Energy-efficient Resource Allocation and Cooperation in Wireless Heterogeneous Networks

by

Rindranirina Ramamonjison

B.Sc., University of Quebec at Trois-Rivieres, 2006
M.Eng., Tokyo Institute of Technology, 2011

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

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2015

Abstract

The deluge of mobile data demands a drastic increase of wireless network capacity. A heterogeneous network design, in which small cells are densely deployed, is required to satisfy this demand. However, it is critical that this dense deployment does not lead to a surge in energy cost. The aim of this thesis is to design energy-efficient resource allocation methods and explore the value of cooperation in terms of energy cost. In particular, three different cooperation schemes are studied. First, a multi-cell coordination scheme is proposed for maximizing the energy efficiency of heterogeneous networks. Although this problem is not convex, convergent algorithms are devised to find an efficient power allocation. We found that this simple coordination can offer a significant energy efficiency gain even in dense networks. Second, a joint energy allocation and energy cooperation is proposed for heterogeneous networks with hybrid power sources and energy storage systems. For this study, an offline optimization problem is considered, in which the cells allocate their energy over time based on average rate contraints, the changing channel conditions and the fluctuating energy arrivals. It is found that an optimal use of the harvested energy significantly improves the energy efficiency. A much larger gain is obtained when energy cooperation is also leveraged, i.e. when the cells can exchange their harvested energy through a smart-grid infrastructure. Third, the trade-off between energy cost and performance is addressed for cooperative clustered small-cell networks. In this cooperative model, the small-cell base stations form a cluster of distributed antennas to collectively serve their mobile users. Hence, a joint optimization of cell clustering and cooperative beamforming is proposed to minimize the total energy cost while satisfying the users' quality of service. The problem is formulated as a mixed-integer convex program

and solved with a decomposition method. For a given clustering, a distributed beamforming algorithm is also designed to achieve near-optimal performance at a small cost of signaling overhead. Through simulations, it is shown that these algorithms converge fast and enable the cooperative small cells to save valuable energy.

Preface

Publications that resulted from the research presented in this thesis are as follows:

- R. Ramamonjison and V. K. Bhargava, "Energy Efficiency Maximization Framework in Cognitive Downlink Two-Tier Networks," *IEEE Transactions on Wireless Communications*, vol.14, no.3, pp.1468-1479, March 2015. (Linked to Chapter 2)

- R. Ramamonjison and V. K. Bhargava, "Sum energy-efficiency maximization for cognitive uplink networks with imperfect CSI," *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, pp.1012-1017, April 2014. (Linked to Chapter 2)

- R. Ramamonjison and V. K. Bhargava, "Energy Allocation and Cooperation for Energy-efficient Wireless Two-Tier Networks," Submitted to a journal for publication consideration, 2015. (Linked to Chapter 3)

- R. Ramamonjison, A. Haghnegahdar and V. K. Bhargava, "Joint Optimization of Clustering and Cooperative Beamforming in Green Cognitive Wireless Networks," *IEEE Transactions on Wireless Communications*, vol.13, no.2, pp.982-997, February 2014. (Linked to Chapter 4)

- R. Ramamonjison and V. K. Bhargava, "Distributed beamforming in cognitive multi-cell wireless systems by fast interference coordination," *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, pp.2208-2213, Sept. 2012. (Linked to Chapter 4)

I am the primary author of these publications and has had the main responsibility in developing the original ideas, the mathematical analysis and the Matlab programs used for the

computer simulations. The contributions of my co-authors are as follows. Prof. Vijay K. Bhargava is my honorable supervisor and he provided valuable guidance and directions in identifying the research problems as well as constructive feedback during the preparation of the manuscripts. Alireza Haghnegahdar was a former Master student in our research group; he contributed in one of the above publications by offering important technical feedback during the formulation of the research problem. He also helped in running some of the simulations and offered editorial support when preparing the manuscript for publication.

Some of the simulation results involving constrained optimisation were obtained using the disciplined convex optimization software CVX developed by Grant, Boyd & Ye [1].

Table of Contents

List of Tables

List of Figures

List of Algorithms

List of Abbreviations

ADMM Alternating Direction Method of Multipliers

AWGN Additive White Gaussian Noise

BS Base Station

C-RAN Centralized Radio Access Network

CCCP Convex-Concave Procedure

CoMP Coordinated Multi-Point

CSI Channel State Information

CU Central Unit

DC Direct Current

eICIC Enhance Inter-Cell Interference Coordination

GBD Generalized Benders Decomposition

MBS Macrocell Base Station

MIMO Multiple-input multiple-output

MM Minorization-Maximization

MU Macrocell User

RAN Radio Access Network

RRU Remote Radio Unit

SBS Small-cell Base Station

SC Small Cell

SINR Signal-to-Noise Power Ratio

SU Small-cell User

Acknowledgements

I feel extremely fortunate to have met and learned from many incredible people at the University of British Columbia during my Ph.D. study. Foremost, I would like to thank my advisor Professor Vijay K. Bhargava for his guidance, patience and support over the past four years. I am very grateful for the tremendous opportunity he has given me. Professor Bhargava's exceptional motivation, vision, generosity and integrity will always be an inspiring role model for my future career.

I would also like to thank Professor David Michelson, Professor Vincent Wong and Professor Alireza Nojeh for serving as committee members during my PhD qualifying examination and for offering practical and constructive feedback.

My gratitude also goes to my colleagues at the CITR research group, who offered me their invaluable support, collaboration, and friendship. Special thanks go to Alireza Haghnegahdar with whom I had the good fortune to collaborate. I have greatly benefited from the many hours of stimulating discussions I had with Alireza Haghnegahdar, Roya Arab Loodarichech and K. N. R. Surya Vara Prasad. I also appreciate very much the enormous advice and help that I received from my senior colleagues namely Umesh Phuyal, Zia Hasan, Shankhanaad Mallick, Hamidreza Boostanimehr and Rajiv Devarajan. I am deeply indebted to my colleagues who boldly reviewed this dissertation and my papers (in chronological order): Alireza Haghnegahdar, Shristi Pradhan, Roya Arab Loodarichech, K. N. R. Surya Vara Prasad, Sudha Lohani and Buddhika Nettasinghe. I also thank our guest visitors Frederik Klement and Dr. Ning Wang for the engaging discussions I had with them.

Special thanks go to the faculty and staff at the UBC ECE department for their great support, including Professor Lutz Lampe, Sonia Dhillon, Brandie Hill, Kristie Henriksen,

David Chu Chong, Sabrina Ho and many others.

I wish to thank also my previous mentors whom I had the chance to meet before and during my graduate studies. I would like to thank Professor Adel Omar Dahmane, Dr. Patrick Marsch and Professor Gerhard Fettweis for introducing me to the fascinating world of wireless communications. I also wish to express my gratitude to Professor Kiyomichi Araki and Professor Kei Sakaguchi who welcomed and supervised me during my Master study and encouraged me to pursue a Ph.D. study.

Last but not the least, I feel very blessed to have my dear parents, my beloved wife and my extended family who not only offered me their love and support but also sacrificed a lot when I have been studying abroad. I would like to thank them for being patient, encouraging and caring during all these years.

This Ph.D. thesis would not have been possible if not for the genereous financial support that I received from the Four Year Fellowship program and from Professor Vijay K. Bhargava. This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC).

To my parents, my wife and my son,

with gratitude for your inspiration, love and support.

Chapter 1

Introduction

The efficient allocation of resources is a constant and primary concern in the design and operation of wireless cellular networks. Reliable wireless communications fundamentally depend on two key resources: radio frequency spectrum and energy. Since the early days of mobile telephony, it is recognized that the spectrum is a very scarce resource and must be efficiently reused [2]. In response, the cellular concept was invented at Bell Labs in 1947 as an effective way to improve the spectrum efficiency, expand the service coverage and increase the network capacity. This is done by installing more base stations, splitting existing cells into smaller ones [3] and by reusing the spectrum across distant cells. In fact, cell splitting is still one of the main drivers for capacity growth in all four generations of cellular networks [4].

Today, network operators face a greater challenge since the proliferation of smartphones and the explosion of mobile multimedia applications have led to an exponential growth of mobile data traffic. Unfortunately, this trend is expected to continue as emerging Internet-of-Things applications will further amplify this traffic growth. Such applications will drive more physical objects, such as transportation vehicles, health monitoring systems or even construction diggers, to be remotely connected to the Internet [5]. In the light of these data-hungry applications, the wireless industry faces a tremendous challenge to support a 1000-fold increase in traffic demand over the next decade [6].

Such capacity crunch will require an intense densification of the networks [4]. This trend will result in a heterogeneous multi-tier design, in which a mixture of small cells and macrocells is used. The small cells are deployed over macrocells to upgrade the capacity in

urban and densely populated areas. As a result, each user is brought closer to the network, ideally within 10-30 meters of the closest small-cell base station. This heterogeneous design has many benefits, which include lowered link power budget, decreased load per cell, higher link data rate and better indoor coverage [7].

However, the dense deployment of small cells also raise important questions about the energy efficiency of such networks. Even if a small cell is served by low-power base stations, the high density of base stations may lead to a surge in energy cost. Recently, operators have realized the importance of managing their networks in an energy efficient way. In fact, studies show that information and communications technology infrastructure contributes to a 2% of the total amount of CO_2 emission levels [8]. Given the exponential increase in data traffic, it becomes critical to maintain the energy cost at its current level. This is possible only if more emphasis is placed on energy efficiency when designing and operating future networks.

In spite of these pressing needs, there is a lack of understanding on how to improve the energy efficiency of cellular networks. For example, the implication of heterogeneous networks on energy cost is not well understood. In fact, previous works have mainly focused on the rate maximization problem. Although there exist some studies on energy-efficient resource allocation, they have considered only the simplest cases such as a point-to-point system [9] or a single-cell downlink system with orthogonal transmissions [10, 11]. In contrast, maximizing the energy efficiency in a multi-cell multi-user scenario, such as small-cell networks, is a challenging open problem. This problem is further complicated by the inevitable presence of the interference due to the need for full spectrum reuse. Thus, new methods are needed to maximize the energy efficiency of heterogeneous networks.

In addition, network operators also need new sustainable energy solutions to increase the capacity in a cost-effective manner. An attractive solution is to harvest inexpensive and clean energy at each cell from renewable sources such as solar, wind or even radio-frequency waves [8]. Given the fluctuation of the harvested energy, it is required to com-

bine the renewable sources with a conventional power source to avoid a detriment of the quality of service. Actually, this hybrid approach has already been applied to power networks in remote or rural areas [12]. For example, network operators have relied on solar panels, batteries and diesel generators due to lack of consistent access to electricity [12]. Nonetheless, further research efforts are needed to study the efficient use of hybrid source in large-scale heterogeneous networks. Previous works on resource allocation with hybrid sources have only considered very simple scenarios [13–15].

Another important consideration is the use of cooperation schemes among the cells. In fact, there are three ways in which cells can cooperate. The first cooperation scheme is coordinated multi-cell power allocation. The second cooperation is multi-cell energy sharing, in which the cells can exchange their harvested energy. This form of cooperation is motivated by recent efforts by the power industry to modernize electrical power systems. In particular, future smart-grid technologies will enable a bidirectional energy flow between distributed energy generators and consumers [16]. This new capability should be leveraged to better utilize the limited energy in future heterogeneous networks. The third cooperation scheme is based on coordinated multi-point (CoMP) transmission, in which the cells share their antennas to collectively serve their users. Using this cooperative transmission scheme, the small cells can better mitigate the interference, achieve a spatial multiplexing gain and counter the negative effects of the wireless channel fading. Although many works have explored the benefits of CoMP for spectrum efficiency improvement [17–23], its value in terms of energy efficiency is not yet completely understood.

This thesis deals with the design of energy-efficient resource allocation methods using a mathematical programming approach. Given that base stations are the most power-hungry equipment in cellular networks [8], we focus on the downlink of heterogeneous networks with multiple cells and multiple users. Our goal is to answer the following two questions. How to achieve energy-efficient transmissions in a multi-cell multi-user environment? What is the value of cooperation when the cells share their energy or their

antennas?

To do that, we consider three specific network models, each mainly characterized by the kind of cooperation used by the cells. These models are chosen in order to explore different solutions for enabling energy-efficient heterogeneous networks, and to capture a variety of deployment and cooperation scenarios. The first model is a small-cell network with a centralized architecture. Specifically, the cells are served by radio remote units, which are connected to a central unit through fronthaul fiber links. In this scenario, the power allocation of the cells is coordinated by the central unit to maximize the sum of their energy efficiencies. The second model is a heterogeneous small-cell network with a hybrid power source. Precisely, each cell is powered by both a conventional power grid and by renewable sources. In addition to this hybrid power supply, an energy cooperation is also used so that the cells can exchange their harvested energy through a smart-grid infrastructure. In this second scenario, the cells also maximize their energy efficiency by efficiently allocating the pool of harvested energy over a finite time horizon. Finally, the third model consists of distributed small-cells with conventional power source. The cells do not share their energy but they can cooperate by sharing their antennas and jointly transmitting to their users. In contrast to energy cooperation, this type of cooperation is used to improve the performance at the expense of energy consumption. In this case, the cells should minimize the total energy cost by optimizing the degree of cooperation and their beamforming design given a target quality of service.

In the rest of this chapter, the model of heterogeneous networks is described, the considered resource allocation and multi-cell cooperation schemes are presented in more details, and the main results of the thesis are outlined.

1.1 Background

1.1.1 Wireless Heterogeneous Networks

A heterogeneous multi-tier design will play an important role to support the escalating traffic demand of future mobile applications [6]. In heterogeneous multi-tier networks, a large number of small cells are added to complement the macrocells. These small cells can be of different sizes and include microcells, picocells, indoor femtocells as well as remote radio units.

There are several benefits to adding small cells [24]. First, they offer a capacity boost in hot spots with high traffic demand. Next, they improve the coverage in areas not covered by the macrocells both outdoors and indoors. They also improve the overall network performance by offloading the traffic from the large macrocells. As a result, heterogeneous networks can offer a much higher peak and average rates per user as well as a higher bit rates per unit area.

Unfortunately, a heterogeneous network architecture also comes with many technical challenges [24, 25]. As explained earlier, it is crucial to minimize the energy cost of the network in spite of the dense small-cell deployment. Because the spectrum is also limited, it must be fully reused while mitigating the resulting interference between the cells to an acceptable level. Another drawback of a heterogeneous design is the need to connect the small cells' base stations to the core network through backhaul or fronthaul links [24].

Given different usage and deployment scenarios, both centralized and distributed architectures have been envisioned for heterogeneous networks [24]. These are illustrated in Figure 1.1.

Distributed network architecture

In a distributed deployment scenario, each small cell has a standalone access point or base station. Each base station has its own power supply. The radio access network functions

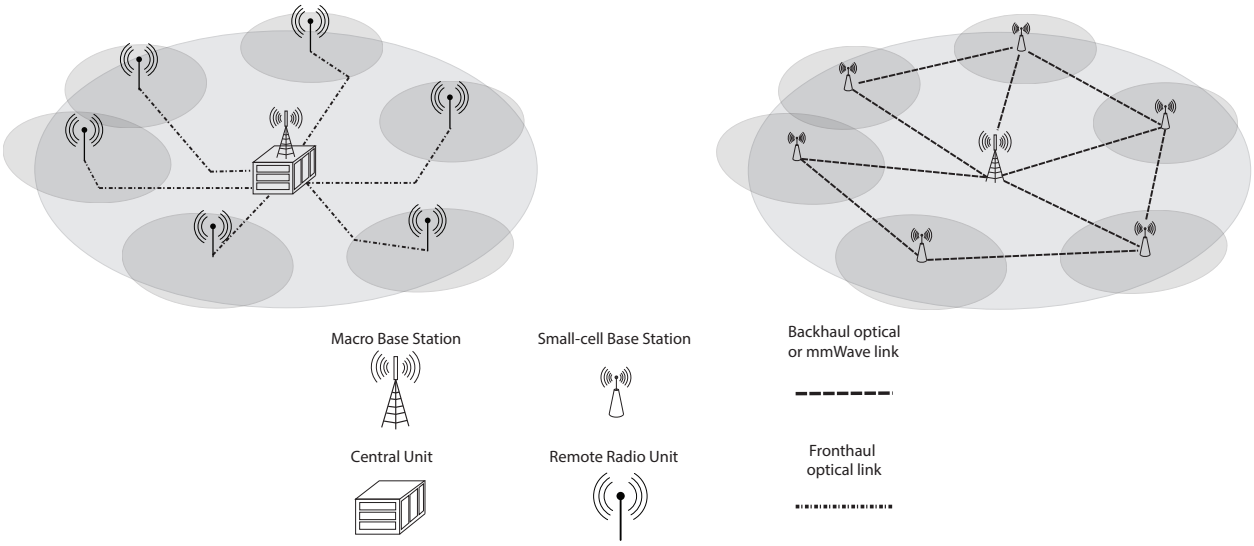


Figure 1.1 – Centralized (*left*) and distributed (*right*) heterogeneous RAN architectures.

of each cell, such as user scheduling, resource allocation and baseband processing, are performed by its own base station. In this case, backhaul links are needed to connect each cell to the core network. When coordination is needed between the cells, some signaling information must also be exchanged through backhaul links. This backhaul infrastructure can be implemented either with optical fiber links or with wireless millimeter-wave links with a large bandwidth [26].

Centralized network architecture

Recently, there is a strong interest to deploy small cells using a centralized radio access network architecture (C-RAN), also known as “cloud” RAN. In this case, each cell is served by a remote radio unit (RRU). The remote units are then connected to a central unit (CU) via fronthaul optical fiber links. In contrast to the distributed deployment, the radio access network functions of the cells are performed by the central unit. As a result, the energy consumption due to processing is controlled by the central unit. This centralized architecture facilitates the coordination between the cells and can also improve the energy efficiency of the network [27, 28].

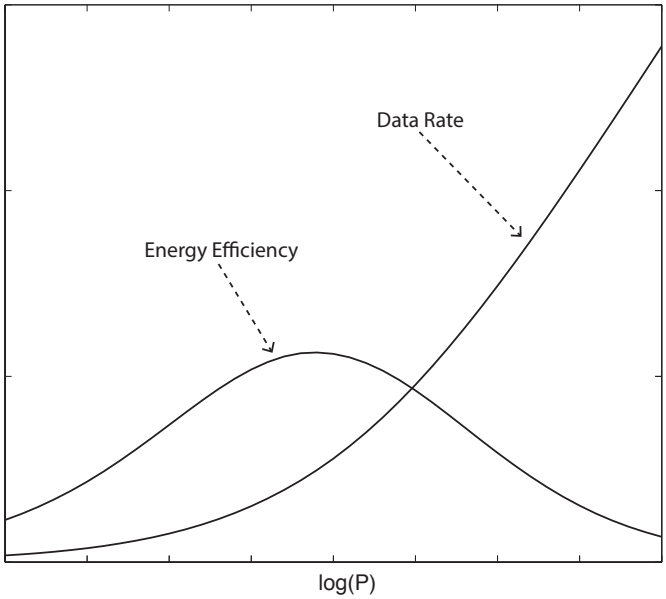


Figure 1.2 – Rate and energy efficiency vs. power (dB)

1.1.2 Resource Allocation Problems in Wireless Networks

The efficient usage of the spectrum and energy is critical to improving the overall experience of mobile users and to reducing the operational costs of the network. Typically, a resource allocation problem is first formulated as an optimization problem based on specific objectives and constraints. Then, algorithms must be developed to find a feasible and efficient solution to the optimization problem.

There are several possible design objectives in wireless communications. To give a simple illustration, let us first consider a single-user single-carrier communication system with a bandwidth W, a transmit power P and a signal-to-noise power ratio $\frac{g}{\sigma^2}$.

A common design goal is to maximize the Shannon's channel capacity C given by [29]:

$$C = W \log_2 \left(1 + \frac{gP}{\sigma^2}\right). \tag{1.1}$$

This capacity function, which defines the maximum achievable data rate for reliable communications is a strictly increasing function of the transmit power.

From energy-limited systems, it is instead desirable to either minimize the energy consumption necessary to satisfy a given rate requirement or to maximize the energy efficiency, also known as the *bits-per-Joule capacity* [9, 30]:

$$\text{EE} = \frac{W}{P_0 + \psi P} \log_2 \left(1 + \frac{gP}{\sigma^2}\right). \tag{1.2}$$

In (1.2), P_0 is the processing power dissipated in the circuit blocks and ψ is the power amplifier efficiency. In contrast to the capacity (1.1), the energy efficiency function (1.2) is not a monotonically increasing function of the transmit power. As shown in Figure 1.2, there is a point at which the effective amount of data transmitted per unit of power is maximum. Another important difference between these two functions is that the rate utility function is concave whereas the energy efficiency function is not. More precisely, (1.2) is a quasi-concave function. This subtle difference makes the energy efficiency maximization problem much more difficult to solve, especially in a multi-user interference-limited scenario. In fact, it has been solved only in simplest scenarios, such as single-user system [9] or in single-cell downlink system with orthogonal multi-access scheme [10]. Unfortunately, the algorithm used [9] and [10] cannot be applied to a multi-cell environment, such as heterogeneous networks.

As previously mentioned, this thesis focuses on the downlink of cellular networks because the energy consumption of base stations is of a greater concern to network operators. Note that our design objective is different from that of wireless sensor networks. Since sensors are powered solely by batteries, the main concerns in these systems are rather to minimize the transmission delay or to maximize the expected amount of data transmitted until the sensor's battery dies [31].

1.1.3 Coordination and Cooperation Schemes in Heterogeneous Networks

In a heterogeneous network, the coordination and cooperation between the cells has a positive impact on the performance in multi-cell environments. We consider three types of cooperation in this thesis.

First, multi-cell coordination is beneficial for power control and interference management. In fact, significant research efforts have been devoted to develop enhanced inter-cell interference management (eICIC) schemes in heterogeneous networks [32–35]. This is due to the fact that traditional frequency reuse and orthogonal multi-access schemes are inefficient or infeasible for a large-scale and heterogeneous deployment [32]. To facilitate the coordination between cells, a centralized network architecture, such as a C-RAN based deployment is desirable. In fact, a distributed coordination is more complicated as signaling information needs to be exchanged between the cells. In this case, the signaling overhead should be kept at a minimum level due to the limited capacity of the backhaul links [36].

Second, coordinated multi-point transmission or CoMP is a more powerful cooperation scheme in which the small-cell base stations form a cluster in order to jointly transmit the data to each user [37]. It can be used both in downlink and uplink. When CoMP is used in a heterogeneous network, a number of cells is involved in the data transmission to or from each user. However, CoMP requires that the transmissions are synchronized and that the user data are available at every point. Hence, it requires backhaul or fronthaul links with a higher capacity. Moreover, a C-RAN based deployment can facilitate its implementation [38]. However, this form of cooperation can incur a large energy consumption due to the multi-point processing.

Third, an energy cooperation scheme can also be envisaged for distributed small cells that are powered by a hybrid power source, i.e. each cell draws energy from both renewable and conventional energy sources. However, previous works that considered the dynamic energy management of hybrid-powered cellular networks only considered the

single-cell scenario [10, 39, 40]. For example, [10] studied the energy-efficient power allocation for a single-cell downlink system. [40] proposed a detailed model of smart hybrid power system and a real-time energy management to reduce the operational cost of a base station. Given the recent developments of smart-grid technologies [16, 41], it will also be important to explore the future possibility for the cells to exchange their harvested renewable through the smart-grid [42].

1.2 Outline of Thesis

The focus of this thesis is to explore the design of energy-efficient resource allocation methods as well as the value of cooperation in heterogeneous multi-cell networks. The outline and the contributions of each chapter are as follows.

Chapter 1, this present chapter, gives the motivation, background and contributions of this thesis.

Chapter 2 develops a simple multi-cell coordination scheme for maximizing the energy efficiency of wireless heterogeneous networks. Precisely, each cell optimizes their power allocation to transmit the maximum number of bits per unit of energy. In this optimization problem, a wide variety of power constraints can be accommodated including deterministic and probabilistic interference constraints. In addition, two scenarios with orthogonal transmissions and full spectrum sharing are both treated. Using a parametric approximation approach, the original complicated non-convex problem is related to a much simpler family of convex parametric problems. Given this result, the energy-efficient power allocation problem can then be solved within the powerful framework of convex optimization theory. In particular, convergent algorithms are derived and proven to converge to global optimum for the case of orthogonal transmissions, and to at least a local optimum for the case of full spectrum sharing. These algorithms are tested in a simulation of small-cell net-

works. The simulation results show that the proposed algorithms have a fast convergence and achieve much better performance compared to an existing game-theoretical scheme.

Chapter 3 considers the joint energy allocation and energy cooperation for heterogeneous networks with hybrid power sources. We assume that each cell has access to a renewable energy source and to the conventional power grid; and has also an energy storage system. The cells thus allocate their energy over time by considering the time-varying channel conditions and energy arrivals. The objective is to maximize the energy efficiency subject to average rate constraints, battery limit constraints and energy causality constraints. Furthermore, an energy cooperation scheme, in which the cells can exchange their harvested renewable energy through a smart-grid infrastructure, is also considered. The goal of this study is to investigate the benefits of using hybrid power source and energy cooperation scheme in heterogeneous networks. Therefore, offline energy-efficient power allocation algorithms are developed for this goal. To do that, the optimization framework of Chapter 2 is extended to handle the dynamic energy constraints and the non-convex rate constraints.

Chapter 4 considers the joint clustering and cooperative beamforming optimization in heterogeneous network. In this model, the cells cooperate not by sharing their energy but instead by sharing their antennas. In other words, they perform a coordinated multi-point transmission by forming a cluster of distributed antennas. This multi-cell cooperation can greatly improve the performance of the cells at the expense of an increased energy cost. Thus, the focus of this chapter is to develop practical methods to achieve an efficient trade-off between energy and performance. Specifically, a convergent algorithm is proposed for finding the joint optimal clustering and beamforming vectors. In addition, a distributed multi-cell beamforming with limited signaling is designed for a given clustering. The convergence and performance of these algorithms are analyzed through simulations.

Finally, Chapter 5 summarizes the main results of the thesis and indicate some interesting directions for future research.

Chapter 2

Energy Efficiency Optimization in Wireless Two-tier Networks

2.1 Introduction

In this chapter, we present a generic optimization framework for the energy-efficient power allocation in wireless heterogeneous two-tier networks. The two-tier base stations control their transmission powers to maximize the sum of their energy efficiency while respecting various power constraints. Moreover, we assume that the two-tier network can share the spectrum with a primary system.

The main contribution in this chapter is an optimization framework for maximizing the sum energy efficiency of the two-tier cells subject to shared and individual power constraints. Before presenting the system model in Section 2.3, we give a detailed review of state-of-the-art research in energy efficient resource allocation in Section 2.2. In contrast to earlier works, we focus on the more difficult multi-cell energy-efficient power allocation problem. We solve this non-convex problem by looking at both the orthogonal multi-access and full spectrum reuse scenarios. In Section 2.5, we first assume that the cells' transmissions are non-interfering or orthogonal. Despite the non-convexity of the problem, we show that it is possible to exploit the problem structure using convex parametric programming. Then, we derive an algorithm which converges to the global optimal power allocation.

Next, we relax the orthogonality constraint in Section 2.6 and allow the cells to interfere with each other while sharing the spectrum. Another algorithm that monotonically converges to at least a local optimum is derived by applying the minorization-maximization principle in conjunction with the Newton method.

We also extend the application of the optimization framework to energy-efficient power allocation problems with imperfect channel state information between the two-tier base stations and the primary users. In this case, the interference power constraints become probabilistic. In Section 2.7, we show how to handle such probabilistic constraints using different robust approximation methods.

Finally, we present some simulation results in Section 2.8, which validate the convergence of these algorithms. In addition, we compare their performance against existing schemes and analyze the effect of network parameters on the energy efficiency.

2.2 Related Works

The design of energy-efficient resource allocation schemes is not a trivial task mainly because of the non-concavity of the energy efficiency utility function.

As mentioned in Chapter 1, the energy efficiency maximization was studied only under simplified network scenarios. First, a single-user point-to-point setting was considered by [9] and [43]. In [9], Isheden et. al. showed that the energy efficiency function is pseudo-concave for a single-user system. As a result, the simple Dinkelbach algorithm [44] can be used to find the optimal solution. The Dinkelbach method could be further applied to more complicated scenarios under two conditions. First, the utility function must consist of a single ratio of the sum rate over total power consumption. Second, the multi-user transmissions should happened over orthogonal non-overlapping channels. For instance, this approach was used in [45] and [46] for the orthogonal single-cell downlink system and in [47] for the single-cell uplink system. Similarly, [48] used these two assumptions

and considered different utility functions for the downlink and uplink systems. For the uplink system, the considered design objective is to maximize the minimum energy efficiency among all the users. In addition, [48] proposed low-complexity algorithms to find suboptimal solutions. In [49], the downlink energy optimization of a simultaneous wireless information and power transfer system is studied. Again, [49] assumed orthogonal transmissions. This study was then extended by [50], which also takes the imperfections of the channel estimation and the multi-user scheduling into account. Finally, similar energy-efficient resource allocation schemes were proposed for OFDMA relay networks. For example, [51] solved a joint relay selection, subcarrier allocation and power allocation problem. The Dinkelbach method was leveraged with an integer relaxation and dual decomposition to derive practical algorithms. Furthermore, the cooperative beamforming problem was studied in [52] for a virtual MIMO system consisting of a single source node, a single destination node and several distributed relay nodes. [52] analyzed the number of relay nodes that maximizes the efficiency for a given outage constraint. In addition, a mode-switching algorithm between cooperative beamforming and direct transmission was studied. Finally, energy-efficient beamforming and scheduling schemes were proposed in [53] for a coordinated multi-cell downlink network. In that model, a zero-forcing beamforming method was used to enforce orthogonality between the transmissions. [53] also chose the single ratio of sum rate over sum power consumption as the objective function.

These aforementioned works share two basic assumptions: orthogonality between transmissions is enforced and the energy efficiency utility metric consists of a single fraction. While such a utility function is perfectly valid for a single-cell downlink transmission, it is not applicable for multi-cell multi-user systems such as a heterogeneous two-tier network. Instead, a more practical utility function is the sum energy efficiency of all users. In fact, the base station of each cell has independent power amplifier and energy source. In such setting, the cells also compete for resources. Therefore, it is more natural to use the sum energy efficiency metric. Actually, it was also shown in [54] that the sum energy

efficiency utility provides higher degrees of freedom for the system design than the ratio of sum rate and total power consumption.

However, maximizing the sum energy efficiency is a very hard problem since this utility function is not even quasiconcave. Hence, the Dinkelbach algorithm can no longer be applied. In addition, the limited spectrum must be reused in heterogeneous two-tier networks. In fact, it was shown in [55] that the resource allocation becomes inefficient when channel orthogonality is enforced in two-tier networks. As a result, the interference between the cells must be mitigated. Nonetheless, the multi-user interference also complicates the resource allocation problem. The previous work in [56] circumvented this difficulty by aiming for a Nash equilibrium. In contrast, we propose in this chapter a novel optimization framework that maximizes the energy efficiency maximization in multi-cell multi-user environments. The framework can handle non-orthogonal transmissions. In contrast to [56], it achieves at least a local optimum.

2.3 System Model

In this chapter, we consider the downlink of a wireless two-tier network shown in Fig. 2.1. Thus, the system is composed of the two-tier cells $\mathcal{C} = \{c_1, \ldots, c_{|\mathcal{C}|}\}$ and of the users $\mathcal{U} = \{u_1, \ldots, u_{|\mathcal{U}|}\}$. Here, the notation $|\mathcal{S}|$ denotes the cardinal of a set $\mathcal{S}$. Each base station (BS) of a cell c serves a subset of users $\mathcal{U}_c$. We assume a centralized radio access network architecture (C-RAN), in which each small cell is served by a radio remote unit (RRU). The RRUs are connected to a central unit (CU) through fronthaul optical or millimeter-wave links. The resource allocation is centralized at the CU. In the rest of this chapter, we use the terms RRU and small-cell base station (SBS) interchangeably. We assume that the BSs and the users are equipped with single antennas. Moreover, the multi-cell transmissions are performed over a set of subcarriers $\mathcal{S} = \{1, \ldots, N\}$.

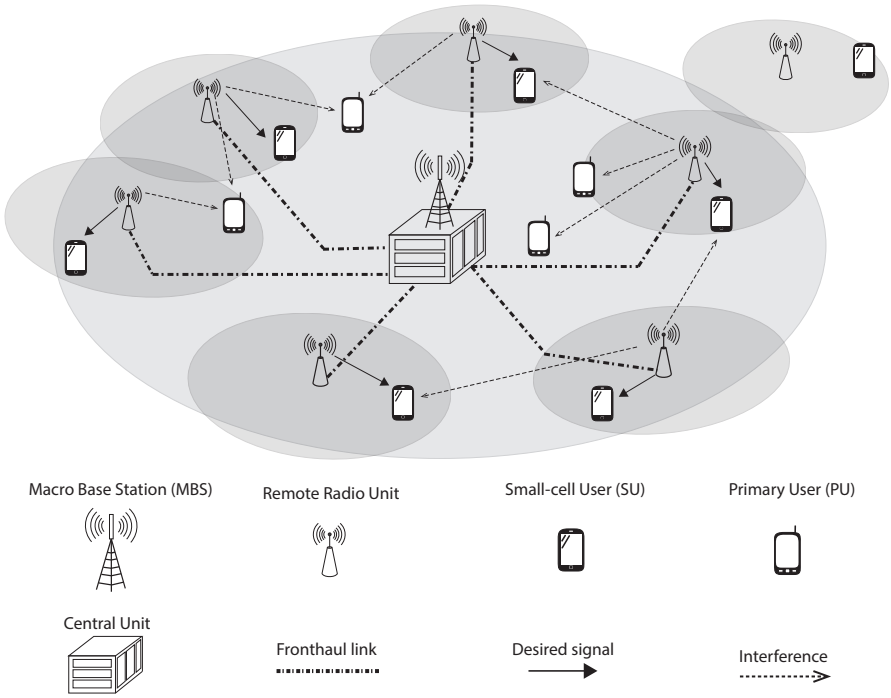


Figure 2.1 – Wireless two-tier network model.

The power allocation vector of all BSs is denoted by the vector $\mathbf{p} = (p_{u,n})_{u\in\mathcal{U},\, n\in\mathcal{S}} \in \mathbb{R}_{+}^{N|\mathcal{U}|}$. Thus, the signal-to-interference-plus-noise ratio (SINR) of user u on subcarrier n is given by:

$$\text{SINR}_{u,n}\left(\mathbf{p}\right) = \frac{p_{u,n}g_{uu,n}}{\sum\limits_{j\in\mathcal{U}\setminus\{u\}} p_{j,n}g_{uj,n} + \sigma_{u,n}^2}, \tag{2.1}$$

where $\sigma_{u,n}^2$ is the variance of an additive white Gaussian noise (AWGN) at user u. In (2.1), $g_{uj,n}$ denotes the channel gain between user u and the BS serving another user j. It accounts for the path loss attenuation, the large-scale shadowing and the small-scale Rayleigh fading with distribution $\mathcal{CN}\left(0,1\right)$ between the BS and the user.

Remark 2.1. We assumed a Rayleigh distribution to model the fading envelope distribution of both the desired and interference channels. This is a somewhat conservative assumption given that, in some small-cell environments such as indoor femto-cells, there can be a dominant line-of-sight path between the BS and the user. Nonetheless, our assumption is meant to simplify the analysis. In fact, more generalized models such as the Rician

fading model would require more knowledge about the specific propagation environment. In particular, the Rician K-factor, which is the power ratio between the fixed and scatter components, is itself a random variable that varies with time, frequency and the location [57].

For convenience, let us express the power allocation of cell c by $\mathbf{p}_c = (p_{u,n})_{u\in\mathcal{U}_c,\ n\in\mathcal{S}} \in \mathbb{R}_+^{N|\mathcal{U}_c|}$ and those of the other cells by the vector $\mathbf{p}_{-c} \triangleq (p_j)_{j\in\mathcal{U}\setminus\mathcal{U}_c,\ n\in\mathcal{S}} \in \mathbb{R}_+^{N(|\mathcal{U}|-|\mathcal{U}_c|)}$. The sum rate of the cell c is thus given by:

$$R_c\left(\mathbf{p}_c, \mathbf{p}_{-c}\right) = \frac{W}{N} \sum_{u\in\mathcal{U}_c} \sum_{n=1}^{N} \log\left(1 + \text{SINR}_{u,n}\right)$$

where W is the total system bandwidth.

By adopting the linear energy consumption model proposed in [58], the overall power cost for cell c is given by

$$C_c\left(\mathbf{p}_c\right) \quad = \quad P_{f,c} + \psi_c \sum_{u\in\mathcal{U}_c} \sum_{n=1}^{N} p_{u,n}$$

where $P_{f,c}$ denotes the fixed power cost due to the BS circuitry and cooling equipment and ψ_c is the BS power amplifier efficiency.

2.4 Problem Formulation

In this chapter, we consider the general problem of maximizing the sum of energy efficiencies of the two-tier cells subject to a set of power constraints:

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \textstyle\sum_{c\in\mathcal{C}} \text{EE}_c\left(\mathbf{p}_c, \mathbf{p}_{-c}\right) \\ \text{subject to} \quad & h_l\left(\mathbf{p}\right) \leq \mathbf{0},\ \forall l \in \mathcal{V}. \end{aligned} \tag{2.2}$$

where each function $h_l,\ \forall l \in \mathcal{V}$ is assumed to be affine. Here, the bit-per-Joule energy efficiency of each cell c is given by the ratio of its sum rate over its power cost:

$$\mathrm{EE}_c\left(\mathbf{p}_c, \mathbf{p}_{-c}\right) = \frac{R_c\left(\mathbf{p}_c, \mathbf{p}_{-c}\right)}{C_c\left(\mathbf{p}_c\right)}. \tag{2.3}$$

In general, the sum energy efficiency is appropriate to measure the overall energy efficiency of two-tier networks assuming the BSs have independent transmissions and power sources. However, the following system energy efficiency metric can also be used

$$\mathrm{EE}_0\left(\mathbf{p}\right) = \frac{\sum_{c\in\mathcal{C}} R_c\left(\mathbf{p}_c, \mathbf{p}_{-c}\right)}{C_c\left(\mathbf{p}_c\right)} \tag{2.4}$$

when the small-cell base stations have a single power source or when they are able share their energy.

For example, in C-RAN heterogeneous networks, the small cells are served by remote radio units which are connected to a central unit. In addition, the system energy efficiency (2.4) is also useful in a future scenario, in which the cells will be able to harvest energy from renewable sources and share their energy through a smart-grid infrastructure.

Under the above assumptions, the optimization problem becomes:

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \mathrm{EE}_0\left(\mathbf{p}\right) \\ \text{subject to} \quad & h_l\left(\mathbf{p}\right) \leq \mathbf{0},\ \forall l \in \mathcal{V}. \end{aligned} \tag{2.5}$$

Due to the interference, neither the sum energy efficiency (2.3) nor the system energy efficiency (2.4) is a concave function. Thus, the above problems cannot be directly solved with standard convex optimization algorithms. In addition, problem (2.2) is more complicated than problem (2.5) since the former maximizes a sum of non-convex fractions. In these optimization problems, we have the following power constraints[1]:

[1]Note that not all these types of power constraints must be imposed. The constraints of the resource allocation will depend on both the network architecture and deployment scenario.

Individual transmit power constraints

Due to regulations and hardware limitations, the macrocell and small-cell base stations are subject to transmit power limitations. For a standard cellular deployment, each base station has its own power supply and is subject to an individual power constraint:

$$\sum_{u \in \mathcal{U}_c} \sum_{n=1}^{N} p_{u,n} \leq P_c, \ \forall c \in \mathcal{C}.$$

Total transmit power constraint

For the case of cloud-radio access network (C-RAN) deployment, the two-tier cells are served by RRUs and may share a total transmit power constraint. In a C-RAN, the low power requirements of RRUs can enable them to be powered by a centralized DC power supply that is located at the central unit [28]. In that case, the cells may share a total transmit power constraint:

$$\sum_{c \in \mathcal{C}} \sum_{u \in \mathcal{U}_c} \sum_{n=1}^{N} p_{u,n} \leq P_{\text{total}}.$$

Interference temperature constraints

It is also possible for future two-tier cells to use a spectrum band that is licensed to a primary system. Using a cognitive underlay model, the primary and secondary systems co-exist and share the same spectrum. Assuming there are primary users $\mathcal{V} = \left\{v_1, \ldots, v_{|\mathcal{V}|}\right\}$, the two-tier cells should not exceed the interference power limit $I_l^{\max}$ measured in Watts imposed by each primary user l as follows

$$\sum_{c \in \mathcal{C}} \sum_{u \in \mathcal{U}_c} \sum_{n=1}^{N} p_{u,n} g_{lc,n} \leq I_l^{\max}, \ \forall l \in \mathcal{V}, \tag{2.6}$$

where $g_{lu,n}$ is the channel gain on subcarrier n between a primary user l and the BS that serves a user u. Note that it is also possible to adopt this cognitive underlay mechanism in the resource allocation of the heterogeneous network. Precisely, the small cells are

considered as secondary users whereas the macrocell users are the primary users. In fact, such spectrum sharing scheme has the benefits of mitigating the inter-cell interference and improve the spectrum efficiency of heterogeneous networks [59, 60].

Probabilistic interference constraints

The aforementioned cognitive underlay scheme works well if the BSs are able to estimate the channels to the primary users. In practice, the channel information can be imperfect due to estimation errors. To take this uncertainty into account, the deterministic interference constraints in (2.6) can be replaced by probablistic interference constraints of the form:

$$\Pr\left\{\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N} p_{u,n}g_{lc,n} \leq I_{\max}\right\} \geq 1-\epsilon,\ \forall l \in \mathcal{V} \tag{2.7}$$

In (2.7), $g_{lc,n}$ represents the uncertain channel gain which depends on the random variable $\zeta_{lc,n}$ in an affine fashion as:

$$g_{lc,n} = \widehat{g}_{lc,n} + \Delta g_{lc,n}\zeta_{lc,n} \tag{2.8}$$

where $\widehat{g}_{lc,n}$ denotes the estimated channel gain and $\Delta g_{lc,n}$ is the absolute value of the maximum channel estimation error. The random error variables $\{\zeta_{lc,n}\}$ are assumed to have mutually independent distributions with $E\left[\zeta_{lc,n}\right] = 0$ and $|\zeta_{lc,n}| \leq 1$, $\forall\,(l,c,n)$. In other words, we assume that the uncertainty region of the random variables $\zeta = (\zeta_{lc,n})_{\forall l,c,n}$ is given by:

$$\mathcal{D} = \{\zeta \mid \|\zeta\|_\infty \leq 1\} \tag{2.9}$$

Note that this bounded uncertainty model is reasonable because, with a limited coordination between the primary and secondary systems, it is difficult for the secondary users to obtain complete information about the probability distributions. Instead, the small-cell BSs can get estimate the error bounds on the channel estimation through empirical measurements. Although the probabilistic contraints (2.7) are non-convex, we will show in Section 2.7 to handle them.

