Enhancing Energy and Spectral Efficiencies in Wireless Cellular Networks

by

Hamza Umit Sokun

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
Department of Systems and Computer Engineering
Carleton University
Ottawa, Ontario
December, 2017

The undersigned hereby recommends to the
Faculty of Graduate and Postdoctoral Affairs
acceptance of the thesis

Enhancing Energy and Spectral Efficiencies in Wireless Cellular Networks

Submitted by **Hamza Umit Sokun**

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

Professor Halim Yanikomeroglu, Thesis Co-supervisor

Professor Ramy Gohary, Thesis Co-supervisor

Professor Ozgur Ercetin, External Examiner
Faculty of Engineering and Natural Sciences,
Sabanci University, Istanbul, Turkey

Professor Yvan Labiche, Department Chair

Ottawa-Carleton Institute for Electrical and Computer Engineering
Department of Systems and Computer Engineering
Carleton University
December, 2017

Abstract

Today's wireless networks are facing crucial challenges, such as radio spectrum scarcity, and excessive energy consumption. The vision of next generation networks lies in overcoming these challenges, and providing seamless wireless communication. This thesis aims to contribute to shaping the vision of next generation wireless networks. Specifically, the thesis investigates how to optimize various network functionalities to improve overall utilization of radio resources, and how to enhance practical significance of the radio resource allocation techniques proposed in academia.

The thesis first focuses on enhancing network energy efficiency (EE) for downlink transmission of macro-only cellular networks. It develops a novel joint design that incorporates resource blocks (RBs) and discrete power levels allocations. The design objective is to determine the assignment of RBs and power levels that maximizes network overall EE. Next, the thesis focuses on enhancing EE fairness (balancing) among individual users for uplink transmission of macro-only cellular networks. Particularly, in such networks, it considers a joint design of RBs and power levels allocations to maximize minimum individual EE. Unfortunately, both the downlink and uplink optimization problems arising from the joint designs are non-convex, and hence difficult to solve. To overcome this difficulty, the thesis reformulates the design problems in a form that is amenable to the standard semidefinite relaxation (SDR) with Gaussian randomization approach. This approach has a polynomial-time complexity, and yields a close-to-optimal performance. Simulation results corroborate the effectiveness of the proposed approach against the optimal benchmark.

The thesis later shifts the focus to enhancing spectral efficiency in multi-tier heterogeneous cellular networks (HetNets), where various low-power base stations (BSs) are deployed within the coverage area of a macro BS. It investigates the problem of user-to-base station (BS) association in downlink transmission of HetNets. More specifically, the thesis proposes a design of user-to-BS association, where the objective

is composed of the linear combination of two quantities: 1) the number of accommodated users with satisfied quality-of-service requirement, which should be maximized to increase revenue, and 2) the aggregate RB usage, which should be minimized to efficiently utilize the network resources. In its simplest form, this problem is non-convex. Hence, to obtain a close-to-optimal user-to-BS association, the thesis uses the efficient two-stage SDR-based technique. Numerical examples confirm that the proposed design accommodates a significantly higher number of users compared to currently available ones.

The last part of the thesis considers a joint design of user-to-BS association, and RBs and power levels allocations in HetNets, allowing RBs to be reused opportunistically. In this joint design, the objective is again to maximize the number of accommodated users with the minimum network resources. The thesis discusses two network instances: with and without time-sharing. In the time-sharing case, user-to-BS association, RBs and power allocations are time-shared, and hence, not fixed throughout the signalling interval. In contrast, in the no time-sharing case, the network functionalities, once determined, are fixed throughout the signalling interval. The optimization problems in both instances are non-convex. To circumvent this difficulty, the thesis generates bounds on the solution of the original problems. To be specific, for the time-sharing case, it develops tight upper and lower bounds, and for the no time-sharing one, a lower bound using the SDR-based technique. Numerical investigations show that the proposed bounds provide a close-to-optimal performance.

Dedicated to my parents, my wife, and my daughter

Acknowledgments

First and foremost, I would like to express my sincere gratitude for my co-supervisors, Dr. Halim Yanikomeroglu and Dr. Ramy H. Gohary, for their dedicated support throughout the development of this work. Both of my co-supervisors have not only deepened my knowledge, but have also taught the mindset of being a productive researcher. This work would not be possible without their patient guidance, and constant encouragement.

I would also like to express my deep gratitude to Dr. Ebrahim Bedeer, Dr. Salama Ikki, Dr. Amr El-Keyi, Dr. Rainer Schoenen, and Dr. Akram Bin Sediq for their constructive advices and suggestions during my PhD research. I must also thank all my colleagues in my research group for their friendship and support. Most particularly, I would like to thank Mehmet Cagri Ilter for always being there when I needed a good laugh or some help.

There are no words can express how deeply grateful I am to my family members. A special gratitude goes to Gizem Altun Sokun, my lovely wife and my best friend, for her patience, encouragement and devoted support. And certainly, I owe a special thank to my little girl, Yaren Sokun, for her understanding. Every day I spend with her is a blessing for me. I must also thank my parents-in-law, Aynur and Murat Altun, for their encouragement and support. Last but not least, I would like to exhibit my sincere gratitude to my sisters, Dilek, Demet, and Buse Sokun, and my parents, Gul and Murat Sokun, for their unconditional love and trust.

Contents

List of Tables

List of Figures

Nomenclature

Acronym	Meaning
EE	Energy Efficiency
SE	Spectral Efficiency
BS	Base-Station
RB	Resource Block
5G	Fifth Generation Wireless
LTE	Long Term Evolution
QoS	Quality-of-Service
SDR	Semidefinite Relaxation
PSD	Positive Semidefinite
IPM	Interior Point Method
SNR	Signal-to-Noise Ratio
SINR	Signal-to-Interference-plus-Noise Ratio
QCQP	Quadratically Constrained Quadratic Programming
3GPP	3rd Generation Partnership Project
OFDMA	Orthogonal Frequency Division Multiple Access
HetNet	Heterogeneous Network
Multi-RAT	Multiple Radio Access Technology

Chapter 1

Introduction

1.1 Research Motivation

Since the day the first mobile phone was developed by Motorola in 1973, mobile phones have evolved significantly over the past forty-five years. The initial purpose of these devices was only to communicate with another remote mobile phone. Today's smartphones have diverse features, such as messaging, internet, and gaming, and they have become an indispensable part of our daily life with a wide range of emerging applications and services, e.g., GPS navigation, e-healthcare, and mobile video conference. However, these advanced applications and services typically require higher data rates, and hence, entail a significant rise in mobile data traffic. Another important factor contributing to the rise in mobile data traffic is the increase in the number of smartphones (i.e., users who subscribe to mobile broadband), too many users demanding too much data. Numerically speaking, according to [1], there will be 8.2 billion mobile devices with 50 percent of these mobile devices will be smartphones by 2020. Additionally, overall mobile data traffic will grow to 30.6 exabytes per month by 2020, and the smartphones will account for 81 percent of mobile data traffic.

Supporting such a massive growth in mobile data traffic is a serious challenge. To meet this challenge, network capacity needs to significantly increase. For increasing network capacity, there are three fundamental approaches. The first approach is to have additional radio spectrum, if available. However, licensed spectrum is a limited resource, and controlled by regulatory bodies. The next one is to improve spectral efficiency (SE) at link level, i.e., having more capacity for the same amount of spectrum (measured in bits per second per Hz (bps/Hz)). To that end, for instance,

5G systems are expected to incorporate link improvement techniques such as coordinated multipoint (CoMP) transmission and reception, full-duplex communication, and massive multiple-input multiple-output (MIMO) [2]. The last approach is to increase the number of cells for spectrum reuse, i.e., improving network SE (measured in bps/Hz per unit area). For improvement of network SE, multi-tier heterogeneous network (HetNet) deployments, which incorporate low-cost low-power base stations (BSs) (micro, pico and femto) to assist a macro BS in the same spectrum, are considered to be the key technology. These deployments not only bring access network closer to end users, but also, provide increased spectrum reuse, and eliminate coverage holes in networks [3].

The other challenge raised by exponential growth in the demand of wireless services is escalation of energy consumption. The escalation of energy consumption causes not only environmental degradations, i.e., increasing greenhouse gas emission, but also, severe economical impacts, i.e., large electricity bills for operators [4]. To alleviate these negative impacts, limited energy resources in networks need to be utilized in the most efficient manner. That being said, driven by environmental and financial concerns, energy efficiency (EE) has become one of the key design objectives in wireless networks. Taking into account that the radio access part of wireless networks accounts for a significant part of total energy consumption [5], enhancing EE (measured in bits per Joule (bits/Joule)) of radio access networks can be an effective approach for improving overall network EE.

On another front, enhancing EE is also critically important for battery-equipped user devices. As the user devices are typically powered by pre-charged batteries [6], high EE in these devices can enable longer battery life. Hence, for all users, it is important to have the highest EE possible to prolong the battery life of their devices. However, this will lead to a problem of EE fairness (balancing) among users, particularly in uplink transmissions, as each individual will want to maximize their own benefits. For current and next-generation wireless networks, ensuring EE fairness among users is another essential design requirement.

Regardless of employed wireless network architecture, i.e., cellular networks or HetNets, and design objective under consideration, i.e., enhancement of SE, EE, or EE fairness, various network functionalities need to be taken into consideration to be able to achieve the desired enhancement in the network performance. For instance, the way in which users are associated to both BSs and time-frequency resource blocks

(RBs) available to the network can have a fundamental impact on the number of users that can be accommodated in the network. Likewise, the way in which BSs allocate their available power across RBs can have a significant impact on the amount of energy required for the network to attain a particular objective.

In this thesis, we address the questions of how to associate users to BSs, and how to allocate network resources, i.e., power levels and RBs, among users in order to enhance SE, EE, and EE fairness. Moreover, we focus on how to bridge the divide between the radio resource allocation techniques proposed in academia and the practical implementation. In particular, we concentrate on enhancement of the design metrics of SE, EE, and EE fairness in orthogonal frequency division multiple access (OFDMA)-based wireless networks, while considering practical design issues, such as discrete power levels, and per-user minimum rate requirements. We begin by investigating overall EE maximization problem in downlink transmission of cellular networks. Next, we study max-min EE fairness problem in uplink transmission of cellular networks. In these problems, EE is defined as the ratio of the total capacity to the total power consumption. That being said, EE accounts for how efficiently energy is used, but does not address how efficiently radio spectrum is utilized. Later, we shift our focus to SE maximization problem in HetNets. In that problem, the concept of SE specifies the maximum number of users accommodated in the network with the minimum network resources. We formulate all design problems as combinatorial optimization problems. Subsequently, to solve the optimization problems, we adapt a technique that constitutes a modified version of the semidefinite relaxation (SDR) with Gaussian randomization. This technique was developed in [7] to provide an approximate solution for the NP-hard maximum cut problem, and is guaranteed to yield an objective value that is at least 87.56 percent of the optimal value. Following [7], variants of SDR with Gaussian approximation have been developed to solve a wide range of combinatorial optimization problems, particularly quadratically constrained quadratic programs (QCQP) [8]. The rationale of SDR with Gaussian randomization is to convert a deterministic optimization problem into a stochastic one via the covariance matrix obtained in the (convex) SDR formulation. To be specific, this technique generates samples from a multivariate Gaussian distribution with the aforementioned covariance matrix, and rounds these samples, verifies feasibility and selects the candidate that yields the largest objective.

1.2 Thesis Overview

1.2.1 Energy-Efficient Resource Allocation in Cellular Networks: More Throughput, Less Energy

Most of the resource allocation literature on energy-efficient wireless networks assume continuous power allocation/control, while in practice the power levels are discrete (such as in 3GPP Long Term Evolution (LTE)). This convenient continuous power assumption has mainly been due to either the limitations of the used optimization tools and/or the high computational complexity involved in addressing the more realistic discrete power allocation/control.

In Chapter 2, we introduce a new optimization framework to maximize the overall EE of the downlink transmission of cellular networks subject to power budget and QoS constraints, while considering discrete power and RBs allocations. The proposed framework consists of two parts: 1) we model the predefined discrete power levels and RBs allocations by a set of binary variables, and 2) we propose a close-to-optimal semidefinite relaxation algorithm with Gaussian randomization to efficiently solve this non-convex combinatorial optimization problem with polynomial time complexity. We notice that a small number of the power levels suffice to approach the EE performance of the continuous power allocation. Based on this observation, we propose an iterative suboptimal heuristic to further reduce the computational complexity. Simulation results show the effectiveness of the proposed schemes in maximizing EE, while considering practical users' quality-of-service (QoS) demands and discrete power levels.

In Chapter 3, we focus on the fairness-oriented energy-efficient resource allocation. Due to the battery-limited nature of mobile devices, improving EE of individual users and ensuring EE fairness among those users are one of the key design issues in uplink transmission of cellular networks.We consider the joint optimization of discrete power and resource blocks allocations to maximize the minimum EE among users subject to individual power budget constraints. The optimization problem is combinatorial. Thus, we propose an efficient algorithm, based on SDR with Gaussian randomization, to solve the resultant non-convex problem in polynomial time complexity. The numerical results show how well the proposed algorithm performs against the optimal one, and indicate the impact of discrete power levels on the fairness-oriented EE optimization.

1.2.2 Spectrally-Efficient Resource Allocation in Heterogeneous Networks: More Users, Less Resource

In Chapter 4, we consider how to determine a close-to-optimal user-to-BS association that maximizes the number of users served by the downlink of a HetNet. Such an association must not only ensure that the number of accommodated users is maximized but also that the network resources are efficiently utilized and the users' QoS demands are met. The optimization problem that underlies this association is combinatorial and the difficulty of solving it is aggravated in HetNets by the disparity of transmit powers and BS computational capabilities. To find close-to-optimal user-to-BS associations, we develop a two-phase method based on SDR with randomization, which is a powerful method for solving a class of combinatorial problems. Unlike the majority of other polynomial complexity techniques, SDR has a provable approximation accuracy. Numerical examples show that, in comparison with other user-to-BS association approaches, the one proposed herein enables more efficient utilization of resources and a significantly higher number of users to be accommodated.

In Chapter 5, we consider joint optimization of user-to-BS association, and RB and power allocation in HetNets. The objective is to develop a design: 1) that maximizes the number of users accommodated in the network while satisfying their QoS demands; and 2) that minimizes usage of the resources required to meet these demands. We investigate two novel instances of HetNets with opportunistic RB-reuse. In the first instance, user-to-BS associations, and power allocations can be time-shared, and the RBs can be reused during the signalling interval. For this instance, it is shown that the design problem can be approximated by a problem that yields tight convex upper and lower bounds on the objective. In contrast, the second instance represents a case in which the RBs can be reused, but the user-to-BS associations, and power allocations are not time-shared, and hence, fixed throughout the signalling interval. The latter case gives rise to a combinatorial optimization problem, which we provide an approximate solution for by using a polynomial complexity two-phase approach based on SDR with randomization.

1.3 Thesis Contributions

Having motivated the work, and provided an overview of the design problems, we now give a brief account of the contributions. Broadly speaking, the key contributions of this thesis are two folds:

- We develop a novel design that incorporates user-to-BS association, RB assignment, and power allocation in both time-sharing and no time-sharing instances of HetNets. The developed design is the first to attempt optimizing these three network functionalities jointly when RBs can be reused, and user-to-BS associations and power allocations can be time-shared. To cast the joint design in an optimization form, we provide a comprehensive framework in which practical design issues, such as RB-reuse, discrete power levels, QoS requirement per-user, and multi-connectivity of single users to multiple BSs, are considered. It is shown that the framework developed herein can be applied to various resource allocation problems in wireless networks.

- The developed optimization framework is amenable to SDR-based optimization. The SDR-based techniques have polynomial-time complexity, and for quadratically-constrained problems, yield close-to-optimal performance [8]. Hence, they have been extensively used for signal processing and physical-layer communications problems. However, the application of these techniques in resource allocation problems is limited. In this thesis, the optimization of resource allocation problems is performed using techniques based on the SDR with Gaussian randomization.

1.4 Publications

Chapters 2-5 present the contributions that have been published, or accepted, whereas Chapter 6 summarizes the contributions, and also, discusses possible extensions.

- The contents of Chapter 2 are published in the following journal paper:

[**J1**] H. U. Sokun, E. Bedeer, R. H. Gohary, and H. Yanikomeroglu, **"Optimization of discrete power and resource block allocation for achieving maximum energy efficiency in OFDMA networks,"** *IEEE Access*, vol. 5, pp. 8648-8658, May 2017.

- The contents of Chapter 3 are published in the following conference paper:

[**C1**] H. U. Sokun, E. Bedeer, R. H. Gohary, and H. Yanikomeroglu, **"Fairness-oriented resource allocation for energy efficiency optimization in uplink OFDMA networks,"** in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018. (acceptance: Dec. 13, 2017)

- The contents of Chapter 4 are published in the following conference paper:

[**C2**] H. U. Sokun, R. H. Gohary, and H. Yanikomeroglu, **"QoS-guaranteed user association in HetNets via semidefinite relaxation,"** in *Proc. IEEE Veh. Tech. Conf. (VTC-Fall)*, pp. 1-5, Sept. 2015.

- The contents of Chapter 5 are published in the following journal paper:

[**J2**] H. U. Sokun, R. H. Gohary, and H. Yanikomeroglu, **"A novel approach for QoS-aware joint user association, resource block and discrete power allocation in HetNets,"** *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7603-7618, Nov. 2017.

Appendix A provides a list of publications produced during the PhD research, but not included in the thesis as they are outside the scope of the thesis.

1.5 Notation

Throughout the work we use bold-faced upper case letters, e.g., $\boldsymbol{X}$, to denote matrices, bold-faced lower case letters, e.g., $\boldsymbol{x}$, to denote column vectors, light-faced italics letters, e.g., x, to denote scalars, and calligraphic letters, e.g., $\mathcal{X}$, to denote sets. $\boldsymbol{I}$ denotes the identity matrix. The vectors of all-ones and all-zeros are denoted by $\mathbf{1}$ and $\mathbf{0}$, respectively, and for ease of exposition, we drop the subscript indicating the dimension of the all-one and the all-zero vectors and matrices. The trace, the rank, and the column vector consisting of the diagonal elements of matrix $\boldsymbol{X}$ are respectively denoted by $\mathrm{Tr}(\boldsymbol{X})$, $\mathrm{rank}(\boldsymbol{X})$, and $\mathrm{diag}(\boldsymbol{X})$. Lastly, $[\cdot]^{\mathrm{T}}$ denotes the transpose operator, $\mathrm{sgn}(\cdot)$ denotes the element-wise signum function, $\mathcal{N}(\cdot, \cdot)$ denotes Gaussian distribution with a particular mean and variance, and $\mathrm{vec}(\cdot)$ denotes the operator that stacks the columns of a matrix on top of each other.

Chapter 2

Optimization of Discrete Power and Resource Block Allocation for Achieving Maximum Energy Efficiency in Downlink Cellular Networks

2.1 Introduction

Cellular communications plays an undeniable role in the daily lives of millions of people worldwide. The demands on data rates are growing exponentially mainly due to smartphones, which are always connected to the cellular network during the day. Due to the hike in energy consumption costs, and ecological, and environmental reasons the increasing demands of data rate cannot be achieved by simply scaling up the transmit power. Instead, this has to be achieved at similar or lower energy consumptions. That said, energy-efficient communications have received a lot of attention from both industry and academia in recent years [9–11].

Orthogonal frequency division multiple access (OFDMA) is adopted in many contemporary wireless standards [12] due to multi-user and frequency diversities. In OFDMA, the frequency spectrum is divided into multiple subcarriers where different groups of subcarriers may be allocated for the transmission of different users depending on the varying channel conditions. Frequency diversity is achieved by activating only subcarriers that can support high quality transmission and nulling subcarriers with poor channel conditions. On the other hand, multi-user diversity is achieved with appropriate user mapping, i.e., a subcarrier may not be assigned to a certain user if the channel between this user and the BS on this subcarrier is in deep fade

and in this case the subcarrier should be assigned to a different user.

Most of the energy-efficient resource allocation algorithms reported in the literature, for various OFDMA system models, are based on continuous power allocation rather than allocation of discrete power levels [13–15]. Additionally, their solution techniques are mainly based on the fractional properties of the energy efficiency (EE) maximization problem [16]. In particular, the EE fractional objective function is transformed into an equivalent weighted sum of rate and power objectives. Then, dual Lagrangian method is applied to achieve the global EE optimal solution in an iterative manner. However, the complexity of finding the Lagrange multipliers (associated with the continuous power allocation) of the dual Lagrangian method is in general of unknown computation complexity [17].

Discrete power control/allocation simplifies the transmitter design, and also, significantly reduces the signalling overhead among nodes [18]. The authors in [19] considered the discrete power and subcarrier allocations to maximize the transmission rate of an OFDMA system, and solved the binary rate maximization problem using concepts of dynamic programming. In [20], the authors investigated the uplink transmission of contention-based synchronization in OFDMA systems, and formulated a constrained finite non-cooperative game to maximize the EE, where each mobile station has its own discrete power levels. For single carrier transmission [21], the authors considered predefined discrete power levels at the BS and proposed a reduced complexity algorithm that maximizes the EE of multi-cell networks. The formulated problem is classified as a fractional discrete optimization problem that is NP-hard to solve. The structure of the fractional problem is investigated and a suboptimal algorithm was proposed to attain an acceptable solution with polynomial time complexity.

2.1.1   Contribution

In this work, we investigate the EE resource allocation problem of the downlink transmission of OFDMA networks subject to power budget and per user quality-of-service (QoS)[1] constraints, while considering practical design issues, i.e., discrete power levels. Such a constraint adds another dimension to the difficulty of the EE maximization problem and we are going to efficiently address in this work. The main contributions of this work are summarized as follows.

---

[1]In this work, the QoS is defined in terms of the minimum data rate that a user requires.

- We introduce a novel optimization framework to efficiently handle such EE maximization problem with discrete power levels and RBs. This framework consists of the following two parts:

  - We model the discrete power levels and discrete RBs by a set of binary variables. We then show that the formulated EE maximization problem is combinatorial non-convex problem that turns out to be NP-hard to solve.

  - To tackle such a non-convexity, we propose a two stage close-to-optimal semidefinite relaxation (SDR)-based algorithm with Gaussian randomization, named COS, to efficiently solve this NP-hard problem with polynomial time complexity. In the first stage, the SDR generates a positive semidefinite covariance matrix together with an upper bound on the EE of the downlink transmission. In the second stage, using Gaussian randomization, we exploit the outputs of the first stage to compute good approximate solutions for the non-convex EE maximization problem with provable approximation accuracy.

- We notice that a small number of the discrete power levels is sufficient to approach the optimal EE performance of the continuous power allocation in [22]. Based on this observation, we propose a reduced-complexity iterative suboptimal heuristic, named SOH, that adopts a single power level.

- Extensive simulations results are provided to show the effectiveness of the proposed schemes in maximizing the EE of the downlink transmission. Results reveal that COS achieves the optimal performance of the exhaustive search. Additionally, results show that SOH strikes a balance between complexity and EE performance.

2.2 System Model

We consider a single-cell OFDMA network, in which a base station (BS) is located at the center of the cell. In this network, K uniformly distributed users communicate with the BS over N RBs[2], each of them with a bandwidth of W_0, and L discrete

---

[2]In 3GPP LTE networks [23], the BS allocates two-dimensional time-frequency resource units, among the users, i.e., scheduling RBs. An RB has a frequency bandwidth of 180 kHz and a time duration of one slot of 0.5 ms.

power levels available to the BS. We respectively denote the set of all users by $\mathcal{K} = \{1, \ldots, K\}$, the set of all RBs by $\mathcal{N} = \{1, \ldots, N\}$, and the set of power levels by $\mathcal{P} = \{p^1, \ldots, p^\ell, \ldots, p^L\}$, where $L = |\mathcal{P}|$ is the cardinality of $\mathcal{P}$. Moreover, we denote the channel gain between the BS and the k-th user on the n-th RB as h_k^n, which includes the path loss, shadowing, and small scale fading.

For such a network, we consider a centralized design, in which a central node collects network parameters and decides on the allocation of RBs between the users, as well as the allocation of power levels over the RBs. Due to such a central node, we assume that each RB is exclusively assigned to a user throughout the signalling interval [14, 24]. Hence, the signal-to-noise ratio (SNR) of the received signal by the k-th user on the n-th RB using the ℓ-th power level can be given as

$$\Gamma_k^{n\ell} = \frac{p^\ell |h_k^n|^2}{W_0 N_0}, \tag{2.1}$$

where N_0 is the power spectral density of the additive white Gaussian noise (AWGN). Hence, the maximum data rate that can be reliably communicated between the BS and the k-th user on the n-th RB using ℓ-th power level is expressed as

$$r_k^{n\ell} = W_0 \log_2 \left(1 + \Gamma_k^{n\ell}\right). \tag{2.2}$$

We note that for a known triplet, (k, n, ℓ), the SNR, $\Gamma_k^{n\ell}$, and the data rate, $r_k^{n\ell}$, can be readily calculated, and they have deterministic values.

2.3 Energy Efficiency Maximization Problem

In this section, we propose an optimization framework that jointly optimizes the RBs and the discrete power allocations to maximize the EE of the downlink transmission of OFDMA systems. We first introduce the RB usage, the power budget and the QoS constrains, and the design objective considered in the system. We then present the EE maximization problem formulation.

2.3.1 System Constraints

2.3.1.1 RB Usage Constraints

We use an indicator variable, $\phi_k^{n\ell}$, and let it represent whether the triplet of (k, n, ℓ), $k \in \mathcal{K}$, $n \in \mathcal{N}$, and $\ell \in \mathcal{P}$, is used for communication or not. If the k-th user is associated with the BS on the n-th RB using the ℓ-th power level, then $\phi_k^{n\ell} = 1$; otherwise, $\phi_k^{n\ell} = 0$. Hence,

$$\phi_k^{n\ell} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \ \forall n \in \mathcal{N}, \text{and } \forall \ell \in \mathcal{P}. \tag{2.3}$$

Using the defined binary variable, the total usage of the n-th RB across the entire network can be shown to be $\sum_{k \in \mathcal{K}} \sum_{\ell \in \mathcal{P}_b} \phi_k^{n\ell}$. To avoid interference, in the system model considered herein, each RB is constrained to be used at most once, and this constraint can be expressed as

$$\sum_{k \in \mathcal{K}} \sum_{\ell \in \mathcal{P}} \phi_k^{n\ell} \leq 1, \quad \forall n \in \mathcal{N}. \tag{2.4}$$

This constraint implies that at a given time instant, at most one power level can be used on each RB, and each RB can at most be used on one BS-to-user link.

2.3.1.2 Power Allocation Constraint

In a practical system, the total power consumption of the BS cannot exceed a maximum total power budget, $P^{\max}$. The total power consumed by the BS can be expressed as $\sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \sum_{\ell \in \mathcal{P}} \phi_k^{n\ell} p^\ell$. Hence, the power budget constraint is expressed as

$$\sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \sum_{\ell \in \mathcal{P}} \phi_k^{n\ell} p^\ell \leq P^{\max}. \tag{2.5}$$

Implicit in (2.5) and the one in (2.4) is that at most one non-zero power level is allowed on an RB at a given time instant.

2.3.1.3 QoS Constraints

Utilizing the expression given in (2.2), the maximum total data rate that the BS can reliably communicate with the k-th user can be expressed as $\sum_{n \in \mathcal{N}} \sum_{\ell \in \mathcal{P}} r_k^{n\ell} \phi_k^{n\ell}$. To ensure that the QoS requirement of the k-th user is met, the following minimum

supported rate constraint should be satisfied:

$$\sum_{n \in \mathcal{N}} \sum_{\ell \in \mathcal{P}} r_k^{n\ell} \phi_k^{n\ell} \geq r_k^{\min}, \quad \forall k \in \mathcal{K}, \tag{2.6}$$

where the QoS demanded by k-th user is denoted by $r_k^{\min}$.

2.3.2 System Design Objective

The EE of the network is defined as the ratio of the total data rate and the total consumed power, with unit of bits/Joule. The total BS power consumption is obtained as

$$P_{\mathrm{T}} = P_{\mathrm{C}} + \epsilon \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \sum_{\ell \in \mathcal{P}} \phi_k^{n\ell} p^\ell, \tag{2.7}$$

where P_{C} is the total circuitry power consumption required to deliver the information from the BS to the users, and ϵ is a constant defined by the inverse of the power amplifier efficiency. The total transmission data rate can be written as

$$r_{\mathrm{T}} = \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \sum_{\ell \in \mathcal{P}} \phi_k^{n\ell} r_k^{n\ell}. \tag{2.8}$$

Hence, the EE metric of the downlink transmission can be expressed as

$$\eta_{\mathrm{EE}} = \frac{r_{\mathrm{T}}}{P_{\mathrm{T}}}. \tag{2.9}$$

The design objective in this work is to maximize the EE metric given in (2.9).

2.3.3 Optimization Problem Formulation

Subsequently, we can cast the joint RB and power allocation problem for EE maximization in the downlink of an OFDMA network in the following form:

$$\max_{\phi_k^{n\ell}} \quad \eta_{\text{EE}} = \frac{\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{\ell\in\mathcal{P}} r_k^{n\ell}\phi_k^{n\ell}}{\epsilon \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{\ell\in\mathcal{P}} p^\ell \phi_k^{n\ell} + P_{\text{C}}}, \tag{2.10a}$$

$$\text{subject to} \quad \sum_{k\in\mathcal{K}}\sum_{\ell\in\mathcal{P}} \phi_k^{n\ell} \leq 1, \quad \forall n \in \mathcal{N}, \tag{2.10b}$$

$$\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{\ell\in\mathcal{P}} \phi_k^{n\ell} p^\ell \leq P^{\max}, \tag{2.10c}$$

$$\sum_{n\in\mathcal{N}}\sum_{\ell\in\mathcal{P}} r_k^{n\ell}\phi_k^{n\ell} \geq r_k^{\min}, \quad \forall k \in \mathcal{K}, \tag{2.10d}$$

$$\phi_k^{n\ell} \in \{0,1\}, \quad \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \text{and } \forall \ell \in \mathcal{P}. \tag{2.10e}$$

The formulation in (2.10) is an integer non-linear program. In particular, it is an integer linear fractional program due to the constraint in (2.10e), in which the optimization variables are restricted to be integer. For solving such a problem optimally, branch-and-bound type algorithms can be used. However, these algorithms have exponential complexity. We note that if the restriction on the variables is removed, then the EE problem in (2.10) becomes a linear fractional program and belongs to the class of quasiconvex programs[3] [25]. Hence, it can be solved efficiently using bisection method, in which a sequence of feasibility problems need to be solved.

2.4 Close-to-Optimal Resource Allocation: A Semidefinite Relaxation-Based Approach

For solving the non-convex problem in (2.10), we use the SDR technique with Gaussian randomization. This technique obtains close-to-optimal solutions for the EE problem in (2.10) with polynomial-time complexity[4]. A similar approach employed herein was considered in [26] for the user association problem in heterogeneous networks

---

[3]Since all constraints are linear, they comprise a convex set. Moreover, the objective function is quasi-linear in the variable $\phi_k^{n\ell}$ as its superlevel and sublevel sets are convex.

[4]In some particular cases, the SDR technique with Gaussian randomization can have a provable approximation accuracy. Finding the bound (quantifying the gap between the performance of the SDR-based technique and the optimal one) is in general an involved problem and it is out of the scope of the current work. However, a summary of some of the major approximation accuracy results is given in Tables I and II in [8].

to maximize the number of the accommodated users while to minimize the number of RBs used in the network. The problem considered in [26] was integer linear programming, in which apart from the integer constraints, the objective function and the constraints were linear. However, in this work, we consider joint optimization of RB and power allocation to maximize the EE in a macro-only network. Here, the problem in (2.10) is integer linear fractional programming, in which the objective function is quasi-convex, the constraints are linear, and the optimization variables are integer.

Before discussing the SDR technique with Gaussian randomization, we first express the EE, the RB usage, and the power budget, and the QoS constraints in vector form. For this purpose, we introduce a 3-dimensional tensor $\boldsymbol{\Phi}$ with entries denoted by $\phi_k^{n\ell}$. We express this tensor in the form of a $N \times KL$ block-partitioned matrix, and particularly, it is written as a matrix of $1 \times K$ blocks,

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \dots & \boldsymbol{\Phi}_K \end{bmatrix}, \tag{2.11}$$

where each blocks of that matrix has $N \times L$ entries, and for $j = 1, \dots, K$, $\boldsymbol{\Phi}_j$ is given as

$$\boldsymbol{\Phi}_j = \begin{bmatrix} \phi_j^{11} & \dots & \phi_j^{1L} \\ \vdots & \ddots & \vdots \\ \phi_j^{N1} & \dots & \phi_j^{NL} \end{bmatrix}. \tag{2.12}$$

We also introduce four additional 3-dimensional tensors: $\boldsymbol{A}_n$, $\boldsymbol{B}_k$, $\boldsymbol{C}$, and $\boldsymbol{D}$. Similar to $\boldsymbol{\Phi}$, these tensors are expressed in the form of a $N \times KL$ block-partitioned matrix with $1 \times K$ blocks, each with $N \times L$ entries. We define these four tensors as follows:

The tensor $\boldsymbol{A}_n$ can be written as

$$\boldsymbol{A}_n = \begin{bmatrix} \boldsymbol{A}_{n_1} & \dots & \boldsymbol{A}_{n_K} \end{bmatrix}, \quad n = 1, \dots, N, \tag{2.13}$$

where, for all $j = 1, \dots, K$, $\boldsymbol{A}_{n_j} = \boldsymbol{e}_n \mathbf{1}_L^T$, where $\boldsymbol{e}_n$ is the n-th column of the $N \times N$ identity matrix $\boldsymbol{I}_N$. Likewise, the tensor $\boldsymbol{B}_k$ can be expressed as

$$\boldsymbol{B}_k = \begin{bmatrix} \boldsymbol{B}_{k_1} & \dots & \boldsymbol{B}_{k_{\hat{k}}} & \dots & \boldsymbol{B}_{k_K} \end{bmatrix}, \quad k = 1, \dots, K, \tag{2.14}$$

where, for $j = 1, \ldots, K$, $\boldsymbol{B}_{k_j} = \boldsymbol{0}_{N \times L}$ when $j \neq \hat{k}$, and, when $j = \hat{k}$,

$$\boldsymbol{B}_{k_{\hat{k}}} = \begin{bmatrix} r_{\hat{k}}^{11} & \cdots & r_{\hat{k}}^{1L} \\ \vdots & \ddots & \vdots \\ r_{\hat{k}}^{N1} & \cdots & r_{\hat{k}}^{NL} \end{bmatrix}. \tag{2.15}$$

Similar to the tensor $\boldsymbol{B}_k$, the tensor $\boldsymbol{C}$ can be expressed as

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 & \cdots & \boldsymbol{C}_K \end{bmatrix}, \tag{2.16}$$

where, for $j = 1, \ldots, K$, $\boldsymbol{C}_j$ can be given as

$$\boldsymbol{C}_j = \begin{bmatrix} r_j^{11} & \cdots & r_j^{1L} \\ \vdots & \ddots & \vdots \\ r_j^{N1} & \cdots & r_j^{NL} \end{bmatrix}. \tag{2.17}$$

The tensor $\boldsymbol{G}$ can be written as

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}_1 & \cdots & \boldsymbol{G}_K \end{bmatrix}, \tag{2.18}$$

where, for $j = 1, \ldots, K$, $\boldsymbol{G}_j$ is given as

$$\boldsymbol{G}_j = \begin{bmatrix} p^1 & \cdots & p^L \\ \vdots & \ddots & \vdots \\ p^1 & \cdots & p^L \end{bmatrix}. \tag{2.19}$$

Lastly, we make these definitions: $\boldsymbol{\phi} \triangleq \text{vec}(\boldsymbol{\Phi}^T)$, $\boldsymbol{a}_n \triangleq \text{vec}(\boldsymbol{A}_n^T)$, $\boldsymbol{b}_k \triangleq \text{vec}(\boldsymbol{B}_k^T)$, $\boldsymbol{c} \triangleq \text{vec}(\boldsymbol{C}^T)$, and $\boldsymbol{g} \triangleq \text{vec}(\boldsymbol{G}^T)$.

