems Journal*, *PP*(99), 1–12. DOI: 10.1109/JSYST.2014.2364252

Li, Y., Cimini, L. J., & Sollenberger, N. R. (1998). Robust channel estimation for OFDM systems with rapid dispersive fading channels. *IEEE Transactions on Communications*, *46*(7), 902–915.

Liang, J., Wang, X., & Zhang, W. (2011). Capacity Theorem and Optimal Power Allocation for Frequency Division Multiple-Access Relay Channels. In *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011* (pp. 1–5). DOI: 10.1109/GLOCOM.2011.6134191

Lin, C., Tao, M., Stüber, G., & Liu, Y. (2013). Distributed cross-layer resource allocation for statistical QoS provisioning in femtocell networks. In *2013 IEEE International Conference on Communications (ICC)* (pp. 5000–5004). DOI: 10.1109/ICC.2013.6655372

Liu, G., Yu, F. R., Ji, H., Leung, V. C. M., & Li, X. (2015). In-Band Full-Duplex Relaying: A Survey, Research Issues and Challenges. *IEEE Communications Surveys Tutorials*, *17*(2), 500–524. DOI: 10.1109/COMST.2015.2394324

Liu, Y., & Chen, W. (2012). Adaptive Resource Allocation for Improved DF Aided Downlink Multi-User OFDM Systems. *IEEE Wireless Communications Letters*, *1*(6), 557–560. DOI: 10.1109/WCL.2012.080112.120377

Liu, Y. F., & Dai, Y. H. (2014). On the Complexity of Joint Subcarrier and Power Allocation for Multi-User OFDMA Systems. *IEEE Transactions on Signal Processing*, *62*(3), 583–596. DOI: 10.1109/TSP.2013.2293130

Loodaricheh, R. A., Mallick, S., & Bhargava, V. K. (2014). Energy-Efficient Resource Allocation for OFDMA Cellular Networks With User Cooperation and QoS Provisioning. *IEEE Transactions on Wireless Communications*, *13*(11), 6132–6146. DOI: 10.1109/TWC.2014.2329877

M. Grant, & S. Boyd. (2014). CVX: Matlab software for disciplined convex programming, version 2.1. Retrieved from http://cvxr.com/cvx

Mao, Z., Wang, X. M., & Lin, J. (2005). Fast optimal radio resource allocation in OFDMA system based on branch-and-bound method. In *Communications, Computers and signal Processing, 2005. PACRIM. 2005 IEEE Pacific Rim Conference on* (pp. 348–351). IEEE. Retrieved from http://IEEExplore.IEEE.org/abstract/document/1517297/

Marks, B. R., & Wright, G. P. (1978). Technical Note—A General Inner Approximation Algorithm for Nonconvex Mathematical Programs. *Operations Research*, *26*(4), 681–683.

Marshoud, H., Otrok, H., Barada, H., Estrada, R., Jarray, A., & Dziong, Z. (2016). Realistic framework for resource allocation in macro–femtocell networks based on genetic algorithm. *Telecommunication Systems*, *63*(1), 99–110. DOI: 10.1007/s11235-015-9976-x

MOSEK Apps. (2014). The MOSEK optimization toolbox for Matlab Manual, version 7.1. Retrieved from https://www.mosek.com/

Neely, M. J. (2006). Energy optimal control for time-varying wireless networks. *IEEE Transactions on Information Theory*, *52*(7), 2915–2934.

Neely, M. J., Modiano, E., & Rohrs, C. E. (2005). Dynamic power allocation and routing for time-varying wireless networks. *Selected Areas in Communications, IEEE Journal on*, *23*(1), 89–103.

Nemhauser, G. L., & Wolsey, L. A. (1988). *Integer and Combinatorial Optimization*. New York, NY, USA: Wiley-Interscience.

