UNIVERSITY OF CALIFORNIA,
IRVINE

Energy Efficiency and Load Balancing in Next-Generation Wireless Cellular Networks

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Electrical and Computer Engineering

by

Kemal Davaslioglu

Dissertation Committee:
Professor Ender Ayanoglu, Chair
Professor Hamid Jafarkhani
Professor Ahmed Eltawil

2015
DEDICATION

To my family for always being there,
Contents

LIST OF FIGURES vi
LIST OF TABLES ix
LIST OF ALGORITHMS x
ACKNOWLEDGMENTS xi
CURRICULUM VITAE xii
ABSTRACT OF THE DISSERTATION xiv

1 Introduction 1

2 Load Balancing via Handovers 8
   2.1 Motivation .............................................. 8
       2.1.1 Contributions ................................... 10
       2.1.2 Related works ................................... 12
   2.2 System Model ........................................... 14
       2.2.1 Notations ......................................... 14
       2.2.2 System Setup ..................................... 15
   2.3 Handovers in LTE and Conventional Cell Selection Methods .......... 19
   2.4 Proposed Interference-Based Handover Decision Policy .............. 22
       2.4.1 Theoretical Analysis ................................ 22
       2.4.2 Implementation Steps and Required Changes ................. 25
       2.4.3 Coverage Area of the Proposed Handover Policy .......... 27
       2.4.4 Differences with Other Handover Policies ................ 28
   2.5 Numerical Results ....................................... 29
   2.6 Conclusion .............................................. 41

3 Efficiency and Fairness Trade-Offs in Schedulers 42
   3.1 Motivation .............................................. 42
       3.1.1 Related Works .................................... 44
       3.1.2 Contributions .................................... 46
   3.2 System Model ............................................ 48
3.3 Utility-Based Resource Allocation ........................................ 53
  3.3.1 Utility Functions ........................................ 53
  3.3.2 The Price of Fairness .................................... 54
  3.3.3 A Case Study ......................................... 56
3.4 FDPS Schedulers .................................................. 58
  3.4.1 Problem Formulation .................................... 58
  3.4.2 Implementation Constraints ................................. 61
  3.4.3 Optimality Conditions .................................. 62
  3.4.4 Set Partitioning Solution ................................ 63
  3.4.5 Finding The Exact Search Spaces ......................... 66
3.5 Numerical Results ................................................ 68
  3.5.1 Simulation Setup ....................................... 68
  3.5.2 Optimization Setup .................................... 70
  3.5.3 Numerical Evaluation ................................... 71
3.6 Conclusion ....................................................... 74

4 Common Rate Maximization in Two-Layer Cellular Radio Systems 76
  4.1 Motivation ................................................... 76
  4.2 System Model ................................................ 77
  4.3 Power Control ............................................... 81
    4.3.1 Single-Layer System ................................ 81
    4.3.2 Two-Layer System .................................. 83
  4.4 LP Solution and Heuristic Common Rate Maximization Algorithms ... 85
  4.5 Simulations .................................................. 88
  4.6 Conclusion .................................................. 92

5 Iterative Water-Filling Algorithms with Pricing for Wireless Network Energy Efficiency 94
  5.1 Motivation ................................................... 94
    5.1.1 Related Works ....................................... 95
    5.1.2 Contributions ...................................... 97
  5.2 Multi-cell Energy Efficiency Maximization Problem with Power Constraints in Single-Tier Networks ................................................. 98
  5.3 Multi-cell Energy Efficiency Maximization Problem with Power Constraints in Two-Tier Networks ............................................. 108
  5.4 Multi-Cell Energy Efficiency Maximization with Rate and Power Constraints ............................................. 110
  5.5 Simulation Results ........................................... 114
    5.5.1 Energy Efficiency and Throughput Results in Single-Tier Networks .......... 115
    5.5.2 Energy Efficiency and Throughput Results in Two-Tier Networks ......... 117
    5.5.3 Incorporating Minimum Rate Constraints in Two-Tier Networks .......... 119
  5.6 Conclusions .................................................. 122
6 Energy-Efficient Resource Allocation for Fractional Frequency Reuse in Heterogeneous Networks 123

6.1 Motivation ................................................................. 123

6.2 System Model ............................................................... 127
  6.2.1 Base Station Power Consumption Models ....................... 131

6.3 Energy-Efficient Resource Allocation Problem ....................... 133
  6.3.1 Proposed Solution .................................................... 134
  6.3.2 Optimality Conditions ............................................... 139
  6.3.3 Gradient Ascent Method ............................................ 140
  6.3.4 Convergence Analysis ............................................... 141

6.4 Energy-Efficient Resource Allocation with Pricing ..................... 142

6.5 Numerical Results ...................................................... 144

6.6 Conclusion .............................................................. 152

6.7 Derivations ............................................................. 153

7 Conclusions and Future Work ............................................ 158

Bibliography .............................................................. 163
List of Figures

2.1 Network signal diagram of a handover procedure. Proposed handover decision algorithm requires changes in Step 3. (MME stands for Mobility Management Entity.) ................................................................. 26

2.2 The simulation layout of a HetNet deployment is depicted. The layout includes a set of 19 hexagonal macrocells with each employing a 3-sector antenna. The setup is overlaid with 2 pico-eNBs per sector. The simulations are carried out for 10 active users per sector. Macrocell and picocell base stations are represented by squares and triangles, respectively, while users are denoted by circles. ................................................................. 29

2.3 Percentage of associated users at macrocell and picocell tiers using the proposed handover policy and conventional CS methods. More users can be offloaded to the picocell tier by increasing the CRE bias. The proposed handover policy achieves the most balanced loading between the macrocell and picocell tiers. ................................................................. 36

2.4 Distribution of the power difference between the RSRP measurements of the closest macrocell and the serving picocell for the picocell associated users with the proposed uplink interference-based handover policy. ................................. 36

2.5 Distribution of PL between the user and its serving base station for the proposed uplink interference-based handover policy and conventional CS schemes. The PL-based CS method provides the cell association with the minimum PL values, whereas the RSRP-based CS yields the one with larger PL values. The proposed handover policy and CRE-based CS methods offer gains within these two ends. ................................................................. 37

2.6 Illustration of cell associations for the users in the first sector of the center cell. The figures are for the single layer (a) and two-layer systems (b)-(d). The cell associations with RSRP- and PL-based CS methods are depicted in (b) and (c), respectively, while the cell association of the proposed policy is shown in (d). ................................................................. 39

3.1 The trade-off between efficiency and fairness is depicted for a family of utility functions. ................................................................. 47

3.2 The transmitter structure of an SC-FDMA system. ......................... 49

3.3 Feasibility region is depicted for a single-cell network with $K_c = 2$ users and $N = 15$ RBs. ................................................................. 57
3.4 The average number of iterations to solve the root problems are depicted for the theoretical upper bounds (dotted), and the empirical results considering the DFT constraints (dashed-lines) and without these constraints (solid lines).  

3.5 The c.d.f. of user rates for different schedulers are depicted. 

3.6 The aggregate user rates, price of fairness, fairness index, and aggregate user transmit power are depicted in (a)-(d), respectively, for different schedulers in a multi-cell multiuser scenario. Solid lines indicate the performance of each scheduler without any constraints, and the dashed-lines are for the schedulers with DFT constraints. 

4.1 Single-layer network layout with 19 hexagonal cells is displayed. Each cell has three-sector 120° directional antennas positioned at the center of the cell and each sector has one randomly placed user. Squares and circles depict macrocell base stations and users, respectively. 

4.2 Two-layer hierarchical network layout with 19 high-power base stations overlaid with 19 low-power base stations placed at predefined locations. Additional low-power base station layer employs omnidirectional antennas. Squares, triangles and circles show macrocell base stations, low-power base stations and users, respectively. 

4.3 The figure shows the average common rate versus system loading in a 57 user system when COST 231 Hata urban propagation model and the COST 231 Walish-Ikegami NLOS model are used. 

4.4 The figure shows the average common rate versus system loading in a 57 user system when (4.18) is used to model propagation in both layers. 

5.1 Single-level water-filling solution for energy efficiency maximization without pricing. 

5.2 Multi-level (or modified) water-filling solution for energy efficiency maximization where the pricing terms determine the water filling level on each subcarrier. 

5.3 Average sector energy efficiency and throughput of a single-tier network using the proposed iterative water-filling algorithms with different initial power levels. IWF stands for iterative water-filling. EE-Max and Throughput-Max correspond to the solutions for the energy efficiency and throughput maximization problems, respectively. The solutions without pricing correspond to the algorithm in [88]. 

5.4 Average power consumption of the proposed iterative water-filling algorithms in a single-tier network. 

5.5 Average sector energy efficiency and throughput of a two-tier network using the proposed iterative water-filling algorithms. CPA corresponds to the constant power allocation algorithm proposed in [89]. 

5.6 Average transmit power consumption of the proposed iterative water-filling algorithms with and without pricing in a two-tier network. 

5.7 Average sector energy efficiency and sector throughput for various minimum rate requirements.
5.8 (a) The outage probability of various minimum rate requirements and (b) the cumulative distribution function of user rates for the minimum rate requirement of 512 kbits/sec. ................................................................. 121

6.1 Illustration of spectrum allocations in a multi-tier FFR network with dynamic cell-center region boundaries in which the MeNBs employ three sector antennas, whereas pico eNBs have omnidirectional antennas. Note that the picocell coverage areas are not depicted in the figure. .................................................. 129

6.2 Illustration of the energy efficiency of a sector and the proposed algorithm solutions using the gradient ascent method. ......................................................... 146

6.3 Average energy efficiency per sector (a) and average sector throughput (b) are depicted for the EBW scheduler. ................................................................. 147

6.4 Average energy efficiency per sector (a) and average sector throughput (b) are depicted for the SRM scheduler. ................................................................. 147

6.5 Average energy efficiency per sector (a) and average sector throughput (b) versus the ratio of the cell-center region radius to cell radius are depicted for EBW scheduler. Note that dashed lines represent the algorithm with pricing, while the solid lines represent power control only. ........................................ 150
List of Tables

2.1 Adaptive Modulation and Coding (AMC) Parameters . . . . . . . . . . . . . 17
2.2 Simulation Parameters . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
2.3 Power Delay Profile of the ETU Channel Model [34] . . . . . . . . . . . . . 30
2.4 Uplink Results for a HetNet Deployment of 2 Picocells per Sector and Uniform User Distribution . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
2.5 Downlink Results for a HetNet Deployment of 2 Picocells per Sector and Uniform User Distribution . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
2.6 Uplink Results for a HetNet Deployment of 2 Picocells per Sector and Non-Uniform User Distribution . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
2.7 Downlink Results for a HetNet Deployment of 2 Picocells per Sector and Non-Uniform User Distribution . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
3.1 Simulation parameters . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 69
5.1 Simulation Parameters . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 115
6.1 Base Station Power Consumption Model Parameter Values [5] . . . . . . . 131
6.2 Simulation Parameters . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 145
# List of Algorithms

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pseudocode for the Proposed Handover Decision Policy</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>Heuristic Power Control Algorithm to Maximize the Common Rate for Single-Layer Networks</td>
<td>87</td>
</tr>
<tr>
<td>3</td>
<td>Heuristic Power Control Algorithm to Maximize the Common Rate for Two-Layer Networks</td>
<td>87</td>
</tr>
<tr>
<td>4</td>
<td>Bisection Method for the Iterative Water-Filling Algorithm</td>
<td>103</td>
</tr>
<tr>
<td>5</td>
<td>Iterative Water-Filling Algorithm with Pricing for Network Energy Efficiency Maximization</td>
<td>106</td>
</tr>
<tr>
<td>6</td>
<td>Iterative Water-Filling Algorithm with Pricing for Network Throughput Maximization</td>
<td>107</td>
</tr>
<tr>
<td>7</td>
<td>Iterative Water-Filling Algorithm with Pricing for Network Energy Efficiency Maximization with Minimum Rate Constraints in Two-Tier Heterogeneous Networks</td>
<td>113</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

“Enter to Learn; Depart to Serve” wrote on the walls of my high school. I have learned a lot here at UC Irvine since my first arrival to U.S. in 2010 and now I am ready to serve as I approach graduation. Although dissertation is an individual effort to write, I should acknowledge some special people whose guidance, help, and supports have enabled me to complete it.

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Ender Ayanoglu. This thesis would not be possible without his guidance and support. Throughout my studies, he has helped me overcome many obstacles and for that I am really grateful. He has been a great example of how an academic should be and I learned a lot from him.

Secondly, I would like to thank my committee members, Professor Hamid Jafarkhani and Professor Ahmet Eltawil, for providing me their valuable comments and suggestions which have significantly improved the quality of this dissertation. I also would like to thank Professor A. Lee Swindlehurst and Assistant Professor Marco Leverato for their valuable feedback and encouraging discussions during my qualifying exam.

I would like to express my thanks to Cemil Can Coskun, with whom I collaborated in the final parts of my research. I enjoyed and benefited from our discussion hours in front of the white board. I would like to thank all my friends in Irvine for being my family abroad. Especially, I owe a debt of gratitude to my friends Dr. Erdem Koyuncu, Dr. Hulya Seferoglu, Dr. Bahattin Tolga Oztan, Dr. Ricardo Valle Rios, Dr. Caner Guclu, Dr. Omer Nebil Naveroglu, Ecehan Uludag, and to many friends that I probably need a lot more space to mention, for all their friendship and support. I will always cherish the great memories we have had in Irvine.

My life would not be the same if I had not met Dr. Thania Muñoz Diaz, who has changed my life. Her love, support, and encouragement were essential during my studies.

Finally, I am deeply grateful to my parents, my grandparents, my uncle, my aunt, and my cousin for their eternal love and support. I would not be here without these amazing people and role models. I have always felt honored to be their sons, grandsons, nieces, and cousins.

My graduate studies was partially supported by fellowships from the Center for Pervasive Communications and Computing and the National Science Foundation under Grant No. 1307551.
CURRICULUM VITAE

Kemal Davaslioglu

EDUCATION

Doctor of Philosophy in Electrical Engineering 2015
University of California, Irvine Irvine, California

Master of Science in Electrical and Electronics Engineering 2010
Bogazici University Istanbul, Turkey

Bachelor of Science in Electrical and Electronics Engineering 2008
Bilkent University Ankara, Turkey

RESEARCH EXPERIENCE

Intern 2015
Broadcom Corporation Irvine, California

Graduate Research Assistant 2013–2015
University of California, Irvine Irvine, California

Intern 2012–2013
Broadcom Corporation Irvine, California

Research Assistant 2008–2010
Bogazici University Istanbul, Turkey

TEACHING EXPERIENCE

Teaching Assistant 2013–2015
University of California, Irvine Irvine, California

Teaching Assistant 2009–2010
Bogazici University Istanbul, Turkey
PREPRINTS


Distributed Interference Pricing Algorithm for Energy-Efficient Heterogeneous Networks with Rate Constraints 2015

JOURNAL PUBLICATIONS

Energy-Efficient Resource Allocation for Fractional Frequency Reuse in Heterogeneous Networks 2015
IEEE Transactions on Wireless Communications

Quantifying Potential Energy Efficiency Gain in Green Cellular Wireless Networks 2014
IEEE Communications Surveys and Tutorials

Efficiency and Fairness Trade-Offs in SC-FDMA Schedulers 2014
IEEE Transactions on Wireless Communications

CONFERENCE PUBLICATIONS

An Energy-Efficient Resource Allocation Algorithm with QoS Constraints for Heterogeneous Networks 2015
IEEE Globecom 2015

Interference-based Cell Selection in Heterogeneous Networks 2013
Information Theory and Applications 2013

Common Rate Maximization in Two-Layer Networks 2012
IEEE Globecom 2012

Power control in multi-layer cellular networks 2012
Available at arXiv:1203:6131[cs:NI]
ABSTRACT OF THE DISSERTATION

Energy Efficiency and Load Balancing in Next-Generation Wireless Cellular Networks

By

Kemal Davaslioglu

Doctor of Philosophy in Electrical and Computer Engineering

University of California, Irvine, 2015

Professor Ender Ayanoglu, Chair

This dissertation focuses on the resource allocation problem in next-generation cellular wireless networks. Our goal is to design and evaluate algorithms and procedures to provide a balanced load and to improve the energy-efficiency of these networks, while satisfying the quality-of-service constraints of the users. The contributions of this dissertation are (i) a new handover policy to balance the load in Long Term Evolution (LTE) Heterogeneous Networks (HetNets), (ii) an analytical characterization of the efficiency and fairness trade-off of LTE uplink schedulers, and (iii) energy-efficient resource allocation algorithms for LTE HetNets with quality-of-service constraints (QoS).

First, we address the load balancing problem in HetNets deployments. We focus on the cell selection and handover policies as more base stations with different properties and functionalities are deployed. Conventional methods such as the strongest cell approach, used in single-layer network architectures, do not offer balanced loading or optimal performance for the HetNets due to the transmit power differences, backhaul, and access constraints of different base station types. Therefore, we propose a new handover decision policy that employs cell breathing, which dynamically adjusts the cell coverage regions based on the uplink interference measurements and current system load. The proposed policy also contributes to the self-adaptive and self-organization goals of the next-generation cellular systems.
Next, we investigate the uplink resource scheduling problem in single carrier frequency-domain multiple access systems. We present an efficient implementation method that translates these scheduling problems into set partitioning problems. Then, we discuss a family of utility functions that enable us to investigate the performance of different frequency domain schedulers such as the sum-rate maximization, proportional fair, and max-min fair schedulers. We use the price of fairness as a metric to analytically quantify the efficiency and fairness trade-offs in the schedulers and provide several upper bounds. We believe that this type of analysis can provide guidelines for the network operators to control the efficiency and fairness trade-off as the data traffic grows.

Finally, we investigate the multi-cell multi-carrier network energy efficiency problem where we propose utility-based energy-efficient resource allocation algorithms. We consider a linearized load-adaptive power consumption model at the base stations. We study the interference pricing mechanisms which include the inter-cell interference contributions and penalize the transmissions based on the interference they create. We propose two types of power control algorithms. First, we propose an iterative multi-level water-filling algorithm for multi-cell wireless networks. Second, we employ the gradient ascent method to control the transmit power of base stations. Both of these frameworks are extended to include QoS constraints such as the minimum rate constraints for each user. We present the optimality conditions and convergence of both algorithms, along with their performance evaluations.
Chapter 1

Introduction

A challenging quest has started in 2010 by a consortium lead by Alcatel Lucent. The so called \textit{1000x challenge} has been proposed by the GreenTouch consortium with the goal of developing key technology enablers, architectures, protocols, and solutions to increase the network energy efficiency by a factor of 1000 compared to 2010 levels [1]. Obviously, this is a challenging quest. Especially, considering the facts that (i) global mobile data traffic has an annual growth rate of 60-70\% [1–3] and (ii) the 110 million new mobile subscriptions just in the third quarter of 2014 globally [3], it is a very important quest with both economical and ecological impacts. Today, the Information and Communication Technologies (ICT) is responsible for 2-4\% of all the Carbon footprint generated by human activity [2]. In general terms, this corresponds to the same amount of carbon footprint of the aviation industry [1,2]. Considering the unprecedented increase in the data rate demand and the massive number connected devices (which include mobile users using smart phones, tablets, laptops, etc. and new devices such as cameras, sensors, and smart home and grid devices), it is clear that network operators have to find sustainable ways to satisfy these constraints. As the GreenTouch project evolved since 2010, the consortium has contributed to developing many solutions. They have recently revisited their main goal. As of 2015, they have concluded in
their research study, called as “Green Meter,” that it is possible to reduce the net network energy consumption by up to 90% by 2020 and achieve energy efficiency improvements up to 1043 times [1]. Clearly, this is a good news despite the fact that more research needs to be done to realize the solutions and fully exploit these gains.

It needs to be noted that GreenTouch is not the only project that addresses the network energy efficiency problem. For example, Energy Aware Radio and neTwork tecHnologies (EARTH) is an European Union founded project which combines fifteen partners from both industry and academia. It is jointly led by Ericsson and Alcatel Lucent. Similar to GreenTouch, the goal of EARTH is to decrease the energy consumption of radio access networks (RAN) by half through employing more efficient hardware components, smart radio resource algorithms, and better network management and deployment strategies [4]. As a part of EARTH project, the Energy Efficiency Evaluation Framework (E3F) has been developed. This framework has provided evaluation methodology for energy consumption. We need to note that several of these tools have been employed in this dissertation. For example, the base station power model for macrocell and picocell base stations has been adopted from [5] which stems from EARTH project. Other similar projects include OPERA-Net, Green Radio [6], and FANTASTIC-5G [7]. In these projects, the common goal is develop solutions that reduce the network power consumption without any performance degradation.

The 1000x challenge is not going to be achieved through a single solution, but rather it will consist of a cumulative gain achieved through new enabling technologies, architecture changes, new energy efficient components, and intelligent algorithms and protocols. While many of the key enabling technologies are already identified, there is still much research that needs to be done to fully develop these technologies for being ready for implementation. Some of the key enablers in next-generation cellular networks are going to be (i) increasing the spectrum to higher frequency bands, (ii) increasing the spectrum allocation flexibility (including methods such as carrier aggregation or implementing cognitive radio
for better spectrum usage), (iii) developing advanced multiple antenna technologies (such as massive multiple-input multiple-output, abbreviated as mMIMO, and distributed antennas systems), (iv) implementing better access and backhaul integration and multi-hop technologies, (v) reducing the link distances through small cells with higher energy efficiency and self-organization capabilities, (vi) using multiple base station coordination to improve link capacities (methods such as coordinated beamforming, joint processing, and intelligent scheduling), and (vii) device-to-device communications. A good summary of these solutions, challenges ahead, and many of the open problems can be found in the survey paper [2].

One of the main themes of this dissertation is the HetNet architecture. This type of an architecture includes a macrocell base station providing mainly the coverage over the site area and service for typically high mobility users. These users can be on their cars commuting or using buses or trains. Underlying the macrocell tier lies the small cell tier which includes densely deployed low-power base stations that provide higher data rates for low mobility users. This type of a network architecture that includes a macrocell and an underlying tier of small cells is called an umbrella network [2]. Macrocell base stations dissipate much more energy compared to the small cells, but they have larger coverage areas. They have higher computational power, are more reliable, and have faster backhaul. The small cells, on the other hand, have less computational capabilities. Depending on the type of the small cells, they may have different type of backhaul properties. Picocells are deployed by the network operators, and in general, have a reliable and fast backhaul. Relays are deployed based on the network operators’ plan, and they have typically wireless (mostly line of sight) backhaul. Femtocells are purchased by the customers or given to the customers by the network operators. They are plug-and-play type of devices that require minimum amount of installation effort. They need to be self-organizing and provide coverage indoors as femtocells target the home or small business market segments. One of the common problems with the femtocells is that they use the customer’s infrastructure which may not always be reliable or be able support the required backhaul traffic. For this reason, the implemented algorithms
need to require as small backhaul traffic as possible. It is clear that small cells are essential in next-generation networks. Here quoting the chairman and CEO of Qualcomm Incorporated, Dr. Paul E. Jacobs, “small cells greatly increase capacity by bringing the network closer to the user, thus enabling operators to serve the anticipated 1000x growth in mobile data traffic and dramatically improving the experience for wireless subscribers [8].” In this dissertation, several problems regarding the HetNet architecture are addressed. For example, in Chapter 2, we study the user and base station association problem, and in Chapters 5-6, we investigate the power control and interference mitigation problems.

In this dissertation, we investigate energy-efficient and load balancing techniques for fourth and future generation cellular wireless networks such as LTE and LTE-Advanced. Through these studies, we aim to contribute to the efforts solving problems in wireless cellular networks such as achieving higher energy-efficiency and providing a balanced load. In particular, we propose a handover/hand-off algorithm to protect the cell-edge users in HetNets, characterize the efficiency and fairness trade-off in LTE uplink schedulers, and study several energy efficient resource allocation schemes for LTE HetNets. This dissertation consists of a collection of seven chapters summarizing our findings.

Chapter 2 addresses the load balancing problem in HetNets deployments. In particular, we focus on the cell selection problem as more base stations with different properties and functionalities are deployed in the next-generation wireless cellular networks. Conventional methods such as the strongest cell approach, used in single-layer network architectures, do not offer balanced loading or optimal performance for the HetNets due to the downlink power level differences of different base station types, backhaul, and access constraints. Therefore, we propose a new handover decision policy that employs cell breathing. The proposed handover decision policy dynamically adjusts the cell coverage regions based on the uplink interference measurements and the current system load by providing better user-base station assignments. The proposed policy also contributes to the self-adaptive and self-organization
goals of the next-generation cellular systems. We present its implementation steps and demonstrate its performance results through simulations. We compare its performance to the conventional cell selection methods. Our simulation results demonstrate that the proposed handover decision policy provides a better load balancing and improved rates to the users.

Chapter 3 investigates the uplink resource scheduling problem in single carrier frequency-domain multiple access systems. In particular, we focus on the efficiency and fairness trade-offs in scheduling and resource allocation for wireless cellular networks. We present an efficient implementation method that translates these scheduling problems into set partitioning problems that are well-studied in the literature. Then, we discuss a family of utility functions that enable us to investigate the performance of different frequency domain schedulers such as the sum-rate maximization, proportional fair, and max-min fair schedulers. We use the price of fairness as a metric to analytically quantify these trade-offs. Based on the intuition that fairness of resource allocation in cellular radio networks corresponds to the prioritization of cell-edge user rates, we demonstrate that the proportional fair scheduler significantly improves fairness among users, and increases the rates offered to the cell-edge and median users when compared to the sum-rate maximization scheduler. This comes at the cost of reducing the cell-center user rates and the aggregate user rate. We present the steps on how to take into account the practical implementation constraints, in particular, those related with the discrete Fourier transform implementation, in the problem formulation. Simulation results that illustrate these trade-offs are also presented. We conclude that this type of analysis can provide guidelines for the network operators to control the efficiency and fairness trade-off as the data traffic grows.

In Chapter 4, we address the common rate maximization problem in two-layer cellular networks where high-power and low-power base stations are co-located in the same geographical area. Interference becomes a serious problem when two or more layers are considered in the same network. For this purpose, power control in the downlink needs to be used to limit the
interference and to fully exploit the benefits of additional layer deployments. We present an analytical framework to the common rate maximization problem both with and without maximum power constraints and propose a heuristic algorithm. We present simulation results for the proposed approaches in a two-layer network setup and observe a significant common rate increase compared to single-layer wireless networks.

Chapter 5 studies the water-filling solutions for multi-carrier systems, which appears in many resource allocation problems. We layout a general framework to solve these problems. Considering realistic base station power consumption models, we formulate a network-wide energy efficiency maximization problem. Using tools from fractional programming, we cast this problem in the framework of bi-criterion optimization where rate maximization and power minimization are weighted accordingly. Interference pricing mechanism is applied to reduce the inter-cell interference and to achieve a higher network performance. We decompose the main problem into subproblems via dual decomposition. These subproblems are independently solved at each sector using limited information exchange between base stations. We first derive our expressions and present algorithms for the single-tier networks that consist of only macrocell base stations. Then, we extend our analysis to two-tier networks where picocell base stations are deployed to improve the network performance and reduce the link distances. Lastly, we extend our framework and include the quality-of-service constraints. We obtain closed-form expressions for the power level updates which are determined by the multi-level water-filling algorithm, or, as it is sometimes called as, the modified water-filling algorithm. We show the dependency of water levels on the interference and dual prices per subcarrier. The performance of the proposed algorithm is evaluated. Based on our simulation results, we demonstrate that the proposed algorithms can outperform the benchmark approaches in terms of energy efficiency by a factor of 2.7. These results, in fact, constitute the upper bound for the algorithms developed in Chapter 6.

Chapter 6 addresses the network energy efficiency problem, where we propose a utility-based
energy-efficient resource allocation algorithm for the downlink transmissions in HetNets. We consider the FFR method in order to mitigate the intra- and inter-cell interference. The proposed algorithm divides the resource allocation problem into frequency and power assignment problems and sequentially solves them. The proposed power control algorithm uses the gradient ascent method to control the transmit power of macrocell base stations as most of the power in the network is consumed there. We present the optimality conditions of the resource allocation problem and the convergence of the proposed algorithm. In order to mitigate the inter-cell interference further, we study the interference pricing mechanisms and obtain an upper bound to the maximum energy efficiency problem including the inter-cell interference contributions. The performance of the proposed algorithm is studied in a LTE system. Our simulation results demonstrate that the proposed algorithm provides substantial improvements in the energy efficiency and throughput of the network. It is also shown that interference pricing provides only marginal improvements over the proposed algorithm.

Chapter 7 summarizes our findings in this dissertation. It includes a summary of our contributions along with possible extensions and future works that can be built upon the presented work. Chapter 7 also provides a point of reference on how the contributions in this dissertation stand in the big picture of wireless networking.
Chapter 2

Load Balancing via Handovers

2.1 Motivation

The global mobile data traffic is expected to grow 13 times between 2012 and 2017 [9]. To answer the unprecedented growth in user data demand, next-generation cellular systems such as Long-Term Evolution (LTE) target significantly increased throughput and capacity requirements. There are enabling technologies in LTE systems such as employing multiple antennas at both base stations and user equipments (UEs), transmit beamforming, and carrier aggregation (CA). However, more will need to be done. The deployment of small cells such as picocells, femtocells, and relays is considered by the network operators in order to increase the network capacity, to avoid coverage holes, and to meet the user traffic demand. In accordance with this direction, a recent study in [10] predicts that each macrocell is expected to be overlaid with an average of three low power nodes (LPNs) by 2017.