Using the defined vectors, the problem in (2.10) can be expressed as

$$\max_{\boldsymbol{\phi}} \quad \frac{\boldsymbol{c}^T\boldsymbol{\phi}}{\epsilon\boldsymbol{g}^T\boldsymbol{\phi} + P_{\mathrm{C}}}, \tag{2.20a}$$

$$\text{subject to} \quad \boldsymbol{a}_n^T\boldsymbol{\phi} \leq 1, \quad \forall n \in \mathcal{N}, \tag{2.20b}$$

$$\boldsymbol{g}^T\boldsymbol{\phi} \leq P^{\max}, \tag{2.20c}$$

$$\boldsymbol{b}_k^T\boldsymbol{\phi} \geq r_k^{\min}, \quad \forall k \in \mathcal{K}, \tag{2.20d}$$

$$\boldsymbol{\phi} \in \{0, 1\}^{KNL}. \tag{2.20e}$$

To express this problem in a form amenable to SDR, we transform the binary optimization variables to antipodal ones. In particular, we introduce the vector $\boldsymbol{\theta} = 2\boldsymbol{\phi} - \mathbf{1}$, which implies that $\boldsymbol{\theta} \in \{-1, 1\}^{KNL}$, and

$$\boldsymbol{\phi} = \frac{1}{2}(\boldsymbol{\theta} + \mathbf{1}). \tag{2.21}$$

Using (2.21), the formulation in (2.20) can be rewritten as:

$$\max_{\boldsymbol{\theta}} \quad \frac{\boldsymbol{c}^T(\boldsymbol{\theta} + \mathbf{1})}{\epsilon\boldsymbol{g}^T(\boldsymbol{\theta} + \mathbf{1}) + 2P_{\mathrm{C}}}, \tag{2.22a}$$

$$\text{subject to} \quad \frac{1}{2}\boldsymbol{a}_n^T(\boldsymbol{\theta} + \mathbf{1}) \leq 1, \quad \forall n \in \mathcal{N}, \tag{2.22b}$$

$$\frac{1}{2}\boldsymbol{g}^T(\boldsymbol{\theta} + \mathbf{1}) \leq P^{\max}, \tag{2.22c}$$

$$\frac{1}{2}\boldsymbol{b}_k^T(\boldsymbol{\theta} + \mathbf{1}) \geq r_k^{\min}, \quad \forall k \in \mathcal{K}, \tag{2.22d}$$

$$\boldsymbol{\theta} \in \{-1, 1\}^{KNL}. \tag{2.22e}$$

To use the SDR-based technique, we consider a homogeneous reformulation of the problem in (2.22), and the optimization variables in (2.22) are constrained to be in the cone of symmetric positive semidefinite (PSD) matrices [8]. To do so, we define the following vectors in $\mathbb{R}^{KNL+1}$, $\hat{\boldsymbol{c}} \triangleq [\boldsymbol{c}^T \ \boldsymbol{c}^T\mathbf{1}]^T$, $\hat{\boldsymbol{g}} \triangleq [\epsilon\boldsymbol{g}^T \ \epsilon\boldsymbol{g}^T\mathbf{1} + 2P_{\mathrm{C}}]^T$, $\hat{\boldsymbol{a}}_n \triangleq [\boldsymbol{a}_n^T \ \boldsymbol{a}_n^T\mathbf{1}]^T$, $n = 1, \ldots, N$, $\hat{\boldsymbol{b}}_k \triangleq [\boldsymbol{b}_k^T \ \boldsymbol{b}_k^T\mathbf{1}]^T$, $k = 1, \ldots, K$, $\hat{\mathbf{1}} \triangleq [\mathbf{1}^T \ \mathbf{1}^T\mathbf{1}]^T$, $\hat{\boldsymbol{\theta}} \triangleq [\boldsymbol{\theta}^T \ 1]^T$ and $\hat{\boldsymbol{f}} \triangleq [\mathbf{0}^T \ 1]^T$. We also define the symmetric matrices $\boldsymbol{\Theta} \in \mathbb{R}^{KNL \times KNL}$ and $\boldsymbol{\Omega} \in \mathbb{R}^{(KNL+1) \times (KNL+1)}$ to be $\boldsymbol{\Theta} \triangleq \boldsymbol{\theta}\boldsymbol{\theta}^T$ and $\boldsymbol{\Omega} = \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^T$, in particular, $\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Theta} & \boldsymbol{\theta} \\ \boldsymbol{\theta}^T & 1 \end{bmatrix}$. Finally, we define the following $(KNL+1) \times (KNL+1)$ matrices $\boldsymbol{H}_1 \triangleq \hat{\boldsymbol{f}}\hat{\mathbf{1}}^T$, $\boldsymbol{H}_{\hat{c}} \triangleq \hat{\boldsymbol{f}}\hat{\boldsymbol{c}}^T$,

$\boldsymbol{H}_{\hat{a}_n} \triangleq \hat{\boldsymbol{f}} \hat{\boldsymbol{a}}_n^T$, and $\boldsymbol{H}_{\hat{g}} \triangleq \hat{\boldsymbol{f}} \hat{\boldsymbol{g}}^T$, $\boldsymbol{H}_{\hat{b}_k} \triangleq \hat{\boldsymbol{f}} \hat{\boldsymbol{b}}_k^T$.

Using the defined matrices, it can be verified that the problem in (2.10) is equivalent to the following optimization problem:

$$\max_{\boldsymbol{\Omega}} \quad \frac{\mathrm{Tr}(\boldsymbol{H}_{\hat{c}} \boldsymbol{\Omega})}{\mathrm{Tr}(\boldsymbol{H}_{\hat{g}} \boldsymbol{\Omega})}, \tag{2.23a}$$

$$\text{subject to} \quad \frac{1}{2} \mathrm{Tr}(\boldsymbol{H}_{\hat{a}_n} \boldsymbol{\Omega}) \leq 1, \quad \forall n \in \mathcal{N}, \tag{2.23b}$$

$$\frac{1}{2} \mathrm{Tr}(\boldsymbol{H}_{\hat{g}} \boldsymbol{\Omega}) \leq \epsilon P^{\mathrm{max}} + P_{\mathrm{C}} \tag{2.23c}$$

$$\frac{1}{2} \mathrm{Tr}(\boldsymbol{H}_{\hat{b}_k} \boldsymbol{\Omega}) \geq r_k^{\mathrm{min}}, \quad \forall k \in \mathcal{K}, \tag{2.23d}$$

$$\boldsymbol{\Omega} \succeq \boldsymbol{0}, \tag{2.23e}$$

$$\mathrm{diag}(\boldsymbol{\Omega}) = 1, \tag{2.23f}$$

$$\mathrm{rank}(\boldsymbol{\Omega}) = 1. \tag{2.23g}$$

The formulation in (2.23) is non-convex due to the rank-1 constraint in (2.23g). To find a close-to-optimal solution, we consider a relaxed version of (2.23) by dropping rank-1 constraint. Let $\boldsymbol{z}$, $\boldsymbol{Z}$ and $\boldsymbol{M}$ be the optimization variables of the relaxed problem corresponding to $\boldsymbol{\theta}$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Omega}$ in the original problem in (2.23), respectively. We then end up with the following formulation

$$\max_{\boldsymbol{M}} \quad \frac{\mathrm{Tr}(\boldsymbol{H}_{\hat{c}} \boldsymbol{M})}{\mathrm{Tr}(\boldsymbol{H}_{\hat{g}} \boldsymbol{M})}, \tag{2.24a}$$

$$\text{subject to} \quad (2.23b) - (2.23f). \tag{2.24b}$$

Though the formulation in (2.24) is still non-convex in $\boldsymbol{M}$, it is quasi-convex due to its linear fractional objective structure. It is known that a quasi-convex problem can be optimally solved using the bisection method. Hence, we solve a sequence of convex feasibility problems, and these feasibility problems are in the following form:

$$\text{find} \quad \boldsymbol{M}, \tag{2.25a}$$

$$\text{subject to} \quad \mathrm{Tr}((\eta_0 \boldsymbol{H}_{\hat{g}} - \boldsymbol{H}_{\hat{c}}) \boldsymbol{M}) \leq 0, \tag{2.25b}$$

$$(2.23b) - (2.23f). \tag{2.25c}$$

For each instance of this problem, the value of η_0 is fixed and represents the EE in the network. The optimal η_0 must lie in $[0, \eta_{\mathrm{max}}]$, where $\eta^{\mathrm{max}} = \boldsymbol{c}^T \boldsymbol{1} / P_{\mathrm{C}}$.

2.4.1 Gaussian Randomization

In the previous section, the optimal value of η_0 is obtained using bisection search. Let $\boldsymbol{z}^*$, $\boldsymbol{Z}^*$ and $\boldsymbol{M}^*$ be the optimal solution of the convex problem in (2.25) corresponding the optimal value of η_0. Since the solution of the relaxed problem is not rank-1 in general, we use the Gaussian randomization approach to obtain a close-to-optimal solution for the problem in (2.23). In the Gaussian randomization approach, the vector $\boldsymbol{z}^*$ generated by solving the relaxed program is considered as the mean of a multivariate Gaussian KNL-dimensional random vector, and $\boldsymbol{Z}^* - \boldsymbol{z}^*\boldsymbol{z}^{*T}$ is considered as the covariance matrix of this random vector. Specifically, a set of J random vector samples is drawn from the Gaussian distribution with mean $\boldsymbol{z}^*$ and covariance $\boldsymbol{Z}^* - \boldsymbol{z}^*\boldsymbol{z}^{*T}$. We denote the set by $\mathcal{R} = \{\boldsymbol{\nu}^j\}_{j=1}^J$, where $\boldsymbol{\nu}^j \sim \mathcal{N}(\boldsymbol{z}^*, \boldsymbol{Z}^* - \boldsymbol{z}^*\boldsymbol{z}^{*T})$, $j = 1, \dots, J$. Letting $\hat{\boldsymbol{\nu}} = [\boldsymbol{\nu}^T\ 1]^T$, and $\hat{\boldsymbol{z}}^* = [\boldsymbol{z}^{*T}\ 1]^T$, it can be seen that the vectors in $\mathcal{R}$ provide an approximate solution to the following stochastic optimization problem:

$$\max_{\substack{\mathbb{E}\{\hat{\boldsymbol{\nu}}\hat{\boldsymbol{\nu}}^T\} = \boldsymbol{M}^*, \\ \mathbb{E}\{\hat{\boldsymbol{\nu}}\} = \hat{\boldsymbol{z}}^*}} \frac{\mathbb{E}\{\hat{\boldsymbol{\nu}}^T \boldsymbol{H}_{\hat{c}} \hat{\boldsymbol{\nu}}\}}{\mathbb{E}\{\hat{\boldsymbol{\nu}}^T \boldsymbol{H}_{\hat{g}} \hat{\boldsymbol{\nu}}\}}, \tag{2.26a}$$

$$\text{subject to} \quad \frac{1}{2}\mathbb{E}\{\hat{\boldsymbol{\nu}}^T \boldsymbol{H}_{\hat{a}_n} \hat{\boldsymbol{\nu}}\} \leq 1, \quad \forall n \in \mathcal{N}, \tag{2.26b}$$

$$\frac{1}{2}\mathbb{E}\{\hat{\boldsymbol{\nu}}^T \boldsymbol{H}_{\hat{g}} \hat{\boldsymbol{\nu}}\} \leq \epsilon P^{\max} + P_{\mathrm{C}}, \tag{2.26c}$$

$$\frac{1}{2}\mathbb{E}\{\hat{\boldsymbol{\nu}}^T \boldsymbol{H}_{\hat{b}_k} \hat{\boldsymbol{\nu}}\} \geq r_k^{\min}, \quad \forall k \in \mathcal{K}, \tag{2.26d}$$

$$\mathbb{E}\{\hat{\boldsymbol{\nu}}_i^2\} = 1, \ i = 1, \dots, KNL. \tag{2.26e}$$

It is important to mention that the Schur complement of the matrix $\boldsymbol{M}^*$ is $\boldsymbol{Z}^* - \boldsymbol{z}^*\boldsymbol{z}^{*T}$, and it is PSD.

We use the vectors in $\mathcal{R}$ to obtain candidate binary solutions $\{\tilde{\boldsymbol{\theta}}^j\}_{j=1}^J$ for the problem in (2.22) by quantizing the entries of each realization of $\{\boldsymbol{\nu}^j\}_{j=1}^J$. Specifically,

$$\tilde{\boldsymbol{\theta}}^j = \mathrm{sgn}(\boldsymbol{\nu}^j), \qquad j = 1, \dots, J. \tag{2.27}$$

Using (2.27) and (2.21), we obtain candidate binary solutions of (2.20), $\tilde{\boldsymbol{\phi}}^j$. The candidate that yields the largest objective and satisfies the constraints in (2.20) is

used for allocating the RBs and the discrete power levels, i.e.,

$$\boldsymbol{\phi}^* = \arg \max_{\mathcal{D}_j} \tilde{\boldsymbol{\phi}}^j, \ \mathcal{D}_j \triangleq \{\tilde{\boldsymbol{\phi}}^j : \ \tilde{\boldsymbol{\phi}}^j \text{ satisfying (2.20b)–(2.20d)}\}. \tag{2.28}$$

The proposed close-to-optimal algorithm based on SDR with randomization, viz., COS, is summarized in Algorithm 1.

Algorithm 1: Proposed COS.

Input: P_C, ϵ, $P^{\max}$, $r_k^{\min}$, J, κ, $\eta^{\min} = 0$, and $\eta^{\max} = \boldsymbol{c}^T \boldsymbol{x}/P_C$
Output: $\boldsymbol{\phi}^*$

1 **while** $\eta^{\max} - \eta^{\min} \geq \kappa$ **do**
2      $\eta_0 = 0.5 \, (\eta^{\min} + \eta^{\max})$.
3   Solve (2.25) to find $\boldsymbol{M}$.
4   **if** *feasible* **then**
5          $\eta^{\min} = \eta^{\max}$
6          $\boldsymbol{M}^* = \boldsymbol{M}$
7   **else**
8          $\eta^{\max} = \eta^{\min}$

9 Find $\boldsymbol{z}^*$, and $\boldsymbol{Z}^*$ from $\boldsymbol{M}^*$.
10 **for** $j = 1 : J$ **do**
11      $\boldsymbol{\nu}^j \sim \mathcal{N}(\boldsymbol{z}^*, \boldsymbol{Z}^* - \boldsymbol{z}^* \boldsymbol{z}^{*T})$
12      $\tilde{\boldsymbol{\theta}}^j = \text{sgn}(\boldsymbol{\nu}^j)$
13      $\tilde{\boldsymbol{\phi}}^j = 0.5(\tilde{\boldsymbol{\theta}}^j + \mathbf{1})$
14   **if** (2.20b)-(2.20d) *satisfied* **then**
15   Record $\tilde{\boldsymbol{\phi}}^j$.

16 $\boldsymbol{\phi}^* = \arg \max_{\mathcal{D}_j} \tilde{\boldsymbol{\phi}}^j$

2.4.2 Computational Complexity Analysis

The joint design problem in (2.20) can be optimally solved using exhaustive search with complexity $\mathcal{O}(2^{KNL})$, which is computationally prohibitive. In contrast with exhaustive search, the two-stage algorithm proposed herein, COS, has a polynomial-time complexity, and hence, it is suitable for solving large-scale RB and power allocation

problems. More specifically, the complexity of solving the relaxation of (2.25), a PSD-constrained convex problem, is $\mathcal{O}\left((KNL)^{3.5}\right)$[5].

The number of iterations required for the convergence of bisection method is $\log(\eta_{\max}/\kappa)$, where $\kappa > 0$, is the solution accuracy of the bisection method. Lastly, for the Gaussian randomization, the complexity of generating and evaluating the objective function corresponding to the J random samples is $\mathcal{O}\left((KNL)^2 J\right)$. Hence, the overall complexity of COS is $\mathcal{O}\left((KNL)^{3.5}\log(\eta_{\max}/\kappa) + (KNL)^2 J\right)$.

2.5 Suboptimal Resource Allocation: A Heuristic-Based Approach

In the previous section, we described the proposed COS that provides a lower computation complexity compared to exhaustive search. However, to further reduce the computational complexity, we propose an iterative suboptimal heuristic SOH that offers a trade-off between the computational complexity and performance.

To reduce the computational complexity involved in solving the EE maximization problem, SOH performs RBs and discrete power allocations, separately. For discrete power allocation, SOH considers a uniform single power level that is obtained with an offline search. The rationale behind this consideration is that as the BS transmit power increases power allocation across RBs tends to become uniform. It is intuitive that when the BS has a low transmit power, power allocation has a pivotal role and increasing the number of power levels affect the performance significantly, whereas it becomes less critical when the BS has a high transmit power. In [27] and [28], it is shown that uniform power allocation is sufficient to approach the maximum rate and maximum EE. Based on our simulation results in Section 2.6, we verify that a single power level can provide a desired level of EE. For instance, in Section 2.6, we provide an example for obtaining a tuned power level that can be then used for allocating the RBs among the users instead of large sets of power levels. Using this observation, when power allocation is assumed to be fixed, the overall complexity of COS given in Section 2.4.2 can be reduced to $\mathcal{O}\left((KN)^{3.5}\log(\eta_{\max}/\kappa) + (KN)^2 J\right)$. However, with the aim of further reducing computational complexity, we use a heuristic approach

---

[5]We note that this technique relies on constructing PSD matrices with a number of entries that scales with the square of the number of variables. The number of variables in a network with K users, N RBs, and L power levels is KNL. Hence, as the network size enlarges, simulation of the network becomes more challenging. Unless special-purpose computers are used, it can be time-consuming.

in SOH to determine the RB allocation between the users. We first assign each RB to the user with highest SNR on that RB until meeting the all users' QoS, as long as the BS power budget is not violated. Afterwards, the remaining RBs are allocated among the users in a greedy manner to maximize EE. In other words, the RB having the highest EE is assigned to corresponding user to maximize EE.

The proposed suboptimal heuristic, SOH, is summarized in Algorithm 2.

Algorithm 2: Proposed SOH.

Input: P_C, ϵ, p^1, P^{max}, $\Gamma_k^{n\ell}$, r_k^{min}, $\bar{\mathcal{K}} = \mathcal{K}$, and $\bar{\mathcal{N}} = \mathcal{N}$
Output: ϕ^*

1   **for** $n = 1 : |\bar{\mathcal{N}}|$ **do**
2   **if** $P^{\mathrm{max}} \geq p^1$ and $|\bar{\mathcal{K}}| \geq 0$ **then**
3       $k^* = \arg\max_{k \in \bar{\mathcal{K}}} \Gamma_k^{n\ell}$
4       $k^* \leftarrow n$
5       $\phi_{k^*}^{n1} = 1$
6       $P^{\mathrm{max}} = P^{\mathrm{max}} - p^1$
7       $\bar{\mathcal{N}} = \bar{\mathcal{N}} \backslash n$
8   **if** $r_{k^*}^{n1} \geq r_{k^*}^{\mathrm{min}}$ **then**
9        $\bar{\mathcal{K}} = \bar{\mathcal{K}} \backslash k^*$

10   **for** $n = 1 : |\bar{\mathcal{N}}|$ **do**
11   **if** $P^{\mathrm{max}} \geq p^1$ **then**
12   **for** $k = 1 : |\bar{\mathcal{K}}|$ **do**
13        $k \leftarrow n$
14        $\phi_k^{n1} = 1$
15   Calculate $\eta_{\mathrm{EE}}^{(k \leftarrow n)}$ using (2.10a).
16      $k^* = \arg\max_{k \in \bar{\mathcal{K}}} \eta_{\mathrm{EE}}^{(k \leftarrow n)}$
17      $k^* \leftarrow n$
18      $\phi_{k^*}^{n1} = 1$
19      $P^{\mathrm{max}} = P^{\mathrm{max}} - p^1$

2.5.1   Computational Complexity Analysis

The computational complexity of SOH can be analyzed as follows:

- Step 1 requires a complexity of $\mathcal{O}(N)$ to evaluate the **for** loop.

- Step 2 requires a complexity of $\mathcal{O}(K)$ to calculate the cardinality of the set of the unsatisfied users with QoS requirements, $\bar{\mathcal{K}}$.

- Step 3 requires a complexity of $\mathcal{O}(K)$ to find the user with the highest SNR on the respective RB.

- Step 7 requires a complexity of $\mathcal{O}(N)$ to remove the used RB from the set of the unused RBs, $\bar{\mathcal{N}}$.

- Step 9 requires a complexity of $\mathcal{O}(K)$ to remove the satisfied user with QoS requirement from the set of the unsatisfied users with QoS requirement.

- Step 10 requires a complexity of $\mathcal{O}(N-K)$ to evaluate the **for** loop for the remaining $(N-K)$ RBs at most.

- Step 12 requires a complexity of $\mathcal{O}(K)$ to evaluate the **for** loop.

- Step 16 requires a complexity of $\mathcal{O}(K)$ to find the user with the highest energy-efficiency on the respective RB.

Finally, Steps 1 to 9 require a complexity of $\mathcal{O}(N(3K+N))$, and Steps 10 to 19 require a complexity of $\mathcal{O}(2(N-K)K)$. Hence, the worst case computational complexity of SOH is calculated as $\mathcal{O}(N(3K+N)+2(N-K)K)$. As it can be seen that SOH has less complexity than COS.

2.6   Simulation Results and Discussions

In this section, we will present the simulation results to verify the proposed algorithms, and provide discussions regarding the obtained results. We consider a single-cell network, in which the BS is located in the center of the cell, and the location of users are randomly generated and uniformly distributed. We use the 3GPP propagation model [23]. According to [23], the path loss is assumed to be given by $PL(d) = 128.1 + 37.6 \log_{10}(d)$ for the link between the BS and the users, where d is the distance in kilometers. The shadowing component is assumed to have a log-normal distribution with a standard deviation of $\sigma_s = 8$ dB. The bandwidth of each RB is assumed to be 180 kHz, and the noise power spectral density is assumed to be -174 dBm/Hz. In addition, the efficiency of the power amplifiers for the BS is assumed to be 38% [29]. Furthermore, the set of power levels used by the BS is assumed to

consist of L equally spaced points in the interval $[P^{\text{lower}},\ P^{\text{upper}}]$, where P^{lower} and P^{upper} are the lower and upper bounds of that set.

We consider Monte Carlo simulations to evaluate the network performance. The reported simulation results are averaged over 100 independent channel realizations. For each channel realization, the solution to the problem of EE maximization in Section 2.4 is obtained using the CVX [30] with the SDPT3 solver, and the number of Gaussian samples used in the randomization stage is set to be $J = 10^4$.

Example 1. (*Lower and Upper Limits for the Set of Power Levels*) In Fig. 2.1, we consider a scenario, where the minimum rate requirement for all users is zero, the BS static power consumption is $P_C = 50$ dBm, the number of users is set to be $K = 4$, the number of RBs is set to be $N = 8$, and the number of power levels is $L = 2$.

Fig. 2.1 investigates the performance of COS with the given different sets of the two power levels. From this figure, it can be seen that the choice of the set of the power levels heavily influences the network performance, and a well-chosen set of power levels can enhance the EE of the system in a positive way. For instance, when $P^{\text{max}} = 45$ dBm, using the set of $\mathcal{P} = \{0.05P^{\text{max}},\ 0.5P^{\text{max}}\}$ rather than the set of $\mathcal{P} = \{0,\ P^{\text{max}}\}$ can improve the EE of the system from 3.436×10^4 bits/Joule to 2.104×10^5 bits/Joule. Another important observation is that the sets of $\mathcal{P} = \{0,\ 0.5P^{\text{max}}\}$, $\mathcal{P} = \{0.5P^{\text{max}},\ P^{\text{max}}\}$, and $\mathcal{P} = \{0.5P^{\text{max}},\ 0.75P^{\text{max}}\}$ show very close performances, since they tend to employ $0.5P^{\text{max}}$ for transmission. Moreover, the best performance for all values of the transmit power can be achieved using the set of $\mathcal{P} = \{0.05P^{\text{max}},\ 0.5P^{\text{max}}\}$. Based on these observations, for the following simulations, unless otherwise stated, the lower and upper bounds for the set of power levels is assumed to be $P^{\text{lower}} = 0.05P^{\text{max}}$, and $P^{\text{upper}} = 0.5P^{\text{max}}$, respectively. $\qquad\square$

Example 2. (*Performance Comparison With Optimal Solution*) In Fig. 2.2, we consider a relatively small scenario, where the minimum rate requirement for all users is zero, the BS static power consumption is $P_C = 50$ dBm, the number of users is $K = 3$, the number of RBs is $N = 4$, and the number of power levels is $L = 2$, viz., $\mathcal{P} = \{0.05P^{\text{max}},\ 0.5P^{\text{max}}\}$.

Fig. 2.2 compares the performance of COS with the optimal EE solution obtained by exhaustive search to validate COS. It is worth to mention that the baseline method looks at every possible triplets of (k, n, ℓ), $k \in \mathcal{K}$, $n \in \mathcal{N}$, and $\ell \in \mathcal{P}$, in order to find which one yields the maximum EE. As seen from Fig. 2.2, the algorithm proposed herein attains the optimal solutions for the entire range of P^{max}. $\qquad\square$

Example 3. (*Effect of Number of Power Levels*) In Fig. 2.3, we consider a scenario with the different number of power levels. In this scenario, the minimum rate requirement for all users is zero, the BS static power consumption is $P_C = 50$ dBm, the number of users is set to be $K = 4$ and the number of RBs is set to be $N = 8$.

Fig. 2.3 illustrates the effect of the number of discrete power levels on the EE. Here, we evaluate the performance of COS through numerical comparisons with the EE maximization scheme proposed in [22] that considers a joint design of the RBs and the continuous power allocation. From this figure, it can be observed that as the number of power levels available at the BS increases, the performance of COS is enhanced, and the gap between the discrete power allocation and the continuous one becomes smaller. It is worth to mention that an increase in the number of power levels helps to improve the performance, especially when the number of power levels is low. For instance, when the number of power level is increased from two to four, the gain in the EE is 4% at $P^{\max} = 38$ dBm, whereas when the number of power level is increased from four to eight, the gain in the EE is only 1.8% at $P^{\max} = 38$ dBm. Hence, in the subsequent figure, we consider four number of power levels at the BS. Another interesting observation is that the algorithm with a carefully chosen single power level reveals a quite good performance even at low $P^{\max}$ values, and its performance improves as $P^{\max}$ increases. Using this observation, we propose SOH that has not only a good performance, but also, a lower complexity. We will show the performance of SOH in comparison to COS below. □

Example 4. (*Effect of Circuitry Power Consumption*) In Fig. 2.4, we consider a scenario, where the minimum rate requirement for all users is zero, the number of users is $K = 8$, the number of RBs is $N = 12$, and the number of power levels is $L = 4$.

Fig. 2.4 investigates the impact of the BS static power consumption on the EE. From this figure it can be seen that the EE decreases with the increase of the BS circuitry power consumption, because transmitting data requires more total power. For instance, when $P^{\max} = 50$ dBm, increasing the circuitry power consumption from 40 dBm to 55 dBm decreases the EE from 1.175×10^6 bits/Joule to 1.137×10^5 bits/Joule for COS. Another important observation is that as the static power consumption increases, the network performance becomes less sensitive to a change in the transmit power, especially at the medium and high transmit power values. □

Example 5. (*Performance Comparison of Proposed Algorithms*) In Fig. 2.5, we consider a scenario, where the number of users is $K = 4$, the number of RBs is $N = 8$, and the number of power levels is $L = 4$.

Fig. 2.5 investigates the effect of increasing the transmit power on the performance of the proposed algorithms, viz., COS and SOH. Particularly, in Fig. 2.5(a), we study the effect of increasing the transmit power at different circuitry power consumptions without the effect being obscured by potentially different minimum rate requirements per users, while in Fig. 2.5(b), we study the effect of increasing the transmit power at different minimum rate requirements per users without the effect being obscured by potentially different circuitry power consumptions.

In Fig. 2.5(a), the EE is plotted as a function of the available power budget at the BS for different static power consumption values. In this figure, we assume that the minimum rate requirement is the same for all users, i.e., $r_k^{\min} = 1$ Mbps, $\forall k \in \mathcal{K}$. One can see that for high power budget values both COS and SOH provide almost identical results, for all static power consumption values. On the other hand, for low power budget values COS outperforms SOH. For instance, when $P^{\max} = 45$ dBm and $P_C = 50$ dBm, the gain achieved using COS over SOH is only 6%. Such trend suggests that the uniform power allocation assumed by SOH is a good approximation, especially for high BS transmit power.