2.5 Special Case: Orthogonal Multi-cell Transmissions

In this section, we first simplify the generic problem 2.2 by assuming orthogonal transmissions for the users. One way to ensure orthogonality is by scheduling the users on different frequency subcarriers. We assume that this is performed before the power allocation. An alternative way to enable orthogonal transmissions is spatial multiplexing scheme but this requires multiple antennas at the base stations so as to cancel the interference between the concurrent transmissions.

2.5.1 Parametric Convex Programming

With orthogonal transmissions, there is no interference between the users at each subcarrier. Thus, the rate function R_c of each cell c is concave and depends only on its power allocation vector $\mathbf{p}_c$. Nonetheless, the energy efficiency maximization problem is still hard to solve since the objective function is a sum of quasiconcave functions.

Let us write this problem in its epigraph form to obtain the equivalent formulation:

$$(P)\begin{cases} \underset{\mathbf{p},\theta}{\text{maximize}} & \sum_{c\in\mathcal{C}} \theta_c \\ \text{subject to} & R_c\left(\mathbf{p}_c\right) - \theta_c C_c\left(\mathbf{p}_c\right) \geq 0,\ \forall c \in \mathcal{C}, \\ & h_l\left(\mathbf{p}\right) \leq \mathbf{0},\ \forall l \in \mathcal{V}, \end{cases} \tag{2.10}$$

in which we introduced some auxilliary variables $\theta \triangleq (\theta_c)_{c\in\mathcal{C}} \in \mathbb{R}^{|\mathcal{C}|}$. To further simplify the notation, let us gather the convex power constraints into a convex set $\mathcal{K}$ of feasible power allocation as follows:

$$\mathcal{K} \triangleq \left\{ \mathbf{p} \in \mathbb{R}^{N|\mathcal{U}|\times 1} :\ \ h_l\left(\mathbf{p}\right) \leq \mathbf{0},\ \forall l \in \mathcal{V} \right\}. \tag{2.11}$$

Next, we introduce the following family of parametric convex problems $(P_{\mathbf{x}})$

$$(P_{\mathbf{x}}) \begin{cases} \underset{\mathbf{p}}{\text{maximize}} & \sum_{c\in\mathcal{C}} \alpha_c \left(R_c\left(\mathbf{p}_c\right) - \theta_c C_c\left(\mathbf{p}_c\right)\right) \\ \text{subject to} & \mathbf{p} \in \mathcal{K}, \end{cases} \tag{2.12}$$

which is parametrized by the following vector of parameters $\mathbf{x} \triangleq (\alpha, \theta) \in \mathbb{R}^{2|\mathcal{C}|}$ with $\alpha \triangleq (\alpha_c)_{c\in\mathcal{U}} \in \mathbb{R}^{|\mathcal{C}|}$. Note that whereas the vector θ corresponds to optimization variables in the original problem (P), it is considered as a vector of parameters in $(P_{\mathbf{x}})$. We can interpret the parametric problem $(P_{\mathbf{x}})$ as a weighted sum rate maximization with penalty terms. Precisely, α is a priority weight vector and θ is a pricing vector for the penalty terms. Since $(P_{\mathbf{x}})$ is convex, its solution can be efficiently and reliably computed using standard algorithms [61]. In our simulations, we employed the disciplined convex optimization software CVX to solve the convex subproblem (2.12) [1].

To find an easier way to solve the non-convex problem (P), we establish the following relation between (P) and the convex parametric problem $(P_{\mathbf{x}})$.

Proposition 2.1. *If $\left(\widehat{\mathbf{p}}, \widehat{\theta}\right)$ is an optimal solution of (P), then there exists $\widehat{\alpha}$ such that $\widehat{\mathbf{p}}$ is a solution of $(P_{\mathbf{x}})$ with $\mathbf{x} = \left(\widehat{\alpha}, \widehat{\theta}\right)$. Moreover, $\widehat{\mathbf{p}}$ must satisfy the following equations:*

$$R_c\left(\widehat{\mathbf{p}}_c\right) - \widehat{\theta}_c C_c\left(\widehat{\mathbf{p}}_c\right) = 0, \ \forall c \in \mathcal{C}, \tag{2.13}$$

$$\widehat{\alpha}_c C_c\left(\widehat{\mathbf{p}}_c\right) - 1 = 0, \ \forall c \in \mathcal{C}. \tag{2.14}$$

Proof. See Appendix A.1. □

Since the energy maximization problem (P) in (2.10) is always feasible, the above results means that there exists at least one parameter $\mathbf{x}$ that satisfies the following two conditions. First, the solution $\widehat{\mathbf{p}}$ of the corresponding convex parametric problem $(P_{\mathbf{x}})$ coincides with the optimal solution of (P). Second, the solution $\widehat{\mathbf{p}}$ must satisfy the system of $2\,|\mathcal{C}|$ nonlinear equations in (2.13) and (2.14).

For conciseness, let us define a function $\widetilde{\mathbf{p}}(\mathbf{x}) : \mathbb{R}^{2|\mathcal{C}|} \to \mathcal{K}$ which takes as input a parameter $\mathbf{x}$ and returns an optimal solution of the parametric problem $(P_{\mathbf{x}})$. Mathematically, we have:

$$\widetilde{\mathbf{p}}(\mathbf{x}) = \underset{h_l(\mathbf{p}) \leq \mathbf{0},\ \forall l \in \mathcal{V}}{\operatorname{argmax}} \sum_{c \in \mathcal{C}} \alpha_c \left(R_c(\mathbf{p}_c) - \theta_c C_c(\mathbf{p}_c) \right) \tag{2.15}$$

Then, we also introduce the following system of nonlinear equations:

$$\mathbf{F}(\mathbf{x}) = \mathbf{0}, \tag{2.16}$$

where the components of the vector-valued function $\mathbf{F} : \mathbb{R}^{2|\mathcal{C}|} \to \mathbb{R}^{2|\mathcal{C}|}$ are defined by:

$$F_c(\mathbf{x}) = \theta_c C_c(\widetilde{\mathbf{p}}(\mathbf{x})) - R_c(\widetilde{\mathbf{p}}(\mathbf{x})), \quad \forall c \in \mathcal{C} \tag{2.17}$$

$$F_{c+|\mathcal{C}|}(\mathbf{x}) = \alpha_c C_c(\widetilde{\mathbf{p}}(\mathbf{x})) - 1, \quad \forall c \in \mathcal{C}. \tag{2.18}$$

Each of these component functions is defined by composite functions of the cell rate R_c or the power cost C_c with the function $\widetilde{\mathbf{p}}$ defined in (2.15).

Given the above definitions, Proposition 2.1 equivalently says the following. If a power allocation $\widehat{\mathbf{p}}$ solves the energy maximization problem (P), then there exists a parameter $\mathbf{x} = \left(\widehat{\alpha}, \widehat{\theta}\right)$ such that $\widehat{\mathbf{p}} = \widetilde{\mathbf{p}}(\mathbf{x})$ and the parameter $\mathbf{x}$ is a root of the system of nonlinear equations $\mathbf{F}(\mathbf{x}) = \mathbf{0}$. This indicates an indirect route for solving the non-convex optimization (P). When (P) is feasible, then at least one root of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ exists. Furthermore, if $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ has a unique solution $\widehat{\mathbf{x}}$, then a global optimal solution of the sum energy maximization problem (P) is exactly given by $\widetilde{\mathbf{p}}(\widehat{\mathbf{x}})$. In the next part, we show that the solution of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ is indeed unique. As a result, it is possible to solve (P) by finding the parameter satisfying $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ using an iterative method.

2.5.2 Uniqueness of the Root of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$

To establish the uniqueness of the solution of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$, we derive some useful properties of $\mathbf{F}(\mathbf{x})$. First, let us define $\mathcal{X}$ as the convex and compact set of parameters:

$$\mathcal{X} \triangleq \left\{ \mathbf{x} = (\alpha, \theta) \in \mathbb{R}_{+}^{2|\mathcal{U}|} \mid -\alpha_c^{\min} \leq \alpha_c \leq \alpha_c^{\max},\ -\theta_c^{\min} \leq \theta_c \leq \theta_c^{\max} \right\}. \tag{2.19}$$

The bounds in (2.19) are defined for each cell $c \in \mathcal{C}$ as:

$$\alpha_c^{\min} = \min_{\mathbf{p} \in \mathcal{K}} \frac{1}{C_c(\mathbf{p}_c)}, \quad \alpha_c^{\max} = \max_{\mathbf{p} \in \mathcal{K}} \frac{1}{C_c(\mathbf{p}_c)}, \tag{2.20}$$

$$\theta_c^{\min} = \min_{\mathbf{p} \in \mathcal{K}} \frac{R_c(\mathbf{p}_c)}{C_c(\mathbf{p}_c)}, \quad \theta_c^{\max} = \max_{\mathbf{p} \in \mathcal{K}} \frac{R_c(\mathbf{p}_c)}{C_c(\mathbf{p}_c)}. \tag{2.21}$$

These bounds are finite and well-defined whenever the feasible set $\mathcal{K}$ in (2.11) is non-empty.

By definition, any solution of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ has to satisfy the conditions (2.13) and (2.14). Thus, given the bounds of $\mathcal{X}$ in (2.20) and (2.21), we can restrict the function $\mathbf{F}(\mathbf{x})$ to be defined over the set $\mathcal{X}$. Effectively, any solution of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ cannot lie outside of $\mathcal{X}$. Then, we can obtain the following key properties for the function $\mathbf{F}(\mathbf{x})$.

Lemma 2.1. *When defined over the set $\mathcal{X}$, the vector function $\mathbf{F}$ defined in* (2.17) *and* (2.18) *must satisfy the following:*

1. $\mathbf{F}$ is differentiable and its Jacobian $\mathbf{F}'$ is a non-singular, diagonal and positive definite matrix given by:

$$\mathbf{F}'(\mathbf{x}) = \begin{bmatrix} \operatorname{diag}\left(\left(C_c\left(\widetilde{\mathbf{p}}(\mathbf{x}) \right) \right)_{c \in \mathcal{C}} \right) & \mathbf{0}_{|\mathcal{C}| \times |\mathcal{C}|} \\ \mathbf{0}_{|\mathcal{C}| \times |\mathcal{C}|} & \operatorname{diag}\left(\left(C_c\left(\widetilde{\mathbf{p}}(\mathbf{x}) \right) \right)_{c \in \mathcal{C}} \right) \end{bmatrix}. \tag{2.22}$$

2. $\mathbf{F}$ is Lipschitz continuous and strongly monotone in $\mathcal{X}$ with constants $M \geq 0$ and $m > 0$ respectively.

3. The Jacobian $\mathbf{F}'$ of $\mathbf{F}$ is Lipschitz continuous and its inverse is bounded.

Proof. The proofs are given in Appendix A.2. □

With the above properties of $\mathbf{F}$, we have the following useful result about the system $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ (2.16).

Proposition 2.2. *There is a unique solution to the nonlinear system* $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ *in the set* $\mathcal{X}$.

Proof. The proof is given in Appendix A.3. □

Consequently, if we can find the unique parameter $\widehat{\mathbf{x}}$ that satisfies $\mathbf{F}(\mathbf{x}) = \mathbf{0}$, the optimal solution of the energy efficiency maximization problem (P) can be determined by solving the convex problem $(P_{\mathbf{x}})$ with its parameter set to $\mathbf{x} = \widehat{\mathbf{x}}$.

2.5.3 Damped Newton Algorithm

Given the results of the previous section, we derive an iterative algorithm for solving problem (P) using a damped Newton method [62]. The proposed algorithm is listed in Algorithm 2.1 on the next page. First, we initialize the power allocation $\mathbf{p}^{(0)} \in \mathcal{K}$ of the users and set the initial parameter $\mathbf{x}^{(0)} = \left(\alpha^{(0)}, \theta^{(0)}\right)$ as follows:

$$\alpha_c^{(0)} = \frac{1}{C_c\left(\mathbf{p}_c^{(0)}\right)},\ \forall c \in \mathcal{C}, \tag{2.23}$$

$$\theta_c^{(0)} = \frac{R_c\left(\mathbf{p}_c^{(0)}\right)}{C_c\left(\mathbf{p}_c^{(0)}\right)},\ \forall c \in \mathcal{C}. \tag{2.24}$$

In each iteration i, we solve the convex parametric subproblem in (2.12) for a given parameter $\mathbf{x}^{(i)}$ to obtain a new power allocation $\widetilde{\mathbf{p}}\left(\mathbf{x}^{(i)}\right)$. After this, if the termination condition $\left\|\mathbf{F}\left(\mathbf{x}^{(i)}\right)\right\| < \epsilon$ is satisfied for a given tolerance $\epsilon > 0$, then we stop the algorithm. Otherwise, we update the parameter $\mathbf{x}^{(i)}$ using a damped Newton method. Precisely, we first calculate the Newton step at the i-th iteration by:

$$\mathbf{d}^{(i)} = -\mathbf{F}'\left(\mathbf{x}^{(i)}\right)^{-1}\mathbf{F}\left(\mathbf{x}^{(i)}\right), \tag{2.25}$$

where $\mathbf{F}'$ is the Jacobian matrix in (2.22) of the vector-valued function $\mathbf{F}$. From Lemma 2.1, $\mathbf{F}'(\mathbf{x})$ is a diagonal positive definite matrix; hence, its inverse can be easily computed.

Then, we use a line search to control the step size $0 < \lambda < 1$ so as to avoid overshooting [63]. In other words, we start with a full Newton step $\lambda = 1$ and reduce it by a factor $0 < \sigma < 1$ until a sufficient decrease in the norm of $\left\|\mathbf{F}\left(\mathbf{x}^{(i)} + \lambda \mathbf{d}^{(i)}\right)\right\|$ is obtained [62]. After this line search, we update the parameter vector as:

$$\mathbf{x}^{(i+1)} \triangleq \mathbf{x}^{(i)} + \lambda \mathbf{d}^{(i)}. \tag{2.26}$$

By substituting (2.17), (2.18) and (2.22) into (2.25), the update equation (2.26) of the parameter $\mathbf{x}^{(i)} = \left(\alpha^{(i)}, \theta^{(i)}\right)$ in step (S.4) is given component-wise by:

$$\alpha_c^{(i+1)} = (1-\lambda)\,\alpha_c^{(i)} + \lambda \frac{1}{C_c\left(\mathbf{p}_c^{(i)}\right)}, \quad \forall c \in \mathcal{C}, \tag{2.27}$$

$$\theta_i^{(i+1)} = (1-\lambda)\,\theta_i^{(i)} + \lambda \frac{R_c\left(\mathbf{p}_c^{(i)}\right)}{C_c\left(\mathbf{p}_c^{(i)}\right)}, \quad \forall c \in \mathcal{C}. \tag{2.28}$$

At the next iteration, we solve again the parametric problem (2.12) with the new parameter $\mathbf{x}^{(i+1)}$. This iterative process is carried on until convergence.

Before presenting the convergence of this algoritm in the following proposition, we should discuss the importance of the parameter updates in (2.27) and (2.28). As mentioned earlier, the parameter α_c plays the role of a priority weight in the parametric problem $(P_{\mathbf{x}})$ whereas θ_c is a pricing factor that corresponds to the power penalty term of cell c. Intuititely, the priority weight $\alpha_c^{(i+1)}$ for the next iteration is calculated as a weighted sum of the most recent value $\alpha_c^{(i)}$ and the inverse of the power cost $C_c\left(\mathbf{p}_c^{(i)}\right)$. The pricing factor $\theta_i^{(i+1)}$ is similarly updated based on the energy efficiency achieved by cell c in the last iteration.

Proposition 2.3. *The iterative damped Newton Algorithm 2.1 converges to an optimal solution of the power allocation problem* (P) *in* (2.10).

Algorithm 2.1: Algorithmic procedure for solving (P)

Initialize: $\epsilon,\ \mathbf{0} < (\mu, \lambda, \sigma) < \mathbf{1},\ \mathbf{p}^{(0)} \in \mathbb{R}^{N|\mathcal{U}|}$.
Set $\mathbf{x}^{(0)}$ using (2.23) and (2.24) and set $i = 0$,
Repeat
(S.1) : Compute $\widetilde{\mathbf{p}}\left(\mathbf{x}^{(i)}\right)$ by solving the parametric problem (2.12),
(S.2) : Initialize $\lambda \leftarrow 1$,
(S.3) : Compute $\mathbf{x}^{(i)} + \lambda \mathbf{d}^{(i)}$ using (2.27)-(2.28),
(S.4) : **while** $\left\|\mathbf{F}\left(\mathbf{x}^{(i)} + \lambda \mathbf{d}^{(i)}\right)\right\| > (1 - \mu\lambda)\left\|\mathbf{F}\left(\mathbf{x}^{(i)}\right)\right\|$ **do** $\lambda \leftarrow \sigma\lambda$
(S.5) : Update $\mathbf{x}^{(i+1)} \leftarrow \mathbf{x}^{(i)} + \lambda \mathbf{d}^{(i)}$
Until $\left\|\mathbf{F}\left(\mathbf{x}^{(i)}\right)\right\| \leq \epsilon$

Proof. According to Proposition 2.1, an optimal solution of (P) must also be a solution of a parametric problem $(P_{\mathbf{x}})$ whose parameter $\mathbf{x}$ is necessarily a solution of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$. In Proposition 2.2, we have shown that $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ has a unique solution. Thus, Algorithm 2.1 can retrieve an optimal solution of (P) by iteratively finding the unique solution of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ and solving the parametric convex problem $(P_{\mathbf{x}})$. A rigorous convergence analysis of the Newton method is available in [62, 63]. □

2.6 General Case: Non-orthogonal Transmissions

In this section, we tackle the energy efficiency maximization problem in (2.2) for the general scenario. In other words, we relax the assumption that the transmissions are orthogonal. In other words, the cells are now assumed to fully reuse the spectrum and may interfere with each other. To tackle the difficulty in solving this optimization problem, we apply the principle of minorization-maximization in conjunction with the previous Newton method. First, let us give the definition of a minorization.

Definition 2.1. Given a function f defined over $\mathcal{K} \subset \mathbb{R}^n$, another function g defined over $\mathcal{K} \times \mathcal{K}$ is said to minorize f if the following conditions are satisfied:

$$f(\mathbf{p}) \geq g(\mathbf{p}, \mathbf{q}) \quad \forall (\mathbf{p}, \mathbf{q}) \in \mathcal{K} \times \mathcal{K}, \tag{2.29}$$

$$f(\mathbf{p}) = g(\mathbf{p}, \mathbf{p}) \quad \forall \mathbf{p} \in \mathcal{K}. \tag{2.30}$$

In other words, g as a function of $\mathbf{p}$ and $\mathbf{q}$ is upper-bounded by f and coincides with f when $\mathbf{q} = \mathbf{p}$. This property is very useful when we want to maximize a complicated function f and have access to a simpler minorizing function g. Instead of solving f directly, we use the following sequential procedure to solve the original problem:

$$\mathbf{p}^{(j+1)} = \underset{\mathbf{p}\in\mathcal{K}}{\operatorname{argmax}}\, g\left(\mathbf{p}, \mathbf{p}^{(j)}\right), \tag{2.31}$$

where at each iteration j, we compute $\mathbf{p}^{(j+1)}$ by maximizing the minorization function g with $\mathbf{q} = \mathbf{p}^{(j)}$. The minorization-maximization (MM) is a general principle for solving non-convex optimization problems. For example, it has been widely used in statistics and machine learning [64]. It has also surfaced in wireless resource allocation and localization problems in [65–67]. In these cited works, the utility function consists of a difference of convex (D.C) functions; and the MM principle directly leads to the convex-concave procedure (CCCP) [68], a special algorithm for D.C programs. In contrast, our problem does not have this form so that the CCCP algorithm is not directly applicable.

2.6.1 Problem Reformulation

As mentioned previously, the sum energy efficiency function is not convex because of the interference coupling in the SINR expression and of its sum-fractional form. Therefore, it is useful to find a minorizing function of the energy efficiency utility function.

For convenience, let us define for each user $u \in \mathcal{U}$ the following two functions:

$$v_u\left(\mathbf{p}_{-u}\right) = \sum_{n=1}^{N} \log\left(\sigma_{u,n}^2 + \sum_{k\in\mathcal{U}\setminus\{u\}} \mathbf{p}_{k,n} g_{k,n}\right), \tag{2.32}$$

$$l_u\left(\mathbf{p}_{-u}, \mathbf{q}_{-u}\right) = v_u\left(\mathbf{q}_{-u}\right) + \nabla v_u\left(\mathbf{q}_{-u}\right)^{\top}\left(\mathbf{p}_{-u} - \mathbf{q}_{-u}\right), \tag{2.33}$$

where $\mathbf{p}_{-u} \in \mathbb{R}^{N(|\mathcal{U}|-1)}$ is the power allocation of all users except user u. The gradient of $v_u\left(\mathbf{q}_{-u}\right)$ in (2.33) is a vector of length $N\left(|\mathcal{U}|-1\right)$ given by

$$\nabla v_u\left(\mathbf{q}_{-u}\right)=\left(\frac{g_{uj,n}}{\sigma_{u,n}^2+\sum\limits_{k\in\mathcal{U}\setminus\{u\}}\mathbf{p}_k g_{ik}}\right)_{j\in\mathcal{U}\setminus\{u\},\ n\in\mathcal{S}}. \tag{2.34}$$

Then, we have the following result which provides a way to solve the energy efficiency maximization (2.2) in its general form.

Lemma 2.2. *A minorization function for the sum energy efficiency function is given by*

$$g\left(\mathbf{p},\mathbf{q}\right)\triangleq\frac{W}{N}\sum_{c\in\mathcal{C}}\frac{\sum_{u\in\mathcal{U}_c}\left[\sum_{n=1}^{N}\log\left(\sigma_{u,n}^2+\sum\limits_{j\in\mathcal{U}}\mathbf{p}_{j,n}g_{uj,n}\right)-l_u\left(\mathbf{p}_{-u},\mathbf{q}_{-u}\right)\right]}{C_c\left(\mathbf{p}_c\right)}. \tag{2.35}$$

Proof. The sum energy efficiency function (2.3) can be rewritten as

$$f\left(\mathbf{p}\right)=\sum_{c\in\mathcal{C}}\frac{\sum_{u\in\mathcal{U}_c}\left[\sum_{n=1}^{N}\log\left(\sigma_{u,n}^2+\sum\limits_{j\in\mathcal{U}}\mathbf{p}_j g_{uj}\right)-v_u\left(\mathbf{p}_{-u}\right)\right]}{C_c\left(\mathbf{p}_c\right)},$$

where the function v_u defined in (2.32) is a concave function of $\mathbf{p}_{-u}$. Therefore, it is bounded above by its first-order Taylor approximation l_u defined in (2.33) as

$$v_u\left(\mathbf{p}_{-u}\right)\leq l_u\left(\mathbf{p}_{-u},\mathbf{q}_{-u}\right),\quad\forall\mathbf{p}_{-u},\mathbf{q}_{-u},\ \forall u\in\mathcal{U} \tag{2.36}$$

and the equality holds at $\mathbf{q}_{-u}$. Furthermore, since the power cost $C_c\left(\mathbf{p}_c\right)$ is always a positive function, then we can deduce that:

$$g\left(\mathbf{p},\mathbf{q}\right)=\frac{W}{N}\sum_{c\in\mathcal{C}}\frac{\sum_{u\in\mathcal{U}_c}\left[\sum_{n=1}^{N}\log\left(\sigma_{u,n}^2+\sum\limits_{j\in\mathcal{U}}\mathbf{p}_{j,n}g_{uj,n}\right)-l_u\left(\mathbf{p}_{-u},\mathbf{q}_{-u}\right)\right]}{C_c\left(\mathbf{p}_c\right)}\leq f\left(\mathbf{p}\right). \tag{2.37}$$

In other words, the function $g\left(\mathbf{p},\mathbf{q}\right)$ defined in (2.35) minorizes the sum energy efficiency function $f\left(\mathbf{p}\right)$ and the equality holds when $\mathbf{q}=\mathbf{p}$. □

The benefit of using g as the objective function of the surrogate maximization problem (2.31) is that we can compute its global optimal solution using the damped Newton

Algorithm 2.2: Minorization-Maximization algorithm based on Newton method

Initialize: $\epsilon, \mathbf{p}^{(0)} \in \mathbb{R}^{N|\mathcal{U}|}$ and set $j = 0$.

Repeat

(S.1) : Minorize the EE utility function f at $\mathbf{p}^{(j)}$ using (2.35)

(S.2) : Update the power allocation as $\mathbf{p}^{(j+1)} = \underset{\mathbf{x}\in\mathcal{K}}{\operatorname{argmax}}\ g\left(\mathbf{p}, \mathbf{p}^{(j)}\right)$

(S.3) : $j \leftarrow j + 1$

Until $\frac{f\left(\mathbf{p}^{(j+1)}\right) - f\left(\mathbf{p}^{(j)}\right)}{f\left(\mathbf{p}^{(j)}\right)} \leq \epsilon$

method presented in Section 2.5. This is explained in more details in the next subsection.

2.6.2 Minorization-Maximization Algorithm

We derive an iterative algorithm for the sum energy maximization problem for the case of non-orthogonal transmissions using the minorization-maximization (MM) procedure. At each iteration j, a minorization function $g\left(\mathbf{p}, \mathbf{p}^{(j)}\right)$ of the sum energy efficiency is obtained using equation (2.35). Then, we maximize this surrogate function $g\left(\mathbf{p}, \mathbf{p}^{(j)}\right)$ using the damped Newton method presented in Section 2.5. Each solution of this subproblem gives us a new power allocation $\mathbf{p}^{(j+1)}$ which is used to provide a new minorizing function $g\left(\mathbf{p}, \mathbf{p}^{(j+1)}\right)$ for the next iteration. This algorithm, which is outlined in Algorithm 2.2, has a nice convergence property as stated by the following proposition.

Proposition 2.4. *The minorization-maximization algorithm monotonically converges to at least a local optimum of the sum energy efficiency maximization problem.*

Proof. To show that the algorithm is monotonically convergent, we prove that:

$$f\left(\mathbf{p}^{(j+1)}\right) \geq f\left(\mathbf{p}^{(j)}\right),\ \forall j. \tag{2.38}$$

According to Lemma 2.2, g minorizes the energy efficiency function f. Thus by (2.29), we have:

$$f\left(\mathbf{p}^{(j+1)}\right) \geq g\left(\mathbf{p}^{(j+1)}, \mathbf{p}^{(j)}\right). \tag{2.39}$$

Next, we show that the damped Newton method can be used to compute:

$$\mathbf{p}^{(j+1)} = \underset{\mathbf{x}\in\mathcal{K}}{\operatorname{argmax}}\; g\left(\mathbf{p}, \mathbf{p}^{(j)}\right). \tag{2.40}$$

Using (2.35), the minorization maximization problem (2.40) can be written in an epigraph form:

$$\begin{cases} \underset{\mathbf{p},\widetilde{\theta}}{\text{maximize}} & \sum_{c\in\mathcal{C}} \widetilde{\theta}_c \\ \text{subject to} & \widetilde{R}_c\left(\mathbf{p}\right) - \widetilde{\theta}_c C_c\left(\mathbf{p}_c\right) \geq 0,\ \forall c \in \mathcal{C}, \\ & h_l\left(\mathbf{p}\right) \leq \mathbf{0},\ \forall l \in \mathcal{V}, \end{cases} \tag{2.41}$$

where $\widetilde{\theta}$ are new auxilliary variables.

The minorized functions $\widetilde{R}_c$ in (2.41) are given by

$$\widetilde{R}_c\left(\mathbf{p}\right) \triangleq \sum_{u\in\mathcal{U}_c}\left[\sum_{n=1}^{N}\log\left(\sigma_{u,n}^2 + \sum_{j\in\mathcal{U}}\mathbf{p}_{j,n}g_{uj,n}\right) - l_u\left(\mathbf{p}_{-u}, \mathbf{q}_{-u}\right)\right],\ \forall c \in \mathcal{C}.$$

In contrast to the cell sum rate function, each minorized function $\widetilde{R}_c$ is strictly concave in $\mathbf{p}$. As a result, the surrogate problem in (2.40) share the same properties as the problem (P) in (2.10). This means that we can also use the convex parametric approach presented in Section 2.5 to solve (2.40). Hence, the statements in Propositions 2.1 and 2.2 also hold for (2.40). Thus, the damped Newton method in Algorithm 2.1 can be used to find the power allocation $\mathbf{p}^{(j+1)}$ that maximizes the function $g\left(\mathbf{p}, \mathbf{p}^{(j)}\right)$.

Therefore, the following holds at the j-th iteration of Algorithm 2.2:

$$g\left(\mathbf{p}^{(j+1)}, \mathbf{p}^{(j)}\right) \geq g\left(\mathbf{p}^{(j)}, \mathbf{p}^{(j)}\right) = f\left(\mathbf{p}^{(j)}\right), \tag{2.42}$$

where the last equality is due to the definition of minorization function in (2.30). By combining (2.39) and (2.42), we proved that $f\left(\mathbf{p}^{(j+1)}\right) \geq f\left(\mathbf{p}^{(j)}\right),\ \forall j$. □

Even if the original problem is not convex, our proposed algorithm can produce an efficient power allocation using the minorization-maximization procedure and Newton

method. In the next section, we verify the convergence properties and analyze the performance of these algorithms.

2.6.3 Practical Issues for Implementation

Given the above-proposed algorithm, we should discuss two aspects that are important for implementation.

First, the proposed resource allocation scheme relies on accurate channel state estimates. This downlink channel estimation can be done through pilot signal training either directly from the base stations to the mobile users in a Frequency Division Duplex system (FDD) or reversely from the mobile users to the base stations in a Time Division Duplex system (TDD). The later is preferable assuming the channel reciprocity is preserved in the small-cell propagation environment. In fact, the direct channel estimation requires a feedback link for the users to communicate the channel estimates to the base stations. This may be expensive in terms of spectrum efficiency especially since it has to be done over the air. Exploiting channel reciprocity in a TDD system, we can avoid this overhead. While we assume perfect channel information between the base stations and served users of the two-tier network, it is more difficult to acquire perfect channel information for the channel towards primary users due to the lack of coordination. In this case, we explicitly consider the channel uncertainty that may be present in the interference constraints in the next section.

Second, it is also crucial that the resource allocation must satisfy a delay requirement that does not exceed the channel coherence time. This coherence time approximately equals to the inverse of Doppler spread and depends on the mobility of the users. We can envisage that small-cells are mostly deployed in urban areas where most communication happens indoor. In such case, the coherence time can be within $500 - 3000$ms at a 2.5Ghz frequency band [69]. Therefore, the channel state information must be shared by the base stations to the central unit before the channel state changes. To meet this re-

quirement, high-capacity front-haul links are able to meet such demand. In fact, fronhaul link requirements, as specified by mobile network operators, should have a delay in the order of 45-250ms with a maximum supporting distance between the RRU and CU of 50km [70]. Our proposed method can implemented by solving the inner problem with standard convex optimization algorithms, such as interior-point methods which have a polynomial-time complexity [61]. As shown in our simulation results, the algorithm converges very fast, within 10 iterations. Therefore, it is practical to be used in the presented centralized framework since the central unit has a much higher processing power than a standard base station [71, 72].

2.7 Power Allocation with Probabilistic Interference Constraints

In the following, we consider the energy efficient resource allocation with probablistic interference power constraints due to the channel uncertainty between the primary users and the small-cell BSs. To handle the intractability of the probabilistic constraints (2.7), we replace them by robust approximation framework [73]. In other words, we replace the probabilistic constraints with conservative yet convex ones. Denote the concatenation of channel uncertainties for each primary user l by the vector $\zeta_l \triangleq (\zeta_{lc,n})_{c\in\mathcal{C},n\in\mathcal{S}}$. By substituting the channel gain (2.8), we can rewrite (2.7) for each c as:

$$\Pr\left\{\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}\Delta g_{lc,n}p_{u,n}\zeta_{lc,n} \leq I_{\max} - \sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}p_{u,n}\widehat{g}_{lc,n}\right\} \geq 1-\epsilon \tag{2.43}$$

Ball-box approximation

The above constraint (2.7) can be replaced by the ball-box approximation given by [73]:

$$\sum_{u\in\mathcal{U}} p_{un}\left(\widehat{g}_{lc,n}+\Delta g_{lc,n}\zeta_{lc,n}\right)\leq I_{\max},\ \forall\zeta_l\in\mathcal{Z}_1 \tag{2.44}$$

$$\mathcal{Z}_1\triangleq\left\{\zeta_l\mid\left\|\zeta_l\right\|_2\leq\Omega,\ \left\|\zeta_l\right\|_\infty\leq 1\right\} \tag{2.45}$$

which is equivalent to the following system of constraints:

$$z_{lc,n}+u_{lc,n}=-\Delta g_{lc,n}p_{u,n},\ \forall u\in\mathcal{U}_c,\ \forall c\in\mathcal{C} \tag{2.46}$$

$$\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}|z_{lc,n}|+\Omega\sqrt{\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}u_{lc,n}^2}\leq I_{\max}-\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}p_{u,n}\widehat{g}_{lc,n} \tag{2.47}$$

where $\{z_{ilc}\}_i$ and $\{u_{ilc}\}_i$ are new slack variables. The following proposition, which is due to [73], states that we can use the above constraints to replace the probabilistic constraints (2.43).

Proposition 2.5. *[73] Every feasible power allocation* $\mathbf{p}$ *of the system of constraints* (2.46)-(2.47) *also satisfies the probabilistic interference constraint* (2.43) *provided that* $\Omega\geq\sqrt{2\ln\left(1/\epsilon\right)}$.

Thus, all feasible power allocations with respect to the above constraints (2.46) and (2.47) satisfy the interference power constraint $\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}p_{u,n}\left(\widehat{g}_{lc,n}+\Delta g_{lc,n}\zeta_{lc,n}\right)\leq I_{\max}$ with a probability of at least $1-\epsilon$.

In contrast to the probabilistic constraint (2.43), constraints (2.46) are more tractable since they are represented by convex conic constraints. Interestingly, constraint (2.46) can also be interpreted as a robust interference constraint in which the channel uncertainty belongs to the intersection of a ball with radius Ω and a box with an edge length of 2 both centered at the origin.

Budgeted robust approximation

Alternatively, we also consider the budgeted robust approximation of the probabilistic constraints (2.43) defined by:

$$\sum_{u\in\mathcal{U}} p_{un}\left(\widehat{g}_{lc,n}+\Delta g_{lc,n}\zeta_{lc,n}\right)\leq I_{\max},\ \forall\zeta_l\in\mathcal{Z}_2 \tag{2.48}$$

$$\mathcal{Z}_2\triangleq\left\{\zeta_l\mid\|\zeta_l\|_\infty\leq 1,\ \sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}|\zeta_{lc,n}|\leq\gamma\right\} \tag{2.49}$$

It can also be shown that the above budgeted robust constraints (2.43) can be equivalently represented by the following finite system of linear constraints [73]:

$$z_{lc,n}+u_{lc,n}=-\Delta g_{lc,n}p_{u,n},\ \forall u\in\mathcal{U}_c,\ \forall c\in\mathcal{C} \tag{2.50}$$

$$\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}|z_{lc,n}|+\gamma\max_{c\in\mathcal{C}}|u_{lc,n}|\leq I_{\max}-\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}p_{u,n}\widehat{g}_{lc,n} \tag{2.51}$$

Then, the following proposition gives the value of the budget parameter γ that guarantees the satisfaction of the probabilistic constraints (2.43) [73].

Proposition 2.6. *[73] Let $(\mathbf{p}_i)_{i\in\mathcal{U}}$ be a feasible solution of the system of constraints* (2.50)-(2.51)*, then it also satisfies the probabilistic interference constraint* (2.43) *provided that we set $\gamma\geq\sqrt{2\left|\mathcal{U}\right|\cdot\left|\mathcal{C}\right|\ln\left(1/\epsilon\right)}$.*

Compared to the ball-box approximation (2.46), the budgeted approximation is more conservative when the uncertainty budget parameter γ is linked to the safety parameter Ω in the ball-box approximation according to $\gamma=\Omega\sqrt{|\mathcal{U}|\cdot|\mathcal{C}|}$. However, the advantage of the budgeted approximation is that we could model the robustness with linear constraints, thus reducing the computational complexity especially if one wants to decompose the optimization problem.

Worst-case robust interference constraints

Finally, it is possible to consider a worst-case robust approach by approximating the interference constraints with:

$$\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N} p_{u,n}\left(\widehat{g}_{lc,n}+\Delta g_{lc,n}\zeta_{lc,n}\right)\leq I_{\max},\ \forall\zeta_l\in\mathcal{D}_l \tag{2.52}$$

$$\mathcal{D}_l\triangleq\{\zeta_l\mid -1\leq\zeta_{il,c}\leq 1,\ \forall\,(i,c)\in\mathcal{U}\times\mathcal{C}\} \tag{2.53}$$

This is also equivalent to:

$$\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}|\Delta g_{lc,n}p_{u,n}|\leq I_{\max}-\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{U}_c}\sum_{n=1}^{N}p_{u,n}\widehat{g}_{lc,n} \tag{2.54}$$

Consequently, the energy efficiency optimization framework can also handle probabilistic power constrains using the robust approximation framework presented above.

2.8 Simulation Results

In this section, we present our simulation results for a two-tier wireless system composed of M primary users served by the macrocell and N secondary users, each of which is served by a hotspot picocell BS. Based on the WINNER models [74], the large-scale path loss attenuation PL (in dB) is calculated as a function of the distance (in meters) and carrier frequency (f_c in GHz):

$$PL=A\log_{10}(d)+B+C\log_{10}\left(\frac{f_c}{5.0}\right)+S\quad[dB] \tag{2.55}$$

where S is the random log-normal shadowing. For the two-tier cellular network, the model parameters for each respective scenario are detailed in Table 2.1. The maximum values of total transmit powers and the small-cell BSs' circuit-power consumption are also listed in

Table 2.1 – Simulation parameters in Chapter 2.

Parameters	Values
Carrier frequency	$f_c = 1.9$ GHz
System bandwidth	5 MHz
Sub-channel bandwidth	180 kHz
Number of primary macro users	5
SBS-to-SU path loss (model A1-LOS)	$A = 1.87,\ B = 46.8,\ C = 20$
SBS-to-PU path loss (model A1-NLOS)	$A = 3.68,\ B = 43.8,\ C = 20$
MBS-to-SU path loss (model C1-NLOS)	$A = 3.36,\ B = 44.36,\ C = 23$
MBS antenna height	$h_{\text{MBS}} = 50$ m
Standard deviation of SBS-to-SU shadowing	$S = 3$ dB
Standard deviation of SBS-to-PU shadowing	$S = 4$ dB
Standard deviation of MBS-to-SU shadowing	$S = 6$ dB
MBS maximum transmit power	46 dBm
SBS maximum transmit power	36 dBm
SBS circuit power consumption	20 W
Noise Power Density	-174 dBm/Hz

Table 2.1. Since we consider only a transmission on one sub-channel, we normalized these values in accordance with the subchannel bandwidth.

With uniform distribution, we randomly placed N small-cell BS in a 1km by 1km grid region, at the center of which the macrocell BS is located. The optimization of small-cell transmissions follows after the cell association. Using a simple weighted max-SNR association rule, we determined the macrocell and small-cell coverage regions as multiplicatively weighted Voronoi tesselations [75]. The weigths were specified by the effective cell association bias β between the macrocell and small cells which can be controlled to expand or shrink the small-cells' range. Precisely, β is a weighted ratio of the received SNR from macrocell BS versus small-cell BS [76]. For simplifying our analysis, we place all small-cell users on their cell edges as illustrated in Fig. 2.2.

2.8.1 Convergence Results

First, we study the convergence of the proposed algorithms for the case of orthogonal transmissions and $I_{\max} = 0$dB with respect to noise level. For simplicity, we set the interference power limit $I_{\max}$ to be equal for every macrocell user. In Fig. 2.3a, a fast convergence is observed for the damped Newton algorithm. Its convergence is displayed in terms of $\|\mathbf{F}(\mathbf{x})\|$. We see that the optimal parameter $\mathbf{x}^*$ can be found within 5 iterations for a specified tolerance $\epsilon = 10^{-4}$.

For the general case, we employ Algorithm 2.2 to find an energy-efficient power allocation. Fig. 2.3b shows that it achieves near-optimal performance within a total of 10 inner iterations.

2.8.2 Performance Benchmark

In the following, we compare the average performance of the proposed spectrum sharing scheme with that of the algorithm in [56], which is based on non-cooperative game theory.