The differences in base station types in heterogeneous network (HetNet) deployments need to be incorporated in network planning to fully exploit the gains. For this reason, network planning in HetNets, such as LTE systems, differs from conventional network planning in...
several aspects. In single-layer networks, users are typically connected to the “strongest cells” based on the reference signal received power (RSRP) measurements at UEs. While this works well for single-layer networks, there are other constraints to be considered for HetNets. First, macrocell and picocell base stations, namely MeNBs and pico-eNBs, differ by almost 16 dB in their downlink transmit power levels [11]. Consequently, RSRP-based cell selection (CS) in HetNets favors MeNBs, even when the path loss (PL) conditions between the pico-eNBs and the UEs are better, leading to highly unbalanced loading. Another constraint is the backhaul capacity. LPNs such as relays that are connected to the network through over-the-air interface and home-deployed femtocells have limited backhaul capacities compared to MeNBs. Also, depending on the network operators, access to some LPNs can be limited. In particular, femtocells can be configured to serve only to a closed subscriber group (CSG). Hence, the CS methods and handover policies in next-generation wireless networks need to consider these power, backhaul, and access constraints such that the network performance can be improved. Networks that include these constraints can achieve robustness, higher throughput, and more energy efficient transmission. Note that energy efficient transmission in the uplink leads to longer battery life for the UEs, while the energy efficient transmission in the downlink can provide significant energy savings for the network operators. These can only be true with a balanced loading in which the resources in MeNBs and LPNs are better utilized.

To balance the load considering the differences in base station types and to efficiently use the LPN deployments, LTE standards include techniques such as cell range extension (CRE)-based CS. In this method, a constant bias is added to the picocell RSRP measurements such that the picocell coverage area can be expanded. This provides significant improvements for UEs in the uplink since the link distances are reduced. However, in the downlink transmissions, picocell users in the range extension area are exposed to severe interference from the MeNB for two reasons. First, the users in picocell extension area are farthest away from the serving pico-eNB. Also, these users are much closer to the aggressor macrocell that conse-
quenty reduces their rates. Therefore, the base station that gives the maximum downlink rates can be different from the one for the uplink. This is referred to as the uplink-downlink imbalance.

2.1.1 Contributions

In this chapter, we study the load balancing problem in HetNet deployments. We identify the role of CS methods in load balancing and extend our previous results in [12]. We propose a new sub-optimal handover policy that dynamically balances the traffic load between the macrocell and picocell tiers. This policy can be seen as a CRE-based cell association algorithm with an adaptive bias. The proposed handover decision policy employs the uplink interference measurements and the RSRP measurements to adjust the cell coverage areas dynamically. Therefore, we refer to it as the uplink interference-based handover policy. It is a distributed policy that requires limited information exchange among the clustering base stations. The distributed nature of the proposed policy also serves the self-organizing and self-adjusting premises of the next-generation HetNet systems. In particular, it offers flexibility under unbalanced traffic conditions and user distributions. In the sequel, we discuss the following findings.

- An adaptive CRE bias that utilizes the channel and interference conditions of the system is derived in Section 2.4 to adaptively adjust the cell coverage regions. For the cells that incur high interference, lower bias values are obtained for their coverage areas to shrink. For the cells that are subject to less interference, their coverage areas need to expand to provide coverage for those areas previously covered by the loaded neighboring cells. The proposed handover decision policy makes the cell association decisions based on the uplink interference in order to increase the uplink transmission rates. However, this may not necessarily provide the best downlink rates, although our
simulation results do not indicate a downlink throughput degradation. The reason for focusing on the uplink interference is due to the homogeneity of the uplink transmission powers. The users transmit at similar power levels, whereas the downlink power levels, especially in HetNets, are subject to significant differences. By focusing on the uplink interference, the cell association and resource allocation problems can be decoupled. Given the cell associations, the resource allocation problem in the uplink and downlink transmissions can be carried out.

- The required changes in the handover procedure are identified. The proposed handover decision policy uses only the signals defined in LTE standards. We first summarize the network signaling for handover procedures for LTE systems in Section 2.4.2. Then, we provide the pseudocode of the proposed handover decision policy. The proposed policy necessitates the users to be connected to base stations. In this chapter, we consider the RSRP-based CS as the initial CS rule. Then, calculating the adaptive CRE bias values, the new cell associations are obtained. If handovers are required, the handover procedure is executed.

- We investigate the performance of the proposed uplink interference-based handover policy in Section 2.5. We compare its performance with the conventional CS methods. In particular, we focus our attention on the aggregate throughput. Our prior work in [12] did not consider small scale fading. In this chapter, we include these effects and employ frequency domain packet scheduling (FDPS) as well. In addition to a comprehensive study of uplink and downlink data rates, we also study the energy efficiencies of these methods. The proposed handover policy achieves 143.2% aggregate throughput improvement over the single-layer homogenous macrocell architecture, 25.7% over RSRP-based CS, and 4.9% over PL-based CS. An important result we show is that the proposed handover policy outperforms all the RSRP-, PL-, and CRE-based CS method both in the uplink and downlink. These improvements come with manageable
computational complexity that do not require a centralized processor, which makes the proposed handover policy more preferable. Our results demonstrate the achievable gains through only cell selection. These gains can be further improved with adaptive resource partitioning among coordinating base stations [13].

2.1.2 Related works

Early works on load balancing with CS investigated cell breathing\footnote{We use the term cell breathing as it is defined in [14].} techniques for Code Division Multiple Access (CDMA) networks are available, e.g., [14, 15]. Due to near-far effects, managing the intra-cell and inter-cell interference was critical for CDMA systems, and cell breathing was initially employed along with fast power control in an interference-limited system. In [14] and [15], the authors realized that cells need to shrink under heavy load and expand with light traffic. They proposed CS algorithms based on the uplink interference at each base station such that cell breathing can be applied. The study in [16] also addressed the cell association problem. It discussed the difficulties related with finding the optimal user-base station assignment based on the downlink transmissions. The authors of [16] proposed to find the cell associations that minimize the uplink transmission powers. Once these are obtained, power control steps are carried out for both uplink and downlink transmissions. In fact, the proposed handover decision policy is based on this idea. We will use the uplink interference values to decide on the cell associations in the proposed dynamic cell range adaptation framework. We should note that we do not consider a joint power control and base station selection procedure as it is not standardized in LTE. Instead, we first employ the proposed handover policy to solve the cell association problem and then proceed with the resource allocation.

Recently, there has been significant work in the literature focusing on CS and handover techniques for LTE networks. A similar work in [17] investigates cell breathing type handovers
in a macrocell-femtocell LTE network. We discuss the distinctions between the work in [17] and the one presented in this chapter in Section 2.4.4. The works in [18, 19] investigate the performance of various CS methods such as RSRP- and CRE-based criteria in LTE HetNet systems. In this work, we also employ these CS criteria as benchmarks to compare the performance of the proposed algorithm. The references in [20, 21] focus on the effects of CRE bias values for load balancing in HetNets and discuss the impact of user clustering and user distribution within the cell. Another related work in [22] studies the uplink downlink imbalance in LTE systems. In this chapter, we will extend this discussion through our simulation results. A joint formulation of cell association and resource allocation problem in a network utility maximization (NUM) framework is studied in [23]. The authors ignore the small-scale fading and employ a log-sum rate maximization (LSRM) scheduler. They show that the optimal resource allocation with LSRM scheduler is the equal distribution of resources to the users connected to the same base station. This enables them to decouple the cell association and resource allocation problems. In [23], the authors also propose a distributed close-to-optimal solution via the dual decomposition method, again ignoring the frequency-selective fading. The joint cell association and resource allocation framework is also studied in [24], although this time in a much smaller network consisting of a macrocell and an underlaying picocell layer. The authors investigate various CS criteria and frequency reuse patterns. They conclude that the universal frequency resource provides the best solution for all the investigated CS criteria in [24]. Albeit the fact that the NUM frameworks investigated in [23, 24] can find the optimal cell association and resource allocation solutions, they are centralized approaches. As it is present in any central processor approaches, the NUM frameworks require excessive traffic signaling, that are often unscalable for dense small cell deployments, and are vulnerable to failures (or crushes) in the central processor which have a network-wide effect. These disadvantages are not present in our proposed handover policy due to its distributed nature. A related study in [13] also investigates the throughput performance of HetNet deployments using similar simulation parameters. The study in [13]
considers the RSRP-based CS criteria and focuses on the macrocell offloading percentage with the deployment of 1, 2, 4, and 10 picocells per macrocell area. Although the picocells are deployed per macrocell area, while our simulation setup considers their deployment per macrocell sector, their results are comparable to those presented in this chapter.

The remainder of the chapter is organized as follows. Section 2.2 introduces the system models for both uplink and downlink transmissions. Section 2.3 describes the commonly used CS criteria such as RSRP, PL, and CRE with constant bias. In Section 2.4, we study the proposed handover decision policy. We first provide its theoretical derivation and then, supply the implementation steps in the LTE handover protocol. Simulation results are presented in Section 2.5. Finally, Section 2.6 summarizes our findings.

2.2 System Model

In this section, we first introduce the notation and definitions used in this chapter. Then, we briefly describe the system models, transmission schemes, and the uplink power control in LTE systems. Uplink transmissions in LTE are designed to achieve high power-efficiency, improved coverage, and reduced power consumption at UEs [25]. For this purpose, single carrier frequency-division multiple access (SC-FDMA) is adopted. We use localized FDMA and employ the set partitioning approach in [26].

2.2.1 Notations

We use the capital calligraphic font to denote the sets, e.g., the set \( \mathcal{N} \) includes \( \mathcal{N} = \{1, \ldots, N\} \). A subset of set \( \mathcal{N} \) is denoted by subscripting it as \( \mathcal{N}_k \), and \( |\mathcal{N}_k| \) denotes the cardinality of the subset. Also, any power value \( P \) is represented in dB with the notation \( P_{dB} \).
2.2.2 System Setup

We consider that the system bandwidth is divided into $M$ orthogonal subcarriers. These subcarriers are grouped into clusters to form $N$ non-overlapping resource blocks (RBs). We denote the set of subcarriers and RBs assigned to user $k$ by $\mathcal{M}_k$ and $\mathcal{N}_k$, respectively. Similarly, the number of subcarriers and RBs assigned to user $k$ are denoted by $M_k$ and $N_k$, respectively. The sampling time $T_s$ and the sampling frequency $f_s$ in LTE systems depend on the system bandwidth. The sampling frequency is given by $f_s = 1/T_s = B_{sc}M$, where the total number of subcarriers $M$ is taken as the IFFT size, and the bandwidth of each subcarrier is fixed to $B_{sc} = 15$ kHz. Typically, $M$ is chosen in powers of two for fast implementation purposes. For example, in a 10 MHz LTE bandwidth, the sampling frequency is $15.36$ MHz [27, p. 70]. There are $M = 1024$ subcarriers, only 600 of which are being used to convey data. These 600 subcarriers make up $N = 50$ RBs where each RB includes $N_{RB}^{sc} = 12$ consecutive subcarriers.

Let $\mathcal{K} = \{1, \ldots, K\}$ and $\mathcal{B} = \{1, \ldots, B\}$ denote the set of users and the set of base stations, respectively. Then, $c_k \in \mathcal{B}$ represents the base station that user $k$ is associated with. We form a $K \times 1$ vector $\mathbf{c}$ to represent all the user-base station assignments in the system. For a given cell association, the uplink closed-loop power control is given by [28]

$$P_{k}^{dB} = \min\{P_{\max}^{dB}, P_0^{dB} + 10\log_{10}(N_k) + \alpha P_{L_{k,c_k}}^{dB} + \Delta_{TF_k} + f_k\}, \quad (2.1)$$

where $P_k^{dB}$ denotes the uplink transmit power of user $k$, $P_{\max}^{dB}$ denotes the maximum UE transmit power, and $P_0^{dB}$ is open-loop transmit power. The total number of contiguous resource blocks (RBs) assigned to user $k$ are represented by $N_k$. The PL between user $k$ and its serving base station $c_k$ is denoted by $P_{L_{k,c_k}}^{dB}$. We assume shadow fading to be static over a frame duration [11], and PL includes the shadow fading of the link. The PL compensation factor $\alpha$ takes its value from the set $\{0,0.4,0.5,0.6,0.7,0.8,0.9,1\}$. The parameter $\Delta_{TF_i}$ is
a user-specific parameter used to adjust the power based on the assigned modulation and coding scheme, and \( f_i \) is a closed-loop power control adjustment parameter. In what follows, we consider the open-loop power control where the terms \( \Delta_T F_k \) and \( f_k \) in (2.1) are omitted as in [25, p. 195]. The UE transmission power is equally distributed on the allocated bandwidth. The corresponding UE transmit power spectral density per subcarrier is expressed as

\[
P_{k,m}^{dB} = P_k^{dB} - 10 \log_{10}(M_k),
\]

(2.2)

where \( P_{k,m} \) denotes the uplink transmit power of user \( k \) on subcarrier \( m \). Equivalently, (2.2) can be expressed in linear scale as \( P_{k,m} = P_k/M_k \) for all \( m \in \mathcal{M}_k \).

In order to investigate the throughput performance of frequency-domain schedulers, we consider small-scale fading and employ frequency selective fading channel models. For example, [29] includes power delay profile models for pedestrian and vehicular users. The channel gain between user \( k \) and base station \( c_k \) on subcarrier \( m \) is denoted by \( H_{k,c_k}(m) \). The channel gain has two components that consist of multipath and PL plus shadowing components. These can be expressed as

\[
H_{k,c_k}(m) = |V_{k,c_k}(m)|^2/PL_{k,c_k}, \quad \text{and} \quad V_{k,c_k}(m) = \sum_{l=1}^{L} v_k(l) \exp \left( -j2\pi m \tau_l/T_s \right),
\]

(2.3)

where \( V_{k,m} \) denotes the multipath effects in the received signal for an \( L \)-path impulse response model. \( T_s \) and \( \tau_l \) denote the sampling time and the delay of \( l \)-th path, respectively. Each multipath component \( v_k(l) \) can be further expressed as

\[
v_k(l) = \begin{cases} 
A \sqrt{P_{rel}(l)} w_k(l) & \text{if } l = \lfloor \tau_l/T_s \rfloor + 1 \\
0 & \text{otherwise},
\end{cases}
\]

(2.4)

where \( w_k(l) \) denotes a zero-mean white Gaussian noise process. \( P_{rel}(l) \) denotes the relative power of \( l \)-th path. The term \( A \) normalizes the average multipath power to unity such that
Table 2.1: Adaptive Modulation and Coding (AMC) Parameters

<table>
<thead>
<tr>
<th>MCS Index</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modulation</td>
<td>×1024</td>
<td>QPSK</td>
<td>16-QAM</td>
<td>64-QAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coding Rate</td>
<td></td>
<td>78 602</td>
<td>567 666</td>
<td>873 772</td>
<td>948 616</td>
<td>616 490</td>
<td>490 378</td>
<td>378 602</td>
<td>602 120</td>
<td>120 78</td>
<td>0.3770</td>
<td>0.6909</td>
<td>0.8770</td>
<td>1.1758</td>
<td>1.4706</td>
<td>1.9411</td>
</tr>
<tr>
<td>SNR (dB)</td>
<td>-6.934</td>
<td>-5.147</td>
<td>-3.148</td>
<td>-1.254</td>
<td>0.761</td>
<td>2.697</td>
<td>4.697</td>
<td>6.528</td>
<td>8.576</td>
<td>10.37</td>
<td>12.3</td>
<td>14.18</td>
<td>15.89</td>
<td>17.82</td>
<td>19.83</td>
<td></td>
</tr>
</tbody>
</table>

\[ E[\sum |v_k(l)|^2] = 1. \]

Let us now define the signal-to-interference-plus-noise ratio (SINR) on each subcarrier \( m \) assigned to user \( k \) as

\[
\gamma_{k,m} = \frac{P_{k,m}H_{k,c_k}(m)}{\sum_{k' \in K_m} P_{k',m}H_{k',c_k}(m) + \sigma^2_{ck}},
\]

where \( K_m \) denotes the set of users that are scheduled to subcarrier \( m \). The interfering users from the neighboring cells on subcarrier \( m \) are denoted by \( k' \in K_m \). It is assumed that co-channel spectrum allocation is used for both macrocell and picocell tiers. The received thermal noise power over the bandwidth of a subcarrier \( m \) at the base station \( c_k \) is given by \( \sigma^2_{ck} \).

In the uplink transmissions, we assume that ideal channel estimates for each link are available. Typically, frequency domain minimum mean square error (MMSE) equalizers are employed at the base station receivers. The wideband SINR for user \( k \) at \( n \)th symbol, \( \gamma_{k,n} \), can be calculated using the individual SINR of each subcarrier \( m \) assigned to user \( k \), \( \gamma_{k,m} \) as [27, pp. 88-91]

\[
\gamma_{k,n} = \left( \frac{1}{M_k} \sum_{m \in M_k} \frac{1}{\gamma_{k,m}} - 1 \right)^{-1}
\]

The wideband SINRs at each symbol can be compressed to a single packet effective SINR value by a method called exponential effective SINR mapping (EESM). In this method, the
mapping is based on the Chernoff bound for the error probability. For a given modulation coding scheme (MCS), the packet effective SINR is given by [30,31]

$$\gamma_{e_{f,k}} = -\beta \ln \left( \frac{1}{N_{sym}} \sum_{n=1}^{N_{sym}} \exp \left( -\frac{\gamma_{k,n}}{\beta} \right) \right),$$

(2.7)

where $\beta$ is an adjusting factor that can be obtained through link-level simulations and $N_{sym}$ denotes the number of symbols within a subframe. In order to map the packet effective SINR to user throughput, we consider the link adaptation framework proposed in standards that employs adaptive modulation and coding (AMC) methods [28]. Table 2.1 presents the modulation schemes, coding rates, and efficiency values in [28]. The base stations use the channel quality indicator (CQI) feedback transmitted by the UE in the uplink that indicate the supported data rate. For each reported CQI value, the scheduler adjusts the modulation and coding rate. The same AMC is used for all RBs scheduled to a UE. The CQI measurements are based on the channel conditions such as the interference, channel gain between UE and base station, and noise level. CQI takes only the integer values between 0 and 15 as also depicted in Table 2.1. We adopted the SINR and $\beta$ threshold values reported in [32] that achieve a block error ratio (BLER) less than 0.1 at the first transmission. We should note that in real systems, these threshold values are not constant and need to be readjusted, probably at a slower rate, in order to include practical constraints such as feedback delay, channel estimation errors, and varying channel statistics [33]. For user $k$, one of the modulation and code rate pairs is assigned based on the CQI feedback. Then, the throughput of user $k$ is expressed as

$$T_k = \frac{N_k N_{sc}^{RB} N_{sym} \varepsilon}{T_{subframe}},$$

(2.8)

where $\varepsilon$ is the efficiency obtained using Table 2.1, which is in bits/symbol. $T_{subframe}$ is the duration of a subframe. The other parameters, $N_k$, $N_{sc}^{RB}$, and $N_{sym}$, in (2.8) are as defined above.
2.3 Handovers in LTE and Conventional Cell Selection Methods

In this section, we first identify the reasons for handovers in LTE systems. Then, we proceed to discuss the conventional CS criteria. We study the RSRP-, PL-, and CRE-based CS methods. Handovers can be divided into three categories as quality-, coverage-, and load-based handovers [34]. First, the quality-based handover is initiated when the UE has better channel quality with a candidate base station than its serving base station. The coverage-based handover occurs when UE is moving towards or already in a coverage hole which often occurs when the user is moving away from an urban area towards a rural area. In this case, the network needs to handover the connection to a different base station or a different radio access technology (RAT) such as Global System for Mobile (GSM) or Universal Mobile Telecommunications System (UMTS) [34]. Finally, with the load-based handovers, the congested network needs to handover some users to a neighboring base station to balance the traffic load. The proposed handover policy, presented in Section 2.4, is a load-based handover type.

The downlink reference signal transmitted by base station $c$ is attenuated by the PL value until it reaches UE $k$, and the RSRP can be expressed mathematically as [28]

$$\text{RSRP}_{k,c}^{dB} = \frac{p_{RS,c}^{dB}}{PL_{k,c}^{dB}},$$

(2.9)

where $p_{RS,c}^{dB}$ denotes the reference signal power of the base station $c$ in decibels. Note that (2.9) can also be represented in the linear domain as $\text{RSRP}_{k,c} = p_{RS,c} / P L_{k,c}$. In this work, we consider that base stations transmit at maximum power for the reference signal transmissions such that $p_{RS,c}^{dB} = 46$ dBm for MeNBs and $p_{RS,c}^{dB} = 30$ dBm for pico-eNBs. Using these definitions, with the strongest CS method, user $k$ selects the base station that maximizes
the received power of the reference signal. This is given by

\[ c_k = \arg \max_{c \in \mathcal{B}} \text{RSRP}_{k,c} = \arg \max_{c \in \mathcal{B}} \frac{p_{RS,c}}{PL_{k,c}}, \tag{2.10} \]

where \( PL_{k,c} \) is the PL between user \( k \) and base station \( c \), and \( p_{RS,c} \) denotes the average power of downlink reference signals for the \( c \)th base station. The reference signals are cell-specific and orthogonal in both time and frequency domains to provide robust channel estimation and sufficient antenna separation [35]. In the rest of this chapter, we will refer to this method as the RSRP-based CS.

In PL-based CS, user \( k \) connects to the base station with the maximum channel gain. Following the notation defined in the previous section, the channel gain includes PL and small-scale fading components. Note that a smaller PL gives a larger channel gain. Also, the channel estimates are again obtained through the cell-specific reference signals. Then, the PL-based CS is defined as

\[ c_k = \arg \min_{c \in \mathcal{B}} PL_{k,c}. \tag{2.11} \]

In single-layer homogenous networks, RSRP- and PL-based CS methods are equivalent since base stations allocate the same power levels for reference signals. A variation of PL-based CS method was used in GSM systems [36], but this criterion is not included in the LTE standards. However, we will still study its performance as a reference since PL-based CS selects closer base stations, and thereby, improves the uplink user rates compared to RSRP-based CS method.

We also investigate the performance of the CRE-based CS scheme. This method extends the picocell coverage with a constant offset to offload macrocell users to picocells. Also, similar to the RSRP-CS method, the downlink RSRP measurements at the UE are used for CS.
The CRE-based CS is given by

\[ c_k = \arg \max_{c \in B} \text{RSRP}_c^{dB} + \text{Bias}_c^{dB}. \]  

(2.12)

For each MeNB, \( \text{Bias}_c = 0 \) is assumed and typical values for pico-eNB offset values are 3, 6, 9, and 12 dB. For instance, user \( k \) chooses a pico-eNB \( p \) with a \( \text{Bias}_p \) instead of an MeNB \( m \) under CRE-based CS if we have

\[ \text{RSRP}_m^{dB} < \text{RSRP}_p^{dB} + \text{Bias}_p^{dB}. \]  

(2.13)

Note that as the pico-eNB offset value \( \text{Bias}_p^{dB} \) increases, the picocell coverage also increases. Let us note that although high pico-eNB offset values increase the uplink rates, these decrease the downlink user rates. In particular, if the MeNBs and pico-eNBs operate on the same frequency bands, picocell users in the range extension region experience severe cross-tier interference from the MeNBs in the downlink. For this reason, the cell coverage regions and offset values are critical parameters that determine the uplink and downlink user rates.

There are several drawbacks of these three CS schemes. First, these depend solely on PL and downlink transmit powers. They do not take into account the instantaneous interference in the system, i.e., they are not dynamic. For instance, when the number of interfering users in neighboring cells increases or the data rate demand of the interfering users is high, all three of these CS schemes yield poor results. Second, the traffic demand in wireless cellular systems is subject to significant temporal and spatial changes [25]. This also indicates that if approaches such as adjusting the constant offsets based on field measurements are pursued, they need to incorporate temporal behavior of the traffic demands at each site which would increase the cost and execution time of these adjustments. In order to eliminate these drawbacks, in the following section, we introduce an adaptive CS criterion.
2.4 Proposed Interference-Based Handover Decision Policy

In this section, we introduce the proposed handover policy. First, we derive its analytical expression and state the handover decision rule. Then, we discuss the implementation steps and required changes to the existing protocols. In particular, we study these for the LTE handover protocols. Finally, we discuss some drawbacks of the proposed method.

2.4.1 Theoretical Analysis

First, we should note that the optimal cell association closely depends on the power control method. For example, the works in [14, 15] proposed optimal CS rules based on the power control rule used in CDMA networks. In this chapter, we focus on LTE systems and employ the open-loop uplink power control rule. In the NUM framework, the objective of the cell association and resource allocation problem for a multicarrier system can be written as

\[ c^* = \arg \max_{c \in \mathcal{M}} \sum_{k \in K} \sum_{m \in \mathcal{M}_k} T_{k,m}, \]  

(2.14)

where \( T_{k,m} \) denotes the throughput of user \( k \) on subcarrier \( m \). The summation of \( T_{k,m} \) over \( m \in \mathcal{M}_k \) gives the throughput of user \( k, T_k \), and we seek to find the best user to base station assignment over the complete set of user to base station assignments. For each user to base station association \( c \), a new frequency domain scheduling is carried out. However, this rule is prohibitively complex to be implemented as it requires a central processor for the cell association, frequency band allocations, and power assignments. Also, in a HetNet system with dense small cell deployments, it is almost impossible to support the backhaul traffic for signalling the instantaneous channel conditions to a central processor. In what follows, we will discuss an alternative cell association method that does not require a central processor,
i.e., that is distributed, and adaptive to instantaneous uplink interference conditions.

Given a set of users and their associated base stations, the proposed handover policy seeks to improve the user SINR in a noncooperative scenario. The subcarrier SINRs, $\gamma_{k,m}$, are mapped to an equivalent wideband SINR $\gamma_{k,n}$ using (2.6), and the following is true

$$\gamma_k \leq \frac{1}{M_k} \sum_{m \in M_k} \gamma_{k,m}.$$  \hfill (2.15)

This expression holds with equality if for all $\gamma_{k,m} = \gamma$ and $\gamma_{k,m} \geq 0$. With the proposed method, user $k$ connects to the base station that satisfies the following

$$\tilde{c}_k = \arg \max_{c \in B} \frac{1}{M_k} \sum_{m \in M_k} \frac{P_{k,m} H_{k,c}(m)}{I_{k,c}(m)},$$  \hfill (2.16)

where

$$I_{k,c}(m) = \sum_{k' \in K_m} P_{k',c} H_{k',c}(m) + \sigma_m^2,$$  \hfill (2.17)

denotes the interference that user $k$ incurs at base station $c$ on subcarrier $m$, and $\sigma_m^2$ is noise per subcarrier as defined previously. Since $M_k$ is constant and we ignore the transmit power of user $k$ to derive a closed form expression, then (2.16) can be rewritten as

$$\tilde{c}_k = \arg \max_{c \in B} \sum_{m \in M_k} \frac{H_{k,c}(m)}{I_{k,c}(m)} = \arg \max_{c \in B} \sum_{m \in M_k} \frac{|V_{k,c}(m)|^2}{PL_{k,c} I_{k,c}(m)}.$$  \hfill (2.18)

Using (2.9) for the expression in (2.18) gives

$$\sum_{m \in M_k} \frac{|V_{k,c}(m)|^2}{PL_{k,c} I_{k,c}(m)} = \sum_{m \in M_k} \frac{\text{RSRP}_{k,c} |V_{k,c}(m)|^2}{p_{RS,c} I_{k,c}(m)} = \frac{\text{RSRP}_{k,c}}{p_{RS,c}} \sum_{m \in M_k} \frac{|V_{k,c}(m)|^2}{I_{k,c}(m)},$$  \hfill (2.18)

which enables us to use the RSRP feedbacks for the handover decisions. Then, the proposed
handover criterion is given by

$$\tilde{c}_k = \arg \max_{c \in \mathcal{B}} \frac{\text{RSRP}_{k,c}}{p_{RS,c}} \sum_{m \in \mathcal{M}_k} \frac{|V_{k,c}(m)|^2}{I_{k,c}(m)}.$$  \hfill (2.19)  

Consider that user $k$ is connected to base station $c$. Using the handover criterion in (2.19), user $k$ is handed over to a candidate base station $p$ if the following is true

$$\frac{\text{RSRP}_{k,p}}{p_{RS,p}} \sum_{m \in \mathcal{M}_k} \frac{|V_{k,p}(m)|^2}{I_{k,p}(m) - P_{k,m}H_{k,p}(m)} > \frac{\text{RSRP}_{k,c}}{p_{RS,c}} \sum_{m \in \mathcal{M}_k} \frac{|V_{k,c}(m)|^2}{I_{k,c}(m)},$$  \hfill (2.20)  

or, equivalently, we have that

$$\text{RSRP}_{k,c} < \text{RSRP}_{k,p} \times \frac{p_{RS,c} \sum_{m \in \mathcal{M}_k} \frac{|V_{k,p}(m)|^2}{I_{k,p}(m) - P_{k,m}H_{k,p}(m)}}{p_{RS,p} \sum_{m \in \mathcal{M}_k} \frac{|V_{k,c}(m)|^2}{I_{k,c}(m)}},$$  \hfill (2.21)  

where the interference of user $k$ is subtracted from the interference incurred at base station $p$. Note that (2.21) can also be expressed in a similar form as in the CRE-based CS criterion in (2.13) such that

$$\text{RSRP}^{dB}_{k,c} < \text{RSRP}^{dB}_{k,p} + \text{Bias}^{dB}_p,$$  \hfill (2.22)  

where $\text{Bias}^{dB}_p$ is an adaptive offset that depends on the interference on the assigned subcarriers, PL, small-scale fading, and the consequent RSRP values. It can be expressed as

$$\text{Bias}^{dB}_p = 10 \log_{10} \left( \frac{p_{RS,p}}{p_{RS,c}} \frac{\sum_{m \in \mathcal{M}_k} \frac{|V_{k,c}(m)|^2}{I_{k,c}(m)}}{\sum_{m \in \mathcal{M}_k} \frac{|V_{k,p}(m)|^2}{I_{k,p}(m) - P_{k,m}H_{k,p}(m)}} \right).$$  \hfill (2.23)  

Note that in order to avoid frequent handover triggering and ping-pong effects on the cell boundaries, a small hysteresis margin can be added to the right-hand side of (2.22).
2.4.2 Implementation Steps and Required Changes

Before introducing the implementation steps of the proposed handover policy, we briefly summarize the radio resource control (RRC), mobility, CS, and handover protocols in LTE networks. First, the RRC protocol handles the Layer 3 control plane signalling by which the Evolved-Universal Terrestrial Radio Access Network (E-UTRAN) controls the UE behavior [37]. It has functions such as system information (SI) broadcasting, connection control including handover within LTE, network-controlled inter-RAT mobility, and measurement configuration and reporting.