In Fig. 2.5(b), the EE is plotted as a function of the available power budget at $P_C = 50$ dBm and for different minimum rate requirements per users. As it can be seen, the performance of the SOH scheme is close to that of the COS scheme for high power budget (although users have different minimum rate requirements), while for low power budget the EE of the COS scheme is higher than its counterpart of the SOH scheme. Fig. 2.5(b) with Fig. 2.5(a) suggest that the uniform power allocation assumption is directly related to the BS power budget. Hence, regardless of both the circuitry power consumption and the minimum rate requirements, the gap between COS and SOH becomes smaller as the BS transmit power increases. □

2.7   Conclusion

This work presented a new optimization framework that efficiently handle practical design issues, i.e., discrete power levels, in resource allocation problems in OFDMA networks. The proposed framework maximized the EE of the downlink transmission

of cellular OFDMA networks subject to RB usage, power budget and per-user QoS constraints. In particular, we modelled both the discrete power levels and discrete RBs by a set of binary variables. Then, we used SDR technique with Gaussian randomization to efficiently solve the combinatorial non-convex problem with polynomial time complexity. We noticed from the solution of the close-to-optimal SDR-based (COS) algorithm that a small number of discrete power levels is sufficient to approach the maximum EE performance of the continuous power allocation solution. Based on this observation, we proposed a low-complexity iterative suboptimal heuristic (SOH) algorithm that relies on a single power level. Simulation results showed that the EE of the COS algorithm approaches that of the exhaustive search. Additionally, the simulation results revealed that the SOH strikes a balance between the performance and complexity.
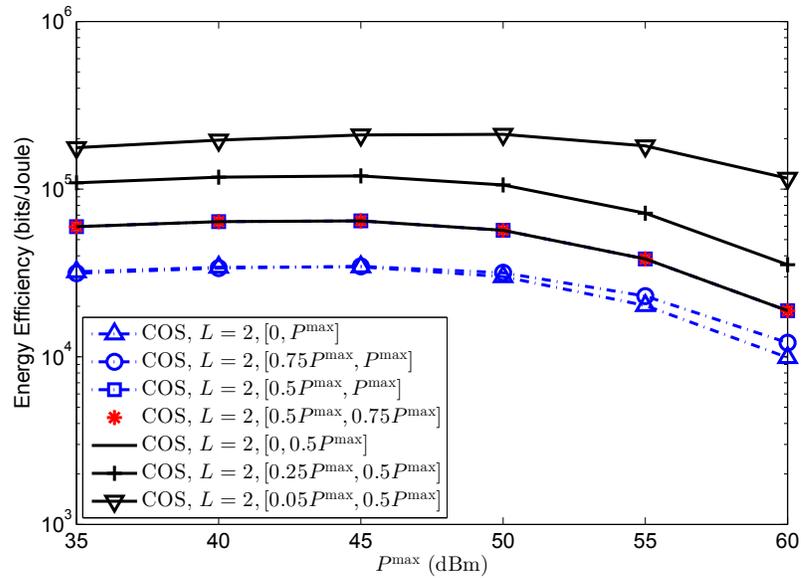
Figure 2.1: Performance of COS with different set of two power levels.
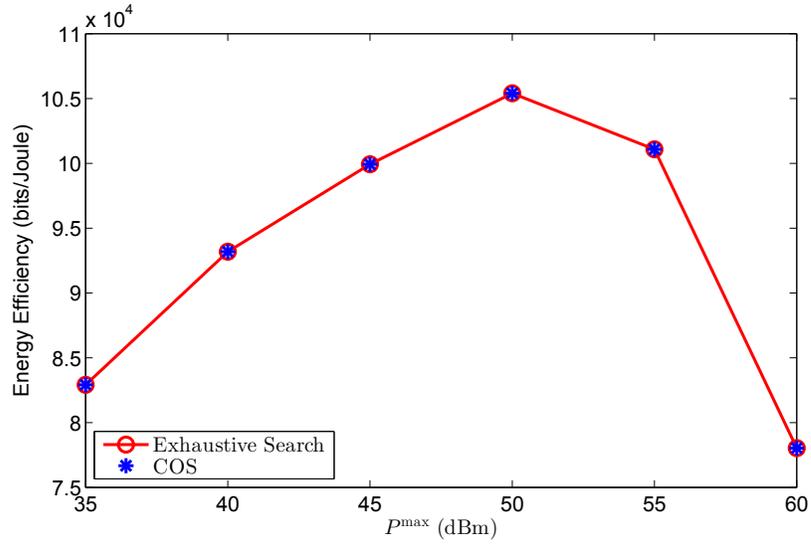
Figure 2.2: Performance of COS in comparison with exhaustive search.
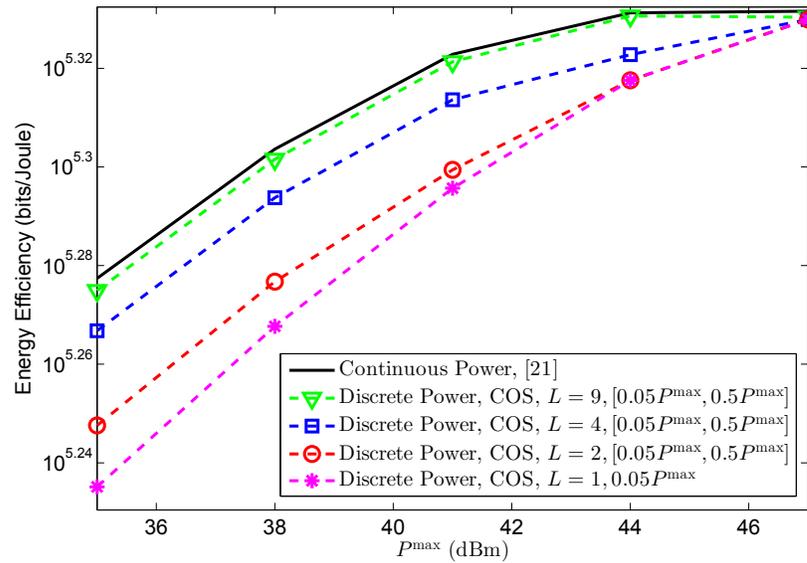
Figure 2.3: Effect of the number of power levels on the EE.
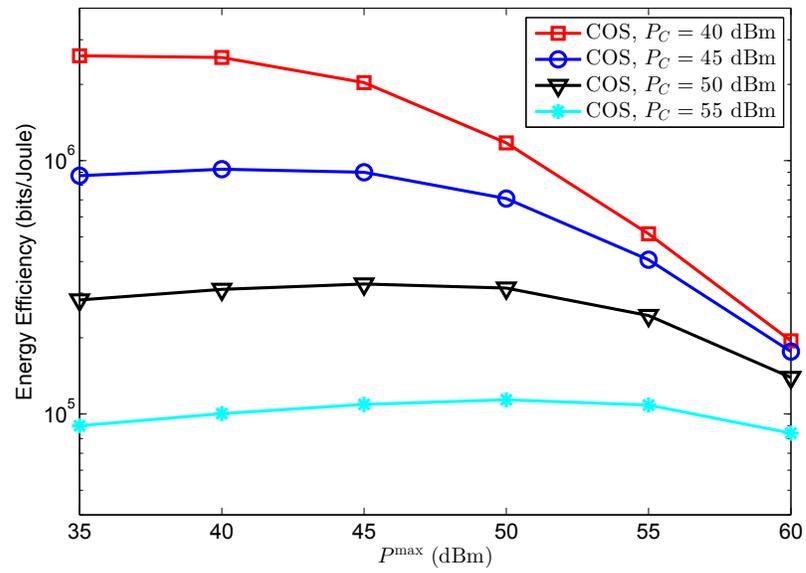
Figure 2.4: Effect of circuitry power consumption on the EE.
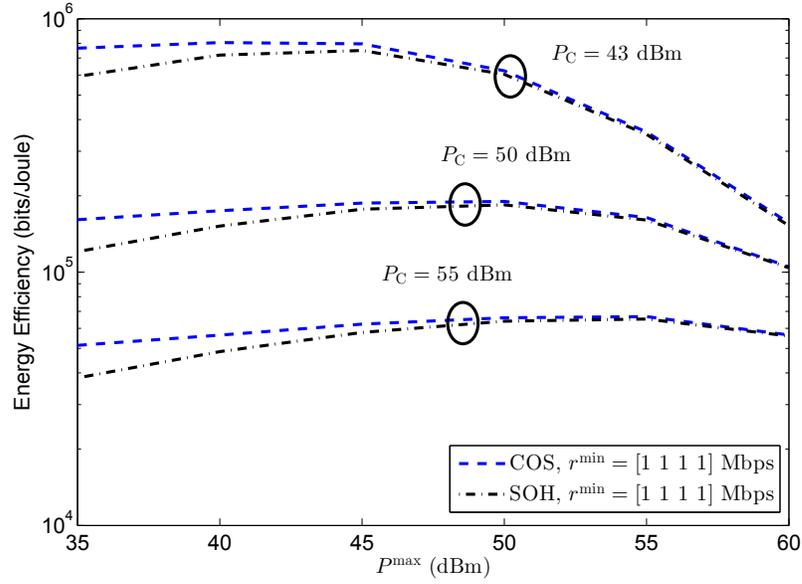
(a) Comparison of the proposed algorithms, considering different circuitry power consumption values.
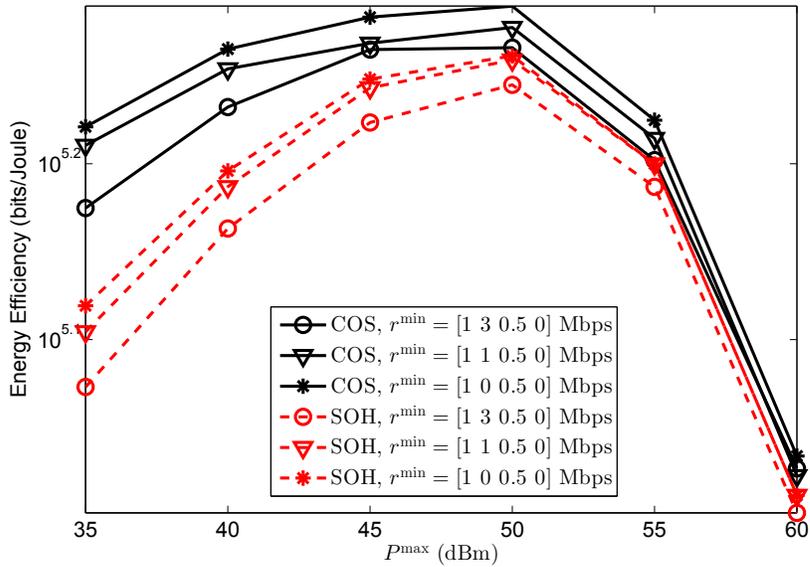
(b) Comparison of the proposed algorithms, considering different minimum rate requirements for the users.

Figure 2.5: Performance comparison of COS and SOH.

Chapter 3

Fairness-Aware Joint Resource Block and Discrete Power Allocation for Energy Efficiency Optimization in Uplink Cellular Networks

3.1 Introduction

Today's mobile devices become an irreplaceable part of our life with a wide range of new emerging applications and services, e.g., GPS navigation, e-healthcare, and mobile video conference. However, these advance applications and services require not only higher data rates [1], but also higher energy consumption. Such a rise in energy consumption adversely affects the battery life of mobile devices that are typically powered by pre-charged batteries. To prolong the battery life of the devices, for all users, it is important to have the highest energy efficiency (EE) possible. However, this will lead to a problem of EE fairness (balancing) among users, particularly in uplink transmissions, as each individual will want to maximize their own benefits. Hence, for current and next-generation wireless networks, not only improving users' EE, but also ensuring EE fairness among users are critically important design requirements.

Generally speaking, for improving the users' EE, their data rate needs to be increased, while at the same time, their energy consumption needs to be decreased. To achieve these objectives, resource allocation is an effective tool [4]. More specifically, optimizing the way how the available resource blocks (RBs) in the network are allocated among the users, and the way how the users allocate their available power

across those RBs can substantially enhance the users' EE [31]. For further enhancement in the users' EE, the interaction between these network functionalities should be considered.

3.1.1 Prior Work and Contribution

Despite the rich literature on the overall EE optimization in the downlink of an orthogonal frequency division multiple access (OFDMA)-based systems, fairness-oriented individual EE optimization in the uplink of OFDMA-based systems is not thoroughly investigated from resource allocation perspective. There are only a limited number of works in the literature studied this problem (see, e.g., [14, 32–34]). In [32], the power optimization problem with fixed RB allocation is considered for uplink transmission in OFDMA systems to maximize the minimum EE of all users. Similarly, in [33], the power optimization problem is discussed for uplink transmission in spectrum-sharing networks to achieve the proportional, harmonic, and max-min fair EE. On the other hand, the problem of the joint optimization of subcarrier and power allocation is addressed in [34] to achieve the max-min EE fairness in the uplink transmission of OFDMA networks. Lastly, in [14], the problem of the joint optimization of subcarrier and and power allocation is investigated to attain the max-min EE fairness in the uplink transmission of amplify-and-forward cooperative OFDMA networks. However, it is important to mention that the resource allocation techniques proposed in [14, 32–34] are based on continuous power allocation, rather than allocation of discrete power levels. Hence, implementing such resource allocation techniques in practical networks supporting discrete power levels, e.g., the the Long Term Evolution (LTE) networks [35], can be problematic.

We propose a joint resource allocation technique to maximize the minimum EE of all users in OFDMA networks, while taking practical implementation constraints into account, i.e., including discrete power levels and per-user power budget. In particular, we consider the joint design of RB and discrete power allocations, and formulate the joint design as a max-min EE optimization problem. Restricting the power levels for being discrete has several advantages, such as simplifying the hardware designs, and enabling a low cost implementation, however, this usually leads to non-convex formulations. To overcome this difficulty, we consider the semidefinite relaxation (SDR)-based Gaussian randomization technique that invokes a relaxation on the rank of the matrix-value optimization variable to obtain near-optimal results in polynomial time.

3.2 Problem Statement and Preliminaries

This section describes the network model and discusses the problem formulation.

3.2.1 System Description

We consider an uplink OFDMA network which consists of a base station (BS), M users, J RBs and L discrete power levels. Let $\mathcal{M} = \{1, \ldots, M\}$ denote the set of users, $\mathcal{J} = \{1, \ldots, J\}$ denote the set of RBs and $\mathcal{P} = \{p^1, \ldots, p^\ell, \ldots, p^L\}$ denote the set of power levels, where $L = |\mathcal{P}|$ is the cardinality of $\mathcal{P}$. Let g_m^j represent the channel gain for the m-th user on the j-th RB. Then, the maximum achievable rate for m-th user on the j-th RB when ℓ-th power level is used can be expressed as

$$r_m^{j\ell} = B \log_2 \left(1 + \frac{p^\ell |g_m^j|^2}{B N_0} \right), \tag{3.1}$$

where each RB has a bandwidth of B and N_0 is the power spectral density of additive white Gaussian noise. Additionally, let binary variable $x_m^{j\ell}$ be an indicator variable that indicates whether the m-th user is associated with the BS on the j-th RB using the ℓ-th power level or not. Using the indicator variables, the overall transmission rate for the m-th user can be cast as $\sum_{j \in \mathcal{J}} \sum_{\ell \in \mathcal{P}} r_m^{j\ell} x_m^{j\ell}$. Moreover, to avoid interference among RBs, each RB is allocated to a single user at most, so that total usage of the j-th RB over the network should satisfy

$$\sum_{m \in \mathcal{M}} \sum_{\ell \in \mathcal{P}} x_m^{j\ell} \leq 1, \quad \forall j \in \mathcal{J}. \tag{3.2}$$

Energy Efficiency Metric   The power consumption for the m-th user is composed of two components. The first is the total transmit power consumption for m-th user to the BS, which can be expressed as $\sum_{j \in \mathcal{J}} \sum_{\ell \in \mathcal{P}} p^\ell x_m^{j\ell}$. Noting that in practice, this consumption is limited by a threshold, $P_m^{\max}$, so that it should satisfy

$$\sum_{j \in \mathcal{J}} \sum_{\ell \in \mathcal{P}} p^\ell x_m^{j\ell} \leq P_m^{\max}, \quad \forall m \in \mathcal{M}. \tag{3.3}$$

The second is circuitry power consumption for the m-th user to the BS, $P_{m,\mathrm{C}}$. As a result, the overall power consumption at the m-th user

$$P_{m,\mathrm{T}} = P_{m,\mathrm{C}} + \frac{1}{\rho} \sum_{j\in\mathcal{J}} \sum_{\ell\in\mathcal{P}} p^\ell x_m^{j\ell}, \quad \forall m \in \mathcal{M}, \tag{3.4}$$

where ρ is the power amplifier efficiency constant.

The EE for each user is defined as the ratio between the total data rate and the total consumed power:

$$\mathcal{E}_{m,\mathrm{EE}} = \frac{\sum_{j\in\mathcal{J}} \sum_{\ell\in\mathcal{P}} r_m^{j\ell} x_m^{j\ell}}{P_{m,\mathrm{T}}} \text{ bits/Joule}, \quad \forall m \in \mathcal{M}. \tag{3.5}$$

Fairness Metric   For measuring the EE fairness between the users with the battery operated devices, we adopt a (non-numeric) qualitative metric, max-min fairness index. When the network reaches max-min EE fairness, any user's EE cannot be increased without decreasing another user's EE.

3.2.2   Problem Formulation for Fairness-Oriented Energy Efficiency Optimization

We consider RB and power allocations jointly to obtain a max-min energy-efficient resource allocation strategy in the uplink OFDMA networks. In particular, this strategy aims to maximize the minimum EE among all users while guaranteeing the total transmit power per user is not exceeded. We formulate the max-min optimization problem as follows:

$$\max_{\{x_m^{j\ell}\}} \quad \min_{m\in\mathcal{M}} \quad \mathcal{E}_{m,\mathrm{EE}} = \frac{\sum_{j\in\mathcal{J}} \sum_{\ell\in\mathcal{P}} r_m^{j\ell} x_m^{j\ell}}{\frac{1}{\rho} \sum_{j\in\mathcal{J}} \sum_{\ell\in\mathcal{P}} p^\ell x_m^{j\ell} + P_{m,\mathrm{C}}}, \tag{3.6a}$$

$$\text{subject to} \quad \sum_{m\in\mathcal{M}} \sum_{\ell\in\mathcal{P}} x_m^{j\ell} \leq 1, \quad \forall j \in \mathcal{J}, \tag{3.6b}$$

$$\sum_{j\in\mathcal{J}} \sum_{\ell\in\mathcal{P}} p^\ell x_m^{j\ell} \leq P_m^{\max}, \quad \forall m \in \mathcal{M}, \tag{3.6c}$$

$$x_m^{j\ell} \in \{0,1\}, \quad \forall m \in \mathcal{M}, \forall j \in \mathcal{J}, \text{and } \forall \ell \in \mathcal{P}. \tag{3.6d}$$

Noting that the objective is fractional, and the constraint in (3.6d) is non-convex. Hence, the optimization problem in (3.6) is a fractional integer non-linear programming problem. Finding out the optimal solution for such a non-convex problem requires exponential complexity and is computationally prohibitive. Hence, we propose a two-steps polynomial-time algorithm that based on the SDR technique with Gaussian randomization [8] to provide an approximate solution with high accuracy. In the next section, we will elaborate on the proposed algorithm.

3.3 Proposed Solution

The idea behind the SDR technique[1] is that the original non-convex problem is first reformulated in a higher dimensional space, and next, the non-convex constraints are relaxed. The resulting convex problem is lastly used to approximate the original one. To facilitate formulating the optimization problem in a higher dimension, we define the following five 3-dimensional tensors: $\boldsymbol{A}_j$, $\boldsymbol{B}_m$, $\boldsymbol{G}_m$, and $\boldsymbol{X}$. These tensors are in the form of a $J \times ML$ block-partitioned matrix with $1 \times M$ blocks, each with $J \times L$ entries. The tensor $\boldsymbol{X}$ can be expressed as $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \ldots & \boldsymbol{X}_M \end{bmatrix}$, where, for

all $m = 1, \ldots, M$, $\boldsymbol{X}_m$ is given as $\boldsymbol{X}_m = \begin{bmatrix} x_m^{11} & \ldots & x_m^{1\ell} \\ \vdots & \ddots & \vdots \\ x_m^{J1} & \ldots & x_m^{JL} \end{bmatrix}$. The tensor $\boldsymbol{A}_j$ can be

given as $\boldsymbol{A}_j = \begin{bmatrix} \boldsymbol{A}_{j_1} & \ldots & \boldsymbol{A}_{j_M} \end{bmatrix}$, $j = 1, \ldots, J$, where for all $m = 1, \ldots, M$, $\boldsymbol{A}_{j_m} = \boldsymbol{e}_j \boldsymbol{1}_L^T$, where $\boldsymbol{e}_j$ is the j-th column of the $J \times J$ identity matrix $\boldsymbol{I}_J$. The tensor $\boldsymbol{B}_m$ can be written as $\boldsymbol{B}_m = \begin{bmatrix} \boldsymbol{B}_{m_1} & \ldots & \boldsymbol{B}_{m_{\hat{m}}} & \ldots & \boldsymbol{B}_{m_M} \end{bmatrix}$, $m = 1, \ldots, M$, where for

$i = 1, \ldots, M$, $\boldsymbol{B}_{m_i} = \boldsymbol{0}_{J \times L}$ when $i \neq \hat{m}$, and, when $i = \hat{m}$, $\boldsymbol{B}_{m_{\hat{m}}} = \begin{bmatrix} r_{\hat{m}}^{11} & \ldots & r_{\hat{m}}^{1\ell} \\ \vdots & \ddots & \vdots \\ r_{\hat{m}}^{J1} & \ldots & r_{\hat{m}}^{JL} \end{bmatrix}$.

The tensor $\boldsymbol{G}_m$ can be expressed as $\boldsymbol{G}_m = \begin{bmatrix} \boldsymbol{G}_{m_1} & \ldots & \boldsymbol{G}_{m_{\hat{m}}} & \ldots & \boldsymbol{G}_{m_M} \end{bmatrix}$, $m =$

---

[1]A somewhat similar technique is used in [36] to address a different optimization problem in a different context. More specifically, in [36], maximizing the network-side overall EE is discussed in the downlink of OFDMA-based systems, whereas herein the design objective is to improve the minimum individual EE as much as possible in the uplink of OFDMA-based systems.

$1, \ldots, M$, where for $i = 1, \ldots, M$, $\boldsymbol{G}_{m_i} = \boldsymbol{0}_{J \times L}$ when $i \neq \hat{m}$, and, when $i = \hat{m}$,

$$\boldsymbol{G}_{m_{\hat{m}}} = \begin{bmatrix} p^1 & \cdots & p^L \\ \vdots & \ddots & \vdots \\ p^1 & \cdots & p^L \end{bmatrix}. \text{ Subsequently, we obtain the following vectors: } \boldsymbol{x} \triangleq \text{vec}(\boldsymbol{X}^T),$$

$\boldsymbol{a}_j \triangleq \text{vec}(\boldsymbol{A}_j^T)$, $\boldsymbol{b}_m \triangleq \text{vec}(\boldsymbol{B}_m^T)$, and $\boldsymbol{g}_m \triangleq \text{vec}(\boldsymbol{G}_m^T)$. Additionally, we introduce the vector $\boldsymbol{x} \in \{0, 1\}^{MJL}$. Using the defined vectors, the optimization problem can be cast as

$$\max_{\boldsymbol{x}} \quad \min_{m \in \mathcal{M}} \quad \frac{\boldsymbol{b}_m^T \boldsymbol{x}}{\frac{1}{\rho} \boldsymbol{g}_m^T \boldsymbol{x} + P_{m,\text{C}}}, \tag{3.7a}$$

$$\text{subject to} \quad \boldsymbol{a}_j^T \boldsymbol{x} \leq 1, \quad \forall j \in \mathcal{J}, \tag{3.7b}$$

$$\boldsymbol{g}_m^T \boldsymbol{x} \leq P_m^{\max}, \quad \forall m \in \mathcal{M}, \tag{3.7c}$$

$$\boldsymbol{x} \in \{0, 1\}^{MJL}. \tag{3.7d}$$

Using the vector $\boldsymbol{y} \in \{-1, 1\}^{MJL}$, where $\boldsymbol{y} = 2\boldsymbol{x} - \boldsymbol{1}$, we reformulate the problem in (3.6) as

$$\max_{\boldsymbol{y}} \quad \min_{m \in \mathcal{M}} \quad \frac{\boldsymbol{b}_m^T (\boldsymbol{y} + \boldsymbol{1})}{\frac{1}{\rho} \boldsymbol{g}_m^T (\boldsymbol{y} + \boldsymbol{1}) + 2P_{m,\text{C}}}, \tag{3.8a}$$

$$\text{subject to} \quad \frac{1}{2} \boldsymbol{a}_j^T (\boldsymbol{y} + \boldsymbol{1}) \leq 1, \quad \forall j \in \mathcal{J}, \tag{3.8b}$$

$$\frac{1}{2} \boldsymbol{g}_m^T (\boldsymbol{y} + \boldsymbol{1}) \leq P_m^{\max}, \quad \forall m \in \mathcal{M}, \tag{3.8c}$$

$$\boldsymbol{y} \in \{-1, 1\}^{MJL}. \tag{3.8d}$$

It is worth to mention that the constraints in (3.8b) and (3.8c) are linear. However, since the objective (3.8a) and the constraint (3.8d) are non-convex, the problem in (3.8) is still non-convex.

The next step for applying the SDR technique is to formulate the optimization problem of (3.8) in a higher dimension, i.e., the optimization variables are replaced with symmetric positive semidefinite (PSD) matrices [8]. Before proceeding with the next step, we also define the following vectors in $\mathbb{R}^{MJL+1}$, $\hat{\boldsymbol{g}}_m \triangleq [\frac{1}{\rho} \boldsymbol{g}_m^T \ \frac{1}{\rho} \boldsymbol{g}_m^T \boldsymbol{1} + 2P_{m,\text{C}}]^T$, $\hat{\boldsymbol{a}}_j \triangleq [\boldsymbol{a}_j^T \ \boldsymbol{a}_j^T \boldsymbol{1}]^T$, $j = 1, \ldots, J$, $\hat{\boldsymbol{b}}_m \triangleq [\boldsymbol{b}_m^T \ \boldsymbol{b}_m^T \boldsymbol{1}]^T$, $m = 1, \ldots, M$, $\hat{\boldsymbol{1}} \triangleq [\boldsymbol{1}^T \ \boldsymbol{1}^T \boldsymbol{1}]^T$, $\hat{\boldsymbol{y}} \triangleq [\boldsymbol{y}^T \ 1]^T$ and $\hat{\boldsymbol{f}} \triangleq [\boldsymbol{0}^T \ 1]^T$. Subsequently, we define the following $(MJL+1) \times (MJL+1)$ matrices $\boldsymbol{K}_{\hat{\boldsymbol{a}}_j} \triangleq \hat{\boldsymbol{f}} \hat{\boldsymbol{a}}_j^T$, $\boldsymbol{K}_{\hat{\boldsymbol{g}}_m} \triangleq \hat{\boldsymbol{f}} \hat{\boldsymbol{g}}_m^T$, and $\boldsymbol{K}_{\hat{\boldsymbol{b}}_m} \triangleq \hat{\boldsymbol{f}} \hat{\boldsymbol{b}}_m^T$, In addition, we define the

symmetric matrices $\boldsymbol{Y} \in \mathbb{R}^{MJL \times MJL}$ and $\boldsymbol{\Omega} \in \mathbb{R}^{(MJL+1) \times (MJL+1)}$ to be $\boldsymbol{Y} \triangleq \boldsymbol{y}\boldsymbol{y}^T$ and $\boldsymbol{\Omega} = \hat{\boldsymbol{y}}\hat{\boldsymbol{y}}^T$, in particular, $\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{y} \\ \boldsymbol{y}^T & 1 \end{bmatrix}$. Using those matrices, the optimization problem in (3.8) can be cast as

$$\max_{\boldsymbol{\Omega}} \quad \min_{m \in \mathcal{M}} \quad \frac{\mathrm{Tr}(\boldsymbol{K}_{\hat{\boldsymbol{b}}_m}\boldsymbol{\Omega})}{\mathrm{Tr}(\boldsymbol{K}_{\hat{\boldsymbol{g}}_m}\boldsymbol{\Omega})}, \tag{3.9a}$$

$$\text{subject to} \quad \frac{1}{2}\mathrm{Tr}(\boldsymbol{K}_{\hat{\boldsymbol{a}}_j}\boldsymbol{\Omega}) \leq 1, \qquad \forall j \in \mathcal{J}, \tag{3.9b}$$

$$\frac{1}{2}\mathrm{Tr}(\boldsymbol{K}_{\hat{\boldsymbol{g}}_m}\boldsymbol{\Omega}) \leq \frac{1}{\rho}P_m^{\max} + P_{m,\mathrm{C}}, \ \forall m \in \mathcal{M}, \tag{3.9c}$$

$$\boldsymbol{\Omega} \succeq \boldsymbol{0}, \tag{3.9d}$$

$$\mathrm{diag}(\boldsymbol{\Omega}) = 1, \tag{3.9e}$$

$$\mathrm{rank}(\boldsymbol{\Omega}) = 1. \tag{3.9f}$$

This formulation is non-convex due to both the objective function (3.9a) and the rank-1 constraint (3.9f). However, noting that after dropping rank-1 constraint, the resulting problem is quasi-convex [25], which can be efficiently tackled using the bisection method. Using a new variable, $\mathcal{E}_0$, we reformulate the relaxed problem as

$$\max_{\mathcal{E}_0, \, \boldsymbol{\Omega}} \quad \mathcal{E}_0, \tag{3.10a}$$

$$\text{subject to} \quad \mathrm{Tr}(\mathcal{E}_0(\boldsymbol{K}_{\hat{\boldsymbol{g}}_m} - \boldsymbol{K}_{\hat{\boldsymbol{b}}_m})\boldsymbol{\Omega}) \leq 0, \ \forall m \in \mathcal{M}, \tag{3.10b}$$

$$(3.9b) - (3.9e). \tag{3.10c}$$

To obtain the maximum value of $\mathcal{E}_0$, we solve a series of convex feasibility problems, each of which is of the form

$$\text{find} \quad \boldsymbol{\Omega}, \tag{3.11a}$$

$$\text{subject to} \quad \mathrm{Tr}(\mathcal{E}_0(\boldsymbol{K}_{\hat{\boldsymbol{g}}_m} - \boldsymbol{K}_{\hat{\boldsymbol{b}}_m})\boldsymbol{\Omega}) \leq 0, \ \forall m \in \mathcal{M}, \tag{3.11b}$$

$$(3.9b) - (3.9e), \tag{3.11c}$$

where $\mathcal{E}_0$ is fixed at each iteration of the bisection method and it lies in $[0, \ \mathcal{E}_{\max}]$, where $\mathcal{E}_{\max} = \min_{m \in \mathcal{M}} \left(\frac{\boldsymbol{b}_m^T \mathbf{1}}{P_{m,\mathrm{C}}} \right)$. Let $\boldsymbol{\Omega}^*$ represent the optimal solution of (3.10) corresponding to the optimal value of $\mathcal{E}_0$. If $\boldsymbol{\Omega}^*$ has a rank one, the relaxation is tight. Otherwise,

we use Gaussian randomization [8] in order to obtain an approximate solution. The idea behind of the Gaussian randomization technique is to tackle a stochastic version of the deterministic problem in (3.10). Particularly, in this technique, several random samples, T, are first generated from a multivariate Gaussian distribution with the co-variance matrix of $\boldsymbol{\Omega}^*$. Then, the random vector that maximizes the objective in (3.6) is selected as the sub-optimal solution.

3.3.1   Complexity Discussion

The proposed algorithm uses the bisection method to solve the problem in (3.10). This method requires $\log_2(\mathcal{E}_{\max}/\kappa)$ number of iterations for the convergence, where $\kappa > 0$ is the solution accuracy. Each iteration involves solving a convex PSD-constrained problem, with the computational complexity of $\mathcal{O}\left((MJL)^{3.5}\right)$. Finally, it uses the Gaussian randomization technique. At this step, the complexity of generating T random samples for which the objective is evaluated is $\mathcal{O}\left((MJL)^2 T\right)$. Therefore, the total computational complexity can be found as $\mathcal{O}\left((MJL)^{3.5}\log_2(\mathcal{E}_{\max}/\kappa) + (MJL)^2 T\right)$. Even though the proposed centralized algorithm has polynomial-time complexity, the order of the polynomial complexity is relatively high. Hence, it is applicable in small-to-medium-size networks. For large-size networks, decentralized algorithms are more desirable in order to ensure reduced complexity and overhead. However, we have the following remarks:

- One of the most effective means of developing optimization-based decentralized designs is to begin with a centralized design and then, using Lagrangian duality, the formulation is decoupled to facilitate the task of each node in the network to optimize its parameters locally. In other words, centralized formulations are usually useful in developing decentralized ones.

- A decentralized formulation can be regarded as a centralized one, but with decoupling constraints. This implies that the solutions generated through the centralized formulation constitute an upper bound on the solutions generated through their decentralized counterparts. In other words, the centralized formulations developed in our work will not only serve as starting points for developing effective decentralized algorithms, but also to serve a benchmark for validating and examining their efficacy.

3.4 Simulation Results

In this section, the EE performance of the proposed resource allocation scheme in an OFDMA network is evaluated with simulations. We assume that the path loss model is expressed as $128.1 + 37.6\log_{10}(d)$ for the user-to-BS links [23], where d denotes the distance in kilometers. The log-normal distribution with a standard deviation of 8 dB is used to model the shadowing factor. The power-level set consists of L equally located points in $[0.05P_m^{\max}, \ 0.5P_m^{\max}]$, and $P_m^{\max} = P^{\max}, \forall m \in \mathcal{M}$, unless otherwise stated. Furthermore, we assume that noise power spectral density is -174 dBm/Hz, the bandwidth of each RB is 180 kHz, and the power amplifier efficiency is 38%. We use 100 Monte Carlo simulations to obtain the average performance. The optimization problem is solved using the SDPT3 package [30] and the randomization technique is implemented using 10^3 independent randomization samples.

Example 6. (*Performance Comparison With Optimal Solution*) Fig. 3.1 considers a network with 4 RBs, 3 users, 2 power levels. In this network, we assume that the static circuitry power consumption is 25 dBm.

In Fig. 3.1, the performance of proposed scheme is compared against the optimal solution derived through exhaustive search. From this figure, it can be readily observed that the performance of the proposed technique is close to the optimal one, especially at high values of the transmit power. For instance, at the transmit power of 31 dBm, the gap between the minimum EE and the one generated by the SDR-technique is 2.97×10^4 bits/Joule, which is about 1% of the objective value obtained by exhaustive search. Moreover, for completeness of presentation, a comparison of these two algorithms in terms of EE fairness is given in Fig. 3.2. For quantifying the EE fairness among multiple users, we adopt Jain's fairness index [37][2]. Since the adopted EE optimization metric, viz., max-min fair EE, improves EE fairness between different users as much as possible, as expected, the achieved EE Jain's fairness index is high, specifically, higher than 0.9 for all values of the transmit power. □

Example 7. (*Effect of Number of Power Levels*) Fig. 3.3 considers a network with 10 RBs, 5 users, and the different number of power levels, i.e., 2, 4, and 8. In this

---

[2]Jain's index is widely used as a (numeric) quantitative fairness metric; a higher value of Jain's index corresponds to a fairer rate allocation. Particularly, it is defined as $F(\mathbf{x}) = \frac{\left(\sum_{m\in\mathcal{M}}\sum_{j\in\mathcal{J}}\sum_{\ell\in\mathcal{P}} r_m^{j\ell}x_m^{j\ell}\right)^2}{M\sum_{m\in\mathcal{M}}(\sum_{j\in\mathcal{J}}\sum_{\ell\in\mathcal{P}} r_m^{j\ell}x_m^{j\ell})^2}$. It is a continuous non-convex function, $F(\mathbf{x}) : \mathbb{R}_m^+ \to \mathbb{R}^+$, with values in the interval $[\frac{1}{M}, \ 1]$.

network, we assume that the static circuitry power consumption is 25 dBm.

Fig. 3.3 demonstrates the impact of the number of power levels on the average network performance. One can see that for a specific transmit power, the minimum EE enhances as the number of power level raises. This is due to the fact that the increase of the number of power level provides extra degrees of freedom in the energy-efficient resource allocation optimization. Another important observation is that the impact of the number of power levels on max-min EE optimization becomes less significant as the transmit power increases. □

Example 8. (*Computational Time*) Fig. 3.4 considers a network with 8 RBs, 4 users. In this network, we assume that the static circuitry power consumption is 25 dBm, and the number of power levels varies from 1 to 11.