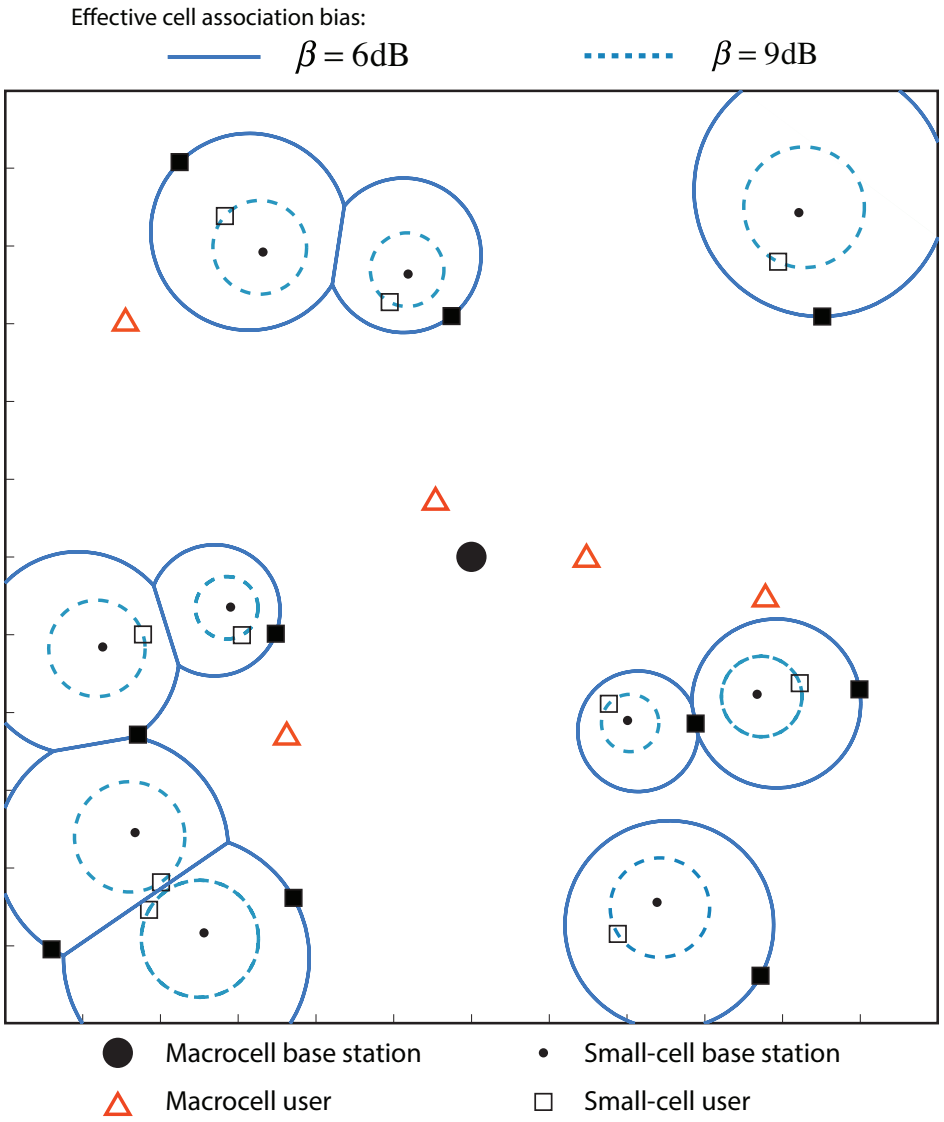
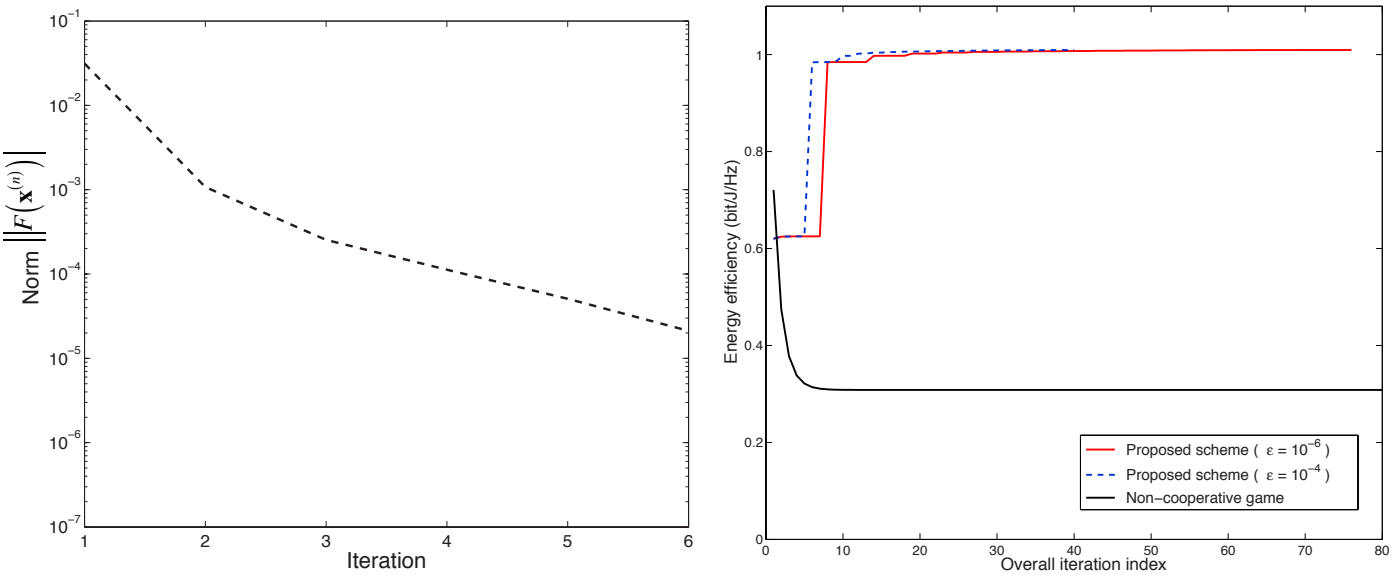


Figure 2.2 – Cell association regions of a two-tier network with 5 macro-cell users and 10 small-cell BSs. Solid and dashed circular lines show the cell edges with effective cell association bias β of 6 dB and 9 dB respectively. Small-cell users are randomly located on the cell edges.



(a) Damped Netwon method with $N = 3$, $M = 2$.

(b) Minorization-Maximization algorithm with $N = 10$ and $M = 5$.

Figure 2.3 – Convergence of proposed algorithms.

The comparison is made for different number of small cells N and for different values of the cell association bias β and of the interference limit $I_{\max}$ (in dB) relative to the noise power. For a given set of simulation parameters $(N, \beta, I_{\max})$, we generate 200 random instances of the network topology and channel realization, over which we average the performance of the proposed and comparative schemes.

First, the average sum energy efficiency is shown in Fig. 2.4. It is seen that our proposed scheme outperforms the non-cooperative approach. When an optimal power allocation is employed, the energy efficiency increases with the number of active small cells. In contrast, the performance achieved by the Nash equilibrium solution decreases when more small cells compete in a selfish way. This result is also reflected in the convergence behavior of the non-cooperative game as shown in Fig. 2.3b. Furthermore, Fig. 2.4 shows that higher energy efficiency can be obtained when small-cell uses shorter cell range. This corresponds to $\beta = 9$dB in our example. Next, Fig. 2.5 shows that our proposed scheme provides a better sum rate performance compared to non-cooperative scheme, especially when the small cells' range are bigger, i.e. with $\beta = 6$dB.

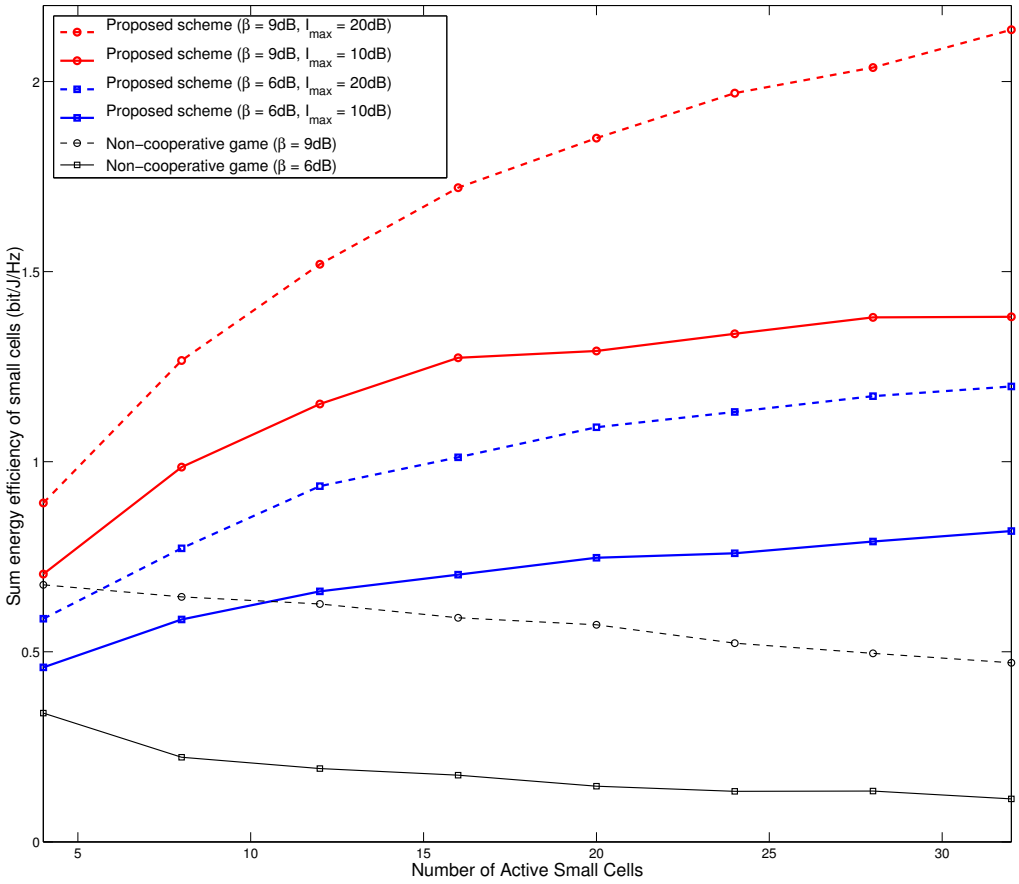


Figure 2.4 – Average sum energy efficiency comparison for different number of small cells.

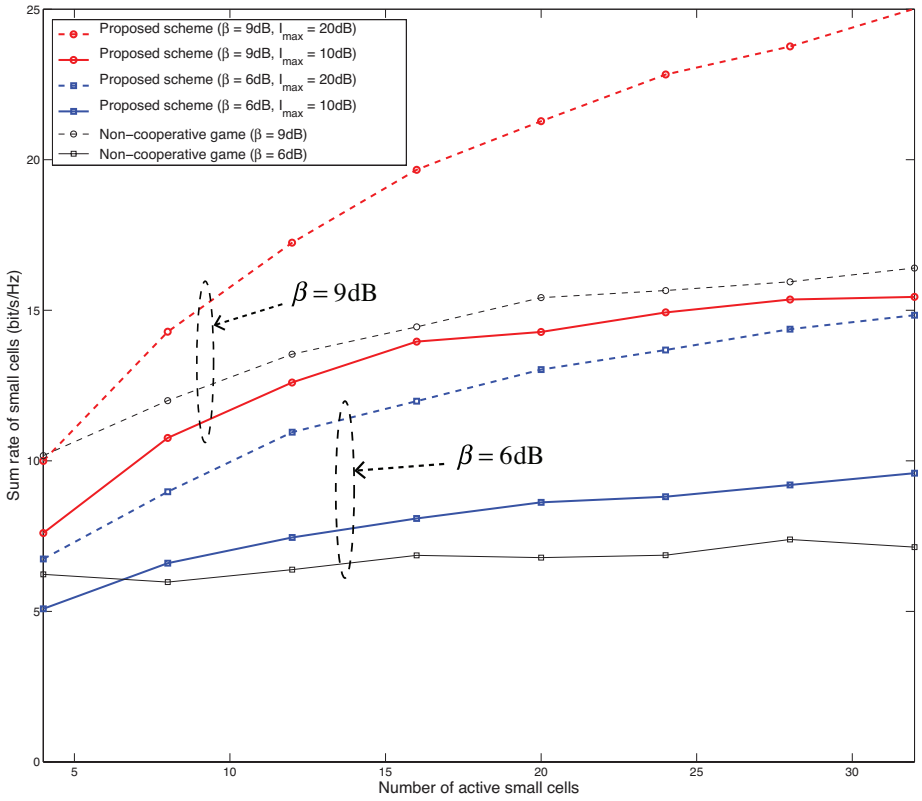


Figure 2.5 – Average sum rate comparison for different number of small cells.

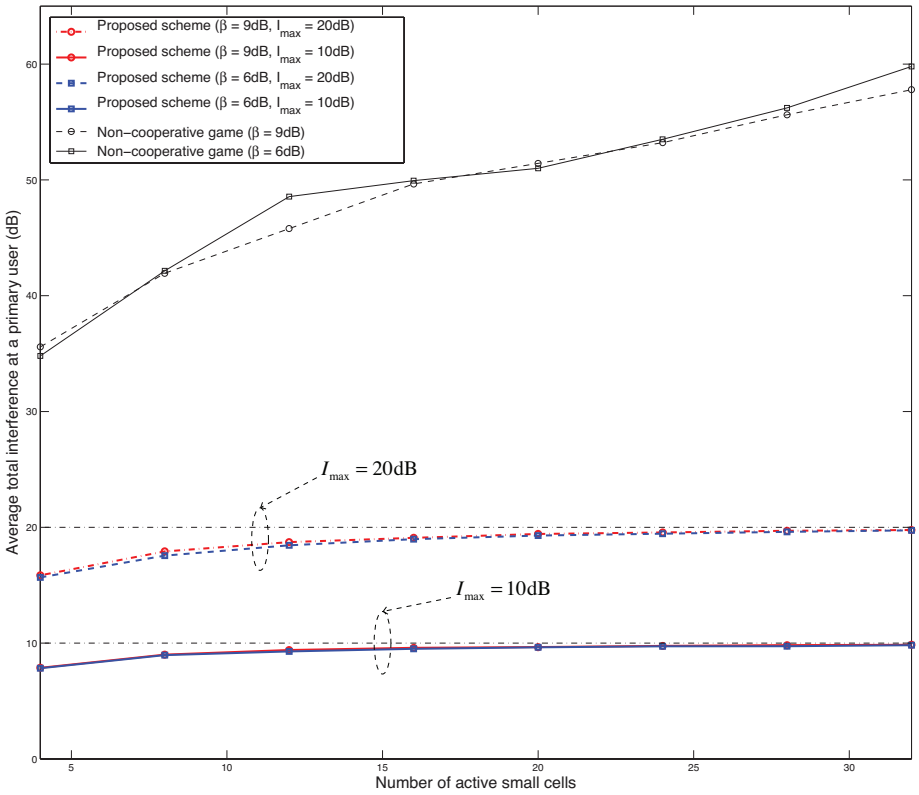


Figure 2.6 – Average total interference power at a macro-cell user.

As we expected, the performance of our scheme slightly decreases when the interference limit imposed by the macrocell users is stringent. For instance, when $I_{\max} = 10$dB and $\beta = 9$dB, it appears that the non-cooperative scheme provides a higher sum rate than our proposed scheme. Nonetheless, the comparison is not fair since the non-cooperative scheme ignores the interference power limits imposed by the macrocell users. Indeed, Fig. 2.6 shows that with the non-cooperative scheme, the interference power received by the primary users always exceed the limit imposed by up to 40dB margin when there are many active small-cells in the two-tier network. In contrast, our proposed scheme always satisfies the total interference limit imposed by the primary users.

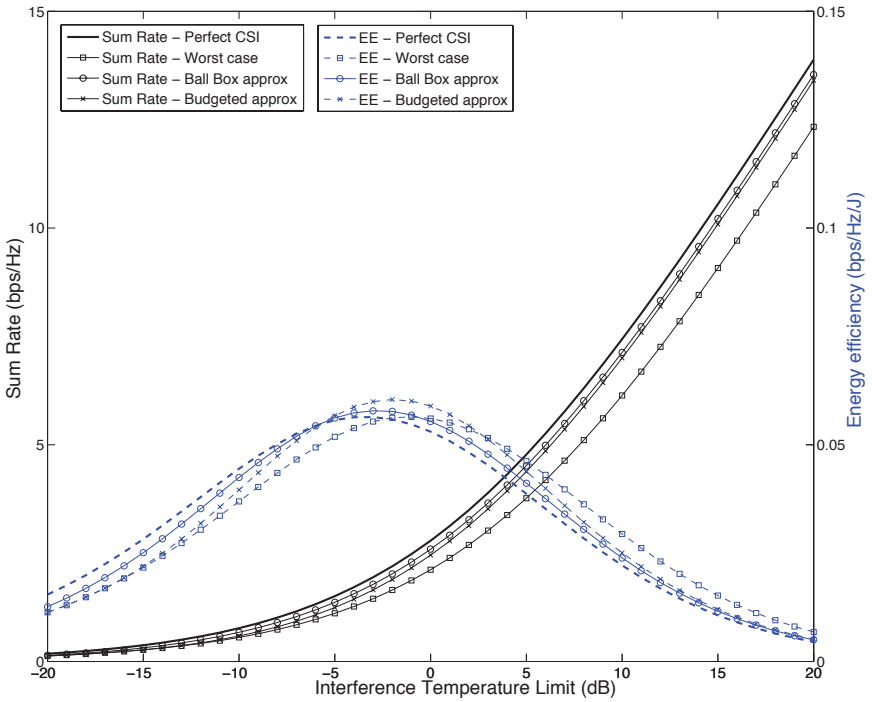


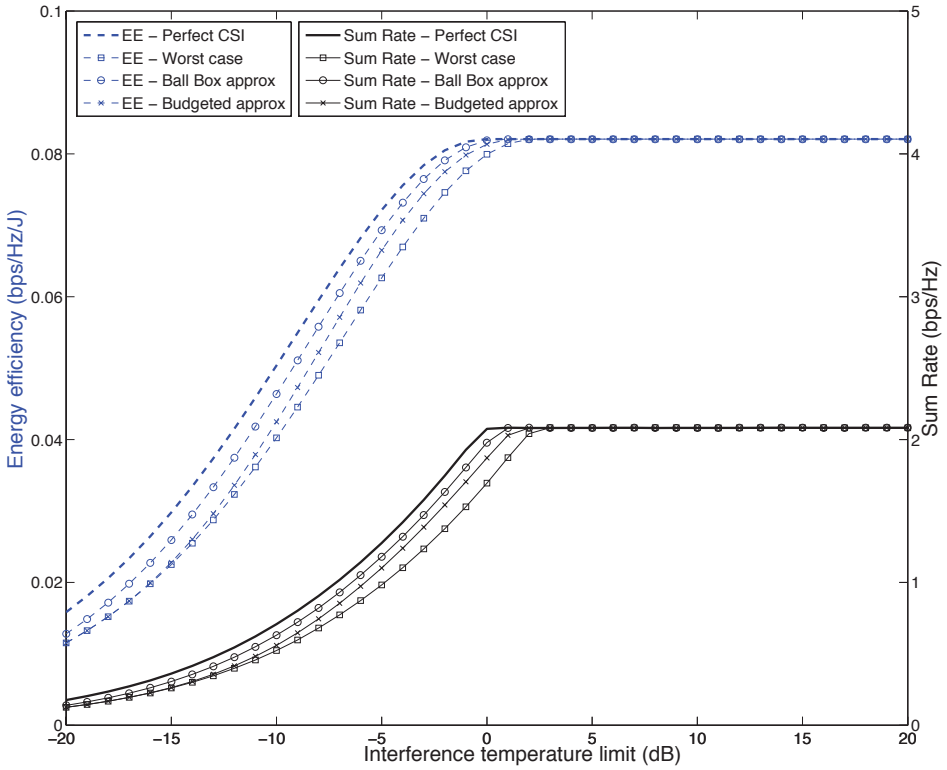
Figure 2.7 – Performance of sum rate-optimal power allocation vs. $I_{\max}$ with $P_0 = 10$ dB.

2.8.3 Impact of Imperfect Channel Information

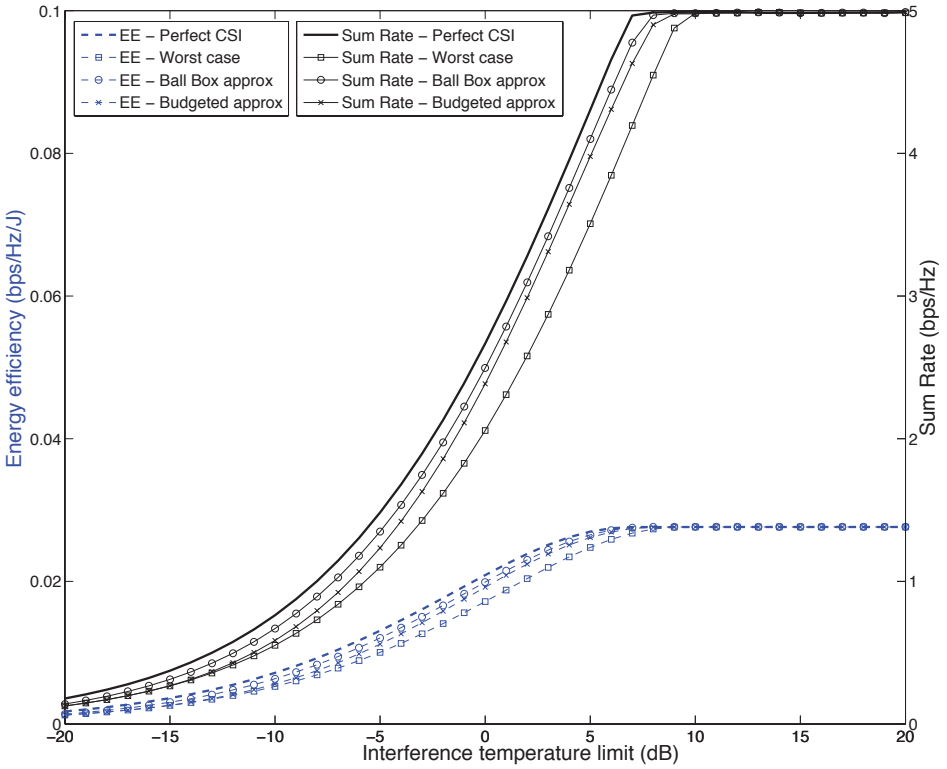
Here, we simulate the resource allocation with probabilistic interference constraints for a system of 3 small-cells and 2 primary users. We use $|\mathcal{C}| = 64$ subcarriers. For the probabilistic interference constraints, we set the threshold to $\epsilon = 0.1$. For simplicity, we assume that the maximum channel estimation error is set to $\Delta g_{lc,n} = 0.7 \times \widehat{g}_{lc,n}$.

First, we plot in Figure 2.7 the performance of the rate-maximizing power allocation. We see that the optimal sum rate increases to infinity with the allowable interference temperature limit. In contrast, the resulting energy efficiency reaches a maximum at $I_{\max}^{\text{thres}} = -4$dB. Increasing the power beyond this threshold will largely reduce the energy efficiency.

In Figure 2.8, we show the performance of the energy-efficiency maximizing power allocation. In accordance with the previous observation, we see that the energy-efficiency and the corresponding sum rate saturate when the interference limit is larger than a certain threshold $I_{\max}^{\text{thres}}$. Beyond this threshold, the performance of the different robust power



(a) Fixed power $P_0 = 10$



(b) Fixed power $P_0 = 100$

Figure 2.8 – Performance of energy-efficient optimal power allocation vs. I_{max} for different P_0.

allocations are equal to that of perfect channel state information (CSI). This is because the mobile users keep their power transmission at the same level in order to avoid reducing the energy efficiency. Therefore, when $I_{\max} > I_{\max}^{\text{thres}}$ the energy-efficient transmission is automatically robust against channel uncertainty. When $I_{\max} \leq I_{\max}^{\text{thres}}$, we can observe that the worst-case robust power allocation has the lowest performance and that the budgeted approximation is more conservative than the ball-box approximation. Comparison between figures 2.8(a) and 2.8(b) show that the optimal energy efficiency decreases when the fixed power cost P_0 increases. However, the mobile users can achieve a higher sum rate up to 5 bps /Hz when $P_0 = 20$ dB and $I_{\max} = 5$dB. These results illustrate the crucial trade-off between spectrum and energy efficiencies.

Chapter 3

Energy Allocation and Cooperation for Hybrid-Powered Two-tier Networks

3.1 Introduction

In this chapter, we investigate the joint energy allocation and cooperation scheme for wireless two-tier networks. First, we study the energy-efficiency optimization problem with each cell having access to both a hybrid power source and an energy storage system. In other words, each base station is powered by both renewable sources and a conventional grid source. As previously mentioned, the large-scale deployment of small-cell base stations can pose a problem to the overall energy efficiency of multi-tier networks [8]. Therefore, new sustainable energy solutions are also needed to increase the network capacity in a cost-efficient manner.

In Section 3.4, we study the energy-efficiency maximization problem with hybrid power source. By utilizing hybrid source and energy storage, each cell can manage its energy over time based on the time-varying channel conditions and renewable energy arrivals. The objective is again to maximize the network energy efficiency by utilizing the available energy from the hybrid source in the most efficient manner over time. In doing so, we also seek to satisfy an average rate constraint at each cell. While the main design target is to improve the energy efficiency of two-tier networks, the spectrum sharing between macrocell and small cells is also enforced because of the scarcity of wireless spectrum [77].

To tackle the fluctuations of renewable energy arrivals, we also investigate the energy cooperation in the multi-cell resource allocation problem in Section 3.5. Specifically, we assume that each cell can transfer some of its harvested energy to other cells through a smart-grid power infrastructure. In this chapter, we focus on offline optimization and assume a non-causal information about the channel and energy arrivals. In fact, our aim is first to analyze the theoretical benefits of using hybrid power source and energy cooperation for wireless two-tier networks. In Section 3.6, we present numerical results and analyze the potential gain of the proposed schemes. As such, the resource optimization used in chapter is also performed in a centralized manner. For practical systems, it is however important to design online and distributed algorithms that coordinate and adapt the energy allocation between the cells over the time. For the design of such online algorithms, we later suggest some solution approaches in Section 5.2.

3.2 Related Works

The use of renewable energy and hybrid energy source for powering cellular networks is a relatively new idea. While many studies have considered pure energy-harvesting systems, only few have explored the use of hybrid power source and its implications. Previously, [10] studied the energy-efficient power allocation for a single-cell downlink system with a hybrid power source. However, [10] assumed an orthogonal access scheme and considered only a single-cell system. Instead, we consider the multi-cell two-tier system with full spectrum reuse. The idea of energy cooperation for cellular networks was initially proposed in [78] and [79] and is motivated by recent progress in renewable energy integration and two-way energy flow in smart grid [42]. However, the study in [78] used a simplified model by abstracting the energy demand at each cell. On the other hand, [79] proposed joint power-spectrum allocation schemes but it assumed that the cells do not store the harvested energy for future use. Instead, they optimized the energy cooperation

assuming independent time frames. In contrast to the previous works, we simultaneously exploit the spatial and time diversities associated with the harvesting of renewable energy across the cells in a dynamic fashion. When the cells have access to an energy storage, it is possible for them to dynamically allocate the stored energy depending on the time fluctuations of the energy harvesting. Similarly, energy sharing between the cells should help to mitigate the imbalance of energy arrivals across geographically separated cells.

3.3 System Model

Let us consider the downlink of a two-tier multicell network which consists of M macrocell base stations (MBSs) and K small-cell base stations (SBSs). Together, these base stations serve a total of N users. Let us denote the set of macrocells and small cells in the network by $\mathcal{M}$ and $\mathcal{K}$ respectively. Similarly, the set of users served by a macrocell m is defined by $\mathcal{U}_m$ whereas that of a small-cell k is defined by $\mathcal{U}_k$. Therefore, the set of all N users is $\mathcal{U} \triangleq \left(\bigcup_{m\in\mathcal{M}} \mathcal{U}_m\right) \cup \left(\bigcup_{k\in\mathcal{K}} \mathcal{U}_k\right)$. We assume that all macrocells and small-cells share the spectrum. Although the resource allocation problem and algorithms studied in this chapter can be readily applied to multi-carrier systems, we consider the transmission on a single channel in the following to simplify the exposition. The signal-to-interference-plus-noise power ratio (SINR) of a macrocell or small-cell user u at the t-th time slot is given by:

$$\gamma_{u,t} = \frac{g_{uu,t}p_{u,t}}{\sum\limits_{j\in\mathcal{U}\setminus\{u\}} g_{uj,t}p_{j,t} + \sigma^2_{u,t}} \tag{3.1}$$

where $g_{uu,t}$ and and $g_{uj,t}$ are the channel gains of user u at time t respectively from its BS serving and from an interfering BS that is serving another user j. Then, σ_u^2 is the variance of the AWGN noise at user u.

In (3.1), the channel gains accounts for the large-scale path loss attenuation, the log-normal shadowing due to obstacles in the environment and the small-scale Rayleigh fad-

Table 3.1 – List of key parameters in Chapter 3.

Symbol	Parameter description
M	Number of macrocell base stations
K	Number of small-cell base stations
N	Total number of users
$\mathcal{M}$	Set of macrocells
$\mathcal{K}$	Set of small cells
c	Cell index $c \in \mathcal{M} \cup \mathcal{K}$
$\mathcal{U}_c$	Set of users served by cell c
t	Time slot index
T_s	Time slot duration
$p_{u,t}$	Power allocated to user u at time slot t
$g_{uj,t}$	Channel gain of user u at time t from the BS serving user j
$\gamma_{u,t}$	SINR of user u at time slot t
$p^g_{u,t}$	Power drawn from the conventional source for user u at time slot t
$p^h_{u,t}$	Power drawn from the renewable source for user u at time slot t
f	Energy harvesting frame index
L	Number of time slots in one frame
N_F	Total number of frames
$H_{c,f}$	Renewable energy harvested by cell c from the renewable source during frame f
$\omega_{c,f}$	Excess power $\omega_{c,i}$ discarded by cell c at the end of frame f
B_c	Capacity of energy storage system at cell c
$\overrightarrow{h}_{c,d,f}$	Renewable energy transferred by cell c to cell d at the end of frame f
η	Energy transfer efficiency factor
$\delta_{c,i}$	Net energy injected by a cell c into the aggregator at the end of frame i

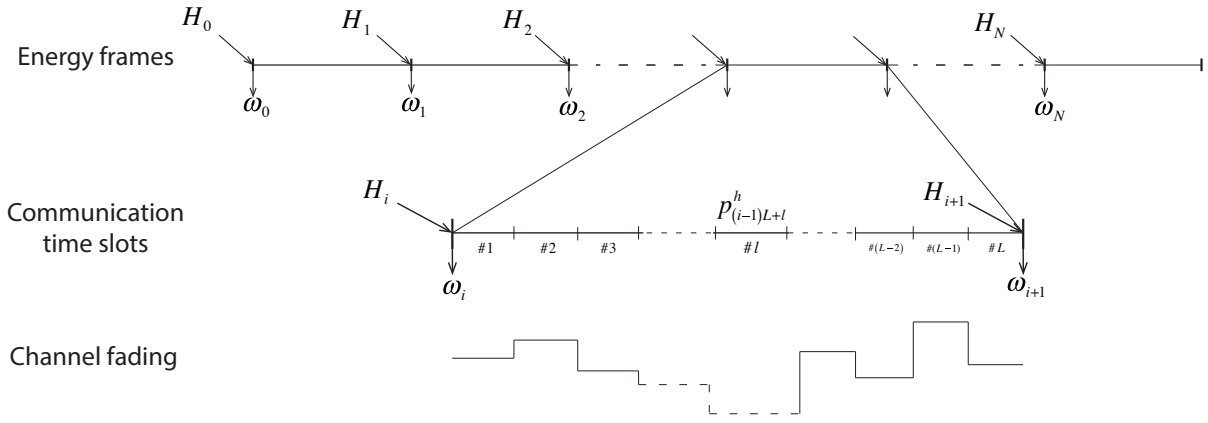
ing. Furthermore, we assume that the small-scale fading is slow and frequency-flat. The time correlation of the fading samples is modeled using an autoregressive model according to Clarke's scattering model [69, 80].

Remark 3.1. The flat fading assumption holds when the maximum excess delay due to multipath propagation is smaller than the symbol period of the transmitted signal. For modern wireless systems, the use of orthogonal frequency division multiplexing technique help reduce a wideband frequency-selective channel into parallel frequency-flat subchannels given the subcarrier bandwidth is sufficiently small. In this case, the coherence bandwidth becomes much larger than the subcarrier bandwidth.

In this work, we assume that every base station receives its energy from the conventional power grid source and from a renewable energy-harvesting source. Moreover, a battery is used by each BS to store the harvested energy from the renewable source. This stored energy will then be used for future transmission. Here, we assume that the battery has no leakage loss because the timeframe for the considered power allocation problem is relatively short. In (3.1), the power allocated to all users at time slot t is denoted by the vector $\mathbf{p}_t = (p_{u,t})_{u\in\mathcal{U}} \in \mathbb{R}_+^N$. Since each BS has a hybrid power source, we need to distinguish between the power $\mathbf{p}_t^g = \left(p_{u,t}^g\right)_{u\in\mathcal{U}} \in \mathbb{R}_+^N$ drawn from the conventional power grid and the transmit power $\mathbf{p}_t^h = \left(p_{u,t}^h\right)_{u\in\mathcal{U}\cup\mathcal{V}} \in \mathbb{R}_+^N$ from the energy-harvesting source. These two power usage vectors must add up to satisfy the equation:

$$p_{u,t}^g + p_{u,t}^h = p_{u,t},\ \forall u \in \mathcal{U},\ \forall t. \tag{3.2}$$

Next, we describe the generic energy management process of the cells. In our model, we have two different time scales for managing the renewable energy and for allocating the transmission power. Precisely, each energy-harvesting frame is divided into L communication time slots. Each time slot has a duration T_s. During the i-th frame, each cell $c \in \mathcal{M} \cup \mathcal{K}$ harvests a total amount of renewable energy $H_{c,i}$ from the renewable source.



H_i : Total harvested power during energy timeframe *i*

ω_i : Discarded power at the end of energy timeframe *i*

$p^h_{(i-1)L+l}$: Renewable power allocated at time slot *l* of the frame *i*

Figure 3.1 – Model for renewable energy management.

At the end of the frame, some of this energy is stored into its battery, whose maximum capacity is denoted by B_c. Then, the BS can allocate some of the stored renewable energy for transmission during the next frame or the subsequent ones. In summary, the BSs manage their harvested energy per energy-harvesting frame and allocate the stored energy for communication over the time slots. In our model, we consider a total transmit period of $T = LN_F$ time slots where N_F is the number of frames. This model is illustrated in Figure 3.1.

As explained above, each cell can store a part of the harvested energy in a battery for future use. To make sure that the cells can use only the energy that is available before the beginning of each frame, the following energy causality constraints need to be satisfied:

$$\sum_{i=1}^{f}\sum_{l=1}^{L}\sum_{u\in\mathcal{U}_c} p^h_{u,(i-1)L+l}T_s \leq \sum_{i=0}^{f-1}\left(H_{c,i}-\omega_{c,i}\right),\ \forall c\in\mathcal{M}\cup\mathcal{K},\ \forall f=1,\ldots,N_F. \tag{3.3}$$

The left-hand side of this inequality is the amount of the energy that was used by each cell c for transmission during the first f frames. It should not exceed the available energy that

was accumulated up until the beginning of the f-th frame.

The auxiliary variable $\omega_{c,i}$ introduced in (3.3) is the excess energy that must be discarded by cell c at the end of the harvesting frame i. This is because cell c has a limited battery capacity B_c and a maximum transmit power limit P_c. As a consequence, it is possible that cell c cannot store all the harvested power at the end of an energy-harvesting frame. This situation arises, for example, when the energy storage is already full at the beginning of a frame and the cell still harvests a large amount of energy that exceeds its transmit power limit P_c. Consequently, cell c has to discard the excess power $\omega_{c,i}$ at the end of each frame i to satisfy its battery capacity constraints:

$$\sum_{i=1}^{f}\left(H_{c,i}-\sum_{l=1}^{L}\sum_{u\in\mathcal{U}_c}p^{h}_{u,(i-1)L+l}T_s-\omega_{c,i}\right)\leq B_c,\ \forall c\in\mathcal{M}\cup\mathcal{K},\ \forall f=1,\ldots,N_F-1. \quad (3.4)$$

Given the above hybrid power source model, we employ the energy efficiency metric defined by:

$$\mathrm{EE}_c\left(\mathbf{p},\mathbf{p}^g_c\right)=\frac{R_c\left(\mathbf{p}\right)}{C_c\left(\mathbf{p}^g_c\right)},\ \forall c\in\mathcal{M}\cup\mathcal{K}.$$

It is calculated as the ratio of its sum rate $R_c\left(\mathbf{p}\right)=\sum\limits_{u\in\mathcal{U}_c}\sum\limits_{t=1}^{T}\log\left(1+\gamma_{u,t}\right)$ over its overall power consumption $C_c\left(\mathbf{p}^c_c\right)$. Specifically, C_c is an affine function defined by [58]:

$$C_c\left(\mathbf{p}^g_c\right)=P^f_c+\psi\sum_{t=1}^{T}\sum_{u\in\mathcal{U}_c}p^g_{u,t} \quad (3.5)$$

where P^f_c denotes the additional power dissipation due to BS's circuitry and cooling system and ψ is the RF power amplifier efficiency respectively. In this model, the cost function accounts only for the energy drawn from the conventional power grid and considers the renewable energy as free. In addition, we assume that the circuit-power consumption P^f_c of each BS is powered by the conventional power grid source.

3.4 Energy Allocation Problem

In the following, we formulate the offline power allocation problem for the hybrid-powered wireless two-tier systems. The first optimization variable is the vector of power allocation $\mathbf{p} = (\mathbf{p}_t)_{t=1,\dots,T} \in \mathbb{R}_+^{NT}$ for all users. In addition, we optimize the following three power usage vectors for the cells: (1) the transmit power $\mathbf{p}^g = (\mathbf{p}_t^g)_{t=1,\dots,T} \in \mathbb{R}_+^{NT}$ used from the power grid, (2) the power $\mathbf{p}^h = \left(\mathbf{p}_t^h\right)_{t=1,\dots,T} \in \mathbb{R}_+^{NT}$ used from the energy-harvesting source, and (3) the renewable power wastage vector $\mathbf{w} = (\omega_{c,f})_{f=1,\dots,N_F,\ k\in\mathcal{U}} \in \mathbb{R}_+^{(M+K)N_F}$ that the cells will not be able to use or store.

Our objective is to maximize the sum of the energy efficiencies of the two-tier cells while satisfying an average quality of service for each cell:

$$
(P)\ \begin{cases}
\underset{\mathbf{p},\mathbf{p}^g,\mathbf{p}^h,\mathbf{w}}{\text{maximize}} & \sum_{c\in\mathcal{M}\cup\mathcal{K}} \mathrm{EE}_c\left(\mathbf{p},\mathbf{p}^g\right) \\
\text{subject to} & \mathrm{c}_1:\ \frac{1}{T}\sum\limits_{t=1}^{T}\sum\limits_{u\in\mathcal{U}_c}\log\left(1+\gamma_{u,t}\right) \geq \rho_c,\ \forall c\in\mathcal{M}\cup\mathcal{K} \\
& \mathrm{c}_2:\ \sum\limits_{i=1}^{f}\sum\limits_{l=1}^{L}\sum_{u\in\mathcal{U}_c} p^h_{u,(i-1)L+l}T_s \leq \sum\limits_{i=0}^{f-1}\left(H_{c,i}-\omega_{c,i}\right),\ \forall c\in\mathcal{M}\cup\mathcal{K},\ \forall f=1,\dots,N_F, \\
& \mathrm{c}_3:\ \sum\limits_{i=1}^{f}\left(H_{c,i}-\sum\limits_{l=1}^{L}\sum_{u\in\mathcal{U}_c} p^h_{u,(i-1)L+l}T_s-\omega_{c,i}\right) \leq B_c,\ \forall c\in\mathcal{M}\cup\mathcal{K},\ \forall f=1,\dots,N_F-1. \\
& \mathrm{c}_4:\ \sum\limits_{u\in\mathcal{U}_c} p_{u,t} \leq P_c,\ \forall c\in\mathcal{M}\cup\mathcal{K},\ \forall t=1,\dots,T, \\
& \mathrm{c}_5:\ p^g_{u,t}+p^h_{u,t}=p_{u,t},\ \forall u\in\mathcal{U}_c,\ \forall c\in\mathcal{M}\cup\mathcal{K},\ \forall t=1,\dots,T, \\
& \mathrm{c}_6:\ \mathbf{p}\geq\mathbf{0},\ \mathbf{p}^g\geq\mathbf{0},\ \mathbf{p}^h\geq\mathbf{0},\ \mathbf{w}\geq\mathbf{0}.
\end{cases}
$$

The first set of constraints c_1 in (P) consists of the average rate constraints for the macrocell and small cells. Next, the causal energy constraint c_2 ensures that the total renewable power used by each BS for transmission up until the n-th frame will not exceed

the available amount that was harvested and stored in the previous $f-1$ frames. On the other hand, constraint c_3 means that the amount of renewable energy that each BS can store at the end of each energy-harvesting frame shall not exceed its battery capacity limit. Consequently, the excess energy must be discarded through the auxiliary variable $\mathbf{w}$. Then, the transmit power constraints for the BSs are given in c_4. Finally, we have the power usage equality constraints c_5 introduced in (3.2) and the non-negativity constraints c_6 of the power vectors.

This offline energy allocation problem described by (P) is hard to solve due to its non-convexity. However, we will show that it is still possible to leverage the tools of convex optimization to find an efficient solution. In particular, we will use sequential and parametric convex programming techniques to derive a convergent algorithm for problem (P).

3.4.1 Minorization-Maximization with Non-convex Rate Constraints

There are intricate difficulties in solving problem (P). First, the feasibility set implied by the rate constraints c_{1-2} is not convex. In fact, the rate functions are not concave due to the interference between the cells. Thereby, the objective function is not concave either because it is defined as a sum of ratios between the rate functions and the affine power costs.

To solve problem (P), we use an extended version of minorization-maximization procedure, which also minorize the non-convex constraints. Given these additional complexities,

we will design a new algorithm that relies on the following convex subproblem $(Q_{\mathbf{q}})$:

$$(Q_{\mathbf{q}}) \begin{cases} \underset{\mathbf{p},\mathbf{p}^g,\mathbf{p}^h,\mathbf{w}}{\text{maximize}} & \sum_{c\in\mathcal{M}\cup\mathcal{K}} \widetilde{\text{EE}}_c\left(\mathbf{p},\mathbf{p}^g;\mathbf{q}\right) \\ \text{subject to} & \text{c}_1:\ \widetilde{R}_c\left(\mathbf{p};\mathbf{q}\right) \geq \rho_c T,\ \forall c\in\mathcal{M}\cup\mathcal{K} \\ & \text{Constraints } \text{c}_{2-6}, \end{cases} \tag{3.6}$$

where the vector $\mathbf{q}\in\mathbb{R}_+^{NT}$ is not an optimization variable but a parameter that defines the objective function of $(Q_{\mathbf{q}})$.