Nesterov, Y., Nemirovskii, A., & Ye, Y. (1994). *Interior-point polynomial algorithms in convex programming* (Vol. 13). SIAM. Retrieved from http://epubs.siam.org/doi/pdf/10.1137/1.9781611970791.fm

Ng, D. W. K., & Schober, R. (2010). Cross-layer scheduling for OFDMA amplify-and-forward relay networks. *Vehicular Technology, IEEE Transactions on*, *59*(3), 1443–1458.

Ng, D. W. K., & Schober, R. (2011). Resource allocation and scheduling in multi-cell OFDMA systems with decode-and-forward relaying. *Wireless Communications, IEEE Transactions on*, *10*(7), 2246–2258.

Novlan, T. D., Ganti, R. K., Ghosh, A., & Andrews, J. G. (2012). Analytical evaluation of fractional frequency reuse for heterogeneous cellular networks. *IEEE Transactions on Communications*, *60*(7), 2029–2039.

Peng, M., Wang, C., Li, J., Xiang, H., & Lau, V. (2015). Recent Advances in Underlay Heterogeneous Networks: Interference Control, Resource Allocation, and Self-Organization. *IEEE Communications Surveys Tutorials*, *17*(2), 700–729. DOI: 10.1109/COMST.2015.2416772

Peng, Mugen, I, C.-L., Tan, C.-W., & Huang, C. (2014). IEEE Access Special Section Editorial: Recent Advances in Cloud Radio Access Networks. *IEEE Access*, *2*, 1683–1685.

Phan, K. T., Le-Ngoc, T., Vorobyov, S. A., & Tellambura, C. (2009). Power allocation in wireless multi-user relay networks. *Wireless Communications, IEEE Transactions on*, *8*(5), 2535–2545.

Qiao, D., & Gursoy, M. C. (2016). Statistical Delay Tradeoffs in Buffer-Aided Two-Hop Wireless Communication Systems. *IEEE Transactions on Communications*, *64*(11), 4563–4577.

Qiao, D., Gursoy, M. C., & Velipasalar, S. (2013). Effective capacity of two-hop wireless communication systems. *IEEE Transactions on Information Theory*, *59*(2), 873–885.

Rappaport, T. (2001). *Wireless Communications: Principles and Practice* (2nd ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.

Rashtchi, R., Gohary, R. H., & Yanikomeroglu, H. (2016). Generalized Cross-Layer Designs for Generic Half-Duplex Multicarrier Wireless Networks With Frequency-Reuse. *IEEE Transactions on Wireless Communications*, *15*(1), 458–471. DOI: 10.1109/TWC.2015.2474864

Rashtchi, R., Gohary, R., & Yanikomeroglu, H. (2014). Routing, scheduling and power allocation in generic OFDMA wireless networks: Optimal design and efficiently computable bounds. *Wireless Communications, IEEE Transactions on*, *13*(4), 2034–2046.

Saad, M., Bettayeb, M., & Boubas, A. Y. (2007). Joint Routing and Radio Resource Management in Multihop Cellular Networks using Genetic Algorithms. In *2007 IEEE International Conference on Signal Processing and Communications* (pp. 1351–1354). DOI: 10.1109/ICSPC.2007.4728578

Saad, M., & Muhaureq, S. A. (2008). Joint routing and radio resource management in multihop cellular networks using particle swarm optimization. In *2008 National Radio Science Conference* (pp. 1–9). DOI: 10.1109/NRSC.2008.4542326

Salem, M., Adinoyi, A., Rahman, M., Yanikomeroglu, H., Falconer, D., & Kim, Y.-D. (2010). Fairness-aware radio resource management in downlink OFDMA cellular relay networks. *Wireless Communications, IEEE Transactions on*, *9*(5), 1628–1639.

Salem, M., Adinoyi, A., Yanikomeroglu, H., & Falconer, D. (2010). Opportunities and challenges in OFDMA-based cellular relay networks: A radio resource management perspective. *Vehicular Technology, IEEE Transactions on*, *59*(5), 2496–2510.