Fig. 3.4 reveals the average computational time required by the proposed algorithm. From this figure, it can be observed that the proposed algorithm requires a polynomial amount of computational time. □

3.5   Conclusion

In this work, we addressed the problem of optimizing resource allocation in uplink OFDMA networks for providing the EE fairness among the users while considering discrete power levels. Particularly, we focused on how to maximize the minimum user EE in the network by jointly optimizing RB and discrete power allocation. We first formulated the optimization problem, but unfortunately, the resulting formulation was not convex. In order to tackle this difficulty, we proposed an SDR-based technique with polynomial time complexity that approximates the solution for the max-min EE problem using a convex program. The numerical results indicated that the proposed technique does not only achieve the desirable improvement on the users' EE, but also ensures EE fairness among the users.
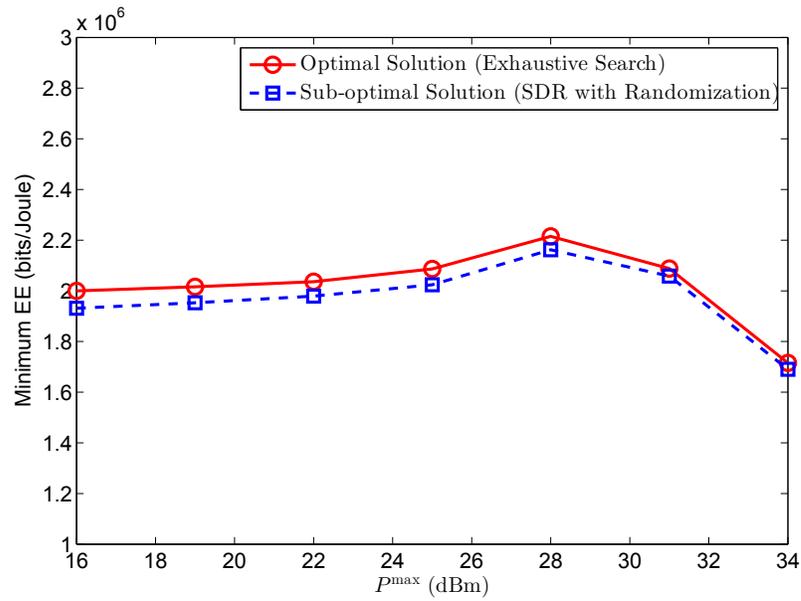
Figure 3.1: Performance of the proposed sub-optimal technique against the optimal one.
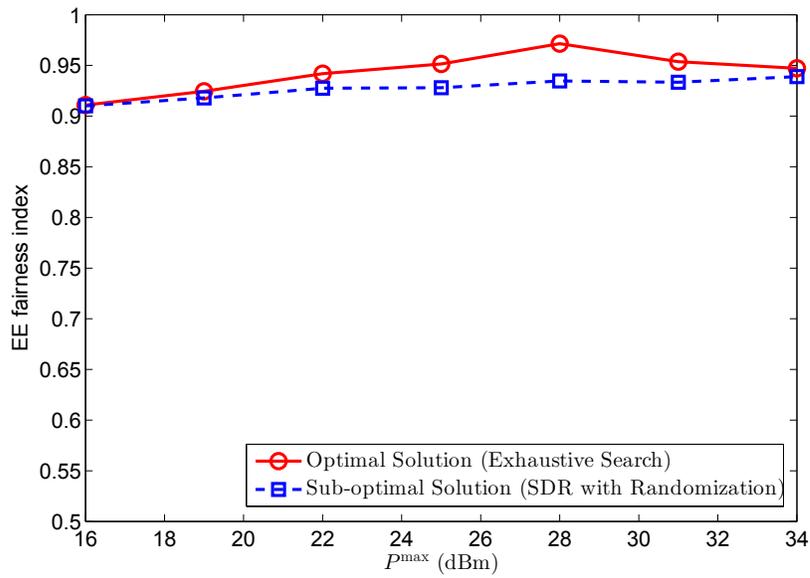

Figure 3.2: EE fairness among the users.
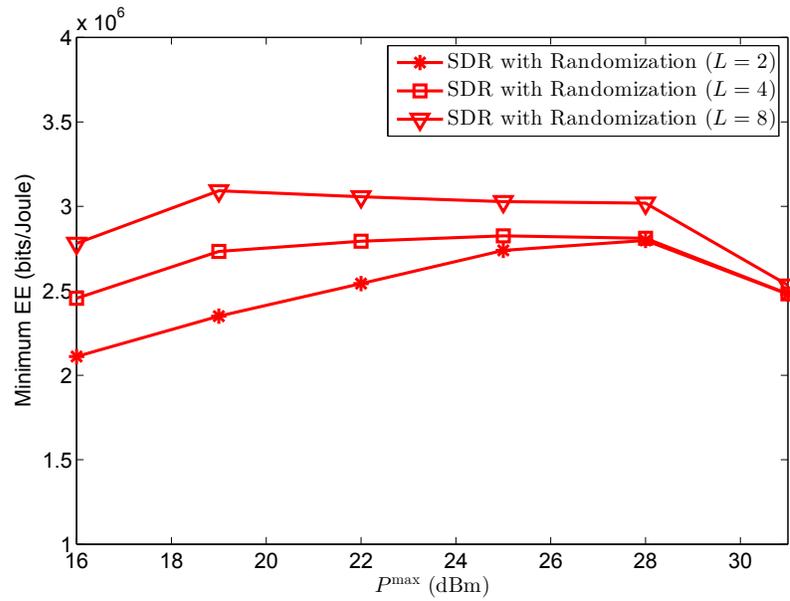
Figure 3.3: Performance of the proposed technique with the different number of power levels.
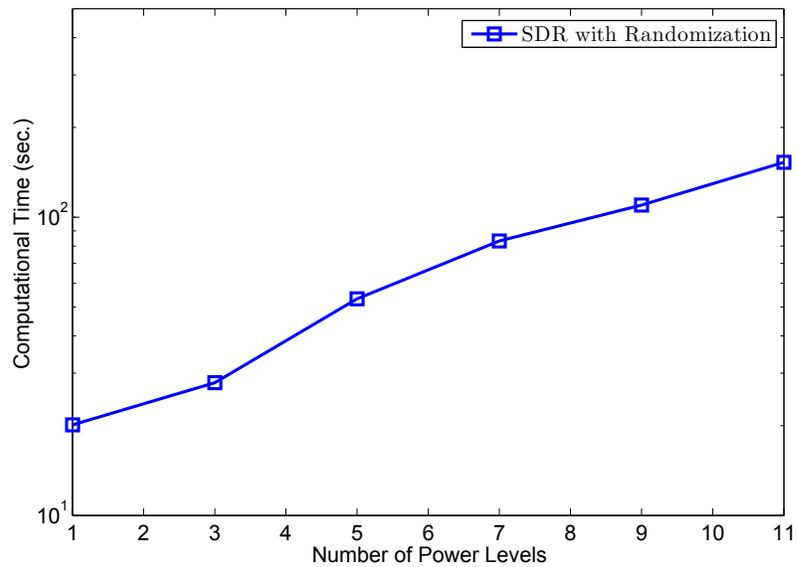
Figure 3.4: Computational time versus network size.

Chapter 4

QoS-Guaranteed User Association in Downlink Heterogeneous Cellular Networks

4.1 Introduction

Despite the ostensible ubiquity of conventional homogeneous cellular networks, their inherent limitations render them incapable of supporting the users' increasing demand for reliable communication of higher data rates. Such limitations include inefficient utilization of spectral resources, poor coverage and excessive power consumption [3]. To mitigate these limitations, cellular networks have evolved to include low-power pico and femto base stations (BSs). Incorporating such BSs to assist homogeneous conventional macro ones gives rise to the so-called heterogeneous networks (HetNets). These networks offer significant advantages over their conventional counterparts. For instance, HetNets are capable of: making efficient utilization of the spectrum available for communication [3]; reducing the overall network power consumption [38]; eliminating coverage holes; and, with proper user-to-BS association, accommodating more users in the system as will be shown herein. Despite the advantages of Het-Nets, the disparate transmit powers and computational capabilities characteristic of the underlying BSs render fair and resource-efficient user-to-BS association a serious challenge [39].

To overcome user-to-BS association challenges in HetNets, several considerations must be taken into account: 1) the channel conditions between the users and the BSs; 2) the load condition of each BS; and 3) the quality-of-service (QoS) demanded by

each user. A conventional approach to associating users to BSs is the one in which each user is associated to the BS with the maximum signal-to-interference-plus-noise ratio (max-SINR). This approach is relatively simple, but can lead to highly imbalanced loading of the BSs and does not guarantee users' QoS. Furthermore, the max-SINR approach does not account for disparities in the transmit powers and BS computational capabilities. Another approach for user-to-BS association is the one based on range expansion (RE) [40]. In that approach, the traffic from macro BSs is offloaded to low-power BSs by adding a positive bias to the measured SINRs of low-power BSs. An advantage of this approach is that it does not depend on the relative location of users and BSs. However, a major weakness of it is the lack of theoretical guidance for choosing the biasing factor. Algorithms for determining user-to-BS associations by finding approximate solutions to relaxed network utility maximization problems using dual pricing methods were proposed in [41] and [42] for the case when the number of users in the system is fixed. Those methods do not account for the QoS requirements and the associations that they yield are not necessarily binary. The latter feature implies that the number of BSs serving a particular user can be practically infeasible.

In this work, we consider user-to-BS association in the downlink of a generic HetNet. Such an association must not only ensure that the number of accommodated users in the network is maximized but also that the network resources are efficiently utilized and the users' QoS demands are met. In its simplest form, this problem is combinatorial NP-hard and the difficulty of solving it in HetNets is aggravated by the disparity of transmit powers and BS computational capabilities; finding the optimal solution even for relatively small homogeneous systems is computationally prohibitive. To find close-to-optimal user-to-BS associations, we propose a method based on SDR with randomization, which is a powerful method for solving a class of combinatorial problems [8]. In the first phase of this method, a convex formulation is used to generate a covariance matrix and an upper bound on the maximum number of users that can be accommodated in the network while minimizing the total network resource consumption and meeting the users' QoS demands. In the second phase, the covariance matrix obtained in the first phase is used to generate a population of random candidates of the solution. The candidate that yields the highest objective is used for associating users to BSs. Unlike other techniques with polynomial complexity, SDR with randomization has a provable approximation accuracy. We show that, in

comparison with other user-to-BS association approaches, SDR with randomization enables the system to accommodate a significantly higher number of satisfied users.

4.2 System Model

We consider the downlink of a static two-tier HetNet shown in Fig. 4.1 [41]. In such a HetNet, tier-1 consists of one or more macro BSs and tier-2 consists of one or more pico BSs. The set of BSs is denoted by $\mathcal{B} = \{1, \cdots, N_B\}$ and the set of users is denoted by $\mathcal{U} = \{1, \cdots, N_U\}$. All BSs are assumed to have full buffers and to
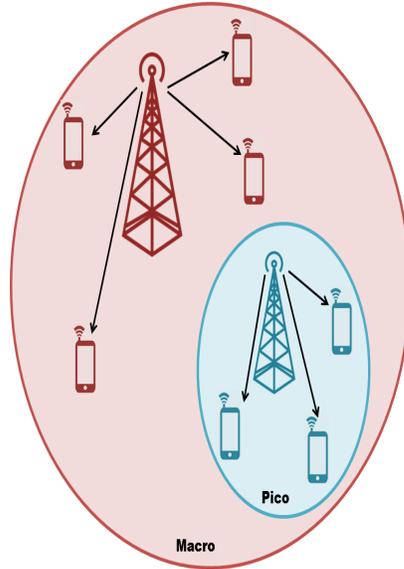
Figure 4.1: A two-tier HetNet; macro BS transmits higher power than pico BS.

have fixed transmit powers throughout the signalling interval. Using g_{ij} to denote the channel gain between the i-th BS and the j-th user, P_i to denote the transmit power of the i-th BS and σ_N to denote the variance of the additive Gaussian noise, the received SINR observed by the j-th user from the i-th BS can be expressed as

$$\gamma_{ij} = \frac{P_i g_{ij}}{\sum\limits_{k \in \mathcal{B}, \, k \neq i} P_k g_{kj} + \sigma_N}, \qquad i \in \mathcal{B}, \, j \in \mathcal{U}. \tag{4.1}$$

The corresponding rate at which the i-th BS can reliably communicate with the j-th user per unit bandwidth is

$$\eta_{ij} = \log_2\left(1 + \gamma_{ij}\right) \quad [\text{bits/s/Hz}]. \tag{4.2}$$

For practical considerations, many communication standards, including the current Long Term Evolution (LTE) one, impose a restriction on the smallest number of time-frequency resource blocks (RBs) that can be assigned to any user [23], and the number of RBs that must be assigned to a user depends on its QoS demand; higher QoS requirement implies a higher number of allocated RBs. To see that, let Q_j be the data rate in bits/s required by the j-th user, $j = 1, \ldots, N_U$, and let W_0 be the bandwidth of each RB. Using this notation, the minimum number of RBs that must be allocated by the i-th BS to meet the demand of the j-th user is given by

$$b_{ij} = \left\lceil \frac{Q_j}{W_0 \eta_{ij}} \right\rceil, \tag{4.3}$$

From this expression, it can be seen that b_{ij} is directly proportional to the QoS required by the j-th user and inversely proportional to the maximum per unit bandwidth rate that can be achieved on the ij-th link.

4.3 Optimization Problem Formulation

Using the characterization in Section 4.2, we now develop an optimization framework for maximizing the number of users accommodated by the network while ensuring efficient utilization of the available resources. Towards that end, we begin by defining the binary association variables $x_{ij} \in \{0, 1\}$ such that $x_{ij} = 1$ if the j-th user is served by the i-th BS and $x_{ij} = 0$ otherwise, $i = 1, \ldots, N_B$, $j = 1, \ldots, N_U$.

Using this notation, the total number of resources used by the i-th BS can be expressed as $\sum_{j \in \mathcal{U}} b_{ij} x_{ij}$ and the total number of resources used in the network can be expressed as $\sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{U}} b_{ij} x_{ij}$. To meet the constraints of the macro and pico BSs, a practically feasible design must ensure that the total number of resources used by the i-th BS does not exceed a given threshold t_i, i.e., $\sum_{j \in \mathcal{U}} b_{ij} x_{ij} \leq t_i, i = 1, \ldots, N_B$. From a practical perspective, it is also desirable for each user to be served by at most one BS. This is because, otherwise, coordination among multiple BSs is necessary, which is usually cumbersome in practice. The requirement that each user is served by

at most one BS can be expressed as the constraint $\sum_{i \in \mathcal{B}} x_{ij} \leq 1$, $j = 1, \ldots, N_U$. When this constraint is enforced, the total number of users in the system can be expressed as $\sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{U}} x_{ij}$. Note that a user $\hat{j} \in \mathcal{U}$ that is not served by any BS will result in $x_{i\hat{j}} = 0$ for every $i \in \mathcal{B}$.

To maximize the number of users while ensuring efficient utilization of resources we consider a linear combination of these two quantities. In particular, let $\rho \in [0, 1]$ be given, and let the objective be to maximize $\rho \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{U}} x_{ij} - (1 - \rho) \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{U}} b_{ij} x_{ij}$. Increasing ρ from 0 to 1 shifts the emphasis of the optimization from minimizing the number of resources used in the network to maximizing the number of users accommodated therein. As such, ρ parametrizes a family of objectives, each with a corresponding number of users and a number of expended resources. To determine the optimal value of ρ, we follow a reasoning similar to the one in [43]. Doing so, it can be shown that any value of $\rho \in \left(\frac{\sum_{i \in \mathcal{B}} t_i}{1 + \sum_{i \in \mathcal{B}} t_i}, 1 \right)$ ensures maximizing the number of served users and simultaneously minimizing the number of expended resources.

Assuming that ρ lies within the aforementioned interval, the downlink user-to-BS association can be formulated as:

$$\max_{x_{ij}} \quad \rho \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{U}} x_{ij} - (1 - \rho) \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{U}} b_{ij} x_{ij} \tag{4.4a}$$

$$\sum_{j \in \mathcal{U}} b_{ij} x_{ij} \leq t_i, \quad i \in \mathcal{B}, \tag{4.4b}$$

$$\sum_{i \in \mathcal{B}} x_{ij} \leq 1, \quad j \in \mathcal{U}, \tag{4.4c}$$

$$x_{ij} \in \{0, 1\}, \quad i \in \mathcal{B}, \ j \in \mathcal{U}. \tag{4.4d}$$

From (4.4), it can be readily verified that apart from the binary constraint in (4.4d), this formulation constitutes a convex linear optimization problem. However, the binary constraint makes this problem an instance of a non-convex integer linear program. In fact, this problem is known as generalized assignment problem [44], which can be solved using various techniques, including those that use the branch-and-bound approach and variations thereof. Unfortunately, the complexity of these algorithms grows exponentially with the number of variables and as such their applicability in the current scenario is questionable. To alleviate this difficulty, herein we consider utilizing an alternate technique for dealing with this class of problems, viz., the one based on semidefinite relaxation with randomization [8].

4.4 User Association via Semidefinite Relaxation

Being an instance of the generalized assignment problem, the optimization of user-to-BS associations can be readily seen to be NP-hard. Hence, solving this problem optimally is computationally infeasible for all, but a relatively small class of Het-Nets with a limited number of users and BSs. To obtain resource-efficient user-to-BS associations, we propose to use the SDR technique with randomization. Unlike the techniques based on the branch-and-bound approach, SDR with randomization has polynomial complexity, which renders it suitable for solving large user-to-BS association problems that may arise in practice. In comparison with other polynomial complexity algorithms, SDR with randomization has provable approximation accuracy [7]. A somewhat similar technique for generating user-to-BS associations was developed in [45] in a different context. In particular, herein we consider maximizing the number of users served by a generic HetNet with an arbitrary number of macro and pico BSs per cell under QoS constraints, whereas in [45] the HetNet is restricted to have one macro and one pico BS per cell and the design objective is to maximize the total sum rate irrespective of the QoSs demanded by the users.

To cast (4.4) in a form that is more amenable to the optimization technique employed hereinafter, we will express the objective and the constraints using vector notation. Towards that end, let $\boldsymbol{X} \in \mathbb{R}^{N_B \times N_U}$ and $\boldsymbol{B} \in \mathbb{R}^{N_B \times N_U}$ be the matrices whose ij-th entries are x_{ij} and b_{ij}, respectively. Let $\boldsymbol{D}_i \in \mathbb{R}^{N_B \times N_U}$ be the all-zero matrix except its i-th column is equal to the i-th column of $\boldsymbol{B}$, $i = 1, \ldots, N_B$, and let $\boldsymbol{E}_j \in \mathbb{R}^{N_B \times N_U}$ be the all-zero matrix except its j-th row is all ones, $j = 1, \ldots, N_u$. Then, we make the following definitions: $\mathbf{x} \triangleq \mathrm{vec}(\boldsymbol{X})$, $\mathbf{b} \triangleq \mathrm{vec}(\boldsymbol{B})$, $\mathbf{d}_i \triangleq \mathrm{vec}(\boldsymbol{D}_i)$ and $\mathbf{e}_j \triangleq \mathrm{vec}(\boldsymbol{E}_j)$. Using this notation the problem in (4.4) can be expressed in the following form:

$$\max_{\mathbf{x}} \quad \rho \mathbf{1}^T \mathbf{x} - (1 - \rho)\mathbf{b}^T \mathbf{x}, \tag{4.5a}$$

$$\text{subject to} \quad \mathbf{d}_i^T \mathbf{x} \le t_i, \quad i \in \mathcal{B}, \tag{4.5b}$$

$$\mathbf{e}_j^T \mathbf{x} \le 1, \quad j \in \mathcal{U}, \tag{4.5c}$$

$$\mathbf{x} \in \{0, 1\}^{N_B N_U}. \tag{4.5d}$$

To cast this problem in the standard form for SDR, the optimization variables must be antipodal rather than binary. Hence, we introduce the vector $\boldsymbol{\beta} = 2\mathbf{x} - \mathbf{1}$, which

48

implies that $\boldsymbol{\beta} \in \{-1, 1\}^{N_B N_U}$, and

$$\mathbf{x} = \frac{1}{2}(\boldsymbol{\beta} + \mathbf{1}). \tag{4.6}$$

Using (4.6), the formulation in (4.5) can be cast as:

$$\max_{\mathbf{x}} \quad \frac{\rho}{2}\mathbf{1}^T(\boldsymbol{\beta} + \mathbf{1}) - \frac{(1 - \rho)}{2}\mathbf{b}^T(\boldsymbol{\beta} + \mathbf{1}), \tag{4.7a}$$

$$\text{subject to} \quad \frac{1}{2}\mathbf{d}_i^T(\boldsymbol{\beta} + \mathbf{1}) \leq t_i, \qquad i \in \mathcal{B}, \tag{4.7b}$$

$$\frac{1}{2}\mathbf{e}_j^T(\boldsymbol{\beta} + \mathbf{1}) \leq 1, \qquad j \in \mathcal{U}, \tag{4.7c}$$

$$\boldsymbol{\beta} \in \{-1, 1\}^{N_B N_U}. \tag{4.7d}$$

To use the SDR technique, the problem in (4.7) is expressed in a form in which the optimization variable is constrained to be in the cone of symmetric positive semidefinite matrices [8]. To do so, we define the following vectors in $\mathbb{R}^{N_B N_U + 1}$, $\hat{\mathbf{b}} \triangleq [\mathbf{b}^T \ \mathbf{b}^T\mathbf{1}]^T$, $\hat{\mathbf{d}}_i \triangleq [\mathbf{d}_i^T \ \mathbf{d}_i^T\mathbf{1}]^T$, $i = 1, \ldots, N_B$, $\hat{\mathbf{e}}_j \triangleq [\mathbf{e}_j^T \ \mathbf{e}_j^T\mathbf{1}]^T$, $j = 1, \ldots, N_U$, $\hat{\mathbf{1}} \triangleq [\mathbf{1}^T \ \mathbf{1}^T\mathbf{1}]^T$, $\hat{\boldsymbol{\beta}} \triangleq [\boldsymbol{\beta} \ 1]^T$ and $\hat{\mathbf{c}} \triangleq [\mathbf{0}_{N_B N_U}^T \ 1]^T$, where $\mathbf{0}_{N_B N_U}$ is the all-zero $N_B N_U \times 1$ vector. We also define the symmetric matrices $\boldsymbol{\Phi} \in \mathbb{R}^{N_B N_U \times N_B N_U}$ and $\boldsymbol{\Psi} \in \mathbb{R}^{N_B N_U + 1 \times N_B N_U + 1}$ to be $\boldsymbol{\Phi} \triangleq \boldsymbol{\beta}\boldsymbol{\beta}^T$ and $\boldsymbol{\Psi} = \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T$; in particular, $\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{\beta} \\ \boldsymbol{\beta}^T & 1 \end{bmatrix}$. Finally, we define the $N_B N_U + 1 \times N_B N_U + 1$ rank-1 matrices $\mathbf{A}_1 \triangleq \hat{\mathbf{c}}\hat{\mathbf{1}}^T$, $\mathbf{A}_b \triangleq \hat{\mathbf{c}}\hat{\mathbf{b}}^T$, $\mathbf{A}_{d_i} \triangleq \hat{\mathbf{c}}\hat{\mathbf{d}}_i^T$, $i = 1, \ldots, N_B$ and $\mathbf{A}_{e_j} \triangleq \hat{\mathbf{c}}\hat{\mathbf{e}}_j^T$, $j = 1, \ldots, N_U$. Using these definitions, it can be ready verified that the problem in (4.7) is equivalent to the following optimization problem.

$$\max_{\boldsymbol{\Psi}} \quad \frac{\rho}{2}\text{Tr}(\mathbf{A}_1\boldsymbol{\Psi}) - \frac{1 - \rho}{2}\text{Tr}(\mathbf{A}_b\boldsymbol{\Psi}), \tag{4.8a}$$

$$\text{subject to} \quad \frac{1}{2}\text{Tr}(\mathbf{A}_{d_i}\boldsymbol{\Psi}) \leq t_i, \qquad i \in \mathcal{B}, \tag{4.8b}$$

$$\frac{1}{2}\text{Tr}(\mathbf{A}_{e_j}\boldsymbol{\Psi}) \leq 1, \qquad j \in \mathcal{U}, \tag{4.8c}$$

$$\boldsymbol{\Psi} \succeq \mathbf{0}, \tag{4.8d}$$

$$\text{diag}(\boldsymbol{\Psi}) = 1, \tag{4.8e}$$

$$\text{rank}(\boldsymbol{\Psi}) = 1. \tag{4.8f}$$

In this formulation, the binary constraint in (4.7d) is replaced with the equivalent constraints in (4.8e) and (4.8f). The constraint in (4.8e) is linear and therefore can be readily incorporated in the optimization framework. However, the rank-1 constraint in (4.8f) is highly non-convex and, in fact, it is this constraint that captures the NP-hardness of the original problem in (4.7). Despite being ostensibly more complex than (4.7), the formulation in (4.8) offers the advantage of being readily amenable to the SDR technique, which we describe next.

4.4.1 SDR with Gaussian Randomization

Our main contribution is to show that maximizing the number of users in a generic HetNet with a given set of available resources can be cast in the form in (4.8). This form is amenable to the standard SDR with Gaussian randomization approach [8]. For completeness, this approach will be tailored below to provide an approximate solution of (4.8).

The main difficulty of solving (4.8) follows from the non-convexity of the rank-1 constraint. By dropping this constraint, the resulting problem can be seen to be linear over the set of matrices $\mathbf{\Psi}$ that lie in the positive semidefinite cone. In other words, the relaxation induced by dropping the rank-1 constraint yields an efficiently solvable convex optimization problem in the form of a semidefinite program. This program forms the basis of the solution approach adopted herein.

Let $\mathbf{z}$, $\mathbf{Z}$ and $\mathbf{R}$ be the optimization variables of the relaxed problem corresponding to $\boldsymbol{\beta}$, $\mathbf{\Phi}$ and $\mathbf{\Psi}$ in the original problem in (4.8), respectively. The convexity of the relaxed version of (4.8) implies that its global optimal solution, $\mathbf{z}^*$, $\mathbf{Z}^*$ and $\mathbf{R}^*$, can be readily obtained. Since the difference between (4.8) and its relaxed version is the rank-1 constraint, it can be readily seen that if the rank of the matrix generated by the relaxed program is 1, i.e., $\text{rank}(\mathbf{R}^*) = 1$, the vector $\mathbf{z}^*$ yielded by solving the relaxed version of (4.8) is binary and constitutes the optimal user-to-BS association. However, the solution of the relaxed problem is in general not rank-1. In this case, the Gaussian randomization approach can be used to obtain a desirable user-to-BS association for the problem in (4.7).

In the Gaussian randomization approach, a stochastic version of the relaxation of (4.8) is conceived. In that version, the vector $\mathbf{z}^*$ generated by solving the relaxed program is considered as the mean of a multivariate Gaussian $N_B N_U$-dimensional random vector, and $\mathbf{Z}^*$ and $\mathbf{Z}^* - \mathbf{z}^* \mathbf{z}^{*T}$ are considered as the correlation and covariance

matrices of this random vector, respectively. In particular, let $\boldsymbol{\Delta}$ be a set of J random vector samples drawn from the Gaussian distribution with mean $\mathbf{z}^*$ and covariance $\mathbf{Z}^* - \mathbf{z}^*\mathbf{z}^{*T}$, i.e., $\boldsymbol{\Delta} = \{\boldsymbol{\delta}^j\}_{j=1}^{J}$, $\boldsymbol{\delta}^j \sim \mathcal{N}(\mathbf{z}^*, \mathbf{Z}^* - \mathbf{z}^*\mathbf{z}^{*T})$, $j = 1, \ldots, J$. Let $\hat{\boldsymbol{\delta}} = [\boldsymbol{\delta}^T \ 1]^T$ and $\hat{\mathbf{z}}^* = [\mathbf{z}^{*T} \ 1]^T$. Hence, the vectors in $\boldsymbol{\Delta}$ can be seen to provide an approximate solution to the following stochastic optimization problem:

$$\max_{\mathbb{E}\{\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^T\}=\mathbf{R}^*, \ \mathbb{E}\{\hat{\boldsymbol{\delta}}\}=\hat{\mathbf{z}}^*} \quad \mathbb{E}\left\{\hat{\boldsymbol{\delta}}^T\left(\frac{\rho}{2}\mathbf{A}_1 - \frac{1-\rho}{2}\mathbf{A}_b\right)\hat{\boldsymbol{\delta}}\right\}, \tag{4.9a}$$

$$\text{subject to} \quad \frac{1}{2}\mathbb{E}\{\hat{\boldsymbol{\delta}}^T\mathbf{A}_{d_i}\hat{\boldsymbol{\delta}}\} \leq t_i, \qquad i \in \mathcal{B}, \tag{4.9b}$$

$$\frac{1}{2}\mathbb{E}\{\hat{\boldsymbol{\delta}}^T\mathbf{A}_{e_j}\hat{\boldsymbol{\delta}}\} \leq 1, \qquad j \in \mathcal{U}. \tag{4.9c}$$

$$\mathbb{E}\{\hat{\boldsymbol{\delta}}_r^2\} = 1, \ r = 1, \ldots, N_U N_B. \tag{4.9d}$$

Notice that because the matrix $\mathbf{R}^* = \begin{bmatrix} \mathbf{Z}^* & \mathbf{z}^* \\ \mathbf{z}^{*T} & 1 \end{bmatrix}$ generated by the relaxed version of (4.8) is positive semidefinite, its Schur complement, $\mathbf{Z}^* - \mathbf{z}^*\mathbf{z}^{*T}$, is also positive semidefinite. In other words, $\mathbf{Z}^* - \mathbf{z}^*\mathbf{z}^{*T}$ constitutes a valid covariance matrix of the random vectors $\{\boldsymbol{\delta}^j\}$. This establishes the equivalence between (4.9) and the relaxed version of (4.8).

Now, the vectors in $\boldsymbol{\Delta}$ are used to find candidate binary solutions $\{\tilde{\boldsymbol{\beta}}^j\}_{j=1}^{J}$ for the problem in (4.7) by quantizing the entries of each realization of $\{\boldsymbol{\delta}^j\}_{j=1}^{J}$. In particular,

$$\tilde{\boldsymbol{\beta}}^j = \text{sgn}(\boldsymbol{\delta}^j), \qquad j = 1, \ldots, J, \tag{4.10}$$

Using $\tilde{\boldsymbol{\beta}}^j$ and (4.6), we obtain candidate solutions of (4.5), $\tilde{\mathbf{x}}^j$. The candidate that yields the largest objective and satisfies the constraints in (4.5) is used for associating the users to the BSs, i.e.,

$$\mathbf{x}^* = \arg\max_{\mathcal{F}_j} \tilde{\mathbf{x}}^j, \quad \mathcal{F}_j \triangleq \{\tilde{\mathbf{x}}^j : \ \tilde{\mathbf{x}}^j \text{ satisfying (4.5b), (4.5c)}\}.$$

4.5 Performance Evaluation

We evaluate the performance of the proposed user-to-BS association method through numerical comparisons with two popular user-to-BS association algorithms, viz., the

max-SINR algorithm and the RE algorithm [46]. We consider a two-tier HetNet scenario consisting of one macro and three pico BSs. The transmit powers of macro and pico BSs are 46 dBm and 35 dBm, respectively. The location of the macro BS is assumed to be fixed, whereas the locations of the pico BSs and the users are assumed to be random and uniformly distributed. For the propagation environment, shadowing is assumed to have a log-normal distribution with a standard deviation of $\sigma_s = 8$ dB and the path loss between the BSs and users is assumed to be $L(d) = 34 + 40 \log_{10}(d)$, where d is the distance between users and BSs in meters. The noise power at all receivers is assumed to be -174 dBm. For all numerical simulations, the results are averaged over 50 independent channel realizations. For practical considerations, users are allocated integer numbers of RBs; fractional RB allocation not allowed. The bandwidth of each RB is set to be 180 kHz [23]. The SDR programs are solved using the CVX package [30] with SDPT3 solver. The number of Gaussian samples used in the randomization phase is set to be $J = 100$ for each channel realization.

Example 9. (*Impact of the Number of Users Accessing the Network*) In this example we consider the scenario in which the users accessing the network, each has a QoS demand of $Q_j = 250$ kilo-bits-per-second (kbps), $j \in U$, and each BS has a total of $t_i = 25$ available RBs, for $i \in B$. The user-to-BS associations are generated using the max-SINR, the RE and the proposed SDR-based algorithms. In Figure 4.2 the number of users that could be accommodated in the network is plotted versus the total number of users who wish to use the network. For the RE algorithm, the biasing parameter for each BS was set to be either 5 dB or 10 dB [46].

From this figure, it can be seen that the proposed SDR-based algorithm outperforms both the max-SNR and the RE algorithms over the entire range of available network users. This is especially true when the system experiences high loading conditions. When the system is lightly loaded, all algorithms can accommodate a high percentage of the available users, and when the system is heavily loaded, the available RBs are exhausted and only a small percentage of users can be accommodated. For instance, when the total number of users is 100, the proposed algorithm can accommodate the demands of about 17% more users than the RE algorithms and about 34% more users than the max-SINR algorithm. □

Example 10. (*Impact of the Users' QoS Requirements*) In this example we investigate the scenario in which the total number of users is $N_U = 100$, and each BS has a total

52

of $t_i = 50$ available RBs, for $i \in B$. Similar to Example 9, we consider the user-to-BS associations generated using the max-SINR, the RE and the proposed SDR-based algorithms. However, in this example, we compare the performance of these algorithms with the increase in the QoS requirements. This comparison is depicted in Figure 4.3.