In fact, this new problem $(Q_{\mathbf{q}})$ is obtained from (P) by replacing the non-concave objective function and the non-concave rate functions in c_{1-2} by their respective concave minorizing functions. These new functions are respectively denoted by $\widetilde{R}_c$ and $\widetilde{\text{EE}}_c$ for the rate and energy efficiency of the cell c. For completeness, let us first recall the definition of a minorization.

Before we can solve $(Q_{\mathbf{q}})$, we need to obtain the minorizations for the rate and energy efficiency functions of each cell. For that, let us first define the following function $\widetilde{r}_u$ for an arbitrary macro-cell or small-cell user $u\in\mathcal{U}$:

$$\widetilde{r}_u\left(\mathbf{p};\mathbf{q}\right) \triangleq \sum_{t=1}^{T} \log\left(\sigma_{u,t}^2 + \sum_{j\in\mathcal{U}} g_{uj,t}p_{j,t}\right) - l_i\left(\mathbf{p}_{-u},\mathbf{q}_{-u}\right) \tag{3.7}$$

where $\mathbf{p}_{-u}\in\mathbb{R}_+^{(N-1)T}$ is the concatenation of the power of all users except user u. The affine function l_u in (3.7) is given by

$$l_u\left(\mathbf{p}_{-u};\mathbf{q}_{-u}\right) = \sum_{t=1}^{T}\left(v_{u,t}\left(\mathbf{q}_{-u,t}\right) + \nabla v_{u,t}\left(\mathbf{q}_{-u,t}\right)^{\top}\left(\mathbf{p}_{-u,t}-\mathbf{q}_{-u,t}\right)\right),\ \forall u\in\mathcal{U} \tag{3.8}$$

where we have

$$v_{u,t}\left(\mathbf{p}_{-u,t}\right)=\log\left(\sigma_{u,t}^{2}+\sum_{j\in\mathcal{U}\setminus\{u\}}g_{uj,t}p_{j,t}\right),\ \forall u\in\mathcal{U},\ \forall t \tag{3.9}$$

and

$$\nabla v_{u,t}\left(\mathbf{q}_{-u,t}\right)=\left(\frac{g_{uj,t}}{\sigma_{u,t}^{2}+\sum\limits_{k\in\mathcal{U}\setminus\{u\}}g_{uk}p_{k,t}}\right)_{j\in\mathcal{U}\setminus\{u\}},\ \forall u\in\mathcal{U},\ \forall t. \tag{3.10}$$

Having defined the function $\widetilde{r}_u$ for all $u\in\mathcal{U}$, the following lemma gives the expressions for the minorizations of each cell's rate and energy efficiency functions.

Lemma 3.1. *The minorizations of the rate functions $\{EE_c\}_{c\in\mathcal{M}\cup\mathcal{K}}$ and of the energy efficiency functions $\{R_c\}_{c\in\mathcal{M}\cup\mathcal{K}}$ of the cells are respectively given by*

$$\widetilde{EE}_c\left(\mathbf{p},\mathbf{p}_c^g;\mathbf{q}\right)=\frac{\sum_{u\in\mathcal{U}_c}\widetilde{r}_u\left(\mathbf{p};\mathbf{q}\right)}{C_c\left(\mathbf{p}_c^g\right)},\ \forall c\in\mathcal{M}\cup\mathcal{K}, \tag{3.11}$$

$$\widetilde{R}_k\left(\mathbf{p};\mathbf{q}\right)=\sum_{u\in\mathcal{U}_c}\widetilde{r}_u\left(\mathbf{p};\mathbf{q}\right),\ \forall c\in\mathcal{M}\cup\mathcal{K}. \tag{3.12}$$

Proof. The proof is similar to that of Lemma 2.2 in Chapter 2. □

After we found the minorizations, we now derive an algorithm for solving the original problem (P) based on a similar procedure presented in Section 2.6. First we initialize the iteration index to $n=0$ and start with a feasible power allocation $\mathbf{p}^{(0)}$. By setting the parameter $\mathbf{q}$ to $\mathbf{p}^{(n)}$, we compute the minorizations of the energy efficiency and rate functions using equations (3.11)-(3.12). Then, we solve at each iteration n the surrogate problem $(Q_{\mathbf{q}})$ whose solution gives us a new power allocation $\mathbf{p}^{(n+1)}$ as well as the corresponding power usage vectors $\left\{\mathbf{p}^{g\ (n+1)},\mathbf{p}^{h\ (n+1)},\mathbf{w}^{(n+1)}\right\}$. The new power allocation $\mathbf{p}^{(n+1)}$ is then used to provide new minorizations that will be used for the next iteration. These steps are repeated until convergence. A reasonable stopping criterion is when the increase of the

Algorithm 3.1: Offline energy allocation algorithm for solving (P)

Initialize: ϵ, $\mathbf{p}^{(0)}, \mathbf{p}^{g\ (0)}, \mathbf{p}^{h\ (0)} \in \mathbb{R}^{NT}$ and $\mathbf{w}^{(0)} \in \mathbb{R}^{(M+K)N_F}$. Set $n = 0$.
Repeat
(S.1) : Set $\mathbf{q} \leftarrow \mathbf{p}^{(n)}$.
(S.2) : Determine the minorizations $\widetilde{R}_c$, and $\widetilde{EE}_c,\ \forall c \in \mathcal{M} \cup \mathcal{K}$ using (3.11)-(3.12).
(S.3) : Update the power allocation $\mathbf{p}^{(n+1)}$ by solving problem $(Q_{\mathbf{q}})$ in (3.6).
(S.4) : $n \leftarrow n + 1$
Until (3.13) *is satisfied*

sum energy efficiency is less than some threshold ϵ:

$$\sum_{c \in \mathcal{M} \cup \mathcal{K}} \mathrm{EE}_c^{(n+1)} - \sum_{c \in \mathcal{M} \cup \mathcal{K}} \mathrm{EE}_c^{(n)} \leq \epsilon. \tag{3.13}$$

This algorithm is summarized in Algorithm 3.1 and its convergence property is summarized below.

Proposition 3.1. *Suppose we start Algorithm 3.1 with feasible power allocation* $\mathbf{p}^{(0)}$ *and power usage vectors* $\left\{\mathbf{p}^{g\ (0)}, \mathbf{p}^{h\ (0)}, \mathbf{w}^{(0)}\right\}$. *Then, all next iterates* $\mathbf{p}^{(n)}$ *and* $\left\{\mathbf{p}^{g\ (n)}, \mathbf{p}^{h\ (n)}, \mathbf{w}^{(n)}\right\}$ *produced in step (S.3) are also feasible* $\forall n \geq 1$. *In addition, the algorithm monotonically converges to at least a local optimum of the sum energy efficiency maximization problem* (P).

Proof. See Appendix B. □

It is therefore important to start with initial power allocation and power usage vectors that satisfy all the constraints of (P). For this purpose, we can find a feasible point by

solving first the following feasibility problem:

$$(F)\begin{cases} \underset{\mathbf{p},\mathbf{p}^g,\mathbf{p}^h,\mathbf{w},\mathbf{z}}{\text{maximize}} & \sum_{c\in\mathcal{M}\cup\mathcal{K}} z_c \\ \text{subject to} & \mathrm{r}_1:\ \frac{1}{T}\sum_{t=1}^{T}\sum_{u\in\mathcal{U}_c}\log\left(1+\gamma_{u,t}\right)+z_c\geq\rho_c,\ \forall c\in\mathcal{M}\cup\mathcal{K} \\ & \text{Constraints } \mathrm{c}_{2-6}, \\ & \mathbf{z}\geq\mathbf{0}. \end{cases}$$

For the feasibility problem (F), we introduced new positive variable $\mathbf{z}\triangleq(z_c)_{c\in\mathcal{M}\cup\mathcal{K}}\in\mathbb{R}_+^{M+K}$. Each slack variable z_c corresponds to the violation of the average rate constraint of cell c. Furthermore, we replaced the average constraints of (P) by the relaxed constraints r_1. In contrast to (P), the feasibility problem (F) always admits a solution. When a solution of (F) satisfies $\sum_{c\in\mathcal{M}\cup\mathcal{K}} z_c=0$, then it must be a feasible solution for (P). Although (F) is also not convex, we can again apply the extended minorization-majorization procedure to find a feasible solution for (P). This procedure is listed in Algorithm 3.2.

The proposed Algorithm 3.1 for maximizing the sum energy efficiency is therefore an ascent algorithm. In particular, it offers the benefits of producing iterates that are always feasible. In addition, it is monotonically convergent. However, the step (S.2) in Algorithm 3.1 depends on the solution of the surrogate problem $(Q_{\mathbf{q}})$. Although problem $(Q_{\mathbf{q}})$ is less complex than (P), $(Q_{\mathbf{q}})$ is still not convex due to its fractional objective function. Therefore, we will next derive another algorithm that finds an optimal solution for problem $(Q_{\mathbf{q}})$.

3.4.2 Parametric Convex Optimization

In this subsection, we apply the convex parametric approach introduced in Section 2.5 to solve the surrogate problem $(Q_{\mathbf{q}})$. By putting $(Q_{\mathbf{q}})$ in its epigraph form, we obtain the

Algorithm 3.2: Solving the feasibility problem (F)

Initialize: ϵ, $\mathbf{p}^{(0)}, \mathbf{p}^{g\ (0)}, \mathbf{p}^{h\ (0)} \in \mathbb{R}^{NT}$ and $\mathbf{w}^{(0)} \in \mathbb{R}^{(M+K)N_F}$. Set $n = 0$.

Repeat

(S.1) : Set $\mathbf{q} \leftarrow \mathbf{p}^{(n)}$.

(S.2) : Determine the minorization $\widetilde{R}_c,\ \forall c \in \mathcal{M} \cup \mathcal{K}$ using (3.12).

(S.3) : Solve the convex optimization problem:

$$\begin{aligned} \underset{\mathbf{p},\mathbf{p}^g,\mathbf{p}^h,\mathbf{w},\mathbf{z}}{\text{maximize}} \quad & \textstyle\sum_{c\in\mathcal{M}\cup\mathcal{K}} z_c \\ \text{subject to} \quad & \widetilde{R}_c\left(\mathbf{p};\mathbf{q}\right) + z_c \geq \rho_c T,\ \forall c \in \mathcal{M} \cup \mathcal{K} \\ & \text{Constraints } \mathrm{c}_{2-6}\text{,} \\ & \mathbf{z} \geq \mathbf{0}. \end{aligned}$$

(S.4) : $n \leftarrow n + 1$

Until $\sum_{c\in\mathcal{M}\cup\mathcal{K}} z_c^{(n)} \leq \epsilon$

Output: Feasible power allocation and usage vectors $\left\{\mathbf{p}, \mathbf{p}^g, \mathbf{p}^h, \mathbf{w}\right\}$.

equivalent reformulation $(E_{\mathbf{q}})$:

$$(E_{\mathbf{q}}) \begin{cases} \underset{\mathbf{p},\mathbf{p}^g,\mathbf{p}^h,\mathbf{w},\theta}{\text{maximize}} \quad \textstyle\sum_{c\in\mathcal{M}\cup\mathcal{K}} \theta_c \\ \text{subject to} \quad \mathrm{c}_0 :\ \widetilde{R}_c\left(\mathbf{p};\mathbf{q}\right) - \theta_c C_c\left(\mathbf{p}_c^g\right) \geq 0,\ \forall c \in \mathcal{M} \cup \mathcal{K}, \\ \qquad\qquad \left(\mathbf{p}, \mathbf{p}^g, \mathbf{p}^h, \mathbf{w}\right) \in \mathcal{S}, \end{cases} \tag{3.14}$$

where the auxiliary variable $\theta \triangleq (\theta_1, \ldots, \theta_{M+K})$ and the new constraints c_0 are introduced to replace the fractional objective problem by a linear term. For convenience, we also replaced the convex constraints c_{1-6} of $(Q_{\mathbf{q}})$ into the convex feasibility set $\mathcal{S}$.

With these manipulations, we have maintained the equivalence between $(E_{\mathbf{q}})$ and $(Q_{\mathbf{q}})$. Nonetheless, the two formulations are different. While $(Q_{\mathbf{q}})$ has a non-convex objective function and convex constraints, the equivalent problem $(E_{\mathbf{q}})$ has a linear objective function, a new variable θ and new non-convex constraints c_0. Next, we will employ the convex parametric optimization approach to construct an algorithm that finds an optimal solution

for $(E_{\mathbf{q}})$, and hence for $(Q_{\mathbf{q}})$.

Let us consider the following family of parametric convex problems $(P_{\mathbf{x}})$:

$$(P_{\mathbf{x}}) \begin{cases} \underset{\mathbf{p},\mathbf{p}^g,\mathbf{p}^h,\mathbf{w}}{\text{maximize}} & \sum_{c\in\mathcal{M}\cup\mathcal{K}} \alpha_c \left(\widetilde{R}_c\left(\mathbf{p};\mathbf{q}\right) - \theta_c C_c\left(\mathbf{p}_c^g\right)\right) \\ \text{subject to} & \left(\mathbf{p},\mathbf{p}^g,\mathbf{p}^h,\mathbf{w}\right) \in \mathcal{S} \end{cases} \tag{3.15}$$

which is parameterized by the vector of parameters $\mathbf{x} \triangleq (\alpha,\theta)$ with $\alpha \triangleq (\alpha_1,\ldots,\alpha_{M+K})$. Observe that the vector θ is not an optimization in $(P_{\mathbf{x}})$ but a parameter. This parametric problem can be interpreted as a weighted sum utility maximization in which each cell's utility is the difference of its sum rate and a pricing penalty term. More importantly, $(P_{\mathbf{x}})$ is a convex problem and can be used as a vehicle to solve $(E_{\mathbf{q}})$ due to the following result.

By applying the result in Proposition 2.1, if $(E_{\mathbf{q}})$ is feasible and admits an optimal solution $\left\{\widehat{\mathbf{p}},\widehat{\mathbf{p}}^g,\widehat{\mathbf{p}}^h,\widehat{\mathbf{w}},\widehat{\theta}\right\}$. Then, there exists $\widehat{\alpha}$ such that $\left\{\widehat{\mathbf{p}},\widehat{\mathbf{p}}^g,\widehat{\mathbf{p}}^h,\widehat{\mathbf{w}}\right\}$ is a solution of $(P_{\mathbf{x}})$ with $\mathbf{x} = \left(\widehat{\alpha},\widehat{\theta}\right)$. Furthermore, the solution $\left\{\widehat{\mathbf{p}},\widehat{\mathbf{p}}^g,\widehat{\mathbf{p}}^h,\widehat{\mathbf{w}}\right\}$ must satisfy the following system of nonlinear equations:

$$\mathbf{F}\left(\mathbf{x};\mathbf{q}\right) = \mathbf{0}, \tag{3.16}$$

where the vector function $\mathbf{F}: \mathbb{R}^{2(K+M)} \to \mathbb{R}^{2(K+M)}$ is defined by its component functions as follows:

$$F_c\left(\mathbf{x};\mathbf{q}\right) = \theta_c C_c\left(\widetilde{\mathbf{p}}_c^g\left(\mathbf{x}\right)\right) - \widetilde{R}_c\left(\widetilde{\mathbf{p}}\left(\mathbf{x}\right);\mathbf{q}\right), \quad \forall c \in \mathcal{M}\cup\mathcal{K}, \tag{3.17}$$

$$F_{c+M+K}\left(\mathbf{x}\right) = \alpha_c C_c\left(\widetilde{\mathbf{p}}_c^g\left(\mathbf{x}\right)\right) - 1, \quad \forall c \in \mathcal{M}\cup\mathcal{K}. \tag{3.18}$$

These component functions are obtained by composing the minorizations $\left\{\widetilde{R}_c\right\}_{c\in\mathcal{M}\cup\mathcal{K}}$ or the power costs $\{C_c\}_{c\in\mathcal{M}\cup\mathcal{K}}$ with the functions $\widetilde{\mathbf{p}}\left(\mathbf{x}\right)$ and $\{\widetilde{\mathbf{p}}_c^g\left(\mathbf{x}\right)\}_{c\in\mathcal{M}\cup\mathcal{K}}$ that return an optimal power allocation of the parametric problem $(P_{\mathbf{x}})$ for a given parameter $\mathbf{x}$.

Furthermore, the uniqueness of the root of the nonlinear system $\mathbf{F}\left(\mathbf{x};\mathbf{q}\right) = \mathbf{0}$ was established in Proposition 2.2. These previous results lead us to a simpler solution method

for the non-convex problem $(Q_{\mathbf{q}})$. In fact, whenever $(Q_{\mathbf{q}})$ is feasible, an optimal solution $\left\{\widehat{\mathbf{p}}, \widehat{\mathbf{p}}^{g}, \widehat{\mathbf{p}}^{h}, \widehat{\mathbf{w}}\right\}$ can be obtained by first solving the system $\mathbf{F}(\mathbf{x}; \mathbf{q}) = \mathbf{0}$ and then by solving the parametric convex problem $(P_{\mathbf{x}})$ with $\mathbf{x}$ set to the unique root of the system. In other words, solving the non-convex problem $(Q_{\mathbf{q}})$ reduces to solving a nonlinear system of equations. In the next subsection, we present an iterative algorithm based on Newton method which enables us to find the root of $\mathbf{F}(\mathbf{x}; \mathbf{q}) = \mathbf{0}$ and solve $(Q_{\mathbf{q}})$.

3.4.3 Damped Newton Algorithm for Solving $(Q_{\mathbf{q}})$

Our proposed algorithm for finding the optimal parameter $\mathbf{x}$ and solving $(Q_{\mathbf{q}})$ is based on a damped Newton method [63]. First, we start with the initial power allocation vector $\mathbf{p}^{(0)}$ of all users and the initial power usage vectors $\left\{\mathbf{p}^{g\ (0)}, \mathbf{p}^{h\ (0)}, \mathbf{w}^{(0)}\right\}$ of all cells. Then, we initialize the parameter $\mathbf{x}^{(0)} = \left(\alpha^{(0)}, \theta^{(0)}\right)$ as follows:

$$\alpha_c^{(0)} = \frac{1}{C_c\left(\widetilde{\mathbf{p}}_c^{g\ (0)}\right)}, \forall c \in \mathcal{M} \cup \mathcal{K}, \tag{3.19}$$

$$\theta_c^{(0)} = \frac{\widetilde{R}_c\left(\mathbf{p}^{(0)}\right)}{C_c\left(\mathbf{p}_c^{g\ (0)}\right)},\ \forall c \in \mathcal{M} \cup \mathcal{K}. \tag{3.20}$$

At each iteration $n \geq 0$ of the algorithm, we solve the convex parametric program $(P_{\mathbf{x}})$ that is parameterized by $\mathbf{x}(n)$. This gives us a new power allocation vector $\widetilde{\mathbf{p}}\left(\mathbf{x}^{(n)}\right)$ and power usage vectors $\left\{\widetilde{\mathbf{p}}^{g}\left(\mathbf{x}^{(n)}\right), \widetilde{\mathbf{p}}^{h}\left(\mathbf{x}^{(n)}\right), \widetilde{\mathbf{w}}\left(\mathbf{x}^{(n)}\right)\right\}$. After that, we check if the stopping criterion $\left\|\mathbf{F}\left(\mathbf{x}^{(n)}\right)\right\| < \epsilon$ is satisfied for a given tolerance $\epsilon > 0$. If it is satisfied, then we stop the algorithm. Otherwise, we update the parameter $\mathbf{x}^{(n)}$ using a damped Newton method. To do so, we first compute the Newton step at the n-th iteration as follows:

$$\mathbf{d}^{(n)} = -\mathbf{F}'\left(\mathbf{x}^{(n)}\right)^{-1}\mathbf{F}\left(\mathbf{x}^{(n)}\right), \tag{3.21}$$

where $\mathbf{F}'$ is the Jacobian matrix of the vector-valued function $\mathbf{F}$. In fact, it can be shown that $\mathbf{F}$ is differentiable and that its Jacobian $\mathbf{F}'$ is a non-singular, diagonal and positive

definite matrix given by [81]:

$$\mathbf{F}'(\mathbf{x}) = \begin{bmatrix} \operatorname{diag}(\mathbf{c}) & \mathbf{0}_{(M+K)\times(M+K)} \\ \mathbf{0}_{(M+K)\times(M+K)} & \operatorname{diag}(\mathbf{c}) \end{bmatrix}. \tag{3.22}$$

where the vector $\mathbf{c}$ concatenates the power costs of all cells as follows:

$$\mathbf{c} = \left(C_c\left(\widetilde{\mathbf{p}}^g\left(\mathbf{x}^{(n)}\right)\right)\right)_{c\in\mathcal{M}\cup\mathcal{K}}^{\top} \in \mathbb{R}_+^{M+K}.$$

Since the matrix $\mathbf{F}'(\mathbf{x})$ is diagonal and positive definite, its inverse can be easily calculated. Once we get the Newton step from (3.21), we perform a line search to find an effective step size $0 < \lambda < 1$. Precisely, we start with the full Newton step, i.e. $\lambda = 1$ and reduce it by a constant factor $0 < \sigma < 1$ until a sufficient decrease in the l_2-norm of $\mathbf{F}\left(\mathbf{x}^{(n)} + \lambda\mathbf{d}^{(n)}\right)$ is obtained. After that, we update the parameter vector as:

$$\mathbf{x}^{(n+1)} \triangleq \mathbf{x}^{(n)} + \lambda\mathbf{d}^{(n)}. \tag{3.23}$$

or component-wise as follows:

$$\alpha_c^{(n+1)} = (1-\lambda)\,\alpha_c^{(n)} + \lambda\frac{1}{C_c\left(\widetilde{\mathbf{p}}^g\left(\mathbf{x}^{(n)}\right)\right)}, \tag{3.24}$$

$$\theta_c^{(n+1)} = (1-\lambda)\,\theta_M^{(n)} + \lambda\frac{\widetilde{R}_c\left(\widetilde{\mathbf{p}}\left(\mathbf{x}^{(n)}\right);\mathbf{q}\right)}{C_c\left(\widetilde{\mathbf{p}}^g\left(\mathbf{x}^{(n)}\right)\right)}, \tag{3.25}$$

Then, we use this new parameter $\mathbf{x}^{(n+1)}$ for the next iteration and we repeat this iterative procedure until convergence. This proposed iterative damped Newton algorithm is summarized in Algorithm 3.3. Its convergence is stated in the following proposition.

According to Proposition 2.3, Algorithm 3.3 converges to an optimal solution of the convex subproblem $(Q_{\mathbf{q}})$ in (3.6).

Algorithm 3.3: Subprocedure for solving the surrogate problem $(Q_{\mathbf{q}})$.

Input: Parameter $\mathbf{q}$ and constants $\epsilon,\ \mathbf{0} < (\mu, \lambda, \sigma) < \mathbf{1}$.
Initialize: $\mathbf{p}^{(0)} \in \mathbb{R}^{NT}$ and set $n = 0$
Repeat
(S.1) : Solve problem $(P_{\mathbf{x}})$ to obtain $\widetilde{\mathbf{p}}\left(\mathbf{x}^{(n)}\right)$, $\widetilde{\mathbf{p}}^{g}\left(\mathbf{x}^{(n)}\right)$, $\widetilde{\mathbf{p}}^{h}\left(\mathbf{x}^{(n)}\right)$ and $\widetilde{\mathbf{w}}\left(\mathbf{x}^{(n)}\right)$.
(S.2) : Initialize $\lambda \leftarrow 1$,
(S.3) : Compute $\mathbf{x}^{(n)} + \lambda \mathbf{d}^{(n)}$ using (3.24)-(3.25).,
(S.4) : **while** $\left\|\mathbf{F}\left(\mathbf{x}^{(n)} + \lambda \mathbf{d}^{(n)}\right)\right\| > (1 - \mu\lambda)\left\|\mathbf{F}\left(\mathbf{x}^{(n)}\right)\right\|$ **do** $\lambda \leftarrow \sigma\lambda$
(S.5) : Update $\mathbf{x}^{(n+1)} \leftarrow \mathbf{x}^{(n)} + \lambda \mathbf{d}^{(n)}$
(S.6) : $n \leftarrow n + 1$
Until $\left\|\mathbf{F}\left(\mathbf{x}^{(n)}\right)\right\| \leq \epsilon$
Result: Root $\mathbf{x}^{(n)}$ of (3.16).

3.5 Joint Energy Allocation and Energy Cooperation

In this section, we propose an energy allocation and cooperation scheme for the wireless two-tier network. Again, we assume that the cells are powered by hybrid power sources. In addition, the cells can exchange their harvested energy through a smart-grid infrastructure. In fact, recent advancements in smart-grid will integrate the distributed generation of renewable energy into conventional power grid. This will be achieved by enabling a two-way energy flow between distributed energy micro-grid generators and the smart-grid [41]. In this work, we explore this future possibility to efficiently power heterogeneous networks. Particularly, if a cell harvests more renewable energy than it can store at a specific frame, then it can inject a portion of that excess energy to the smart-grid instead of wasting it. On the other hand, a cell that could not harvest or store enough energy can benefit by drawing some of the transferred renewable energy. As illustrated in Figure 3.2, we assume that there exists a micro-grid aggregator that controls the energy sharing between the base stations. Thus, the aggregator helps realize the two-way energy flow between the cells using real-time information about their energy usage. Here, we assume that the energy transfer happens on a per-frame basis. At the end of each frame, if there is still an excess energy that cannot be stored by all cells, that energy will be diverted by the aggregator into the smart-grid for other uses.

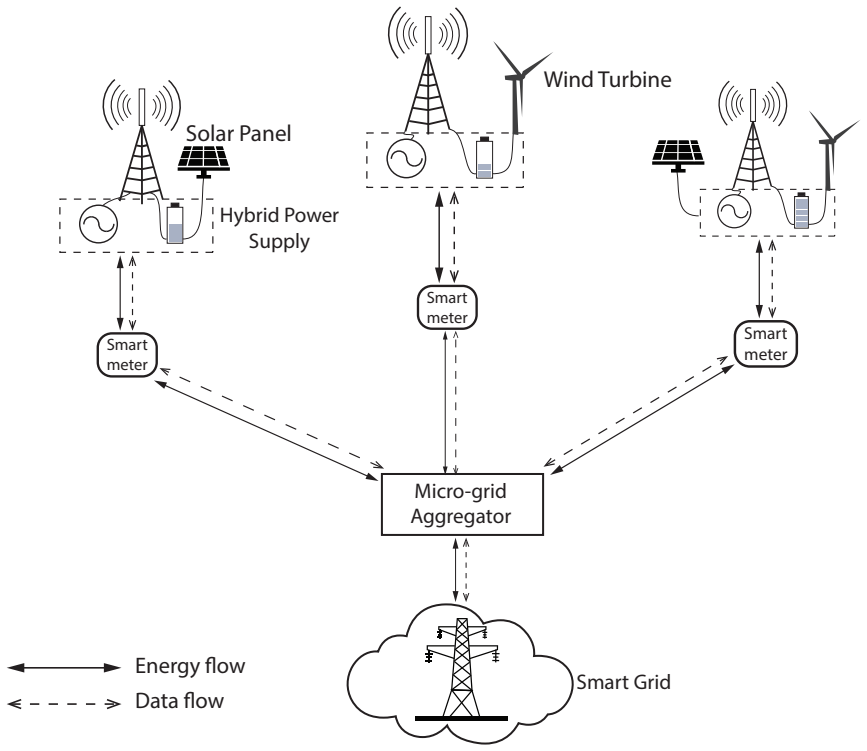


Figure 3.2 – A model of cellular networks with hybrid power source and energy sharing through a smart micro-grid.

3.5.1 Energy Cooperation between Two-tier Cells

To model the energy cooperation, we introduce the following array of energy exchange variables $\overrightarrow{\mathbf{H}} = \left(\overrightarrow{h}_{c,d,i}\right)_{c,d\in\mathcal{M}\cup\mathcal{K},i\in 1,\ldots,N_F} \in \mathbb{R}_{+}^{(M+K)\times(M+K)\times N_F}$. Each element $\overrightarrow{h}_{c,d,i} \geq 0$ denotes the amount of harvested renewable energy that a cell c transfers to another cell d at the end of frame i. When cell c transfers an energy $\overrightarrow{h}_{c,d,i}$ through the smart-grid power network, we assume that there is an energy loss that is incurred by the transfer. As a result, the effective energy that cell d can receive is $\eta\overrightarrow{h}_{c,d,i}$ with $0 < \eta < 1$ being the transfer efficiency factor.

Next, we define by the auxiliary variable $\delta_{c,i}$ as the net energy injected by a cell c into the aggregator at the end of frame i as:

$$\delta_{c,i} = \sum_{d\in\mathcal{M}\cup\mathcal{K}\setminus\{c\}} \left(\overrightarrow{h}_{c,d,i} - \eta\overrightarrow{h}_{d,c,i}\right),\ \forall c \in \mathcal{M}\cup\mathcal{K},\ \forall i = 0,...,N_F-1. \tag{3.26}$$

If $\delta_{c,i}$ is positive, then the cell c is effectively injecting some renewable energy into the smart-grid. Conversely, it is drawing renewable energy from the smart-grid when $\delta_{c,i}$ is negative. Note that our model does not explicitly preclude a cell from drawing and injecting some renewable energy from/into the grid at the same time. Since our focus here is on offline optimization, we need not explicitly add this constraint in our problem formulation.

After defining these auxiliary variables, we modify the energy causality constraints for the energy-cooperative two-tier cells as follows:

$$\sum_{i=1}^{f}\sum_{l=1}^{L}\sum_{u\in\mathcal{U}_c} p^h_{u,(i-1)L+l}T_s \leq \sum_{i=0}^{f-1}\left(H_{c,i}-\omega_{c,i}-\delta_{c,i}\right),\ \forall c\in\mathcal{M}\cup\mathcal{K},\ \forall f=1,\ldots,N_F. \quad (3.27)$$

Similarly, the battery constraints at the cells become:

$$\sum_{i=1}^{f}\left(H_{c,i}-\sum_{l=1}^{L}\sum_{u\in\mathcal{U}_c} p^h_{u,(i-1)L+l}T_s-\omega_{c,i}-\delta_{c,i}\right)\leq B_c,\ \forall c\in\mathcal{M}\cup\mathcal{K},\ \forall f=1,\ldots,N_F-1 \quad (3.28)$$

For the energy cooperation model, we use the following network energy efficiency metric:

$$\mathrm{EE}_0\left(\mathbf{p},\mathbf{p}^g\right)\triangleq\frac{\sum_{c\in\mathcal{M}\cup\mathcal{K}}R_c\left(\mathbf{p}\right)}{C_0\left(\mathbf{p}^g\right)}$$

where the total network power consumption C_0 is equal to:

$$C_0\left(\mathbf{p}^g\right)=\sum_{c\in\mathcal{M}\cup\mathcal{K}}\left[P_c^f+\psi\sum_{t=1}^{T}\sum_{u\in\mathcal{U}_c}p^g_{u,t}\right]. \quad (3.29)$$

Because of the energy cooperation between the cells, we use this network energy efficiency metric as a new objective function. By cooperating, the cells can exchange their harvested renewable energy and share their battery storage.

As a result, the problem formulation for the energy allocation and cooperation is given below:

$$
(C)\begin{cases}
\text{variables} & \mathbf{p}, \mathbf{p}^g, \mathbf{p}^h, \mathbf{w}, \overrightarrow{\mathbf{H}}, \delta \\
\text{maximize} & \mathrm{EE}_0\left(\mathbf{p}, \mathbf{p}^g\right) \\
\text{subject to} & \mathrm{d}_1: \ \frac{1}{T}\sum\limits_{t=1}^{T}\sum\limits_{u\in\mathcal{U}_c}\log\left(1+\gamma_{u,t}\right) \geq \rho_c, \ \forall c \in \mathcal{M}\cup\mathcal{K}, \\
& \mathrm{d}_2: \ \sum\limits_{i=1}^{f}\sum\limits_{l=1}^{L}\sum_{u\in\mathcal{U}_c} p^h_{u,(i-1)L+l}T_s \leq \sum\limits_{i=0}^{f-1}\left(H_{c,i}-\omega_{c,i}-\delta_{c,i}\right), \ \forall c \in \mathcal{M}\cup\mathcal{K}, \ \forall f = 1,\ldots,N_F, \\
& \mathrm{d}_3: \ \sum\limits_{i=1}^{f}\left(H_{c,i}-\sum\limits_{l=1}^{L}\sum_{u\in\mathcal{U}_c} p^h_{u,(i-1)L+l}T_s-\omega_{c,i}-\delta_{c,i}\right) \leq B_c, \ \forall c \in \mathcal{M}\cup\mathcal{K}, \ \forall f = 1,\ldots,N_F-1, \\
& \mathrm{d}_4: \ \sum\limits_{u\in\mathcal{U}_c} p_{u,t} \leq P_c, \ \forall c \in \mathcal{M}\cup\mathcal{K}, \ \forall t = 1,\ldots,T, \\
& \mathrm{d}_5: \ p^g_{u,t}+p^h_{u,t} = p_{u,t}, \ \forall u \in \mathcal{U}_c, \ \forall c \in \mathcal{M}\cup\mathcal{K}, \ \forall t = 1,\ldots,T, \\
& \mathrm{d}_6: \ \sum_{d\in\mathcal{M}\cup\mathcal{K}\setminus\{c\}}\left(\overrightarrow{h}_{c,d,i}-\eta\,\overrightarrow{h}_{d,c,i}\right) = \delta_{c,i}, \ \forall c \in \mathcal{M}\cup\mathcal{K}, \ \forall i = 0,...,N_F-1, \\
& \mathrm{d}_7: \ \mathbf{p}\geq\mathbf{0}, \ \mathbf{p}^g\geq\mathbf{0}, \ \mathbf{p}^h\geq\mathbf{0}, \ \mathbf{w}\geq\mathbf{0}, \ \overrightarrow{\mathbf{H}}\geq\mathbf{0}.
\end{cases}
$$

In next subsection, we present an algorithm for solving this new problem.

3.5.2 Offline Optimization Algorithm

Similar to (P), the problem (C) above is also non-convex. However, it is more tractable as its objective function consists of one fractional term. Following the approach in the previous section, we handle the non-convexities in (C) using the extended minorization-majorization procedure. After that, we use the simple Dinkelbach method [9] to solve the minorized subproblem.

First, let us write the expression for the minorization $\widetilde{\mathrm{EE}}_0$ of the objective function of (C) as:

$$\widetilde{\mathrm{EE}}_0\left(\mathbf{p},\mathbf{p}^g;\mathbf{q}\right)=\frac{\sum_{c\in\mathcal{M}\cup\mathcal{K}}\widetilde{R}_c\left(\mathbf{p}\right)}{C_0\left(\mathbf{p}^g\right)}. \tag{3.30}$$

Next, we consider the following minorized subproblem associated to (C):

$$(M_{\mathbf{q}})\begin{cases}\text{variables} & \mathbf{p},\mathbf{p}^g,\mathbf{p}^h,\mathbf{w},\overrightarrow{\mathbf{H}},\delta,\theta\\ \text{maximize} & \theta\\ \text{subject to} & \sum_{c\in\mathcal{M}\cup\mathcal{K}}\widetilde{R}_c\left(\mathbf{p};\mathbf{q}\right)-\theta C_0\left(\mathbf{p}^g\right)\geq 0,\\ & \widetilde{R}_c\left(\mathbf{p};\mathbf{q}\right)\geq\rho_c T,\ \forall c\in\mathcal{M}\cup\mathcal{K},\\ & \text{Constraints } \mathrm{d}_{2-7},\end{cases} \tag{3.31}$$

which is obtained by first minorizing the rate functions in (C) at $\mathbf{q}$, and then by rewriting the resulting problem in its epigraph form. This problem $(M_{\mathbf{q}})$ is analogous to problem $(Q_{\mathbf{q}})$ in (3.6). To solve $(M_{\mathbf{q}})$, we use the following convex parametric subproblem, which is parameterized by both θ and $\mathbf{q}$:

$$(D_{\theta,\mathbf{q}})\begin{cases}\text{variables} & \mathbf{p},\mathbf{p}^g,\mathbf{p}^h,\mathbf{w},\overrightarrow{\mathbf{H}},\delta\\ \text{maximize} & \sum_{c\in\mathcal{M}\cup\mathcal{K}}\widetilde{R}_c\left(\mathbf{p};\mathbf{q}\right)-\theta C_0\left(\mathbf{p}^g\right)\\ \text{subject to} & \widetilde{R}_c\left(\mathbf{p};\mathbf{q}\right)\geq\rho_c T,\ \forall c\in\mathcal{M}\cup\mathcal{K},\\ & \text{Constraints } \mathrm{d}_{2-7}.\end{cases} \tag{3.32}$$

After formulating the subproblem $(D_{\theta,\mathbf{q}})$, we now propose the Algorithm 3.4 for solving

Algorithm 3.4: Algorithm for solving the energy management and cooperation problem (C)

Input: Constants ϵ and δ

Initialize: $\mathbf{p}^{(0)}, \mathbf{p}^{g\ (0)}, \mathbf{p}^{h\ (0)} \in \mathbb{R}^{NT}$ and $\mathbf{w}^{(0)} \in \mathbb{R}^{(M+K)N_F}$.
$\overrightarrow{\mathbf{H}}^{(0)} \in \mathbb{R}_{+}^{(M+K)\times(M+K)\times N_F}$ with feasible values for (C). Set $n = 0$, $k = 0$.

Repeat

(S.1) : Set the parameter $\mathbf{q}^{(n)} \leftarrow \mathbf{p}^{(k)}$.

(S.2) : Compute the minorizations $\left\{\widetilde{R}_c\right\}_{c\in\mathcal{M}\cup\mathcal{K}}$ and $\widetilde{EE}_0$ at $\mathbf{q}^{(n)}$ using (3.12) and (3.30).

(S.3) : Initialize $\theta^{(k)} = \sum_{c\in\mathcal{M}\cup\mathcal{K}} \widetilde{R}_c\left(\mathbf{p}^{(k)}; \mathbf{q}^{(n)}\right) / C_0\left(\mathbf{p}^{g\ (k)}\right)$.

(S.4) : **while** $\left\|\sum_{c\in\mathcal{M}\cup\mathcal{K}} \widetilde{R}_c\left(\mathbf{p}^{(k)}; \mathbf{q}^{(n)}\right) - \theta^{(k)} C_0\left(\mathbf{p}^{g\ (k)}\right)\right\| > \epsilon$ **do**

Solve the convex problem $\left(D_{\lambda^{(k)},\mathbf{q}^{(n)}}\right)$ to obtain $\left\{\widehat{\mathbf{p}}, \widehat{\mathbf{p}}^g, \widehat{\mathbf{w}}, \widehat{\overrightarrow{\mathbf{H}}}\right\}$,

Update $\lambda^{(k+1)} \leftarrow \lambda^{(k)} - \frac{\sum_{c\in\mathcal{M}\cup\mathcal{K}} \widetilde{R}_c\left(\widehat{\mathbf{p}}; \mathbf{q}^{(n)}\right)}{C_0(\widehat{\mathbf{p}}^c)}$,

Update $\mathbf{p}^{(k+1)} \leftarrow \widehat{\mathbf{p}}$,

$k \leftarrow k + 1$

(S.5) : $n \leftarrow n + 1$

Until $\frac{EE_0^{(n+1)} - EE_0^{(n)}}{EE_0^{(n)}} \leq \delta$

Output: Power allocation $\mathbf{p}$ and power usage vectors $\left\{\mathbf{p}, \mathbf{p}^g, \mathbf{p}^h, \mathbf{w}, \overrightarrow{\mathbf{H}}\right\}$.

(C), which is listed in the next page. It consists of two loops. In each iteration k of the inner loop, we solve a convex subproblem $\left(D_{\lambda^{(k)},\mathbf{q}^{(n)}}\right)$ that is parameterized by both $\lambda^{(k)}$ and $\mathbf{q}^{(n)}$. At the end of that iteration, we update the parameter $\lambda^{(k)}$ using a full Newton step. In fact, a linear search is not required in the Dinkelbach method to guarantee theoretical convergence to an optimal solution of $(M_{\mathbf{q}})$. We exit the inner loop when the value of the objective function of $\left(D_{\lambda^{(k)},\mathbf{q}^{(n)}}\right)$ becomes zero. After that, we update the parameter $\mathbf{q}^{(n)}$ at each iteration n of the outer loop and compute new minorizations of the rate and objective functions using the new parameter. Then, the process is repeated until the increase in the network energy efficiency is smaller than a specified tolerance δ. Similar to Algorithm 3.1, this new algorithm enjoys the same convergence properties as those stated in Proposition 2.4.

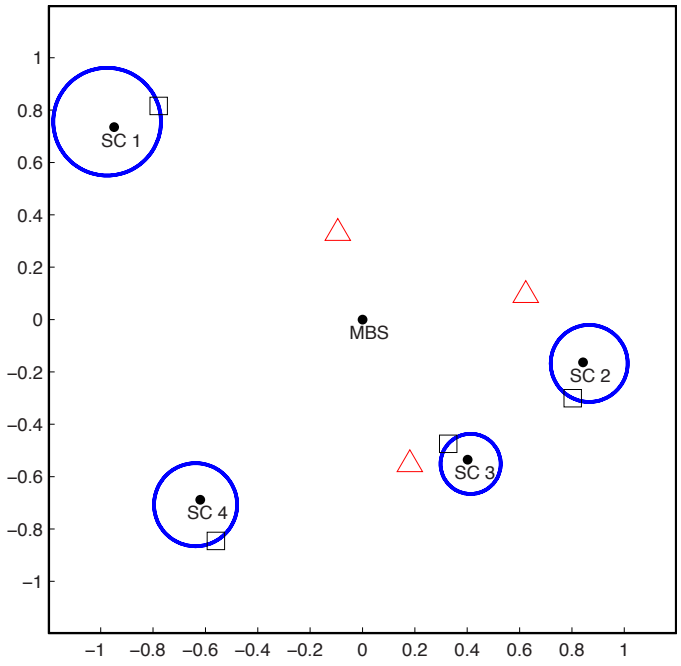


Figure 3.3 – Base stations and user locations.