The mobility functions of the RRC include UE measurement reporting and control, handover, and the control of cell association and reselection [34]. The RRC has two states as RRC-connected and RRC-idle. In RRC-connected mode, the UE is connected to an E-UTRAN, and the E-UTRAN knows the cell that UE is connected to. In this state, the UE can transmit and receive data from the network, report CQI and feedback information to its eNB, and carry out neighbor cell measurements. The user handovers are carried out by the E-UTRAN. These decisions are based on factors such as UE capability, subscriber type, and access restrictions. The E-UTRAN provides a list of neighboring frequencies and cells such that the UE can report its measurements for the cell association procedure. Two lists are maintained. The white-list of neighboring cell list (NCL) indicates the set of neighboring base stations that UE can connect to, while the black-list indicates those that UE cannot. The black-list of NCL typically includes closed-access groups such as home-deployed femtocells. Also, E-UTRAN is not required to indicate all the neighboring cells to the UE [34]. The RRC-idle state is UE-controlled and used during cell reselection. In this state, the UE cannot transmit or receive data since it does not have the RRC context stored in any eNB.

We focus on the RRC-connected mode since the proposed handover decision policy requires the UE to be connected to an eNB. The handover procedure in LTE is summarized in
Figure 2.1: Network signal diagram of a handover procedure. Proposed handover decision algorithm requires changes in Step 3. (MME stands for Mobility Management Entity.)

Fig. 2.1. It includes three main phases as handover preparation (steps 1-6), handover execution (steps 7-11), and handover completion (steps 12-18). The details of each step are discussed in [25, pp. 172-73]. The proposed handover decision policy requires changes in the handover decision (Step 3). The pseudocode in Algorithm 1 summarizes the decision criteria in detail. First, the user and base station association is carried out using the RSRP-based
Algorithm 1 Pseudocode for the Proposed Handover Decision Policy

1: Initialize: Use RSRP CS rule to obtain $c$
2: for each user $k \in \mathcal{K}$ do
3: Based on RSRP measurements, obtain $\mathcal{B}_q$
4: for each base station $c \in \mathcal{B}_q$ do
5: Calculate $\text{Bias}_{dB}^c$
6: end for
7: $c_k^* \leftarrow \max_{c \in \mathcal{B}_q} RSRP_{dB}^c + \text{Bias}_{dB}^c$
8: if $c_k \neq c_k^*$ then
9: Handover request, $c_k \leftarrow c_k^*$
10: end if
11: end for

CS protocol. Based on the provided white-list of NCL, the UE periodically feeds back the RSRP measurements to its eNB. We denote the set of base stations in the white-list of NCL as $\mathcal{B}_q$ where $q$ is an integer. For example, $\mathcal{B}_8$ includes eight neighboring eNBs. In the simulations, the white-list of NCL is obtained using the following procedure. First, the RSRP measurements at a UE are sorted in decreasing order. Then, the highest $q$ eNBs are selected in the white-list of NCL. For the sake of simplicity, let us consider the same $q$ for each user. We assume that $q = 8$ eNBs. Given the NCL, the UEs feed back the RSRP measurements to the eNBs. Then, using $I_{k,c}(m)$ and $RSRP_{dB}^{k,c}$ measurements, the RRC calculates $\text{Bias}_{dB}^c$ for all $c \in \mathcal{B}_q$ and selects the eNB that correspond to the largest sum of $RSRP_{dB}^c$ and $\text{Bias}_{dB}^c$. If a handover is required, then the handover steps are executed. The same process is repeated for each user.

2.4.3 Coverage Area of the Proposed Handover Policy

In conventional CS methods such as RSRP-, PL-, and CRE-based CS, the coverage area of a site only depends on channel conditions, and does not consider the current system load. However, with the proposed handover decision policy, the coverage area of base stations depends also on the interference incurred. Under heavy traffic, the coverage area with the proposed handover policy shrinks to offload extra traffic to the neighboring cells. On the
other hand, under light traffic load, the coverage area expands to provide coverage. Thereby, the proposed handover policy incorporates cell breathing to the next-generation wireless cellular systems by including the interference to the handover procedure.

2.4.4 Differences with Other Handover Policies

The proposed method is similar to the CRE-based CS method which is employed in standards. In both policies, the RSRP measurements and bias values are used to consider the heterogeneity of different base stations types in the CS procedure. The novelty that we introduce is to employ an adaptive bias value which provides better load balancing in a dense HetNet environment. In Section 2.5, we demonstrate the achievable throughput gains using this handover policy.

Also, the proposed policy in this chapter differs from the one in [17] in two aspects. First, the work in [17] considers a macrocell-femtocell layout in which macrocells operate at a single band and femtocells can select one of the three subbands. Only one of the subbands that femtocells operate overlap with the macrocell subband. Hence, the model in [17] allocates protected subbands for the femtocells. The proposed handover policy does not impose this constraint. Instead, we study the co-channel allocation among macrocells and picocells. When determining the adaptive cell range bias values, we use the interference measurements per RB at the base stations, and dynamically adjust the cell coverage accordingly. Therefore, our work considers a more efficient and realistic model considering LTE systems. The second remark is that we employ the standardized LTE uplink power control defined in [28], and use the above expressions to dynamically adjust the bias values for cell association. Reference [17] uses cell breathing for both power control and cell selection towards achieving power savings.
Figure 2.2: The simulation layout of a HetNet deployment is depicted. The layout includes a set of 19 hexagonal macrocells with each employing a 3-sector antenna. The setup is overlaid with 2 pico-eNBs per sector. The simulations are carried out for 10 active users per sector. Macrocell and picocell base stations are represented by squares and triangles, respectively, while users are denoted by circles.

2.5 Numerical Results

In this section, the performance of the proposed handover decision policy is studied. Its performance is compared with the RSRP-, PL-, and CRE-based CS methods in a HetNet deployment scenario. First, the number of associated users at each HetNet tier is presented in order to investigate the offloading ratios. Then, the cumulative distribution functions (c.d.f.) of user rates, aggregate throughput, and energy efficiency aspects of the uplink and downlink transmissions are studied. Finally, the distribution of adaptive CRE bias values obtained by the proposed policy is presented.
### Table 2.2: Simulation Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell layout</td>
<td>Hexagonal grid, 19 cells, 3-sectors per site</td>
</tr>
<tr>
<td>Center frequency</td>
<td>2 GHz</td>
</tr>
<tr>
<td>Channel bandwidth</td>
<td>10 MHz</td>
</tr>
<tr>
<td>Freq. selective channel model</td>
<td>ETU Channel Model</td>
</tr>
<tr>
<td>UE to MeNB PL model</td>
<td>$128.1 + 37.6 \log_{10}(d)$</td>
</tr>
<tr>
<td>UE to pico-eNB PL model</td>
<td>$140.7 + 36.7 \log_{10}(d)$</td>
</tr>
<tr>
<td>Inter-site distance</td>
<td>500 m</td>
</tr>
<tr>
<td>Minimum macro- to user distance</td>
<td>50 m</td>
</tr>
<tr>
<td>Minimum pico- to user distance</td>
<td>10 m</td>
</tr>
<tr>
<td>Minimum pico- to macro- distance</td>
<td>75 m</td>
</tr>
<tr>
<td>Minimum pico- to pico- distance</td>
<td>40 m</td>
</tr>
<tr>
<td>Total number of data RBs</td>
<td>50 RBs</td>
</tr>
<tr>
<td>MeNB Maximum Tx Power</td>
<td>46 dBm</td>
</tr>
<tr>
<td>Pico-eNB Maximum Tx Power</td>
<td>30 dBm</td>
</tr>
<tr>
<td>UE Maximum Tx Power</td>
<td>23 dBm</td>
</tr>
<tr>
<td>Effective thermal noise power</td>
<td>$-174 \text{ dBm/Hz}$</td>
</tr>
<tr>
<td>UE and eNB noise figures</td>
<td>9 and 5 dB, respectively</td>
</tr>
<tr>
<td>MeNB antenna gain</td>
<td>15 dBi</td>
</tr>
<tr>
<td>Pico-eNB antenna gain</td>
<td>5 dBi</td>
</tr>
<tr>
<td>UE antenna gain</td>
<td>0 dBi</td>
</tr>
<tr>
<td>Antenna horizontal pattern, $A(\theta)$</td>
<td>$-\min(12(\theta/\theta_{3dB})^2, A_m)$</td>
</tr>
<tr>
<td>$A_m$ and $\theta_{3dB}$</td>
<td>20 dB and 70°</td>
</tr>
<tr>
<td>Penetration loss</td>
<td>20 dB</td>
</tr>
<tr>
<td>Macrocell shadowing std. dev.</td>
<td>8 dB</td>
</tr>
<tr>
<td>Picocell shadowing std. dev.</td>
<td>10 dB</td>
</tr>
<tr>
<td>Uplink power control</td>
<td>$(P_0, \alpha) = (-60, 0.8)$</td>
</tr>
<tr>
<td>Traffic model</td>
<td>Full buffer</td>
</tr>
<tr>
<td>Scheduling algorithm</td>
<td>SRM scheduler</td>
</tr>
<tr>
<td>Link to system mapping</td>
<td>EESM Method</td>
</tr>
<tr>
<td>Macrocell Minimum Coupling Loss</td>
<td>$-70 \text{ dB [34]}$</td>
</tr>
<tr>
<td>Picocell Minimum Coupling Loss</td>
<td>$-45 \text{ dB [34]}$</td>
</tr>
</tbody>
</table>

### Table 2.3: Power Delay Profile of the ETU Channel Model [34]

<table>
<thead>
<tr>
<th>$P_{rel}(\tau_l)$ (dB)</th>
<th>-7</th>
<th>-5</th>
<th>-3</th>
<th>0</th>
<th>0</th>
<th>-1</th>
<th>-1</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_l$ (usec)</td>
<td>0</td>
<td>50</td>
<td>120</td>
<td>200</td>
<td>230</td>
<td>500</td>
<td>1600</td>
<td>2300</td>
<td>5000</td>
</tr>
</tbody>
</table>
Consider the LTE network depicted in Fig. 2.2. It includes a HetNet deployment of 19 macrocells and 2 picocells per sector. The macrocell base stations are employed with 3-sector antennas, whereas the picocell base stations are equipped with omnidirectional antennas. Ten users per sector is assumed. In order to observe the user clustering effects, two types of user distributions are considered in this chapter. In the first scenario, uniform user distribution is assumed such that users are randomly generated within each macrocell sector area. The second scenario is the non-uniform user distribution, which aims to observe the hotspot capacity enhancement with the picocell deployments [11]. It assumes that two users per picocell are randomly generated within the initial radius of each picocell, which is taken as 40 meters. The remaining users are randomly generated within each macrocell sector area. Although the users are initially dropped close to the picocells, they can still be associated with another base station depending on the CS method. The downlink reference signal powers are taken as 46 and 30 dBm for macro- and picocell base stations, respectively. It is assumed that all users are always active and have data to transmit. This model is known as the full buffer traffic model [11]. The simulation models and parameters are summarized in Table 2.2. These follow the standard models in [11] and [38] for the heterogeneous system simulation baseline parameters. The correlated shadow fading parameters are generated using the procedure described in Section 5.6 of [29]. Small scale fading is also considered in this work. Table 2.3 shows the power delay profile of the Extended Typical Urban (ETU) channel model. In order to exploit frequency and multiuser diversity, FDPS is employed. The resources are scheduled in the frequency domain with an RB granularity. For the uplink transmissions, our prior work in [26] discussed an adaptive bandwidth channel-aware sum rate maximization (SRM) FDPS. In this work, we employ this scheduler with equal bandwidths. The resources at each base station are equally divided to its associated users such that the aggregate rate is maximized. In the downlink, the same scheduler without the contiguity constraint is used. Power control in the downlink is not considered. It is assumed that a base station transmits at maximum power which is equally allocated among its subcarriers. In order to
avoid user outages, we employ inter-cell interference control (ICIC) methods. In the case of an outage, the transmit power of the strongest interfering base station or user is reduced by a certain amount. The outage users are typically located at cell boundaries where they are exposed to high interference from neighboring cells. By employing this simple ICIC method, the interference coupled to cell-edge users can be significantly reduced. These messages can be conveyed as the Relative Narrowband Transmit Power (RNTP) indicators in the downlink, which are defined per RBs in LTE Release 8 for downlink transmissions, and as the Overload Indicator (OI) message in the uplink [25, pp. 204-207]. These signals can be signaled through the X2 interface in between base stations. We assumed that in the case of an outage, the transmit power of the aggressor is reduced by 3 and 6 dB in uplink and downlink, respectively. Also, in our simulations, the wrap-around technique is employed to mitigate edge effects [11,12], and the simulations are repeated for 10 independent runs.

Tables 2.4-2.5 present the uplink and downlink simulation results with the proposed policy and conventional CS methods for the uniform user distribution scenario, and Tables 2.6-2.7 show the results for the non-uniform user distribution case. The throughput performance of a single-layer macrocell system is also included as a baseline for comparison. In order to observe the throughput of cell-edge, median, and cell-center users, 10th, 50th, and 90th percentiles are shown, respectively. First, it can be observed that the deployment of picocells brings significant throughput increases for both user distributions. When two picocells per sector are deployed, the RSRP-based CS method provides 1.9x to 3.6x increase in aggregate uplink sector throughput compared to the baseline depending on the user distribution. The throughput of CRE- and PL-based CS methods are better than RSRP-based CS due to better load balancing among tiers. The proposed handover policy provides the highest throughput compared to other CS methods. For example, with the uniform user distribution, the proposed policy increases the average uplink sum throughput per sector of baseline by 143.2%, of RSRP-based CS method by 25.7%, and of PL-based CS method by 4.9%, as shown in Table 2.4. Similar observations can be made for the non-uniform user distribution
### Table 2.4: Uplink Results for a HetNet Deployment of 2 Picocells per Sector and Uniform User Distribution

<table>
<thead>
<tr>
<th>CS Method</th>
<th>Percentile Throughput (Mbits/sec)</th>
<th>Macrocell Assoc. Sector Throughput (Mbits/sec)</th>
<th>Picocell Assoc. Sector Throughput (Mbits/sec)</th>
<th>Aggregate Sector Throughput (Mbits/sec)</th>
<th>Average Energy Efficiency (Mbits/Joule)</th>
<th>Percentage of Picocell Users (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10th 50th 90th</td>
<td>Throughput Assoc. Sector Throughput</td>
<td>Throughput Assoc. Sector Throughput</td>
<td>Aggregate Sector Throughput</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.16 0.96 2.80</td>
<td>12.79</td>
<td>–</td>
<td>12.79</td>
<td>1.38</td>
<td>–</td>
</tr>
<tr>
<td>RSRP-based CS</td>
<td>0.20 1.18 4.59</td>
<td>12.30</td>
<td>12.44</td>
<td>24.74</td>
<td>2.68</td>
<td>8.9</td>
</tr>
<tr>
<td>CRE-based CS, 3 dB Bias</td>
<td>0.20 1.19 5.16</td>
<td>11.99</td>
<td>14.65</td>
<td>26.65</td>
<td>2.89</td>
<td>12.3</td>
</tr>
<tr>
<td>CRE-based CS, 6 dB Bias</td>
<td>0.20 1.24 6.08</td>
<td>11.68</td>
<td>15.82</td>
<td>27.50</td>
<td>2.98</td>
<td>16.4</td>
</tr>
<tr>
<td>CRE-based CS, 9 dB Bias</td>
<td>0.24 1.33 6.70</td>
<td>11.56</td>
<td>17.14</td>
<td>28.70</td>
<td>3.10</td>
<td>21.4</td>
</tr>
<tr>
<td>CRE-based CS, 12 dB Bias</td>
<td>0.24 1.49 6.87</td>
<td>11.09</td>
<td>18.40</td>
<td>29.49</td>
<td>3.18</td>
<td>27.4</td>
</tr>
<tr>
<td>CRE-based CS, 15 dB Bias</td>
<td>0.24 1.61 7.26</td>
<td>10.71</td>
<td>19.07</td>
<td>29.78</td>
<td>3.21</td>
<td>33.5</td>
</tr>
<tr>
<td>PL-based CS</td>
<td>0.24 1.72 6.87</td>
<td>10.48</td>
<td>19.19</td>
<td>29.66</td>
<td>3.19</td>
<td>36.1</td>
</tr>
<tr>
<td>Proposed Handover Policy</td>
<td>0.24 1.62 7.26</td>
<td>9.43</td>
<td>21.69</td>
<td>31.11</td>
<td>3.35</td>
<td>40.6</td>
</tr>
</tbody>
</table>

### Table 2.5: Downlink Results for a HetNet Deployment of 2 Picocells per Sector and Uniform User Distribution

<table>
<thead>
<tr>
<th>CS Method</th>
<th>Percentile Throughput (Mbits/sec)</th>
<th>Macrocell Assoc. Sector Throughput (Mbits/sec)</th>
<th>Picocell Assoc. Sector Throughput (Mbits/sec)</th>
<th>Aggregate Sector Throughput (Mbits/sec)</th>
<th>Average Energy Efficiency (Mbits/Joule)</th>
<th>Percentage of Picocell Users (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10th 50th 90th</td>
<td>Throughput Assoc. Sector Throughput</td>
<td>Throughput Assoc. Sector Throughput</td>
<td>Aggregate Sector Throughput</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.16 1.29 4.05</td>
<td>17.41</td>
<td>–</td>
<td>17.41</td>
<td>0.09</td>
<td>–</td>
</tr>
<tr>
<td>RSRP-based CS</td>
<td>0.20 1.64 6.27</td>
<td>17.82</td>
<td>18.35</td>
<td>36.17</td>
<td>0.18</td>
<td>8.9</td>
</tr>
<tr>
<td>CRE-based CS, 3 dB Bias</td>
<td>0.23 1.76 6.82</td>
<td>17.86</td>
<td>20.35</td>
<td>38.21</td>
<td>0.19</td>
<td>12.3</td>
</tr>
<tr>
<td>CRE-based CS, 6 dB Bias</td>
<td>0.26 1.95 7.47</td>
<td>18.14</td>
<td>21.78</td>
<td>39.92</td>
<td>0.19</td>
<td>16.4</td>
</tr>
<tr>
<td>CRE-based CS, 9 dB Bias</td>
<td>0.30 2.11 8.15</td>
<td>18.23</td>
<td>22.38</td>
<td>40.61</td>
<td>0.20</td>
<td>21.4</td>
</tr>
<tr>
<td>CRE-based CS, 12 dB Bias</td>
<td>0.31 2.25 8.73</td>
<td>18.68</td>
<td>23.07</td>
<td>41.75</td>
<td>0.20</td>
<td>27.4</td>
</tr>
<tr>
<td>CRE-based CS, 15 dB Bias</td>
<td>0.31 2.43 8.96</td>
<td>18.95</td>
<td>22.84</td>
<td>41.79</td>
<td>0.20</td>
<td>33.5</td>
</tr>
<tr>
<td>PL-based CS</td>
<td>0.32 2.54 9.09</td>
<td>19.19</td>
<td>22.89</td>
<td>42.08</td>
<td>0.20</td>
<td>36.1</td>
</tr>
<tr>
<td>Proposed Handover Policy</td>
<td>0.26 2.28 9.45</td>
<td>18.60</td>
<td>27.04</td>
<td>45.64</td>
<td>0.23</td>
<td>40.6</td>
</tr>
</tbody>
</table>

The average uplink throughput per sector of the proposed policy is 49.54 Mbps, while the RSRP- and PL-based CS achieves 41.68 Mbps and 48.14 Mbps, respectively, in non-uniform user distribution. We observe that the proposed policy increases all of the cell-edge, median, and cell-center user rates for the uniform user distribution case. However, for the nonuniform user distribution, the cell-center user rates are significantly higher at the cost of slight throughput decreases for the cell-edge and median user rates. In the downlink, the gains through better cell associations with the proposed policy can also be observed. The aggregate sector throughput of the proposed policy is 69.57 Mbps, while the throughput of the baseline, RSRP-, and PL-based CS methods are 16.51, 63.77, and 65.07 Mbps, respectively, for the non-uniform user distribution scenario. Note that these gains are achieved through only better cell associations and they can be further improved by...
employing the inter-cell coordination with joint RB scheduling techniques [13]. Our results demonstrate that, although the highest channel gains are always selected with the PL-based CS method, it does not necessarily provide the highest uplink rates. Furthermore, the PL-based CS method can yield lower throughput in the downlink compared to the CRE-based CS schemes due to the power difference between macrocell and picocell tiers, which we will study next.

The uplink-downlink imbalance discussed in Section 2.1 can be observed for the non-uniform user distribution. The aggregate uplink throughput increases as the CRE bias increases, whereas this is not always true in the downlink. For example, the average downlink sum throughput increases from 63.77 Mbps to 66.77 Mbps when RSRP-based CS and 6 dB CRE
bias are employed, respectively. As the CRE bias increases beyond this value, the average downlink sum rate decreases, for example, to 65.47 Mbps at 15 dB CRE bias. Notice that the proposed handover policy achieves the highest average downlink sum throughput per sector with 69.57 Mbps. These gains are due to managing the interference better compared to other methods. Unlike the non-uniform user distribution, the uplink-downlink imbalance is not observed for the uniform user distribution. The reason for this is as the CRE bias increases, the users in the extension region are subject to high macrocell interference in the downlink. When users are generated in the close proximity of picocells as in the non-uniform distribution, downlink throughput of a picocell associated user can be significantly improved with low or moderate CRE bias values. However, this is not true for the uniform user distribution. In this case, low or moderate CRE bias values do not provide sufficient macrocell offloading nor base station diversity to outperform the downlink performance of the higher CRE bias values.

The deployment of picocells significantly increases the energy efficiency of the transmissions in both directions as shown in Tables 2.4-2.7. For instance, HetNet deployments provide 1.9x to 2.4x energy efficient uplink transmissions with uniform user distribution, while these gains are around 3.8x to 4.6x with the non-uniform case. The reason for the significant gain with the non-uniform user distribution is due the clustering effect around hotspots and the consequent reduced link distances. In the downlink, it can also be observed that the energy efficiency is significantly improved with the deployment of small cell base stations. The reason that energy efficiency of downlink transmissions is less than that of the uplink in our results is that downlink power control is not implemented. We should also note that only the transmit energy is considered in the energy efficiency calculations. The static energy consumptions of macro- and picocells are not taken into account.

Figure 2.3 illustrates the cell association characteristics of the proposed handover policy and conventional CS methods. Since an MeNB reference signal transmit power is larger
Figure 2.3: Percentage of associated users at macrocell and picocell tiers using the proposed handover policy and conventional CS methods. More users can be offloaded to the picocell tier by increasing the CRE bias. The proposed handover policy achieves the most balanced loading between the macrocell and picocell tiers.

Figure 2.4: Distribution of the power difference between the RSRP measurements of the closest macrocell and the serving picocell for the picocell associated users with the proposed uplink interference-based handover policy.
Figure 2.5: Distribution of PL between the user and its serving base station for the proposed uplink interference-based handover policy and conventional CS schemes. The PL-based CS method provides the cell association with the minimum PL values, whereas the RSRP-based CS yields the one with larger PL values. The proposed handover policy and CRE-based CS methods offer gains within these two ends.

compared to that of a pico-eNB and the channel conditions of the picocell environment are more severe, macrocells have larger coverage area than picocells. Consequently, there are more macrocell associated users. Due to the clustering around hotspots, more users can be offloaded from the macrocells with the non-uniform user distribution compared to the uniform case. For example, with the RSRP-based CS method, 23.4% of users can be offloaded from macrocell tier to picocell tier in the non-uniform user distribution, while only 8.9% of users can be offloaded in the uniform user distribution. As the CRE bias increases, more users are associated with picocells. This creates a more balanced loading among the tiers in a network. When PL-based CS is considered, we observe that 36.1% and 54.5% of users are offloaded from MeNBs to picocells in the uniform and non-uniform user distributions, respectively. The proposed handover policy offers high macrocell offloading as well. We observe that 40.6% and 49.5% of users can be offloaded from the macrocell layer with the proposed handover policy in the uniform and non-uniform cases, respectively.

As discussed in Section 2.4, we propose to use an adaptive CRE bias in order to dynamically balance the user load per tier. In order to show the distribution of the adaptive CRE bias
values, the power difference between the RSRP measurements of the closest macrocell and the serving picocell for each user is plotted in Fig. 2.4 for the non-uniform user distribution case. Similar results are obtained for the uniform user distribution scenario. In Fig. 2.4, the power differences that are greater than zero correspond to the picocell associated users in the extension area, while those less than zero are the users within the boundaries defined by the RSRP. We observe that 52.7% of the picocell users are located in the CRE areas and 75% of the CRE bias values are located between $-10$ and $20$ dB. This demonstrates the cell breathing capabilities of the proposed policy such that the coverage region of a picocell expands or shrinks, or equivalently, it takes positive and negative CRE values, depending on the uplink interference. In order to give the reader more intuition on Fig. 2.4, the same figure would be upper bounded at a constant bias value if the CRE-based CS method was employed. Similarly, it would be upper bounded at zero if the RSRP-based CS method was used. Hence, with the employment of dynamic CRE bias, the proposed handover policy is able to dynamically adjust the picocell coverage areas and offload significant number of macrocell users to the picocell tier, which increases the throughput performance in both the uplink and downlink.

In Fig. 2.5, the PL between the serving base stations and users in the non-uniform user distribution scenario are shown. Note that these values also include the shadow fading and antenna gains as well. The deployment of the picocell tier significantly decreases the PL due to the increase in base station diversity and reduced link distances. It can be observed that the PL of the median user decreases by $4-8$ dB depending on the CS method when 2 picocells per sector are deployed. These gains still exist even for the cell-center users. For example, the PL of the proposed handover policy and PL-based CS are 100.2 and 99.0 dB for the 10th percentile users, respectively, while it is 106.9 dB for the single-layer baseline system. Note that the PL-based CS method always provides the cell association with the lowest PL values.

The proposed handover policy does not necessarily associate the users to the closest base stations, even when the users are dropped to the close vicinity of the picocells. Instead, it
Figure 2.6: Illustration of cell associations for the users in the first sector of the center cell. The figures are for the single layer (a) and two-layer systems (b)-(d). The cell associations with RSRP- and PL-based CS methods are depicted in (b) and (c), respectively, while the cell association of the proposed policy is shown in (d).

takes the interference into consideration during the cell association procedure. This not only balances the loading, but also improves both the uplink and downlink throughput results.

Finally, we investigate the performance of the proposed policy under imbalanced interference conditions. In order to do so, we generated different number of users in the center cell. In Fig. 2.6, we depict the cell association of users with different CS methods. To provide neater figures, we present a smaller network. Initially, 4 users per sector are randomly generated. Then, the number of users in one of the sectors in the center cell are doubled to 8. The newly generated users are shown by filled circles, whereas the initial users are shown with empty circles. This type of user generation creates an imbalance in traffic and interference.
to the neighboring base stations. Before presenting the throughput results, we will use this example to illustrate some discussions on the cell coverage. In Sections 2.1 and 2.3, we mentioned that a user can be associated with a macrocell with higher transmit power, although its PL to a nearby picocell is less. For example, UEs 6 and 8 in Fig. 2.6(b) are connected to a MeNB with a higher RS transmit power using the RSRP-based CS method, while they are connected to a pico-eNB using the PL-based CS in Fig. 2.6(c). Although we use the hexagonal regions to generate users and picocells, due to the random nature of shadow fading, the users can be associated with a nearby base station that is located in another hexagonal area. UEs 1, 5, and 6, that are located around the cell borders, are such examples. For instance, UE 6 is connected to a neighboring macrocell with the RSRP-based CS method as depicted in Fig. 2.6(a)-(b). In the proposed policy, users are initially associated with base stations based on the RSRP-based CS rule. Then, calculating the adaptive CRE bias values, new cell associations are obtained. We observed that the new cell associations are similar to the PL-based CS criteria except for UEs 2 and 7. Both of these users are connected to macrocells with the PL-base CS rule, whereas they are offloaded to the picocells with the proposed policy. As a result of this new cell association and macrocell offloading, the throughput of UE 2 increases from 1.58 Mbps to 2.53 Mbps, while UE 7 gets the same rate 0.31 Mbps in both cases. Note that although the PL of UE 2 to the MeNB and pico-eNB are 108 and 111 dB, respectively, UE 2 is associated with the pico-eNB which has a higher CRE bias of 21.4 dB, compared to the CRE bias of 5.7 dB with the MeNB. Before introducing new users to the system, the average uplink sum rates in the network are 13.2, 22.4, 27.4, and 29.2 Mbps with the baseline, RSRP- and PL-based CS methods, and the proposed policy, respectively. When the new users are introduced to the system, they changed to 12.7, 22.8, 27.9, and 29.8 Mbps, in the same order as before. Notice that in the baseline, the uplink throughput decreases due to the fact that the same number of resources are now shared by more users. This indicates the lack of base station diversity in the baseline system. For the HetNet architectures, the deployment of picocells increases the base station
diversity such that the aggregate sector uplink throughput also increases for each scheme. Due to better cell associations, the proposed handover policy again outperforms the other CS methods.