Salem, M., Adinoyi, A., Yanikomeroglu, H., & Falconer, D. (2011). Fair resource allocation toward ubiquitous coverage in OFDMA-based cellular relay networks with asymmetric traffic. *Vehicular Technology, IEEE Transactions on*, *60*(5), 2280–2292.

Sendonaris, A., Erkip, E., & Aazhang, B. (2003). User cooperation diversity. Part I. System description. *IEEE Transactions on Communications*, *51*(11), 1927–1938. DOI: 10.1109/TCOMM.2003.818096

Sierksma, G., Dam, P. van, & Tijssen, G. A. (1996). *Linear and integer programming: theory and practice*. New York: Marcel Dekker.

Simoens, S., Muńoz-Medina, O., Vidal, J., & Coso, A. D. (2010). Compress-and-forward cooperative MIMO relaying with full channel state information. *Signal Processing, IEEE Transactions on*, *58*(2), 781–791.

Son, K., Lee, S., Yi, Y., & Chong, S. (2011). REFIM: A Practical Interference Management in Heterogeneous Wireless Access Networks. *IEEE Journal on Selected Areas in Communications*, *29*(6), 1260–1272. DOI: 10.1109/JSAC.2011.110613

Song, G., Li, Y., & Cimini, L. J. (2009). Joint channel-and queue-aware scheduling for multiuser diversity in wireless OFDMA networks. *IEEE Transactions on Communications*, *57*(7). Retrieved from http://IEEExplore.IEEE.org/abstract/document/5165406/

Song, G., & Li, Y. G. (2005). Cross-layer optimization for OFDM wireless networks-part II: algorithm development. *Wireless Communications, IEEE Transactions on*, *4*(2), 625–634.

Stolyar, A. L., & Viswanathan, H. (2009). Self-organizing Dynamic Fractional Frequency Reuse for Best-Effort Traffic Through Distributed Inter-cell Coordination. In *IEEE Infocom 2009 - IEEE Conference on Computer Communications, Vols 1-5* (pp. 1287–1295). New York: IEEE.

Suraweera, H. A., Louie, R. H., Li, Y., Karagiannidis, G. K., & Vucetic, B. (2009). Two hop amplify-and-forward transmission in mixed Rayleigh and Rician fading channels.

IEEE Communications Letters, *13*(4). Retrieved from http://IEEExplore.IEEE.org/abstract/document/4815065/

Tang, J., & Zhang, X. (2007). Cross-layer resource allocation over wireless relay networks for quality of service provisioning. *IEEE Journal on Selected Areas in Communications*, *25*(4), 645–656. DOI: 10.1109/JSAC.2007.070502

Tang, Z., & Wei, G. (2009). Resource Allocation with Fairness Consideration in OFDMA-Based Relay Networks. In *2009 IEEE Wireless Communications and Networking Conference* (pp. 1–5). DOI: 10.1109/WCNC.2009.4917952

Tassiulas, L., & Ephremides, A. (1992). Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *Automatic Control, IEEE Transactions on*, *37*(12), 1936–1948.

Tassiulas, L., & Ephremides, A. (1994). Dynamic scheduling for minimum delay in tandem and parallel constrained queueing models. *Annals of Operations Research*, *48*(4), 333–355. DOI: 10.1007/BF02024520

Trinh, P. V., & Pham, A. T. (2015). Outage Performance of Dual-Hop AF Relaying Systems with Mixed MMW RF and FSO Links. In *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)* (pp. 1–5). DOI: 10.1109/VTCFall.2015.7391061

Universal Mobile Telecommunications System (UMTS); Spacial channel model for Multiple Input Multiple Output (MIMO) simulations. (n.d.). (3GPP TR 25.996 version 11.0.0 2012 No. 3GPP TR 25.996 version 11.0.0 2012).