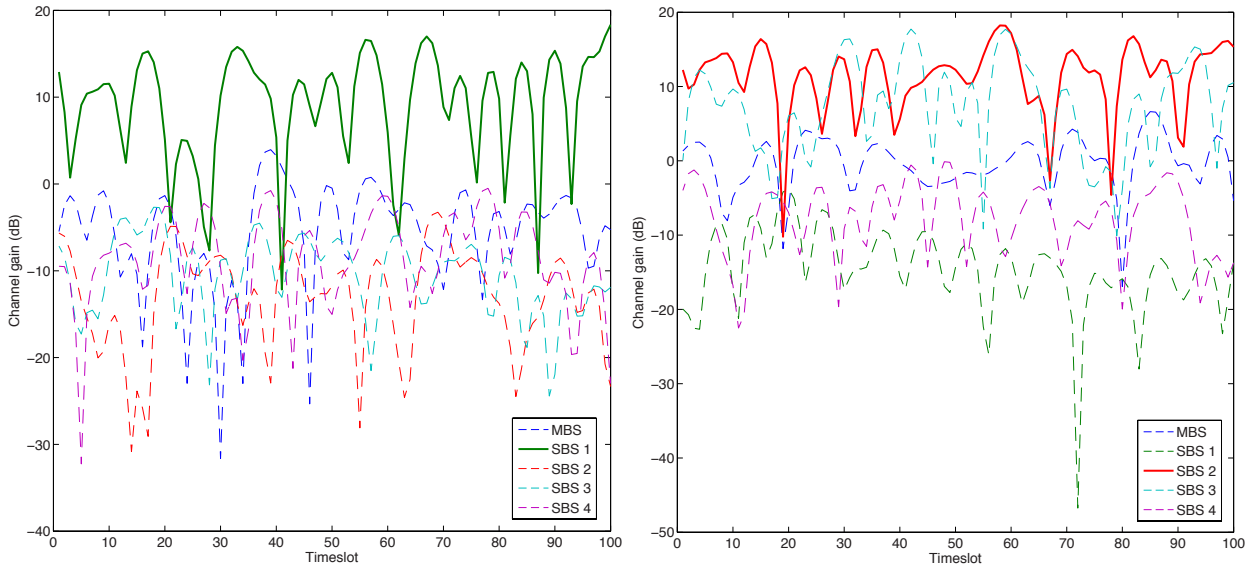
3.6 Simulation Results

Next, we present our simulation results for a wireless two-tier networks with $M = 1$ macrocell and $K = 4$ small cells. The MBS serves 3 macrocell users (MUs) whereas the SBSs each serve one small-cell user (SU). The SBSs and all users are randomly placed in a 1km by 1km grid region as shown in Figure 3.3. The MUs and SUs are denoted by triangles and rectangles respectively.

We model the large-scale path loss attenuation PL (in dB) using the WINNER model [74] described in (2.55). The simulation parameters are listed in Table 3.2. In addition, we model the time-varying small-scale fading as a correlated Rayleigh process in accordance with Clarke's scattering model [69]. This was simulated using the autoregressive modeling technique in [80]. Taking these into account, the time variations of the channel gains for the first and second SUs are, for instance, illustrated in Figure 3.4.

Table 3.2 – Path loss model and system parameters in Chapter 3.

Parameters	Values
Carrier frequency	$f_c = 1.9$ GHz
SBS-to-SU path loss (model A1-LOS)	$A = 1.87,\ B = 46.8,\ C = 20$
SBS-to-MU path loss (model A1-NLOS)	$A = 3.68,\ B = 43.8,\ C = 20$
MBS-to-SU path loss (model C1-NLOS)	$A = 3.36,\ B = 44.36,\ C = 23$
MBS-to-SU path loss (model C1-NLOS)	$A = 3.36,\ B = 44.36,\ C = 23$
Standard deviation of SBS-to-SU shadowing	$S = 3$ dB
Standard deviation of SBS-to-PU shadowing	$S = 4$ dB
Standard deviation of MBS-to-SU shadowing	$S = 6$ dB
MBS antenna power gain	12dB
SBS antenna power gain	5dB
Normalized maximum Doppler frequency	$f_d T_S = 0.01$
Total number of frames	$N_F = 10$
Time slot duration	$T_s = 1$sec
Length of energy-harvesting frame	$L = 10$ time slots



(a) Channels of SU #1 **(b)** Channels of SU #2

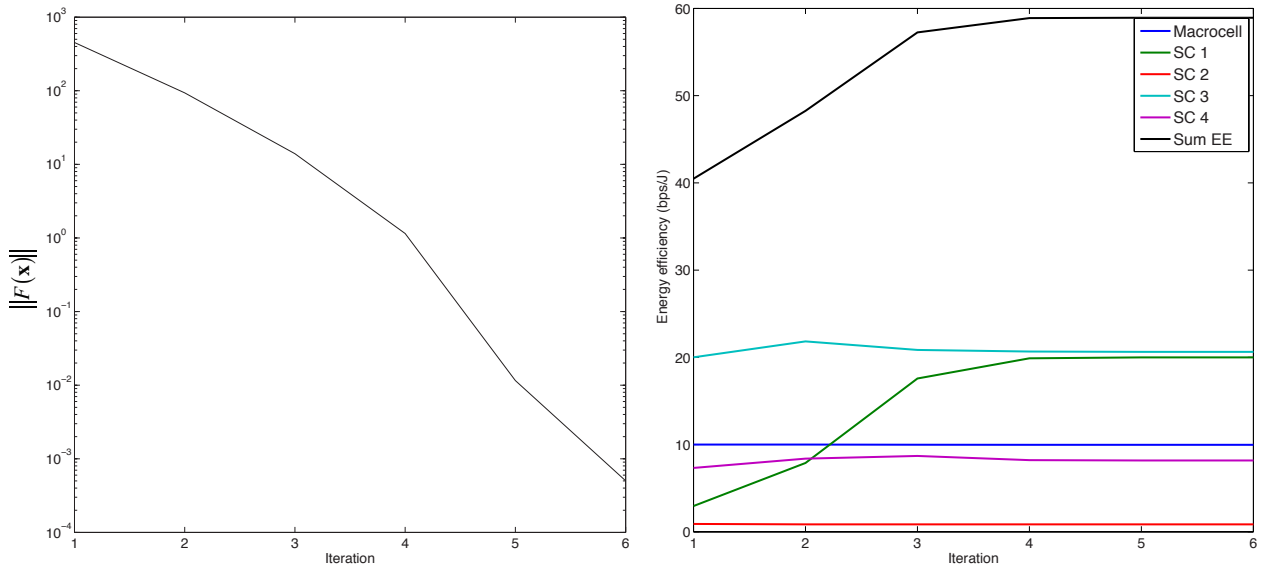
Figure 3.4 – Channel gains for the desired signal (solid line) and inter-cell interference (dotted line) received by SUs.

3.6.1 Convergence of Algorithms

First, we present the convergence of the proposed Algorithm 3.1. These algorithms were implemented in Matlab and we used the disciplined convex optimization software CVX to solve the convex parametric subproblems [1]. In Figure 3.5(a), we observed a fast convergence as the norm of $\mathbf{F}(\mathbf{x})$, which is defined in (3.17)-(3.18), rapidly decreased towards zero. Figure 3.5(b) shows that the energy efficiency of each cell also converges accordingly. For this example, only one inner iteration of the damped newton algorithm was needed to solve each subproblem $(Q_{\mathbf{q}})$. In other words, a full Newton step satisfied the norm-decrease condition. However, the line search is in general required to guarantee convergence. In our experiments, the algorithm converges within 10 to 20 inner iterations in most cases.

3.6.2 Analysis of Energy Efficiency Performance

Next, we analyze the energy efficiency of the proposed schemes. In our simulation, we normalized the transmit power P_c, the battery capacity $B_{\max}$ and the energy-harvesting rate



(a) Value of the norm $\left\|F\left(\mathbf{x}^{(n)}\right)\right\|$ **(b)** Value of minorized energy efficiency per cell

Figure 3.5 – Convergence of Algorithm 3.1.

of the renewable source by the noise power σ^2. Without loss of generality, we assume that the time slot duration $T_s = 1$sec to simplify the analysis. We set the maximum normalized transmit power of the SBS and MBS to 10 dB and 13 dB respectively. Then, the normalized battery capacity $B_{\max}$ is set to be equal to 10 dB and 17 dB. The main reasons to normalize the energy arrival rate and battery capacity to a decibel scale are to simplify the analysis and to facilitate reproducibility of our simulations. Since the contributions of this chapter is mainly theoretical, future studies could use the results presented here for performance benchmark. In such case, normalized energy levels instead of absolute energy in Watts helps abstract away implementation-specific details such as the type of energy storage system or energy sources.

In our simulations, we consider three different models for the renewable energy arrival per frame: (1) constant rate, (2) linearly increasing rate and (3) Poisson random variable. Although the energy arrival is commonly modeled as a Poisson renewal process [10, 82], the constant and linear rate models help gain some insight on the energy allocation. For fair comparison, we set the average harvested energy to be equal for these three scenarios. In addition, the energy rate is assumed to be equal for each cell. For the Poisson model,

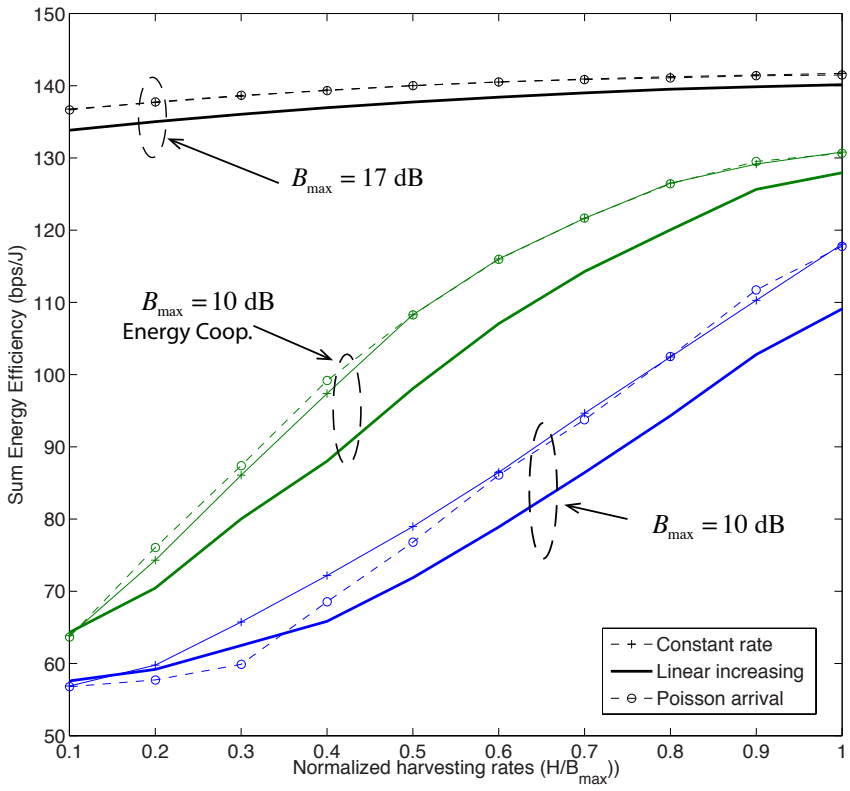
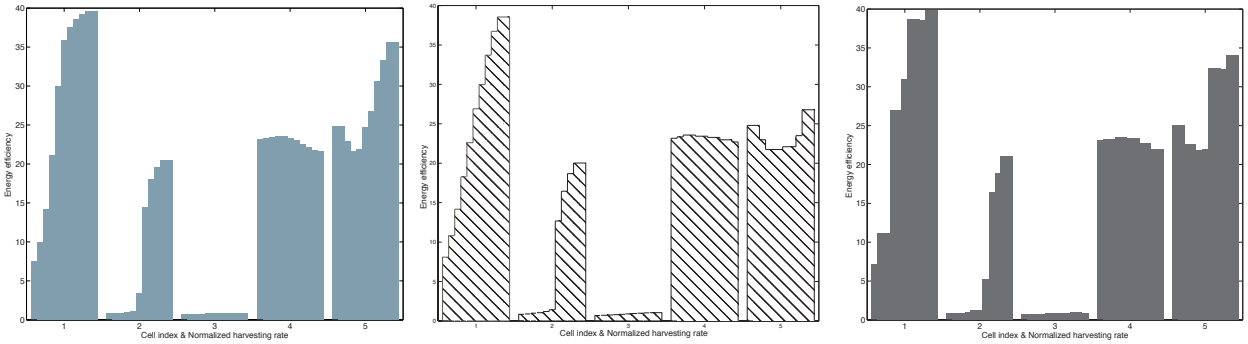


Figure 3.6 – Comparison of sum energy efficiency for different energy harvesting rates and arrivals.

we average the achieved energy efficiency over 100 different realizations of the energy arrivals.

Figure 3.6 shows the achieved sum energy efficiency for different energy harvesting rates. As expected, the energy efficiency performance increases with the energy arrival rate as this later is increased from $0.1B_{\max}$ to $B_{\max}$ per frame. Moreover, we see that increasing the capacity of the energy storage at each cell to $B_{\max} = 17$ dB significantly improves the energy efficiency. Finally, Figure 3.6 shows that under the same capacity limit $B_{\max} = 10$ dB, the energy-efficiency performance can be improved by exploiting the energy cooperation between the cells.

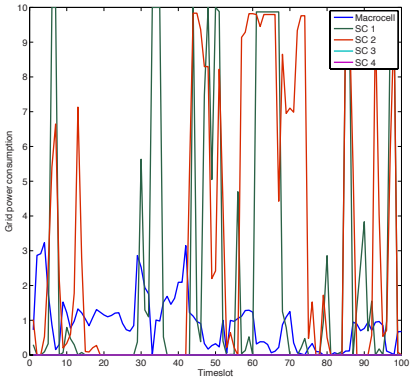
In the next Figure 3.7, we look at the energy efficiency at each cell when arrival rate is increased from $0.1B_{\max}$ to $B_{\max}$ per frame. We notice similar patterns for the per-cell energy efficiency under the three different assumptions for the energy arrival rates. Since SU#1 has favorable channel conditions (see Figure 3.4), the energy efficiency gain of



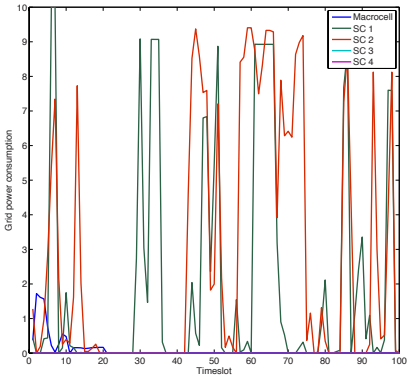
(a) Constant arrival rate **(b)** Linearly increasing rate **(c)** Poisson arrival

Figure 3.7 – Energy efficiency per cell for energy-harvesting rates between 0 dB and 10 dB.

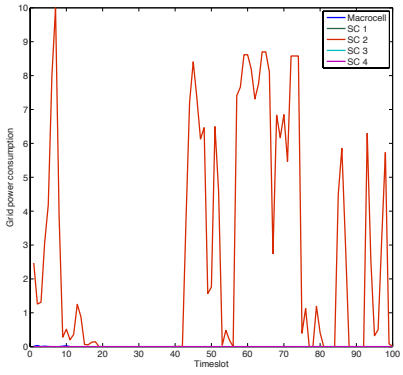
small-cell (SC) #1 (cell index 2 in Figure 3.7) is very high when the arrival rate increases from $0.1B_{\max}$ to $B_{\max}$. In contrast, small-cell #2 (cell index 3) suffers from strong inter-cell interference and its energy efficiency remain the same. Finally, Figure 3.8 shows the grid power consumption of each cell over time. It is evident from Figure 3.8 that energy cooperation decreases the required energy drawn from the conventional grid source.



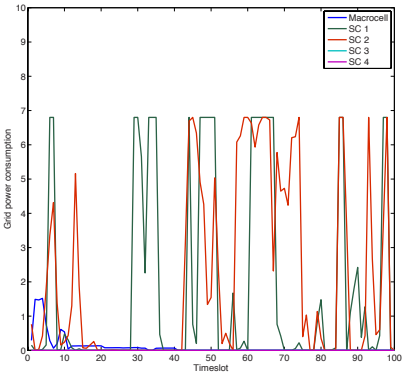
(b) No Cooperation - Rate $E_H = 0$dB



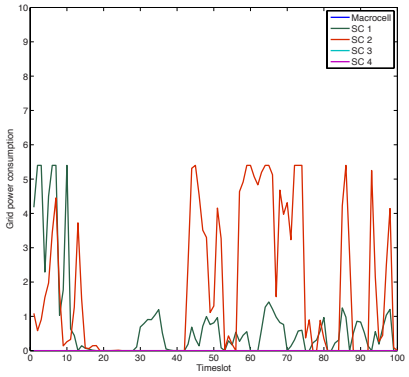
(c) No Cooperation - Rate $E_H = 6$dB



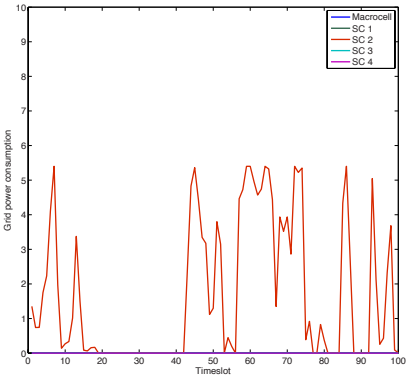
(d) No Cooperation - Rate $E_H = 10$dB



(f) Energy Cooperation - Rate $E_H = 0$dB



(g) Energy Cooperation - Rate $E_H = 6$dB



(h) Energy Cooperation - Rate $E_H = 10$dB

Figure 3.8 – Grid power consumed per cell for different Poisson arrival rates.

Chapter 4

Power minimization of Cooperative Clustered Small-cell Networks

4.1 Introduction

In Chapter 2 and Chapter 3, we consider two-tier networks in which the base stations have single antennas. In this chapter, we assume that the base stations can have multiple antennas. Therefore, we extend the power allocation to a cooperative multi-cell beamforming scheme, which is also known as coordinated multi-point transmission (CoMP) [83] or network MIMO. By enabling cooperation between the two-tier cells, we can reap larger capacity and diversity gains [84] and also to mitigate the interference [85]. This is achieved thanks to the increased spatial degrees of freedom of CoMP systems.

Despite these benefits of CoMP, it can also come with significant costs [11, 86]. First, the cooperating BSs must be connected by a backhaul through which they exchange the channel knowledge and user data. Second, cooperation largely increase the energy consumption due to the extra signal processing [11]. Although the theoretical benefits and practical issues of CoMP have been studied in many works, only few have looked into its energy efficient implementation.

In this chapter, we thus propose a flexible cooperation scheme which balances the trade-off between performance and energy efficiency. Instead of maximizing performance as in previous works [87–89], our goal is to enable the cells to satisfy their users and protect the primary users while minimizing their overall power consumption.

In Section 4.6, we present the state-of-the-art research on green CoMP and distributed antenna systems. Then, in Section 4.3, we propose a flexible and energy-aware cooperation scheme for cellular two-tier systems. Using the energy consumption model in [11], we formulate a cross-layer optimization of cooperative beamforming and clustering as a mixed integer program. By jointly optimizing the clustering and beamforming, we provide an extra degree of freedom for the cells to adapt their transmission according to changing parameters such as the users' locations, service requirements and channel conditions. We also propose a cognitive resource allocation mechanism to enable the spectrum sharing between the macrocell and small cells.

In Section 4.4, we present a framework for decomposing this problem into a clustering master problem and a beamforming subproblem. Then, we derive an optimally convergent algorithm using the generalized Benders decomposition (GBD) method [90]. We also propose simple techniques for accelerating the convergence of this algorithm.

Then, we propose a decentralized optimization algorithm for the cooperative beamforming when the clustering is fixed in Section 4.5. Assuming that the clusters do not overlap, we exploit the quasi-separability of the problem and derive a distributed algorithm based on the alternating direction method of multipliers (ADMM) [91]. We also propose limited signaling schemes that can considerably reduce the overhead.

Finally, numerical simulations are presented in Section 4.6 to analyze the trade-off between performance and energy consumption in a cooperative two-tier systems. In addition, we examine the convergence and effectiveness of the proposed algorithms. In Section 4.6, we also study the performance of the proposed decentralized beamforming and signaling schemes.

4.2 Related Works

The study of multi-cell joint-transmission system started with the landmark paper in [84]. Within the decade that followed [84], many research efforts followed on looking at the theoretical benefits of CoMP system from the viewpoint of spectrum efficiency [83]. Several papers have also been published on the practical aspects and issues, such as the impact of limited backhaul signaling [36] or the need for antenna selection and clustering [92].

However, the energy efficiency of CoMP has not received much attention until later. For instance, [11] was the first to look at the energy consumption model of cooperating base stations. In [93], an energy efficiency analysis of an idealized CoMP system was presented. The trade-off between energy efficiency and spectrum efficiency has been studied by [94]. Then, [45] studied the energy-efficient power allocation in distributed antenna systems. However, [45] assumed a zero-forcing strategy and did not try to optimize the cooperative beamforming. However, zero-forcing beamforming is limited by the number of antennas available at each base station. Therefore, a small number of users can be accomodated to avoid interference between the multi-cell transmissions. [95] tackled the difficult problem of energy-efficient transmit precoding design using a parameterization of the Pareto boundary. They proposed an iterative method in which each user optimally solves a parameterized energy-efficiency maximization problem and updates the inter-user interference parameters at each iteration. Nonetheless, they could not prove the optimal convergence of such algorithm. In contrast, we propose in this chapter a multi-cell beamforming optimization scheme for small-cell networks. Instead of solving an energy efficiency maximization problem, we aim to minimize the overall energy consumption of the CoMP system. In contrast to previous works, we propose a scheme that jointly optimizes the clustering and the beamforming.

After the results of this chapter were published in [96], there were other papers that considered the energy efficiency of CoMP. For instance, [97] gave a performance analysis and designed an optimization scheme for green heterogeneous CoMP networks. Another

study in [98], which is based on a similar idea proposed in this chapter, investigates different base station sleeping strategies for energy-efficient small-cells.

4.3 System Model

We consider the downlink of a clustered cooperative small-cell networks whose base stations $\mathcal{B} = \{b_1, \ldots, b_{|\mathcal{B}|}\}$ are connected through a backhaul interface. Each base station has m antennas. By sharing their antennas, the small-cell base stations can cooperate to serve a pool of users $\mathcal{U} = \{u_1, \ldots, u_{|\mathcal{U}|}\}$. Here, we make the following assumptions. The small cells also co-exist with a primary macro-cell system using a cognitive underlay strategy. In other words, they share the spectrum spectrum under the condition that the secondary small-cell transmissions do not exceed specific interference limits that are imposed by the primary macrocell users.

We assume perfect channel information and perfect synchronization for the cooperative beamforming. Next, the backhaul interface is assumed to have enough capacity for exchanging the user data and signaling information. Hence, we consider the following cooperation schemes for the secondary small cells:

1. **Full cooperation** is achieved when all base stations cooperate to jointly serve the users $\mathcal{U}$. Essentially, this strategy corresponds to a $m\,|\mathcal{B}| \times |\mathcal{U}|$ network-MIMO transmission [84]. It provides the highest spatial degrees of freedom for multiplexing and for interference mitigation but incurs the highest cost in terms of signaling and energy consumption.

2. **Inter-cell coordination** is the simplest scheme as each user is served by a single base station. In this case, the interference is mitigated by having the base stations coordinate the base station-to-user assignment and beamforming.

3. **Flexible cooperation** allows each user u to be served either by a single base station or by multiple cooperating ones. With this flexible strategy, it is thus possible to balance performance and energy saving.

In our model, we select the flexible cooperation as a general strategy and consider full cooperation and inter-cell coordination as special cases. Let us define a binary vector $\mathbf{x} = (x_{bu})_{b\in\mathcal{B},\ u\in\mathcal{U}} \in \{0,1\}^{|\mathcal{B}|\cdot|\mathcal{U}|\times 1}$ to model the clustering, where $|\mathcal{S}|$ denotes the cardinal of a set $\mathcal{S}$. The component x_{bu} of $\mathbf{x}$ determines whether base station b serves the user u or not:

$$x_{bu} = \begin{cases} 1 & \text{if user } u \text{ is served by BS } b \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

Thus, the special case of *full cooperation* corresponds to $\mathbf{x} = \mathbf{1}_{|\mathcal{B}|\cdot|\mathcal{U}|\times 1}$ so that each user is jointly served by all base stations. On the other hand, the special case of *inter-cell coordination* requires the following equality to hold:

$$\sum_{b\in\mathcal{B}} x_{bu} = 1 \quad \forall u \in \mathcal{U} \tag{4.2}$$

to enforce that one and only one base station will be assigned to each user u.

Next, let us describe the signal model by first assuming every base stations serve each user. Let $\mathbf{h}_u \triangleq (\mathbf{h}_{bu})_{b\in\mathcal{B}} \in \mathbb{C}^{m|\mathcal{B}|\times 1}$ be the channel vector from all base stations to a user u where $\mathbf{h}_{bu} \in \mathbb{C}^{m\times 1}$ is the channel between base station b and user u. Each channel vector $\mathbf{h}_{bu}$ accounts for both the large-scale path loss attenuation as well as the small-scale Rayleigh fading. Similarly, $\mathbf{v}_u \triangleq (\mathbf{v}_{bu})_{b\in\mathcal{B}} \in \mathbb{C}^{m|\mathcal{B}|\times 1}$ is the concatenated beamforming vector for user u with $\mathbf{v}_{bu} \in \mathbb{C}^{m\times 1}$. Then, all base stations cooperatively transmit the data symbol s_u of a user u through the channel $\mathbf{h}_u$ (see Figure 4.1). Its received signal is thus given by:

$$y_u = \mathbf{h}_u^H \mathbf{v}_u s_u + \sum_{j\in\mathcal{U}\setminus\{u\}} \mathbf{h}_u^H \mathbf{v}_j s_j + n_u \tag{4.3}$$

Table 4.1 – List of key parameters in Chapter 4.

Symbol	**Parameter description**
$\mathcal{B}$	Set of base stations
$\mathcal{U}$	Set of users
$\lvert\mathcal{B}\rvert$	Number of base stations
$\lvert\mathcal{U}\rvert$	Total number of users
x_{bu}	Binary indicator for user u being served by BS b
y_b	Binary activity status of BS b
$\mathbf{h}_u$	Channel vector from all base stations to a user u
$\mathbf{h}_{bu}$	Channel vector between base station b and user u.
$\mathbf{v}_u$	Concatenated beamforming vector for user u
$\mathbf{v}_{bu}$	Beamforming vector of BS b for user u
SINR_u	SINR achieved by user u
ρ_u	Minimum SINR requirement for user u
R_u	Achieved rate of user u
P_b	Transmit power limit of BS b
$\mathcal{V}$	Set of primary users
$\mathbf{g}_{bv}$	Channel from the BS b to the primary user v.
P_{total}	Total power consumption of the system
P_{tx}	Transmit power radiated by all base stations
P_{sp}	Total power cost due to multi-cell processing
P_{bh}	Backhaul power cost
P_0	Total fixed power cost of the BSs

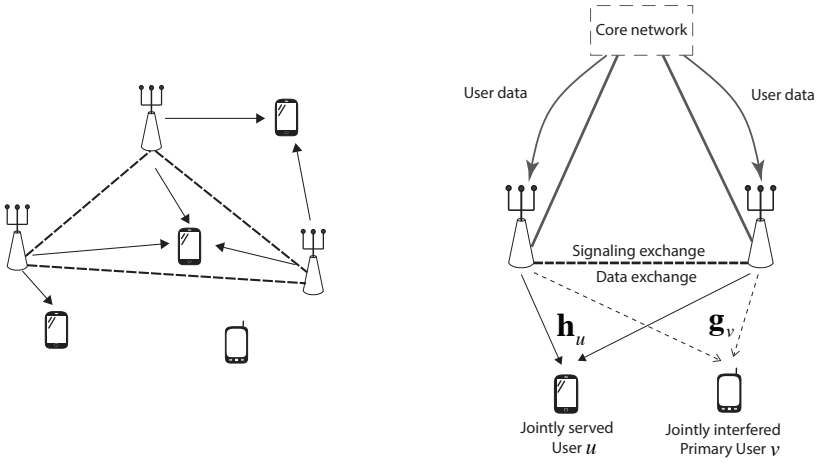


Figure 4.1 – Coordinated multi-point transmission model in distributed and cognitive small-cell networks.

where the symbol $(\cdot)^H$ means the conjugate transpose. Moreover, $n_u \sim \mathcal{CN}(0, \sigma_u^2)$ is a circularly complex Gaussian noise with zero-mean and a variance σ_u^2. Therefore, the signal-to-interference-plus-noise ratio (SINR) of the received signal y_u is:

$$\text{SINR}_u = \frac{\left|\mathbf{h}_u^H \mathbf{v}_u\right|^2}{\sum_{j \in \mathcal{U} \setminus \{u\}} \left|\mathbf{h}_u^H \mathbf{v}_j\right|^2 + \sigma_u^2} \tag{4.4}$$

and its achievable rate is given by:

$$R_u = \log\left(1 + \text{SINR}_u\right) \quad [\text{bps/Hz}] \tag{4.5}$$

4.4 Joint Clustering and Beamforming Optimization

Our objective is to minimize the total power consumption of the system given SINR targets of the small-cell users and interference power constraints at the primary users.

4.4.1 Beamforming Power Constraints

Each small-cell base station is subject to power limit constraints:

$$\sum_{u \in \mathcal{U}} \|\mathbf{v}_{bu}\|_2^2 \leq P_b, \ \forall b \in \mathcal{B} \tag{4.6}$$

To model the *flexible cooperation* scheme, we make sure that if a specific base station b does not serve user u , i.e. when $x_{bu} = 0$, then the corresponding beamforming vector $\mathbf{v}_{bu}$ becomes a zero vector. In other words, we enforce that a base station b allocates a non-zero power for user u only if $x_{bu} = 1$, through the mixed-integer constraints:

$$\|\mathbf{v}_{bu}\|_2^2 \leq x_{bu} \cdot P_b, \ \forall u \in \mathcal{U}, \ \forall b \in \mathcal{B} \tag{4.7}$$

With constraints (4.6), the mixed integer constraint (4.7) simply becomes redundant when $x_{bu} = 1$. Although the SINR definition in (2.1) initially assumes *full cooperation*, it is also valid for the *flexible cooperation* scheme after we added constraints (4.7). In fact, when a subset of BSs $\overline{\mathcal{B}} \subset \mathcal{B}$ does not serve user u, then constraints (4.7) enforce $\|\mathbf{v}_{bu}\|^2 = 0, \ \forall b \in \overline{\mathcal{B}}$.

In addition, we assume that the two-tier network can co-exist with a primary system whose users are $\mathcal{V} = \{v_1, \ldots, v_{|\mathcal{V}|}\}$. Using a cognitive underlay strategy, the small-cell base stations opportunistically use the same channel licensed to the primary system under the condition that they do not exceed a total interference temperature limit $I_v \geq 0$ imposed by each primary user v as follows:

$$\sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \left| \mathbf{g}_{bv}^H \mathbf{v}_{bu} \right|^2 \leq I_v, \quad \forall v \in \mathcal{V} \tag{4.8}$$

with $\mathbf{g}_{bv} \in \mathbb{C}^{m \times 1}$ being the channel from the small-cell base station b to the primary user v. Note that these constraints couple the transmissions of the small-cell base stations.

4.4.2 Energy Consumption Model

A linear model for the energy consumption of a CoMP system was given by [11, 58] as follows:

$$P_{\text{total}} = P_{\text{tx}} + P_{\text{sp}} + P_{\text{bh}} + P_0 \tag{4.9}$$

where P_{total}, P_{tx}, P_{sp}, P_{bh} and P_0 denote the total power consumption, the transmit power radiated by all base stations, the signal processing power cost, the backhaul power cost and the fixed power cost respectively. Except for the static consumed power, these costs scale dynamically with the load, i.e. the number of wireless links between the users and their serving base stations, as follows [11, 58]:

$$P_{\text{tx}} = \frac{\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \|\mathbf{v}_{bu}\|_2^2}{\mu_{PA}} (1 + C_{\text{PS}})$$
$$P_{\text{sp}} = \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} x_{bu} p_{sp}$$
$$P_{\text{bh}} = \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} x_{bu} r_u p_{bh}$$

where r_u is the data rate for user u, p_{bh} is the power consumed for transmitting one bit of information over the backhaul and p_{sp} is the power consumed per link for the cooperative processing. Note that in contrast to the cooperation schemes studied in Chapter 2 and Chapter 3, the multi-cell cooperation considered in this chapter may lead to significant energy cost and this cost scales with the cluster size of the cooperating base stations [11]. Finally, the static power P_0 is the fixed power consumed by the active base stations:

$$P_0 = \sum_{b\in\mathcal{B}} y_b p_{\text{static}} \tag{4.10}$$

where the binary variable y_b defines the activity status of base station b such that:

$$y_b = \begin{cases} 1 & \text{if base station } b \text{ is actively serving some user(s)} \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

Therefore, we can define the overall power consumption of the network by a cost function f of the beamforming variable $\mathbf{v}$ and of the binary variables $\mathbf{x}$ and $\mathbf{y}$:

$$f(\mathbf{x},\mathbf{y},\mathbf{v}) = \frac{1 + C_{\text{PS}}}{\mu_{PA}} \cdot \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \|\mathbf{v}_{bu}\|_2^2 + \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} x_{bu}(p_{sp} + r_u p_{bh}) + \sum_{b\in\mathcal{B}} y_b p_{\text{static}} \tag{4.12}$$
$$= f_1(\mathbf{v}) + f_2(\mathbf{x},\mathbf{y}) \tag{4.13}$$

where the term f_1 in (4.12) is the transmit-dependent power and the term f_2 is the sum of the cooperative processing power and fixed power. Note that $\mathbf{v} \triangleq (\mathbf{v}_u)_{u\in\mathcal{U}}$ concatenates all beamforming vectors of small-cell users.

4.4.3 Problem Formulation

Finally, we formulate the joint optimization of clustering and cooperative beamforming as:

$$\begin{aligned}
&\text{variables } \mathbf{x}\ \mathbf{y}\ \mathbf{v} \\
&\text{minimize } f(\mathbf{x},\mathbf{y},\mathbf{v}) && (4.14)\\
&\text{subject to } x_{bu} \in \{0,1\},\ \forall u \in \mathcal{U},\ \forall b \in \mathcal{B} && (4.15)\\
&\qquad y_b \in \{0,1\},\ \forall b \in \mathcal{B} && (4.16)\\
&\qquad y_b \leq \sum_{u\in\mathcal{U}} x_{bu},\ \forall b \in \mathcal{B} && (4.17)\\
&\qquad y_b \geq x_{bu},\ \forall u \in \mathcal{U},\ \forall b \in \mathcal{B} && (4.18)\\
&\qquad \text{SINR}_u(\mathbf{v}) \geq \rho_u,\ \forall u \in \mathcal{U} && (4.19)\\
&\qquad \sum_{u\in\mathcal{U}} \|\mathbf{v}_{bu}\|_2^2 \leq P_b,\ \forall b \in \mathcal{B} && (4.20)\\
&\qquad \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \left|\mathbf{g}_{bv}^H \mathbf{v}_{bu}\right|^2 \leq I_v, \quad \forall v \in \mathcal{V} && (4.21)\\
&\qquad \|\mathbf{v}_{bu}\|_2^2 \leq x_{bu}\cdot P_b,\ \forall u \in \mathcal{U},\ \forall b \in \mathcal{B} && (4.22)
\end{aligned}$$

where constraints (4.17)-(4.18) ensure that a base station is active if and only if it serves some users. In (4.19), ρ_u denotes the target SINR of user u. The above problem is hard to solve due to the non-convexity of the SINR constraints and to its combinatorial nature.

Let us create the matrix $\mathbf{V} = \left[\mathbf{v}_{u_1}, \ldots, \mathbf{v}_{u_{|\mathcal{U}|}}\right] \in \mathbb{C}^{m|\mathcal{B}|\times|\mathcal{U}|}$ which concatenates all beamforming vectors and define the following set $\mathcal{F}_{\mathbf{v}}$ of beamforming vectors:

$$\mathcal{F}_{\mathbf{v}} = \left\{ \mathbf{v} \in \mathbb{C}^{m|\mathcal{B}|\cdot|\mathcal{U}|\times 1} : \begin{array}{l} \left\| \begin{bmatrix} \mathbf{h}_u^H \mathbf{V} & \sigma_u \end{bmatrix} \right\| \leq \sqrt{1+\frac{1}{\rho_u}} \mathbf{h}_u^H \mathbf{v}_u, \ \forall u \in \mathcal{U} \\ \text{Re}\left(\mathbf{h}_u^H \mathbf{v}_u\right) \geq 0, \ \forall u \in \mathcal{U} \\ \text{Im}\left(\mathbf{h}_u^H \mathbf{v}_u\right) = 0, \ \forall u \in \mathcal{U} \\ \sum_{u\in\mathcal{U}} \|\mathbf{v}_{bu}\|_2^2 \leq P_b^{\max}, \ \forall b \in \mathcal{B} \\ \sum_{b\in\mathcal{B}} \sum_{u\in\mathcal{U}} \left|\mathbf{g}_{bv}^H \mathbf{v}_{bu}\right|^2 \leq I_v, \quad \forall v \in \mathcal{V} \end{array} \right\} \tag{4.23}$$

Thus, the following lemma tells us how to reduce the complexity of the problem.

Lemma 4.1. *By fixing the values of the clustering variables to* $(\mathbf{x}^k,\mathbf{y}^k)$*, problem* (4.14) *can be reduced to a convex beamforming optimization problem* $\left(S^k\right)$ *defined by:*

$$\begin{aligned} &\textit{variable } \mathbf{v} \\ &\textit{minimize } f_1(\mathbf{v}) + f_2\left(\mathbf{x}^k,\mathbf{y}^k\right) && (4.24) \\ &\textit{subject to } \mathbf{v} \in \mathcal{F}_{\mathbf{v}} && (4.25) \\ &\qquad \|\mathbf{v}_{bu}\|_2^2 \leq x_{bu}^k \cdot P_b^{max}, \ \forall (u,b) \in \mathcal{U}\times\mathcal{B} && (4.26) \end{aligned}$$

and the constraint set $\mathcal{F}_{\mathbf{v}}$ *defined in* (4.23) *is convex and compact.*

Proof. See Appendix C.1 □

4.4.4 Generalized Benders Decomposition Method

Given the previous results, we will next derive an provably-convergent iterative algorithm for solving problem (4.14). Using the Benders decomposition method [90], we will separate the clustering optimization and the cooperative beamforming design. Define $\mathcal{H}$ as the set of clustering variables for which the corresponding beamforming problem (S) admits a feasible solution:

$$\mathcal{H} = \left\{ (\mathbf{x}, \mathbf{y}) : \exists \mathbf{v} \in \mathcal{F}_{\mathbf{v}} \ \left\| \mathbf{v}_{bu} \right\|_2^2 \leq x_{bu} \cdot P_b^{\max}, \ \forall (u,b) \in \mathcal{U} \times \mathcal{B} \right\} \tag{4.27}$$

Then, we can rewrite the original problem as a clustering master problem (M):

$$(M) \begin{cases} \underset{\mathbf{x}\ \mathbf{y}}{\text{minimize}} & g(\mathbf{x}, \mathbf{y}) \\ \text{subject to} & (\mathbf{x},\mathbf{y}) \in \mathcal{H} \cap \mathcal{F}_{(\mathbf{x},\mathbf{y})} \end{cases} \tag{4.28}$$

For any arbitrary $(\mathbf{x},\mathbf{y})$, the objective function g of problem (M) returns the optimal value of the beamforming optimization (S) in (4.24). In (4.28), the set $\mathcal{F}_{(\mathbf{x},\mathbf{y})}$ gathers the constraints that are only related to the clustering variables as:

$$\mathcal{F}_{(\mathbf{x},\mathbf{y})} = \left\{ (\mathbf{x},\mathbf{y}) : \begin{array}{l} x_{bu} \in \{0,1\}, \ \forall (u,b) \in \mathcal{U} \times \mathcal{B} \\ y_b \in \{0,1\}, \ \forall b \in \mathcal{B} \\ y_b \leq \sum_{u \in \mathcal{U}} x_{bu}, \ \forall (u,b) \in \mathcal{U} \times \mathcal{B} \\ y_b \geq x_{bu}, \ \forall (u,b) \in \mathcal{U} \times \mathcal{B} \end{array} \right\} \tag{4.29}$$

The equivalence between the original problem (4.14) and (4.28) is stated below.

Proposition 4.1. *If* $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{v}^*)$ *is an optimal solution of the original mixed-integer problem in* (4.14)*, then* $(\mathbf{x}^*, \mathbf{y}^*)$ *is optimal for the clustering master problem* (M) in (4.28). *Conversely, if* $(\mathbf{x}^*, \mathbf{y}^*)$ *solves problem* (M) *and if* $\mathbf{v}^*$ *solves the beamforming problem* (S) *in* (4.24) *with* $(\mathbf{x},\mathbf{y}) = (\mathbf{x}^*,\mathbf{y}^*)$*, then* $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{v}^*)$ *solves the original problem* (4.14). *Moreover, the original problem is infeasible if and only if* (M) is infeasible.

Proof. The proof of the above proposition follows directly from the partitioning procedure described above and from the definitions of the function g and of the master problem (M). □

To exploit the convexity of (S), we derive an iterative algorithm in which we solve the subproblem (S) and update the clustering variables at each iteration. But the master problem (M) cannot be solved in its current form (4.28) since the function g is only known implicitly. To circumvent this, we will rewrite (M) in its dual form and apply the GBD method to solve it iteratively.

Dual representation of the master problem

Let us consider the subproblem (S) and define its partial Lagrangian with respect to the *complicating constraints*:

$$\mathcal{L}\left(\mathbf{v}, \lambda; \mathbf{x}, \mathbf{z}\right) \triangleq f\left(\mathbf{v}; \mathbf{x}, \mathbf{z}\right) + \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \lambda_{bu} \left(\|\mathbf{v}_{bu}\|_2^2 - x_{bu} \cdot P_b \right) \tag{4.30}$$

where the Lagrange multipliers satisfy $\lambda_{bu} \geq 0, \ \forall \left(u, b\right) \in \mathcal{U} \times \mathcal{B}$.