2.6 Conclusion

The deployment of heterogeneous base stations provides substantial gains on the cellular network performance in terms of increased data rates, improvements in cell coverage, and significantly reduced user outages. In order to fully utilize the benefits of heterogeneous base station deployments, a different approach in network planning than conventional single-layer network architectures needs to be pursued. For this purpose, we identified the critical roles of the CS criteria and handover policies. The current state-of-the-art methods such as CRE-based CS heavily rely upon offline simulations or field measurements. They are not self-adaptive or self-optimizing to the traffic and user distributions. The proposed uplink interference-based handover policy does not have these disadvantages. It dynamically adjusts the cell coverage and associates users to base stations based on the instantaneous statistics. It is distributed and can be employed in the standardized LTE handover procedure with limited information exchange between base stations. The proposed policy and conventional CS methods are tested in both uniform and non-uniform user distributions within a cell. Based on our simulation results, we conclude that in both scenarios, the proposed uplink interference-based handover policy outperforms all conventional CS methods, including the CRE- and PL-based CS criteria, in terms of both uplink and downlink throughput due to better load balancing between tiers in HetNet architectures.
Chapter 3

Efficiency and Fairness Trade-Offs in Schedulers

3.1 Motivation

The global mobile data traffic grew 70 percent in 2012 [9], and this trend is expected to continue in the next decade. These data rate demands are not only challenging to achieve, but also raise concerns on power consumption of next-generation wireless systems. In this chapter, we address the concerns on efficiency, fairness, and power consumption for different schedulers. We discuss these in a Long-Term Evolution (LTE) system. LTE systems employ single carrier frequency domain multiple access (SC-FDMA) transmissions in the uplink transmissions to achieve high power-efficiency, improved coverage, and reduced power consumption at user equipments (UEs) [25]. This transmission scheme also provides a smaller peak-to-average power ratio (PAPR) compared to the orthogonal frequency division multiplexing access (OFDMA) that is used in the downlink, and thereby achieves significant cost and battery life savings for UEs.
In terms of resource management, there are four main algorithms in LTE. These are admission control, packet scheduling, power control, and interference control [25]. In this chapter, we focus on the packet scheduling and power control algorithms. Let us note that LTE systems are designed to offer large flexibility in packet scheduling in time, frequency, and spatial domains. Through control channel signals, base stations schedule each transmission, so that the users are allocated orthogonal resources without any overlap. In order to keep the required signaling overhead manageable, especially for the uplink transmissions, the subcarriers are scheduled in groups in the frequency domain. Each subcarrier that occupies a bandwidth of 15 kHz is grouped in 12 subcarriers. These 12 subcarriers are called as a resource block (RB). Resource scheduling is then carried out with an allocation granularity of 180 kHz in the frequency domain [25, p. 78].

Frequency domain packet scheduling (FDPS) is a method used in LTE systems to allocate radio resources in the frequency domain such that the capacity of the system and the user experience are improved under some quality of service (QoS) and fairness constraints. When resource allocation for multiple users is considered, each user experiences different channel conditions and has different QoS requirements. FDPS allows assigning the resources such that the system performance can be optimized for different QoS requirements and efficiency versus fairness trade-offs. Typically, RBs are allocated to the users with the highest channel gains. Contrary to the best channel assignment strategy in OFDMA systems, only consecutive resources are assigned to users in SC-FDMA systems.

The works in [39, 40] investigate resource allocation problems in air traffic, health care, organ donor, and call center scheduling applications. In all these areas, utilities are shared among multiple users, and naturally, trade-offs between efficiency and fairness arise. The price of fairness is a metric introduced in [39, 40] to quantify the losses incurred to achieve fairness in the system. This metric can be used by the central decision maker to characterize and balance these trade-offs. Same observations can be made in scheduling and resource
allocation in cellular radio networks. Especially, these trade-offs characterize the cell-edge user performance. In cases where the cell-center users are favored to increase the aggregate cell throughput, this may lead to cell-edge user rate starvation cases. In this chapter, we use the price of fairness to characterize these trade-offs in an analytical framework.

Scheduling granularity per RB and contiguous resource assignment constraints provide significant benefits to reduce the control signal overhead and the search space of the SC-FDMA resource allocation problem. However, there are implementation constraints to be considered. In particular, we focus on the discrete Fourier transform (DFT) implementation constraints. The radices of DFT operations are typically chosen as prime factors [41]. In the early phases of the LTE standardization process, it was proposed to limit the DFT sizes to radices of 2, 3, and 5 [42].

### 3.1.1 Related Works

Related works in literature in channel dependent scheduling in SC-FDMA systems include [43–49]. It is shown in [43] that the FDPS problem in SC-FDMA systems with the contiguity constraint is NP-hard. The authors of [43] considered maximizing the proportional fair (PF) metric of users and presented four suboptimal algorithms with RB granularity. The same problem is also addressed in [44] on a per subcarrier basis where the authors presented two algorithms with different complexities. The first algorithm achieved the optimal solution with relatively high computational complexity, and the second yielded a suboptimal solution. The works in [45] and [46] proposed suboptimal greedy heuristic algorithms to maximize utility functions based on channel capacity and PF metrics, respectively. However, they did not consider the contiguity constraint. Another related work in [47] proposed three suboptimal algorithms based on the PF metric with RB granularity in varying levels of complexity. The efficiency and fairness of these schedulers are further studied in [48, 49]. In particular,
the authors in [49] compared the performance of the heuristic algorithms presented in [43] and [47] to solve the sum-rate maximization (SRM) and PF scheduling problems. Although the work in [49] also identified the efficiency and fairness trade-offs in these schedulers, it did not quantify the efficiency losses or provide bounds for these trade-offs. We need to note that, unlike the heuristic solutions in [43] and [47–49], the work in this chapter satisfies the optimality conditions, albeit with some increased complexity. Furthermore, the work in [49] considers a single-cell simulation setup whereas in this chapter, we consider a more realistic multi-cell multiuser scenario that is proposed in the standards [11,29].

The FDPS problem is also addressed in LTE downlink systems. The work in [50] presents linear programming (LP) solutions for various schedulers. The authors of [50] assume that the signal-to-noise-ratio (SNR) to the channel quality indicator (CQI) mapping for a fixed block error ratio (BLER) is linear. The study in [50] shows that the throughput can be estimated based on the uplink CQI feedback reports, and these estimates can be employed in the scheduling problem formulation. The constraint matrices of the LP problems in [50] are similar to those of [44] except for the contiguity constraint. Also, in [51] and [52], the authors present a cross-layer optimization framework for the utility-based scheduling problem in OFDM networks. The analysis in [51] considers an infinite number of subcarriers, whereas a finite number of subcarriers and more realistic conditions are investigated in [52]. The two-part paper [51,52] studies the necessary and sufficient conditions for the dynamic subcarrier assignment (DSA), adaptive power allocation (APA), and their joint allocation schemes. The results in [51] identify that DSA offers a significant improvement over the fixed subcarrier assignment (FSA) scheme, while APA provides a limited improvement over FSA. The joint DSA and APA allocation has a marginal gain over the DSA scheme. However, in [52], with a finite number of subcarriers, the improvements of APA are more significant compared to the DSA scheme. This time, their joint allocation offers a substantial gain over both schemes. Since APA within RBs is not standardized in LTE systems, in this chapter, we will only consider equal power allocation for the subcarriers in RBs. Similar to [52], we
will also propose a DSA method in order to improve the sum utility of the users. In [53], the authors investigate the effects of finite and full buffer traffic models using the utility maximization framework in OFDMA networks. The study in [53] uses the same family of utility functions as the one we use in this chapter. The results in [53] show that as the parameter characterizing the family of utility functions, $\alpha$, decreases, the cell-edge user rates increase, and the authors propose to use a specific value of $\alpha = 0.6$ as the best trade-off point. As we will show in the sequel, as $\alpha$ increases, the fairness of user rates improves, and the fairness in wireless networks determines the cell-edge user rates [12]. The study in [53] also presents the gains of an $\alpha$-fair scheduler over PF scheduler at different network loads. In this chapter, along with a numerical study, we provide the upper bounds for these gains such that the network operator (as the decision maker) can make an informed decision on the efficiency and fairness trade-off before its implementation.

### 3.1.2 Contributions

In this chapter, we study the FDPS problem in an SC-FDMA system. We follow the optimal solution framework in [44], but employ it with RB granularity in the frequency domain to reduce computational complexity by two orders of magnitude. Our solution approach involves a utility-based resource allocation scheme such that we can exploit the multiuser diversity. We formulate the FDPS problem as a set partitioning problem which can be solved by branch-and-cut methods. We investigate a family of utility functions called as the $\alpha$-fair utility function. In particular, we study the SRM, PF, and max-min fair (MMF) utilities. We consider the functions of user capacity as the utility to be maximized, and define the system efficiency as the sum of user rates. By using these utility functions, we solve the FDPS problem. We identify the optimality conditions, and present the size of the solution spaces and computational complexities of these schedulers. Moreover, we propose to use the price of fairness as a comparison metric between schedulers. The price of fairness of a
Figure 3.1: The trade-off between efficiency and fairness is depicted for a family of utility functions.

scheduler defines the amount of aggregate loss of a scheduler when compared to the SRM scheduler. Along with a fairness index (such as Jain’s fairness index [54]), the decision maker can acquire how much system efficiency loss is incurred while improving fairness by a certain amount. To the best of our knowledge, this metric has not been explored in a multiuser cellular radio environment. We depict the trade-offs between efficiency and fairness for $\alpha$-fair schedulers in Fig. 3.1. The SRM scheduler achieves the highest system efficiency of 1, i.e., the price of fairness of SRM scheduler is 0, but it lacks fairness among the users. When fairness is introduced to the system, the system efficiency decreases, or equivalently, the price of fairness increases. At the highest fairness, MMF scheduler maximizes the minimum user rate. Unfortunately, the $(1, 0)$ point on Fig. 3.1 cannot be achieved since the most efficient solution lacks fairness, and the most fair solution lacks efficiency. Also, note that these points correspond to the optimal scheduler solutions that are Pareto optimal which means that there is no user whose rate can be increased without decreasing the rates of other
users. With the use of price of fairness metric, this chapter answers the following questions: How do the cell-edge users get affected when the scheduler tries to maximize the sum-rate per cell? How can we improve fairness among users, especially between cell-center and cell-edge users? And, how much aggregate rate loss is incurred while introducing fairness? We present upper bounds on the aggregate cell rate loss for the PF scheduler. We describe how to include DFT constraints to construct RB assignment patterns and how to form the constraint matrix. The optimality conditions of the investigated problems are also presented. The analytical framework is supported with the numerical results of these FDPS problems. We derive conclusions based on the trade-offs between different types of schedulers.

The remainder of this chapter is organized as follows. Section 3.2 introduces the system model and discusses the uplink power control in LTE systems. Section 3.3 presents the utility-based resource allocation framework and discusses $\alpha$-fair utility functions. Also, we define and provide the upper bounds on the price of fairness in this section to quantify the scheduler efficiencies. In Section 3.4, we study and formulate the scheduling problems. We identify the necessary optimality conditions for the $\alpha$-fair schedulers. An effective solution method is also presented. The implementation constraints and exact dimensions of the search spaces are also investigated. Sections 3.5-3.6 discuss the results of proposed FDPS problems in a realistic simulation environment along with some concluding remarks.

### 3.2 System Model

In this section, SC-FDMA uplink system model and the uplink power control mechanism in LTE standards are described. First, we start with introducing the notations used in this chapter. The vectors and matrices are represented by boldface characters, e.g., $\mathbf{x}, \mathbf{A}$. The dimensions of the vectors are shown by subscripts, e.g., $\mathbf{x}_N$ is an $N \times 1$ column vector. Vectors in a set are enumerated by superscripts, e.g., $\mathbf{x}_N^1, \ldots, \mathbf{x}_N^K$. We use $\mathbf{1}_N$ and $\mathbf{0}_N$ to
denote all ones and all zeros column vectors, respectively. The transpose of a column vector \( \mathbf{x} \) is given by \( \mathbf{x}^T \). The sets are shown by capital calligraphic font, e.g., the set \( \mathcal{N} \) includes \( \mathcal{N} = \{1, \ldots, N\} \). A subset of set \( \mathcal{N} \) is denoted by subscripting it as \( \mathcal{N}_k \), and \( |\mathcal{N}_k| \) denotes the cardinality of the subset. The total number of ones in vector \( \mathbf{x} \) is denoted by \( w(\mathbf{x}) \) where \( \mathbf{x} \) consists of binary elements 0 and 1. Finally, any power value \( P \) is represented in dB with the notation \( P_{dB} \).

Assume that the system bandwidth consists of \( M \) orthogonal subcarriers. These subcarriers can be grouped into clusters to form \( N \) non-overlapping RBs. For example, according to LTE specifications, a system with 10 MHz bandwidth occupies \( M = 600 \) subcarriers, or equivalently, \( N = 50 \) RBs, and each RB consists of \( N_{sc}^{RB} = 12 \) consecutive subcarriers [25].

Let the total set of subcarriers and RBs be denoted by \( \mathcal{M} \) and \( \mathcal{N} \), respectively. Then, \( \mathcal{M}_k \) and \( \mathcal{N}_k \) represent the set of subcarriers and RBs assigned to user \( k \), respectively. Also, the complete set of users in the system is represented as \( \mathcal{K} = \{1, \ldots, K\} \), and those associated with base station \( c \) are given by the subset \( \mathcal{K}_c \). The cardinality of this subset, \( K_c = |\mathcal{K}_c| \), denotes the total number of users associated with base station \( c \). When all the RBs in the system are scheduled, \( N = |\mathcal{N}| = M/N_{sc}^{RB} = |\mathcal{N}_1| + \cdots + |\mathcal{N}_{K_c}| = N_1 + \cdots + N_{K_c} \) holds true. Furthermore, let \( \mathcal{K}_m \) denote the set of users that are assigned to subcarrier \( m \), and note that the set of users in \( \mathcal{K}_m \) are located in different cells in a multi-cell multiuser scenario.

The transmitter structure of an uplink SC-FDMA system is depicted in Fig. 3.2. It includes DFT, subcarrier mapping, and inverse fast Fourier Transform (IFFT) operations. In LTE systems, the orthogonal resource assignment over the time, frequency and spatial domains.
helps to avoid the near-far problem that existed in Wideband Code Division Multiple Access (WCDMA). For this reason, power control in LTE is carried out at a slower rate [25]. For a given user-base station assignment, the fractional open-loop power control is expressed as [28]

\[ P_{dB}^k = \min\{P_{dB}^{max}, P_0^{dB} + 10 \log_{10}(N_k) + \beta PL_{dB}^{k,c_k}\}, \quad (3.1) \]

where \( P_{dB}^k \) denotes the uplink transmit power of user \( k \), and \( P_{dB}^{max} \) denotes the maximum UE transmit power. \( P_0^{dB} \) is the open loop transmit power. \( PL_{dB}^{k,c_k} \) is the path loss between user \( k \) and its serving base station \( c_k \). We consider that the path loss includes the shadow fading of the link. Since shadow fading is a slow variation process for pedestrian speeds, we assume it to be static over a frame duration [11]. The path loss compensation factor \( \beta \) takes its value from the set \( \{0,0.4,0.5,0.6,0.7,0.8,0.9,1\} \). It determines the fairness within the cell such that the network enables cell-edge users with high path loss values to transmit at high power levels for \( \beta = 1 \). The fairness in the system decreases as the path loss compensation approaches zero, i.e., \( \beta \to 0 \), since the high path losses for cell-edge users are not compensated for. However, this improves the rates for the cell-center and median users due to reduced intercell interference compared to the full path loss compensation case. The UE transmission power is equally distributed on the allocated bandwidth such that the UE transmit power per subcarrier is \( P_{k,m} = P_k/M_k \) in linear scale.

Furthermore, in order to investigate the performance of the frequency-domain scheduler, frequency selective fading channel models are considered. For example, [29] provides power delay profile models for pedestrian and vehicular users. We denote the channel gain between user \( k \) and base station \( c \) on subcarrier \( m \) by \( H_{k,c}(m) \). The channel gain has two components, consisting of multipath and path loss plus shadowing components. They can be expressed
as

\[
H_{k,c}(m) = \frac{|V_{k,c}(m)|^2}{P L_{k,c}}, \quad V_{k,c}(m) = \sum_{l=1}^{L} v_k(l) \exp \left( -\frac{j2\pi m \tau_l}{T_s} \right), \tag{3.2}
\]

where \( V_{k,c}(m) \) includes the multipath effects in the received signal for an \( L \)-path impulse response model. \( T_s \) and \( \tau_l \) denote the sampling time and the delay of \( l \)-th path, respectively. Each multipath component \( v_k(l) \) can be further expressed as

\[
v_k(l) = \begin{cases} A \sqrt{P_{\text{rel}}(l)} w_k(l) & \text{if } l = \lfloor \tau_l/T_s \rfloor + 1 \\
0 & \text{otherwise}, \end{cases} \tag{3.3}
\]

where \( w_k(l) \) denotes a zero-mean Gaussian noise process. \( P_{\text{rel}}(l) \) denotes the relative power of the \( l \)-th path. The term \( A \) normalizes the average multipath power to unity such that \( E[\sum |v_k(l)|^2] = 1 \). The sampling time \( T_s \) and the sampling frequency \( f_s \) in LTE systems depend on the system bandwidth. The sampling frequency is given by \( f_s = 1/T_s = B_{sc} \times M \), where the total number of subcarriers \( M \) is also the IFFT size and the bandwidth of each subcarrier is fixed to \( B_{sc} = 15 \) kHz. For example, in a 10 MHz bandwidth, there are 600 subcarriers and \( M = 1024 \). Then, the sampling frequency becomes \( f_s = 15.36 \) MHz [27, p. 70].

The SNR on each subcarrier \( m \) assigned to user \( k \) is given by

\[
\gamma_{k,m} = \frac{P_{k,m} |H_{k,c}(m)|^2}{\sigma_c^2} = \frac{P_{k,m} |V_{k,c}(m)|^2}{P L_{k,c} \sigma_c^2}, \tag{3.4}
\]

where \( \sigma_c^2 \) is the thermal noise effective on a subcarrier at base station \( c \). Similarly, the signal-to-interference-plus-noise ratio (SINR) of user \( k \) on subcarrier \( m \) is given by

\[
\Gamma_{k,m} = \frac{P_{k,m} H_{k,c}(m)}{\sum_{j \in \mathcal{K}_m, j \neq k} P_{j,m} H_{j,c}(m) + \sigma_c^2}. \tag{3.5}
\]

We assume that ideal channel estimates for each link are available. Typically, frequency
domain minimum mean square error (MMSE) equalizers are employed at the base station receivers. The wideband SNR for user $k$, $\gamma_k$, can be calculated using the individual SNR of each subcarrier $m$ assigned to user $k$, $\gamma_{k,m}$ as [27,55]

$$\gamma_k = \left( \frac{1}{\sum_{m \in M_k} \frac{\gamma_{k,m}}{\gamma_{k,m} + 1}} + 1 \right)^{-1}. \quad (3.6)$$

Similarly, the wideband SINR of user $k$, $\Gamma_k$, can be defined by replacing $\gamma_{k,m}$ with $\Gamma_{k,m}$ in (3.6). The channel capacity of user $k$ defines the maximum reliable communication rate over a channel. However, in practical communication systems, one needs to account for other factors (e.g., reference signals, control signal overhead, frame retransmissions, etc.) to derive the actual throughput of a user. Therefore, the throughput of a user depends on the system parameters and protocols. It is shown in [56] that these factors can be accounted for by scaling the bandwidth and the SINR such that the throughput of user $k$ is given by

$$C_k(\Gamma) = b_1 B_{sc} N_k N_{sc} \log_2 (1 + \frac{\Gamma_k}{b_2}), \quad (3.7)$$

where $b_1$ accounts for the bandwidth losses due to control signals and $b_2$ considers the SINR implementation efficiency of LTE. However, in a multi-cell scenario, estimating the exact interference levels per subframe requires excessive traffic signaling among base stations. Therefore, in this chapter, we use the throughput estimates based on the SNR of each link during scheduling phase. This corresponds to noncooperative scheduling in game theory [57]. Hence, the scheduler at each base station uses the following utility

$$C_k(\gamma) = b_1 B_{sc} N_k N_{sc} \log_2 (1 + \frac{\gamma_k}{b_2}), \quad (3.8)$$

to solve the FDPS problem. The same type of scheduling is also studied in [43–52].
3.3 Utility-Based Resource Allocation

In this section, we present the utility-based resource allocation and study a family of utility functions that enable us to investigate the SRM, PF, and MMF schedulers. We also discuss the price of fairness metric and investigate the corresponding upper bounds for each of these schedulers.

3.3.1 Utility Functions

To analyze the utility-based frequency-domain resource allocation in noncooperative cells, we use the $\alpha$-fair utility function that is defined in [58] as

$$U_\alpha (C_k) = \begin{cases} \log (C_k) & \text{if } \alpha = 1 \\ C_k^{1-\alpha} / (1 - \alpha) & \text{if } \alpha \neq 1, \alpha \geq 0, \end{cases} \quad (3.9)$$

where $\alpha$ is the parameter that characterizes the efficiency and fairness trade-off. When $\alpha = 0$ is considered, the scheduler simply maximizes the sum of individual utilities. This is also called as the utilitarian solution in optimization theory and it has the maximum efficiency [39]. In this chapter, we refer to this allocation as the SRM scheduler. In cellular radio systems, this corresponds to the best effort solution.

When $\alpha = 1$, the scheduler is referred to as the PF scheduler. This scheduler is well-studied in the literature [39, 40, 58–60]. In [59], it is defined that a vector of rates $\mathbf{r}^*$ is proportionally fair if it is feasible, and if for any other feasible vector of rates $\mathbf{r}$, the aggregate proportional change is zero or negative. This is represented by

$$\sum_{k=1}^{K_c} \frac{r_k - r_k^*}{r_k^*} \leq 0, \quad (3.10)$$
where \( r_k \) and \( r_k^* \) are the \( k \)th elements of the rate vectors \( \mathbf{r} \) and \( \mathbf{r}^* \), respectively.

In general, as \( \alpha \) increases, the efficiency of the system decreases \([39, 40]\) (see \([61]\) for a unique counter example). The MMF scheduler maximizes the minimum rate offered to the users and achieves the maximum fairness. Besides the SRM, PF, and MMF schedulers, we will also look into \( \alpha \) values in the range \( 0 < \alpha < 1 \) to further analyze the trade-offs between efficiency and fairness.

In order to provide the reader some intuition about utility functions, the following observations are important. Notice that the \( \alpha \)-fair utility function in (3.9) is a nondecreasing concave up function for \( \alpha > 0 \). This makes this utility function achieve fairness as \( \alpha \) increases. This can be shown as follows. We know that if \( U_\alpha(C) \) is a nondecreasing concave up function, then \( U'_\alpha(C_k) \geq U'_\alpha(C_j) \) for \( C_k \leq C_j \), where \( U'_\alpha(C) \) denotes the derivative of the utility function. This means that for any nondecreasing concave up utility function, the increases in small rates are more favored compared to the larger rates. Note that we are using the terms rate and utility interchangeably since we focus on the user rate as the utility to be maximized. Similarly, any nondecreasing convex (concave down) utility function that one defines will favor the increases in the larger utilities. When the utility function is linear, the increases in all utilities are favored equal.

### 3.3.2 The Price of Fairness

We can now introduce the price of fairness as a metric to compare scheduler efficiencies. When resources are shared among users, natural questions on how to allocate resources and how to balance the trade-off of efficiency versus fairness arise. In fact, this general problem can be observed in many areas of economics, finance, and social welfare. In the sequel, we define the expressions for the price of fairness and derive the upper bounds in both single-cell and multi-cell environments. In particular, the detrimental effects of interference need to be
considered in the multi-cell scenario. Let us first start with defining the system efficiency of
the SRM scheduler in a single-cell scenario as

\[ U_{SRM}(\gamma) = \sum_{k \in \mathcal{K}_c} C^*_k(\gamma), \]  

(3.11)

where \( C^*_k(\gamma) \) represents the throughput of users associated with base station \( c \) in the optimal
solution. Similarly, the aggregate throughput of the optimal solutions for the \( \alpha \)-fair, PF,
and MMF schedulers are denoted by \( U_\alpha, U_{PF}, \) and \( U_{MMF} \), respectively. Next, the price of
fairness is defined as the percentage loss in the efficiency of a scheduler when compared to
the utilitarian solution \([39,40]\). In a single-cell scenario, it is given by

\[ \text{PoF}(U(\gamma)) = \frac{U_{SRM}(\gamma) - U(\gamma)}{U_{SRM}(\gamma)}, \]  

(3.12)

where \( U(\gamma) \) denotes the sum of utilities for any scheduler, and \( \text{PoF}(U(\gamma)) \) ranges between
0 and 1. A lower price of fairness means higher efficiency. It achieves its lowest value of zero
at the utilitarian solution, i.e., \( U = U_{SRM} \). In general, as the system efficiency decreases,
fairness in the system increases, and consequently, the price of fairness increases. At the worst
case, when the scheduler does not schedule any resources, the price of fairness becomes 1.

The application of this metric in wireless cellular radio communications enables us to quantify
the performance of different schedulers and select the parameter \( \alpha \). Let us note that in
cellular radio communications, each user experiences different channel conditions due to the
random nature of wireless radio. Therefore, they have different maximum achievable utilities.
For instance, cell-center users achieve significantly higher peak rates compared to those of
cell-edge users. In \([40]\), upper bounds are derived for the price of fairness of \( \alpha \)-fair schedulers
where users have unequal achievable utilities. They are given by

\[ \text{PoF}(U_\alpha) \leq 1 - \min_{x \in [1, K_c]} \frac{(\frac{B}{L})^{\frac{1}{\alpha}} x^{1+\frac{1}{\alpha}} + K_c - x}{(\frac{B}{L})^{\frac{1}{\alpha}} x^{1+\frac{1}{\alpha}} + \frac{B}{L}(K_c - x)x}, \]  

(3.13)
where \( B \) and \( L \) represent the highest and lowest maximum achievable utilities over all users, respectively. In order to find appropriate \( B \) and \( L \) parameters, we make the following observation. In an adaptive bandwidth system, the maximum achievable utility of user \( k \) is achieved when full bandwidth is scheduled to user \( k \). Then, \( B \) and \( L \) are given as

\[
B = b_1 B_{sc} M \log_2 \left( 1 + \frac{\gamma_l}{b_2} \right), \text{ s.t. } l = \arg \max_{k \in K_c} C_k(\gamma)
\]

\[
L = b_1 B_{sc} M \log_2 \left( 1 + \frac{\gamma_m}{b_2} \right), \text{ s.t. } m = \arg \min_{k \in K_c} C_k(\gamma),
\]

where \( \gamma_l \) and \( \gamma_m \) denote the wideband SNR values when the whole bandwidth is scheduled to users \( l \) and \( m \), respectively. As the ratio of the highest to lowest maximum achievable utility, \( B/L \), increases, the bound in (3.13) loosens [40]. Similarly, the price of fairness in a multi-cell scenario is defined as

\[
\text{PoF} \left( U(\Gamma) \right) = \frac{U_{SRM}(\Gamma) - U(\Gamma)}{U_{SRM}(\Gamma)},
\]

where \( U_{SRM}(\Gamma) \) and \( U(\Gamma) \) denote the sum of user throughput including the interference such as \( U(\Gamma) = \sum_{k \in K_c} C_k(\Gamma) \). In Section 3.5, we demonstrate that employing this metric and the bounds on each \( \alpha \)-fair scheduler give network operators an analytical tool to compare and quantify the efficiency and fairness trade-offs.