From Figure 4.3, it can be seen that the proposed SDR-based algorithm performs better than both the max-SINR and the RE algorithms over the entire range of QoS requirements. However, its advantage over these algorithms is more pronounced at high QoS requirements. For instance, when the users' QoS requirement is 1000 kbps, the proposed algorithm can accommodate the demands of about 32% more users than the RE algorithm with the biasing parameter of 5 dB, about 36% more users than the RE algorithm with the biasing parameter of 10 dB, and about 29% more users than the max-SINR algorithm. At low QoSs, all algorithms can accommodate a large percentage of users, and at high QoSs, the number of available of RBs becomes the main bottleneck, thereby limiting the performance of the three algorithms. $\qquad\square$

4.6 Conclusion

This work investigated the problem of associating users to BSs in a resource-limited downlink scenario of a multi-tier HetNet. In this scenario, the users may have disparate QoS demands and the BSs may have disparate powers and disparate number of available RBs. We proposed a QoS-guaranteed user-to-BS association algorithm based on SDR and Gaussian randomization. This algorithm has polynomial complexity and provable approximation accuracy. In comparison with currently available techniques, the proposed algorithm enables the accommodation of more users with less resources.
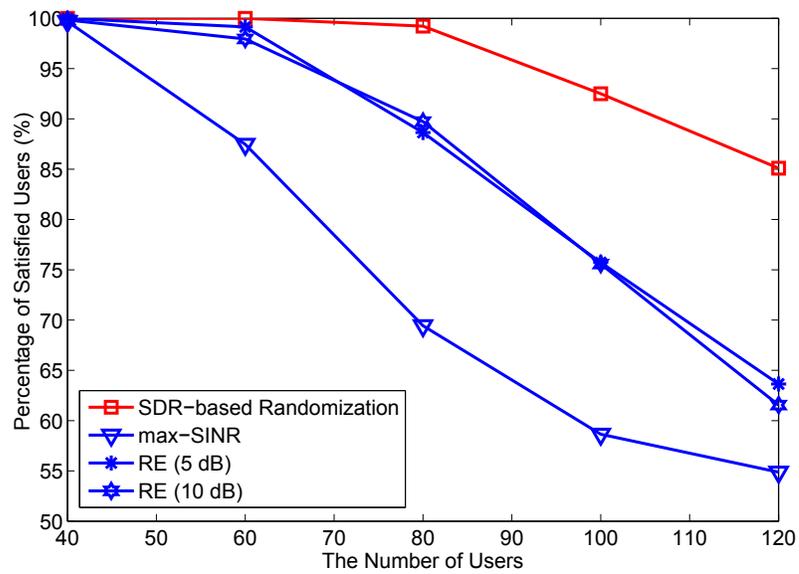
Figure 4.2: Percentage of satisfied users versus the total number of user accessing the network.
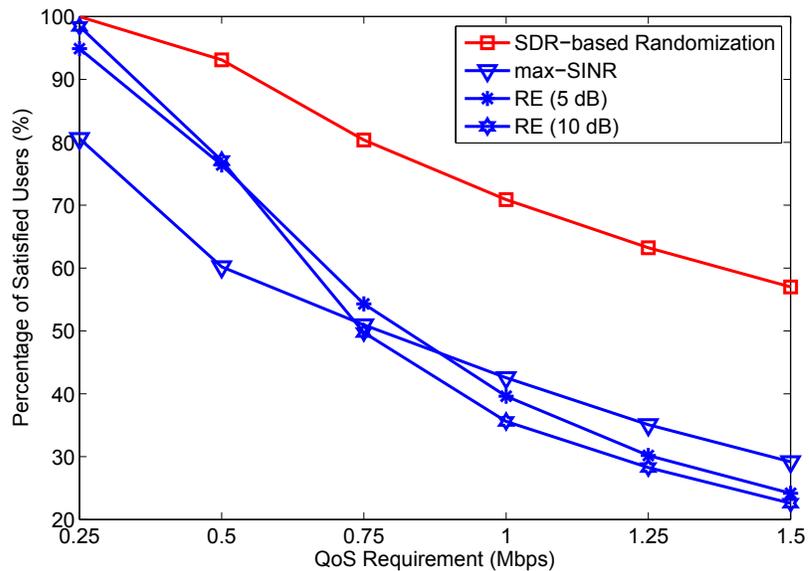
Figure 4.3: Percentage of satisfied users versus QoS requirements.

Chapter 5

A Novel Approach for QoS-Aware Joint User Association, Resource Block and Discrete Power Allocation in Downlink Heterogeneous Cellular Networks

5.1 Introduction

Heterogeneous networks (HetNets) are wireless networks that incorporate low-power pico and femto base stations (BSs) in addition to conventional macro BSs. The low cost and flexibility of deploying such networks can not only be used to increase their capacity, but also to extend their coverage and to provide better user experience. To realize the benefits of HetNets and to attain the desired network performance, the interplay between various network functionalities must be taken into consideration. For instance, the way in which users are associated to the BSs and the time-frequency resource blocks (RBs) available to the system can have a fundamental impact on the number of users that can be accommodated. Likewise, the way in which the BSs allocate their available power across the RBs can have a significant impact on the amount of physical resources required for the network to attain particular performance goals.

For the users to be associated to the BSs in a resource-efficient manner, several considerations must be taken into account. These include the channel conditions between the users and the BSs, the load condition of each BS, and the quality-of-service (QoS) demanded by each user.[1] A conventional approach for associating users to BSs is the one in which each user is associated to the BS with the maximum

---

[1]Similar to [47], herein we assume that QoS is measured by the data rate delivered to the users.

signal-to-interference-plus-noise ratio (max-SINR). This approach is simple, but can lead to highly imbalanced BS loading and does not guarantee users' QoS. In particular, this methodology can result in severely overloading the high-power macro BSs while leaving the low-power pico and femto BSs largely underutilized. Such load imbalance does not take advantage of the pico and femto BSs in supporting more users in the network.

In addition to user-to-BS association, the number of users that can be accommodated in the network is affected by the way in which the RBs are allocated to each user [48, 49]. For instance, granting a user the right to utilize its preferred RB can result in depriving other users of the opportunity to communicate altogether. Hence, it can be seen that accommodating more users with QoS constraints in the network requires a mechanism that takes into consideration not only the local network conditions of individual users, but rather the global conditions of all users. In a complementary fashion, it is desirable for the design to utilize minimal physical resources to accommodate a given network load. For instance, even if the number of users is fixed and the optimal user-to-BS associations are known *a priori*, the available RBs must be allocated to the users in a manner that avoids excessive consumption of valuable physical resources.

Efficient utilization of resources is coupled with the way in which power is allocated across the available RBs. For practical considerations, transmit powers usually assume discrete, rather than continuous, levels. For instance, networks based on the Long Term Evolution (LTE) standards can only support discrete power levels in the downlink [18, 35]. Restricting the power allocations to be discrete simplifies the hardware design and reduces the cost of practical transmitters. To see the effect of the available power on the allocation of RBs, we note that a set of RBs can be allocated to a particular user only if the power available at the BS to transmit on those RBs suffices to meet the user's demanded QoS. Hence, allocating power across RBs is coupled by the way in which the RBs are assigned to the users, and by the way in which the users are assigned to the BSs. Sequential consideration of these functionalities simplifies design, at the expense of reducing the user capacity and exhausting valuable network resources. To avoid these drawbacks it is essential to jointly optimize the user-to-BS associations and the RBs and the power allocations to ensure that the maximum number of QoS-constrained users can be accommodated with the least possible physical resources.

In this work, two instances of downlink HetNets with opportunistic RB-reuse are investigated. In the first instance the user-to-BS associations, and the RBs and power level allocations are time-shared [50], which implies that the corresponding assignment variables assume continuous values. Using this, we develop convex formulations for tight upper and lower bounds on the design objective. In the second instance, no time-sharing is considered and the user-to-BS associations, and the RBs and power level allocations, once determined, are restricted to be fixed during the entire signalling interval. In this case the assignment variables assume binary values and the resulting optimization problem is non-convex. In both instances, the formulations automatically invoke opportunistic RB-reuse. In other words, for each RB, they allow each BS to either use the RB exclusively at any given instance or to share it concurrently with another BS, i.e., they do not enforce the BSs to reuse RBs. Moreover, in both instances each user can be associated by one or multiple BSs as in [41, 51, 52]. The case with time-sharing generalizes the one without time-sharing. Hence, it offers the potential of achieving significant performance advantages, however at the expense of increasing signalling overhead. This is especially true in networks with time-varying channels [50]. In such situations, it might be more practical to restrict the RB schedules to be binary, thereby implying that each RB is assigned to a particular user for the entire signalling interval [48, 53]. Generally speaking, user-to-BS association typically takes place at a time scale larger than the one in which the resources are allocated. However, jointly optimizing the user-to-BS associations and the resource allocations can be used at system initialization and at periodic, yet relatively wide, intervals thereafter. This will provide the system with the optimal signalling parameters at those periodic instants. In between times, subproblems for optimizing power and RB allocations with fixed user-to-BS associations can be readily derived from the formulations provided in our work. Furthermore, joint optimization of the user-to-BS association and power and RB allocations, even though not always feasible, provides a benchmark for the maximum number of users that can be supported in the network.

5.1.1 Related Work

Methods for determining 'good' user-to-BS associations have been considered in the literature, mostly in isolation from other network design aspects. For instance, associating users to the BSs with the highest SINR, as in the max-SINR technique, was the conventional method up to LTE release-8 [47]. Despite being arguably the simplest

approach, it does not take the effect of loading into account and can therefore result in severe load imbalance across the system. A modified approach that attempts to alleviate this drawback was proposed in [54]. In this approach, a user is associated with the BSs that would yield the maximum expected throughput rather than the SNR. Since the expected throughput depends on the signal strength and the population of users served by the BS, this heuristic approach provides better load balancing in comparison to the max-SINR one. Another heuristic technique for user-to-BS association is the one based on range expansion (RE) [40]. In that approach, the traffic from macro BSs is offloaded to low-power BSs by adding a positive bias to their measured SINRs. An advantage of this approach is that it does not depend on the relative location of users and BSs. However, a major weakness of it is the lack of guidance for choosing the biasing factor. Algorithms for determining user-to-BS associations by finding approximate solutions to relaxed network utility maximization problems using dual pricing methods were proposed in [41] and [42]. Those methods do not account for the QoS requirements. Other approaches that invoke game-theoretic techniques for generating user-to-BS associations have been considered in [55] and [56]. The techniques that underlie those approaches do not necessarily converge and the associations that they yield can be highly suboptimal. In [45], semidefinite relaxation (SDR) technique were employed in a HetNet scenario, which is restricted to have one macro and one pico BS per cell and the design objective was to maximize the total sum rate irrespective of the QoSs demanded by the users. Despite the potential benefits offered by these approaches, the fact that they do not incorporate other key design aspects, such as RB and power allocations, usually limits the extent of their practical effectiveness.

To increase their effectiveness, the user-to-BS association techniques have been considered in conjunction with power allocation ones in [57–59]. Such techniques were developed for various HetNet design objectives, but with fixed RB allocations, which usually facilitates formulating the design problem. For instance, in [57] the design problem corresponding to maximizing the number of users in the system considering user-to-BS associations and continuous power allocations was formulated for the case in which the users have prescribed QoS demands. The resulting problem is non-convex, but was solved in [57] using a technique based on a branch-and-bound type approach called Bender's decomposition. This technique is efficient in many practical scenarios, but its worst case complexity is non-polynomial. A somewhat related problem, is the one considered in [58]. Therein, the user-to-BS associations and the continuous power

allocations that maximize the minimum SINR of the users of the downlink of a HetNet were investigated, but without QoS guarantees. It was shown in [58] that this problem is NP-hard unless the number of BSs is equal to the number of users. In the latter case, the problem has polynomial complexity. Other instances of jointly optimizing user-to-BS associations and power allocations in HetNets were considered in, e.g., [59]. Finding the global solution in such instances constitutes a formidable task, but local solutions could be obtained efficiently using iterative techniques.

The efficiency with which the physical resources are utilized can be further enhanced by jointly considering user-to-BS associations along with both RB and power allocations. Although no results seem to be available for the practical case of discrete power allocations, the case with continuous power allocations seem to have attracted attention. For instance, such power allocations are jointly optimized with user-to-BS associations and RB allocations using a greedy approach in [39] to maximize a proportional fairness metric of the rates delivered to the HetNet users. Similarly, user-to-BS associations, RB and continuous power allocations are jointly optimized in [60] using a relaxed formulation and an iterative approach that aims at maximizing the throughput of the HetNet. Another approach that considers optimizing these aspects jointly is developed in [61] using, the so-called, Gibbs sampler. This approach aims at minimizing the system-wide delay experienced by the users, but without considering their QoS demands.

5.1.2   Contribution and Comparison with Related Work

Having provided some background on the designs that consider the optimization of user-to-BS associations, and RBs and power allocations either separately or jointly, we now provide a brief account of our contributions. Specifically, the contributions of the work are the following: 1) We introduce a novel framework for jointly optimizing the user-to-BS associations, and the RBs and power allocations in HetNets with potentially disparate power budgets and spectral resources. 2) We consider two cases: with and without time-sharing. For both instances, we develop optimization frameworks and polynomial-complexity algorithms that maximizes the number of users accommodated in the network and, at the same time, minimizes the network resources consumed to accommodate those users. The designs proposed herein for the two instances possess the following distinguishing features in comparison with related work.

- The design objective considered herein is to jointly maximize the number of network users while ensuring minimal consumption of the physical resources. Hence, this objective relates directly to the revenue of the system operator and the satisfaction of a larger set of users [57]. This is in contrast with other objectives that are commonly considered in the literature. For instance, sum-rate maximization may result in a small number of users being served by the network and may therefore impact the revenue of the service provider.

- The transmit power levels considered herein are constrained to assume discrete, rather than continuous, values. This feature renders the algorithms developed herein amenable to practical systems. Although most existing resource alloca-tion schemes consider continuous transmit power, applying those schemes in practical systems can be problematic; rounding does not guarantee optimality. Hence, it can be seen that our work makes a step towards bridging the divide between theoretical-designs and practical implementation of resource allocation algorithms. We show that, when time-sharing is considered, imposing the re-striction on the power levels to be discrete does not affect the convexity of the formulations for upper and lower bounds. However, when time-sharing is not considered, the restriction on the power levels to be discrete usually re-sults in difficult-to-solve non-convex formulations. This difficulty is alleviated here by invoking the SDR technique which usually generates close-to-optimal performance by invoking a careful relaxation on the rank of the matrix-value optimization variable. This technique enables to perform joint optimization for HetNets with a relatively large number of BSs, RBs, users, and power levels. Such an optimization would be computationally prohibitive to perform with exhaustive search.

- The proposed algorithms in the time-sharing and the no time-sharing cases allow the RBs to be reused opportunistically. Hence, these algorithms offer valuable design flexibility and enable higher spectral efficiencies to be achieved.

5.2   System Model

In this section we describe the downlink of a two-tier HetNet, in which tier-1 consists of one or more macro BSs and tier-2 consists of one or more pico BSs. The main difference

between these BSs is the available power budgets; macro BSs have power budgets that are manifold of the power budgets of their pico counterparts. Subsequently, the area of the cell covered by a pico BS is usually much smaller than that covered by a macro BS.

It is desirable in such a HetNet scenario to optimize the network parameters in order to maximize the number of users supported by the network, while minimizing the amount of consumed resources. To achieve this goal, we consider a centralized design whereby a central entity that is aware of the network parameters controls the association of users to BSs, the allocation of RBs among users, and the allocation of available power budget across RBs. In our work, we consider a setup similar to the one considered in [41] and [42], wherein the BSs, macro or pico, share a common pool of RBs. This setup offers design flexibility and avoids the penalty of partitioning the available frequency band into non-overlapping sections. Now, depending on the power budget of each BS and the channel to prospective users, it may be beneficial for one BS to serve one user on a particular RB, but not on another one. The availability of the RB depends on whether it is being used to serve other users, and whether RB-reuse is allowed. In a complementary fashion, it can be seen that an RB may be suitable to serve a particular user from one BS, but not from another BS, depending on the aforementioned factors. Hence, it can be seen that, whether RB-reuse is allowed or not, RB assignment and BS association are tightly coupled; a change in one aspect can/will readily affect the other aspect.

To expose the relationship between the network parameters and the design objective, we denote the set of all BSs, including macro and pico ones, by $\mathcal{B} = \{1, \ldots, B\}$ and the set of all users by $\mathcal{U} = \{1, \ldots, U\}$. The wide range of services, e.g., voice and non-voice applications, requested by those users render their QoS demands rather disparate. We denote the QoS demanded by user u by Q_u, $u \in \mathcal{U}$, and it will be measured in megabits per second (Mbps). To satisfy these demands, the service provider allocates a different number of RBs to each user; higher QoS requirement implies higher usage of RBs. In addition, the usage of RBs depends on the channel conditions observed by the user; weaker channels imply higher usage of RBs. For notational convenience, we will denote the set of RBs available to all the BSs in the system by $\mathcal{S} = \{1, \ldots, S\}$.

For practical considerations, it is desirable for each BS to use a finite set of discrete power levels [18]. We will denote the set of power levels utilized by the b-th BSs by

$\mathcal{P}_b = \{p_b^1, \ldots, p_b^L\}$, where $L = |\mathcal{P}_b|$ is the cardinality of $\mathcal{P}_b$, $b \in \mathcal{B}$. Note that, because we are considering a scenario with a heterogeneous infrastructure, the set $\mathcal{P}_b$ may vary across BSs. For instance, this set may depend on the total amount of transmit power available at each BS.

To complete the characterization of the considered HetNet, we will denote the channel gain between the b-th BS and the u-th user on the s-th RB by h_{ub}^s, $u \in \mathcal{U}$, $b \in \mathcal{B}$, $s \in \mathcal{S}$. These gains account for path-loss, log-normal shadowing, and fading and will be assumed to be random and fixed during the signalling interval. We will use W_0 to denote the bandwidth of each RB, and N_0 to denote the power spectral density of the additive white Gaussian noise at the receiver.

In the scenario considered herein, a central entity is assumed to have access to accurate channel estimates and power budgets. The central entity optimizes the user-to-BS associations, and the RB and power allocations and sends its decisions to the users and the BSs. Two cases can be considered, viz., without and with RB-reuse. In the first case, the central entity opts for design and processing simplicity by avoiding RB-reuse and assigning each RB exclusively to one user throughout [62]. In contrast, in the second case, the central entity opts for higher spectral efficiency by allowing RB-reuse and concurrently assigning each RB to potentially multiple users. The second case subsumes the first one. In other words, allowing RB-reuse, in the sense of the forthcoming formulations, encompasses the case with no RB-reuse. Hence, we will focus on the case with RB-reuse, and allude to the other case as necessary. To avoid overly complicated exposition, we will restrict attention on the case in which at most two users within the HetNet can share the same RB at any given instant. Allowing each RB to be shared concurrently by multiple users results in contaminating the users' signal with interference from neighbouring BSs. In this case, the power of the signal received by the u-th user, $u \in \mathcal{U}$, on the s-th RB, $s \in \mathcal{S}$, from the b-th BS, $b \in \mathcal{B}$, when it uses the ℓ-th power level, $\ell \in \mathcal{P}_b$, is given by $p_b^\ell |h_{bu}^s|^2$. The interference signal that this user observes from the k-th BS, $k \in \mathcal{B} \setminus b$, when it uses the n-th power level, $n \in \mathcal{P}_k$, is given by $p_k^n |h_{ku}^s|^2$. Subsequently, the SINR of the signal can be expressed as

$$\gamma_{bu}^{s\ell kn} = \frac{p_b^\ell |h_{bu}^s|^2}{W_0 N_0 + p_k^n |h_{ku}^s|^2}, \tag{5.1}$$

and the corresponding rate is given by

$$R_{bu}^{s\ell kn} = W_0 \log_2(1 + \gamma_{bu}^{s\ell kn}) \quad \text{[bits per second]}. \tag{5.2}$$

For notational convenience, we will use $\gamma_{bu}^{s\ell00} = \frac{p_b^\ell |h_{bu}^s|^2}{W_0 N_0}$ to represent the signal-to-noise ratio (SNR) when the s-th RB is not reused. To proceed, we note that $\gamma_{bu}^{s\ell kn}$ and $R_{bu}^{s\ell kn}$ are determined by the ordered sextet (b, u, s, ℓ, k, n), that is, a given sextet (b, u, s, ℓ, k, n) corresponds to values of $\gamma_{bu}^{s\ell kn}$ and $R_{bu}^{s\ell kn}$ that can be determined *a priori*. We will use this observation in the prospective formulation to distill the optimization of user-to-BS associations, and RBs and power allocations to the optimization of sextets as will be elucidated in the next section.

5.3   Objective and System Constraints

In this section, we will use the sextet (b, u, s, ℓ, k, n) to describe the design objective and the system constraints. We will show that the problem of associating users to BSs and allocating RBs and power levels amounts to determining an optimal set of sextets from among all admissible ones. To determine admissible sextets, we introduce an indicator variable $y_{bu}^{s\ell kn}$ for each sextet (b, u, s, ℓ, k, n), $b \in \mathcal{B}$, $u \in \mathcal{U}$, $s \in \mathcal{S}$, $\ell \in \mathcal{P}_b$, $k \in \{0\} \cup \mathcal{B} \setminus b$, and $n \in \mathcal{P}_k$. When time-sharing is considered, $y_{bu}^{s\ell kn}$ represents the fraction of time during which the sextet (b, u, s, ℓ, k, n) is used for transmission, and hence, $y_{bu}^{s\ell kn} \in [0, 1]$. In contrast, when time-sharing is not considered, $y_{bu}^{s\ell kn}$ is 1 if the sextet (b, u, s, ℓ, k, n) is used for transmission and is zero otherwise, i.e., $y_{bu}^{s\ell kn} \in \{0, 1\}$, see Fig. 5.1.
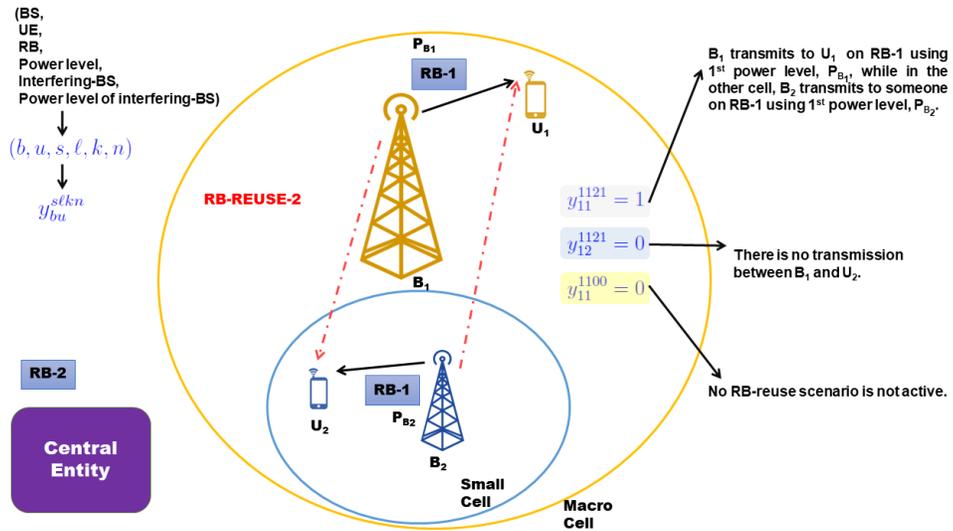

Figure 5.1: Indicator variables used to determine admissible sextets.

5.3.1 System Constraints

BS Power Budget Constraints In a HetNet, the total transmission power budget may vary significantly between macro and pico BSs. We will denote such a budget by $P_b^{\max}$, $b \in B$. Hence, the total power transmitted by the b-th BS to all the users in the network must satisfy

$$\sum_{u \in \mathcal{U}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} \right) p_b^\ell \leq P_b^{\max}, \quad b \in \mathcal{B}. \tag{5.3}$$

QoS Constraints The objective of the design developed herein is to maximize the number of users. Hence, it must ensure that the users' QoS demands are met. Otherwise, infinitely many users can be trivially accommodated, i.e., with vanishing QoSs. To avoid such a possibility, the QoS metric is taken to be the data rate that can be reliably communicated to each user, e.g., [47]. Using (5.2), the maximum aggregate data rate that can be reliably communicated to user $u \in \mathcal{U}$, can be expressed as $\sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} R_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right)$. Hence, ensuring that the QoS requirement of user u is satisfied is equivalent to ensuring that

$$\sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} R_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right) \geq Q_u, \quad u \in \mathcal{U}. \tag{5.4}$$

This constraint must be satisfied only for those users accommodated in the network. Otherwise, it must be excluded from the formulation. However, since the design objective is to maximize the number of users, the users that will be actually accommodated in the network are not known *a priori*. To circumvent this difficulty, in Section 5.4.1, we will provide a mathematically convenient means for characterizing this conditional constraint automatically.

RB Allocation Constraints To derive a constraint on the RB allocation, let us first consider the case without RB-reuse. In this case, restricting each RB to be used at most once at any given time instant can be enforced by constraining the summation of $\{y_{bu}^{s\ell 00}\}$ over all BSs, users and power levels to be at most equal to one. Next, we consider the case with RB-reuse in which two BSs use the same RB. In this case, the summation of $\{y_{bu}^{s\ell kn}\}$ over all BSs, users, power levels and combinations of interfering

BSs with their power levels to be at most equal to two. Since our formulation allows each RB to be reused opportunistically, i.e., it can be used exclusively by one BS during a fraction of the signalling interval and RB-reused during the remainder of the signalling interval, the following constraint must be satisfied.

$$\sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} y_{bu}^{s\ell00} + \frac{1}{2} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}} \sum_{k \in \mathcal{B} \setminus b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} \leq 1, \quad s \in \mathcal{S}. \tag{5.5}$$

RB-Reuse Constraints   Since in our formulation at most two BSs can concurrently use the same RB, and each BS can use one power level at any time instant, we have:

$$y_{bu}^{s\ell00} + \sum_{k \in \mathcal{B} \setminus b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} \leq 1, \quad b \in \mathcal{B}, u \in \mathcal{U}, s \in \mathcal{S}, \ell \in \mathcal{P}_b. \tag{5.6}$$

This constraint enables RB-reuse to be invoked opportunistically; the case with no RB-reuse corresponds to $y_{bu}^{s\ell kn} = 0$, $b \in \mathcal{B}$, $\ell \in \mathcal{P}_b$, $u \in \mathcal{U}$, $k \in \mathcal{B} \setminus b$ and $n \in \mathcal{P}_k$.

Interference-Coupling Constraints   Let us consider a time-sharing case in which a particular time fraction $y_{bu}^{s\ell kn}$ is strictly greater than zero (For the corresponding case without time-sharing, $y_{bu}^{s\ell kn} = 1$.). Such $y_{bu}^{s\ell kn}$ implies that the k-th BS is transmitting a signal with power level n for at least a time fraction of $y_{bu}^{s\ell kn}$, $b \in \mathcal{B}$, $\ell \in \mathcal{P}_b$, $u \in \mathcal{U}$, $k \in \mathcal{B} \setminus b$ and $n \in \mathcal{P}_k$. This implies that $y_{k\tilde{u}}^{snb\ell}$ must be strictly greater than zero for at least one $\tilde{u} \in \mathcal{U} \setminus u$. But, since a BS can transmit to multiple users at different time instants, it follows that $y_{bu}^{s\ell kn}$ must satisfy the following constraint:

$$y_{bu}^{s\ell kn} \leq \sum_{\tilde{u} \in \mathcal{U} \setminus u} y_{k\tilde{u}}^{snb\ell}, \quad b \in \mathcal{B}, u \in \mathcal{U}, s \in \mathcal{S}, \ell \in \mathcal{P}_b, k \in \mathcal{B} \setminus b, n \in \mathcal{P}_k. \tag{5.7}$$

We make two observations. First, for the case with RB-reuse and no time-sharing when $y_{bu}^{s\ell kn} = 1$, the constraint in (5.7) ensures that $y_{k\tilde{u}}^{snb\ell} = 1$ for exactly one user $\tilde{u} \in \mathcal{U} \setminus u$. Second, (5.7) is relevant when $k \neq 0$, and is trivial otherwise, i.e., for RBs that are not reused.

User Association Constraints   In our framework, a BS cannot simultaneously use multiple power levels nor broadcast to multiple users on the same RB. This yields

the following constraint:

$$\sum_{\ell \in \mathcal{P}_b} \sum_{u \in \mathcal{U}} \left(y_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} \right) \leq 1, \quad b \in \mathcal{B}, s \in \mathcal{S}. \tag{5.8}$$

BS Association Constraints   At any time instant and RB, each user can be served by at most one BS with one power level. This yields the following constraint:

$$\sum_{b \in \mathcal{B}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} \right) \leq 1, \quad u \in \mathcal{U}, s \in \mathcal{S}. \tag{5.9}$$

5.3.2   Design Objective

The design objective considered in this work is composed of two components. The first characterizes the number of users accommodated by the system. Maximizing this component directly contributes to maximizing revenue and utility. The second component characterizes RB usage and ought to be minimized provided that the accommodated users' QoS constraints are satisfied. We will begin by providing the mathematical description of the second component.

Aggregate RB Usage   The aggregate usage of RBs assigned to the u-th user can be expressed as $\sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} \right)$, $u \in \mathcal{U}$. This usage has slightly different meanings in the cases with and without time-sharing. To see that, consider the case in which there is only one RB. When time-sharing is considered, the summation of $\{y_{bu}^{s\ell kn}\}$ corresponds to the normalized total time over which this RB is used. Conversely, when time-sharing is not considered, this summation serves as an indicator whether this RB is used or not. Hence, the aggregate usage of RBs in the network can be expressed as $\sum_{u \in \mathcal{U}} \sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} \right)$. Minimizing this summation directly reduces the amount of resources exhausted by the system, and indirectly contributes to increasing revenue.

Total Number of Accommodated Users   The number of users accommodated is indirectly related to RB usage, and the number of BSs and power levels. To see that, we recall that in the current system each user may be served by multiple BSs over multiple RBs with different power levels at every time instant. Hence, a user

$u \in \mathcal{U}$ is guaranteed to be served whenever at least one of the assignment variables in the set $\{y_{bu}^{s\ell kn}\}$, $s \in \mathcal{S}$, $\ell \in \mathcal{P}_b$, $b \in \mathcal{B}$, $k \in \{0\} \cup \mathcal{B} \setminus b$, $n \in \mathcal{P}_k$, is strictly greater than zero. But since $\{y_{bu}^{s\ell kn}\}$ are restricted to be non-negative, it can be readily seen that a user $u \in \mathcal{U}$ is guaranteed to be served if and only if $\sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} R_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \setminus b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right) > 0$. In other words, a user $u \in \mathcal{U}$ is not accommodated in the system if and only if $\sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} R_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \setminus b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right) = 0$. Hence, to indicate whether a particular user $u \in \mathcal{U}$ is accommodated in the network or not, we define the indicator function $\iota(x)$ such that $\iota(x) = 1$ for $x > 0$ and $\iota(x) = 0$ for $x = 0$. To use $\iota(x)$ as a user accommodation indicator, the variable x is replaced with the normalized[2] aggregate rate of the u-th user, $\sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \frac{1}{Q_u} \left(y_{bu}^{s\ell 00} R_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \setminus b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right)$. Unfortunately, the indicator function is not smooth and hence cannot be readily incorporated in the optimization framework. To circumvent this difficulty, we use an alternate representation. We express it as

$$\iota(x) = \lim_{\sigma \to \infty} 1 - e^{-\sigma x}. \tag{5.10}$$

This representation implies that if σ is chosen to be sufficiently large, $1 - e^{-\sigma x}$ will provide a smooth approximation of $\iota(x)$. A proper choice of σ depends on whether time-sharing is considered or not. In particular, when time-sharing is considered, we will show that the choice of a good value of σ is straightforward, whereas when time-sharing is not considered, a judicious choice of this value must be made. We will elaborate on the choice of the value of σ for the latter case in Section 5.5.2.

A Composite Objective   Our design objective is composed of the linear combination of two quantities: 1) the number of accommodated users, which ought to be maximized to increase revenue, and 2) the aggregate RB usage, which ought to be minimized to efficiently utilize the network resources, see Fig. 5.2. To synthesize this objective, we will weight each of its components by an appropriate scalar $\rho \in [0, 1]$. Using ρ, we define our design objective to be maximizing $\rho \sum_{u \in \mathcal{U}} \left(1 - e^{-\frac{\sigma}{Q_u} \sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} R_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \setminus b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right)} \right) - (1 - \rho) \sum_{u \in \mathcal{U}} \sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \setminus b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} \right)$. Increasing ρ from 0 to 1 in this objective shifts the emphasis of the optimization from minimizing the usage of the resources used in the network to maximizing the number of users accommodated

---

[2]Normalization by the QoS of individual users is introduced for numerical convenience.
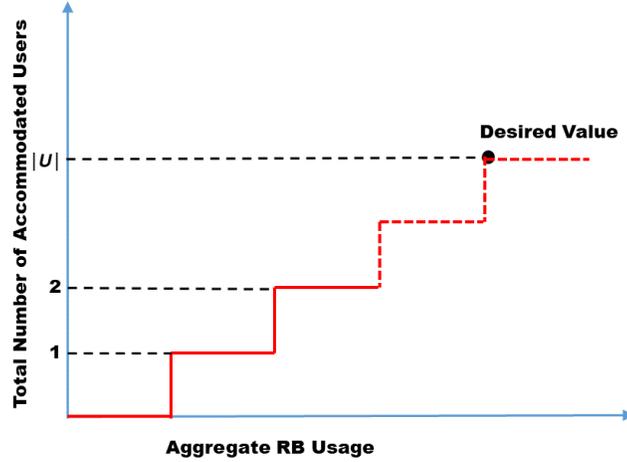
Figure 5.2: Illustration of the considered composite design objective.

therein. As such, ρ parameterizes a family of objectives, each of which corresponds to a number of accommodated users and a level of aggregate RB usage. In order to ensure that the number of users accommodated in the network is maximized, the optimal choice of the value of ρ is stated in the following proposition.