Moreover, we define a second type of Lagrangian function $\overline{\mathcal{L}}$ with a parameter $\mathbf{x}^k$ as:

$$\overline{\mathcal{L}}\left(\mathbf{v}, \mu, \mathbf{x}^k\right) \triangleq \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \mu_{bu} \left(\|\mathbf{v}_{bu}\|_2^2 - x_{bu}^k \cdot P_b^{\max} \right) \tag{4.31}$$

where the multiplier μ also satisfies $\mu_{bu} \geq 0, \ \forall \left(u, b\right) \in \mathcal{U} \times \mathcal{B}$. Note that any convex programming algorithm such as interior point methods can produce the optimal multiplier vector λ^k along with $\mathbf{v}^k$ without any extra computations [61]. Using these Lagrangian functions, we can obtain a more explicit reformulation of problem (M) by using the following result.

Proposition 4.2. *The master problem* (M) *in* (4.28) *is equivalent to the following problem* (P')*:*

$$(P') \begin{cases} \textit{variables} & \mathbf{x}\ \mathbf{y}\ \gamma \\ \textit{minimize} & \gamma \\ & (\mathbf{x},\mathbf{y}) \in \mathcal{F}_{(\mathbf{x},\mathbf{y})} \\ & \gamma \geq \xi\left(\mathbf{x},\mathbf{y};\lambda\right),\ \forall \lambda \geq 0 \\ & 0 \geq \overline{\xi}\left(\mathbf{x};\mu\right),\ \forall \mu \geq 0:\ \sum_{(b,u)\in\mathcal{B}\times\mathcal{U}} \mu_{bu} = 1 \end{cases} \tag{4.32}$$

where the support functions ξ and $\overline{\xi}$ are defined as the minimum of the Lagrange functions with respect to $\mathbf{v}$:

$$\xi\left(\mathbf{x},\mathbf{y};\lambda\right) = \min_{\mathbf{v}\in\mathcal{F}_{\mathbf{v}}} \mathcal{L}\left(\mathbf{v},\lambda,\mathbf{x},\mathbf{y}\right) \tag{4.33}$$

$$\overline{\xi}\left(\mathbf{x};\mu\right) = \min_{\mathbf{v}\in\mathcal{F}_{\mathbf{v}}} \overline{\mathcal{L}}\left(\mathbf{v},\mu,\mathbf{x}\right) \tag{4.34}$$

Proof. See Appendix C.2. □

Although the equivalent formulation (P') is more explicit than (M), there are still two issues with (P'). First, the functions ξ and $\overline{\xi}$ defining constraints (4.32) are represented as inner minimization problems. Second, (P') involves an infinite number of constraints which makes it computationally impossible to solve. To overcome these, we first obtain an explicit expression for each support function ξ and $\overline{\xi}$.

Lemma 4.2. *If $\mathbf{v}^k$ and λ^k are optimal primal and dual solutions for the convex subproblem $\left(S^k\right)$, then the function ξ of $(\mathbf{x},\mathbf{y})$, when parameterized by λ^k,* is explicitly given by*:*

$$\xi\left(\mathbf{x},\mathbf{y};\lambda^k\right) = \mathcal{L}\left(\mathbf{v}^k,\lambda^k,\mathbf{x},\mathbf{y}\right),\quad \forall\left(\mathbf{x},\mathbf{y}\right) \tag{4.35}$$

Proof. See Appendix C.3. □

In a similar way, we can also obtain an explicit expression of the function $\overline{\xi}$ as explained next.

Lemma 4.3. *Suppose that* $\left(\overline{\mathbf{v}}^k, \alpha^k\right)$ *and* μ^k *are optimal primal and dual solutions for the following convex feasibility problem* $\left(F^k\right)$*:*

$$\left(F^k\right) \begin{cases} \textit{variables} & \mathbf{v}\ \alpha \\ \textit{minimize} & \alpha \\ \textit{subject to} & \mathbf{v} \in \mathcal{F}_{\mathbf{v}} \\ & \alpha \geq 0 \\ & \|\mathbf{v}_{bu}\|_2^2 - x_{bu}^k \cdot P_b^{max} \leq \alpha,\ \forall\, (u,b) \in \mathcal{U} \times \mathcal{B} \end{cases} \tag{4.36}$$

Then, the function $\overline{\xi}$ *of* $\mathbf{x}$*, when parameterized by* μ^k*, is explicitly given by:*

$$\overline{\xi}\left(\mathbf{x}; \mu^k\right) = \overline{\mathcal{L}}\left(\overline{\mathbf{v}}^k, \mu^k, \mathbf{x}\right), \quad \forall \mathbf{x} \tag{4.37}$$

and the optimal multiplier μ^k *must satisfy* $\sum_{(b,u)\in\mathcal{B}\times\mathcal{U}} \mu_{bu} = 1$.

In the feasibility problem $\left(F^k\right)$ in (4.36), we relax the complicating constraints (4.22) and minimize the maximum constraint violation α. Note that $\left(S^k\right)$ is infeasible if and only if the optimal value of $\left(F^k\right)$ is strictly positive. Next, we will use *relaxation* to solve the master problem (P') iteratively with a finite number of constraints.

Algorithmic procedure

The basic idea of the GBD method is to generate, at each iteration, an upper bound and a lower bound on the optimum of the original problem until those bounds meet. The lower bound is obtained by solving a relaxed version of problem (P') and the upper bound is

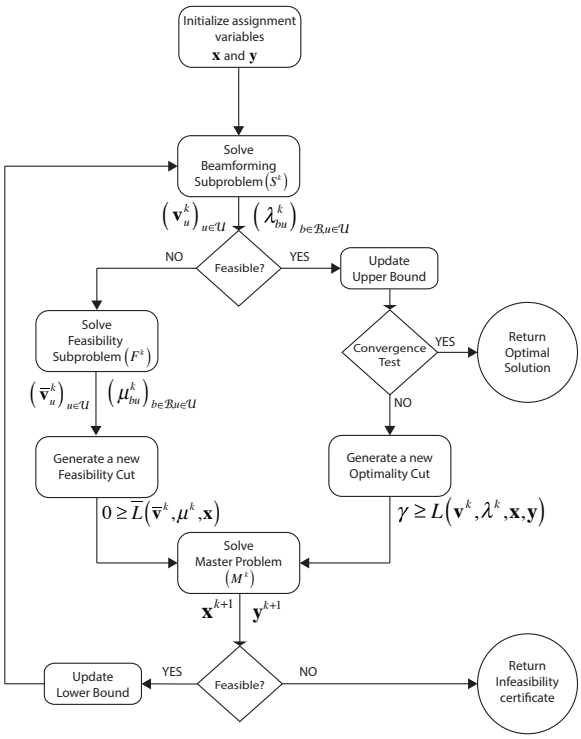


Figure 4.2 – Generalized Benders Decomposition procedure for joint BS assignment and multi-cell beamforming problem.

computed from the optimal values of beamforming subproblems $\left(S^k\right)$. The overall algorithm is presented in the flowchart in Figure 4.2.

First, we start with an initial assignment $\mathbf{x} = \mathbf{x}^0$ and $\mathbf{y} = \mathbf{y}^0$. Next, we solve the subproblem $\left(S^k\right)$ which, when feasible, will give us an optimal pair $\left(\mathbf{v}^k, \lambda^k\right)$ of beamformers and Lagrange multipliers. Then, we use these to generate the following optimality cut $\gamma \geq \mathcal{L}\left(\mathbf{v}^k, \lambda^k, \mathbf{x}, \mathbf{y}\right)$. Otherwise, when subproblem $\left(S^k\right)$ is infeasible, we solve the feasibility problem $\left(F^k\right)$ and use the solution and multiplier $\left(\overline{\mathbf{v}}^k, \mu^k\right)$ to generate a feasibility cut $0 \geq \overline{\mathcal{L}}\left(\overline{\mathbf{v}}^k, \mu^k, \mathbf{x}\right)$. Depending on whether $\left(S^k\right)$ is feasible or not, an optimality or feasibility cut is added to the following relaxed version of the master problem (M) at each iteration k:

$$
\left(M^k\right)\begin{cases}
\text{variables} & \gamma, \mathbf{x}, \mathbf{y} \\
\text{minimize} & \gamma \\
\text{subject to} & (\mathbf{x}, \mathbf{y}) \in \mathcal{F}_{(\mathbf{x},\mathbf{y})} \\
& \gamma \geq \mathcal{L}\left(\mathbf{v}^l, \lambda^l, \mathbf{x}, \mathbf{y}\right), \quad \forall l \in \mathcal{I}^k_{\text{feas}} \\
& 0 \geq \overline{\mathcal{L}}\left(\overline{\mathbf{v}}^j, \mu^j, \mathbf{x}\right), \quad \forall j \in \mathcal{I}^k_{\text{infeas}}
\end{cases}
$$

where $\mathcal{I}^k_{\text{feas}} = \left\{l \in \{1, \ldots, k\} \mid \left(S^l\right) \text{ feasible}\right\}$ and $\mathcal{I}^k_{\text{infeas}} = \left\{l \in \{1, \ldots, k\} \mid \left(\mathcal{S}^l\right) \text{ infeasible}\right\}$ are the sets of iteration indices up to k for which the subproblems $\left\{\left(S^l\right)\right\}_{l=1,\ldots,k}$ were feasible or infeasible respectively.

Solving the relaxed problem $\left(M^k\right)$ will give new binary clustering variables $\left(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\right)$ for the next iteration $k+1$. In addition, the optimal value γ^k of $\left(M^k\right)$ gives a lower bound $\text{LB}_k = \gamma^k$ for the optimal value of the original problem. This lower bound is non-decreasing with k because adding a new cut at each iteration can only shrink the feasibility set. On the other hand, the upper bound is updated at each iteration k as $\text{UB}_k \triangleq \min_{l \in \mathcal{I}^k_{\text{feas}}} f\left(\mathbf{x}^l, \mathbf{y}^l, \mathbf{v}^l\right)$, which is clearly non-increasing with k. This iterative process is carried on until the difference between these bounds is less than a prespecified tolerance $\epsilon \geq 0$. From [90], the finite convergence property of this algorithm is restated below.

Proposition 4.3. *For any given $\epsilon \geq 0$, the GBD method algorithm terminates in a finite number of iterations.*

Proof. The proof directly follows Lemma 4.1 and from the results of Geoffrion in [90]. □

Convergence acceleration techniques Although the optimal convergence of the Benders' algorithm is guaranteed, it can exhibit slow convergence in practice. In fact, the

binary assignment variables $\left(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}\right)$ obtained from $\left(M^k\right)$ does not necessarily guarantee that $\left(S^{k+1}\right)$ will be feasible at each iteration k. Another undesirable feature of cutting plane algorithms such as Benders decomposition is that in early iterations, the solutions tend to oscillate from one region of the feasible set to another, thus slowing convergence [99]. Thus, we propose two techniques to speed up the convergence of the classical Benders' algorithm.

- First, when the subproblem $\left(S^k\right)$ becomes infeasible, we solve the feasibility problem $\left(F^k\right)$ and construct a feasibility cut as prescribed by the classical algorithm. Then, we check all pairs $\mathcal{A}$ of base station and user that create the maximum violation $\hat{\alpha}$:

$$\mathcal{A} = \left\{(b,u) \in \mathcal{B} \times \mathcal{U} \mid x_{bu}^k = 0,\ \|\mathbf{v}_{bu}\|_2^2 - x_{bu}^k \cdot P_b = \hat{\alpha}\right\} \tag{4.38}$$

  After that, we select a new feasible assignment variables $\mathbf{x}$ by setting $x_{bu} = 1,\ \forall\,(b,u) \in \mathcal{A}$, solve the corresponding subproblem to get an extra optimality cut, and add this later to the master problem with the feasibility cut. The idea is to restore the feasibility whenever we encounter an infeasible assignment and to generate as many optimality cuts as possible in the first iterations.

- Second, we add valid cuts to the relaxed master problem to speed up the algorithm. Adding extra constraints shrinks the feasibility set and helps find feasible and more efficient solution in early iterations. Using our knowledge of the clustering and cooperative beamforming problem, we add the following constraints:

$$\begin{cases} \sum_{b \in \mathcal{B}} x_{bu} \geq 1,\ \forall u \in \mathcal{U} \\ \\ x_{\bar{b}_u u} \geq x_{bu},\ \forall b \in \mathcal{B} \backslash \left\{\bar{b}_u\right\}\ \forall u \in \mathcal{U} \end{cases}$$

  The first constraint enforces that each user must be served by some base stations whereas the second one ensures that the base station $\bar{b}_u$ that has the strongest chan-

nel to a user u first serve that user. Only when cooperation is required, the other base stations can jointly serve user u with $\bar{b}_u$.

4.5 Distributed Multi-cell Cooperative Beamforming

In this section, we propose a decentralized algorithm for cooperative beamforming when the clustering is fixed. In fact, the centralized optimization is impractical since it requires the exchange of global channel information over the backhaul and may result in large overhead [86]. Instead, our proposed distributed algorithm requires only local channel information and some limited signaling.

4.5.1 Problem Reformulation

Denote by $\mathcal{C}$ the set of all clusters in the small-cell system. Each cluster c has $\mathcal{B}_c \subseteq \mathcal{B}$ as its set of cooperating BSs that jointly serve the users $\mathcal{U}_c \subset \mathcal{U}$. To simplify the design of a decentralized algorithm, we assume that the clusters do not overlap, i.e. we have $\mathcal{B}_c \cap \mathcal{B}_{c'} = \emptyset$ and $\mathcal{U}_c \cap \mathcal{U}_{c'} = \emptyset$ for each pair of clusters $(c, c') \in \mathcal{C}^2$. With this assumption, the received signal for user u served by a cluster c becomes:

$$y_u = \underbrace{\left(\mathbf{h}_u^c\right)^H \mathbf{v}_u^c s_u}_{\text{Desired signal}} + \underbrace{\sum_{j \in \mathcal{U}_c \setminus \{u\}} \left(\mathbf{h}_u^c\right)^H \mathbf{v}_j^c s_j}_{\text{Intra-cluster interference}} + \sum_{c' \in \mathcal{C} \setminus \{c\}} \underbrace{\sum_{k \in \mathcal{U}_{c'}} \left(\mathbf{h}_u^{c'}\right)^H \mathbf{v}_k^{c'} s_k}_{\text{Inter-cluster interference}} + n_u \tag{4.39}$$

where $\mathbf{h}_u^c \triangleq \left(\mathbf{h}_{bu}\right)_{b \in \mathcal{B}_c} \in \mathbb{C}^{m|\mathcal{B}_c| \times 1}$ and $\mathbf{v}_u^c \triangleq \left(\mathbf{v}_{bu}\right)_{b \in \mathcal{B}_c} \in \mathbb{C}^{m|\mathcal{B}_c| \times 1}$.

Our goal is to make the beamforming optimization separable between the clusters. To this end, we first transform the SINR of each user u as:

$$\text{SINR}_u = \frac{\left|\left(\mathbf{h}_u^c\right)^H \mathbf{v}_u^c\right|^2}{\sum_{j \in \mathcal{U}_c \setminus \{u\}} \left|\left(\mathbf{h}_u^c\right)^H \mathbf{v}_j^c\right|^2 + \sum_{c' \in \mathcal{C} \setminus \{c\}} z_{c'u}^2 + \sigma_u^2} \tag{4.40}$$

by replacing the inter-cluster interference term received by u from each $c' \in \mathcal{C} \setminus \{c\}$ with a slack variable $z_{c'u}^2 \geq 0$. To keep the formulation coherent, we also need the constraints:

$$\sum_{k\in\mathcal{U}_{c'}} \left| \left(\mathbf{h}_u^{c'}\right)^H \mathbf{v}_k^{c'} \right|^2 \leq z_{c'u}^2,\ \forall \left(c', u\right) \in \mathcal{C}\backslash\{c\} \times \mathcal{U}_c,\ \forall c \in \mathcal{C} \tag{4.41}$$

Note that each variable $z_{c'u}$ exactly couples a pair of clusters: user u's serving cluster c and its interfering cluster c'. Thus, the interference constraints (4.41) are equivalent to:

$$\sum_{l\in\mathcal{U}_c} \left| \left(\mathbf{h}_j^c\right)^H \mathbf{v}_l^c \right|^2 \leq z_{cj}^2,\ \forall j \in \mathcal{U}\backslash\mathcal{U}_c,\ \forall c \in \mathcal{C} \tag{4.42}$$

Despite the equivalence, having constraints (4.41) at each cluster c is not useful because cluster c cannot control the interference power $z_{c'u}^2$ received by its user $u \in \mathcal{U}$. Instead, cluster c can control the interference power z_{cj}^2 that it creates to a non-intended user $j \in \mathcal{U}\backslash\mathcal{U}_c$ through (4.42). To simplify the notation, we collect all coupling inter-cluster interference variables into $\mathbf{z} \triangleq \left(z_{cj}\right)_{j\in\mathcal{U}\backslash\mathcal{U}_c,\ c\in\mathcal{C}} \in \mathbb{R}_+^{(|\mathcal{C}|-1)|\mathcal{U}|}$.

Similarly, we create a new slack variable $\mathbf{i}_c \triangleq \left(i_{cv}\right)_{v\in\mathcal{V}} \in \mathbb{R}_+^{|\mathcal{V}|}$ that collects the interference temperature created by cluster c to each primary user. Then, we decompose the total interference temperature constraints (4.21) as follows:

$$\begin{cases} \sum_{l\in\mathcal{U}_c} \left| \left(\mathbf{g}_v^c\right)^H \mathbf{v}_l^c \right|^2 \leq i_{cv},\ \forall v \in \mathcal{V},\ \forall c \in \mathcal{C} \\ \\ \sum_{c\in\mathcal{C}} i_{cv} \leq I_v, \quad \forall v \in \mathcal{V} \end{cases} \tag{4.43}$$

As a result, the beamforming optimization problem becomes:

$$
\begin{cases}
\text{variables} & (\mathbf{v}_c)_{c\in\mathcal{C}}\,,\ \mathbf{z},\ (\mathbf{i}_c)_{c\in C} \\
\text{minimize} & \sum_{c\in\mathcal{C}}\sum_{k\in\mathcal{U}_c}\|\mathbf{v}_k^c\|^2 \\
\text{subject to} & \text{SINR}_u\left(\mathbf{v}_c,(z_{c'u})_{c'\in\mathcal{C}\setminus\{c\}}\right)\geq\rho_u,\ \forall u\in\mathcal{U}_c,\ \forall c\in\mathcal{C} \\
& \sum_{l\in\mathcal{U}_c}\left|\left(\mathbf{h}_j^c\right)^H\mathbf{v}_l^c\right|^2\leq z_{cj}^2,\ \forall j\in\mathcal{U}\backslash\mathcal{U}_c,\ \forall c\in\mathcal{C} \\
& \sum_{l\in\mathcal{U}_c}\left|\left(\mathbf{g}_v^c\right)^H\mathbf{v}_l^c\right|^2\leq i_{cv},\quad \forall v\in\mathcal{V},\ \forall c\in\mathcal{C} \\
& \sum_{l\in\mathcal{U}_c}\|\mathbf{v}_{bl}\|_2^2\leq P_b^{\max},\ \forall b\in\mathcal{B}_c,\ \forall c\in\mathcal{C} \\
& \sum_{c\in\mathcal{C}} i_{cv}\leq I_v,\quad \forall v\in\mathcal{V}
\end{cases}
\tag{4.44}
$$

Adding local interference variables and consistency constraints

In (4.44), the clusters are still coupled through the variables $\mathbf{z}$, $\{\mathbf{i}_c\}_{c\in\mathcal{C}}$ and the shared constraints $\sum_{c\in\mathcal{C}} i_{cv}\leq I_v,\ v\in\mathcal{V}$. To remove the coupling through $\mathbf{z}$, we create a local variable $\widetilde{\mathbf{z}}_c$ for each cluster c, which is defined by $\widetilde{\mathbf{z}}_c\triangleq(\widetilde{\mathbf{z}}_c^{\text{rx}},\widetilde{\mathbf{z}}_c^{\text{tx}})\in\mathbb{R}_+^{(|\mathcal{C}|-2)|\mathcal{U}_c|+|\mathcal{U}|}$. Thus, it contains only the following inter-cluster interference variables relevant to cluster c:

$$
\widetilde{\mathbf{z}}_c^{\text{rx}}\triangleq\left(\widetilde{z}_{c'u}^{\text{rx}}\right)_{(c',u)\in\mathcal{C}\setminus\{c\}\times\mathcal{U}_c}\in\mathbb{R}_+^{(|\mathcal{C}|-1)|\mathcal{U}_c|} \tag{4.45}
$$

$$
\widetilde{\mathbf{z}}_c^{\text{tx}}\triangleq\left(\widetilde{z}_{cj}^{\text{tx}}\right)_{j\in\mathcal{U}\setminus\mathcal{U}_c}\in\mathbb{R}_+^{(|\mathcal{U}|-|\mathcal{U}_c|)} \tag{4.46}
$$

The local variable $\widetilde{z}_{c'u}^{\text{rx}}$ of cluster c denotes the *desired* interference that cluster c wants its user $u\in\mathcal{U}_c$ to receive from an interfering cluster c' whereas the local variable $\widetilde{z}_{cj}^{\text{tx}}$ of cluster c denotes the interference that cluster c *desires* to transmit to a non-intended user j. To enforce a consistency between the local and coupling variables, we add the constraints:

$$
\widetilde{\mathbf{z}}_c=\mathbf{B}_c\mathbf{z},\ \forall c\in\mathcal{C} \tag{4.47}
$$

where $\mathbf{B}_c$ is a binary matrix that maps the local $\widetilde{\mathbf{z}}_c$ to the part of $\mathbf{z}$ relevant to cluster c.

Similarly, we introduce local interference temperature variable $\widetilde{\mathbf{i}}_c$ for each c with the consistency constraints:

$$\widetilde{\mathbf{i}}_c = \mathbf{i}_c, \ \forall c \in \mathcal{C} \tag{4.48}$$

After this step, the beamforming problem (4.24) is equivalent to:

$$\begin{cases} \text{variables} & \left(\mathbf{v}_c, \widetilde{\mathbf{z}}_c, \widetilde{\mathbf{i}}_c\right)_{c\in\mathcal{C}}, \ \mathbf{z}, \ (\mathbf{i}_c)_{c\in\mathcal{C}} \\ \text{minimize} & \sum_{c\in\mathcal{C}} \sum_{k\in\mathcal{U}_c} \left\|\mathbf{v}_k^c\right\|^2 \\ \text{subject to} & \left(\mathbf{v}_c, \widetilde{\mathbf{z}}_c, \widetilde{\mathbf{i}}_c\right) \in \mathcal{F}_c, \ \forall c \in \mathcal{C} \\ & \sum_{c\in\mathcal{C}} i_{cv} \leq I_v, \quad \forall v \in \mathcal{V} \\ & \widetilde{\mathbf{z}}_c = \mathbf{B}_c\mathbf{z}, \ \forall c \in \mathcal{C} \\ & \widetilde{\mathbf{i}}_c = \mathbf{i}_c, \ \forall c \in \mathcal{C} \end{cases} \tag{4.49}$$

In (4.49), the local variables $\left(\mathbf{v}_c, \widetilde{\mathbf{z}}_c, \widetilde{\mathbf{i}}_c\right)$ belongs to a convex set $\mathcal{F}_c$ defined by:

$$\mathcal{F}_c = \left\{ \left(\mathbf{v}_c, \widetilde{\mathbf{z}}_c, \widetilde{\mathbf{i}}_c\right) : \begin{array}{l} \left\| \begin{bmatrix} (\mathbf{h}_u^c)^H \mathbf{V}^c & \widetilde{\zeta}_u & \sigma_u \end{bmatrix} \right\| \leq \sqrt{1+\frac{1}{\rho_u}}\, (\mathbf{h}_u^c)^H \mathbf{v}_u^c,\ \forall u \in \mathcal{U}_c \\ \mathrm{Re}\left((\mathbf{h}_u^c)^H \mathbf{v}_u^c\right) \geq 0,\ \forall u \in \mathcal{U}_c \\ \mathrm{Im}\left((\mathbf{h}_u^c)^H \mathbf{v}_u^c\right) = 0,\ \forall u \in \mathcal{U}_c \\ \sum_{l\in\mathcal{U}_c} \left|\left(\mathbf{h}_j^c\right)^H \mathbf{v}_l^c\right|^2 \leq \widetilde{z}_{cj}^2,\ \forall j \in \mathcal{U}\backslash\mathcal{U}_c \\ \sum_{l\in\mathcal{U}_c} \left|(\mathbf{g}_v^c)^H \mathbf{v}_l^c\right|^2 \leq \widetilde{i}_{cv}, \quad \forall v \in \mathcal{V} \\ \sum_{l\in\mathcal{U}_c} \|\mathbf{v}_{bl}\|_2^2 \leq P_b^{\max},\ \forall b \in \mathcal{B}_c \end{array} \right\} \tag{4.50}$$

in which the matrix $\mathbf{V}^c \triangleq \left(\mathbf{v}_j^c\right)_{j\in\mathcal{U}_c} \in \mathbb{C}^{m|\mathcal{B}_c|\times|\mathcal{U}_c|}$ concatenates the beamforming vectors of cluster c and the vector $\widetilde{\zeta}_u \triangleq (\widetilde{z}_{c'u})_{c'\in\mathcal{C}\backslash\{c\}} \in \mathbb{R}_+^{1\times(|\mathcal{C}|-1)}$ collects the local variables of the interference received by its user u. The new formulation (4.49) is obtained after using the manipulations in (4.40)-(4.43), introducing the local variables and consistency constraints (4.47)-(4.48), and rewriting the SINR constraints in a conic form. As a result, the equivalent beamforming problem (4.49) is now convex and quasi-separable.

4.5.2 Alternating Direction Method of Multiplier (ADMM)

Next, we show how to solve (4.49) using the ADMM method. First, let us form the augmented Lagrangian $\mathcal{L}_{\text{aug}}$:

$$\mathcal{L}_{\text{aug}}\left(\left(\mathbf{v}_c, \widetilde{\mathbf{z}}_c, \widetilde{\mathbf{i}}_c\right)_{c\in\mathcal{C}}, \mathbf{z}, (\mathbf{i}_c)_{c\in\mathcal{C}}\right) = \sum_{c\in\mathcal{C}} \left(\sum_{k\in\mathcal{U}_c} \|\mathbf{v}_k^c\|^2 + p_c^{\text{IC}}\left(\mathbf{B}_c\mathbf{z}, \widetilde{\mathbf{z}}_c\right) + p_c^{\text{IT}}\left(\mathbf{i}_c, \widetilde{\mathbf{i}}_c\right) \right) \tag{4.51}$$

where the penalty functions p_b^{IC} and p_b^{IT} in (4.51) are defined by:

$$p_c^{\text{IC}}\left(\mathbf{B}_c\mathbf{z}, \widetilde{\mathbf{z}}_c\right) \triangleq \lambda_c^\top \left(\mathbf{B}_c\mathbf{z} - \widetilde{\mathbf{z}}_c\right) + \frac{\omega^{\text{IC}}}{2} \|\mathbf{B}_c\mathbf{z} - \widetilde{\mathbf{z}}_c\|_2^2 \tag{4.52}$$

$$p_c^{\text{IT}}\left(\mathbf{i}_c, \widetilde{\mathbf{i}}_c\right) \triangleq \mu_c^\top \left(\mathbf{i}_c - \widetilde{\mathbf{i}}_c\right) + \frac{\omega^{\text{IT}}}{2} \left\|\mathbf{i}_c - \widetilde{\mathbf{i}}_c\right\|_2^2 \tag{4.53}$$

The Langrangian dual variables $\lambda_c = [\lambda_c^{\text{rx}}, \lambda_c^{\text{tx}}] \in \mathbb{R}^{(|\mathcal{C}|-2)|\mathcal{U}_c|+|\mathcal{U}|}$ and $\mu_c \in \mathbb{R}^{|\mathcal{V}|}$ in (4.52) and (4.53) play the role of consistency prices, and $\omega^{\text{IC}}, \omega^{\text{IT}} \geq 0$ are fixed penalty weights for the quadratic regularization terms.

Using the ADMM method, we can solve problem (4.49) by sequentially updating the local variables $\left(\mathbf{v}_c, \widetilde{\mathbf{z}}_c, \widetilde{\mathbf{i}}_c\right)_{c\in\mathcal{C}}$, the coupling interference variables $\left(\mathbf{z}, (\mathbf{i}_c)_{c\in\mathcal{C}}\right)$ and the dual variables (λ_c, μ_c) at each iteration n with the following three steps:

(S.1) Update the local variables:

$$\left(\mathbf{v}_c^{(n+1)}, \widetilde{\mathbf{z}}_c^{(n+1)}, \widetilde{\mathbf{i}}_c^{(n+1)}\right)_{c\in\mathcal{C}} = \underset{\left(\mathbf{v}_c, \widetilde{\mathbf{z}}_c, \widetilde{\mathbf{i}}_c\right)\in\mathcal{F}_c,\ \forall c\in\mathcal{C}}{\operatorname{argmin}} \mathcal{L}_{\text{aug}}\left(\left(\mathbf{v}_c, \widetilde{\mathbf{z}}_c, \widetilde{\mathbf{i}}_c\right)_{c\in\mathcal{C}}, \mathbf{z}^{(n)}, \left(\mathbf{i}_c^{(n)}\right)_{c\in\mathcal{C}}\right) \tag{4.54}$$

(S.2) Update the coupling variables:

$$\left(\mathbf{z}^{(n+1)}, \left(\mathbf{i}_c^{(n+1)}\right)_{c\in\mathcal{C}}\right) = \underset{\mathbf{z}, (\mathbf{i}_c)_{c\in\mathcal{C}}}{\operatorname{argmin}} \ \mathcal{L}_{\text{aug}}\left(\left(\mathbf{v}_c^{(n+1)}, \widetilde{\mathbf{z}}_c^{(n+1)}, \widetilde{\mathbf{i}}_c^{(n+1)}\right)_{c\in\mathcal{C}}, \mathbf{z}, (\mathbf{i}_c)_{c\in\mathcal{C}}\right) \tag{4.55}$$

$$\text{subject to} \quad \sum_{c\in\mathcal{C}} i_{cv} \leq I_v, \quad \forall v \in \mathcal{V} \tag{4.56}$$

(S.3) Update the consistency prices:

$$\lambda_c^{(n+1)} = \lambda_c^{(n)} + \omega^{\text{IC}}\left(\mathbf{B}_c\mathbf{z}^{(n+1)} - \widetilde{\mathbf{z}}_c^{(n+1)}\right) \tag{4.57}$$

$$\mu_c^{(n+1)} = \mu_c^{(n)} + \omega^{\text{IT}}\left(\mathbf{i}_c^{(n+1)} - \widetilde{\mathbf{i}}_c^{(n+1)}\right) \tag{4.58}$$

Given the convexity of problem (4.49), the convergence properties of the ADMM algorithm are given as follows.

Proposition 4.4. *If problem* (4.49) *has a non-empty feasible set, then the ADMM-based algorithm converges as follows: (i) the total transmit power converges to the optimal value* P_T^* *as* $n \to \infty$; (ii) *a consensus will be reached between the clusters, i.e.* $\mathbf{B}_c\mathbf{z}^{(n)} - \widetilde{\mathbf{z}}_c^{(n)} \to \mathbf{0},\ \forall c$ *and* $\mathbf{i}_c^{(n)} - \widetilde{\mathbf{i}}_c^{(n)} \to \mathbf{0},\ \forall c$ *as* $n \to \infty$, *and (iii) the iterates* $\left(\lambda_c^{(n)}, \mu_c^{(n)}\right)$ *will converge to the optimal multipliers of problem* (4.49).

This result follows from the convexity of problem (4.49). Assuming that the problem instance is feasible and Slater's constraint qualification holds, then problem (4.49) satisfies the required assumptions under which ADMM is guaranteed to converge [100].

4.5.3 Decentralized Beamforming and Signaling Schemes

The previous ADMM-based algorithm can be carried out in a decentralized way by allowing the clusters to exchange some signaling information. First, the optimization problem in step **(S.1)** is completely separable. Thus, each cluster c can independently update its beamformers and local interference variables.

For the step **(S.2)** in (4.55), $\mathbf{z}$ and $(\mathbf{i}_c)_{c\in\mathcal{C}}$ can be updated separately with:

$$\mathbf{z}^{(n+1)} = \frac{1}{2}\sum_{c\in\mathcal{C}} \mathbf{B}_c^\top \left(\widetilde{\mathbf{z}}_c^{(n+1)} - \frac{1}{\omega^{\text{IC}}}\lambda_c^{(n)}\right) \tag{4.59}$$

and the following quadratic program as:

$$\left(\mathbf{i}_c^{(n+1)}\right)_{c\in\mathcal{C}} = \underset{\sum_{c\in\mathcal{C}} i_{cv}\le I_v,\ v\in\mathcal{V}}{\text{argmin}} \ \sum_{c\in\mathcal{C}} \mu_c^\top \mathbf{i}_c + \sum_{c\in\mathcal{C}} \frac{\omega^{\text{IT}}}{2}\left\|\mathbf{i}_c - \widetilde{\mathbf{i}}_c^{(n+1)}\right\|^2 \tag{4.60}$$

Next, the previous $\mathbf{z}$-update and $\mathbf{i}$-update can be carried out locally at each cluster. In fact, it can be shown that the $\mathbf{z}$-update for cluster c is done component-wise as follows:

$$z_{c'u}^{(n+1)} = \frac{1}{2}\left(\widetilde{z}_{c'u}^{\text{rx}(n+1)} - \frac{1}{\omega^{\text{IC}}}\lambda_{c'u}^{\text{rx}(n)}\right) + \frac{1}{2}\left(\widetilde{z}_{c'u}^{\text{tx}(n+1)} - \frac{1}{\omega^{\text{IC}}}\lambda_{c'u}^{\text{tx}(n)}\right),\ \forall u\in\mathcal{U}_c,\ \forall c'\in\mathcal{C}\backslash\{c\} \tag{4.61}$$

$$z_{cj}^{(n+1)} = \frac{1}{2}\left(\widetilde{z}_{cj}^{\text{tx}(n+1)} - \frac{1}{\omega^{\text{IC}}}\lambda_{cj}^{\text{tx}(n)}\right) + \frac{1}{2}\left(\widetilde{z}_{cj}^{\text{rx}(n+1)} - \frac{1}{\omega^{\text{IC}}}\lambda_{cj}^{\text{rx}(n)}\right),\ \forall j\in\mathcal{U}_{c'},\ \forall c'\in\mathcal{C}\backslash\{c\} \tag{4.62}$$

For this to be possible, each pair of clusters (c, c') needs to exchange some signaling information. Cluster c' needs to send to cluster c the real vectors $\left(\widetilde{z}_{c'u}^{\text{tx}(n+1)} - \frac{1}{\omega^{\text{IC}}}\lambda_{c'u}^{\text{tx}(n)}\right)_{u\in\mathcal{U}_c}$ and $\left(\widetilde{z}_{cj}^{\text{rx}(n+1)} - \frac{1}{\omega^{\text{IC}}}\lambda_{cj}^{\text{rx}(n)}\right)_{j\in\mathcal{U}_{c'}}$ and vice versa. Similarly, the coupling variable $(\mathbf{i}_c)_{c\in\mathcal{C}}$ can be locally updated at each cluster with (4.60) after having each pair of clusters (c, c') exchange their local variables $\left(\widetilde{\mathbf{i}}_c^{(n+1)}, \widetilde{\mathbf{i}}_{c'}^{(n+1)}\right) \in \mathbb{R}^{2|\mathcal{V}|}$.

For step **(S.3)**, each cluster c can update the consistency prices (λ_c, μ_c) independently from (4.57)-(4.58) with no extra signaling cost. As a result, the total signaling overhead

per iteration required for achieving optimality is:

$$\sum_{c\in\mathcal{C}}\sum_{c'\in\mathcal{C}\setminus\{c\}}\left(|\mathcal{U}_c|+|\mathcal{U}_{c'}|+|\mathcal{V}|\right)=2\left(|\mathcal{C}|-1\right)\left(|\mathcal{U}|+|\mathcal{V}|\right) \tag{4.63}$$

In practice, it is however desirable to limit the signaling and aim for a near-optimal performance. To do that, we first limit the number of iterations as done in [101]. Specifically, the clusters can agree to fix their local interference to the consented values $\left(\mathbf{B}_c\mathbf{z}^{(n+1)},\mathbf{i}_c^{(n+1)}\right)_{c\in\mathcal{C}}$ after step **(S.2)** of an intermediate iteration n. Then, they recompute their beamformers as:

$$\begin{cases}\underset{\mathbf{v}_c}{\text{minimize}} & \sum_{k\in\mathcal{U}_c}\|\mathbf{v}_k^c\|^2\\ \text{subject to} & \left(\mathbf{v}_c,\mathbf{B}_c\mathbf{z}^{(n+1)},\mathbf{i}_c^{(n+1)}\right)\in\mathcal{F}_c\end{cases} \tag{4.64}$$

This extra computation (4.64) enables to obtain a feasible solution at an intermediate iteration provided that the consented values $\left(\mathbf{B}_c\mathbf{z}^{(n+1)},\mathbf{i}_c^{(n+1)}\right)_{c\in\mathcal{C}}$ render the subproblems (4.64) feasible for all clusters. Otherwise, they need to carry more iterations. Our simulation results however shows that tens of iterations are enough to achieve near-optimal performance.

Second, we reduce the signaling overhead per iteration through user grouping. Based on the local channel information $\{\mathbf{h}_u^c\}_{u\in\mathcal{U}_c}$ (and/or $\{\mathbf{g}_v^c\}_{u\in\mathcal{V}}$), each cluster c can partition the small-cell users (and/or the primary users) into two groups as $\mathcal{U}_c=\mathcal{U}_c^1\cup\mathcal{U}_c^2$ (and/or $\mathcal{V}=\mathcal{V}_c^1\cup\mathcal{V}_c^2$). The users in the first group $\mathcal{U}_c^1$ are those whose desired channel gains from c are weaker compared to the users in the second group $\mathcal{U}_c^2$. On the other hand, the primary users in $\mathcal{V}_c^1$ are those whose interference channels are stronger compared to $\mathcal{V}_c^2$. Using different restriction rules, the following limited signaling schemes are proposed:

- **Fractional signaling:** Since $\mathcal{U}_c^1$ and $\mathcal{V}_c^1$ are more vulnerable to interference, the fractional signaling scheme does not impose any restriction on their interference variables. Instead, each cluster c restricts the inter-cluster interference (and/or interference temperature) to be equal for all small-cell (and/or primary) users in the second

Table 4.2 – Overhead per iteration of limited signaling schemes.

Signaling schemes	Signaling restriction on inter-cluster interference only	Signaling restriction on inter-cluster interference and interference temperature
Fractional	$\sum_{c\in\mathcal{C}}\sum_{c'\in\mathcal{C}\setminus\{c\}}\left(\left\|\mathcal{U}_c^1\right\|+\left\|\mathcal{U}_{c'}^1\right\|+\left\|\mathcal{V}\right\|+2\right)$	$\sum_{c\in\mathcal{C}}\sum_{c'\in\mathcal{C}\setminus\{c\}}\left(\left\|\mathcal{U}_c^1\right\|+\left\|\mathcal{U}_{c'}^1\right\|+\left\|\mathcal{V}_c^1\right\|+\left\|\mathcal{V}_{c'}^1\right\|+4\right)$
Group-wise	$\sum_{c\in\mathcal{C}}\sum_{c'\in\mathcal{C}\setminus\{c\}}\left(4+\left\|\mathcal{V}\right\|\right)$	$16\left(\left\|\mathcal{C}\right\|-1\right)$
Cluster-wise	$\sum_{c\in\mathcal{C}}\sum_{c'\in\mathcal{C}\setminus\{c\}}\left(2+\left\|\mathcal{V}\right\|\right)$	$8\left(\left\|\mathcal{C}\right\|-1\right)$

group $\mathcal{U}_c^2$ (or $\mathcal{V}_c^2$). This can be achieved by adding extra constraints in the optimization problem of step (S.1) and by setting the initial consistency prices to be equal for the users in the second group. With that, it can be verified by induction that for every iteration $n > 0$, the signaling information exchanged between (c, c') become equal for all small-cell users in $\mathcal{U}_c^2$ (and/or all primary users in $\mathcal{V}_c^2$), thereby reducing the overhead.

- **Group-wise signaling:** As before, each cluster c partitions its users into two groups. However, the inter-cluster interference (or interference temperature) of the small-cell (or primary) users are restricted to be equal in each of the groups.

- **Cluster-wise signaling:** Here, each cluster c does not divide the primary or small-cell users into two groups but simply enforces the interference received by all small-cell users and/or all primary users to be equal.

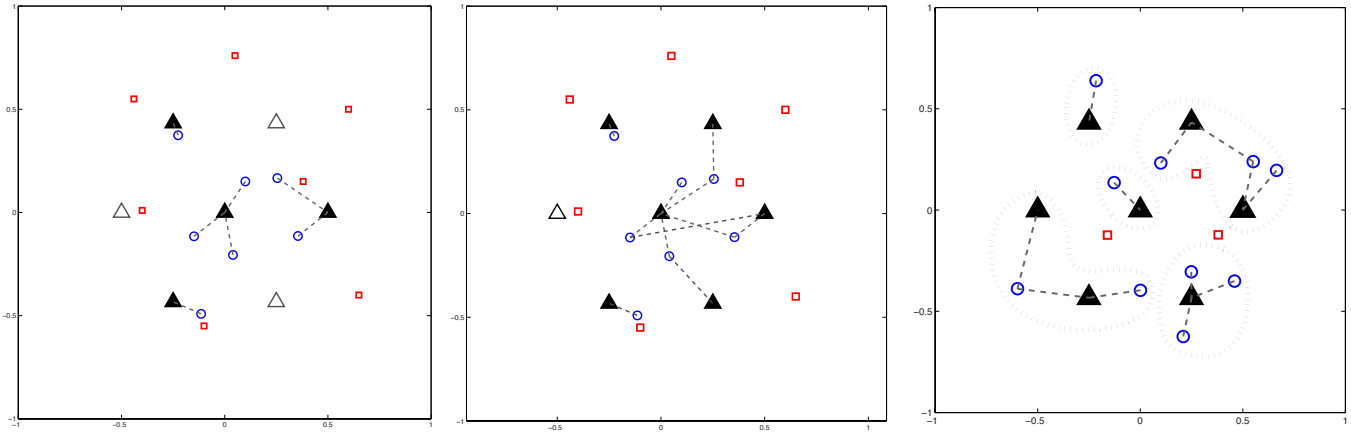
The overhead per iteration of these signaling schemes is summarized in Table 4.2. For comparison, the total overhead of the central optimization due to the exchange of global channel information is given by $\sum_{c\in\mathcal{C}}\sum_{c'\in\mathcal{C}\setminus\{c\}} 2m\cdot\left|\mathcal{B}_{c'}\right|\left(\left|\mathcal{U}\right|+\left|\mathcal{V}\right|\right)$ assuming that each complex channel coefficient is sent over the backhaul as two real scalars.