### 3.3.3 A Case Study

Consider a network where the SRM scheduler satisfies the QoS constraints of users and the network operator wants to increase the fairness in the network without violating these constraints. Let us assume that the network operator can tolerate a system efficiency loss up to 20% compared to the SRM scheduler in order to increase the fairness and we want to determine an appropriate \( \alpha \) value such that the efficiency and fairness trade-off can be
Figure 3.3: Feasibility region is depicted for a single-cell network with $K_c = 2$ users and $N = 15$ RBs.

balanced. This scenario may occur in cases where the operators need to improve the user satisfaction considering the QoS constraints. In this example, we consider a simplified simulation setup with $K_c = 2$ users and $N = 15$ RBs in a single-cell scenario. Fig. 3.3 depicts the feasibility region for different $\alpha$ values. We identify the critical points in the feasibility region that correspond to optimal solutions of different $\alpha$-fair schedulers. Note that the boundaries of the feasibility region are Pareto optimal, which was discussed in Section 3.1.2. We assume that $P_{dB}^0 = -70$ dBm and $\beta = 1$ is employed. Furthermore, let the path loss values of user 1 and 2 to be 60 dB and 90 dB, respectively, and ignore the frequency fading and shadow fading for the sake of simplicity. The corresponding solutions and the feasibility region are depicted in Fig. 3.3. It can be observed that the SRM scheduler maximizes the sum rate without any fairness concerns. It assigns all the resources to user 1 which achieves 55.11 Mbps, while user 2 is not served. The PF scheduler ignores the solution where any user is not assigned any resources since the payoff of such a solution will be $\log(0) = -\infty$. Instead, the PF scheduler always allocates a resource to each user and achieves a more fair distribution of resources compared to the SRM scheduler. In this case, it assigns 8 RBs and
7 RBs to users 1 and 2, respectively, and their throughput are 29.39 Mbps and 17.16 Mbps. It can be observed that as the fairness is introduced to the system, the system efficiency is reduced. Also, we observe that the MMF scheduler allocates more RBs to the users with low channel gains in order to increase the minimum rate user. This is achieved, again, at the expense of system efficiency. With the MMF scheduler, users 1 and 2 are assigned 6 RBs and 9 RBs, respectively, and they are served at 22.05 Mbps and 21.48 Mbps.

The maximum achievable utilities of users 1 and 2 are 55.1 Mbps and 33.8 Mbps, respectively. The theoretical upper bounds of the price of fairness for \(\alpha = \{0.25, 0.5, 0.75, 1\}\) and the MMF scheduler are \{0.073, 0.147, 0.179, 0.199, 0.282\}, respectively, using the bounds in (3.13). The empirical values of the price of fairness are \{0.055, 0.103, 0.129, 0.155, 0.210\}, in the same order as before. Notice that the bounds are tight. Also, for the given system efficiency tolerance, the network operator decides to employ \(\alpha = 1\) by only using the maximum achievable utilities of its users without the need of extensive simulations.

3.4 FDPS Schedulers

In this section, we present the FDPS problem formulation, implementation constraints, optimality conditions, set partitioning method, and solution space dimensions.

3.4.1 Problem Formulation

We can now discuss the uplink FDPS problem for an SC-FDMA system. As shown earlier, we investigate a family of objectives parameterized by a single variable, \(\alpha\). Given the utility
function defined in (3.9), the FDPS problem can be mathematically expressed as follows

\[
P_1: \max_{\{N_1, \ldots, N_{K_c}\} \in \mathcal{N}} \sum_{k=1}^{K_c} \omega_k U_\alpha(C_k(\gamma)) \tag{3.16a}
\]

s.t. \(\mathcal{N}_k \cap \mathcal{N}_j = \emptyset, \ \forall k \neq j\) \tag{3.16b}

\(\mathcal{N}_1 \cup \cdots \cup \mathcal{N}_{K_c} \subseteq \mathcal{N}\) \tag{3.16c}

\(N_k \in \{1, \ldots, N\}, \ \forall k\) \tag{3.16d}

where \(\omega_k\) denotes the nonnegative QoS weight for user \(k\). The first constraint assigns each RB to only one user without any overlap, and the second constraint ensures that all the resources are assigned. The third constraint denotes that the cardinality of the set of RBs scheduled to user \(k\) can take any integer values up to \(N\). Note that the construction of RB assignment sets, \(N_k\), ensures the contiguity of RB assignments.

In a multiuser scenario, there may exist assignments that yield cases where some users do not get any resources. We refer to this as the user rate starvation, and it creates a significant problem for network operators in terms of satisfying user experience and QoS requirements. Typically, user rate starvation occurs when the cell-edge users that experience low channel gains are not served at the benefit of improving the rates of cell-center users. In order to avoid user rate starvation cases, we consider the following assumption.

**Assumption 1.** At every subframe, each user \(k \in K_c\), connected to base station \(c_k\), is assigned at least one RB.

Assumption 1 can be incorporated into problem P1 by introducing an additional constraint.

59
Then, we rewrite the scheduling problem as

\begin{align}
\text{P2:} \quad & \max_{\{N_1, \ldots, N_{K_c}\} \in \mathcal{N}} \sum_{k=1}^{K_c} \omega_k U_\alpha (C_k(\gamma)) \\
\text{s.t.} \quad & \mathcal{N}_k \cap \mathcal{N}_j = \emptyset, \ \forall \ k \neq j \quad (3.17b) \\
& \mathcal{N}_1 \cup \cdots \cup \mathcal{N}_{K_c} \subseteq \mathcal{N} \quad (3.17c) \\
& 1 \leq N_k \leq N - K_c + 1, \ \forall k \in \mathcal{K}_c \quad (3.17d)
\end{align}

where the third constraint includes Assumption 1. We use P2 to solve the PF scheduling problem efficiently. Section 3.4.5 discusses that this constraint significantly reduces the search space and results in a faster algorithm. Note that for the PF scheduler, any user that does not get any resources contributes to the objective as \(\log(0) = -\infty\), and the optimal solution does not allow such an allocation.

The MMF scheduling problem considers the same constraints in P2 but it has a different objective. This problem can be written as

\begin{align}
\text{P3:} \quad & \max_{\{N_1, \ldots, N_{K_c}\} \in \mathcal{N}} \min_{k \in \mathcal{K}_c} \omega_k U_\alpha (C_k(\gamma)) \\
\text{s.t.} \quad & (3.17b) - (3.17d).
\end{align}

If a user is not allocated any resource, then the objective of the MMF scheduler will be zero which does not occur unless \(K_c > N\), and we avoid this condition in this chapter. Hence, using P2-P3 instead of P1 for the PF and MMF schedulers still achieves the optimal solution when \(K_c \leq N\) is satisfied.
3.4.2 Implementation Constraints

The practical constraints for the SC-FDMA system include those of the DFT implementations. In order to support an efficient DFT design, it was agreed in [62] to restrict the largest prime-factor that needs to be supported. In this chapter, we consider the implementation of radix set \( \{2, 3, 5\} \). Then, the number of RBs assigned to a user needs to be divisible by 2, 3, and 5. Note that, although we have only considered these three radices for a computationally fast implementation, this can be extended to include any radices. Hence, the number of resources allocated to user \( k \) in the uplink can only take the following values [63, p. 17] [27],

\[
N_k = |\mathcal{N}_k| = 2^{\nu_1} \times 3^{\nu_2} \times 5^{\nu_3}, \quad \forall k \in \mathcal{K}_c
\] (3.19)

where \( \nu_1, \nu_2, \) and \( \nu_3 \) are non-negative integers. It is shown in [62] that the DFT constraints of the above radix set reduce the number of complex multiplications more than five-folds compared to the unconstrained case.

The DFT constraints can be added to the FDPS problem as follows

\[
P_4: \max_{\{N_1, \ldots, N_{K_c}\} \in \mathcal{N}} \sum_{k=1}^{K_c} \omega_k U_\alpha (C_k(\gamma))
\] (3.20a)

s.t. \( \mathcal{N}_k \cap \mathcal{N}_j = \emptyset, \quad \forall k \neq j \) (3.20b)

\( \mathcal{N}_1 \cup \cdots \cup \mathcal{N}_{K_c} \subseteq \mathcal{N} \) (3.20c)

\( N_k = 2^{\nu_1} 3^{\nu_2} 5^{\nu_3}, \quad \forall k \in \mathcal{K}_c \) (3.20d)

\( 1 \leq N_k \leq N - K_c + 1, \quad \forall k \in \mathcal{K}_c \) (3.20e)

where the constraint in (3.20d) includes the DFT constraint that the number of RBs that can be scheduled to a user needs to be a multiple of the radix set. The last constraint in (3.20e) ensures that every user gets at least one RB. Section 3.5 discusses the simulation results with and without the DFT constraints in order to observe the effects of implementation.
3.4.3 Optimality Conditions

Let $\mathcal{U}$ denote the set of feasible throughput vectors. Suppose $\mathbf{r}, \mathbf{r}^* \in \mathcal{U}$ denote two feasible rate vectors, then $\mathbf{r}^*$ is optimal if it satisfies the first-order optimality condition, that is,

$$
\nabla U_\alpha(\mathbf{r}^*)(\mathbf{r} - \mathbf{r}^*) = \sum_{k=1}^{K_c} \frac{(r_k - r_k^*)}{(r_k^*)^\alpha} \leq 0, \quad r_k, r_k^* \in \mathcal{U}, \forall k \in \mathcal{K}_c
$$

(3.21)

where $\nabla U_\alpha(\cdot)$ denotes the derivative of the utility function with respect to $C_k$ and for $\alpha \geq 0$ [64]. In the special case, when $\alpha = 1$, then (3.21) satisfies the proportional fairness condition in (3.10). The optimality condition in (3.21) can be rewritten as

$$
\mathbf{g}^T \mathbf{r} \leq 1, \quad \forall \mathbf{r} \in \mathcal{U}
$$

(3.22)

where the elements of $\mathbf{g}$ are given by

$$
g_k = \frac{(r_k^*)^{-\alpha}}{\sum_{k=1}^{K_c} (r_k^*)^{1-\alpha}}, \quad \forall k \in \mathcal{K}_c.
$$

(3.23)

Moreover, since the $\alpha$-fair utility function is twice differentiable, the second-order sufficiency condition for $\mathbf{r}^*$ is that

$$
(\mathbf{r} - \mathbf{r}^*)^T \nabla^2 U_\alpha(\mathbf{r}^*)(\mathbf{r} - \mathbf{r}^*) = -\alpha \sum_{k=1}^{K_c} \frac{(r_k - r_k^*)^2}{(r_k^*)^{1+\alpha}} < 0
$$

(3.24)

is satisfied for all $\mathbf{r} \in \mathcal{U}, \mathbf{r} \neq \mathbf{r}^*$, where $\nabla^2 U_\alpha(\cdot)$ denotes the Hessian matrix of the utility function [64].

constraints on the system efficiencies.
3.4.4 Set Partitioning Solution

The problems investigated above can be cast as binary integer programming problems. The following special form is referred to as the set partitioning problem [65]

\[
\max_x \ f^T x \\
\text{s.t. } Ax = 1, \ x_i \in \{0, 1\}, \forall i
\]

where \( A \) denotes the constraint matrix with binary elements 0 and 1. The vectors \( f \) and \( x \) are the weighting and the binary assignment vectors, respectively. Both vectors have dimension \( JK_c \times 1 \), where \( J \) denotes the total number of possible contiguous resource allocation patterns for each user, and it will be further explained in the sequel. The binary variable \( x_{jk} \in x \) is associated with RB assignment pattern \( j \) for user \( k \). We assign \( x_{jk} = 1 \) if RB assignment pattern \( j \) is assigned to user \( k \), and 0 if not. Only one RB assignment pattern, corresponding to a column of \( A \), can be selected per user.

The objective vector \( f \) requires constant channel state information (CSI) updates per sub-frame in order to adapt to the varying channel conditions. The \( \alpha \)-fair utility function for \( 0 \leq \alpha < 1 \) uses the following objective function

\[
f_{k,m} = \sum_{m \in \mathcal{M}_k} \omega_k \frac{(b_1 B_{sc} N_k N_{sc}^{RB} \log_2 (1 + \gamma_k / b_2))^{1-\alpha}}{1 - \alpha}.
\]

(3.26)

In the special case, when \( \alpha = 0 \), the SRM scheduler maximizes the aggregate throughput. Similarly, the objective function for the PF scheduler is

\[
f_{k,m} = \sum_{m \in \mathcal{M}_k} w_k \log (b_1 B_{sc} N_k N_{sc}^{RB} \log_2 (1 + \gamma_k / b_2))
\]

(3.27)

where the objective is to maximize the sum of the logarithm of utilities. Note that the base
of the logarithm operator is not critical. Unlike previous schedulers, an auxiliary variable $\nu$ needs to be introduced to solve the MMF scheduler. Then, using the auxiliary variable, the MMF scheduler problem can be translated into a mixed integer program such that

$$\begin{align*}
\max_x \nu \\
\text{s.t. } \nu \leq f^T x \\
Ax = 1, \quad x_i \in \{0, 1\}, \forall i
\end{align*}$$

where $\nu$ is a free variable and $f$ is given in (3.26) with $\alpha = 0$.

The constraint matrix $A$ is created once, and does not require any updates as long as the number of users connected to the base station stays the same. The matrix $A$ has dimension $(K_c + N) \times JK_c$, and it can be expressed as

$$A = \begin{bmatrix}
A_1 & A_2 & \cdots & A_{K_c} \\
1_j^T & 0_j^T & \cdots & 0_j^T \\
0_j^T & 1_j^T & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0_j^T \\
0_j^T & \cdots & 0_j^T & 1_j^T
\end{bmatrix},$$

where each submatrix $A_k$ has dimension $N \times J$. For illustration purposes, the constraint matrix $A$ is divided into two parts in (3.29). The upper portion of $A$ is composed of concatenation of submatrices, $[A_1, \ldots, A_{K_c}]$. Without loss of generality, we consider that $A_k$’s are the same for each user. Also, each submatrix, $A_k$ represents the set of all RB assignment patterns with the contiguity constraint, and it can be written as

$$A_k = \begin{bmatrix}
q_{1N}^k, \ldots, q_{N_j}^k
\end{bmatrix}, \forall k \in K_c$$

(3.30)
where each column of $A_k$, that is $q^j_N$, $j \in \{1, \ldots, J\}$ represents one possible contiguous RB assignment pattern. For example, the binary vector $q^1_N = [1, 0, \ldots , 0]^T$ assigns the first RB to user $k$. Also, notice that the weight of the assignment pattern, $w(q^j_N)$, is equal to the cardinality of RB assignment set $|N_k|$, that is $N_k$, and we express this as $w(q^j_N) = |N_k| = N_k$, $\forall k \in \mathcal{K}_c$. Lower portion of the constraint matrix $A$ in (3.29) has $K_c$ rows, and it ensures that each user is assigned only one RB assignment pattern. For this purpose, it is structured in a staircase form where each row of $A_k$ has $J$ consecutive entries of ones and each column has only a single one entry. The remaining entries of $A_k$ are all zeros. In fact, the construction of the constraint matrix $A$ with the contiguity constraint is what distinguishes single carrier localized-FDMA schedulers from OFDM schedulers. As an example, assume that $N = 4$ RBs and $M = 2$ users. Then, the constraint submatrix $A_k$ for problem P1 is given by

$$A_k = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1
\end{bmatrix}$$

(3.31)

where the columns of $A_k$ correspond to RB assignment patterns, $q^j_N$, $j \in \{1, \ldots, J\}$, with $J = 11$. When we impose Assumption 1, then the first and last columns of $A_k$ in (3.31) need to be removed. The DFT constraints can be included to the constraint submatrix $A_k$ for P3 by removing the columns that does not satisfy $w(q^j_N) = |N_k| = 2^3 3^2 5^3$. Then, the constraint submatrix for P2 and P3 are

$$A_k = \begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1
\end{bmatrix}$$

(3.32)

where $J = 9$. In the next section, we shall present, in detail, how to find the total number
of resource allocation patterns for each of the above problems.

### 3.4.5 Finding The Exact Search Spaces

Now, we can discuss the total number of resource allocation patterns for the problems P1-P4 when the set partitioning approach is applied. First, let us consider the FDPS problem in P1. The total number of RB allocation patterns for the SRM scheduler is

\[
J_1 = 1 + \sum_{i=1}^{N} (N - (i - 1)) = \frac{N^2}{2} + \frac{N}{2} + 1 \tag{3.33}
\]

where \( N \geq K_c \), and \( J_1 \) increases as the number of RBs increase. Note that \( J_1 \) is independent of the number of users in the system.

Next, we look at the total number of RB assignment patterns for P2. This case corresponds to the PF and MMF schedulers, and it includes Assumption 1. Then, the total number of RB allocation patterns is given by

\[
J_2 = \sum_{i=1}^{N-K_c+1} (N - (i - 1)) = (N - K_c + 1) \left( \frac{N}{2} + \frac{K_c}{2} \right) \tag{3.34}
\]

where \( N \geq K_c \). Also, in order to investigate the effects of DFT implementation constraints, we revisit the SRM, PF, and MMF scheduling problems. When we impose the DFT constraints for the SRM scheduler, the total number of RB allocation patterns is

\[
J_3 = \frac{N^2}{2} + \frac{N}{2} + 1 - \sum_{i \in \mathcal{E}_1} (N - (i - 1)) \tag{3.35}
\]

where the set \( \mathcal{E}_1 \) denotes the set of numbers that cannot be achieved using the powers of radix set. An example set is \( \mathcal{E}_1 = \{7, 11, 13, 17, 19, \ldots\} \). Note that the largest element in \( \mathcal{E}_1 \) can be at most \( N \). Similarly, when we consider Assumption 1 and the DFT constraints, the
The total number of RB allocation patterns for the PF and MMF schedulers become

\[ J_4 = \sum_{i=1}^{N-K_c+1} (N - (i - 1)) - \sum_{i \in E_2} (N - (i - 1)) \]

\[ = (N - K_c + 1)(N/2 + K_c/2) - \sum_{i \in E_2} (N - (i - 1)) \]  

(3.36)

where \( N \geq K_c \). \( E_2 \) includes the set of numbers that are not multiples of the radix set. The largest element of \( E_2 \) is at most equal to \( N - K_c + 1 \) due to Assumption 1.

As a final remark, if this resource allocation problem is solved with subcarrier granularity instead of RB granularity, the total number of subcarrier assignment patterns per user becomes

\[ J_5 = 1 + \sum_{i=1}^{M} M - (i - 1) = \frac{M^2}{2} + \frac{M}{2} + 1 \]  

(3.37)

where \( M \geq K_c \). Obviously, resource allocation pattern size with subcarrier granularity in (3.37) is prohibitively complex to be implemented in real time.

In order to give the reader a sense of the dimensions of the problems investigated in this chapter, we give an example assignment in an LTE system. Assume a bandwidth of 10 MHz with 50 RBs, and 12 users per sector, \( J_1, J_2, J_3, J_4, \) and \( J_5 \) are 1276, 1209, 739, 717, and 180301, respectively. Since the problem size closely depends on the total number of RB allocation patterns, \( J \) is a critical factor when multiple schedulers are compared. With the DFT constraints and Assumption 1, it is clear to see that both the total number of RB allocation patterns, \( J \), and the problem search space are significantly reduced.
3.5 Numerical Results

3.5.1 Simulation Setup

In this section, we present the simulation results for the $\alpha$-fair, SRM, PF, and MMF schedulers in a multi-cell multiuser scenario. Our goal is to identify the efficiency and fairness trade-offs of these schedulers along with the characterization of the aggregate user rate and power. In our simulation model, we consider 19 macrocells, in which each cell is employed with 3-sector antennas. The users are randomly dropped within the macrocell sector area and each user is equipped with a single omni-directional antenna. We investigate a range of user numbers from 2 to 16. Furthermore, we consider that all users are always active and have data to transmit. This model is known as the full buffer traffic model [11]. Also, shadow fading of a user to each base station is considered to be spatially correlated, and the procedure to generate this parameter is detailed in [29]. Shadow fading parameters among different users are assumed to be uncorrelated, although in practice if mobiles are close to each other, this would not hold [29]. The simulation models and parameters are summarized in Table 3.1. These are in accordance with the standard models and baseline simulation setups in [11, 29]. We assume an uplink system with a bandwidth of 10 MHz, and included frequency selective fading in order to observe the effects of channel dependent scheduling. It is assumed that the power delay profile follows the modified Pedestrian B channel model in [29]. The channel state is assumed to be static during a frame and it is independent from one to another. We consider $b_1 = 0.75$ and $b_2 = 1.25$ as suggested in [56]. Finally, the wrap-around technique is used to avoid edge effects.

In each simulation, the data are collected from all 57 cells and the experiment is repeated 10 times to obtain reliable statistics. As discussed in Section 3.2, noncooperative schedulers are used at each base station. In particular, we investigate five properties of each scheduler: the price of fairness, the fairness index, the aggregate user rates, the cumulative distribution
function (c.d.f) of user rates, and the aggregate power. We use the fairness index definition in [54] such that

$$F = \left( \frac{K_c}{\sum_{k=1}^{K_c} C_k(\gamma)} \right)^2 / \left( K_c \sum_{k=1}^{K_c} C_k^2(\gamma) \right).$$  \hspace{1cm} (3.38)$$

It attains its maximum value 1 when each user has the same data rate, and its minimum value $1/K_c$ when all the resources are allocated to a single user only.
3.5.2 Optimization Setup

To solve the optimization problems, we used IBM ILOG CPLEX Optimization Studio v12.4. This tool employs the dynamic search method, a variant of the branch-and-cut method [66]. In the branch-and-cut method, the binary integer problem is first reduced to the LP problem by relaxing the binary integer constraint. The relaxed LP problem is referred to as the root problem, and used as a starting point to divide the main problem into subproblems. The solution to the root problem provides the upper bound for the solution to the binary integer problem. By introducing new cuts, using methods such as cutting-plane algorithms, new subproblems are formed until the integer solution that maximizes the objective is obtained.

The average number of iterations required to solve the root problem is no more than $(JK_c/2)$ for the binary integer problems [65, p. 128]. Unfortunately, we cannot provide any upper bounds for the mixed integer problem used to solve the MMF scheduler. Fig. 3.4 depicts the empirical results for the average number of iterations to solve the root problems for the

Figure 3.4: The average number of iterations to solve the root problems are depicted for the theoretical upper bounds (dotted), and the empirical results considering the DFT constraints (dashed-lines) and without these constraints (solid lines).
3.5.3 Numerical Evaluation

Fig. 3.5 summarizes the c.d.f. of user rates for different schedulers for $K_c = 8$ users. First, we observe that the c.d.f. of each scheduler with DFT constraints (dashed-lines) closely follows the performance of each scheduler without any constraints (solid lines). Second, we can clearly see the user rate starvation case for the SRM scheduler. On average, 9% of the users do not get any resources with the SRM scheduler. On the other hand, cell-center users achieve significantly more rates with the SRM scheduler, especially in the region above the
70th percentile. Second, we observe that the PF scheduler improves user rates at and below the median, and those rates typically correspond to the cell-edge and median users. For instance, when we look at 5, 10, and 50th percentile user rates, the PF schedulers provide 1.63, 2.17, and 4.13 Mbps, respectively. The SRM scheduler only achieves 0.53 and 2.64 Mbps for 10 and 50th percentile users. Around 70th percentile, the cross-over occurs, and the SRM scheduler starts to provide higher rates compared to the PF scheduler. Third, the MMF scheduler assigns most of the resources to the cell-edge users to maximize the rate for the minimum rate user, and this significantly reduces the median and cell-center user rates. Note that these results are expected since the fairness in the cellular radio networks corresponds to the trade-off between the rates provided to cell-edge and cell-center users. Thus, on the two edges of the efficiency and fairness trade-off lies the SRM and MMF schedulers, respectively, and the PF scheduler balances both properties.

In Fig. 3.6(a)-(b), the aggregate user rates and the price of fairness for each scheduler are depicted for various number of users per sector. First, we observe that the performance losses incurred by considering the DFT constraints are almost negligible for each scheduler in both figures. In Fig. 3.6(a), we depict the aggregate user rates of each scheduler for different number of users. This plot also demonstrates the multiuser diversity such that as the number of users increases, the aggregate user rate increases. In Fig. 3.6(b), we observe that the price of fairness of the \( \alpha \)-fair schedulers (for \( \alpha \leq 1 \)) are less than 0.2 in a multi-cell multiuser scenario. Unfortunately, the upper bounds for the price of fairness of most \( \alpha \)-fair schedulers (\( \alpha > 0.25 \)) are loose. This is due to the large \( B/L \) factor discussed in Section 3.3.2. However, we observe that the bounds are tighter for the \( \alpha = 0.25 \) case. The price of fairness values for the MMF scheduler are significantly high such that it ranges between 0.1-0.3. Notice that this means that the system efficiency loss between the SRM and MMF scheduler is between 10-30%. Also, it needs to be emphasized that Figs. 3.6(a)-(b) are closely related with each other. The percentage loss in the aggregate user rates depicted in Fig. 3.6(a) is, in fact, the price of fairness of each scheduler in Fig. 3.6(b) compared to the aggregate user
rate of the SRM scheduler. Thereby, we can argue that the price of fairness metric gives the network operator a meaningful and reliable metric to compare the efficiencies of any schedulers.

Fig. 3.6(c) depicts the fairness of each scheduler versus the number of users. Previously in Section 3.3, we discussed that as $\alpha$ increases, the fairness increases as well. We can observe this in Fig. 3.6(c). The $\alpha$-fair and MMF schedulers provide highly fair distribution of user
rates compared to the SRM scheduler. The fairness of the $\alpha$-fair schedulers ($\alpha > 0$) ranges between 0.7 and 0.9, whereas the fairness of the SRM scheduler is always less than 0.6. Moreover, the fairness gradually decreases as the number of users increases for the SRM scheduler. This is due to the fact that as the number of users increases, it is more probable that the SRM scheduler allocates more RBs to the users with good channel conditions, and only a few RBs are scheduled to cell-edge users in order to increase the aggregate data rate. At the worst case, the fairness of the SRM schedulers is 0.34 for $K_c = 16$ users, and at this point, the PF scheduler provides roughly three times more fair distribution of user rates compared to the SRM scheduler.

Fig. 3.6(d) shows the aggregate uplink transmit power versus the number of users for each scheduler. Note that the total dissipated power at UEs is a critical factor due to battery life concerns. We observe that the SRM scheduler always results in an allocation where significantly less power is required compared to the other schedulers. This is due to the fact that the SRM scheduler does not necessarily allocate resources to the cell-edge users that have high path loss values. Therefore, it avoids cases where cell-edge users transmit at high power levels. Also, the gap between the aggregate power of SRM and other schedulers gradually increases as the number of users increase.

### 3.6 Conclusion

In this chapter, we investigated the uplink resource scheduling problem for SC-FDMA systems. We studied a family of utility functions called as the $\alpha$-fair utility function. In particular, we focused on the SRM, PF, and MMF schedulers, and highlighted the system efficiency, fairness, and power trade-offs. We identified that the DFT implementation constraints result in only a slight degradation in the system efficiency and fairness. We introduced a general framework to compare the performance of different schedulers by using concise metrics such
as the price of fairness and fairness index. We observed that the \( \alpha \)-fair schedulers for \( \alpha > 0 \) provide fairer distributions of resources at the cost of some efficiency losses and increased transmit power. However, finding the appropriate \( \alpha \) parameter that satisfies the QoS constraints and improves the cell-edge user rates is a challenge. The presented analysis provides network operators with a greater provisioning to identify these trade-offs associated with resource allocation for different schedulers and optimize the user experience in the system. Although it is not considered in this work, the frequency domain schedulers investigated here can be used to complement time domain schedulers. This would require an appropriate weighting rule, including the past user rates over a time window, in the scheduling objective. Thereby, different QoS requirements of each user such as delay and guaranteed bit rate, possibly due to different types of applications, can be included in the scheduling algorithm.
Chapter 4

Common Rate Maximization in
Two-Layer Cellular Radio Systems

4.1 Motivation

In cellular networks, employing low-power base stations to umbrella macrocells offers more opportunities to provide increased system capacity and throughput, enhanced coverage and seamless service to reach high data rates [67]. Therefore, the application of low-power base stations are proposed in recent standards such as LTE and LTE-Advanced [68, 69]. There are two types of low-power base stations considered in this chapter and they are classified according to the existence of the backhaul connections with the high-power base station layers. First, we consider a microcell base station overlay that is connected to the macrocell layer through a fast backhaul. Second types of low-power base stations that we consider here are the relays that only have decode-and-forward capabilities and have no backhaul connections to the macrocell layer.

The throughput maximization problem in wireless networks has been vastly studied in the
literature for various optimization objectives and constraints. In [70], Chiang et al. analyze the sum-rate maximization, the worst-case user rate maximization and specific user rate maximization objectives, and propose geometric programming (GP) and heuristic solutions depending on the signal-to-interference ratio (SIR) regime. In [71], Gjendemsj et al. consider the weighted sum rate maximization problem in single-layer wireless networks. Julian et al. study the power allocation problem in single-layer networks to maximize the minimum SIR with certain quality of service and fairness constraints [72]. Karakayali et al. solve the common rate maximization problem in single-layer networks in [73]. On the other hand, there are fewer works on maximizing throughput in two-layer wireless networks. In [74], Raman et al. propose a linear programming (LP) solution to the same problem in a two-layer network system that consists of macrocell and relay layers. Our proposed algorithm differs from the aforementioned papers in two ways. We first present an analytical solution to solve the power control problem in two-layer wireless networks without any maximum power constraints and identify the necessary conditions for feasible power levels. We use this framework to define boundary points and apply LP to heuristically solve the common rate maximization problem in two-layer systems where we consider macrocell-microcell and macrocell-relay systems. For comparison, we also investigate the maximum common rate of an uncoordinated system without power control.

4.2 System Model

In this section, we present three different system models used in this chapter. In all three system models, we assume a central processor with sufficiently high computational capability, as in [73–75]. First, as a reference scenario, we consider a single-layer radio network where only macrocells are present in the system. As in [73,74] assume an idealized 19-cell hexagonal layout shown in Figure 4.1. In each hexagonal cell, the macrocell base stations are located
at the center of the cell. Every macrocell base station is equipped with three sector antennas such that each sector covers 120° within the cell. We consider universal frequency reuse such that all the base stations operate on the same frequency band. In order to avoid edge effects, we employ the wrap-around technique [74].