Proposition 1. *Any value of $\rho \in \left(\frac{2S}{1+2S}, 1 \right)$ ensures maximizing the number of users accommodated in the network.*

Proof. See Appendix B. ∎

5.4 Optimization Based Formulation With and Without Time-Sharing

5.4.1 Network Instance With Time-Sharing

We consider a network instance with time-sharing, where the indicator variables $\{y_{bu}^{s\ell kn}\}$ are continuous on the interval $[0,1]$. Using a value of ρ within the range suggested in Proposition 1, we can now provide a mathematical formulation for the design that jointly optimizes the user-to-BS associations, the RBs and the power allocations. Before we do that, we provide a mathematically convenient means for characterizing the QoS constraint in (5.4). To do that, we will use the approximation

68

of $\iota(\cdot)$ in Section 5.3.2. In particular, the constraint in (5.4) is equivalent to multiplying the right hand side of the inequality therein with the indicator function used in the design objective with σ sufficiently large. Using this characterization, the joint design problem can be cast in the following form:

$$\max_{\mathbf{y},\,\mathbf{t}} \quad \rho \sum_{u\in\mathcal{U}}(1-t_u) - (1-\rho)\sum_{s\in\mathcal{S}}\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k}y_{bu}^{s\ell kn}\right),$$

$$(5.11\text{a})$$

$$\text{subject to} \quad \sum_{u\in\mathcal{U}}\sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k}y_{bu}^{s\ell kn}\right)p_b^\ell \leq P_b^{\max}, \quad b\in\mathcal{B}, \quad (5.11\text{b})$$

$$\sum_{b\in\mathcal{B}}\sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00}R_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k}y_{bu}^{s\ell kn}R_{bu}^{s\ell kn}\right) \geq Q_u(1-t_u), \quad u\in\mathcal{U},$$

$$(5.11\text{c})$$

$$t_u = e^{-\frac{\sigma}{Q_u}\sum_{b\in\mathcal{B}}\sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00}R_{bu}^{s\ell00}+\sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k}y_{bu}^{s\ell kn}R_{bu}^{s\ell kn}\right)}, \quad u\in\mathcal{U},$$

$$(5.11\text{d})$$

$$\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}}\sum_{\ell\in\mathcal{P}_b}y_{bu}^{s\ell00} + \frac{1}{2}\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}}\sum_{\ell\in\mathcal{P}_k}\sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}}y_{bu}^{s\ell kn} \leq 1, \quad s\in\mathcal{S}, \quad (5.11\text{e})$$

$$y_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k}y_{bu}^{s\ell kn} \leq 1, \quad b\in\mathcal{B},\ u\in\mathcal{U}, s\in\mathcal{S}, \ell\in\mathcal{P}_b, \quad (5.11\text{f})$$

$$y_{bu}^{s\ell kn} \leq \sum_{\tilde{u}\in\mathcal{U}\backslash u}y_{k\tilde{u}}^{snb\ell}, \quad b\in\mathcal{B},\ u\in\mathcal{U},\ s\in\mathcal{S},\ \ell\in\mathcal{P}_b,\ k\in\mathcal{B}\backslash b,\ n\in\mathcal{P}_k,$$

$$(5.11\text{g})$$

$$\sum_{\ell\in\mathcal{P}_b}\sum_{u\in\mathcal{U}}\left(y_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k}y_{bu}^{s\ell kn}\right) \leq 1, \quad b\in\mathcal{B}, s\in\mathcal{S}, \quad (5.11\text{h})$$

$$\sum_{b\in\mathcal{B}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k}y_{bu}^{s\ell kn}\right) \leq 1, \quad u\in\mathcal{U}, s\in\mathcal{S}, \quad (5.11\text{i})$$

$$t_u \geq 0, \quad u\in\mathcal{U}, \quad (5.11\text{j})$$

$$y_{bu}^{s\ell kn} \in [0,1], \quad b\in\mathcal{B},\ u\in\mathcal{U},\ s\in\mathcal{S},\ \ell\in\mathcal{P}_b,\ k\in\{0\}\cup\mathcal{B}\backslash b,\ n\in\mathcal{P}_k,$$

$$(5.11\text{k})$$

where (5.11c) and (5.11d) are equivalent to (5.4) when $\sigma \to \infty$. In other words, similar to (5.4), for large σ, (5.11c) and (5.11d) ensure that the QoS constraints are enforced only for the users accommodated in the network and are immaterial otherwise; unaccommodated users cannot impose QoS constraints. When $\sigma \to \infty$, t_u

in (5.11d) indicates whether the u-th user $u \in \mathcal{U}$ is accommodated in the network or not. If accommodated, $t_u = 0$, otherwise $t_u = 1$.

Before proceeding with our formulation, we note that it is usually possible for the optimization of time-shares to be cast in a convex form because time-sharing of two feasible solutions is equivalent to finding a point in their convex hull. However, the objective of the optimization at hand is not convex because of the (discrete) number of users to be maximized. This discreteness is captured by (5.11d) with large σ. Hence, although time-sharing results in a convex feasible set, the design problem remains non-convex because of the discrete component of the objective.

The accuracy of the approximation in the objective, (5.11a), and the constraints in (5.11c) and (5.11d) is controlled by σ; the value of σ is generally not critical, and choosing σ to be large, e.g., $\sigma = 100$, ensures that each summand in the first term of the objective and the right hand side of (5.11c) is essentially equal to one whenever the summation in the exponent is greater than zero. In other words, choosing σ in (5.11d) to be large ensures that whenever the aggregate user rate normalized to its QoS is bounded away from zero, $t_u \approx 0$ and (5.11c) is enforced; when the normalized aggregate user rate is close to zero $t_u \approx 1$ and (5.11c) is trivial.

From the formulation in (5.11), it can be seen that all constraints are linear except the non-affine equality constraint in (5.11d), which renders the formulation non-convex, and hence difficult to solve. To circumvent this difficulty, we relax the constraint (5.11d) to be an inequality in the form

$$t_u \geq e^{-\frac{\sigma}{Q_u} \sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}} \left(y_{bu}^{s\ell00} R_{bu}^{s\ell00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}} y_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right)}, \quad u \in \mathcal{U}. \tag{5.12}$$

This constraint together with the objective imply that when the normalized rate of the u-th user is zero, t_u is equal to one and (5.11c) holds with equality. However, when the normalized rate of the u-th user is greater than zero, the new constraint in (5.12) may be inactive, and t_u can take on a value between 0 and 1. In this case, the fact that the objective is monotonically increasing in $(1 - t_u)$, $u \in \mathcal{U}$, implies that (5.11c) will be active. In other words, for every user, $u \in \mathcal{U}$, at least one of the two inequality constraints in (5.11c), and (5.12) will be satisfied with equality. More specifically, let us define $\alpha_u \triangleq \frac{\sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}} \left(y_{bu}^{s\ell00} R_{bu}^{s\ell00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}} y_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right)}{Q_u}$. We have the following three operating regions:

1. When $\alpha_u = 0$, there is no resource allocated to user u and its rate is zero. In

this case, both (5.11c), and (5.12) hold with equality and the relaxation is tight.

2. When $\alpha_u \geq 1$, the QoS requirement of user u is achieved. In this case, both (5.11c), and (5.12) hold with equality and the relaxation is tight.

3. When $\alpha_u \in (0, 1)$, the QoS requirement of user u is neither low enough to guarantee its support in the network nor high enough to be definitely denied access to the network. In this case, either (5.11c), or (5.12) holds with equality and the other one holds with strict inequality, and the relaxation is not tight.

Now we will show how (5.12) will be used to obtain upper and lower bounds on the number of supported users. Using (5.12), the relaxed version of the problem in (5.11) can be expressed as

$$
\max_{\mathbf{y},\ \mathbf{t}} \quad \rho \sum_{u \in \mathcal{U}} (1 - t_u) - (1 - \rho) \sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} \right),
$$
(5.13a)

subject to   (5.11b), (5.11c), (5.11e)–(5.11k) and (5.12).   (5.13b)

The formulation in (5.13) comprises an efficiently solvable convex problem; objective and constraints are convex. To see how (5.13) can be used to generate upper and lower bounds on the objective, we consider the following possibilities:

- If (5.12) holds with equality for every $u \in \mathcal{U}$, then (5.13) is solved optimally, and the cardinality of the set of u's for which the corresponding $t_u = 0$ is equal to the maximum number of users that can be supported in the network.

- If (5.12) does not hold with equality for a set of users in, say $\tilde{\mathcal{U}} \subset \mathcal{U}$, we consider the cardinality of this set, i.e., $|\tilde{\mathcal{U}}|$. We have two cases: either $|\tilde{\mathcal{U}}| = 1$, or $|\tilde{\mathcal{U}}| > 1$.

  – If $|\tilde{\mathcal{U}}| = 1$, then t_u assumes a fractional value for exactly one user, i.e., $t_u \in (0, 1)$. But, since one component of the objective is to minimize $\sum_{u \in \mathcal{U}} t_u$, it can be readily seen that not allowing the user with fractional t_u to join the network is optimal, and the cardinality of the set of u's for which the corresponding $t_u = 0$ is equal to the maximum number of users that can be supported in the network.

  – If $|\tilde{\mathcal{U}}| > 1$, then the cardinality of the set of u's for which the corresponding $t_u = 0$ provides a lower bound on the maximum number of users that can be

supported. In this case, an upper bound on the maximum number of users can be obtained by assuming that the users in $\tilde{\mathcal{U}}$ can also be incorporated in the network.

In particular, suppose that the maximum number of users than can be accommodated in the network is given by U^*, i.e., $U^* = \sum_{u \in \mathcal{U}} (1 - t_u^*)$, where t_u^* is generated by (5.11). To obtain an upper bound on U^*, we note that when $\{t_u\}$ generated by (5.13) assume fractional values, the first component of the objective in (5.13), $\sum_{u \in \mathcal{U}} (1 - t_u)$, also assumes a fractional value and an upper bound on U^* is given by $\left\lfloor \sum_{u \in \mathcal{U}} (1 - t_u) \right\rfloor$. To obtain a lower bound, we round each non-zero value of t_u to 1. This yields that a lower bound on the number of users that can be accommodated in the system is given by $\sum_{u \in \mathcal{U}} \lfloor (1 - t_u) \rfloor$. In other words, the maximum number of users than can be accommodated in the network, U^*, satisfies the following inequalities:

$$\sum_{u \in \mathcal{U}} \lfloor (1 - t_u) \rfloor \leq U^* \leq \left\lfloor \sum_{u \in \mathcal{U}} (1 - t_u) \right\rfloor. \tag{5.14}$$

We conclude by noting that, even though the power levels in (5.13) are constrained to be discrete, the fact that these levels are determined by indicator functions over a continuous interval enables us to develop a convex framework for generating tight bounds on the objective. This framework represents an instance in which design complexity is traded for dimensionality.

5.4.2   Network Instance Without Time-Sharing

In contrast with the time-sharing-based formulation, herein we consider an instance of a network without time-sharing. This framework yields a formulation analogous to that in (5.11) but with the continuous constraint in (5.11k) replaced with the binary constraint $y_{bu}^{s\ell kn} \in \{0, 1\}$, $\forall b, u, s, \ell, k, n$. We have the following optimization problem:

$$\max_{\mathbf{y}, \, \mathbf{t}} \quad \rho \sum_{u \in \mathcal{U}} (1 - t_u) - (1 - \rho) \sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} \right), \tag{5.15a}$$

$$\text{subject to} \quad \text{(5.11b)–(5.11j)}, \tag{5.15b}$$

$$y_{bu}^{s\ell kn} \in \{0, 1\}, \ b \in \mathcal{B}, \ u \in \mathcal{U}, \ s \in \mathcal{S}, \ \ell \in \mathcal{P}_b, \ k \in \{0\} \cup \mathcal{B} \backslash b, \ n \in \mathcal{P}_k. \tag{5.15c}$$

The formulation in (5.15) is non-convex because of (5.15c) and (5.11d) in (5.15b). Similar to (5.13), an upper bound on (5.15) can be obtained by relaxing (5.11d), which yields the following formulation:

$$\max_{\mathbf{y},\,\mathbf{t}} \qquad \rho \sum_{u\in\mathcal{U}}(1-t_u) - (1-\rho)\sum_{s\in\mathcal{S}}\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn}\right),$$
$$(5.16\text{a})$$

$$\text{subject to} \qquad (5.11\text{b}),\ (5.11\text{c}),\ (5.11\text{e})\text{--}(5.11\text{j})\text{ and }(5.12), \qquad\qquad (5.16\text{b})$$

$$y_{bu}^{s\ell kn} \in \{0,1\},\ b\in\mathcal{B},\ u\in\mathcal{U},\ s\in\mathcal{S},\ \ell\in\mathcal{P}_b,\ k\in\{0\}\cup\mathcal{B}\backslash b,\ n\in\mathcal{P}_k.$$
$$(5.16\text{c})$$

Apart from the binary constraint in (5.16c), the formulation in (5.16) constitutes a convex optimization problem. However, because of the binary constraint, this problem represents an instance of a non-convex mixed integer non-linear program (MINLP). The complexity of algorithms for solving such MINLP problems is generally high, rendering their applicability in the current scenario questionable. To alleviate this difficulty, we will use a polynomial-complexity technique based on SDR with randomization [8]. However, unlike the time-sharing case, the effectiveness of the SDR-based technique depends on the choice of σ in (5.12), cf. Section 5.5.2 below.

5.5 Solution Approaches

5.5.1 Network Instance With Time-Sharing

The formulation for the optimization problem in (5.13) is convex; the objective is linear, and the constraints in (5.13b) are linear except the constraint (5.12), which is convex. Hence, it can be solved reliably and efficiently in polynomial time using, e.g., interior-point methods (IPM) [63].

Computational Complexity Analysis

The problem in (5.13) is solved only once, and the computational complexity of solving it with IPM can be obtained using the analysis pertaining to self-concordant functions. For such functions the total number of Newton steps is proportional to the number of inequality constraints. Fortunately, when the logarithmic barrier function

is invoked, the problem in (5.13) possesses the self-concordance property. We have the following result.

Proposition 2. *The computational complexity of solving the problem in* (5.13) *can be bounded by* $\mathcal{O}(n^{3.5})$, *where* $n = B + S + 3U + BS + US + BUSL + BUSL(2(B-1)L+1)$.

Proof. See Appendix C. ∎

It can be seen from this result that the computational complexity of solving (5.13) is polynomial in B, S, U, and L, which implies that it can be efficiently solved for relatively large networks.

Impact of RB-Reuse Order on Computational Complexity

Increasing the order of RB-reuse can offer a substantial increase in throughput and utility. However, this comes at a large computational cost. In particular, the number of optimization variables increases quickly with the order of reuse. This renders solving the formulation corresponding to the general case of arbitrary RB-reuse excessively time consuming, thereby reducing its practical significance. To see that, we note that the number of variables corresponding to an RB-reuse of order q can be shown to be given by $BUSL \sum_{r=0}^{q-1} \frac{(B-1)!}{(B-r-1)!} L^r$. This number grows at least exponentially with q, and hence results in a computationally prohibitive formulation except for small q. Notwithstanding the magnitude of the number of variables, our formulation is readily extensible to general RB-reuse orders, but with more optimization variables. To show that, in Appendix D, we provided the formulation corresponding to an RB-reuse order of 3.

5.5.2   Network Instance Without Time-Sharing

For room considerations, in this section we will provide explicit development of the SDR-based technique for the case with no time-sharing and no RB-reuse. Carrying over this development to the case with RB-reuse is straightforward, but consumes unaffordable room. The formulation for the case with no RB-reuse can be readily derived from (5.16) by dropping all the $y_{bu}^{s\ell kn}$ terms for which $k \neq 0$. For notational convenience, in this case, the $k = n = 0$ in the superscript can be dropped, and since the constraints in (5.11h) and (5.11i) are automatically incorporated in (5.11e), they can be dropped as well.

Similar to its RB-reuse counterpart, the no RB-reuse variant of (5.16) represents an instance of an MINLP problem, which is computationally infeasible to solve optimally for all, but a relatively small class of HetNets with a limited number of users, BSs, RBs, and power levels. To obtain resource-efficient solutions for larger networks, we will use an SDR-based technique, which, unlike techniques based on the branch-and-bound approach, has polynomial complexity. This renders it suitable for solving (5.16) and its no RB-reuse variant for more practical network scenarios. In comparison with other polynomial-complexity algorithms, when constraints and objective are quadratic, the SDR-based technique has a guaranteed approximation accuracy. For instance, for the (NP-hard) maximum-cut problem, the SDR-based technique is guaranteed to yield at least 87.56% of the optimal objective value [7].

To cast the formulation for the case with no RB-reuse in a form that is more amenable to the proposed SDR-based technique, we will express the objective and the constraints using vector notation. Towards that end, let Y be a 4-dimensional tensor with the entries given by $y_{bu}^{s\ell}$, $b \in \mathcal{B}$, $u \in \mathcal{U}$, $s \in \mathcal{S}$, $\ell \in \mathcal{P}_b$. For ease of representation, we express this tensor in the form of a $BS \times UL$ block-partitioned matrix. In particular, Y is written as a matrix of $B \times U$ blocks, $Y = \begin{bmatrix} Y_{11} & \dots & Y_{1U} \\ \vdots & \ddots & \vdots \\ Y_{B1} & \dots & Y_{BU} \end{bmatrix}$,

where $Y_{bu} = \begin{bmatrix} y_{bu}^{11} & \dots & y_{bu}^{1L} \\ \vdots & \ddots & \vdots \\ y_{bu}^{S1} & \dots & y_{bu}^{SL} \end{bmatrix}$, $b \in \mathcal{B}$, $u \in \mathcal{U}$. To facilitate subsequent analysis, we introduce three additional 4-dimensional tensors:D_s, E_u, and F_b. These tensors will serve as indicators for users, BSs, RBs and power levels, respectively. Similar to Y, these tensors are expressed in the form of a $BS \times UL$ block-partitioned matrix with $B \times U$ blocks, each with $S \times L$ entries. We define these tensors as follows: the tensor D_s can be expressed as $D_s = \begin{bmatrix} D_{s_{11}} & \cdots & D_{s_{1U}} \\ \vdots & \ddots & \vdots \\ D_{s_{B1}} & \cdots & D_{s_{BU}} \end{bmatrix}$, $s \in \mathcal{S}$, where, for all $i \in \mathcal{B}$ and $j \in \mathcal{U}$, $D_{s_{ij}} = \boldsymbol{i}_s \boldsymbol{1}_L^T$, where $\boldsymbol{i}_s$ is the s-th column of the $S \times S$ identity matrix $\boldsymbol{I}_S$ and $\boldsymbol{1}_L$ is the L-dimensional all-one vector. The tensor E_u can be expressed as

$$E_u = \begin{bmatrix} E_{u_{11}} & \cdots & E_{u_{1U}} \\ \vdots & \ddots & \vdots \\ E_{u_{B1}} & \cdots & E_{u_{BU}} \end{bmatrix}, \ u \in \mathcal{U}, \text{ where } E_{u_{ij}} = \begin{bmatrix} R_{iu}^{11} & \cdots & R_{iu}^{1L} \\ \vdots & \ddots & \vdots \\ R_{iu}^{S1} & \cdots & R_{iu}^{SL} \end{bmatrix} \text{ for } i \in \mathcal{B} \text{ and}$$

$j \in \mathcal{U}$, when $j = u$, and $E_{u_{ij}} = \mathbf{0}_{S \times L}$ when $j \neq u$. Finally, the tensor F_b can be

$$\text{written as } F_b = \begin{bmatrix} F_{b_{11}} & \cdots & F_{b_{1U}} \\ \vdots & \ddots & \vdots \\ F_{b_{B1}} & \cdots & F_{b_{BU}} \end{bmatrix}, \ b \in \mathcal{B}, \text{ where } F_{b_{ij}} = \begin{bmatrix} p_b^1 & \cdots & p_b^L \\ \vdots & \ddots & \vdots \\ p_b^1 & \cdots & p_b^L \end{bmatrix} \text{ for } i \in \mathcal{B} \text{ and}$$

$j \in \mathcal{U}$, when $i = b$, and $F_{b_{ij}} = \mathbf{0}_{S \times L}$ when $i \neq b$.

Using vec($\cdot$) to denote the operator that stacks the columns of a matrix on top of each other, we make the following definitions: $\mathbf{y} \triangleq \mathrm{vec}(Y^T)$, $\mathbf{d}_s \triangleq \mathrm{vec}(D_s^T)$, $\mathbf{e}_u \triangleq \mathrm{vec}(E_u^T)$ and $\mathbf{f}_b \triangleq \mathrm{vec}(F_b^T)$. For ease of exposition, in the forthcoming formulations we will drop the subscript indicating the dimension of the all-one and the all-zero vectors and matrices.

To cast the optimization problem (5.16) in a form amenable to SDR, we introduce the vector $\boldsymbol{\beta} \in \{-1, 1\}^{BUSL}$, $\boldsymbol{\beta} = 2\mathbf{y} - \mathbf{1}$, where $\boldsymbol{y} \in \{0, 1\}^{BUSL}$, i.e.,

$$\mathbf{y} = \frac{1}{2}(\boldsymbol{\beta} + \mathbf{1}). \tag{5.17}$$

Using (5.17), the formulation in (5.16) can be cast as

$$\max_{\mathbf{y}, \, t_u} \quad \rho \sum_{u \in \mathcal{U}} (1 - t_u) - (1 - \rho)\frac{\mathbf{1}^T}{2}(\boldsymbol{\beta} + \mathbf{1}), \tag{5.18a}$$

$$\text{subject to} \quad \frac{\mathbf{d}_s^T}{2}(\boldsymbol{\beta} + \mathbf{1}) \leq 1, \quad s \in \mathcal{S}, \tag{5.18b}$$

$$\frac{\mathbf{f}_b^T}{2}(\boldsymbol{\beta} + \mathbf{1}) \leq P_b^{\max}, \quad b \in \mathcal{B}, \tag{5.18c}$$

$$\frac{\mathbf{e}_u^T}{2}(\boldsymbol{\beta} + \mathbf{1}) \geq Q_u(1 - t_u), \quad u \in \mathcal{U}, \tag{5.18d}$$

$$t_u \geq e^{-\sigma \frac{\mathbf{e}_u^T(\boldsymbol{\beta} + \mathbf{1})}{2Q_u}}, \quad u \in \mathcal{U}, \tag{5.18e}$$

$$\boldsymbol{\beta} \in \{-1, 1\}^{BUSL}. \tag{5.18f}$$

We note that this problem is not convex because of the constraint in (5.18f). We also note that, for the solution of the relaxed problem in (5.18) to be optimal for the original problem in (5.15), the constraints in (5.18e) must be satisfied with equality

$\forall u \in \mathcal{U}$. We will now, show how the SDR technique can be used to provide an approximate solution for (5.18).

To use the SDR technique, the optimization variables in (5.18) are constrained to be in the cone of symmetric positive semidefinite (PSD) matrices [8]. Towards that end, we define the following vectors in $\mathbb{R}^{BUSL+1}$, $\hat{\mathbf{d}}_s \triangleq [\mathbf{d}_s^T \ \mathbf{d}_s^T \mathbf{1}]^T$, $s \in \mathcal{S}$, $\hat{\mathbf{f}}_b \triangleq [\mathbf{f}_b^T \ \mathbf{f}_b^T \mathbf{1}]^T$, $b \in \mathcal{B}$, $\hat{\mathbf{e}}_u \triangleq [\mathbf{e}_u^T \ \mathbf{e}_u^T \mathbf{1}]^T$, $u \in \mathcal{U}$, $\hat{\mathbf{1}} \triangleq [\mathbf{1}^T \ \mathbf{1}^T \mathbf{1}]^T$, $\hat{\boldsymbol{\beta}} \triangleq [\boldsymbol{\beta} \ 1]^T$ and $\hat{\mathbf{k}} \triangleq [\mathbf{0}^T \ 1]^T$, and the PSD matrices $\boldsymbol{\Phi} \in \mathbb{R}^{BUSL \times BUSL}$ and $\boldsymbol{\Psi} \in \mathbb{R}^{(BUSL+1) \times (BUSL+1)}$ to be $\boldsymbol{\Phi} \triangleq \boldsymbol{\beta}\boldsymbol{\beta}^T$ and $\boldsymbol{\Psi} = \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T$, i.e., $\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{\beta} \\ \boldsymbol{\beta}^T & 1 \end{bmatrix}$. Finally, we define the $(BUSL+1) \times (BUSL+1)$ matrices $\mathbf{A}_1 \triangleq \hat{\mathbf{k}}\hat{\mathbf{1}}^T$, $\mathbf{A}_{\hat{\mathbf{d}}_s} \triangleq \hat{\mathbf{k}}\hat{\mathbf{d}}_s^T$, $\mathbf{A}_{\hat{\mathbf{f}}_b} \triangleq \hat{\mathbf{k}}\hat{\mathbf{f}}_b^T$ and $\mathbf{A}_{\hat{\mathbf{e}}_u} \triangleq \hat{\mathbf{k}}\hat{\mathbf{e}}_u^T$. Using these definitions, it can be readily verified that (5.16) is equivalent to

$$\max_{\boldsymbol{\Psi}, \ t_u} \quad \rho \sum_{u \in \mathcal{U}} (1 - t_u) - \frac{(1-\rho)}{2} \operatorname{Tr}(\mathbf{A}_1 \boldsymbol{\Psi}), \tag{5.19a}$$

$$\text{subject to} \quad \frac{1}{2} \operatorname{Tr}(\mathbf{A}_{\hat{\mathbf{d}}_s} \boldsymbol{\Psi}) \leq 1, \quad s \in \mathcal{S}, \tag{5.19b}$$

$$\frac{1}{2} \operatorname{Tr}(\mathbf{A}_{\hat{\mathbf{f}}_b} \boldsymbol{\Psi}) \leq P_b^{\max}, \quad b \in \mathcal{B}, \tag{5.19c}$$

$$\frac{1}{2} \operatorname{Tr}(\mathbf{A}_{\hat{\mathbf{e}}_u} \boldsymbol{\Psi}) \geq Q_u(1 - t_u), \quad u \in \mathcal{U}, \tag{5.19d}$$

$$\operatorname{diag}(\boldsymbol{\Psi}) = 1, \tag{5.19e}$$

$$t_u \geq e^{-\frac{\sigma}{2Q_u} \operatorname{Tr}(\mathbf{A}_{\hat{\mathbf{e}}_u} \boldsymbol{\Psi})}, \quad u \in \mathcal{U}, \tag{5.19f}$$

$$\boldsymbol{\Psi} \succeq \mathbf{0}, \tag{5.19g}$$

$$\operatorname{rank}(\boldsymbol{\Psi}) = 1. \tag{5.19h}$$

Similar to the case in Section 5.4.1, it can be seen that a user $u \in \mathcal{U}$ is not accommodated in the system if and only if $\operatorname{Tr}(\mathbf{A}_{\hat{\mathbf{e}}_u} \boldsymbol{\Psi}) = 0$. In other words, a user $u \in \mathcal{U}$ is guaranteed to be served if and only if $\operatorname{Tr}(\mathbf{A}_{\hat{\mathbf{e}}_u} \boldsymbol{\Psi}) > 0$. A sufficient condition for this to happen is that $t_u = 0$, cf. (5.19f). In other words, the QoS demands of unaccommodated users need not be enforced.

The formulation in (5.19) is non-convex because of the rank-1 constraint (5.19h), even though the objective in (5.19a) and the constraints in (5.19b)–(5.19g) are convex in $\boldsymbol{\Psi}$ and $\{t_u\}$. Hence, apart from the rank-1 constraint, the formulation in (5.19) can be efficiently solved over PSD matrices, and it is this observation that forms the basis of the proposed SDR-based technique.

The philosophy of the SDR-based technique is to convert the deterministic optimization problem in (5.19) into a stochastic one by treating the PSD matrix $\mathbf{\Psi}$ generated by (5.19) as the covariance matrix of a multivariate Gaussian distribution in the prospective stochastic formulation [64]. The technique proceeds by drawing random samples from the aforementioned distribution. The samples are rounded to obtain candidate antipodal solutions and the feasible candidate that yields the largest objective is selected by the algorithm.

We conclude our presentation of the proposed SDR-based technique with a discussion on the choice of σ in (5.19f) and its impact on the performance. To begin with, we note that in this technique, candidate solutions are obtained by quantizing samples drawn from the multivariate Gaussian distribution. Now, if σ is large, users with normalized aggregate rate close to zero will be treated as if they were accommodated in the relaxed formulation in the (5.11c) and (5.11d) constraints in (5.16b). However, in the subsequent randomization step, such users are likely to be not accommodated. Hence, if σ is chosen to be too large, it would result in an overly restrictive formulation wherein users that are not accommodated in the network may enforce their QoS constraints in the relaxed formulation. Hence, it can be seen that the value of σ must be carefully chosen in order to ensure that, when the Gaussian sample vectors generated via the solution of (5.16), are quantized, the constraints are only enforced for those users who are actually accommodated in the network. Finding an appropriate value for σ analytically appears to be difficult. Hence, such a value will be obtained numerically in Section 5.6.

Computational Complexity Analysis

We first examine the complexity required for solving the relaxed version of (5.19), i.e., when the rank-1 constraint is dropped. The relaxation of (5.19) is solved only once, and the complexity of solving it, which is a PSD-constrained convex problem, is obtained using an approach similar to the one given in Section 5.5.1, and requires $\mathcal{O}\left((BUSL + 2U)^{3.5}\right)$. For the Gaussian randomization procedure that is used after solving the relaxation of (5.19), the complexity of generating and evaluating the objective function corresponding to the J random samples is $\mathcal{O}\left((BUSL)^2 J\right)$ [8]. Combining these observations, the complexity of the algorithm is $\mathcal{O}\left((BUSL + 2U)^{3.5} + (BUSL)^2 J\right)$. Note that the number of randomization samples does not affect the *grosso modo* complexity order [65]. This complexity is significantly

less than that of exhaustive search one. The latter is $\mathcal{O}(2^{BUSL})$.

An important observation is that the SDR-based technique relies on constructing PSD matrices with a number of entries that scales with the square of the number of variables. For the case without RB-reuse, this number is $BUSL$ and for the case with RB-reuse of $q = 1$ this number is $BUSL(1 + (B - 1)L)$. Hence, it can be seen that, despite being polynomial, the complexity of the SDR-based technique grows fast with the size of the network, rendering simulation of large networks a challenging task, especially for the case with RB-reuse. In other words, while our algorithm is fundamentally more efficient to implement than exhaustive search, implementing it for large networks can be computationally prohibitive, at least on non-special-purpose computers. Hence, the formulation proposed in our work is readily applicable in small-to-medium networks. For larger networks, the given formulation can be used to develop decentralized designs, which are likely to be more practical in those cases.

5.6   Numerical Results

We investigate the performance of the proposed joint design approaches, and provide numerical examples to illustrate the merits of these approaches for network instances with and without time-sharing and with and without RB-reuse[3]. We consider the 3GPP propagation model in [23] in a two-tier HetNet scenario with one macro BS with a maximum transmit power of 46 dBm, and several pico BSs with a maximum transmit power of 35 dBm. The location of the macro BS is assumed to be fixed, whereas the locations of the pico BSs and the users are assumed to be randomly distributed. The users have identical QoS demands, and the power levels used by the b-th BS are taken to be L equally spaced points in the interval $[0.05P_b^{\max},\ 0.5P_b^{\max}]$, unless otherwise stated. As per the 3GPP propagation model in [23], shadowing is assumed to have a log-normal distribution with a standard deviation of 8 dB for the macro BS, and 10 dB for the pico BSs. Denoting the distance between users and BSs by d (km), we express the path losses for the link between the macro BS and the users, and the link between the pico BSs and the users by $PL(d) = 128.1 + 37.6\log_{10}(d)$, and $PL(d) = 140.7 + 36.7\log_{10}(d)$, respectively. The bandwidth of each RB is assumed to be 180 kHz, and the noise power spectral density is assumed to be -174 dBm/Hz.

---

[3]Preliminary results on this problem were reported in [26] for a simplified variant of the joint design problem with neither time-sharing nor RB-reuse and with fixed RBs and power allocations.

For all simulations, results are averaged over 100 independent channel realizations, and the mathematical programs are solved using the CVX package [30] with the SeDuMi solver. The number of Gaussian samples used in the randomization phase of the SDR programs is set to be $J = 10^4$. For the case with time-sharing, the value of σ in (5.13) is chosen to be 100, whereas for the case without time-sharing, numerical investigations suggested that setting $\sigma \in [1.5, 2]$ in (5.16) tends to generate favourable results; in our simulations we set $\sigma = 1.8$.

Example 11. (*Tightness of Bounds—Time-Sharing Case*) In this example, we consider a time-sharing scenario without RB-reuse in a network with $B = 5$ BSs, one macro and four picos, $U = 16$ users, $S = 12$ RBs, and $L = 3$ power levels, yielding a total of 2880 variables.

Fig. 5.3(a) investigates the tightness of the bounds generated by the algorithms in Section 5.5.1. From this figure, it can be seen that the gap between the bounds is generally small, especially at low QoS values. For instance, for QoS values below 1 Mbps, the bounds coincide. For higher QoS values, the maximum gap between the bounds is about 0.9 user on average.

To complement the preceding discussion, in Fig. 5.3(b), we investigate the aggregate usage of the RBs corresponding to the aforementioned bounds. From this figure, it can be seen that, although the bounds on the number of users are tight, the difference in the corresponding RB usage can be relatively large, especially at high QoS demands. To understand this phenomenon, we note that the cost of accommodating an extra user increases with the QoS demand. Hence, at high QoS demands, the difference between the number of users generated by the upper and lower bounds corresponds to a large gap between the number of RBs required to realize these demands. This gap becomes smaller when the gap between the bounds on the number of users is small and disappears when the bounds coincide. For instance, for QoS demands less than 1 Mbps, the bounds on the number of users in Fig. 5.3(a) coincide and so do the corresponding RB usages in Fig. 5.3(b). In other words, the gap between the bounds on RB usage scales with QoS demands and the gap between the bounds on accommodated users. □

Example 12. (*Opportunistic RB-Reuse*) In this example, we provide and discuss the solution generated by the approach proposed in Section 5.5.1 for a particular realization of a network with time-sharing, 2 BSs (b_1 and b_2), 2 users (u_1, and u_2), 2 RBs (s_1 and s_2), and 2 power levels (ℓ_1 and ℓ_2). We construct two tables for the data

	(u_1,ℓ_1,n_1)	(u_1,ℓ_1,n_2)	(u_1,ℓ_2,n_1)	(u_1,ℓ_2,n_2)	(u_2,ℓ_1,n_1)	(u_2,ℓ_1,n_2)	(u_2,ℓ_2,n_1)	(u_2,ℓ_2,n_2)
(b_1,s_1,k_2)	4.6390	4.6280	5.0570	5.0459	0.0001	0.0001	0.0003	0.0001
(b_1,s_2,k_2)	5.1236	5.1235	5.5415	5.5415	0.0001	0.0000	0.0002	0.0001
(b_2,s_1,k_1)	0.0001	0.0000	0.0003	0.0000	2.6915	2.2788	3.1094	2.6967
(b_2,s_2,k_1)	0.0002	0.0000	0.0005	0.0000	2.0166	1.6037	2.4344	2.0212

Table 5.1: Rates in RB-reuse case

rate that can be communicated by each user-to-BS association on each RB: Table 5.1 pertains to the case with RB-reuse. In this table, k_i refers to the BS interfering with the transmission of BS b_j, and n_i refers to the power level used by BS k_i, $i,j \in \{1,2\}$, $i \neq j$. Table 5.2 pertains to the case with no RB-reuse.