Table 4.3 – Power cost and system parameters in Chapter 4.

Simulation parameters	**Values**
Number of antenna at base stations m	3
Number of base stations	7
Inter-base station site distance	$500m$
Maximum transmit power, P_{max}	$46dBm$
System bandwidth	$5MHz$
Channel bandwidth	$200KHz$
Thermal noise power density	$-174dBm/Hz$
Power amplifier efficiency, μ_{PA}	35%
Static power cost, P_{static}	$6.8W$
Power supply battery backup, C_{PS}	11%
Dynamic signal processing power, p_{sp}	$5.8W$
Power efficiency of backhaul	$0.1W/Mbps$

4.6 Simulation Results

Next, we simulate a two-tier network with a total of seven base stations. We assume that the small cells are the secondary users and they share the spectrum with the primary macro-cell users. Then, we apply the cognitive resource allocation framework presented above to this network. As a result, we use the terms macro-cell user and primary user interchangeably. Similarly, small-cell user and secondary user are used interchangeably in this section. All these users are uniformly distributed within a circular area with a radius of 1km from the center base station. Our simulation parameters are listed in Table 4.3. The maximum base station transmit power, the cooperative processing power and fixed consumed power parameters are normalized to the subcarrier bandwidth. In our simulations, we solved the convex beamforming subproblems using the convex optimization software CVX [1].



(a) 7 small-cell users with SINR 10dB

(b) 7 small-cell users with SINR 20dB

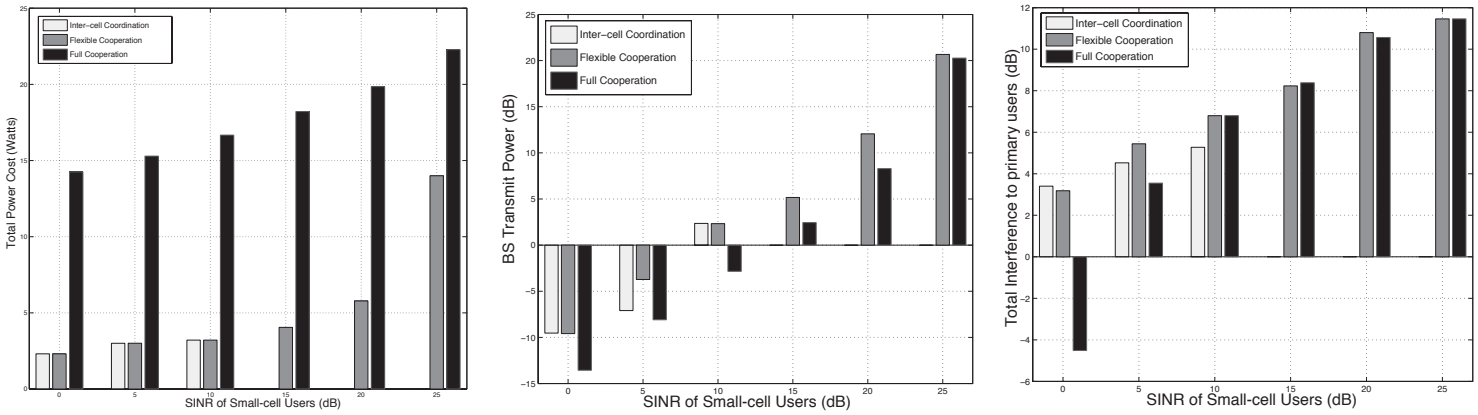
(c) 10 small-cell users with SINR 10dB

Figure 4.3 – Optimal clustering for different network topologies and SINR targets

4.6.1 Performance of Joint Clustering and Beamforming

In Figure 4.3, we illustrate the optimal clustering for two different network topologies. The primary users are denoted by squares whereas small-cell users by circles. The total interference power limit at each primary user is $I_v = 3$dB above the noise level. In Figure 4.3(a), the simple inter-cell coordination scheme is optimal when $|\mathcal{U}| = |\mathcal{V}| = 7$ and $\rho = 10$dB. However, some BS cooperation is required to support a higher SINR target $\rho = 20$dB (Figure 4.3(b)) or to serve $|\mathcal{U}| = 10$ small-cell users (Figure 4.3(c)). Thus, the flexible cooperation scheme allows the BSs to share their antennas only when they need to jointly serve some cell-edge users with a high quality of service or when their limited spatial degrees of freedom do not allow them to individually serve more small-cell users while protecting the primary users.

For the network topology in Figure 4.3(a), we compare the performance and power cost of the flexible cooperation with those of inter-cell coordination and full cooperation schemes in Figure 4.4. While the simple inter-cell coordination is optimal at low SINR, it is no longer feasible when $\rho > 10$dB. We also observe from Figures 4.4(a)-(b) that full cooperation achieves the lowest transmit power thanks to the macro-diversity gain. But since the cooperative power cost is high, significant energy can be saved by resorting to the



(a) Total power consumption

(b) Total transmit power

(c) Total interference temperature power

Figure 4.4 – Performance comparison and effect of SINR requirements of small-cell users.

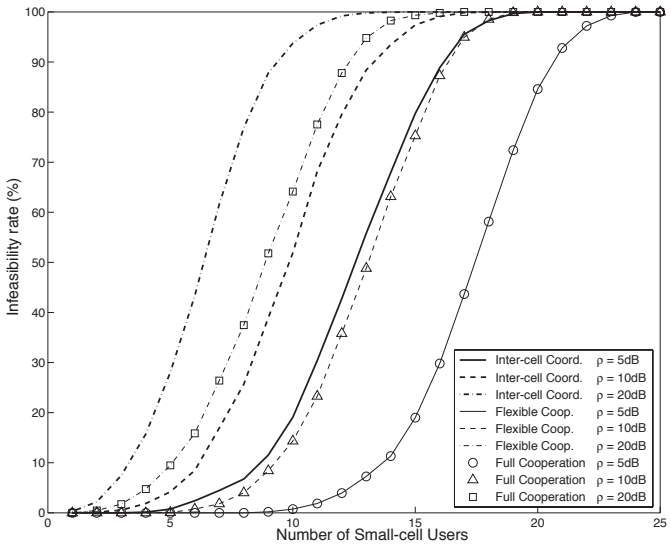
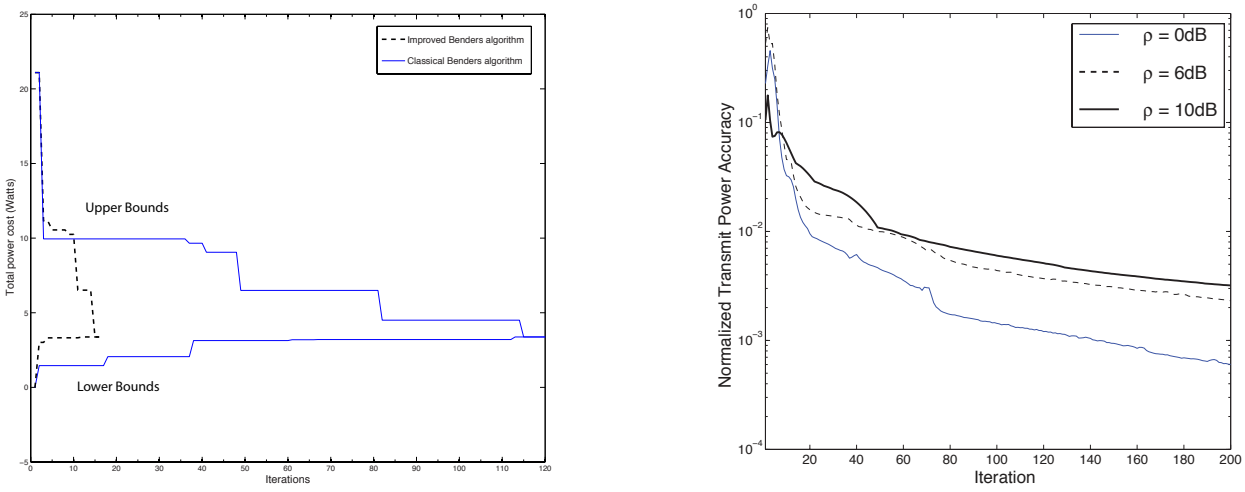


Figure 4.5 – Effect of the number of small-cell users with 10dB SINR target

flexible cooperation. Figure 4.4(c) shows that the interference to primary users increases with the SINR target until the limit is reached.

In Figure 4.5, we present the effect of the number of small-cell users on the feasibility of supporting the primary and small-cell users' requirements. Here, we have $|\mathcal{V}| = 7$ primary users. For a given number of secondary users, we generate 1000 random network instances with different user locations and channels. With $|\mathcal{U}| = 10$ and $\rho = 10$dB, only 15% infeasible cases were encountered when cooperation was used. Otherwise, 65% of the problem instances were infeasible with the simple inter-cell coordination scheme. Thus,



(a) Classical and improved Benders algorithms

(b) ADMM-based decentralized algorithm

Figure 4.6 – Convergence of proposed algorithms

the flexible cooperation scheme enhances the capabilities of the base stations to combat interference. Although it has the same infeasibility rate as the full cooperation scheme, it can also save power by reducing the degree of cooperation whenever possible.

4.6.2 Convergence of Proposed Algorithms

Next, we analyze the convergence of our proposed algorithms for scenario in Figure 4.3(c). For the joint clustering and beamforming problem, we show the convergence of the Benders algorithm when optimizing the clustering and beamforming for the scenario in Figure 4.3(c). We see that the convergence rate of the algorithm is largely improved by applying feasibility restoration and by inserting valid cuts. In fact, those two simple techniques enabled to strengthen the lower bounds and generate as many optimality cuts as possible in the early iterations. The typical convergence of the ADMM-based algorithm is shown in Figure 4.6(b). Note that the normalized power accuracy is defined as $\frac{\left|\sum_{c\in\mathcal{C}}\sum_{b\in\mathcal{B}_c} P_b - P_{\mathrm{T}}^*\right|}{P_{\mathrm{T}}^*}$.

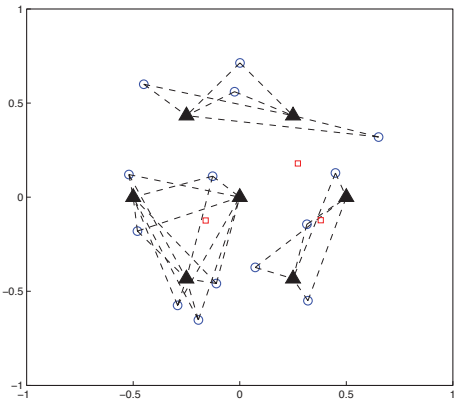
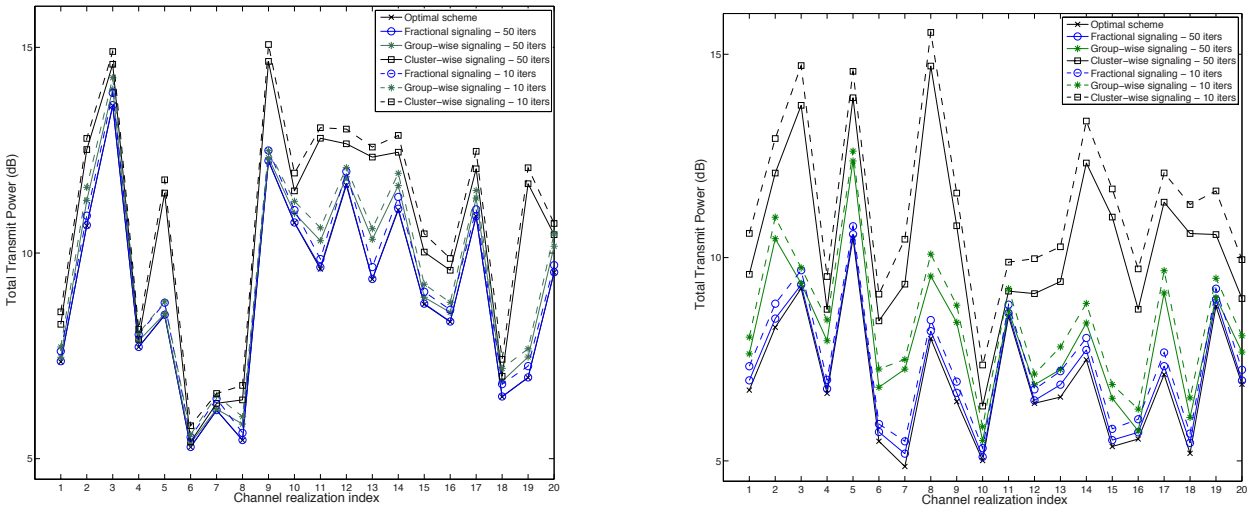


Figure 4.7 – Cooperative network with 3 clusters and $|\mathcal{U}| = 14$, $|\mathcal{V}| = 3$.



(a) $|\mathcal{V}| = 3$ and $\rho = 6$dB **(b)** $|\mathcal{V}| = 6$ and $\rho = 3$dB

Figure 4.8 – Performance comparison of limited signaling schemes with $I = 3$dB.

4.6.3 Performance of Decentralized Scheme with Limited Signaling

We consider the network in Figure 4.7 with three clusters, $|\mathcal{B}| = 7$, $|\mathcal{U}| = 14$, $\rho = 6$dB and $I_v = 3$dB. We compare the performance of the decentralized schemes with limited signaling for different channel realizations. We use the extra computation in (4.64) to obtain a feasible solution after 10 and 50 iterations. In Figure 4.8(a), we have $|\mathcal{V}| = 3$ and the signaling restriction is imposed only on the inter-cluster interference. In Figure 4.8(b), we increase the number of primary users to $|\mathcal{V}| = 6$ and add restriction on the interference temperature. Each cluster partitions the users into two groups of equal size for the fractional and group-wise signaling schemes. Figures 4.8(a) and (b) show that the fractional signaling, followed by group-wise signaling, perform near the optimum. On the other hand, the cluster-wise signaling creates a large optimality gap especially when the channels are badly conditioned. Finally, using (4.63) and Table 4.2, we found that the fractional and group-wise signaling can reduce the overhead per iteration by about 30% and 60% respectively compared to the optimal ADMM scheme while achieving near-optimal performance. When compared to the centralized optimization, the total overhead is reduced by about 67% and 80% respectively with 10 iterations.

Chapter 5

Summary, Conclusions and Future Work

In order to support the exploding mobile data traffic, wireless operators must drastically increase the network capacity through an ultra-densification of the network. A huge challenge for dense heterogeneous deployments is to keep the energy consumption at the same level as today's. In this thesis, we studied different resource allocation and cooperation schemes for enabling energy-efficient wireless heterogeneous two-tier networks. In particular, we focused on the coordinated multi-cell power allocation, energy cooperation and cooperative multi-point beamforming in such networks. Throughout this thesis, we leveraged different convex optimization techniques to solve these non-convex resource allocation problems. Our proposed algorithms are efficient, provably-convergent and applicable to various heterogeneous network deployments. In the following, we present a summary of the contributions made in this thesis along with some concluding remarks.

5.1 Summary and Conclusions

In Chapter 2, we proposed a multi-cell coordination framework for maximizing the energy efficiency of wireless heterogeneous networks. To enable the coordination between the cells, the channel state information needs to be shared through some backhaul links. Alternatively, a centralized RAN architecture can also facilitate this coordination by having a central unit controlling the power allocation of the remote radio units. When developing energy-efficient power allocation algorithms, we considered two different scenarios: orthogonal multiple access and full spectrum sharing. Even with orthogonal transmis-

sions, the resulting optimization problem is not convex and cannot be solved by standard algorithms. Nonetheless, we derived a useful convex parameterization to handle the non-convexity. Specifically, an efficient solution of the non-convex energy-efficiency maximization problem is found by solving a sequence of convex parametric problems instead. Moreover, we showed that the optimal parameters of this surrogate problems turn out to be the unique root of a non-linear system of equations. As a result, an algorithm based on a damped Newton method is derived and proved to converge to the global optimum. In addition, we also solved the general scenario of non-orthogonal transmissions, i.e. when the cells fully share the spectrum. This setup is more difficult due to the multi-user interference. Thus, another algorithm was derived using the minorization-maximization principle and it was shown to monotonically converges to at least a local optimum. These proposed optimization algorithms can be used with various power constraints such as individual transmit power, total power constraint and interference constraints. Moreover, we showed that the proposed framework can be useful even with the presence of probabilistic interference constraints. Our simulation results validated the convergence and effectiveness of the proposed algorithms. By comparing the proposed scheme with a baseline non-cooperative scheme, we found that a simple coordination between the cells can drastically improve the energy efficiency of heterogeneous networks. In fact, the achieved energy efficiency increases with the number of active small cells. These results confirm the need for multi-cell coordination to achieve an efficient operation of dense heterogeneous networks.

In Chapter 3, we studied the energy allocation problem in hybrid-powered wireless two-tier networks. The dense deployment of base stations in future networks pose a tremendous challenge to the energy cost. Therefore, it is desirable to use advanced yet sustainable energy solutions. Specifically, we assumed that each cell has a hybrid energy source and an energy storage system. The hybrid power source allows each cell to draw energy from both renewable source and the conventional power grid. With this setup, we formulated an optimization problem, in which the cells maximize the sum of their

energy efficiency by allocating their energy over a finite time horizon. Further, each cell imposes an average sum rate constraint. In contrast to Chapter 2, the new problem formulation takes into account the possible fluctuations of renewable energy arrivals at the cells. Our objective was to investigate the benefits of the hybrid power source in cellular two-tier networks. Thus, we focused on offline optimization and assumed a non-causal knowledge of the channel conditions and renewable energy arrivals. The algorithms proposed in Chapter 2 were extended to handle the non-convex average cell rate constraints. Whenever the new algorithm starts with a feasible power allocation, the subsequent iterates necessarily satisfy the rate constraints and monotonically converges to at least a local optimum. Furthermore, we studied the joint energy allocation and energy cooperation in wireless two-tier networks. Precisely, we extended the energy-saving capability of the network by allowing the cells to exchange their harvested energy through a smart-grid power infrastructure. In this case, we propose to exploit both the spatial and time diversity of the renewable energy arrivals across the cells. Our numerical results showed that both the hybrid power source and the energy cooperation scheme can significantly improve the energy efficiency of cellular two-tier networks. A much higher gain can be obtained when they are used together. Hence, energy cooperation is a viable approach for reducing both the operational and environmental costs of future heterogeneous networks. Nonetheless, there are still many practical issues that need further investigation to fully realize such possibility. Some of these open problems are discussed in the next section.

In Chapter 4, we investigated the joint optimization of the clustering and cooperative beamforming for minimizing the total energy cost in clustered two-tier systems. In contrast to Chapter 2 and Chapter 3, we assumed here that the base stations can have multiple antennas. In addition, they cooperate by sharing their antennas and by jointly transmitting to their users. To achieve this form of cooperation, the cells must be connected by backhaul links through which they exchange the user data and/or signaling information. The advantages of using this coordinated multi-point transmission scheme are the robustness

against interference and the spatial multiplexing gain. However, this cooperation scheme may incur some significant energy cost due to the multi-cell processing and the backhaul signaling. Therefore, we propose a flexible cooperation scheme, in which the cooperative cluster for each user is optimized to adapt the degree of cooperation of the base stations. In fact, most existing works in the literature have overlooked this trade-off. Instead, we explicitly modeled this optimization problem as a mixed-integer program and solved it using the generalized Benders decomposition method. Precisely, we divided the problem into a beamforming subproblem and a clustering master problem. The proposed Benders algorithm optimally converges in a finite number of iterations but it can exhibit a slow convergence. Therefore, we derived simple techniques to accelerate its convergence. We also proposed a decentralized cooperative beamforming algorithm when the clusters are fixed. In fact, the use of distributed algorithms is of practical importance when the cells are not connected through a central unit. With this distributed coordination, the clusters were able to manage their interference through consensus. Given the limited capacity of the backhaul links, it is also important to limit the signaling overhead. By employing some limited signaling schemes, it is possible to reduce the total overhead by about 30% and 60% respectively at the cost of a marginal increase in total energy consumption. More importantly, we showed that it is possible to achieve an efficient trade-off between energy cost and performance when employing a coordinated multi-point transmission in heterogeneous networks.

5.2 Future Research Directions

As a result of the work done in this thesis, we present some possible directions of future research as described in the following sections.

Distributed algorithms for energy-efficient device-to-device networks

The optimization framework presented in Chapter 2 is suitable for heterogeneous networks that have a centralized architecture. An example of this is a cloud-RAN based networks in which the cells are served by remote radio units that are connected to a central unit through optical fronthaul links. Alternatively, the centralized energy-efficiency optimization can also be performed for distributed cells under the condition that they can exchange their channel information through some backhaul links. Nonetheless, it is preferable if not necessary in certain scenarios to employ distributed algorithms instead of a centralized one. For instance, device-to-device communication systems do not have a centralized unit nor backhaul links that connect the devices. In such case, they need to coordinate their resource allocation in a self-organized manner. Clearly, distributed algorithms will need to incur some signaling overhead to enable the devices or cells to achieve good performance. When designing distributed algorithms, an important criteria is to achieve a good trade-off between signaling overhead and performance. For this purpose, the centralized algorithm proposed in Chapter 2 can serve as a benchmark. Moreover, the computational and run-time complexities of the distributed algorithms should be at least comparable to that of a centralized one. This should not be too hard if a parallel update method of the Jacobi type is employed [91]. Nonetheless, it is a challenging yet interesting problem to derive parallel, convergent and efficient algorithms for the non-convex energy-efficiency maximization. To our knowledge, only the algorithm in [56], which converges to a Nash equilibrium, has attempted such parallel updates for this problem. Nonetheless, we showed in Chapter 2 of this thesis that the achieved Nash equilibrium is not efficient for this problem. In addition, the distributed algorithm should be able to handle individual as well as shared power constraints. Finally, another technical challenge is to analyze the convergence of the designed distributed algorithm under asynchronous updates. This is an important practical issue for device-to-device networks given the lack of centralized coordination.

Stochastic energy-efficiency maximization with imperfect channel information

In Chapter 2, we addressed the possibility to have imperfect channel information between the cells and the co-existing primary users in the system. For this, we proposed to use probabilistic constraints and handle the tractability issue through robust approximation methods. However, it is also possible that the channel state information between the cells themselves are uncertain. In this case, the uncertainty is present in the utility function and our new goal is to maximize the expected energy efficiency. However, even if the exact probability distribution of the channels is known, it is intractable to obtain a closed-form expression of such expected utility function. In this case, it is more practical to use a stochastic approximation approach [102]. In this line, a method was recently proposed in [103] for stochastic non-convex rate maximization problems. Based on partial linearization [104] and a gradient averaging approach [105], the algorithm consists of solving a sequence of simpler surrogate subproblems. It was shown to converge to a stationary point [103]. However, the approximate objective function of the subproblems need to be strongly convex. Unfortunately, it is not easy to find such approximation for the non-convex energy-efficiency function. Nonetheless, the application of stochastic approximation and averaging methods is a promising direction for solving the stochastic energy-efficiency maximization problem.

Online adaptive optimization for joint energy allocation and energy cooperation

Another interesting open problem is to design online algorithms for the joint energy allocation and energy cooperation problem. In practice, only information about past channel conditions as well as past/current energy arrivals are available at each cell. Therefore, a new framework is needed to solve this optimization problem in an adaptive manner. For-

tunately, there is a growing interest for online algorithms not only for communications but also for different applications such as smart-grid and machine learning on data streams. To our knowledge, some of the frameworks that were previously used to design adaptive algorithms are dynamic programming [106], Lyapunov optimization [107] and online convex optimization [108]. The traditional approach based on dynamic programming is difficult to apply here because the algorithm may suffer from the curse of dimensionality [106]. On the other hand, the algorithms derived from online convex optimization framework (e.g. mirror descent [108]) first need to be extended to handle the non-convexity of the energy efficiency function. Moreover, it should be able to satisfy both the time-varying energy constraints and the average rate constraints. Therefore, further research efforts will be needed to tackle this interesting problem.

Energy-efficient load balancing and adaptive clustering for cooperative small-cell networks

With regard to Chapter 4, we assumed that each user has a fixed quality of service requirement. This is reasonable if the amount of data traffic created by the users is slowly varying and predictable. A more applications will be moved to the mobile cloud, future wireless networks will have to manage a fast-changing and highly volatile data traffic across the network. Cooperation between the small-cells would be needed in order to balance the data load and achieve a satisfactory level of service for the users. Hence, the joint clustering and resource allocation problem, studied in Chapter 4, can be extended to take into account the load balancing of the cells. A new model and a new utility function will be needed to capture the trade-off between energy cost and service delay.

Bibliography

[1] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*, `http://cvxr.com/cvx`, Mar. 2014.

[2] S. Ginn, "Personal communication services: Expanding the freedom to communicate", *IEEE Communications Magazine*, vol. 29, no. 2, pp. 30–32, 1991.

[3] V. Donald, "Advanced mobile phone service: The cellular concept", *The Bell System Technical Journal*, vol. 58, no. 1, pp. 15–41, 1979.

[4] Q. Li, H. Niu, A. Papathanassiou, and G. Wu, "5G Network Capacity: Key Elements and Technologies", *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 71–78, 2014.

[5] P. Vlacheas, R. Giaffreda, V. Stavroulaki, D. Kelaidonis, V. Foteinos, G. Poulios, P. Demestichas, A. Somov, A. Biswas, and K. Moessner, "Enabling smart cities through a cognitive management framework for the Internet of things", *IEEE Communications Magazine*, vol. 51, no. 6, pp. 102–111, 2013.

[6] J. Zander and P. Mähönen, "Riding the data tsunami in the cloud: Myths and challenges in future wireless access", *IEEE Communications Magazine*, vol. 51, no. 3, pp. 145–151, 2013.

[7] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A Lozano, A. Soong, and J. Zhang, "What will 5G be?", *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[8] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges", *IEEE Communications Surveys and Tutorials*, vol. 13, no. 4, pp. 524–540, 2011.

[9] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter", *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2946–2957, Aug. 2012.

[10] D. W. K. Ng, E. S. Lo, and R Schober, "Energy-efficient resource allocation in OFDMA systems with hybrid energy harvesting base station", *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3412–3427, 2013.

[11] A. Fehske, P. Marsch, and G. Fettweis, "Bit per Joule efficiency of cooperating base stations in cellular networks", in *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, 2010, pp. 1406–1411.

[12] V. Mancuso and S. Alouf, "Reducing costs and pollution in cellular networks", *IEEE Communications Magazine*, vol. 49, no. 8, pp. 63–71, 2011.

[13] I. Ahmed, A. Ikhlef, D. W. K. Ng, and R. Schober, "Power allocation for an energy harvesting transmitter with hybrid energy sources", *IEEE Transactions on Wireless Communications*, vol. 12, no. 12, pp. 6255–6267, 2013.

[14] J. Gong, S. Zhou, and Z. Niu, "Optimal power allocation for energy harvesting and power grid coexisting wireless communication systems", *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 3040–3049, 2013.

[15] X. Kang, Y.-K. Chia, C. K. Ho, and S. Sun, "Cost minimization for fading channels with energy harvesting and conventional energy", *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4586–4598, 2014.

[16] M. Erol-Kantarci and H. Mouftah, "Energy-efficient information and communication infrastructures in the smart grid: A survey on interactions and open issues", *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 179–197, 2015.

[17] C. Yang, S. Han, X. Hou, and A. Molisch, "How do we design CoMP to achieve its promised potential?", *IEEE Wireless Communications Magazine*, vol. 20, no. 1, pp. 67–74, 2013.

[18] P. Marsch and G. Fettweis, "Uplink CoMP under a constrained backhaul and imperfect channel knowledge", *IEEE Transactions on Wireless Communications*, vol. 10, no. 6, pp. 1730–1742, 2011.

[19] Q. Zhang, C. Yang, and A. Molisch, "Cooperative downlink transmission mode selection under limited-capacity backhaul", in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, 2012, pp. 1082–1087.

[20] M. Grieger, P. Marsch, Z. Rong, and G. Fettweis, "Field trial results for a coordinated multi-point (CoMP) uplink in cellular systems", in *Smart Antennas (WSA), 2010 International ITG Workshop on*, 2010, pp. 46–51.

[21] S. Kaviani and W. Krzymien, "Multicell scheduling in network MIMO", in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, 2010, pp. 1–5.

[22] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming intercell interference in indoor wireless systems", in *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, 2007, pp. 83–87.

[23] K. Hosseini, W. Yu, and R. Adve, "Large-scale MIMO versus network MIMO for multicell interference mitigation", *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 930–941, 2014.

[24] M. Peng, Y. Li, J. Jiang, J. Li, and C. Wang, "Heterogeneous cloud radio access networks: A new perspective for enhancing spectral and energy efficiencies", *IEEE Wireless Communications*, vol. 21, no. 6, pp. 126–135, 2014.

[25] G. K. Tran, S. Tajima, R. Ramamonjison, K. Sakaguchi, K. Araki, S. Kaneko, N. Miyazaki, S. Konishi, and Y. Kishi, "Study on resource optimization for heterogeneous networks", *IEICE Transactions*, vol. 95-B, no. 4, pp. 1198–1207, 2012.

[26] M. Coldrey, U. Engstrom, K. W. Helmersson, M. Hashemi, L. Manholm, and P. Wallentin, "Wireless backhaul in future heterogeneous networks", *Ericsson Review*, pp. 2–10, Oct. 2014.

[27] D. Wubben, P. Rost, J. Bartelt, M. Lalam, V. Savin, M. Gorgoglione, A. Dekorsy, and G. Fettweis, "Benefits and impact of cloud computing on 5G signal processing: Flexible centralization through cloud-RAN", *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 35–44, 2014.

[28] P. Frenger, C. Friberg, Y. Jading, M. Olssom, and O. Persson, "Radio network energy performance: shifting focus from power to precision", *Ericsson Review*, pp. 2–9, Feb. 2014.

[29] C. Shannon, "A mathematical theory of communication", *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[30] V. Rodoplu and T. Meng, "Bits-per-Joule capacity of energy-limited wireless networks", *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 857–865, 2007.

[31] S. Mao, M. H. Cheung, and V. Wong, "Joint energy allocation for sensing and transmission in rechargeable wireless sensor networks", *IEEE Transactions on Vehicular Technology*, vol. 63, no. 6, pp. 2862–2875, 2014.

[32] H. Zhang, S. Chen, X. Li, H. Ji, and X. Du, "Interference management for heterogeneous networks with spectral efficiency improvement", *IEEE Wireless Communications*, vol. 22, no. 2, pp. 101–107, 2015.

[33] A. Liu, V. Lau, L. Ruan, J. Chen, and D. Xiao, "Hierarchical radio resource optimization for heterogeneous networks with enhanced inter-cell interference coordination (eICIC)", *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1684–1693, 2014.

[34] D. Lopez-Perez, X. Chu, and I. Guvenc, "On the expanded region of picocells in heterogeneous networks", *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 3, pp. 281–294, 2012.

[35] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks", *IEEE Wireless Communications*, vol. 18, no. 3, pp. 22–30, 2011.

[36] P. Marsch and G. Fettweis, "A decentralized optimization approach to backhaul-constrained distributed antenna systems", in *Mobile and Wireless Communications Summit, 2007. 16th IST*, 2007, pp. 1–5.

[37] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results", *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, 2011.

[38] A. Davydov, G. Morozov, I. Bolotin, and A. Papathanassiou, "Evaluation of joint transmission CoMP in C-RAN based LTE-A HetNets with large coordination areas", in *Globecom Workshops (GC Wkshps), 2013 IEEE*, 2013, pp. 801–806.

[39] D Niyato, X. Lu, and P. Wang, "Adaptive power management for wireless base stations in a smart grid environment", *IEEE Wireless Communications*, vol. 19, no. 6, pp. 44–51, 2012.

[40] P. Diamantoulakis, A. Ghassemi, and G. Karagiannidis, "Smart hybrid power system for base transceiver stations with real-time energy management", in *Global Communications Conference (GLOBECOM), 2013 IEEE*, 2013, pp. 2773–2778.

[41] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart Grid - the new and improved power grid: A survey", *IEEE Communications Surveys and Tutorials*, vol. 14, no. 4, pp. 944–980, 2012.

[42] J. Xu, L. Duan, and R. Zhang, "Cost-aware green cellular networks with energy and communication cooperation", *IEEE Communications Magazine*, vol. 53, no. 5, pp. 257–263, 2015.

[43] G. Miao, N Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels", *IEEE Transactions on Communications*, vol. 58, no. 2, pp. 545–554, 2010.

[44] W. Dinkelbach, "On nonlinear fractional programming", *Management Science*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[45] X. Chen, X. Xu, and X. Tao, "Energy efficient power allocation in generalized distributed antenna system", *IEEE Communications Letters*, vol. 16, no. 7, pp. 1022–1025, Jul. 2012.

[46] M Naeem, K Illanko, A Karmokar, A Anpalagan, and M Jaseemuddin, "Iterative power allocation for downlink green cognitive radio network", *Globecom Workshops (GC Wkshps), 2012 IEEE*, pp. 163–167, 2012.

[47] G. Miao, N. Himayat, G. Li, and S. Talwar, "Low-complexity energy-efficient scheduling for uplink ofdma", *IEEE Transactions on Communications*, vol. 60, no. 1, pp. 112–120, 2012.

[48] C. Xiong, G. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in OFDMA networks", *IEEE Transactions on Communications*, vol. 60, no. 12, pp. 3767–3778, 2012.

[49] D. Ng, E. Lo, and R. Schober, "Energy-efficient power allocation in OFDM systems with wireless information and power transfer", in *IEEE International Conference on Communications (ICC), 2013*, 2013, pp. 4125–4130.

[50] K. Lee and J. Hong, "Energy efficient resource allocation for simultaneous information and energy transfer with imperfect channel estimation", *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.

[51] R. Loodaricheh, S. Mallick, and V. Bhargava, "Energy-efficient resource allocation for OFDMA cellular networks with user cooperation and qos provisioning", *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6132–6146, 2014.

[52] G. Lim and J. Cimini L.J., "Energy-efficient cooperative beamforming in clustered wireless networks", *IEEE Transactions on Wireless Communications*, vol. 12, no. 3, pp. 1376–1385, 2013.

[53] X. Wang, F. Zheng, P. Zhu, and X. You, "Energy-efficient resource allocation in coordinated downlink multi-cell OFDMA systems", *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.

[54] L. Venturino, C. Risi, S. Buzzi, and A. Zappone, "Energy-efficient coordinated user scheduling and power control in downlink multi-cell OFDMA networks", in *IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), 2013*, 2013, pp. 1655–1659.

[55] G. K. Tran, S. Tajima, R. Ramamonjison, K. Sakaguchi, K. Araki, S. Kaneko, N. Miyazaki, S. Konishi, and Y. Kishi, "Study on resource optimization for heterogeneous networks", *IEICE Transactions*, vol. 95-B, no. 4, pp. 1198–1207, 2012.

[56] G. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Distributed interference-aware energy-efficient power optimization", *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1323–1333, Apr. 2011.

[57] L. Greenstein, S. Ghassemzadeh, V. Erceg, and D. Michelson, "Ricean k-factors in narrow-band fixed wireless channels: theory, experiments, and statistical models", *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 4000–4012, 2009.

[58] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks", in *Future Network and Mobile Summit*, 2010, pp. 1–8.

[59] Y. S. Soh, T. Quek, M. Kountouris, and G. Caire, "Cognitive hybrid division duplex for two-tier femtocell networks", *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 4852–4865, 2013.

[60] B. Ma, M. Cheung, V. Wong, and J. Huang, "Hybrid overlay/underlay cognitive femtocell networks: a game theoretic approach", *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3259–3270, 2015.

[61] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[62] C. T. Kelley, *Solving nonlinear equations with Newton's method*, ser. Fundamentals of Algorithms. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2003.

[63] ——, *Iterative methods for linear and nonlinear Equations*, ser. Frontiers in Applied Mathematics 16. SIAM, 1995.

[64] D. R. Hunter and K. Lange, "A tutorial on MM algorithms", *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[65] N. Vucic, S. Shi, and M. Schubert, "DC programming approach for resource allocation in wireless networks", in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2010 Proceedings of the 8th International Symposium on*, 2010, pp. 380–386.

[66] Q. Li, Q. Zhang, and J. Qin, "Secure relay beamforming for simultaneous wireless information and power transfer in non-regenerative relay networks", *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, pp. 2462–2467, 2014.

[67] M. Gholami, S. Gezici, and E. Ström, "A concave-convex procedure for TDOA based positioning", *IEEE Communications Letters*, vol. 17, no. 4, pp. 765–768, 2013.

[68] A. L. Yuille and A. Rangarajan, "The concave-convex procedure", *Neural Comput.*, vol. 15, no. 4, pp. 915–936, Apr. 2003.

[69] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd. Upper Saddle River, USA: Prentice Hall, 2001.

[70] NGMN RAN-EV D1 Document, "C-ran fronthaul requirements", 2015.

[71] NGMN P-CRAN D1 Document, "An analysis of ran cost-structure", 2012.

[72] NGMN P-CRAN D2 Document, "General requirements for c-ran", 2012.

[73] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust optimization*, ser. Princeton Series in Applied Mathematics. Princeton University Press, 2009.

[74] P Kyösti, J. Meinilä, L. Hentilä, X. Zhao, T. Jämsä, C. Schneider, M. Narandzić, M. Milojević, A. Hong, J. Ylitalo, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. de Jong, and T. Rautiainen, "WINNER II channel models", Tech. Rep., 2007.

[75] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey", *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 996–1019, 2013.

[76] J. N. Portela and M. Alencar, "Cellular network as a multiplicatively weighted voronoi diagram", in *3rd IEEE Consumer Communications and Networking Conference, 2006. CCNC 2006.*, vol. 2, 2006, pp. 913–917.

[77] R. Ramamonjison, G. K. Tran, K. Sakaguchi, K. Araki, S. Kaneko, Y. Kishi, and N. Miyazaki, "Spectrum allocation strategies for heterogeneous networks", in *6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, IEEE, May 2012.

[78] Y.-K. Chia, S. Sun, and R. Zhang, "Energy cooperation in cellular networks with renewable powered base Stations", *IEEE Transactions on Wireless Communications*, vol. 13, no. 12, pp. 6996–7010, 2014.

[79] Y. Guo, J. Xu, L. Duan, and R. Zhang, "Joint energy and spectrum cooperation for cellular communication systems", *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3678–3691, 2014.

[80] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation", *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1650–1662, 2005.

[81] R. Ramamonjison and V. Bhargava, "Energy efficiency maximization framework in cognitive downlink two-tier networks", *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1468–1479, 2015.

[82] O Ozel, K Tutuncuoglu, J. Yang, S. Ulukus, and A Yener, "Transmission with energy harvesting nodes in fading wireless channels", *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1732–1743, 2011.

[83] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference", *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, 2010.

[84] S. Shamai and B. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end", in *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*, vol. 3, 2001, pp. 1745–1749.

[85] R. Zhang and Y.-C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks", *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 88–102, Feb. 2008.

[86] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results", *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[87] G. Scutari and D. Palomar, "MIMO cognitive radio: A game theoretical approach", *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 761–780, 2010.

[88] K. Cumanan, R. Krishna, L. Musavian, and S. Lambotharan, "Joint Beamforming and User Maximization Techniques for Cognitive Radio Networks Based on Branch and Bound Method", *IEEE Transactions on Wireless Communications*, vol. 9, no. 10, pp. 3082–3092, 2010.

[89] M.-L. Ku, L.-C. Wang, and Y.-T. Su, "Toward optimal multiuser antenna beamforming for hierarchical cognitive radio systems", *IEEE Transactions on Communications*, vol. 60, no. 10, pp. 2872–2885, 2012.

[90] A. Geoffrion, "Generalized Benders decomposition", *Journal of Optimization Theory and Applications*, vol. 10, no. 4, pp. 237–260, 1972.

[91] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1989.

[92] P. Marsch and G. Fettweis, "Static clustering for cooperative multi-point (CoMP) in mobile communications", in *ICC 2011 - 2011 IEEE International Conference on Communications*, IEEE, pp. 1–6.

[93] F. Heliot, M. Imran, and R. Tafazolli, "Energy efficiency analysis of idealized coordinated multi-point communication system", in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, 2011, pp. 1–5.

[94] C. He, B. Sheng, P. Zhu, X. You, and G. Li, "Energy- and spectral-efficiency tradeoff for distributed antenna systems with proportional fairness", *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 5, pp. 894–902, 2013.

[95] Y. Huang, J. Xu, and L. Qiu, "Energy efficient coordinated beamforming for multi-cell MISO systems", *arXiv.org*, Jul. 2013. arXiv:1307.2421v1 [cs.IT].

[96] R. Ramamonjison, A. Haghnegahdar, and V. Bhargava, "Joint optimization of clustering and cooperative beamforming in green cognitive wireless networks", *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 982–997, 2014.

[97] K. Huq, S. Mumtaz, J. Bachmatiuk, J. Rodriguez, X. Wang, and R. Aguiar, "Green hetnet comp: energy efficiency analysis and optimization", *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2014.

[98] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency", *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.

[99] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms: Part II: Advanced Theory and Bundle Methods*. Berlin: Springer, 1993.

[100] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers", *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[101] A. Tolli, H. Pennanen, and P. Komulainen, "Decentralized minimum power multi-cell beamforming with limited backhaul signaling", *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 570–580, 2011.

[102] H. Robbins and S. Monro, "A stochastic approximation method", *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, Sep. 1951.

[103] Y. Yang, G. Scutari, and D. Palomar, "Parallel stochastic decomposition algorithms for multi-agent systems", in *Signal Processing Advances in Wireless Communications (SPAWC), 2013 IEEE 14th Workshop on*, 2013, pp. 180–184.

[104] M. Patriksson, "Partial linearization methods in nonlinear programming", *J. Optim. Theory Appl.*, vol. 78, no. 2, pp. 227–246, Aug. 1993.