In our simulations, we place the users one-by-one randomly to each cell such that each sector only serves a single user. During the user generation process, the only condition we seek to satisfy is that each user has the highest received signal strength from its associated base station and not from any of the neighboring base stations. If this condition is not satisfied and the user needs to be handed over to the neighboring base station, that user is discarded and another one is generated. This method is the same as in the ones in [73,74].

The first transmission is devoted to estimating channel parameters for all the users in which pilot-assisted channel estimation methods such as least-squares (LS) or minimum mean square error (MMSE) estimators are used. We refer the reader to [76–78] and references therein for details on the channel estimation methods for various systems. Since in this chapter we only evaluate the performance increase with low-power base station deployment, we assume perfect channel estimation and leave the effects of the channel estimation error as future work.

We investigate a wide range of system loads and outage rates which are two important design parameters that closely affect the system throughput in wireless networks. Due to severe path losses and shadowing effects, the network performance can be significantly degraded when 100% system load is considered. For this reason, network design engineers use outage percentage as a design parameter in order to guarantee a common rate to the remaining users. In this chapter, we investigate a wide range of system loads from 80% to 100%. In the procedure to decide the users in outage, we discard users one-by-one from the system based on the worst SINR values until the designed system load is reached. By this way, coupling among the users are avoided. The user discarding procedure here is similar to the
ones in [73,74].

In the next transmissions, power control is applied to reduce intercell interference and the common rate maximization problem is solved by the central processor that can handle the high computational load in assigning power levels to the base stations. In the following sections, we present these techniques used here in detail. At the end of this step, the baseline system is completed.

The second and the third systems we evaluate include additional overlay of low-power base stations in order to help macrocell base stations to conserve power, increase coverage, and possibly improve system throughput. For the macrocell-microcell system, we assume that backhaul connections exist to connect both layers to the central processor where the channel gains of the users are analyzed to decide on downlink transmit powers. The macrocell-relay system includes no backhaul to connect both layers, only the macrocell layer base stations are connected among each other. The relays in the system turn on if they can decode the message from the macrocell base station they are associated with, otherwise they are turned off. Furthermore, we assume that through a control channel, the active relays notify the associated base stations about their status and the base stations send the assigned power levels to the relays in return.

Figure 4.2 depicts our two-layer cellular system layout where high-power and low-power base stations are deployed in the same area. Note that the position and the quantity of low-power base stations are very important while evaluating the system performance. In this work, we do not seek to place them in optimal locations. Instead, low-power base stations are located at a half distance from the cell center to cell edge in the boresight direction of the sector antennas.

The user generation methodology and user discarding procedures are the same as in single-layer networks. Note that the generated channel matrix is an augmented matrix compared to
Figure 4.1: Single-layer network layout with 19 hexagonal cells is displayed. Each cell has three-sector 120° directional antennas positioned at the center of the cell and each sector has one randomly placed user. Squares and circles depict macrocell base stations and users, respectively.

Figure 4.2: Two-layer hierarchical network layout with 19 high-power base stations overlaid with 19 low-power base stations placed at predefined locations. Additional low-power base station layer employs omnidirectional antennas. Squares, triangles and circles show macrocell base stations, low-power base stations and users, respectively.

the baseline case such that it additionally considers parameters regarding the second layer. Once the channel conditions are learned, the central processor applies power control, solves the rate-maximization problem, assigns power levels to all the base stations and notifies them through backhaul in the case of macrocell-microcell systems or through short notification messages in the control channel for the macrocell-relay systems. In the sequel, we describe
the analytical framework for the power control processes, identify necessary conditions for feasible power levels and analyze the common rate maximization problem for both single-layer and two-layer cellular networks.

4.3 Power Control

4.3.1 Single-Layer System

In the single-layer system, only high-power macrocell base stations are considered. First, we identify the target signal-to-interference-plus-noise-ratio (SINR) for user \( i \) as

\[
\gamma_i = \frac{g_{ii} p_i}{\sum_{j \neq i} g_{ij} p_j + \sigma_i^2}
\]  

(4.1)

where \( p_i \) is the transmit power of the \( i \)th base station, \( \sigma_i^2 \) denotes the noise power, and \( g_{ij} \) denotes the channel coefficient between user \( i \) and base station \( j \) and it includes the path loss and the shadowing. When we rearrange the terms above and divide by \( g_{ii} \), we get

\[
p_i = \gamma_i \sum_{j \neq i} \frac{g_{ij}}{g_{ii}} p_j + \gamma_i \frac{\sigma_i^2}{g_{ii}}.
\]  

(4.2)

In vector-matrix form, the above equation can be written to include all users as follows

\[
p = DFp + Du
\]  

(4.3)

where \( u_i = \sigma_i^2/g_{ii} \), \( D = \text{diag}\{\gamma_1, \ldots, \gamma_N\} \),

\[
F = \begin{cases} 
g_{ij}/g_{ii} & \text{if } j \neq i \\
0 & \text{if } j = i, \end{cases}
\]  

(4.4)
and we refer to $\mathbf{F}$ as the normalized channel gain matrix. The matrix $\mathbf{D}$ includes the target SINR values for each user on its diagonal entries. In the case of the common-rate maximization problem, all the users target the same SINR. Therefore, $\mathbf{D}$ can be reduced to a scalar $\gamma_0$ such that the Additive White Gaussian Noise (AWGN) channel capacity is $\log_2(1 + \gamma_0) = r_0$. Hence, the optimum power levels for a single-layer system providing a common rate target of $r_0$ can be shown as

$$
\mathbf{p}^* = (\mathbf{I}_N - \gamma_0 \mathbf{F})^{-1} \gamma_0 \mathbf{u}
$$

(4.5)

where $\mathbf{I}_N$ is an $N \times N$ identity matrix. In the following, we identify the nonnegativity of the assigned power levels to macrocell base stations $\mathbf{p}^*$ in single-layer. First, we assume that entries of channel gain matrix $\mathbf{F}$ are realizations of the underlying stochastic processes, namely they are all random path loss variables such that they are all independent and $\mathbf{F}$ is a full-rank matrix. Then, we apply the Perron-Frobenious theorem [79] which states that a real square matrix with nonnegative entries has a unique largest eigenvalue and its corresponding eigenvector has strictly nonnegative components. Since this theorem only applies to irreducible matrices, we need to question the irreducibility of $\mathbf{F}$. Note here that a square matrix is reducible if and only if it can be placed into block upper triangular form by simultaneous row-column permutations.

Following the same argument in [75], $\mathbf{F}$ becomes reducible if and only if there exists more than one 0 element on one row. Since we include the channel gains of every base station to every user using the wrap-around technique, we can conclude that $\mathbf{F}$ is irreducible. Hence, for an irreducible nonnegative matrix $\mathbf{F}$, there always exists a positive real eigenvalue of $\mathbf{F}$, $\lambda^*$ such that $\lambda^* = \max\{\lambda\}_{i=1}^{N} = \rho(\mathbf{F})$, which is called the spectral radius of $\mathbf{F}$. The eigenvector associated with this eigenvalue is element-wise nonnegative [79]. Using these results, we can
rewrite (4.5) such that

\[ p^* = (I_N - \gamma_0 F)^{-1}\gamma_0 u = \frac{1}{1 - \gamma_0 \rho(F)} \gamma_0 u \geq 0 \] (4.6)

and for the convergence of the solution, we seek that the spectral radius of \( F \) needs to be less than \( 1/\gamma_0 \), \( \rho(F) < 1/\gamma_0 \). Here, we note that this condition was already identified in [75, 80–83].

### 4.3.2 Two-Layer System

In the two-layer system, we consider cross-layer interference from both macrocell and microcell layers and update our target SINR definition such that

\[ \gamma_i = \frac{g_{ii}p_i + h_{ii}q_i}{\sum_{j\neq i} (g_{ij}p_j + h_{ij}q_j) + \sigma_i^2} \] (4.7)

where \( p_i \) denotes the power transmitted from the \( i \)th macrocell base station and \( q_i \) is the transmit power of the microcell base station \( i \). The channel coefficient representing the path loss and shadowing from the macrocell base station \( j \) to user \( i \) is denoted by \( g_{ij} \) and for the microcell \( j \) to user \( i \) is denoted by \( h_{ij} \). The noise power at receiver \( i \) is represented as \( \sigma_i^2 \). Using these, the above equation can also be expressed as the following when we rearrange terms and divide every term by \( g_{ii} \)

\[ \frac{p_i + h_{ii}q_i}{g_{ii}} = \gamma_i \sum_{j \neq i} \left( \frac{g_{ij}}{g_{ii}} p_j + \frac{h_{ij}}{g_{ii}} q_j \right) + \gamma_i \frac{\sigma_i^2}{g_{ii}}. \] (4.8)
One can rewrite the above equation in vector-matrix form as

\[
\begin{bmatrix}
I_N & C_{N \times N} \\
A_{N \times 2N} & 0 & \mathbf{p} \\
\mathbf{q} & B_{N \times 2N} & \mathbf{x}_{2N \times 1}
\end{bmatrix}
\begin{bmatrix}
\mathbf{p} \\
\mathbf{q} \\
\mathbf{x}_{2N \times 1}
\end{bmatrix}
= 
\begin{bmatrix}
D & F_{N \times N} & G_{N \times N} \\
B_{N \times 2N} & \mathbf{p} & \mathbf{x}_{2N \times 1}
\end{bmatrix}
\begin{bmatrix}
\mathbf{p} \\
\mathbf{q} \\
\mathbf{x}_{2N \times 1}
\end{bmatrix}
+ D
\begin{bmatrix}
\sigma^2_1 \\
\frac{g_{11}}{g_{ii}} \\
\vdots \\
\frac{g^2_N}{g_{NN}} \\
\mathbf{u}_{N \times 1}
\end{bmatrix}
\]

where \( N \) denotes the number of users in the system, \( I_N \) is an \( N \times N \) identity matrix and \( C \) and \( G \) matrices are as shown below

\[
C = \begin{cases} 
\frac{h_{ii}}{g_{ii}} & \text{if } j = i \\
0 & \text{if } j \neq i
\end{cases}, \quad G = \begin{cases} 
\frac{h_{ij}}{g_{ii}} & \text{if } j \neq i \\
0 & \text{if } j = i.
\end{cases}
\] (4.9)

We refer to \( G \) as the normalized channel gain matrix for the low-power base station layer where the normalization is carried out with respect to macrocell layer path loss values, \( g_{ii} \).

Similar to the single-layer case, for the common-rate maximization problem, \( D \) can be reduced into a scalar \( \gamma_0 \) and to determine the optimum power levels for the two-layer system, we need to solve \( A\mathbf{x} = \gamma_0 B\mathbf{x} + \gamma_0 \mathbf{u} \). When we rearrange the terms on each side, the power control problem can be expressed as

\[
A \left( I_{2N} - \gamma_0 \tilde{B} \right) \mathbf{x} = \gamma_0 \mathbf{u}
\] (4.10)

where \( \tilde{B} = A^{-1}B \) such that \( B = A\mathbf{\tilde{B}} \) and \( A^{-1} \) denotes the adjoint matrix of the rectangular matrix \( A \). Then, the optimal solution for the two-layer cellular system becomes

\[
\mathbf{x}^* = \gamma_0 (I_{2N} - \gamma_0 \mathbf{\tilde{B}})^{-1} A^{-1} \mathbf{u}.
\] (4.11)

To analyze the existence and nonnegativity of the optimal solution vector \( \mathbf{x}^* \), we follow a similar analysis as in the single-layer case and apply the Perron-Frobenious theorem. Then,
we see that there always exists a componentwise nonnegative power vector \( \mathbf{x}^* \) as long as the spectral radius of \( \tilde{\mathbf{B}} \) is less than \( 1/\gamma_0 \). Hence, we conclude that the necessary condition for feasibility in two-layer networks is \( \rho(\tilde{\mathbf{B}}) < 1/\gamma_0 \).

### 4.4 LP Solution and Heuristic Common Rate Maximization Algorithms

In this section, we consider maximum power constraints for each base station due to the physical limitations of radio amplifiers in the base stations and introduce our common rate maximization algorithm. In cases where the analytical solutions obtained using (4.5) or (4.11) exceed these levels, we need to pursue a different approach. Raman et al. have proposed a linear programming solution to this problem in [74] where the maximum power level constraints are introduced to the common rate maximization problem. Note that the two-layer system considered in [74] is our third system model where macrocell cellular network is overlaid with a relay layer.

In what follows, we introduce the common rate maximization problem in two-layer cellular networks and propose our heuristic solution. The common rate maximization problem for the macrocell-microcell system can be stated as

\[
\max_{\mathbf{p}, \mathbf{q}} \quad r_0 \\
\text{s.t.} \quad \log_2 \left( 1 + \frac{g_{i,p_i} + h_{i,q_i}}{\sum_{j \neq i} (g_{i,j}p_j + h_{i,j}q_j) + \sigma_n^2} \right) \geq r_0, \quad \forall i \\
0 \leq p_i \leq p_{\text{max}}, \quad \forall i \\
0 \leq q_i \leq q_{\text{max}}, \quad \forall i
\]

(4.12)

where \( r_0 = \log_2(1 + \gamma_0) \) bits/sec/Hz denotes the common rate provided to the users in the system, and \( p_{\text{max}} \) and \( q_{\text{max}} \) are the maximum transmit power levels of macrocells and
microcells, respectively. The first constraint ensures that the users get at least their target SINR as in [70,74] and the last two constraints impose the maximum power level constraints. The maximization problem solution in (4.12) yields a new set of power levels considering the maximum power constraints in each layer.

For the macrocell-relay system only those relays that can fully decode the first transmission are included in the solution. Then, we update the power constraints in (4.12) as

\[
0 \leq p_i \leq p_{\text{max}}, \quad \forall i \\
0 \leq q_i \leq q_{\text{max}}, \quad i \in S \\
q_i = 0, \quad i \in S^c
\]  \hspace{1cm} (4.13)

where \( S \) denotes the set of relays that can decode the first macrocell base station transmission and \( S^c \) denotes its complement set.

Although we state the common-rate maximization problem in (4.12) and (4.13), the proposed heuristic common rate maximization algorithm solves the following objective to employ power control

\[
\min_{p,q} \quad 1^T \begin{bmatrix} p \\ q \end{bmatrix}
\]  \hspace{1cm} (4.14)

with the associated constraints in (4.12) or (4.13). In the first step, we target a small but a feasible common rate \( r_0 \) and increase it in small \( \Delta r \) increments until the maximum power constraints are violated in either layer. The reason that we use the heuristic algorithm rather than the analytical solution is that the analytical solution reduces the macrocell power levels and increases the transmit power levels for the low-power base station above the permissible power levels since the users are typically closer to low-power base stations on average due to the inherent geometry. For this reason, we employed the LP solution in (4.14) to solve the common rate maximization problem under the maximum power constraints.
**Algorithm 2** Heuristic Power Control Algorithm to Maximize the Common Rate for Single-Layer Networks

1: while $\gamma_0 < 1/\rho(F)$ do  
2: \hspace{1em} $p = (I_N - \gamma_0 F)^{-1}\gamma_0 u$;  
3: \hspace{1em} if $p_i > p_{\text{max}}, \ i \in \{1, \ldots, N\}$ then  
4: \hspace{2em} $r_0 = r_0 - \Delta r$; $p = (I_N - \gamma_0 F)^{-1}u$; Exit  
5: \hspace{1em} else  
6: \hspace{2em} $r_0 = r_0 + \Delta r$;  
7: \hspace{1em} end if  
8: end while  
9: return $r_0, p^*$

**Algorithm 3** Heuristic Power Control Algorithm to Maximize the Common Rate for Two-Layer Networks

1: while $\gamma_1 < 1/\rho(\tilde{B})$ do  
2: \hspace{1em} Solve (4.14) with the constraints in (4.12) or (4.13)  
3: \hspace{1em} if $p_i > p_{\text{max}} \lor q_i > q_{\text{max}}, \ i \in \{1, \ldots, N\}$ then  
4: \hspace{2em} $r_1 = r_1 - \Delta r$  
5: \hspace{2em} Solve (4.14) with the constraints in (4.12) or (4.13)  
6: \hspace{2em} Exit  
7: \hspace{1em} else  
8: \hspace{2em} $r_1 = r_1 + \Delta r$;  
9: \hspace{1em} end if  
10: end while  
11: return $r_1, p^*, q^*$
Algorithms 2 and 3 are used to solve the heuristic common rate maximization problem in single-layer networks and two-layer networks. In the latter algorithm, we solve the objective in (4.14) with the convex constraints in (4.12) for the macrocell-microcell system and use (4.14) with the nonconvex power constraints in (4.13) for the macrocell-relay system solution. For comparison purposes, we record the maximum common rates \( r_0 \) and \( r_1 \) at the end of each while loop. The solution converges for both cases since we generate an increasing sequence of rates that are bounded [74].

4.5 Simulations

In this section, we present our simulation results for the common rate maximization problem when two-layer cellular systems are considered. We follow the simulation setup described in Section 4.2, and without loss of generality, consider the 19 hexagonal cell layout in Figure 4.1. In our simulations, we consider macrocell base stations with three sector antennas, each covering 120° within the cell. We assume an idealized cell radius of 1 km and employ the wrap-around technique to avoid the edge effects. The horizontal radiation pattern used for the three-sector antenna is

\[
A(\theta) = -\min \left( 12 \left( \frac{\theta}{\theta_{3dB}} \right)^2, A_{max} \right) \text{ (dBi)}
\]

(4.15)

where \(-180 \leq \theta \leq 180\), \(\theta_{3dB}\) and \(A_{max}\) denote the 3 dB beam width and the maximum attenuation, respectively, and they are taken as \(\theta_{3dB} = 65^\circ\) and \(A_{max} = 20\) dB [84]. For the microcell and relay base stations, we consider only omnidirectional antennas.

Following the user placement procedure described in Section 4.2, 57 users are placed in the system such that they share the same resource. In the first transmission, the channel estimation for all users is carried out and the central processor forms the channel gain matrix.
Based on this information, the system loading is adjusted. In our simulations, we sweep a wide range of system loading 80% to 100%.

We consider different channel models to model propagation in macrocell and microcell environments while adopting the proposed parameters specified in [84]. The path loss model for urban macrocells is based on modified COST 231 Hata urban propagation model and the microcell non-line of sight (NLOS) environment is based on COST 231 Walfish-Ikegami NLOS model. Furthermore, in our simulations, we assume the macrocell base stations height as $h_{BS} = 32$ m, mobile height as $h_{MS} = 1.5$ m, and carrier frequency as $f_c = 1900$ MHz. Then, the macrocell path loss model becomes

$$PL_{Macro}(dB) = 34.5 + 35 \log_{10}(d/m)$$ (4.16)

and we consider log-normal shadowing with a standard deviation of 8 dB. For NLOS microcell path loss model, we take microcell base station height as $h_{BS} = 12.5$ m, average building height as 12 m, mobile height as 1.5 m, orientation for all paths as $\phi = 30^\circ$, building to building separation as 50 m and street widths as 25 m. The resulting path loss model for NLOS microcells at $f_c = 1900$ MHz is

$$PL_{Micro}(dB) = 34.53 + 38 \log_{10}(d/m)$$ (4.17)

and log-normal shadowing with a standard deviation of 10 dB.

Moreover, we assume the maximum downlink power levels for macrocell base stations, microcell base stations and relays are 43 dBm, 33 dBm and 30 dBm as in [68,74]. Also, the base station and user antenna gains are taken as 15 dB and $-1$ dB, respectively. We also consider the other losses such as due to cabling losses, and it is taken as 10 dB. A major difference between our proposed algorithm and [74] is that the same path loss model is used
Figure 4.3: The figure shows the average common rate versus system loading in a 57 user system when COST 231 Hata urban propagation model and the COST 231 Walfish-Ikegami NLOS model are used.

to model both macrocell and relay environments in [74] and it is given by

\[
PL(dB) = 31.5 + 38 \log_{10} (d/m). \tag{4.18}
\]

where the log-normal shadowing parameter has a standard deviation of 8 dB. We note here that the transmission power levels of the base stations bound the permissible power level ranges and the antenna gains determine the received power levels along with the path loss models used. Hence, these parameters closely affect the maximum achievable common rate.

In our simulations, we used \( r_0 = 0.1 \) bits/sec/Hz and step size \( \Delta r = 0.1 \) bits/sec/Hz. Then, the heuristic algorithms for the common-rate maximization are applied. In Fig. 4.3, we plot the maximum common rate versus various system loads when the COST 231 Hata urban propagation model and the COST 231 Walfish-Ikegami NLOS model are used. The first observation we make is that by employing low-power base stations to support high-power macrocell base stations brings 23.52% to 68.03% increase in common rate throughput compared to the single-layer system. For instance, when 90% system load is considered, the maximum common rate increased from 1.48 bps/Hz to 1.85 bps/Hz (25.45% increase)
Figure 4.4: The figure shows the average common rate versus system loading in a 57 user system when (4.18) is used to model propagation in both layers.

and to 1.91 bps/Hz (29.14% increase) for macrocell-microcell and macrocell-relay systems, respectively.

Second, as the system load increases, the maximum common rate decreases due to the increase in both intercell and intracell interference. Obviously, when the system load is reduced by the central processor, the outage in the system increases and the availability of the service decreases. Hence, the maximum allowable system load in the system is a design parameter for the network design engineers to trade-off between the maximum common rate and the outage of the system. This parameter can also be used in admission control to make sure a certain level of common rate is offered to the users at all times.

Third, we see that the macrocell-microcell system offers slightly more common rate compared to the macrocell-relay system. In the former system, microcells are connected to the macrocell layer through backhaul, whereas in the latter, the relays need to be able to decode the message from the macrocell base stations. In our simulations, we observe that on the average 35 relays are active. In cases where the relays cannot decode the macrocell transmission, they are not included in the solution and this clearly reduces the degrees of freedom in the solution.
Also, for comparison purposes, we simulated the uncoordinated transmission where all the base stations transmit at full power instead of implementing power control. This is to observe the tradeoff between complexity and performance increase that power control brings. We see that even in the uncoordinated network system, employing low-power base station overlay to macrocell system improves the common rate performance. As an example, at 85% system load, the uncoordinated single-layer system consisting of only macrocell base stations can only provide up to 0.15 bps/Hz and this rate increases to 0.33 bps/Hz for the macrocell-microcell and to 0.28 bps/Hz for the macrocell-relay systems. We clearly see that almost half an order of magnitude increase in common rate can be achieved when power control is employed. An important observation is that the system load cannot exceed 95% without coordination in all three systems.

Figure 4.4 depicts the results for the same analysis when the path loss model in (4.18) is used to model the propagation loss in both layers. We observed that the macrocell-microcell and macrocell-relay networks provided 28.87% to 127.09% and 25.32% to 114.16% common rate increase, respectively. For instance, at 90% system load, the maximum common rate increased from 1.52 bps/Hz to 2.03 bps/Hz (33.83% increase) and to 1.98 bps/Hz (30.5% increase) for the macrocell-microcell and macrocell-relay systems, respectively. We see that our simulation results and the results presented in [74] are consistent.

4.6 Conclusion

The deployment of low-power base station overlay to high-power base station layers offer increased common rate throughput in the cellular radio systems. These additional low-power base stations offer more opportunities to hand over the transmissions from macrocells to low-power layers where the same data rates can be achieved with less transmit power in the downlink. This advantage reduces the interference, brings significant power savings to the
operators and provides solutions to the coverage problems. In this chapter, we presented an analytical solution framework to solve the power control problem in two-layer cellular networks and outlined the feasibility conditions for determining the power levels. We proposed a heuristic solution to maximize the common rate offered in the two-layer systems. Through simulations, we showed that significant increase in common rate can be achieved for macrocell-microcell and macrocell-relay systems.
Chapter 5

Iterative Water-Filling Algorithms with Pricing for Wireless Network Energy Efficiency

5.1 Motivation

With the rapid increase in the number of mobile connected devices and continuing demand on higher data rates, there is a need for energy-efficient solutions for wireless networks. The energy efficiency is not going to be achieved through a single solution, but rather will consist of a cumulative effect of several solutions. These solutions will come in many different flavors such as new enabling technologies (e.g., massive MIMO, device-to-device communications), new architectural changes (e.g., deployment of heterogeneous networks), energy-efficient equipment (e.g., the advances in power amplifiers), protocol changes (e.g., IEEE 802.3az standard), etc. [2]. Energy efficiency is important for both the network operators and the end-users. From the perspective of a network operator, energy efficiency means low-
ering the operational expenses, improving environmental sustainability, and reducing their carbon footprint. From the point of view of an end-user, an energy-efficient equipment means a longer battery life and mitigation of the energy trap problem, see [2]. In this chapter, we address these concerns on network power consumption and throughput, and we design algorithms to improve the network energy efficiency.

5.1.1 Related Works

Related works on energy efficiency maximization problem similar to the one studied in this chapter include [85–89]. In [85–87], the authors study maximizing the energy efficiency of single links, that consist of a transmit and a receiver pair, rather than maximizing the throughput. In their energy efficiency definition, they consider both the transmit power and power consumed in the circuitry. They demonstrate that the solution of the energy efficiency maximization problem does not always overlap with the one that maximizes throughput. In our chapter, we will pursue the same approach and consider the static power consumption of base stations along with the transmit power. In the framework proposed in [88], the authors formulate the energy efficiency maximization problem in a similar way to the one proposed in this chapter. Through the use of methods for fractional programming, the study in [88] defines the energy efficiency maximization problem as a bi-criterion optimization in which the rate maximization and power minimization problems are weighted. They use the Dinkelbach method for finding the roots of this bi-criterion optimization problem, which yields very simple and robust iterations. In our study, we also pursue the same approach for casting the bi-criterion optimization and employ the Dinkelbach method for root finding. Despite some similarities, our work differs from [88] in two major points. First, the problem in [88] is defined for a single-cell energy efficiency maximization, whereas we formulate a multi-cell energy efficiency maximization problem via introducing pricing mechanisms. From a game theoretical point of view, our formulation corresponds to a case where competing players
(base stations) cooperate to achieve a higher optimum solution for the sum of their individual profits (energy efficiencies), whereas the one in [88] corresponds to a non-cooperative scenario where the players compete for resources. This means that the solution proposed in this chapter will outperform the one in [88], as also demonstrated in our simulation results. The second difference is that, in our formulation, we extend the energy efficiency maximizing problem to include other constraints such as the minimum rate, total power constraints, and spectral mask constraints per subcarrier. Our prior work in [89] formulates the same problem of the multi-cell energy efficiency maximization, but it uses constant power allocation across subbands. It employs two variables to characterize the power transmissions per sector, which are to be optimized. In the sequel, we will pursue a different approach and consider allocating different power levels per subcarrier. We will obtain closed-form expressions which are iterated to find the optimal solutions. The fact that we use different power levels on each subcarrier brings an additional gain over the constant power allocation scenario. Hence, the framework proposed in this study provides an upper bound for the one in [89]. Another difference is that, in [89], we only implemented power control in macrocell base stations and did not consider it for picocell base stations. In this chapter, we will employ power control in both tiers.

Pricing in the resource allocation problem for wireless networks has been widely studied in the literature, see e.g., [90–96]. Especially, in cellular networks with dense base station deployments, inter-cell interference becomes a limiting factor that needs to be accounted for. Pricing mechanisms offer effective solutions to alleviate interference such that a higher network optimum solution can be achieved. To reduce the interference, for example, the studies in [90, 91] penalize the transmission of a base station proportional to their transmit power. However, this method requires a precise tuning of these penalty functions and does not offer any optimality. On the other hand, the studies in [92–95] propose to penalize the transmissions based on the interference they create. In order to convey the interference information, called as interference prices, limited information exchange between base stations
is required. The major difference between our work and the studies in [92–94], which also employ interference pricing, is that we incorporate the interference pricing terms to determine the water-filling levels, whereas those studies have not done so. Thus, we do not need to take any derivatives. In terms of optimality, as the studies in [94,95] also pointed out, the power control problem for the multi-cell networks is a non-convex problem. Due this non-convex nature, convergence to a global maximum is hard to achieve [95]. The obtained solutions satisfy the Karush-Kuhn-Tucker (KKT) conditions that guarantee convergence to a local maximum.

5.1.2 Contributions

In this chapter, we study the multi-cell multi-carrier network energy efficiency maximization problem. We take into account the transmit power and static power consumption of base stations. The linearized load-dependent power consumption model in [5] is employed. This model considers the contributions of the power amplifier, radio-frequency small-signal transceiver module, baseband receiver unit, power supply, and cooling. Using methods from fractional programming, we reformulate the energy efficiency maximization problem as a bi-criterion optimization problem in which the minimum power and maximum throughput problems are weighted accordingly. We obtain closed-form expressions for the water-filling algorithm. Using dual decomposition and the interference pricing mechanism, we decouple the network-wide energy efficiency problem into subproblems which are solved independently at each sector using limited information exchange. In addition, we incorporate several practical constraints in our formulation. We consider the total transmit power of a base station and the maximum power levels per subcarrier to account for different spectral masks and power amplifier constraints. We also incorporate the minimum rate constraints per user to account for different quality-of-service levels. Lastly, we apply the same framework for the network throughput maximization problem to investigate the differences between these
two problems. Since the proposed algorithms employ the closed-form expressions for the power updates and do not require any derivatives, their implementation complexities are significantly low compared to the works in [89,92–94]. We test our findings in a Long Term Evolution (LTE) network simulation testbed. In our simulation results, we evaluate the performance of the proposed algorithm and compare its performance with the algorithms proposed in [88] and [89]. We demonstrate that the proposed algorithm outperforms both of these studies.