	(u_1,ℓ_1)	(u_1,ℓ_2)	(u_2,ℓ_1)	(u_2,ℓ_2)
(b_1,s_1)	4.6418	5.0598	0.9595	1.3722
(b_1,s_2)	5.1236	5.5416	0.9891	1.4024
(b_2,s_1)	0.0028	0.0139	3.6510	4.0689
(b_2,s_2)	0.0000	0.0001	3.0056	3.4235

Table 5.2: Rates in no RB-reuse case

The rate values in these tables are in Mbps and obtained using the expression in (5.2) with b_1 and b_2 being macro and pico BSs, respectively. If the users' QoS demands are set to 6 Mbps, one can verify that two users cannot be accommodated when both RBs are reused or neither of them is reused over the entire signalling interval. In particular, the fact that no entries in Table 5.2 exceeds 6 Mbps directly implies that aggregate RB usage of each user should be greater than 1 to be accommodated. However, since the aggregate RB usage in the network cannot be greater than 2, the two users cannot be accommodated in the no RB-reuse case. In a similar fashion, Table 5.1 shows that the sum of any two entries corresponding to the u_2-th user cannot add up to 6 Mbps. Hence, user u_2 cannot be accommodated if RB-reuse is enforced on all RBs. Next, we will show that in this example, both users can be accommodated if the RBs are allowed to be reused opportunistically. For instance, these users can be accommodated if the first RB, s_1, is reused for 25% of time, and not reused for 75% of time, and the second RB, s_2, is reused over the entire signalling interval. In this case, $y_{b_1u_1}^{s_1\ell_1k_2n_2} = 0.25$, $y_{b_1u_1}^{s_2\ell_1k_2n_2} = 1$, $y_{b_2u_2}^{s_1\ell_2k_1n_1} = 0.25$, $y_{b_2u_2}^{s_2\ell_2k_1n_1} = 1$ and $y_{b_2u_2}^{s_1\ell_200} = 0.75$. Using these values yields a QoS of 6.2805 Mbps for user u_1 and

6.2634 Mbps for user u_2. The corresponding aggregate RB usage is 3.25.

To confirm the above observations, we perform the optimization proposed in Section 5.5.1. This optimization yielded the following solution: In the RB-reuse case, we have $y_{b_1u_1}^{s_1\ell_1k_2n_2} = 0.1770$, $y_{b_1u_1}^{s_2\ell_1k_2n_2} = 0.8630$, $y_{b_2u_2}^{s_1\ell_2k_1n_1} = 0.1770$, and $y_{b_2u_2}^{s_2\ell_2k_1n_1} = 0.8630$, and in the no RB-reuse case, we have $y_{b_1u_1}^{s_2\ell_200} = 0.1370$ and $y_{b_2u_2}^{s_1\ell_200} = 0.8230$. Since the proposed algorithm opportunistically switches between the RB-reuse case and the no RB-reuse one, this algorithm offers valuable design flexibility, and enables the two users to be accommodated in the network. The QoS of the users generated by the algorithm in this case is 6 Mbps per users and the aggregate RB usage is 3.04. Hence, it can be seen that the algorithm ensures that the users' QoS demands while minimizing aggregate RB usage. Fig. 5.4 provides a comparison between the algorithms with opportunistic RB-reuse and without RB-reuse for the rate values given in Table 5.1 and Table 5.2 for a range of QoS values. From this figure, it can be seen that the algorithm with opportunistic RB-reuse shows a superior performance for QoS values between 5 and 6 Mbps. □

Example 13. (*Tightness of Bounds—No Time-Sharing Case*) In this example, we compare the lower bound on the number of users generated by the algorithm in Section 5.5.2 with the optimal number of accommodated users obtained by exhaustive search for the case with no time-sharing and no RB-reuse. Since exhaustive search is computationally expensive for large networks, we consider a relatively small network with $B = 2$ BSs, one macro and one pico, $U = 3$ users, $S = 4$ RBs, and $L = 1$ power level of $\{0.25P_b^{\max}\}$.

The optimal number of users obtained by exhaustive search and the lower bound generated by the algorithm in Section 5.5.2 are depicted in Fig. 5.5(a). From this figure, it can be seen that the technique proposed in Section 5.5.2 yields a close-to-optimal user-to-BS association and RBs allocations especially at low QoS values. For instance, for QoS demands below 3 Mbps, the average number of users accommodated by the proposed technique coincides with that accommodated by exhaustive search. For QoS demands between 3 and 7 Mbps, the average number of users accommodated by the proposed technique is about 75% of that accommodated by exhaustive search. It is worth noting that the gap between the maximum number of accommodated users and the one generated by the SDR-based technique is at most 1 user. In fact, for low QoS the gap is zero and for high QoS numerical experiments indicated that this gap does not exceed 1 user, i.e., the smallest nonzero gap. Although we have not been

able to develop an analytical proof of the latter statement, numerical analysis confirms that the proposed SDR-based technique can yield close-to-optimal performance.

Analogous to the case in Example 11, we now investigate the aggregate usage of the RBs in the network without time-sharing. In Fig. 5.5(b), we compare the RB usage when both exhaustive search and the algorithm proposed in Section 5.5.2 are used. Similar to the case in Example 11, we note that, as the difference between the optimal number of users and the lower bound in Fig. 5.5(a) increases, the gap between the corresponding RB usages in Fig. 5.5(b) widens; a zero difference in the number of accommodated users corresponds to a zero RB usage gap. However, unlike Fig. 5.3(b), Fig. 5.5(b) exhibits a non-monotonic behaviour. This is because in this example time-sharing is not allowed. This implies that the association and allocation variables are discrete, which bars the users from occupying fractions of the signalling interval. Hence, for some QoS demands, there maybe extra RBs that, although not used, do not suffice to accommodate additional users. As the QoS demand increases, these extra RBs can be used by the existing users, which explains the rise in the RB usage curves observed at a QoS demand of 5 Mbps. As the QoS continues to increase, less users can be accommodated and the RB drops to zero.

As in Example 12, we now compare the optimal solution with the solution generated by the SDR-based technique for a network realization with 2 BSs (b_1 and b_2), 3 users (u_1, u_2 and u_3) and 4 RBs ($s_1, \ldots, s_4$). We use (5.2) to construct the table of the data rates (in Mbps) that can be communicated for every user-to-BS association on each RB, cf. the first table in (5.20).

If the users' QoS demands are set to 3 Mbps, one can readily verify that the 3 users can be accommodated with user u_1 associated with BS b_1 on any one of the four RBs, i.e., s_1, s_2, s_3, and s_4, user u_2 can be associated with BS b_2 on any one of the remaining three RBs, and user u_3 can be associated with BS b_1 on any pair of the remaining two RBs.

	u_1	u_2	u_3
(b_1, s_1)	3.0463	2.0456	2.2549
(b_1, s_2)	3.7359	1.5271	2.8976
(b_1, s_3)	3.0241	1.9794	2.2589
(b_1, s_4)	3.3721	1.2696	2.7857
(b_2, s_1)	0.1003	3.8978	0.0003
(b_2, s_2)	0.6106	4.7605	0.0001
(b_2, s_3)	0.0505	4.6056	0.0002
(b_2, s_4)	0.1971	4.4528	0.0001

	u_1	u_2	u_3
(b_1, s_1)	1	0	0
(b_1, s_2)	0	0	1
(b_1, s_3)	0	0	1
(b_1, s_4)	0	0	0
(b_2, s_1)	0	0	0
(b_2, s_2)	0	0	0
(b_2, s_3)	0	0	0
(b_2, s_4)	0	1	0

	u_1	u_2	u_3
(b_1, s_1)	1	0	0
(b_1, s_2)	0	0	1
(b_1, s_3)	0	0	0
(b_1, s_4)	0	0	1
(b_2, s_1)	0	0	0
(b_2, s_2)	0	0	0
(b_2, s_3)	0	1	0
(b_2, s_4)	0	0	0

$$(5.20)$$

A confirmation of the above observations is obtained by performing exhaustive search, which yielded the association table in the middle of (5.20). The corresponding matrix generated by the proposed SDR-based technique is given in the association table on the right of (5.20). Comparing the association tables, it can be seen that in both tables, the users are associated with the same BSs using the same number of RBs, albeit with different ones. In both tables, the number of RBs exhausted to accommodate the users is 4. □

Example 14. (*Impact of BS Transmit Powers—No Time-Sharing Case*) We investigate the performance of the SDR-based technique in a network with no time-sharing, no RB-reuse and with $B = 5$ BSs, one macro and two picos, $U = 6$ users, $S = 10$ RBs and $Q_u = 2$ Mbps, $\forall u$. We plot in Fig. 5.6 the number of accommodated users versus the transmit power of the macro BS, P_{macro}, assuming that $P_{\text{pico}} = P_{\text{macro}} - 11$ dBm.

From Fig. 5.6, it can be seen that increasing the number of power levels improves the network performance. For instance, for all transmit powers, the algorithm with $L = 3$ power levels shows a performance advantage of 15 dBm over the case with $L = 2$ power levels for an accommodated number of users of $U = 5$. □

Example 15. (*Impact of Power Levels—No Time-Sharing Case*) We investigate the impact of the number of power levels on the performance of the SDR-based technique. To do so, we consider a no time-sharing scenario with no RB-reuse in a network with $B = 2$ BSs, one macro and one pico, $U = 4$ users and $S = 6$ RBs. In Fig. 5.7, we plot the number of served users versus the number of power levels for various QoS demands.

From this figure, it can be seen that for low QoS demands, there is no perceptible change in the number of accommodated users with the increase of the number of power level. This is because, for low QoS demands, there are enough RBs to support all users. However, as the QoS demands increase, the increase in the number of power levels yields a large increase in the number of users accommodated in the network. For instance, for a QoS of 5 Mbps, the number of accommodated users doubles as the number of power levels increases from two levels to four. Our simulations suggest that this phenomenon continues to prevail for larger networks. □

Example 16. (*Computational Time*) We consider a no RB-reuse scenario with and without time-sharing in a network with $B = 3$ BSs, one macro and two picos, $U = 4$ users, $S = 6$ RBs, and $Q_u = 2$ Mbps, $\forall u$. In this setting, we vary the number of power levels from 1 to 9.

Fig. 5.8 shows the average computational time required by the algorithms proposed in Section 5.5.1 and Section 5.5.2 against the number of power levels. From this figure, it can be seen that both algorithms require a polynomial amount of computational time, and the algorithm proposed in Section 5.5.1 requires much lower computational time than the one proposed in Section 5.5.2. □

5.7   Conclusion

We considered the joint optimization of the user-to-BS associations, and the RBs and the power allocations in HetNets. We investigated this optimization for two novel network instances with opportunistic RB-reuse: one in which the RBs can be reused and the user-to-BS associations and the power allocations can be time-shared throughout the signalling interval, and one in which the RBs can be reused but the user-to-BS associations and the power allocations cannot be time-shared. Unfortunately, both design problems are non-convex, and therefore, difficult to solve. To deal with this difficulty, we provided a formulation, in which the discrete number of users is approximated by a continuous function. We then developed tight upper and lower bounds for the time-shared version of this formulation, and for the no time-shared one we developed a lower bound using the SDR with randomization technique. Numerical
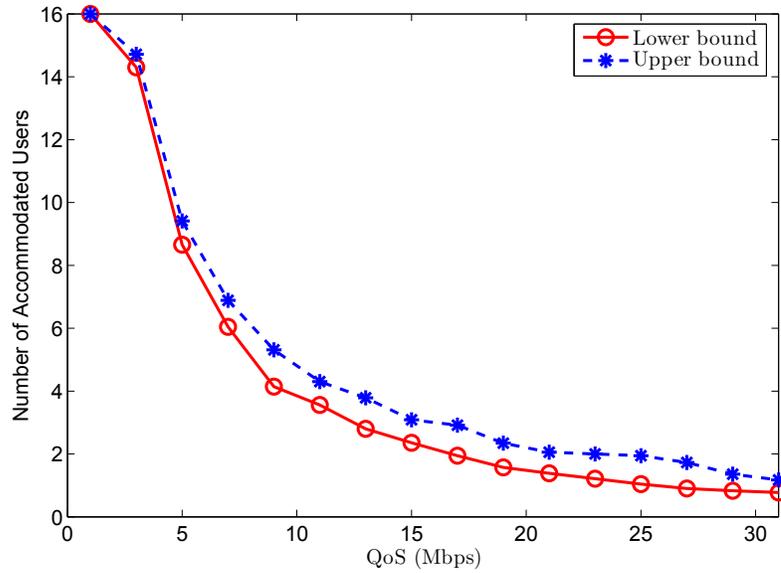
(a) Upper and lower bounds on the average number of users.
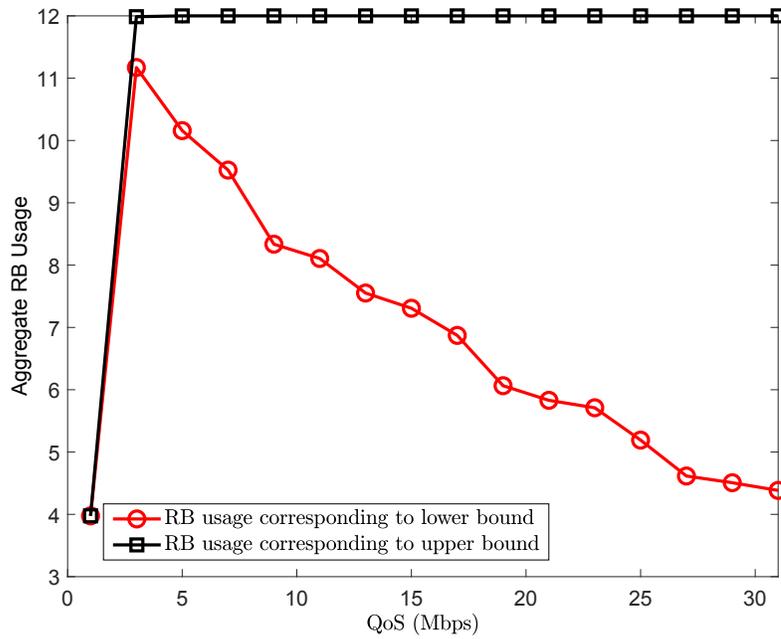
(b) Average aggregate RB usage corresponding to the upper and lower bounds.

Figure 5.3: Upper and lower bounds for the network with time-sharing and without RB-reuse.
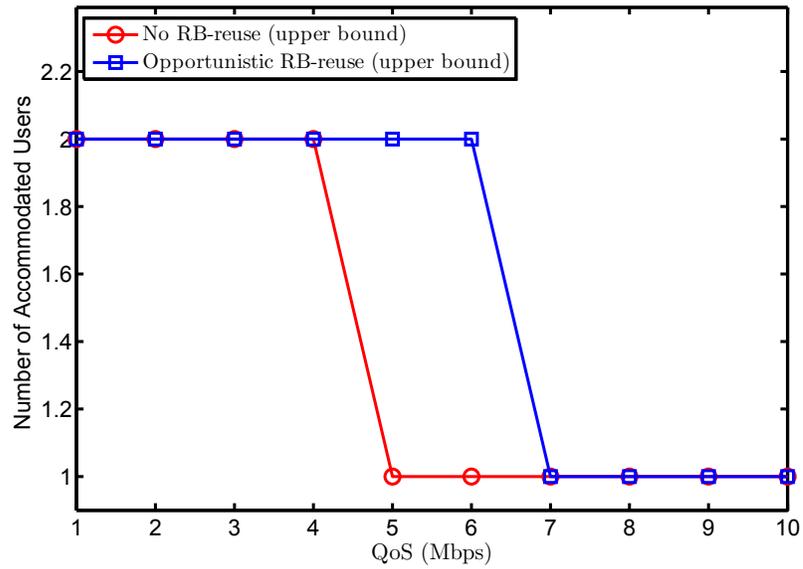
Figure 5.4: Opportunistic RB-reuse on the number of users.

results confirm that the developed bounds yield close-to-optimal performance.
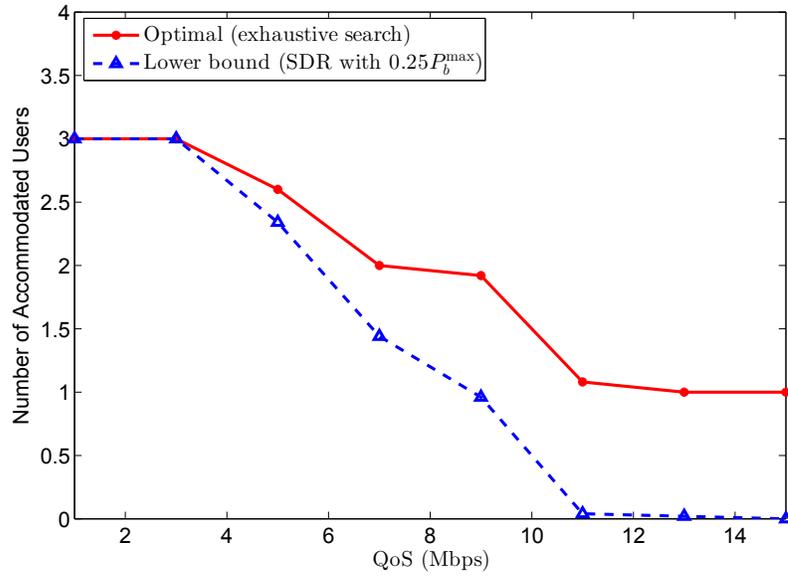
(a) Optimal and lower bound on the average number of users.
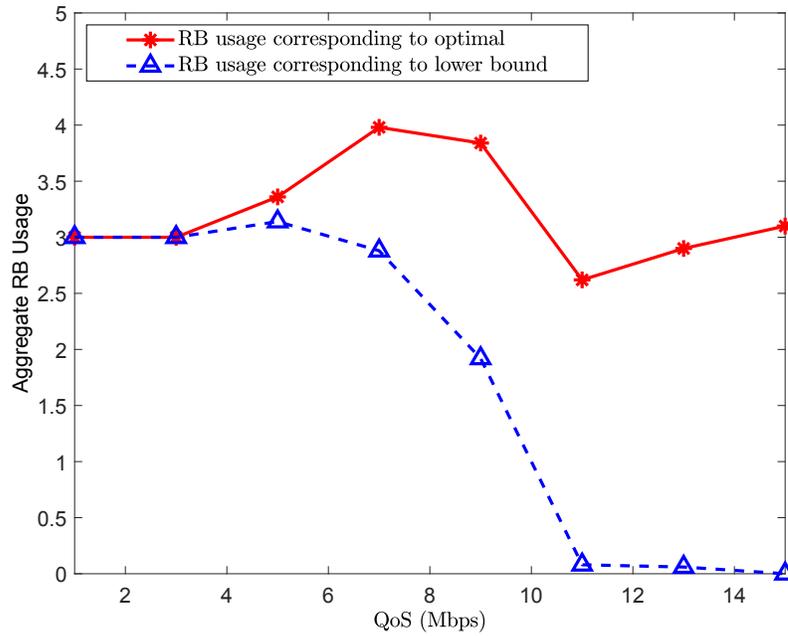
(b) Corresponding average RB usage

**Figure 5.5:** Optimal and lower bound for the network with no time-sharing and no RB-reuse.
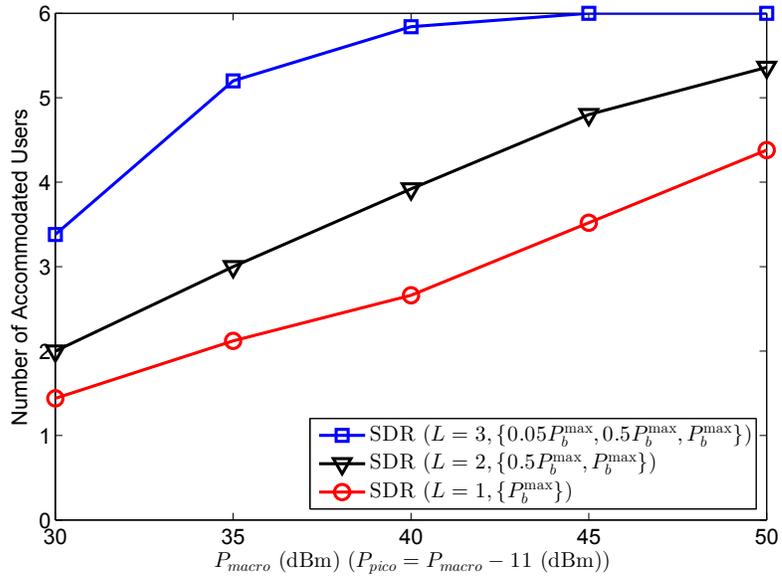
**Figure 5.6:** Impact of the BS transmit power on performance.
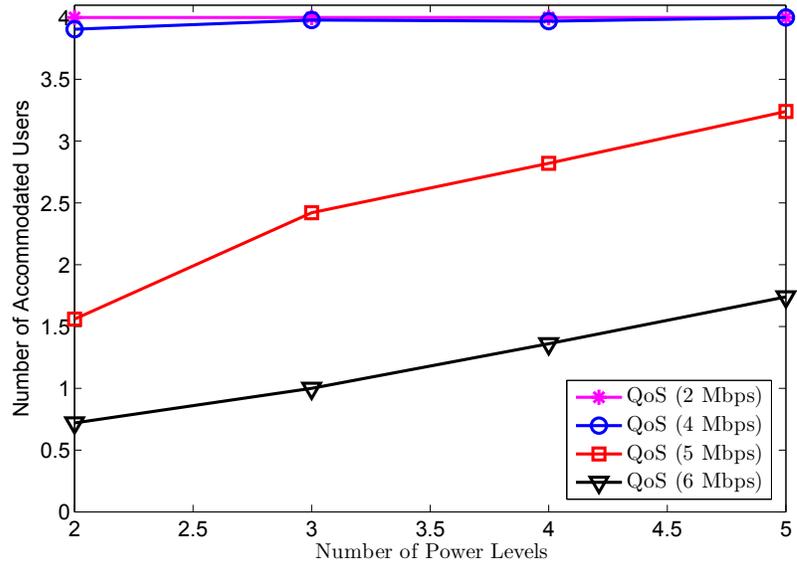
**Figure 5.7:** Impact of the number of power levels on performance.
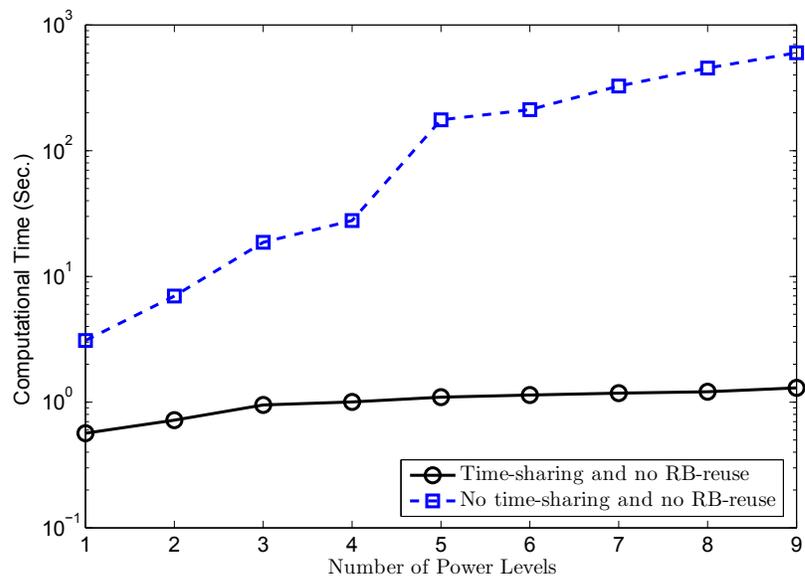
Figure 5.8: Computational time versus network size.

Chapter 6

Summary and Future Work

In this chapter, we first summarize the contribution of our work, and then, we discuss various directions for future work.

6.1 Summary

The following observations can be made from the work presented in this thesis.

6.1.1 Energy-Efficient Resource Allocation in Cellular Networks

- Resource allocation is an effective method to enhance EE of wireless networks. In Chapter 2, we considered the joint optimization of RBs and power levels allocations to maximize overall network EE in downlink cellular networks. Moreover, in Chapter 3, we focused on improving EE fairness among users in uplink cellular networks as users are equipped with a battery that has a limited power supply. In particular, we considered the joint optimization of RBs and power levels allocations to maximize minimum EE among users.

- In these works, the transmit power levels were constrained to assume discrete values. This constraint is implicit in many practical systems. Allowing these levels to be continuous simplifies the formulation, but cannot be readily extended to the case of discrete power levels.

- Due to the discreteness of the optimization programs, solving optimally is often computationally prohibitive. Hence, we adopted the SDR-based Gaussian

randomization approach to find close-to-optimal solutions.

- Through extensive simulations, we showed the effectiveness of the proposed techniques in achieving higher gains in overall EE, and EE fairness.

- In Chapter 2, we also observed that using few power levels suffices to achieve the EE performance of the continuous power allocation. Owing to such an observation, we developed a suboptimal heuristic to further decrease the computationally complexity.

6.1.2 User Association and Resource Allocation in Heterogeneous Networks

- In Chapter 4, we investigated SE maximization problem in HetNets. We adopt a multi-objective optimization approach wherein the number of users accommodated in the network is maximized and the RB usage is minimized. To combine these objectives, we weighted each of its components by an appropriate scalar. Moreover, to ensure that the number of users accommodated in the network is maximized, we showed the optimal choice of weights for each objectives. This composite objective is arguably more challenging than other traditional ones, such as sum-rate maximization, since it accounts for not only maximization of network SE, but also admission control.

- We proposed a load-aware user-to-BS association scheme that finds the close-to-optimal solution. This scheme employed a variant of the SDR-based Gaussian randomization approach used for solving the EE optimization problems. Numerical investigations suggested that the scheme proposed herein yields a number of accommodated users that is superior to the ones yielded by currently available schemes.

- Performing user-to-BS association in isolation might incur a performance loss. To realize the full benefits of the HetNet architecture, the interplay between various network functionalities must be taken into consideration. Hence, in Chapter 5, we proposed another scheme that jointly optimizes user-to-BS association, and the RB and the discrete power allocations to achieve the same objective considered in Chapter 4.

- In Chapter 5, we first proposed a novel function for obtaining the number of users accommodated in the network. In particular, to indicate whether a particular user is accommodated in the network or not, we defined a smooth indicator function, which can be readily incorporated in the optimization framework.

- We incorporated RB-reuse in the design and accounted for the resultant interference, which enriched the design, and enabled higher spectral efficiencies to be achieved.

- We investigated two instances of HetNets with opportunistic RB-reuse; with and without time-sharing. The case with time-sharing can be regarded as a generalization of the case without time-sharing, and hence offers the potential of superior performance.

- For the network instance with time-sharing, it was shown that the design problem can be approximated by a problem that yield tight upper and lower bounds on the objective.

- For the network instance with no-time-sharing, the design problem can be cast as a combinatorial problem. Hence, to obtain a close-to-optimal solution, we used the SDR-based Gaussian randomization with the polynomial complexity.

- Lastly, in Chapter 5, we considered the possibility that each user can be associated with more than one BS, i.e., multi-connectivity (also known as multi-homing). This feature provided additional degrees of design freedom, which would not be available if each user is restricted to be served solely by one BS as in Chapter 4. The design problem in the latter case is simplified by the fact that the number of users in that case is equal to the number of connections. Hence, the framework in Chapter 5 can be regarded as a generalization of the framework in Chapter 4.

6.2 Future Work

There are various directions in which our work can be readily extended.

- In order to exploit the full potential of the proposed algorithms, there are inherent limitations of the underlying technique that have to be overcome. These

include high complexity, centrality, and limited scope. To address these limitations, the following directions can be pursued in the future:

– *Prohibitive computational complexity*: Despite its efficacy in small-to-medium networks, the proposed algorithms do not scale well for larger networks. This is due to excessive communication and computation overheads. Hence, for such networks, it is necessary to develop a scalable algorithm with lower computational complexity and acceptable accuracy. To that end, an interesting research problem is to develop an efficient and low-complexity algorithm that is akin to the algorithms in big data.

– *Distributed approach*: For larger networks, an effective solution can be obtained using a distributed approach in which a large scale problem is decomposed into smaller subproblems, and the subproblems are solved separately. Then, the solution of subproblems is coordinated to achieve globally optimal solution. That type of algorithms is practically desirable. This is because they can provide significant gains in the computational efficiency and reductions in the communication overhead for lager networks. Hence, the future research plans can put emphasis on distribute solutions.

– *Joint backhaul complexity*: As the size of the network grows, the backhaul capacity of the BSs becomes the bottleneck of the system. Previously, we assumed that all BSs have backhaul connections that can carry all the traffic generated by the users. However, in practice, all BSs, including macro, pico and femto BSs, have backhaul connections with limited capacities. Due to that fact, an interesting research problem is to incorporate the consideration of backhaul constraints into the optimization of the network functionalities.

- An important extension of our work is to apply a variant of the proposed optimization approach to other advanced network architectures, such as multiple radio access technology (Multi-RAT) heterogeneous networks. The architecture of Multi-RAT HetNet enables the utilization of both licensed and unlicensed spectrum to support inter-RAT multi-connectivity and to be flexible for adaptation for new user cases. Despite its advantages, the complex multi-layer/multi-band/multi-mode structure (i.e., including different access protocols, frequency bands, BSs with different cell sizes, and association modes, i.e., single/multiple

RAT association) of Multi-RAT HetNet renders radio resource allocation a serious challenge. An interesting research problem is to develop a joint design of user association, and RB and discrete power allocation for maximizing overall EE in downlink multi-RAT networks.

List of References

[1] "Cisco visual networking index: Global mobile data traffic forecast update, 2015–2020," White Paper, Cisco Systems Inc., Feb. 2016.

[2] E. Hossain and M. Hasan, "5G cellular: Key enabling technologies and research challenges," *IEEE Instrum. Meas. Mag.*, vol. 18, no. 3, pp. 11–21, Jun. 2015.

[3] A. Damnjanovic, J. Montojo, J. Cho, H. Ji, J. Yang, and P. Zong, "UE's role in LTE-Advanced heterogeneous networks," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 164–176, Feb. 2012.

[4] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Commun. Surv. Tuts.*, vol. 15, no. 1, pp. 167–178, Feb. 2013.

[5] A. Farrokhi and O. Ercetin, "QoE based random sleep-awake scheduling in heterogeneous cellular networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 1–6, Apr. 2016.

[6] A. M. Ibrahim, O. Ercetin, and T. ElBatt, "Stability analysis of slotted aloha with opportunistic RF energy harvesting," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1477–1490, May 2016.

[7] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.

[8] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[9] "Sustainable energy use in mobile communications," White Paper, Ericsson Inc., Aug. 2007.

[10] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy efficiency in the future internet: A survey of existing approaches and trends in energy-aware fixed network infrastructures," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 2, pp. 223–244, Jul. 2011.

[11] C. Han, T. Harrold, S. Armour, I. Krikidis, S. Videv, P. M. Grant, H. Haas, J. S. Thompson, I. Ku, C.-X. Wang, T. A. Le, M. R. Nakhai, J. Zhang, and L. Hanzo, "Green radio: Radio techniques to enable energy-efficient wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 46–54, May 2011.

[12] K. Fazel and S. Kaiser, *Multi-Carrier and Spread Spectrum Systems: From OFDM and MC-CDMA to LTE and WiMAX*. John Wiley & Sons Inc., 2008.

[13] W. Wu, Q. Yang, P. Gong, and K. S. Kwak, "Energy-efficient resource optimization for OFDMA-based multi-homing heterogenous wireless networks," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 5901–5913, Nov. 2016.

[14] E. Bedeer, A. Alorainy, M. J. Hossain, O. Amin, and M.-S. Alouini, "Fairness-aware energy-efficient resource allocation for AF co-operative OFDMA networks," *IEEE J. Select. Areas Commun.*, vol. 33, no. 12, pp. 2478–2493, Dec. 2015.

[15] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1037–1050, Apr. 2016.

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

[17] M. Hajiaghayi, M. Dong, and B. Liang, "Jointly optimal channel and power assignment for dual-hop multi-channel multi-user relaying," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 9, pp. 1806–1814, Oct. 2012.

[18] H. Zhang, L. Venturino, N. Prasad, P. Li, S. Rangarajan, and X. Wang, "Weighted sum-rate maximization in multi-cell networks via coordinated scheduling and discrete power control," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1214–1224, Jun. 2011.

[19] L. Lei, V. Angelakis, and D. Yuan, "Performance analysis of chunk-based resource allocation in wireless OFDMA systems," in *Proc. Int. Wkshp. on Computer Aided Modeling and Design of Commun. Links and Ntwk. (CAMAD)*, pp. 90–94, Sept. 2012.

[20] G. Bacci, L. Sanguinetti, M. Luise, and H. V. Poor, "Energy-efficient contention-based synchronization in OFDMA systems with discrete powers and limited feedback," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 4197–4202, Apr. 2013.

[21] H.-H. Nguyen and W.-J. Hwang, "Distributed scheduling and discrete power control for energy efficiency in multi-cell networks," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2198–2201, Dec. 2015.

[22] X. Xiao, X. Tao, and J. Lu, "QoS-aware energy-efficient radio resource scheduling in multi-user OFDMA systems," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 75–78, Jan. 2013.

[23] 3rd Generation Partnership Project (3GPP) TR 36.839, "Evolve universal terrestrial radio access (E-UTRA); Mobility enhancements in heterogeneous networks," v11.1.0, Jan. 2013.

[24] Z. Ren, S. Chen, B. Hu, and W. Ma, "Energy-efficient resource allocation in downlink OFDM wireless systems with proportional rate constraints," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2139–2150, Jun. 2014.

[25] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge Univ. Press, 2004.