[105] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging", *SIAM J. Control Optim.*, vol. 30, no. 4, pp. 838–855, Jul. 1992.

[106] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd. Athena Scientific, 2000.

[107] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan and Claypool Publishers, 2010.

[108] S. Shalev-Shwartz, "Online learning and online convex optimization", *Found. Trends Mach. Learn.*, vol. 4, no. 2, pp. 107–194, Feb. 2012.

[109] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, 2003.

[110] A. Ben-Tal and A. S. Nemirovski, *Lectures on modern convex optimization: Analysis, algorithms, and engineering applications*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2001.

[111] W. Hager, "Lipschitz continuity for constrained processes", *SIAM Journal on Control and Optimization*, vol. 17, no. 3, pp. 321–338, 1979.

[112] K. Malanowski, "Higher order sensitivity of solutions to convex programming problems without strict complementarity", in *System modelling and optimization*, ser. Lecture Notes in Control and Information Sciences, vol. 113, Springer, 1988, pp. 148–164.

[113] K. Eriksson, D. Estep, and C. Johnson, "Applied mathematics: Body and soul", *Springer-Verlag Publishing*, vol. I, 2004.

[114] D. Bertsekas, *Convex Optimization Theory*, ser. Athena Scientific optimization and computation series. Athena Scientific, 2009.

[115] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers", *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 161–176, 2006.

Appendices

Appendix A

Proof of Results in Chapter 2

A.1 Proof of Proposition 2.1

To prove Proposition 2.1, let us define the following function:

$$L\left(\theta, \mathbf{p}; \alpha, \mathbf{w}, \lambda\right) = \gamma \sum_{c \in \mathcal{C}} \theta_c + \sum_{c \in \mathcal{C}} \alpha_c \left(r_c\left(\mathbf{p}_c\right) - \theta_c C_c\left(\mathbf{p}_c\right)\right) + \sum_{l \in \mathcal{V}} \lambda_l h_l\left(\mathbf{p}\right), \tag{A.1}$$

which is parametrized by $(\gamma, \alpha, \lambda)$, with $\lambda \triangleq (\lambda_l)_{l \in \mathcal{V}}$.

Proof. Suppose that $\left(\widehat{\mathbf{p}}, \widehat{\theta}\right)$ is an optimal solution of the original problem (P) formulated in (2.10). Then, using the Fritz-John optimality conditions [109], there exists a triple $\left(\overline{\gamma}, \overline{\alpha}, \overline{\lambda}\right)$ that satisfies:

$$\frac{\partial L}{\partial \mathbf{p}} = \sum_{c \in \mathcal{C}} \overline{\alpha}_c \left(\nabla_{\mathbf{p}} R_c\left(\widehat{\mathbf{p}}_c\right) - \widehat{\theta}_c \nabla_{\mathbf{p}} C_c\left(\widehat{\mathbf{p}}_c\right)\right) + \sum_{l \in \mathcal{V}} \overline{\lambda}_l \nabla_{\mathbf{p}} h_l\left(\widehat{\mathbf{p}}\right) = \mathbf{0}, \tag{A.2}$$

$$\frac{\partial L}{\partial \theta_i} = \overline{\gamma} - \overline{\alpha}_c C_c\left(\widehat{\mathbf{p}}_c\right) = 0, \ \forall c \in \mathcal{C}, \tag{A.3}$$

$$\overline{\alpha}_c \left(R_c\left(\widehat{\mathbf{p}}_i\right) - \widehat{\theta}_c C_c\left(\widehat{\mathbf{p}}_c\right)\right) = 0, \ \forall c \in \mathcal{C}, \tag{A.4}$$

$$\overline{\lambda}_l h_l\left(\widehat{\mathbf{p}}\right) = 0, \ \forall l \in \mathcal{V}, \tag{A.5}$$

$$R_c\left(\widehat{\mathbf{p}}_c\right) - \widehat{\theta}_c C_c\left(\widehat{\mathbf{p}}_c\right) \leq 0, \ \forall c \in \mathcal{C}, \tag{A.6}$$

$$h_l\left(\widehat{\mathbf{p}}\right) \leq 0, \ \forall l \in \mathcal{V}, \tag{A.7}$$

$$\overline{\gamma} \geq 0, \overline{\alpha} \geq \mathbf{0}, \overline{\lambda} \geq \mathbf{0}, \tag{A.8}$$

$$\left(\overline{\gamma}, \overline{\alpha}, \overline{\lambda}\right) \neq \left(0, \mathbf{0}, \mathbf{0}\right). \tag{A.9}$$

Next, we show by contraposition that $\overline{\gamma} \neq 0$ necessarily holds. Suppose that $\overline{\gamma} = 0$. Hence, we obtain from conditions (A.3) that $\overline{\alpha}_c = 0,\ \forall c \in \mathcal{U}$ since $C_c\left(\widehat{\mathbf{p}}_c\right) > 0,\ \forall c \in \mathcal{C}$. Due to the complementary slackness conditions in (A.5), we also have $\overline{\lambda}_l = 0$ if $l \notin \mathcal{V}_a$. Here, $\mathcal{V}_a = \{l \in \mathcal{V} \mid h_l\left(\widehat{\mathbf{p}}\right) = 0\}$ is the subset of primary users for which the interference constraints are active. Thus, (A.2) implies:

$$\sum_{l \in \mathcal{V}_a} \overline{\lambda}_l \nabla_{\mathbf{p}} h_l\left(\widehat{\mathbf{p}}\right) = 0. \tag{A.10}$$

Furthermore, if $\overline{\gamma} = 0$ and $\overline{\alpha}_c = 0,\ \forall c \in \mathcal{C}$, then the condition (A.9) implies that

$$\sum_{l \in \mathcal{V}_a} \overline{\lambda}_l > 0. \tag{A.11}$$

Since the power constraint functions $h_l,\ \forall l \in \mathcal{V}$ are affine, we can assume that $\nabla_{\mathbf{p}} h_l\left(\mathbf{p}\right)$ is a positive vector. Hence, we obtain from (A.11) that $\sum_{l \in \mathcal{V}_a} \overline{\lambda}_l \nabla_{\mathbf{p}} h_l\left(\widehat{\mathbf{p}}\right) > 0$, which contradicts equation (A.10). Consequently, $\overline{\gamma} = 0$ cannot hold.

Since $\overline{\gamma} \neq 0$, we can now redefine the parameters as $\widehat{\mathbf{w}} = \frac{\overline{\mathbf{w}}}{\overline{\gamma}}$ and $\widehat{\lambda} = \frac{\overline{\lambda}}{\overline{\gamma}}$ and divide equations (A.2)-(A.5) by $\overline{\gamma}$ to obtain the equivalent equations (A.12)-(A.15):

$$\sum_{c \in \mathcal{C}} \widehat{\alpha}_c \left(\nabla_{\mathbf{p}} R_c\left(\widehat{\mathbf{p}}_c\right) - \widehat{\theta}_c \nabla_{\mathbf{p}} C_c\left(\widehat{\mathbf{p}}_c\right)\right) + \sum_{l=1}^{L} \widehat{\lambda}_l \nabla_{\mathbf{p}} h_l\left(\widehat{\mathbf{p}}\right) = \mathbf{0}, \tag{A.12}$$

$$1 - \widehat{\alpha}_c C_c\left(\widehat{\mathbf{p}}_c\right) = 0,\ \forall c \in \mathcal{C}, \tag{A.13}$$

$$\widehat{\alpha}_c \left(R_c\left(\widehat{\mathbf{p}}_c\right) - \widehat{\theta}_c C_c\left(\widehat{\mathbf{p}}_c\right)\right) = 0,\ \forall c \in \mathcal{C}, \tag{A.14}$$

$$\widehat{\lambda}_l h_l\left(\widehat{\mathbf{p}}\right) = 0,\ \forall l \in \mathcal{V}, \tag{A.15}$$

$$R_c\left(\widehat{\mathbf{p}}_c\right) - \widehat{\theta}_c C_c\left(\widehat{\mathbf{p}}_c\right) \leq 0,\ \forall l \in \mathcal{V}, \tag{A.16}$$

$$h_l\left(\widehat{\mathbf{p}}\right) \leq 0,\ \forall l \in \mathcal{V}, \tag{A.17}$$

$$\widehat{\alpha} \geq \mathbf{0}, \tag{A.18}$$

$$\widehat{\lambda} \geq \mathbf{0}. \tag{A.19}$$

Now, let us consider the parametric problem $(P_{\mathbf{x}})$ with a parameter $\mathbf{x} = (\alpha, \theta)$. Its Lagrangian is given by:

$$\mathcal{L}_{\mathbf{x}}(\mathbf{p}; \lambda) = \sum_{c \in \mathcal{C}} \alpha_c \left(R_c(\mathbf{p}_c) - \theta_c C_c(\mathbf{p}_c) \right) + \sum_{\forall l \in \mathcal{V},} \lambda_l h_l(\mathbf{p}). \tag{A.20}$$

We can identify equations (A.12), (A.15), (A.17) and (A.19) as the Karush-Kuhn-Tucker (KKT) optimality conditions [61] of $(P_{\mathbf{x}})$ with $\mathbf{x} = \left(\widehat{\alpha}, \widehat{\theta}\right)$. Since $(P_{\mathbf{x}})$ is convex, these KKT conditions are sufficient to prove that $\widehat{\mathbf{p}}$ is also an optimal solution of $(P_{\mathbf{x}})$ [61].

Finally, we obtain (2.14) from equation (A.13), i.e. $\widehat{\mathbf{p}}$ satisfies:

$$\widehat{\alpha}_c C_c(\widehat{\mathbf{p}}_c) - 1 = 0, \ \forall c \in \mathcal{C}. \tag{A.21}$$

By definition, we have $C_c(\mathbf{p}_c) > 0, \ \forall c \in \mathcal{C}$, and with (A.21), we also have $\widehat{\alpha}_c > 0, \ \forall c \in \mathcal{C}$. Consequently, the complementary slackness conditions in (A.14) simplify to:

$$R_c(\widehat{\mathbf{p}}_c) - \widehat{\theta}_c C_c(\widehat{\mathbf{p}}_c) = 0, \ \forall c \in \mathcal{C}. \tag{A.22}$$

Therefore, it is proved that an optimal solution of (P) must satisfy the equations in (2.13) and (2.14). □

A.2 Proof of Lemma 2.1

Here, we are going to prove the three properties of the matrix-valued function $\mathbf{F}(\mathbf{x})$ stated in Lemma 2.1. Recall that a function $f : \mathcal{X} \to \mathcal{Y}$ is said Lipschitz continuous on a set $\mathcal{X}$ if and only if there exists a constant $K \geq 0$ such that $\|f(\mathbf{x}_1) - f(\mathbf{x}_2)\| \leq K \|\mathbf{x}_1 - \mathbf{x}_2\|$, $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$.

First, we prove that the Jacobian of $\mathbf{F}$ is given by (2.22). We can verify that the objective function of the parametric problem $(P_{\mathbf{x}})$ is strictly convex by computing its second derivative. Thus, its solution is unique for a given $\mathbf{x}$, and we get from (2.17) and (2.18)

that:

$$\frac{\partial F_c(\mathbf{x})}{\partial \alpha_d} = 0,\ \forall d \in \mathcal{C}, \qquad \frac{\partial F_c(\mathbf{x})}{\partial \theta_d} = 0,\ \forall d \in \mathcal{C} \backslash \{c\},$$

$$\frac{\partial F_{c+|\mathcal{C}|}(\mathbf{x})}{\partial \theta_d} = 0,\ \forall d \in \mathcal{C}, \qquad \frac{\partial F_{c+|\mathcal{C}|}(\mathbf{x})}{\partial \alpha_c} = C_c\left(\widetilde{\mathbf{p}}_c\left(\mathbf{x}\right)\right),$$

$$\frac{\partial F_c(\mathbf{x})}{\partial \theta_c} = C_c\left(\widetilde{\mathbf{p}}_c\left(\mathbf{x}\right)\right), \qquad \frac{\partial F_c(\mathbf{x})}{\partial \alpha_d} = 0,\ \forall d \in \mathcal{C} \backslash \{c\}.$$

Correspondingly, the Jacobian matrix $\mathbf{F}'\left(\mathbf{x}\right)$ is given by the diagonal matrix in (2.22) and it is positive definite since $C_c\left(\mathbf{p}_c\right) > 0,\ \forall c \in \mathcal{C}$ by definition.

Next, we prove the Lipschitz continuity and strong monotonicity of the matrix-valued function $\mathbf{F}$. By the definition in (2.17) and (2.18), $\mathbf{F}$ can be written as a composite function $\mathbf{G} \circ \widetilde{\mathbf{p}}$ where $\mathbf{G}:\ \mathbb{R}^{|\mathcal{U}|} \to \mathbb{R}^{2|\mathcal{U}|}$ is defined by the convex functions:

$$G_c\left(\mathbf{p}\right) = \theta_c C_c\left(\mathbf{p}\right) - r_c\left(\mathbf{p}\right), \quad \forall c \in \mathcal{C}, \tag{A.23}$$

$$G_{c+|\mathcal{C}|}\left(\mathbf{p}\right) = \alpha_c C_c\left(\mathbf{p}\right) - 1, \quad \forall c \in \mathcal{C}, \tag{A.24}$$

and $\widetilde{\mathbf{p}}\left(\mathbf{x}\right)$ is a function that returns an optimal solution of the parametric convex problem $(P_{\mathbf{x}})$ for a given $\mathbf{x}$.

First, the component functions G_c and $G_{c+|\mathcal{C}|}$ are all Lipschitz continuous in the set $\mathcal{K}$ defined in (2.11) since $\mathcal{K}$ is compact and G_c and $G_{c+|\mathcal{U}|}$ are convex in $\mathcal{K}$ for each $c \in \mathcal{C}$ (see Theorem C.4.1 in [110]). Therefore, the matrix-valued function $\mathbf{G}$ is also Lipschitz continuous in the set $\mathcal{K}$. Furthermore, by invoking Theorem D.1 in [111], it can be shown that the function $\widetilde{\mathbf{p}}\left(\mathbf{x}\right)$ is Lipschitz continuous in the set $\mathcal{X}$ defined in (2.19) since the parametric problem $(P_{\mathbf{x}})$ admits a unique solution for each $\mathbf{x}$ by strict convexity and since the set $\mathcal{X}$ is compact and convex (see also Lemma 2.1 in [112]). If $\widetilde{\mathbf{p}}$ and $\mathbf{G}$ are Lipschitz continuous in $\mathcal{X}$ and $\mathcal{K}$ respectively, then it implies that $\mathbf{F} = \mathbf{G} \circ \widetilde{\mathbf{p}}$ is also Lipschitz continuous in $\mathcal{K}$ with a certain constant M. This result is due to Theorem 12.6 in [113]. Finally,

to show that $\mathbf{F}$ is strongly monotone, we write:

$$\mathbf{y}^{\top}\mathbf{F}'(\mathbf{x})\,\mathbf{y}=\sum_{c\in\mathcal{C}}\left(y_c^2+y_{c+|\mathcal{U}|}^2\right)C_c\left(\widetilde{\mathbf{p}}_c(\mathbf{x})\right)\geq\sum_{c=1}^{2|\mathcal{C}|}y_c^2 m_c\geq m\left\|\mathbf{y}\right\|^2,\ \forall\mathbf{y}\in\mathbb{R}^{2|\mathcal{C}|} \tag{A.25}$$

where $m_c\triangleq\min_{\mathbf{p}\in\mathcal{K}}C_c(\mathbf{p})=P_{f,c}$ for each $c\in\mathcal{C}$ and $m=\min_{c\in\mathcal{U}}P_{f,c}>0$. The last inequality in (A.25) implies that the function $\mathbf{F}$ is strongly monotone with a constant $m>0$.

Finally, we use the established properties above to show that $\mathbf{F}'$ is also Lipschitz continuous and that its inverse $\mathbf{F}'^{-1}$ is bounded. As shown previously, the function $\widetilde{\mathbf{p}}(\mathbf{x})$ is Lipschitz continuous in the compact and convex set $\mathcal{X}$. Thus, there is a constant $L\geq 0$ such that $\|\widetilde{\mathbf{p}}(\mathbf{x}_1)-\widetilde{\mathbf{p}}(\mathbf{x}_2)\|\leq L\|\mathbf{x}_1-\mathbf{x}_2\|$ for all $(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{X}\times\mathcal{X}$, and we obtain:

$$\left|C_c\left(\widetilde{\mathbf{p}}(\mathbf{x}_1)\right)-C_c\left(\widetilde{\mathbf{p}}(\mathbf{x}_2)\right)\right|=\psi_c\left|\widetilde{\mathbf{p}}_c(\mathbf{x}_1)-\widetilde{\mathbf{p}}_c(\mathbf{x}_2)\right|\leq\psi_c L\left\|\mathbf{x}_1-\mathbf{x}_2\right\|,\ \forall c\in\mathcal{C}. \tag{A.26}$$

From (2.22) and (A.26), we then have:

$$\left\|\mathbf{F}'(\mathbf{x}_1)-\mathbf{F}'(\mathbf{x}_2)\right\|\leq\sqrt{2\left|\mathcal{C}\right|\sum_{c\in\mathcal{C}}\psi_c^2}L\left\|\mathbf{x}_1-\mathbf{x}_2\right\|, \tag{A.27}$$

which means that $\mathbf{F}'$ is also Lipschitz continuous in $\mathcal{X}$. Moreover, its inverse $\mathbf{F}'^{-1}$ equals:

$$\mathbf{F}'^{-1}(\mathbf{x})=\begin{bmatrix}\operatorname{diag}\left(\left(\frac{1}{C_c(\widetilde{\mathbf{p}}(\mathbf{x}))}\right)_{c\in\mathcal{U}}\right) & \mathbf{0}_{|\mathcal{U}|\times|\mathcal{U}|}\\ \mathbf{0}_{|\mathcal{U}|\times|\mathcal{U}|} & \operatorname{diag}\left(\left(\frac{1}{C_c(\widetilde{\mathbf{p}}(\mathbf{x}))}\right)_{c\in\mathcal{U}}\right)\end{bmatrix}. \tag{A.28}$$

Given the bounds of the set $\mathcal{X}$ in (2.20) and (2.21), each element $\frac{1}{C_c(\widetilde{\mathbf{p}}(\mathbf{x}))}$ of $\mathbf{F}'^{-1}$ is bounded from above by $\alpha_c^{\max}$. Hence, there exists a constant $B>0$ such that $\|\mathbf{F}'^{-1}(\mathbf{x})\|\leq B$.

A.3 Proof of Proposition 2.2

Here, we prove in two steps that the system of equations $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ has a unique solution. Before doing that, let us define the mapping $P(\mathbf{x}) = \text{Proj}_{\mathcal{X}}(\mathbf{x} - \mu\mathbf{F}(\mathbf{x}))$ where $\text{Proj}_{\mathcal{X}}$ denotes an orthogonal projection onto the convex set $\mathcal{X}$. In addition, we select the parameter $\mu \in \left(0, \frac{2m}{M^2}\right)$ where m and M are the constants associated with the strong monotonicity and Lipschitz continuity of $\mathbf{F}$ as previously shown.

The first step in proving Proposition 2.2 is to show that P is a contractive mapping in $\mathcal{X}$ for all $\mu \in \left(0, \frac{2m}{M^2}\right)$. In this case, P is guaranteed to have a unique fixed point. The second step is then to prove that $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ is equivalent to the fixed-point equation $P(\mathbf{x}) = \mathbf{x}$ for all $\mu \in \left(0, \frac{2m}{M^2}\right)$. As a result, the solution of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ will also be unique.

Proof that $P(\mathbf{x})$ is a contraction

Let us prove that P is a contraction. From the properties of F in Lemma 2.1, we can write:

$$\|P(\mathbf{x}_1) - P(\mathbf{x}_2)\|^2 = \|\text{Proj}_{\mathcal{X}}(\mathbf{x}_1 - \mu\mathbf{F}(\mathbf{x}_1)) - \text{Proj}_{\mathcal{X}}(\mathbf{x}_2 - \mu\mathbf{F}(\mathbf{x}_2))\|^2 \tag{A.29}$$

$$\leq \|(\mathbf{x}_1 - \mu\mathbf{F}(\mathbf{x}_1)) - (\mathbf{x}_2 - \mu\mathbf{F}(\mathbf{x}_2))\|^2 \tag{A.30}$$

$$= \|\mathbf{x}_1 - \mathbf{x}_2\|^2 - 2\mu(\mathbf{x}_1 - \mathbf{x}_2)^\top(\mathbf{F}(\mathbf{x}_1) - \mathbf{F}(\mathbf{x}_2)) + \mu^2\|\mathbf{F}(\mathbf{x}_1) - \mathbf{F}(\mathbf{x}_2)\|^2 \tag{A.31}$$

$$\leq \|\mathbf{x}_1 - \mathbf{x}_2\|^2 - 2\mu m\|\mathbf{x}_1 - \mathbf{x}_2\|^2 + \mu^2\|\mathbf{F}(\mathbf{x}_1) - \mathbf{F}(\mathbf{x}_2)\| \tag{A.32}$$

$$= \left(1 - 2\mu m + \mu^2 M^2\right)\|\mathbf{x}_1 - \mathbf{x}_2\|^2, \tag{A.33}$$

where the inequality (A.30) results from the non-expansiveness of the projection operator [114]. The inequalities (A.32) and (A.33) are respectively obtained from the strong monotonicity and Lipschitz continuity of $\mathbf{F}$ according to Lemma 2.1.

From (A.33), we obtain $\|P(\mathbf{x}_1) - P(\mathbf{x}_2)\| \leq \sqrt{1 - 2\mu m + \mu^2 M^2}\,\|\mathbf{x}_1 - \mathbf{x}_2\|$. Since $\mu \in$

$\left(0, \frac{2m}{M^2}\right)$, we have $\sqrt{1-2\mu m+\mu^2 M^2} < 1$. Thus, we showed that P is a contractive mapping.

Proof of equivalence between $P(\mathbf{x}) = \mathbf{x}$ and $\mathbf{F}(\mathbf{x}) = \mathbf{0}$

First, suppose that $\widehat{\mathbf{x}} \in \mathcal{X}$ satisfies $\mathbf{F}(\widehat{\mathbf{x}}) = \mathbf{0}$. Then, we have:

$$P(\widehat{\mathbf{x}}) = \text{Proj}_{\mathcal{X}}(\widehat{\mathbf{x}} - \mu\mathbf{F}(\widehat{\mathbf{x}})) = \text{Proj}_{\mathcal{X}}(\widehat{\mathbf{x}}) = \widehat{\mathbf{x}}$$

which means that $\widehat{\mathbf{x}}$ is a fixed point of $P(\mathbf{x})$.

Conversely, suppose that $\widehat{\mathbf{x}}$ is the unique fixed point of the contractive mapping P for a given $\mu \in \left(0, \frac{2m}{M^2}\right)$. In other words, we have $P(\widehat{\mathbf{x}}) = \text{Proj}_{\mathcal{X}}(\widehat{\mathbf{x}} - \mu\mathbf{F}(\widehat{\mathbf{x}})) = \widehat{\mathbf{x}}$. Then, we next show that $\widehat{\mathbf{x}}$ also satisfies $\mathbf{F}(\widehat{\mathbf{x}}) = \mathbf{0}$. Since we assume that the problem (P) is feasible, we know that there exists $\widetilde{\mathbf{x}} \in \mathcal{X}$ such that $\mathbf{F}(\widetilde{\mathbf{x}}) = \mathbf{0}$. We then need to prove that $\widehat{\mathbf{x}} = \widetilde{\mathbf{x}}$. For that, we use the projection theorem (see Proposition 1.1.9 in [114]) which states that:

$$(\text{Proj}_{\mathcal{X}}(\mathbf{z}) - \mathbf{z})^\top (\text{Proj}_{\mathcal{X}}(\mathbf{z}) - \mathbf{y}) \leq 0, \ \forall \mathbf{y} \in \mathcal{X}.$$

Setting $\mathbf{z} := \widehat{\mathbf{x}} - \mu\mathbf{F}(\widehat{\mathbf{x}})$, we have:

$$(\text{Proj}_{\mathcal{X}}(\widehat{\mathbf{x}} - \mu\mathbf{F}(\widehat{\mathbf{x}})) - (\widehat{\mathbf{x}} - \mu\mathbf{F}(\widehat{\mathbf{x}})))^\top (\text{Proj}_{\mathcal{X}}(\widehat{\mathbf{x}} - \mu\mathbf{F}(\widehat{\mathbf{x}})) - \mathbf{y}) \leq 0,$$

which simplifies to $\mathbf{F}(\widehat{\mathbf{x}})^\top (\widehat{\mathbf{x}} - \mathbf{y}) \leq 0, \ \forall \mathbf{y} \in \mathcal{X}$ since $\mu > 0$. Next, using the fact that $\mathbf{F}(\widetilde{\mathbf{x}}) = \mathbf{0}$ and replacing $\mathbf{y} := \widetilde{\mathbf{x}}$, we obtain:

$$(\mathbf{F}(\widehat{\mathbf{x}}) - \mathbf{F}(\widetilde{\mathbf{x}}))^\top (\widehat{\mathbf{x}} - \widetilde{\mathbf{x}}) \leq 0. \tag{A.34}$$

According to Lemma 2.1, $\mathbf{F}$ is strongly monotone with a constant $m > 0$. It means that:

$$(\mathbf{F}(\widehat{\mathbf{x}}) - \mathbf{F}(\widetilde{\mathbf{x}}))^\top (\widehat{\mathbf{x}} - \widetilde{\mathbf{x}}) \geq m \|\widehat{\mathbf{x}} - \widetilde{\mathbf{x}}\|^2. \tag{A.35}$$

Finally, we combine (A.34) and (A.35) into:

$$0 \leq \left(\mathbf{F}\left(\widetilde{\mathbf{x}}\right) - \mathbf{F}\left(\widehat{\mathbf{x}}\right)\right)^{\top} \left(\widehat{\mathbf{x}} - \widetilde{\mathbf{x}}\right) \leq -m \left\|\widehat{\mathbf{x}} - \widetilde{\mathbf{x}}\right\|^{2}, \tag{A.36}$$

which implies that $\widehat{\mathbf{x}} = \widetilde{\mathbf{x}}$ since $m > 0$.

In summary, we showed that the system $\mathbf{F}\left(\mathbf{x}\right) = \mathbf{0}$ is equivalent to the fixed point equation $P\left(\mathbf{x}\right) = \mathbf{x}$ for any $\mu \in \left(0, \frac{2m}{M^2}\right)$. Furthermore, the solution of $\mathbf{F}\left(\mathbf{x}\right) = \mathbf{0}$ is unique because the $P\left(\mathbf{x}\right)$ is a contraction and has a unique fixed point.

Appendix B

Proof of Result in Chapter 3

Proof of Proposition 3.1

First, we prove by induction that if we start with a feasible point of (P) then the iterates generated by the step (S.3) of Algorithm 3.1 are all feasible. Note that for any power allocation $\mathbf{p}$ that satisfies the rate constraints c_1 and the transmit power constraints c_4, we can always find a set of feasible power usage vectors $\left\{\mathbf{p}^g, \mathbf{p}^h, \mathbf{w}\right\}$ that satisfies the other constraints $\mathrm{c}_{2,3,5,6}$. Thus, we need only to prove that the iterates $\left\{\mathbf{p}^{(n)}\right\}$ of the power allocation are feasible. To do that, let us assume that $\mathbf{p}^{(n)}$ is feasible for (P), i.e for each $c \in \mathcal{M} \cup \mathcal{K}$, we have $R_c\left(\mathbf{p}^{(n)}\right) \geq \rho_c T$ and $\sum\limits_{u \in \mathcal{U}_c} p_{u,t} \leq P_c$, $\forall t = 1, \ldots, T$. Due to Lemma 3.1, we obtain:

$$R_c\left(\mathbf{p}^{(n)}\right) = \widetilde{R}_c\left(\mathbf{p}^{(n)}; \mathbf{p}^{(n)}\right) \geq \rho_c T, \ \forall c \in \mathcal{M} \cup \mathcal{K}. \tag{B.1}$$

(B.1) means that $\mathbf{p}^{(n)}$ is also feasible for the subproblem $\left(Q_{\mathbf{p}^{(n)}}\right)$, which further implies that $\left(Q_{\mathbf{p}^{(n)}}\right)$ has a non-empty feasible set.

Then, step (S.2) of the algorithm gives us an optimal solution $\mathbf{p}^{(n+1)}$ of subproblem $\left(Q_{\mathbf{p}^{(n)}}\right)$. This solution must also be feasible to $\left(Q_{\mathbf{p}^{(n)}}\right)$ and satisfies:

$$\widetilde{R}_c\left(\mathbf{p}^{(n+1)}; \mathbf{p}^{(n)}\right) \geq \rho_c T, \ \forall c \in \mathcal{M} \cup \mathcal{K}. \tag{B.2}$$

According to Lemma 3.1, the following inequality must also hold:

$$R_c\left(\mathbf{p}^{(n+1)}\right) \geq \widetilde{R}_c\left(\mathbf{p}^{(n+1)}; \mathbf{p}^{(n)}\right), \ \forall c \in \mathcal{M} \cup \mathcal{K} \tag{B.3}$$

since $\widetilde{R}_c$ minorizes R_c. Combining (B.2) and (B.3), we obtain $R_c\left(\mathbf{p}^{(n+1)}\right) \geq \rho_c T$, i.e. the new iterate $\mathbf{p}^{(n+1)}$ is feasible for (P). By induction, if the initial power allocation $\mathbf{p}^{(0)}$ is feasible, then all subsequent iterates must remain feasible.

Next, we prove that the proposed algorithm is a monotonically convergent and ascent algorithm. In other words, we will show that:

$$\sum_{c\in\mathcal{M}\cup\mathcal{K}} \mathrm{EE}_c^{(n+1)} \geq \sum_{c\in\mathcal{M}\cup\mathcal{K}} \mathrm{EE}_c^{(n)},\ \forall n \in \mathbb{N}$$

where $\mathrm{EE}_c^{(n+1)}$ is the energy efficiency of cell c achieved with the optimal solution $\mathbf{p}^{(n+1)}$ of subproblem $\left(Q_{\mathbf{p}^{(n)}}\right)$, whose parameter is $\mathbf{q} = \mathbf{p}^{(n)}$. Using Lemma 1 again, we have the minorization relationship:

$$\sum_{c\in\mathcal{M}\cup\mathcal{K}} \mathrm{EE}_c^{(n+1)} \geq \sum_{c\in\mathcal{M}\cup\mathcal{K}} \widetilde{\mathrm{EE}}_c\left(\mathbf{p}^{(n+1)}, \mathbf{p}_c^{g\ (n+1)}; \mathbf{p}^{(n)}\right). \tag{B.4}$$

Moreover, $\mathbf{p}^{(n+1)}$ maximizes the objective function of subproblem $\left(Q_{\mathbf{p}^{(n)}}\right)$. Hence, we have:

$$\sum_{c\in\mathcal{M}\cup\mathcal{K}} \widetilde{\mathrm{EE}}_c\left(\mathbf{p}^{(n+1)}, \mathbf{p}_c^{g\ (n+1)}; \mathbf{p}^{(n)}\right) \geq \sum_{c\in\mathcal{M}\cup\mathcal{K}} \widetilde{\mathrm{EE}}_c\left(\mathbf{p}^{(n)}, \mathbf{p}_c^{g\ (n)}; \mathbf{p}^{(n)}\right). \tag{B.5}$$

The right-hand side expression of (B.3) is equal to $\sum_{c\in\mathcal{M}\cup\mathcal{K}} \mathrm{EE}_c^{(n)}$ according to Lemma 1. Consequently, the last two inequalities (B.2) and (B.3) imply that $\sum_{c\in\mathcal{M}\cup\mathcal{K}} \mathrm{EE}_c^{(n+1)} \geq \sum_{c\in\mathcal{M}\cup\mathcal{K}} \mathrm{EE}_c^{(n)}$.

Appendix C

Proof of Results in Chapter 4

C.1 Proof of Lemma 4.1

After fixing the clustering variables to $\left(\mathbf{x}^k, \mathbf{y}^k\right)$, the objective function becomes a convex function of only the beamforming variable $\mathbf{v}$. Next, we reformulate the non-convex SINR constraints in conic form [115]. Note that with simple manipulations, the SINR constraints in (4.19) can be rewritten as:

$$\left(1+\frac{1}{\rho_u}\right)\left|\mathbf{h}_u^H\mathbf{v}_u\right|^2 \geq \left\|\begin{bmatrix}\mathbf{h}_u^H\mathbf{V} & \sigma_u\end{bmatrix}\right\|^2,\ \forall u \tag{C.1}$$

Moreover, any phase rotation on a beamforming vector $\mathbf{v}_u$ preserves the feasibility and the objective function value. It means we can multiply an optimal $\mathbf{v}_u^*$ by an appropriate $e^{j\theta_u}$ and obtain another feasible and optimal beamformer $\hat{\mathbf{v}}_u^* = e^{j\theta_u}\mathbf{v}_u$. In particular, we can choose θ_u to make the inner product $\mathbf{h}_u^H\mathbf{v}_u$ real and positive for each user u. Since such phase rotation does not change the problem, (C.1) is thus equivalent to:

$$\left\|\begin{bmatrix}\mathbf{h}_u^H\mathbf{V} & \sigma_u\end{bmatrix}\right\| \leq \sqrt{1+\frac{1}{\rho_u}}\mathbf{h}_u^H\mathbf{v}_u,\ \forall u \in \mathcal{U} \tag{C.2}$$

$$\text{Re}\left(\mathbf{h}_u^H\mathbf{v}_u\right) \geq 0,\ \forall u \in \mathcal{U} \tag{C.3}$$

$$\text{Im}\left(\mathbf{h}_u^H\mathbf{v}_u\right) = 0,\ \forall u \in \mathcal{U} \tag{C.4}$$

with (C.2) defining a second-order conic constraint [61]. As a result, $\left(S^k\right)$ is convex. Since the second-order cone in (C.2) is closed and the power and interference limit constraints

define compact sets, their intersection, i.e. $\mathcal{F}_{\mathbf{v}}$, is bounded and closed, thus compact.

C.2 Proof of Proposition 4.2

In short, the equivalence between (M) and (P') is obtained from the dual representations of the function g and of the feasible set $\mathcal{H}$ defined in (4.27)-(4.28). First, recall from Lemma 4.1 that the set $\mathcal{F}_{\mathbf{v}}$ in (4.23) is convex, closed and bounded and that constraints (4.26) are convex for any fixed $\left(\mathbf{x}^k,\mathbf{y}^k\right) \in \mathcal{F}_{(\mathbf{x},\mathbf{y})}$. Therefore, we can apply Theorem 2.2 in [90] to express that $(\mathbf{x},\mathbf{y}) \in \mathcal{H}$ if and only if it satisfies the infinite system:

$$\inf_{\mathbf{v}\in\mathcal{F}_{\mathbf{v}}} \overline{\mathcal{L}}\left(\mathbf{v},\mu,\mathbf{x}\right) \leq 0,\ \forall \mu \geq 0 : \sum_{(b,u)\in\mathcal{B}\times\mathcal{U}} \mu_{bu} = 1$$

Since (S) is convex by Lemma 4.1, the strong duality theorem gives $g\left(\mathbf{x},\mathbf{y}\right) = \sup_{\lambda\geq 0} \inf_{\mathbf{v}\in\mathcal{F}_{\mathbf{v}}} \mathcal{L}\left(\mathbf{v},\lambda,\mathbf{x},\mathbf{y}\right)$, $\forall\left(\mathbf{x},\mathbf{y}\right) \in \mathcal{H} \cap \mathcal{F}_{(\mathbf{x},\mathbf{y})}$. Substituting g and $\mathcal{H}$ with their dual representations, problem (M) thus becomes:

$$\begin{aligned} \underset{\mathbf{x}\ \mathbf{y}}{\text{minimize}} \quad & \sup_{\lambda\geq 0} \inf_{\mathbf{v}\in\mathcal{F}_{\mathbf{v}}} \mathcal{L}\left(\mathbf{v},\lambda,\mathbf{x},\mathbf{y}\right) \\ \text{subject to} \quad & (\mathbf{x},\mathbf{y}) \in \mathcal{F}_{(\mathbf{x},\mathbf{y})} \\ & 0 \geq \inf_{\mathbf{v}\in\mathcal{F}_{\mathbf{v}}} \overline{\mathcal{L}}\left(\mathbf{v},\mu,\mathbf{x}\right),\ \forall \mu \geq 0 :\ \textstyle\sum_{(b,u)\in\mathcal{B}\times\mathcal{U}} \mu_{bu} = 1 \end{aligned} \tag{C.5}$$

Since $\mathcal{F}_{\mathbf{v}}$ is convex and compact, the optimal solution of (S) is bounded. Thus, we can replace the infimum in (C.5) with minimum. Using the definition of supremum as the smallest upper-bound, we can obtain (4.32).

C.3 Proof of Lemma 4.2

Since $\left(\mathbf{v}^k, \lambda^k\right)$ is a pair of optimal solution and multiplier for $\left(S^k\right)$, we have:

$$\mathbf{v}^k = \underset{\mathbf{v}\in\mathcal{F}_\mathbf{v}}{\operatorname{argmin}}\ \mathcal{L}\left(\mathbf{v}, \lambda^k, \mathbf{x}^k, \mathbf{y}^k\right) \tag{C.6}$$

$$= \underset{\mathbf{v}\in\mathcal{F}_\mathbf{v}}{\operatorname{argmin}}\ \left[f_1\left(\mathbf{v}\right) + f_2\left(\mathbf{x}^k, \mathbf{y}^k\right) + \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \lambda_{bu}^k \left(\|\mathbf{v}_{bu}\|_2^2 - x_{bu}^k \cdot P_b^{\max}\right)\right] \tag{C.7}$$

$$= \underset{\mathbf{v}\in\mathcal{F}_\mathbf{v}}{\operatorname{argmin}}\ \left(f_1\left(\mathbf{v}\right) + \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \lambda_{bu}^k \|\mathbf{v}_{bu}\|_2^2\right) \tag{C.8}$$

where (C.6) is due to the optimality of $\left(\mathbf{v}^k, \lambda^k\right)$. After replacing the expression of $\mathcal{L}$ in (4.30) into (C.7) and removing the constant terms, we obtain the last equation (C.7).

Now, we can rewrite the function ξ in (4.35) as:

$$\xi\left(\mathbf{x}, \mathbf{y}; \lambda^k\right) = \min_{\mathbf{v}\in\mathcal{F}_\mathbf{v}} \left[f_1\left(\mathbf{v}\right) + f_2\left(\mathbf{x}, \mathbf{y}\right) + \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \lambda_{bu}^k \left(\|\mathbf{v}_{bu}\|_2^2 - x_{bu} \cdot P_b^{\max}\right)\right] \tag{C.9}$$

$$= f_2\left(\mathbf{x}, \mathbf{y}\right) - \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} P_b^{\max}\lambda_{bu}^k x_{bu} + \min_{\mathbf{v}\in\mathcal{F}_\mathbf{v}} \left(f_1\left(\mathbf{v}\right) + \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \lambda_{bu}^k \|\mathbf{v}_{bu}\|_2^2\right) \tag{C.10}$$

From (C.7), we see that the primal solution $\mathbf{v}^k$ of $\left(S^k\right)$ also solves the minimization term in (C.10). Thus, we have:

$$\xi\left(\mathbf{x}, \mathbf{y}; \lambda^k\right) = f_2\left(\mathbf{x}, \mathbf{y}\right) - \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} P_b^{\max}\lambda_{bu}^k x_{bu} + f_1\left(\mathbf{v}^k\right) + \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \lambda_{bu}^k \left\|\mathbf{v}_{bu}^k\right\|_2^2 \tag{C.11}$$

After rearranging the terms in (C.11), we obtain the result $\xi\left(\mathbf{x}, \mathbf{y}; \lambda^k\right) = \mathcal{L}\left(\mathbf{v}^k, \lambda^k, \mathbf{x}, \mathbf{y}\right)$.

C.4 Proof of Lemma 4.3

First, we write the partial Langragian $\mathcal{L}_F$ of the feasibility problem $\left(F^k\right)$ as:

$$\mathcal{L}_F\left(\mathbf{v}, \alpha, \mu, x\right) = \alpha + \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \mu_{bu}\left(\|\mathbf{v}_{bu}\|_2^2 - x_{bu} \cdot P_b^{\max} - \alpha\right) \tag{C.12}$$

Since $\left(F^k\right)$ is convex and $\left(\overline{\mathbf{v}}^k, \alpha^k\right)$ and μ^k are its optimal primal and dual solutions, then we have:

$$\left(\overline{\mathbf{v}}^k, \alpha^k\right) = \underset{\mathbf{v}\in\mathcal{F}_\mathbf{v},\ \alpha\geq 0}{\operatorname{argmin}} \mathcal{L}_F\left(\mathbf{v}, \alpha, \mu^k, x^k\right) \tag{C.13}$$

$$= \underset{\mathbf{v}\in\mathcal{F}_\mathbf{v},\ \alpha\geq 0}{\operatorname{argmin}} \alpha\left(1 - \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \mu_{bu}^k\right) + \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \mu_{bu}^k\left(\|\mathbf{v}_{bu}\|_2^2 - x_{bu}^k \cdot P_b^{\max}\right) \tag{C.14}$$

after rearranging the expression of $\mathcal{L}_F$ in (C.12). With the optimality condition $\frac{\partial \mathcal{L}_F\left(\mathbf{v},\alpha,\mu^k,x^k\right)}{\partial \alpha} = 0$, we can show that μ^k must satisfy:

$$\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \mu_{bu}^k = 1 \tag{C.15}$$

By using the condition (C.15) in (C.14) and removing the constant terms in (C.14), we obtain:

$$\overline{\mathbf{v}}^k = \underset{\mathbf{v}\in\mathcal{F}_\mathbf{v}}{\operatorname{argmin}} \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}} \mu_{bu}^k \|\mathbf{v}_{bu}\|_2^2 \tag{C.16}$$

Following the same procedure (C.9)-(C.11) as in the proof of Lemma 4.2, we can exploit the linear separability of $\overline{\mathcal{L}}$ with the fact (C.16) to show that $\overline{\xi}\left(\mathbf{x}; \mu^k\right) = \overline{\mathcal{L}}\left(\overline{\mathbf{v}}^k, \mu^k, \mathbf{x}\right)$.