The remainder of this chapter is organized as follows. In Section 5.2, we formulate the multi-cell energy efficiency maximization problem with power constraints. We derive the corresponding iterative water-filling solution. We present the proposed algorithm and discuss its steps in detail. We provide an algorithm for the throughput maximization problem in Section 5.2. We study the same problem for two-tier networks in Section 5.3. We extend the preceding framework to include minimum rate constraints in Section 5.4. We derive the corresponding solution and present its implementation steps. Section 5.5 presents our simulation results, where we evaluate the performance of the proposed algorithm and compare its performance with the several benchmarks to quantify the additional gains. Finally, Section 5.6 provides the concluding remarks.

5.2 Multi-cell Energy Efficiency Maximization Problem with Power Constraints in Single-Tier Networks

In this section, we discuss the energy efficiency maximization problem for the multi-cell multi-carrier systems in a single-tier network. This means that there are only macrocell base stations in the network. We consider three-sector antennas at macrocell base stations.
To model the power consumption at a base station, we employ the load-dependent power consumption model proposed in [5]. Our objective is to maximize the sum of sector energy efficiencies in the network subject to the power constraints at each base station. In what follows, we first obtain the power consumption expression in each sector and then define the energy efficiency maximization problem. We denote the power consumed at each macrocell base station sector $s$ by

$$P_{\text{Macro},s} = P_{0,s} + \Delta_M \sum_{n \in \mathcal{N}} p_s^{(n)}$$ (5.1)

where $P_{0,s}$ is the power consumption at the minimum non-zero output power of a macrocell sector $s$ and $\Delta_M$ is the slope of the load-dependent power consumption of macrocell base station sector [5]. The power per subcarrier $n$ at sector $s$ is represented by $p_s^{(n)}$ and $\mathcal{N}$ denotes the set of subcarriers. Using the power consumption model, we can formulate the multi-cell multi-carrier network energy efficiency maximization problem for a single-tier network as follows

$$\max_{s \in \mathcal{S}_m} \left[ \frac{\left( \sum_{n \in \mathcal{N}} \Delta_f \log_2 \left( 1 + p_s^{(n)} \chi_k^{(n)} \right) \right)}{P_{\text{Macro},s}} \right]$$

s.t. $\sum_{n \in \mathcal{N}} p_s^{(n)} \leq P_{\text{Total},s}$ for all $s \in \mathcal{S}_m$ (5.2)

$$P_{\text{max},s}^{(n)} \geq p_s^{(n)} \geq 0 \text{ for all } n \in \mathcal{N} \text{ and for all } s \in \mathcal{S}_m$$

where $\Delta_f$ is the bandwidth of a subcarrier and the set of all macrocell sectors is denoted by $\mathcal{S}_m$, the total transmit power of a macrocell base station sector is $P_{\text{Total},s}$, and the maximum transmit power per subcarrier is denoted by $P_{\text{max},s}^{(n)}$. Note that the quantity maximized in (5.2) has units bits/Joule. The channel-to-interference-plus-noise ratio (CINR) of user $k$ is
given by

\[ \chi_k^{(n)} = \frac{g_{k,s}^{(n)}}{\sigma^2 + I_k^{(n)}} = \frac{g_{k,s}^{(n)}}{\left(\sigma^2 + \sum_{s' \neq s, s' \in S^{(n)}} p_{s'}^{(n)} g_{k,s'}^{(n)}\right)}, \]  

(5.3)

where \( g_{k,s}^{(n)} \) is the channel gain between user \( k \) and macrocell sector \( s \), and \( I_k^{(n)} \) is the interference incurred by user \( k \) on subcarrier \( n \). The set \( S^{(n)} \) is the set of base stations that transmit on subcarrier \( n \). Using this notation, \( s' \neq s, s' \in S^{(n)} \) denotes the set of base stations that creates the interference to user \( k \) on subcarrier \( n \). In (5.2), we maximize the aggregate energy efficiencies of sectors with respect to the total power constraints and per subcarrier power constraints. The first constraint in (5.2) is due to the maximum power limitations on the base station analog front-end [97], which are defined by the standards. The second constraint in (5.2) arises due to the spectral masks [97]. The work in [98] shows how to relate a fractional program to a parametric program and develops an effective and simple algorithm. In this chapter, we will employ the same approach as in [98] such that the problem in (5.2) is translated into the following equivalent form by introducing a new parameter \( \lambda_s \) per sector

\[
\begin{align*}
\max & \quad \sum_{s \in S_m} \left[ \sum_{n \in \mathcal{N}} \Delta_f \log_2 \left( 1 + p_s^{(n)} \chi_k^{(n)} \right) - \lambda_s P_{\text{Macro},s} \right] \\
\text{s.t.} & \quad \sum_{n \in \mathcal{N}} p_s^{(n)} \leq P_{\text{Total},s} \quad \text{for all } s \in S_m \\
& \quad P_{\text{max},s}^{(n)} \geq p_s^{(n)} \geq 0 \quad \text{for all } n \in \mathcal{N} \text{ and for all } s \in S_m.
\end{align*}
\]  

(5.4)

This type of formulation enables us to obtain closed form expressions. From an optimization perspective, this corresponds to a bi-criterion optimization problem in which both the rate maximization and power consumption minimization are two objectives that we want to jointly solve [99]. In other words, we want to maximize the aggregate throughput while minimizing the power consumption in which the new objective in (5.4) weights the rate maximization objective with one and power consumption minimization objective by \(-\lambda_s\) at
each sector $s$. When we write the Lagrangian of the problem (5.4), we obtain

$$
\mathcal{L}(p_s, \lambda_s, \mu_s, v_{l,s}, v_{u,s}) = \sum_{s \in S_n} \left[ \sum_{n \in \mathcal{N}} \Delta_f \log_2 \left( 1 + p_s^{(n)} \chi_k \right) - \lambda_s P_{\text{Macro},s} 
+ \mu_s \left( P_{\text{Total},s} - \sum_{n \in \mathcal{N}} p_s^{(n)} \right)
+ \sum_{n \in \mathcal{N}} \left( v_{l,s}^{(n)} p_s^{(n)} + v_{u,s}^{(n)} \left( P_{\text{max},s}^{(n)} - p_s^{(n)} \right) \right) \right],
$$

(5.5)

where $\lambda = [\lambda_1, \cdots, \lambda_S]$, $\mu_s = [\mu_1, \cdots, \mu_S]$, $v_{l,s} = [v_{l,1}^{(n)}, \cdots, v_{l,S}^{(n)}]$, and $v_{u,s} = [v_{u,1}^{(n)}, \cdots, v_{u,S}^{(n)}]$.

The vectors $\mu$, $v_{l,s}$, and $v_{u,s}$ are the non-negative Lagrange variables associated with the total power at each base station, and the constraints on the minimum and maximum power levels per subcarrier, respectively. The transmit powers of all macrocell sectors over all subcarriers are denoted by $p$. Next, we take the derivative of (5.5) with respect to $p_s^{(n)}$ and equate the corresponding equation to zero. Then, we obtain

$$
\frac{\partial \mathcal{L}}{\partial p_s^{(n)}} = \frac{\Delta_f}{\log(2)} \cdot \frac{\chi_k^{(n)}}{1 + p_s^{(n)} \chi_k} - \frac{\Delta_f}{\log(2)} \sum_{j \neq k, j \in \mathcal{K}^{(n)}} \pi_{k,j}^{(n)} \cdot \Delta_M - \mu_s + v_{l,s}^{(n)} - v_{u,s}^{(n)} = 0,
$$

(5.6)

where the interference pricing terms are expressed as

$$
\pi_{k,j}^{(n)} = \frac{\gamma_j^{(n)}}{\gamma_j^{(n)} + 1} \cdot \frac{g_{j,s}^{(n)}}{I_j^{(n)} + \sigma^2},
$$

(5.7)

and where $\gamma_j^{(n)}$ is the signal-to-noise-ratio of user $j$ on subcarrier $n$. The set of users assigned to subcarrier $n$ is given by $\mathcal{K}^{(n)}$. Then, the set $j \neq k, j \in \mathcal{K}^{(n)}$ denotes the set of users that sector $s$ interferes on subcarrier $n$ while transmitting to its associated user $k$. Rearranging terms, we have the following closed-form expression for the transmit power allocated to
user \( k \) of sector \( s \) on subcarrier \( n \) as

\[
p_s^{(n)} = \left[ \frac{1}{\log(2) \cdot \Delta_f \cdot \Delta_M + \mu_s} + \sum_{j \neq k, j \in K^{(n)}} \pi_{k,j}^{(n)} - \frac{1}{\chi_k^{(n)}} \right]_{0}^{P_{\text{max},s}^{(n)}},
\]

where \([x]_{0}^{P_{\text{max},s}^{(n)}}\) denotes that \( x \) is lower bounded by 0 and upper bounded by \( P_{\text{max},s}^{(n)} \). Note that by projecting \( p_s^{(n)} \) onto the interval between zero and \( P_{\text{max},s}^{(n)} \), we omitted the \( v_{l,s}^{(n)} \) and \( v_{u,s}^{(n)} \) terms. Equation (5.8) suggests that whenever the transmissions of a sector creates high interference to the users in neighboring cells, the water-filling levels are reduced, and the corresponding transmissions are decreased. The closed-form expression in (5.8) closely depends on the value of \( \mu_s \). It is straightforward to show that the following is true

\[
\sum_{n \in \mathcal{N}} p_s^{(n)}(\mu_s) \leq P_{\text{Total},s},
\]

where the transmit power at subcarrier \( n \) is a function of \( \mu_s \). As the value of \( \mu_s \) increases, the aggregate transmit power monotonically decreases. To find the optimal \( \mu_s^* \) that satisfies the sum power constraints, we employ a one-dimensional search such as the bisection algorithm.

We summarize the corresponding bisection algorithm for the iterative water-filling algorithm under the heading Algorithm 4. For the bisection algorithm, we first need to determine the search domain. The lower bound \( \mu_{s,l} \) is set to zero, while the upper bound \( \mu_{s,u} \) is increased to the powers of two until the aggregate transmit power is below \( P_{\text{Total},s} \). When \( \mu_{s,u} \) is found, the algorithm proceeds to the classical binary search loop. The loop terminates when the difference between the upper and lower values is less than the threshold. Finally, if the sum of transmit powers is less than the total power constraint, then \( \mu_{s,mid} \) needs to be set to zero. This comes from the complementary slackness condition [99], and it was also pointed out in [95].
Algorithm 4 Bisection Method for the Iterative Water-Filling Algorithm

1: Let $\epsilon$ denote the tolerance and $l_{\text{max}}$ be the maximum number of iterations
2: Initialize $\mu_{s,l} = 0$ and $\mu_{s,u} = 1$
3: Calculate $p^{(n)}_s(\mu_{s,u})$
4: while $\sum_{n \in \mathcal{N}} p^{(n)}_s(\mu_{s,u}) > P_{\text{Total},s}$ do
5: $\mu_{s,u} = 2 \times \mu_{s,u}$
6: end while
7: while $|\mu_{s,u} - \mu_{s,l}| > \epsilon$ do
8: $\mu_{\text{mid}} = (\mu_{s,l} + \mu_{s,u})/2$
9: Calculate $p^{(n)}_s(\mu_{\text{mid}})$
10: if $\text{sign}(\sum_{n \in \mathcal{N}} p^{(n)}_s(\mu_{\text{mid}}) - P_{\text{Total},s}) = \text{sign}(\sum_{n \in \mathcal{N}} p^{(n)}_s(\mu_{s,l}) - P_{\text{Total},s})$ then
11: $\mu_{s,l} = \mu_{\text{mid}}$
12: else
13: $\mu_{s,u} = \mu_{\text{mid}}$
14: end if
15: end while
16: if $\sum_{n \in \mathcal{N}} p^{(n)}_s(\mu_{s,mid}) < P_{\text{Total},s}$ then
17: $\mu_{s,mid} = 0$
18: end if

Let the optimal cut-off value in the water-filling solution be defined as

$$\Omega^{*}_{EE,P} = \frac{\log(2)}{\Delta_f} \cdot (\lambda_s \cdot \Delta_M + \mu_s) + \sum_{j \neq k, j \in \mathcal{K}^{(n)}} \pi_{k,j}^{(n)},$$

(5.10)

where the initials EE and P stand for energy efficiency maximization and pricing, respectively. In the water-filling solution, this cut-off value can be interpreted as the threshold that determines if the subcarrier is used or not. Any subcarrier $n$ with the CINR, $\chi^{(n)}_k$, is not used if its magnitude is below the optimal cut-off value $\Omega^{*}_{EE,P}$. Mathematically, we can express this condition as

$$p^{(n)}_s > 0 \text{ if } \Omega^{*}_{EE,P} < \chi^{(n)}_k \text{ and } p^{(n)}_s = 0 \text{ if } \Omega^{*}_{EE,P} \geq \chi^{(n)}_k.$$  

(5.11)

Notice that the cut-off value depends both on frequency-dependent and frequency-independent terms. Frequency-dependent terms come from the interference pricing values, denoted by $\pi_{k,j}^{(n)}$, while the frequency-independent terms are system related parameters such as $\Delta_M, \Delta_f,$
and $\mu_s$. In an interference-dominated region, the water-filling levels are adjusted based on the interference pricing terms.

The closed-form expression in (5.8) corresponds to the solution for the energy-efficient maximization problem with pricing. For the case without pricing, the closed-form expression reduces to

$$p_s^{(n)} = \left[ \frac{1}{\frac{\log(2)}{\Delta_f} \cdot (\lambda_s \cdot \Delta_M + \mu_s)} - \frac{1}{\chi_k^{(n)}} \right]_{P_{\text{max},s}}^{(n)}$$

(5.12)

Similarly, the cut-off value for the case without pricing is given by

$$\Omega_{EE,NP}^* = \frac{1}{\Delta_f} \cdot (\lambda_s \cdot \Delta_M + \mu_s)$$

(5.13)

where the initials NP stand for no pricing case. Note that, in the case without pricing, the cut-off value is constant for all subcarriers and it has no frequency dependency since interference pricing terms are omitted in the solution. In Figures 5.1 and 5.2, we illustrate examples of water-filling energy efficiency maximization solutions without and with pricing, respectively. In Fig. 5.1, we observe that the optimal water-filling level is constant throughout the subcarriers, and thus, there is a single-level for water-filling. For example, when we incorporate interference pricing, we observe that there are multiple levels for water-filling level on each subcarrier. When the created interference is high on particular subcarriers, i.e., higher interference prices, the water-filling levels are lowered.

Another way of relating the problems in (5.2) and (5.4) is as follows. Let $q_s^*$ and $\lambda_s^*$ denote the respective solutions for these two problems in the same order as before. For each sector, consider the following function

$$F_s(\lambda_s) = \max_{p_s} \sum_{n \in N} \Delta_f \log_2 \left( 1 + p_s^{(n)} \lambda_k^{(n)} \right) - \lambda_s \cdot P_{\text{Macro},s}$$

(5.14)
where the vector $\mathbf{p}_s$ satisfies the feasibility conditions, i.e., \( \{\mathbf{p}_s \in \mathcal{P} | \sum_{n \in \mathcal{N}} p_s^{(n)} \leq P_{\text{Total},s}; P_{\text{max},s}^{(n)} \geq p_s^{(n)} \geq 0\} \) for all $n \in \mathcal{N}$ and $s \in \mathcal{S}_m$. Then, the following statements are true [88,98]:

\[
\begin{align*}
F_s(\lambda_s) &> 0, \quad \text{if } \lambda_s < q_s^* \\
F_s(\lambda_s) &= 0, \quad \text{if } \lambda_s = q_s^* \\
F_s(\lambda_s) &< 0, \quad \text{if } \lambda_s > q_s^*.
\end{align*}
\] (5.15)

Hence, solving problem (5.2) is equivalent to finding the roots of $F_s(\lambda_s)$, and the correspond-
Algorithm 5 Iterative Water-Filling Algorithm with Pricing for Network Energy Efficiency Maximization

1: Let $\epsilon$ denote the tolerance, $l_{\text{max}}$ be the maximum number of iterations, and $t = 0$. Initialize transmit power levels and interference prices. Solve the following at each sector $s$

2: while $|F_s(\lambda_s)| > \epsilon$ and $l < l_{\text{max}}$ do

3: $\lambda_s(l) = \left(\sum_{n \in \mathcal{N}} \Delta_f \log_2 \left(1 + p_s^{(n)} \chi_k^{(n)}\right)\right)/P_{\text{Macro},s}$

4: Obtain $\mu_s$ using the bisection method

5: For all $n \in \mathcal{N}$, solve for $p_s^{(n)}$

6: Calculate $F_s(\lambda_s)$

7: $l = l + 1$

8: end while

9: Update the power levels using

$$p_s^{(n)}(t + 1) = (1 - \delta) \cdot p_s^{(n)}(t) + \delta \cdot p_{\text{Next}}^{(n)}$$ (5.18)

10: Distribute the interference prices, $\{\pi_{k,j}^{(n)}\}$, among base stations

11: Go to Step 2 and repeat for $t = t + 1$

The optimal condition is

$$F_s(\lambda_s^*) = \max_p \left(\sum_{n \in \mathcal{N}} \Delta_f \log_2 \left(1 + p_s^{(n)} \chi_k^{(n)}\right) - \lambda_s^* P_{\text{Macro},s}\right) = 0.$$ (5.16)

We summarize the iterative energy-efficient water-filling algorithm with pricing under the heading Algorithm 5 in which the variables $p$, $\lambda_s$, and $\mu_s$ are iteratively updated. We use the Dinkelbach method to update $\lambda_s$ at each sector, which is an application of the classical Newton’s method for root finding [100]. This method has the following iterations

$$\lambda_s(l + 1) = \lambda_s(l) - \frac{F_s(\lambda_s)}{F_s'(\lambda_s)} = \frac{\sum_{n \in \mathcal{N}} \Delta_f \log_2 \left(1 + p_s^{(n)} \chi_k^{(n)}\right)}{P_{\text{Macro},s}},$$ (5.21)

where $F_s'(\lambda_s)$ denotes the derivative of $F_s(\lambda_s)$ with respect to $\lambda_s$. Next, we use $\lambda_s(l)$ to
Algorithm 6 Iterative Water-Filling Algorithm with Pricing for Network Throughput Maximization

1: Set \( t = 0 \) and initialize transmit power levels and interference prices. For each sector \( s \), solve the following:
2: Obtain \( \mu_s \) using the bisection method
3: For all \( n \in \mathcal{N} \), solve for \( p_{\text{Next}}^{(n)} \) using

\[
p_{\text{Next}}^{(n)} = \left[ \frac{1}{\log(2)/\Delta_f \cdot \mu_s + \sum_{j \neq k \in \mathcal{K}} \frac{n_{k,j}^{(n)}}{X_k^{(n)}}} - \frac{1}{\chi_k^{(n)}} \right] P_{\text{max},s}^{(n)}
\] (5.19)

4: Update the power levels using

\[
p_s^{(n)}(t + 1) = (1 - \delta) \cdot p_s^{(n)}(t) + \delta \cdot p_{\text{Next}}^{(n)}
\] (5.20)

5: Distribute the interference prices, \( \{\pi_{k,j}^{(n)}\} \), among base stations
6: Go to Step 2, and repeat for \( t = t + 1 \)

determine the power levels, \( p_s^{(n)} \), and use Algorithm 4 to find the optimal \( \mu_s \) value. Lastly, in order to avoid rapid fluctuations which may cause unstability in the system, we use the following technique where the power control parameters are updated as

\[
p_s^{(n)}(t + 1) = (1 - \delta) \cdot p_s^{(n)}(t) + \delta \cdot p_{\text{Next}}^{n}
\] (5.22)

where \( \delta \) satisfies [57, p. 286]

\[
\delta(t = 0) = 1, \quad \delta(t) \in (0, 1) \text{ for } t > 0, \quad \sum_{t=0}^{\infty} \delta(t) = \infty.
\] (5.23)

In general, \( \delta(t) \) is chosen as

\[
\delta(t) = \frac{t}{2t + 1}, \quad \text{for } t > 0.
\] (5.24)

As the limit goes to infinity, \( \delta(t) \) approaches \( 1/2 \). This iterative update method is called as the Mann iterative method [57]. Empirically, we have also observed that, without this iterative
method, the power level updates yield significantly large oscillations and sometimes do not converge. We also present an algorithm for the network throughput maximization problem with pricing under the heading Algorithm 6. Comparing the solutions of two problems, the major difference is that the terms $\lambda_s \cdot \Delta_M$ do not exist in the closed-form expression of the power updates in the throughput maximization problem as terms related to power consumption do not appear in the objective.

5.3 Multi-cell Energy Efficiency Maximization Problem with Power Constraints in Two-Tier Networks

Next, we consider network energy efficiency maximization for two-tier networks. We consider picocell deployments underlying the macrocell tier and our objective is to maximize the sum of energy efficiencies of all sectors. First, we need to express the total power consumed in a sector as

$$\psi_s = P_{\text{Macro},s} + \sum_{p \in S_{P,s}} P_{\text{Pico},p}, \quad (5.25)$$

where $P_{\text{Pico},p}$ is total power consumption of a picocell base station $p$ and the set $S_{P,s}$ is the set of picocell base stations in sector $s$. The power consumption at a picocell base station is given by

$$P_{\text{Pico},p} = P_{0,p} + \Delta_P \sum_{n \in N} P_p^{(n)}, \quad (5.26)$$

where $P_{0,p}$ and $\Delta_P$ are the power consumption at the minimum non-zero output power and the slope of the power consumption of a picocell base station $p$, respectively. The transmit power per subcarrier $n$ at picocell base station $p$ is represented by $P_p^{(n)}$. We can now formulate
the multi-cell energy efficiency maximization for two-tier networks as follows:

$$\text{max} \sum_{s \in S_m} \left[ \sum_{n \in \mathcal{N}} \Delta_f \log_2 \left( 1 + p_s^{(n)} \chi_k^{(n)} \right) + \sum_{p \in S_{P,s}} \sum_{n \in \mathcal{N}} \Delta_f \log_2 \left( 1 + p_p^{(n)} \chi_k^{(n)} \right) - \lambda_s \psi_s \right]$$

s.t. $$\sum_{n \in \mathcal{N}} p_s^{(n)} \leq P_{\text{Total},s} \quad \text{for all } s \in S_m$$

$$\sum_{n \in \mathcal{N}} p_p^{(n)} \leq P_{\text{Total},p} \quad \text{for all } s \in S_{P,s}$$

$$P_{\text{max},s}^{(n)} \geq p_s^{(n)} \geq 0 \quad \text{for all } n \in \mathcal{N} \text{ and for all } s \in S_m$$

$$P_{\text{max},p}^{(n)} \geq p_p^{(n)} \geq 0 \quad \text{for all } n \in \mathcal{N} \text{ and for all } s \in S_{P,s},$$

(5.27)

where $$P_{\text{Total},p}$$ and $$P_{\text{max},p}^{(n)}$$ are the total transmit power of a picocell base station $$p$$ and maximum transmit power of $$p$$ on subcarrier $$n$$, respectively. When we apply Lagrangian relaxation, take the derivative with respect to $$p_p^{(n)}$$, equate to zero, and rearrange the terms, we obtain the following closed-form expression of the iterative power updates for picocell base stations, which are given by

$$p_p^{(n)} = \left[ \frac{1}{\Delta_f \cdot \left( \lambda_s \cdot \Delta_P + \mu_p \right) + \sum_{j \neq k,j \in K^{(n)}} \pi_{k,j}^{(n)} - \chi_k^{(n)}} \right]_{0}^{P_{\text{max},p}^{(n)}},$$

(5.28)

where $$\mu_p$$ is the Lagrangian dual variable associated with the total power constraint of a picocell base station $$p$$. Note that the expression for $$p_s^{(n)}$$ remains the same as in (5.8). We present the proposed algorithm for energy maximization problem in two-tier heterogeneous networks in the next section where we also account for the minimum rate constraints.
5.4 Multi-Cell Energy Efficiency Maximization with Rate and Power Constraints

We now extend the preceding framework and include the minimum rate constraints per user. The multi-cell multi-carrier network energy efficiency maximization with power constraints and minimum rate constraints can be formulated as

\[
\begin{align*}
\max & \sum_{s \in S_m} \left[ \sum_{n \in N} \Delta_f \log_2 \left( 1 + p_s^{(n)} \chi_k^{(n)} \right) + \sum_{p \in S_p, n \in N} \Delta_f \log_2 \left( 1 + p_p^{(n)} \chi_k^{(n)} \right) - \lambda_s \psi_s \right] \\
\text{s.t.} & \sum_{n \in N_k} r_k^{(n)} \geq R_{\min, k}, \text{ for all } k \in K \\
& \sum_{n \in N} p_s^{(n)} \leq P_{\text{Total}, s} \text{ for all } s \in S_m \\
& \sum_{n \in N} p_p^{(n)} \leq P_{\text{Total}, p} \text{ for all } p \in S_p, \\
& P_{\max,s}^{(n)} \geq p_s^{(n)} \geq 0 \text{ for all } n \in N \text{ and for all } s \in S_m \\
& P_{\max,p}^{(n)} \geq p_p^{(n)} \geq 0 \text{ for all } n \in N \text{ and for all } p \in S_p, \\
\end{align*}
\]

(5.29)

where \( R_{\min, k} \) denotes the minimum rate requirement of user \( k \). As we consider multi-carrier systems, the aggregate throughput of subcarriers assigned to a user defines its rate. First constraint in (5.29) ensures that a user gets at least its minimum rate requirement. Similar to our previous discussion, we introduce \( \lambda_s \) per sector and the corresponding Lagrangian of
the problem (5.29) can be written as

\[
\mathcal{L}(\mathbf{p}, \mathbf{\lambda}, \mathbf{\tau}, \mathbf{\mu}) = \sum_{s \in \mathcal{S}_m} \left[ \Delta f \log_2 \left(1 + p_s^{(n)} \chi_k^{(n)}\right) + \sum_{n \in \mathcal{N}} \sum_{p \in \mathcal{S}_{p,s}} \Delta f \log_2 \left(1 + p_p^{(n)} \chi_k^{(n)}\right) - \lambda_s \psi_s + \sum_{k \in \mathcal{K}_s} \tau_k \left( \sum_{n \in \mathcal{N}_k} r_k^{(n)} - R_{\text{min},k} \right) + \mu_s \left( P_{\text{Total},s} - \sum_{n \in \mathcal{N}} p_s^{(n)} \right) + \sum_{p \in \mathcal{S}_{p,s}} \mu_p \left( P_{\text{Total},p} - \sum_{n \in \mathcal{N}} p_p^{(n)} \right) \right],
\]

(5.30)

where \( \mathbf{\tau} = [\tau_1, \cdots, \tau_K] \) denotes the vector of Lagrange multipliers associated with the minimum rate constraints and \( K \) is the total number of users in the system. The throughput of a user is the sum throughput of subcarriers assigned to this user. For a macrocell-associated user, \( r_k^{(n)} = \Delta f \log_2 \left(1 + p_s^{(n)} \chi_k^{(n)}\right) \), whereas for a picocell-associated user it is defined as \( r_k^{(n)} = \Delta f \log_2 \left(1 + p_p^{(n)} \chi_k^{(n)}\right) \). To obtain the closed-form expressions for the macrocell base station power updates, we take the derivative of (5.30) with respect to \( p_s^{(n)} \), equate it to zero, rearrange terms, and obtain the following closed-form expression for the power levels on each subcarrier

\[
\begin{align*}
p_s^{(n)} = & \left[ \frac{(1 + \tau_k)}{\Delta f} \cdot (\lambda_s \cdot \Delta M + \mu_s) + \sum_{j \neq k, j \in \mathcal{K}^{(n)}} (1 + \tau_j) \pi_{k,j}^{(n)} \right] \frac{\chi_k^{(n)} - 1}{\pi_{s,m}^{(n)}}. \\
\end{align*}
\]

(5.31)

The picocell base station power updates are given as

\[
\begin{align*}
p_p^{(n)} = & \left[ \frac{(1 + \tau_k)}{\Delta f} \cdot (\lambda_s \cdot \Delta P + \mu_s) + \sum_{j \neq k, j \in \mathcal{K}^{(n)}} (1 + \tau_j) \pi_{k,j}^{(n)} \right] \frac{\chi_k^{(n)} - 1}{\pi_{p,m}^{(n)}}. \\
\end{align*}
\]

(5.32)

When the user minimum rate constraint is satisfied, its corresponding Lagrangian multiplier is zero. In that case, (5.31) reduces to (5.8). In addition, we can express the optimal cut-off
value for the energy efficiency maximization case with rate constraints using (5.31) as

$$\Omega_{EE, P, RC}^{*\text{(n)}} = \left( \frac{\log(2)}{\Delta_f} \cdot (\lambda_s \cdot \Delta_M + \mu_s) + \sum_{j \neq k, j \in \mathcal{K}^{(n)}} (1 + \tau_j) \pi_{k,j}^{(n)} \right) / (1 + \tau_k), \quad (5.33)$$

where the initials RC stand for the rate constraints. Notice that when all the rate constraints are satisfied, the optimal cut-off value in (5.33) reduces to (5.10).