[26] H. Sokun, R. Gohary, and H. Yanikomeroglu, "QoS-guaranteed user association in HetNets via semidefinite relaxation," in *Proc. IEEE Veh. Tech. Conf. (VTC-Fall)*, pp. 1–5, Sept. 2015.

[27] E. Bedeer and M. J. Hossain, "Performance of low-complexity uniform power loading OFDM systems with reduced feedback over rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3783–3795, Jun. 2016.

[28] S. Nagaraj, "An extension to the ordered subcarrier selection algorithm (ossa)," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1159–1163, MAr. 2009.

[29] M. Deruyck, E. Tanghe, W. Joseph, W. Vereecken, M. Pickavet, L. Martens, and B. Dhoedt, "Model for power consumption of wireless access networks," *IET Science, Measurement and Technology*, vol. 5, no. 4, pp. 155–161, Aug. 2011.

[30] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Mar. 2014. Available at:http://cvxr.com/cvx.

[31] H. Sokun, R. H. Gohary, and H. Yanikomeroglu, "A novel approach for QoS-aware joint user association, resource block and discrete power allocation in HetNets," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7603–7618, Nov. 2017.

[32] Y. Li, M. Sheng, X. Wang, Y. Zhang, and J. Wen, "Max–min energy-efficient power allocation in interference-limited wireless networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4321–4326, Sep. 2015.

[33] C. Guo, B. Liao, L. Huang, X. Lin, and J. Zhang, "On convexity of fairness-aware energy-efficient power allocation in spectrum-sharing networks," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 534–537, Mar. 2016.

[34] Y. Li, M. Sheng, C. W. Tan, Y. Zhang, Y. Sun, X. Wang, Y. Shi, and J. Li, "Energy-efficient subcarrier assignment and power allocation in OFDMA systems with max-min fairness guarantees," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3183–3195, Sep. 2015.

[35] S. Sesia, I. Toufik, and M. Baker, *LTE – The UMTS Long Term Evolution.* John Wiley & Sons, 2009.

[36] H. Sokun, E. Bedeer, R. Gohary, and H. Yanikomeroglu, "Optimization of discrete power and resource block allocation for achieving maximum energy efficiency in OFDMA networks," *IEEE Access*, vol. 5, pp. 8648–8658, May 2017.

[37] R. Jain, D.-M. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," Tech. Rep. DEC-TR-301, Digital Equipment Corporation, 1984.

[38] C. Foster, I. Dickie, G. Maile, H. Simth, and M. Crisp, "Understanding the environmental impact of communication systems," *Ofcom Final Report*, Dec. 2009.

[39] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1479–1489, Dec. 2010.

[40] I. Guvenc, M.-R. Jeong, I. Demirdogen, B. Kecicioglu, and F. Watanabe, "Range expansion and inter-cell interference coordination (ICIC) for picocell networks," in *Proc. IEEE Veh. Tech. Conf. (VTC-Fall)*, pp. 1–6, Sept. 2011.

[41] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, Jun. 2013.

[42] K. Shen and W. Yu, "Distributed pricing-based user association for downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1100–1113, Jun. 2014.

[43] E. Matskani, N. D. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "Convex approximation techniques for joint multiuser downlink beamforming and admission control," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2682–2693, Jul. 2008.

[44] D. Morales and H. Romeijn, "The generalized assignment problem and extensions," in *Handbook of Combinatorial Optimization*, pp. 259–311, Springer US, 2005.

[45] S. Corroy and R. Mathar, "Semidefinite relaxation and randomization for dynamic cell association in heterogeneous networks," in *Proc. IEEE Glob. Commun. Conf. (Globecom)*, pp. 2373–2378, Dec. 2012.

[46] "On cell range expansion—R1-105294," in *The Mobile Broadband Standard—3GPP Meeting R1-62b*, Texas Instruments, Oct. 2010. Available at: `http://www.3gpp.org/DynaReport/TDocExMtg--R1-62b--28041.htm`.

[47] H. Boostanimehr and V. K. Bhargava, "Unified and distributed QoS-driven cell association algorithms in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1650–1662, Mar. 2015.

[48] B. Zhuang, D. Guo, and M. L. Honig, "Energy-efficient cell activation, user association, and spectrum allocation in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 823–831, Apr. 2016.

[49] G. Hegde, O. D. Ramos-Cantor, Y. Cheng, and M. Pesavento, "Optimal resource block allocation and muting in heterogeneous networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 3581–3585, Mar. 2016.

[50] R. Rashtchi, R. H. Gohary, and H. Yanikomeroglu, "Generalized cross-layer designs for generic half-duplex multicarrier wireless networks with frequency-reuse," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 458–471, Jan. 2016.

[51] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "Optimal user-cell association for massive MIMO wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1835–1850, May 2016.

[52] F. Boccardi, J. Andrews, H. Elshaer, M. Dohler, S. Parkvall, P. Popovski, and S. Singh, "Why to decouple the uplink and downlink in cellular networks and how to do it," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 110–117, Mar. 2016.

[53] A. T. Gamage, H. Liang, and X. Shen, "Two time-scale cross-layer scheduling for cellular/WLAN interworking," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2773–2789, Aug. 2014.

[54] K. Son, S. Chong, and G. Veciana, "Dynamic association for load balancing and interference avoidance in multi-cell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3566–3576, Jul. 2009.

[55] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "RAT selection games in HetNets," in *Proc. IEEE Int. Conf. Comp. Commun. (INFOCOM)*, pp. 998–1006, Apr. 2013.

[56] M. Hong and A. Garcia, "Mechanism design for base station association and resource allocation in downlink OFDMA network," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2238–2250, Dec. 2012.

[57] L. P. Qian, Y. J. Zhang, Y. Wu, and J. Chen, "Joint base station association and power control via Benders' decomposition," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1651–1665, Apr. 2013.

[58] R. Sun, M. Hong, and Z.-Q. Luo, "Joint downlink base station association and power control for max-min fairness: Computation and complexity," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1040–1054, Jun. 2015.

[59] V. N. Ha and L. B. Le, "Distributed base station association and power control for heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 282–296, Jan. 2014.

[60] Q. Kuang, J. Speidel, and H. Droste, "Joint base-station association, channel assignment, beamforming and power control in heterogeneous networks," in *Proc. IEEE Veh. Tech. Conf. (VTC-Spring)*, pp. 1–5, May 2012.

[61] C. S. Chen, F. Baccelli, and L. Roullet, "Joint optimization of radio resources in small and macro cell networks," in *Proc. IEEE Veh. Tech. Conf. (VTC-Spring)*, pp. 1–5, Mar. 2011.

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

[63] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optim. Meth. and Softw.*, vol. 11, no. 1-4, pp. 625–653, Aug. 1999.

[64] A. Wiesel, Y. C. Eldar, and S. Shamai, "Semidefinite relaxation for detection of 16-QAM signaling in MIMO channels," *IEEE Signal Process. Lett.*, vol. 12, no. 9, pp. 653–656, Sep. 2005.

[65] N. D. Sidiropoulos and Z.-Q. Luo, "A semidefinite relaxation approach to MIMO detection for high-order QAM constellations," *IEEE Signal Process. Lett.*, vol. 13, no. 9, pp. 525–528, Sep. 2006.

Appendix A

Lines of Research—Not Included in the Thesis

In the research work, we have touched upon a number of topics. Particularly, we have worked on three fundamentally different problems, besides the one investigated herein. The first pertains to designing spectrally efficient transmission schemes for cooperative communications. The second pertains to examining the signal space diversity in two-way cooperative systems. The third pertains to investigating the co-existence performance of multiple protocols in the same Internet of Things bands in C++ based NS-3 network simulator. We will now give more details about each of those research projects.

A.1 Spectrally-Efficient Transmission Scheme for Cooperative Communication

Despite the rich literature on cooperative networks, employment of different modulation levels by the source and relay terminals has not been investigated thoroughly from the physical layer perspective. We first investigate the bit error rate (BER) performance of selective relaying in a multi-relay decode-and-forward (DF) cooperative network where the source and the relays transmit using different modulation levels. Specifically, we derive a closed form expression for the end-to-end (E2E) (uncoded) BER. To draw further insights on the BER performance, we also provide a simpler approximate BER expression that is accurate in the high signal-to-noise ratio.

Next, we focus on the systematic optimization of power allocation in this relaying scheme with imperfect channel estimation. We obtain an asymptotic expression

for the error probability in the presence of Gaussian imperfect channel estimations. By using this expression, we derive the power allocation scheme that minimizes the asymptotic error probability. We show that the power allocation problem can suitably be presented as a geometric programming problem that is solved by an efficient convex programming technique.

Lastly, it is known that multi-relay cooperative relaying enables spatial diversity often at the expense of spectral efficiency (SE). To alleviate the loss in SE due to half-duplex relaying and transmission over orthogonal channels, we propose a novel transmission scheme. Particularly, we obtain a closed-form expression for both E2E SE in such scheme. Subsequently, using this closed-form expression and average channel statistics, we perform joint optimization of power allocation and modulation level selection to maximize the E2E SE while maintaining a target E2E average error probability, and a set of transmit power constraints. Simulation results demonstrate that the transmission scheme proposed improves the E2E SE significantly in comparison with the conventional adaptive DF transmission scheme.

Our findings in this line of research are published in the following journal papers and conference paper.

[**J3**] H. U. Sokun and H. Yanikomeroglu, **"On the spectral efficiency of selective DF relaying with different modulation levels,"** *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4500-4506, May 2017.

[**J4**] H. U. Sokun, A. Bin Sediq, H. Yanikomeroglu, and S. Ikki, **"Power allocation optimization in selective DF relaying with different modulation levels in the presence of imperfect channel estimations,"** *IEEE Commun. Lett.*, vol. 19, no. 5, pp. 867-870, May 2015.

[**C3**] H. U. Sokun, A. Bin Sediq, S. Ikki, and H. Yanikomeroglu, **"Selective DF relaying in multi-relay networks with different modulation levels,"** in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 5035-5041, June 2014.

A.2 Signal Space Diversity in Two-Way Cooperative Systems

This work proposes a new transmission scheme for multi-relay two-way relaying systems, where two end-sources communicate with each other via a set of relays. In

the proposed scheme, we consider signal space diversity and time division broadcast protocol jointly to increase SE without sacrificing the system reliability. The idea behind this consideration is that original data symbols are rotated before transmission, and then the in-phase and quadrature components of these rotated symbols are transmitted through the cooperation of the end-sources and the relays. By doing so, the end-sources exchange four symbols over three time slots instead of two, i.e., doubling the number of transmitted symbols. In addition, relay selection is incorporated into this scheme to achieve a further increase in SE. For relay selection, two different strategies are discussed: reactive and proactive relay selections. These strategies differ depending on whether relay selection is performed after or before the start of transmission. Specifically, for each relay-selection strategy, we first obtain a closed-form expression for the E2E error probability with an arbitrary constellation, which accounts for all the resulting non-uniform constellation cases due to constellation rotation. Subsequently, with the derived expressions, we then formulate an optimization problem that considers the joint optimization of the rotation angle and the transmit powers at the end-sources and the relays. The objective of the optimization problem is to minimize the E2E error probability of one of the end-sources, while satisfying a set of total and individual transmit power constraints, and a predefined threshold for the E2E error probability of the other end-source. Numerical results verify the theoretical analysis, and show that the scheme proposed herein provides not only higher SE, but also more reliable transmission.

Our findings in this line of research are published in the following journal paper and conference paper.

[**J5**] H. U. Sokun, M. C. Ilter, S. Ikki, and H. Yanikomeroglu, **"A spectrally efficient signal space diversity-based two-way relaying system,"** *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6215-6230, July 2017.

[**C4**] H. U. Sokun, M. C. Ilter, S. Ikki, and H. Yanikomeroglu, **"A signal space diversity based time division broadcast protocol in two-way relay systems,"** in *Proc. IEEE Glob. Commun. Conf. (Globecom)*, pp. 1-6, Dec. 2015.

A.3 On the Coexistence of IEEE 802.11 and IEEE 802.15.4: Simulation Results Using NS-3 Network Simulator

We study coexistence between IEEE 802.11 and IEEE 802.15.4 protocols in 2.4 GHz band. We utilize NS-3 for investigating the effect of co-channel interference on the performance of the two standards. Although NS-3 was not built for such coexistence studies, we have managed to modify some modules in NS-3 to enable devices using the two standards to share the same channel and visualize the interaction between them. We have also developed a probabilistic indoor path loss model based on collected data during measurement campaign and integrated the model into the NS-3 library.

We have utilized several simulation scenarios to investigate the effect of interference between IEEE 802.11 and IEEE 802.15.4 protocols on the packet success rate and throughput. Our results indicate that unless proper coordination is performed between two devices, simultaneous transmission of IEEE 802.11 and IEEE 802.15.4 devices can degrade the performance of IEEE 802.15.4 devices due to the relatively high transmission power of IEEE 802.11 devices. The amount of degradation depends on density of IEEE 802.15.4 devices and their packet transmission rate as well. Particularly, if the total transmission rate of IEEE 802.15.4 devices is smaller than 500 packet/sec, there is no significant degradation in system performance. Since transmission rate of IEEE 802.15.4 devices reported during experimental measurements was around 0.03 packet/sec, we conjecture that over 16000 IEEE 802.15.4 devices employing IEEE 802.15.4 can be accommodated on the same transmission channel. The presence of interference from IEEE 802.11 transmission leads to decreasing the number of IEEE 802.15.4 devices that can operate simultaneously. The effect of IEEE 802.11 on IEEE 802.15.4 transmissions depends on data rate of IEEE 802.11 stations and the number of stations accessing network simultaneously.

Our findings in this line of research are published in the following conference paper.

[**C5**] A. El-Keyi, H. Sokun, T. Nguyen, Q. Ye, H. Zhu, and H. Yanikomeroglu, **"A novel probabilistic path loss model for simulating coexistence between 802.11 and 802.15.4 networks in smart home environments,"** in *Proc. IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun. (PIMRC)*, pp. 1-6, Sept. 2017.

Appendix B

Proof of Proposition 1

The proof of this proposition uses a contradiction argument analogous to the one in [43] and [57] but for a different design objective, network functionalities, and network instances.

Let $\{\tilde{y}_{bu}^{s\ell kn}\}$ be the optimal solution for the objective in Section 5.3.2 and $\forall b, u, s, \ell, k, n$, let $\{\hat{y}_{bu}^{s\ell kn}\}$, be a feasible solution which accommodates one user more than the optimal solution, i.e., $f(\hat{\mathbf{y}}) = f(\tilde{\mathbf{y}}) + 1$, $f(\hat{\mathbf{y}}) = U - \sum_{u \in \mathcal{U}} e^{-(\sigma/Q_u) \sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(\hat{y}_{bu}^{s\ell 00} R_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} \hat{y}_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right)}$, and $f(\tilde{\mathbf{y}}) = U - \sum_{u \in \mathcal{U}} e^{-(\sigma/Q_u) \sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(\tilde{y}_{bu}^{s\ell 00} R_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} \tilde{y}_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right)}$. Then, the objective function of the feasible solution can be expressed as

$$\rho f(\hat{\mathbf{y}}) - (1-\rho)g(\hat{\mathbf{y}}) \geq \rho f(\tilde{\mathbf{y}}) + \rho - 2(1-\rho)S \geq \rho f(\tilde{\mathbf{y}}) \geq \rho f(\tilde{\mathbf{y}}) - (1-\rho)g(\tilde{\mathbf{y}}), \text{ (B.1)}$$

where the aggregate usage of RBs in the network for the feasible and optimal solutions are respectively given by $g(\hat{\mathbf{y}}) = \sum_{u \in \mathcal{U}} \sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(\hat{y}_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} \hat{y}_{bu}^{s\ell kn} \right)$, and $g(\tilde{\mathbf{y}}) = \sum_{u \in \mathcal{U}} \sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(\tilde{y}_{bu}^{s\ell 00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} \tilde{y}_{bu}^{s\ell kn} \right)$. Now, the total usage of RBs in the system can be at most $2S$. Supposing that $\rho - 2(1-\rho)S > 0$, i.e., $\rho > \frac{2S}{1+2S}$, it can be seen that the value of the objective function of the feasible solution is greater than the value of the objective at the optimal solution, that is, $\rho f(\hat{\mathbf{y}}) - (1-\rho)g(\hat{\mathbf{y}}) \geq \rho f(\tilde{\mathbf{y}}) - (1-\rho)g(\tilde{\mathbf{y}})$. This contradicts the optimality of $\{\tilde{y}_{bu}^{s\ell kn}\}$, and hence, when $\rho \in \left(\frac{2S}{1+2S}, 1 \right)$, the composite objective ensures that the maximum number of users is accommodated with the least number of RBs.

Appendix C

Proof of Proposition 2

A function $f : \mathbb{R}^n \to \mathbb{R}$ is self-concordant if it satisfies $|\frac{\partial^3 f(x+kv)}{\partial k^3}| \leq 2\frac{\partial^2}{\partial k^2} f(x + kv)^{3/2}$, for all $x \in \mathbf{dom}\ f$ and for all v [25]. Now, the log-barrier function for (5.13) is given by

$$\Omega = k\left(\rho \sum_{u\in\mathcal{U}}(1 - t_u) - (1 - \rho)\sum_{s\in\mathcal{S}}\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn}\right)\right) + \phi, \quad \text{(C.1)}$$

where ϕ is the logarithmic barrier defined as

$$
\begin{aligned}
\phi = &-\sum_{b\in\mathcal{B}}\log\left(P_b^{\max} - \sum_{u\in\mathcal{U}}\sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn}\right)p_b^\ell\right)\\
&-\sum_{u\in\mathcal{U}}\log\left(\sum_{b\in\mathcal{B}}\sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00}R_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn}R_{bu}^{s\ell kn}\right) - Q_u(1 - t_u)\right)\\
&-\sum_{s\in\mathcal{S}}\log\left(1 - \sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}}\sum_{\ell\in\mathcal{P}_b}y_{bu}^{s\ell00} - \frac{1}{2}\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}}\sum_{\ell\in\mathcal{P}_k}\sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}} y_{bu}^{s\ell kn}\right)\\
&-\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}}\sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{P}_b}\left(\sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k}\log\left(\sum_{\tilde{u}\in\mathcal{U}\backslash u} y_{k\tilde{u}}^{snb\ell} - y_{bu}^{s\ell kn}\right)\right.\\
&\left.+ \log\left(1 - y_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn}\right)\right)\\
&-\sum_{u\in\mathcal{U}}\sum_{s\in\mathcal{S}}\log\left(1 - \sum_{b\in\mathcal{B}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00} - \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn}\right)\right)\\
&-\sum_{b\in\mathcal{B}}\sum_{s\in\mathcal{S}}\log\left(1 - \sum_{\ell\in\mathcal{P}_b}\sum_{u\in\mathcal{U}}\left(y_{bu}^{s\ell00} - \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn}\right)\right)\\
&-\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}}\sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{P}_b}\sum_{k\in\{0\}\cup\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k}\log\left(y_{bu}^{s\ell kn}\right) - \sum_{u\in\mathcal{U}}\log\left(\log(t_u)\right.\\
&+ \frac{\sigma}{Q_u}\sum_{b\in\mathcal{B}}\sum_{s\in\mathcal{S}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell00}R_{bu}^{s\ell00} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn}R_{bu}^{s\ell kn}\right)\right) - \log(t_u).
\end{aligned}
\tag{C.2}
$$

Self-concordance of the first eight terms can be readily verified. For the remaining two terms, we utilize a known self-concordant form, $-\log(\log(a) + b) - \log(a)$ [25], where $a = \log(t_u)$, and $b = \frac{\sigma}{Q_u} \sum_{b \in \mathcal{B}} \sum_{s \in \mathcal{S}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell00} R_{bu}^{s\ell00} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn} R_{bu}^{s\ell kn} \right)$, which shows that Ω is self-concordant. Using self-concordance of the log-barrier function, the number of Newton steps required to solve (5.13) can be shown to be proportional to $\sqrt{n}$, where n is the number of inequality constraints. Since each Newton step requires cubic complexity, the complexity of solving (5.13) is proportional to $n^{3.5}$, which establishes the claim of the proposition.

Appendix D

Formulation for Opportunistic RB-Reuse 3—Time Sharing Case

For the case in which each RB is allowed to be reused at most three times in the network, we revise the optimization variables and the constraints. In particular, we will use the octet $(b, u, s, \ell, k, n, m, r)$, $b \in \mathcal{B}$, $u \in \mathcal{U}$, $s \in \mathcal{S}$, $\ell \in \mathcal{P}_b$, $k \in \{0\} \cup \mathcal{B} \backslash b$, $n \in \mathcal{P}_k$, $m \in \{0\} \cup \mathcal{B} \backslash b \backslash k$, $r \in \mathcal{P}_m$, to describe the design objective and the system constraints. We will show that the problem of associating users to BSs and allocating RBs and power levels amounts to determining an optimal set of octets from among all admissible ones. To determine admissible octets, we introduce an indicator variable $y_{bu}^{s\ell knmr}$ for each octet $(b, u, s, \ell, k, n, m, r)$. Specifically, $y_{bu}^{s\ell knmr}$ represents the fraction of time during which the octet $(b, u, s, \ell, k, n, m, r)$ is used for transmission, and hence, $y_{bu}^{s\ell knmr} \in [0, 1]$.

D.0.1 System Constraints

QoS Constraints: Ensuring that the QoS requirement of user u is satisfied is equivalent to ensuring that

$$
\sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 0000} R_{bu}^{s\ell 0000} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn00} R_{bu}^{s\ell kn00} \right.
$$
$$
\left. + \sum_{m \in \mathcal{B} \backslash b \backslash k} \sum_{r \in \mathcal{P}_m} y_{bu}^{s\ell knmr} R_{bu}^{s\ell knmr} \right) \geq Q_u, \ u \in \mathcal{U}.
$$
(D.1)

This constraint must be satisfied only for those users accommodated in the network. Otherwise, it must be excluded from the formulation. Please see the formulation in (D.10) for mathematical characterization of this conditional constraint. In this

constraint, the first term corresponds to an RB-reuse case of order $q = 1$, the second term corresponds to an RB-reuse case of order $q = 2$, and the third term corresponds to an RB-reuse case of order $q = 3$.

RB Allocation Constraint: Our formulation allows each RB to be reused opportunistically. In particular, each RB can be used exclusively by one BS during a fraction of the signalling interval and reused during the remainder of the signalling interval. Now, in the cases where RB-reused is considered, each RB can be reused either two or three times. Hence, the following constraint must be satisfied.

$$\sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 0000} + \frac{1}{2} \sum_{k \in \mathcal{B} \setminus b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn00} + \frac{1}{3} \sum_{m \in \mathcal{B} \setminus b \setminus k} \sum_{r \in \mathcal{P}_m} y_{bu}^{s\ell knmr} \right) \leq 1, \quad s \in \mathcal{S}. \tag{D.2}$$

In this constraint, the first term corresponds ton an RB-reuse case of order $q = 1$, the second term corresponds to an RB-reuse case of order $q = 2$, and the third term corresponds to an RB-reuse case of order $q = 3$.

Interference-Coupling Constraints: Let us consider a case in which a particular time fraction $y_{bu}^{s\ell knmr}$ is strictly greater than zero. Such $y_{bu}^{s\ell knmr}$ implies that the k-th and m-th BSs are respectively transmitting signals with power level n and r for at least a time fraction of $y_{bu}^{s\ell knmr}$, $b \in \mathcal{B}$, $u \in \mathcal{U}$, $s \in \mathcal{S}$, $\ell \in \mathcal{P}_b$, $k \in \{0\} \cup \mathcal{B} \setminus b$, $n \in \mathcal{P}_k$, $m \in \{0\} \cup \mathcal{B} \setminus b \setminus k$, $r \in \mathcal{P}_m$. Since a BS can transmit to multiple users at different time instants, it follows that $y_{bu}^{s\ell knmr}$ must satisfy the following constraint:

$$y_{bu}^{s\ell kn00} \leq \sum_{\tilde{u} \in \mathcal{U} \setminus u} y_{k\tilde{u}}^{snb\ell 00}, \ b \in \mathcal{B}, u \in \mathcal{U}, s \in \mathcal{S}, \ell \in \mathcal{P}_b, k \in \mathcal{B} \setminus b, \ n \in \mathcal{P}_k, \tag{D.3}$$

$$y_{bu}^{s\ell knmr} \leq \sum_{\tilde{u} \in \mathcal{U} \setminus u} y_{k\tilde{u}}^{snb\ell mr}, \ b \in \mathcal{B}, u \in \mathcal{U}, s \in \mathcal{S}, \ell \in \mathcal{P}_b, k \in \mathcal{B} \setminus b \setminus m, \ n \in \mathcal{P}_k, \tag{D.4}$$
$$m \in \mathcal{B} \setminus b \setminus k, r \in \mathcal{P}_m,$$

$$y_{bu}^{s\ell knmr} \leq \sum_{\tilde{u} \in \mathcal{U} \setminus u} y_{m\tilde{u}}^{srb\ell kn}, \ b \in \mathcal{B}, u \in \mathcal{U}, s \in \mathcal{S}, \ell \in \mathcal{P}_b, k \in \mathcal{B} \setminus b \setminus m, \ n \in \mathcal{P}_k, \tag{D.5}$$
$$m \in \mathcal{B} \setminus b \setminus k, r \in \mathcal{P}_m.$$

The constraint in (D.3) corresponds to an RB-reuse case of order $q = 2$, while the constraints in (D.4) and (D.5) correspond to an RB-reuse case of order $q = 3$. Moreover, it is worth noting that the constraints in (D.3), (D.4) and (D.5) are only relevant

110

when $k \neq 0$ and $m \neq 0$.

RB-Reuse Constraints: Since in our formulation at most three BSs can concurrently use the same RB, and each BS can use one power level at any time instant, we have

$$y_{bu}^{s\ell0000} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn00} + \sum_{m\in\mathcal{B}\backslash b\backslash k}\sum_{r\in\mathcal{P}_m} y_{bu}^{s\ell knmr} \leq 1, \quad b\in\mathcal{B}, \ u\in\mathcal{U}, s\in\mathcal{S}, \ell\in\mathcal{P}_b.$$
(D.6)

This constraint enables RB-reuse to be invoked opportunistically, wherein the first term corresponds to an RB-reuse case of order $q=1$, the second term corresponds to an RB-reuse case of order $q=2$, and the third term corresponds to an RB-reuse case of order $q=3$.

User Association Constraints: In our framework, a BS cannot simultaneously use multiple power levels nor broadcast to multiple users on the same RB. This yields the following constraint:

$$\sum_{\ell\in\mathcal{P}_b}\sum_{u\in\mathcal{U}}\left(y_{bu}^{s\ell0000} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn00} + \sum_{m\in\mathcal{B}\backslash b\backslash k}\sum_{r\in\mathcal{P}_m} y_{bu}^{s\ell knmr}\right) \leq 1, \quad b\in\mathcal{B}, s\in\mathcal{S}. \text{ (D.7)}$$

BS Association Constraints: At any time instant and RB, each user can be served by at most one BS with one power level. This yields the following constraint:

$$\sum_{b\in\mathcal{B}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell0000} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn00} + \sum_{m\in\mathcal{B}\backslash b\backslash k}\sum_{r\in\mathcal{P}_m} y_{bu}^{s\ell knmr}\right) \leq 1, \quad u\in\mathcal{U}, s\in\mathcal{S}. \text{ (D.8)}$$

BS Power Budget Constraints: The total power transmitted by the b-th BS to all the users in the network must satisfy

$$\sum_{s\in\mathcal{S}}\sum_{b\in\mathcal{B}}\sum_{u\in\mathcal{U}}\sum_{\ell\in\mathcal{P}_b}\left(y_{bu}^{s\ell0000} + \sum_{k\in\mathcal{B}\backslash b}\sum_{n\in\mathcal{P}_k} y_{bu}^{s\ell kn00} + \sum_{m\in\mathcal{B}\backslash b\backslash k}\sum_{r\in\mathcal{P}_m} y_{bu}^{s\ell knmr}\right)p_b^\ell \leq P_b^{\max}, \quad b\in\mathcal{B}.$$
(D.9)

D.0.2 Design Objective

The design objective is composed of two components. The first characterizes the number of users accommodated by the system, and should be maximized to maximize

revenue. The second component characterizes RB usage, and should be minimized provided that the accommodated users' QoS constraints are satisfied.

Aggregate RB Usage: The summation of $\{y_{bu}^{s\ell knmr}\}$ corresponds to the normalized total time over which this RB is used. The aggregate usage of RBs assigned to the u-th user can be given as $\sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{\ell \in \mathcal{P}_b} (y_{bu}^{s\ell 0000} R_{bu}^{s\ell 0000} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn00} R_{bu}^{s\ell kn00} + \sum_{m \in \mathcal{B} \backslash b \backslash k} \sum_{r \in \mathcal{P}_m} y_{bu}^{s\ell knmr} R_{bu}^{s\ell knmr})$. Hence, the aggregate usage of RBs in the network can be expressed as $\sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} (y_{bu}^{s\ell 0000} R_{bu}^{s\ell 0000} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn00} R_{bu}^{s\ell kn00} + \sum_{m \in \mathcal{B} \backslash b \backslash k} \sum_{r \in \mathcal{P}_m} y_{bu}^{s\ell knmr} R_{bu}^{s\ell knmr})$.

Total Number of Accommodated Users: The number of users accommodated is indirectly related to RB usage, and the number of BSs and power levels. Hence, a user $u \in \mathcal{U}$ is guaranteed to be served whenever at least one of the assignment variables in the set $\{y_{bu}^{s\ell knmr}\}$, $b \in \mathcal{B}$, $u \in \mathcal{U}$, $s \in \mathcal{S}$, $\ell \in \mathcal{P}_b$, $k \in \{0\} \cup \mathcal{B} \backslash b$, $n \in \mathcal{P}_k$, $m \in \{0\} \cup \mathcal{B} \backslash b \backslash k$, $r \in \mathcal{P}_m$, is strictly greater than zero. Using the smooth indicator function defined in (10) in the revised manuscript, the total number of accommodated user can be expressed as $\sum_{u \in \mathcal{U}} (1 - \exp(-\frac{\sigma}{Q_u} \sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} y_{bu}^{s\ell 0000} R_{bu}^{s\ell 0000} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn00} R_{bu}^{s\ell kn00} + \sum_{m \in \mathcal{B} \backslash b \backslash k} \sum_{r \in \mathcal{P}_m} y_{bu}^{s\ell knmr} R_{bu}^{s\ell knmr}))$.

A Composite Objective: Our design objective is composed of the linear combination of two quantities: 1) the number of accommodated users, which ought to be maximized to increase revenue, and 2) the aggregate RB usage, which ought to be minimized to efficiently utilize the network resources. To synthesize this objective, we will weight each of its components by an appropriate scalar $\rho \in [0, 1]$. Using ρ, we define our design objective to be maximizing $\rho \sum_{u \in \mathcal{U}} (1 - \exp(-\frac{\sigma}{Q_u} \sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} y_{bu}^{s\ell 0000} R_{bu}^{s\ell 0000} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn00} R_{bu}^{s\ell kn00} + \sum_{m \in \mathcal{B} \backslash b \backslash k} \sum_{r \in \mathcal{P}_m} y_{bu}^{s\ell knmr} R_{bu}^{s\ell knmr})) - (1 - \rho) \sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} y_{bu}^{s\ell 0000} R_{bu}^{s\ell 0000} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn00} R_{bu}^{s\ell kn00} + \sum_{m \in \mathcal{B} \backslash b \backslash k} \sum_{r \in \mathcal{P}_m} y_{bu}^{s\ell knmr} R_{bu}^{s\ell knmr}$.

D.0.3   Network Instance With Time-Sharing

Using the characterization, the joint design problem can be cast in the following form:

$$
\max_{\mathbf{y},\,\mathbf{t}} \quad \rho \sum_{u \in \mathcal{U}} (1 - t_u) - (1 - \rho) \sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 0000} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn 00} \right.
$$

$$
\left. + \sum_{m \in \mathcal{B} \backslash b \backslash k} \sum_{r \in \mathcal{P}_m} y_{bu}^{s\ell knmr} \right),
$$

$$\tag{D.10a}$$

$$\text{subject to} \quad \text{(D.2)} - \text{(D.9)}, \tag{D.10b}$$

$$
\sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} \left(y_{bu}^{s\ell 0000} R_{bu}^{s\ell 0000} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn 00} R_{bu}^{s\ell kn 00} \right.
$$

$$
\left. + \sum_{m \in \mathcal{B} \backslash b \backslash k} \sum_{r \in \mathcal{P}_m} y_{bu}^{s\ell knmr} R_{bu}^{s\ell knmr} \right) \geq Q_u (1 - t_u), \quad u \in \mathcal{U}, \tag{D.10c}
$$

$$
t_u \geq \exp\Big(- \frac{\sigma}{Q_u} \sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \sum_{\ell \in \mathcal{P}_b} y_{bu}^{s\ell 0000} R_{bu}^{s\ell 0000} + \sum_{k \in \mathcal{B} \backslash b} \sum_{n \in \mathcal{P}_k} y_{bu}^{s\ell kn 00} R_{bu}^{s\ell kn 00}
$$

$$
+ \sum_{m \in \mathcal{B} \backslash b \backslash k} \sum_{r \in \mathcal{P}_m} y_{bu}^{s\ell knmr} R_{bu}^{s\ell knmr} \Big), \quad u \in \mathcal{U}, \tag{D.10d}
$$

$$t_u \geq 0, \quad u \in \mathcal{U}, \tag{D.10e}$$

$$y_{bu}^{s\ell knmr} \in [0,1], \quad b \in \mathcal{B}, \ u \in \mathcal{U}, \ s \in \mathcal{S}, \ \ell \in \mathcal{P}_b, \ k \in \{0\} \cup \mathcal{B} \backslash b, \ n \in \mathcal{P}_k,$$

$$m \in \{0\} \cup \mathcal{B} \backslash b \backslash k, \ r \in \mathcal{P}_m. \tag{D.10f}$$

The formulation given above is structurally identical to the one in (5.13), except with more optimization variables. Specifically, the number of the variables for the case that each RBs is allowed to be reused at most three times is $BUSL(1 + (B-1)L + (B-1)(B-2)L^2)$.