Uygungelen, S., Auer, G., & Bharucha, Z. (2011). Graph-Based Dynamic Frequency Reuse in Femtocell Networks. In *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)* (pp. 1–6). DOI: 10.1109/VETECS.2011.5956438

Venturino, L., Prasad, N., & Wang, X. (2009). Coordinated Scheduling and Power Allocation in Downlink Multicell OFDMA Networks. *IEEE Transactions on Vehicular Technology*, *58*(6), 2835–2848. DOI: 10.1109/TVT.2009.2013233

Viswanath, P., Tse, D. N. C., & Laroia, R. (2002). Opportunistic beamforming using dumb antennas. *IEEE Transactions on Information Theory*, *48*(6), 1277–1294. DOI: 10.1109/TIT.2002.1003822

Viswanathan, H., & Mukherjee, S. (2005). Performance of cellular networks with relays and centralized scheduling. *Wireless Communications, IEEE Transactions on*, *4*(5), 2318–2328.

Wang, J., Li, Y., Zhong, B., Zhang, Z., & Long, K. (2013). Adaptive Power Allocation for Decode-and-Forward OFDM Transmission with Multi-Hop Relaying. *2013 IEEE/Cic International Conference on Communications in China (Iccc)*, 345–350.

Wang, N., & Gulliver, T. A. (2015). Queue-aware transmission scheduling for cooperative wireless communications. *IEEE Transactions on Communications*, *63*(4), 1149–1161.

Wong, C. Y., Cheng, R. S., Ben Letaief, K., & Murch, R. D. (1999). Multiuser OFDM with adaptive subcarrier, bit, and power allocation. *IEEE Journal on Selected Areas in Communications*, *17*(10), 1747–1758.

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

Wunder, G., Zhou, C., Bakker, H.-E., & Kaminski, S. (2008). Throughput maximization under rate requirements for the OFDMA downlink channel with limited feedback. *Eurasip Journal on Wireless Communications and Networking*, 437921. DOI: 10.1155/2008/437921

Yu, W., & Lui, R. (2006). Dual methods for nonconvex spectrum optimization of multicarrier systems. *Communications, IEEE Transactions on*, *54*(7), 1310–1322.

Zaki, A. N., & Fapojuwo, A. O. (2011). Optimal and Efficient Graph-Based Resource Allocation Algorithms for Multiservice Frame-Based OFDMA Networks. *IEEE Transactions on Mobile Computing*, *10*(8), 1175–1186. DOI: 10.1109/TMC.2010.227

Zarakovitis, C. C., Ni, Q., Skordoulis, D. E., & Hadjinicolaou, M. G. (2012). Power-Efficient Cross-Layer Design for OFDMA Systems With Heterogeneous QoS, Imperfect CSI, and Outage Considerations. *IEEE Transactions on Vehicular Technology*, *61*(2), 781–798. DOI: 10.1109/TVT.2011.2179817

Zhang, G., Yang, K., Wu, S., Mei, X., & Zhao, Z. (2015). Efficient power control for half-duplex relay based D2D networks under sum power constraints. *Wireless Networks*, *21*(7), 2345–2355.

Zhang, H., Jiang, C., Beaulieu, N. C., Chu, X., Wen, X., & Tao, M. (2014). Resource Allocation in Spectrum-Sharing OFDMA Femtocells With Heterogeneous Services. *IEEE Transactions on Communications*, *62*(7), 2366–2377. DOI: 10.1109/TCOMM.2014.2328574

Zhang, Y. J., & Ben Letaief, K. (2004). Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems. *IEEE Transactions on Wireless Communications*, *3*(5), 1566–1575. DOI: 10.1109/TWC.2004.833501

Zhou, L., Hu, X., Ngai, E. C.-H., Zhao, H., Wang, S., Wei, J., & Leung, V. C. M. (2016). A Dynamic Graph-Based Scheduling and Interference Coordination Approach in Heterogeneous Cellular Networks. *IEEE Transactions on Vehicular Technology*, *65*(5), 3735–3748. DOI: 10.1109/TVT.2015.2435746