We need to emphasize that this type of formulation enables network operators to satisfy two contradicting objectives of maximizing the average energy efficiency (or similarly, the aggregate throughput) and introducing the fairness among users. For example, users who are subject to high interference conditions or low channel gains are typically allocated low power levels due to the water-filling principle. For this reason, their throughput values are typically low. The formulation in (5.29) solves this problem by increasing their power levels through the Lagrangian variables associated with the minimum rate requirements. Thus, it ensures that the system fairness is increased.

The iterative water-filling algorithm for network energy efficiency maximization problem with minimum rate constraints in two-tier networks is given under the heading Algorithm 7. The dual prices associated with the minimum rate constraints are updated using

$$\tau_k^{(l+1)} = \left[ \tau_k^{(l)} - \alpha^{(l)} \left( \sum_{n \in \mathcal{N}_k} r_{k}^{(n)} - R_{\text{min},k} \right) \right]^+, \quad (5.37)$$

where the operator $[x]^+$ denotes max$(0, x)$. The step size at $l$th iteration is denoted by $\alpha^{(l)}$. We employ an adaptive step size selection algorithm such that [101]

$$\alpha^{(l)} = \begin{cases} 
\beta \alpha^{(l-1)} & \text{if } \left( R_{\text{min},k} - \sum_{n \in \mathcal{N}_k} r_{k}^{(n,l)} \right) > \kappa \left( R_{\text{min},k} - \sum_{n \in \mathcal{N}_k} r_{k}^{(n,l-1)} \right) \\
\alpha^{(l-1)} & \text{otherwise}, 
\end{cases} \quad (5.38)$$
Algorithm 7 Iterative Water-Filling Algorithm with Pricing for Network Energy Efficiency Maximization with Minimum Rate Constraints in Two-Tier Heterogeneous Networks

1: Let $\epsilon$ denote the tolerance, $l_{\text{max}}$ be the maximum number of iterations, and $t = 0$. Set the initial transmit power levels, interference prices, and dual prices. At each sector, solve

2: while $|F_s(\lambda_s)| > \epsilon$ and $l < l_{\text{max}}$ do

3: Determine $\lambda_s$ using the following

$$\lambda_s = \left( \sum_{n \in \mathcal{N}} \Delta_f \log_2 (1 + p_s^{(n)} \chi_k^{(n)}) + \sum_{p \in \mathcal{S}_{P,s}} \sum_{n \in \mathcal{N}} \Delta_f \log_2 (1 + p_p^{(n)} \chi_k^{(n)}) \right) / \psi_s$$ (5.34)

4: Obtain $\mu_s$ using the bisection method in Algorithm 4

5: For all $n \in \mathcal{N}$, solve for $p_{\text{Next}}^{(n)}$ using (5.31)

6: for all $p \in \mathcal{S}_{P,s}$ do

7: Obtain $\mu_p$ using the bisection method in Algorithm 4

8: Solve for $p_{\text{Next},p}^{(n)}$ for all $n \in \mathcal{N}$ using (5.32)

9: end for

10: Update the dual prices, $\tau_k$ for all $k \in \mathcal{K}_s$, using (5.37)

11: Calculate the following

$$F_s(\lambda_s) = \sum_{n \in \mathcal{N}} \Delta_f \log_2 (1 + p_s^{(n)} \chi_k^{(n)}) + \sum_{p \in \mathcal{S}_{P,s}} \sum_{n \in \mathcal{N}} \Delta_f \log_2 (1 + p_p^{(n)} \chi_k^{(n)}) - \lambda_s \psi_s$$ (5.35)

12: $l = l + 1$

13: end while

14: Update the power levels using

$$p_s^{(n)}(t + 1) = (1 - \delta) \cdot p_s^{(n)}(t) + \delta \cdot p_{\text{Next}}^{(n)}$$ and

$$p_p^{(n)}(t + 1) = (1 - \delta) \cdot p_p^{(n)}(t) + \delta \cdot p_{\text{Next},p}^{(n)}$$ for all $p \in \mathcal{S}_{P,s}$ (5.36)

15: Distribute the interference prices, $\{\pi_{k,j}^{(n)}\}$, among base stations

16: Go to Step 2 and repeat for $t = t + 1$

where $r_k^{(n,l)}$ and $r_k^{(n,l-1)}$ are the throughput of user $k$ on subcarrier $n$ at iterations $l$ and $l - 1$, respectively. The scalar $\beta$ increases the step size if the difference between the minimum rate requirement and the throughput of a user is not decreased by a factor of $\kappa$ in the next time instant. In the simulations, we take the step size as $\alpha^{(0)} = 2.5 \times 10^{-4}$, the increment factor $\beta$ as 2, and the comparison threshold $\kappa$ as 0.9. This step size rule is studied more in detail in [101, p. 123] to update the dual prices in constrained optimization problems.
The proposed algorithm starts transmitting at an initial transmit power. Dual prices and interference prices are taken as zero initially. The algorithm calculates $\lambda_s$, and using this value, power levels for the macrocell and picocell base stations are determined. We update the dual prices and repeat this process until the convergence criterion is satisfied. To avoid rapid fluctuations, we use the Mann iterations as in Algorithm 5. Finally, interference prices are measured at the user and these measurements are fed back to the base stations, where they are distributed among base stations using the fast and reliable backhaul (for example, through the X2-interface in LTE, see [25]), and the process is repeated in the next time slot.

5.5 Simulation Results

In this section we present the simulation results for the single-tier and two-tier energy-efficiency maximization problems. In the simulation model, we follow the simulation models and parameters suggested in [11] as a baseline simulation for LTE heterogeneous networks. We consider a network consisting of 19 hexagonal macrocell deployments and each macrocell has 3-sector antennas. In each sector, 30 users are randomly generated within the macrocell sector area and each user is equipped with a single omni-directional antenna. This corresponds to the uniform user distribution scenario in [11]. For the two-tier simulation model, we deploy four picocells per sector. We consider a non-uniform user distribution where two users are initially dropped within a 40 meter radius per picocell base station and the remaining users are randomly generated. This model is also proposed in [11]. Simulation parameters and models are summarized in Table 5.1. For the scheduler, we employ the Equal Bandwidth Scheduler which is detailed in [11, 89, 102]. For the spectrum allocation, we consider the fractional frequency reuse scheme in [89], which is shown to achieve very high energy efficiency performance in two-tier heterogeneous networks compared to other benchmark spectrum allocations. We will investigate two problems: energy efficiency and
throughput maximization, and for each problem we consider the non-pricing and pricing scenarios. For the macrocell base stations, $P_{\text{Total},s} = 46 \text{ dBm}$ and $P_{\text{Total},p} = 30 \text{ dBm}$ [11]. Also, for simplicity, we take $P_{\text{max},s}$ and $P_{\text{max},p}$ as zero. The base station power consumption model parameters are taken as $P_{0,m} = 130 \text{ W}$, $P_{0,p} = 56 \text{ W}$, $\Delta_{M} = 4.7$, and $\Delta_{P} = 2.6$ as in [88]. Note that when a picocell base station has no associated users, we consider that it is in dormant mode and it consumes $P_{\text{Sleep},p} = 6.3 \text{ W}$.

### 5.5.1 Energy Efficiency and Throughput Results in Single-Tier Networks

In Figs. 5.3(a)-(d), we present the average sector energy efficiency and throughput results for a single-tier network. These four figures investigate different initial power levels for warm-up. In Figs. 5.3(a)-(b), we start the simulations with initially transmitting at maximum power levels, whereas the power levels are determined without any interference price information. Figs. 5.3(c)-(d). It can be observed that both initial power levels converge to the same
Figure 5.3: Average sector energy efficiency and throughput of a single-tier network using the proposed iterative water-filling algorithms with different initial power levels. IWF stands for iterative water-filling. EE-Max and Throughput-Max correspond to the solutions for the energy efficiency and throughput maximization problems, respectively. The solutions without pricing correspond to the algorithm in [88].

point after 40 time instants. Also, we observe that power control improves the energy efficiency and throughput by factors of 2.53 and 1.10 for the energy efficiency maximization problem, respectively, and 22% in energy efficiency and 16% in throughput for the throughput maximization problem.

When we compare the resource allocation with and without interference pricing, we observe that interference pricing brings 40% and 13% additional improvements in terms of energy efficiency for the energy efficiency and throughput maximization problems, respectively. Note that the scenario without interference pricing corresponds the algorithm proposed in [88].
Hence, it can be concluded that the proposed algorithm outperforms the one in [88].

Figure 5.4 illustrates another advantage of the proposed algorithm that it brings significant power savings. When we apply power control without interference pricing, the average transmit power reduces from 39.8 W (46 dBm) to 11.89 W (40.75 dBm), which corresponds to a reduction of 3.35 times. It is worth noting that when base stations communicate among each other to exchange interference prices, it can bring additional power savings. For example, in the energy efficiency maximization problem, average transmit power of macrocell base station reduces from 11.89 W (40.75 dBm) to 1.66 W (32.20 dBm) when pricing is introduced. Thus, we observe that interference pricing brings a power reduction of 7 times compared to the case without pricing and 24 times compared to the case without power control, which are very significant.

### 5.5.2 Energy Efficiency and Throughput Results in Two-Tier Networks

Figures 5.5(a)-(b) depict the average energy efficiency and aggregate sector throughput for the iterative water-filling algorithm with and without pricing in two-tier heterogeneous net-
Figure 5.5: Average sector energy efficiency and throughput of a two-tier network using the proposed iterative water-filling algorithms. CPA corresponds to the constant power allocation algorithm proposed in [89].

Figure 5.6: Average transmit power consumption of the proposed iterative water-filling algorithms with and without pricing in a two-tier network.
works. Note that the case without pricing corresponds to the algorithm in [88]. Also, for comparison, we evaluate the performance of the maximum power case and constant power allocation with pricing which was proposed in [89]. We observe that power control improves the energy efficiency and throughput by factors of 2.68 and 1.77, respectively. Interference pricing brings 39% improvement in energy efficiency and 29% in throughput over the case without pricing.

Figure 5.6 presents the transmit power consumption of each tier using the above algorithms. First, we observe that significant power savings can be achieved in the macrocell tier, whereas picocell base stations typically operate close to the maximum power levels. For example, in the case with pricing, iterative water-filling algorithm reduces the power consumption from the initial maximum power level of 46 dBm down to 20.2 dBm, whereas the average transmit power of a picocell base station is slightly reduced from 30 dBm to 28.4 dBm. Also, pricing mechanism brings an additional 3.6 times average transmit power saving per sector compared to the case without pricing, reducing it from 40.2 dBm to 34.6 dBm, which is very significant. These results illustrate why picocells should be deployed as an underlying tier such that users can be offloaded from the macrocell tier to the small cell tiers where the link distances are smaller and higher rates can be achieved.

5.5.3 Incorporating Minimum Rate Constraints in Two-Tier Networks

Finally, we extend the iterative water-filling algorithm for the energy efficiency maximization problem and we incorporate the minimum rate constraints. For simplicity, the same target rate is considered for all users. We need to note that, in real applications, users may have different rate requirements. For example, [25] considers a mixture of different traffic requirements consisting of best-effort users and users with strict rate requirements. Fig. 5.7
First, we observe that as the rate requirement increases, the average sector energy efficiency decreases. As we have derived in (5.33), the rate requirements are enforced through adjusting the water-filling levels. However, this comes at the expense of reductions in energy efficiency and throughput. For example, the average energy efficiency is 284.5 kbits/Joule without any rate constraints and it reduces to 187.6 kbits/Joule for rate constraints of 512 kbits/sec. When the rate requirements are not satisfied, the users are considered to be in outage.
Figure 5.8: (a) The outage probability of various minimum rate requirements and (b) the cumulative distribution function of user rates for the minimum rate requirement of 512 kbits/sec.

Fig. 5.8(a) presents the outage probability of users for different minimum rate requirements. As expected, a higher rate requirement yields a higher outage probability. As the dual prices are updated and interference prices are distributed, the number of users in outage decreases significantly. For example, when power control is not employed for the 512 kbits/sec case, the outage probability is 25%. Using the proposed algorithm, the outage probability gradually decreases to 7% at the end of 40 iterations. Fig. 5.8(b) illustrates the cumulative distribution of user rates. It depicts how the user rate distribution is improved using the proposed algorithm. We observe that the proposed algorithm outperforms the case without
power control, shifting every point of the cumulative distribution to the right.

5.6 Conclusions

Resource allocation in multi-cell networks is an important aspect for cellular wireless systems. In this chapter, we have investigated the energy efficiency maximization problem from a power control perspective. We have considered a realistic load-adaptive base station power consumption model capturing the characteristics of a macrocell and a picocell base station. We have obtained closed-form expressions for the water-filling solutions using methods from fractional programming. We have proposed numerous iterative water-filling algorithms for LTE networks with single-tier and two-tier deployments. We have incorporated interference pricing mechanism in which base stations communicate among themselves to exchange limited information. Then, the preceding framework has been extended to incorporate the minimum rate constraints per user. The corresponding closed-form expressions for the case with minimum rate constraints have been derived as well. The average energy efficiency, throughput, and transmit power consumption performance of the proposed algorithms have been evaluated and compared to other baseline works. The numerical results have demonstrated that the proposed algorithms can achieve significant gains and outperform the baseline methods.
Chapter 6

Energy-Efficient Resource Allocation for Fractional Frequency Reuse in Heterogeneous Networks

6.1 Motivation

Energy efficiency is an important issue in the next generation wireless networks. In particular, the high data rate demands of the LTE standards bring the hidden cost of increasing energy consumption. These have been studied in the literature under the topic of “Green Communications” [2]. Network operators are also seeking green solutions in order to reduce their operational expenses. Several methods have been investigated and standardized to increase the energy efficiency of the networks, see, e.g., [2,103–106]. Among these solutions, we study the FFR method in this chapter to reduce both the intra- and inter-cell interference.

In the FFR method, cells are divided into cell-center and cell-edge regions and orthogonal subbands are allocated in these regions. In an interference-dominated region, increasing the
transmit power slightly improves the throughput, but it significantly degrades the energy efficiency. The frequency allocation of the FFR method significantly increases the energy efficiency and reduces the outages.

The resource allocation problem for multicell networks has been widely investigated in the literature, see, e.g., [85–87,107,108]. Reference [107] investigates resource allocation in time, frequency, and both time and frequency domains. In [108], a heuristic algorithm, which minimizes the duality gap, is presented to solve the sum rate maximization problem using constant energy allocation across subbands. Considering the complexity of the problem and practical constraints, both references [107] and [108] propose to use constant energy allocation across subbands instead of using a multi-user water-filling algorithm. It needs to be noted that the same constraints also apply to the LTE systems, where the standards define the smallest scheduling granularity to be per resource block (RB) and constant power is allocated to the subcarriers within an RB [103]. In this chapter, we also employ this approach and find the optimal power levels that maximize the energy efficiency of the network. Recently, several studies have been reported investigating energy-efficient resource allocation in multicell networks, see, e.g. [85–87]. These works consider static power consumption along with transmit power to develop better link adaptation schemes. This problem is also addressed in this chapter, but in a large scale such that the energy efficiency of the network is maximized considering the static and dynamic power consumption at MeNBs.

The FFR scheme has been studied extensively in the context of Orthogonal Frequency Division Multiple Access (OFDMA) systems, see, e.g., [109–114]. The capacity and coverage characteristics of various FFR schemes are analytically investigated in [109] using stochastic geometry methods. Similar to our proposed algorithm, the study in [109] also uses constant power allocation across subbands. However, the FFR scheme employed in [109] is prone to high cross-tier interference for small cells, especially those close to the MeNB, as they are not protected from the MeNB downlink transmissions. In FFR schemes, the system performance
closely depends on the cell-center region radius. Reference [110] investigates the dependence of optimal cell-center radius on the number of users. It identifies that as the number of users increases, the optimal cell-center radius needs to decrease. The study in [111] finds that the network sum throughput is maximized when the radius of the cell-center region is chosen as 0.6 times the cell radius for single-layer networks, employing omnidirectional antennas. A similar study is presented in [112] for two-layer networks with three and six sector antennas in which 0.61 and 0.54 times the cell radius, respectively, are determined to be the cell-center region radii that maximize the network throughput. In contrast to the studies in [111,112] which consider only the network throughput, in this chapter, we also investigate the cell-center radius that maximizes the energy efficiency. This subject has not been investigated in the literature. Recent studies in [112–114] propose a novel FFR scheme for multi-tier HetNet deployments with sectorized MeNBs. In this chapter, we employ the same FFR scheme. Also, our proposed algorithm differs from [109–114] in that power control across subbands is employed to optimize the system performance in terms of energy efficiency.

In this chapter, we address the resource allocation problem in OFDMA systems employing the FFR method in HetNets deployments. We propose a novel resource allocation algorithm in which the objective is to maximize the energy efficiency in each sector. We account for both the transmit and static power consumed at base stations. The proposed algorithm divides the resource allocation problem into frequency and power allocation problems. To solve the frequency assignment problem, we investigate two well-known schedulers, the sum rate maximization (SRM) and equal bandwidth (EBW) schedulers, see, e.g., [102]. For power control assignments, the gradient ascent method is applied. We note that power control is applied only at the MeNB and not at the picocell base station (pico eNB). The reasons that power control is not applied at the pico eNBs are twofold. First, the maximum power transmission of pico eNBs ensures full coverage is provided in the small cells. Second, the backhaul congestion can be avoided as the network becomes denser with small cell deployments and excessive traffic overhead can be eliminated. The proposed framework can
be considered as a non-cooperative resource allocation in which only the intra-cell interference information is required.

In dense urban scenarios with overlapping base station coverage, however, the inter-cell interference may become a critical factor. In such scenarios, non-cooperative resource allocation does not account for the loss in utility one causes to other users due to interference. This is called as negative externality in economics [93]. One of the mechanisms that addresses this problem is to employ the interference pricing method. We study this method to provide an upper bound to the preceding framework in terms of energy efficiency and account for the negative externalities in the power allocation step. In interference pricing, downlink transmissions of base stations are penalized based on the inter-cell interference that they create. In general, this topic is investigated under the game-theoretic pricing research area in the literature, see, e.g., [90–94, 115]. In [92–94], the authors study a linear pricing function and propose an asynchronous distributed pricing (ADP) algorithm for ad-hoc networks. The pricing function reflects the marginal change in the utility per unit interference. Each terminal announces an interference price and then updates its power. The bottleneck of the ADP algorithm is that each terminal requires the interference price information of all other users, which is often difficult to distribute in practice. The authors of [92–94] also show that the solution satisfies the Karush-Kuhn-Tucker (KKT) conditions. Note that the power control problem in multicell networks is a non-convex problem which means that it can have multiple local optima and the KKT conditions can only guarantee local optimality. However, in general, convergence to the global optimal point is difficult to achieve. We applied the ADP approach to our algorithm in order to obtain an upper bound to its performance. Through simulations, we demonstrated that pricing brings only marginal improvements. This shows that the proposed algorithm reaches close to the optimal performance while requiring only intra-cell information exchange.

The remainder of the chapter is organized as follows. Section 6.2 introduces the system
model and presents the base station power consumption models. Section 6.3 formulates the energy-efficient resource allocation in LTE systems employing FFR method. The algorithm is proposed in Section 6.3, along with implementation steps, optimality conditions, and convergence proof. Section 6.4 introduces the pricing methods that consider inter-cell interference in the resource allocation. Numerical results are presented in Section 6.5 and concluding remarks are made in Section 6.6.

6.2 System Model

In this section, we first present the FFR method used in this chapter. Then we identify the interference conditions in each region and present the system model. Finally we discuss the base station power consumption models that are used to derive energy-efficient algorithms.

Interference is a key problem in mobile wireless communication systems. In today’s mobile communication networks, base station distances are typically on the order of less than a kilometer for urban deployments. This poses a challenging interference problem for HetNet deployments due to the large downlink transmit power differences. For example, the MeNBs and pico eNBs differ by 16 dB in their transmit power levels [103]. In order to mitigate interference, several methods have been proposed such as FFR [112], enhanced inter-cell interference cancellation [103], coordinated multipoint transmissions [116], and carrier aggregation [103]. In the FFR method, different subbands are allocated in different regions. Each cell is divided into two regions as the cell-center and cell-edge region. Depending on the locations, the macrocell and picocell associated users, abbreviated as MUEs and PUEs, respectively, are assigned to different subbands. This orthogonal frequency assignment significantly reduces both the intra- and inter-cell interference. In this chapter, we denote the cell-center region radius as $r_{th}$. This distance sets the region boundaries. For example, users that are closer than $r_{th}$ to the MeNB are considered to be in the cell-center and those that
are farther are considered to be in the cell-edge. Similarly, users connected to a pico eNB located closer than $r_{th}$ to the MeNB are considered to be in the cell-center.

Consider the FFR scheme depicted in Fig. 6.1, which is designed for the MeNBs employing three sector antennas with the objective of mitigating the interference. The system bandwidth is partitioned into four subbands. Subband $A$ is assigned to the cell-center region for the macrocell tier, and the rest of the spectrum is divided into three subbands, one subband for each sector is used in the macrocell cell-edge for the MUEs, as shown in Fig. 6.1. This allocation means that frequency reuse (FR) of three is used in the macrocell tier for the cell-edge. Since the inter-cell interference is mitigated, the outages at the cell-edge MUEs are reduced. In the picocell tier, two of the remaining subbands are allocated to the PUEs in the cell-center region, while subband $A$ and the same two subbands are used in the cell-edge region. With the pico eNBs in cell-edge region using the rest of the subbands, better use of spectrum resources can be achieved. Due to different spectrum allocations, interference conditions vary depending on the cell region and the associated tier. Let us identify the interference conditions for the MUEs and PUEs in the cell-center and cell-edge regions in different sectors. For example, in Sector 1, the MUEs in the cell-center region are assigned to subband $A$, while the PUEs in the same region are allocated to subbands $C$ and $D$. On the other hand, in the cell-edge region, the MUEs are scheduled on subband $B$, whereas the PUEs operate on subbands $A$, $C$, and $D$.

First, consider the MUEs in the cell-center region that are scheduled on subband $A$. The strongest interfering base stations for the users in this region are the six MeNBs in the first ring surrounding this cell. Also, note that the surrounding 12 MeNBs in the second ring create interference although at smaller magnitudes. As for the cross-tier interference, the cell-edge pico eNBs transmitting on subband $A$ create interference for these users. Then, we can express the signal-to-interference-plus-noise ratio (SINR) of a cell-center MUE $k$ on
subcarrier $n$ as

$$
\gamma_k^{(n)} = \frac{P_M^{(n)} g_{k,b}^{(n)}}{\sum_{b' \in \mathcal{B}_M^A, b' \neq b} P_M^{(n)} g_{k,b'}^{(n)} + \sum_{b'' \in \mathcal{B}_P^A} P_P^{(n)} g_{k,b''}^{(n)} + N_0 \Delta_f} \tag{6.1}
$$

where $P_M^{(n)}$ and $P_P^{(n)}$ denote the downlink transmit powers of macrocell $M$ and picocell $P$ on subcarrier $n$, respectively. The channel gain between user $k$ and base station $b$ is denoted by $g_{k,b}^{(n)}$ on subcarrier $n$. Also, $\mathcal{B}_M^A$ and $\mathcal{B}_P^A$ denote the set of MeNBs and pico eNBs operating on subband $A$. The thermal noise power per Hz is denoted by $N_0$. The bandwidth of a subcarrier is represented as $\Delta_f$, $\Delta_f = 15$ kHz for LTE systems [103]. It is straightforward to derive the interference conditions for the cell-center MUEs in Sectors 2 and 3.

The SINR of the MUEs in the cell-edge regions can be obtained similarly. In this region, the number of interfering MeNBs in the first ring is reduced from six to two due to the frequency reuse of $1/3$. It comes from the sectorization at the MeNBs. As in the cell-center region, the picocells operating on subband $B$ still interfere with the cell-edge MUEs. The five MeNBs in the second ring also create interference, but relatively less compared to those from the first tier.
Let us now discuss the interference conditions for the PUEs. Assume that a PUE \(k\) is located in the cell-center region of Sector 1. This user can be scheduled on the resources over the subbands \(C\) and \(D\). The interference from the picocell tier comes from the pico eNBs operating over these subbands. These correspond to the pico eNBs within the same cell and those in the neighboring cells. The interference from the macrocell tier is mainly caused by the four MeNBs in the first ring, in which each MeNB creates interference for either one of these subbands. Note that the closest MeNB does not interfere with the pico eNBs in the cell-center region as they are allocated to different subbands. Hence, this spectrum allocation significantly decreases the number of interfering MeNBs for the picocell tier, and thereby reduces the cross-tier interference. The SINR of a PUE in the cell-center region scheduled to the subcarriers in subband \(C\) is

\[
\gamma_k^{(n)} = \frac{P_p^{(n)} g_{k,b}^{(n)}}{\sum_{b' \in B_M^C} P_M^{(n)} g_{k,b'}^{(n)} + \sum_{b'' \in B_P^C, b'' \neq b} P_P^{(n)} g_{k,b''}^{(n)} + N_0 \Delta_f}
\]  

(6.2)

where \(B_M^C\) and \(B_P^C\) denote the set of MeNBs and pico eNBs operating on subband \(C\). Similar expressions can be obtained for the other PUEs, in which the only differences will be the interfering base stations.

Constant power allocation over subbands is considered in this chapter, which follows the standards for LTE systems [103]. The proposed algorithm in Section 6.3.1 introduces two power control parameters \(\beta\) and \(\varepsilon\) to adjust the downlink transmit power levels. The first parameter \(\beta\) scales the transmit power of each MeNB. The second parameter \(\varepsilon\) denotes the ratio of the power allocated to the subcarriers in the cell-edge region to those in the cell-center region. By this way, \(\varepsilon\) characterizes the fairness between the MUEs in the cell-center and cell-edge regions. Let \(N_A\), \(N_B\), \(N_C\), and \(N_D\) denote the total number of subcarriers in subbands \(A\), \(B\), \(C\), and \(D\), respectively. Then, for Sector 1, the maximum transmit power
of an MeNB satisfies

$$\beta P_{\text{max},M} = P_M N_A + \varepsilon P_M N_B,$$ (6.3)

and therefore, $P_M = \beta P_{\text{max},M} / (N_A + \varepsilon N_B)$, where $P_M$ and $P_{\text{max},M}$ are the MeNB transmit power per subcarrier in the cell-center region and the MeNB maximum transmit power, respectively. The term $\varepsilon P_M$ denotes the transmit power per subcarrier for the MUEs in the cell-edge region. Similar expressions can be obtained for Sectors 2 and 3 by replacing $N_B$ with $N_C$ and $N_D$, respectively. For completeness, we also express the picocell transmit power per subcarrier. For a pico eNB in Sector 1, the transmit power per subcarrier, denoted by $P_P$, depends on the picocell’s location in the cell, and it is given by

$$P_P = \begin{cases} P_{\text{max},P} / (N_C + N_D) & \text{if } d_p \leq r_{th} \\ P_{\text{max},P} / (N_A + N_C + N_D) & \text{if } d_p > r_{th} \end{cases}$$ (6.4)

where $P_P$ and $P_{\text{max},P}$ are the transmit power of a pico eNB per subcarrier and the maximum transmit power of a pico eNB, respectively. The distance between the closest MeNB and pico eNB is denoted by $d_p$. Similar expressions can be obtained for Sectors 2 and 3.

### 6.2.1 Base Station Power Consumption Models

Recent studies have quantified the energy consumption of a base station down to the component level and several power consumption models have been proposed, see, e.g., [5, 117, 131]...
These models include the contributions of the power amplifier, radio frequency (RF) transceiver parts, baseband unit, power supply, and cooling devices [5]. Using these models each component’s contribution can be identified and efficient methods can be developed to introduce energy savings [2]. In this chapter, we study the load-dependent power consumption model presented in [5], which is

\[
P_{\text{Total}} = \begin{cases} 
N_{\text{TRX}} (P_0 + \Delta \cdot P_{TX}) & 0 < P_{TX} \leq P_{\max} \\
N_{\text{TRX}} P_{\text{sleep}} & P_{TX} = 0 
\end{cases} 
\]  

(6.5)

where \(P_{\text{Total}}\) and \(P_{TX}\) are the overall base station power consumption and RF transmit output powers, respectively. \(N_{\text{TRX}}\) is the number of transceiver chains, \(P_0\) is the power consumption at the minimum non-zero output power, and \(\Delta\) is the slope of the load-dependent power consumption. \(P_{\text{sleep}}\) denotes the power consumption of the sleep mode. Notice that the power consumption at a base station depends on the RF transmit power, \(P_{TX}\), and thereby this model is referred to as load-dependent power consumption model. Using this definition, the power consumption at an MeNB and a pico eNB can be expressed as

\[
P_{\text{Macro}} = N_{\text{TRX},M} (P_{0,M} + \Delta_M P_{TX,M}) \quad \text{and} \quad P_{\text{Pico}} = N_{\text{TRX},P} (P_{0,P} + \Delta_P P_{TX,P}) 
\]  

(6.6)

where \(P_{0,M}\), \(P_{0,P}\), \(P_{\text{Macro}}\), and \(P_{\text{Pico}}\) are the power consumption at the minimum non-zero output power and the total power consumption of the MeNBs and pico eNBs, respectively. \(N_{\text{TRX},M}\) and \(N_{\text{TRX},P}\) represent the number of transceiver chains at the MeNBs and pico eNBs, respectively. The corresponding slopes of the load-dependent power consumption are denoted as \(\Delta_M\) and \(\Delta_P\), in the same order as before. Note that (6.6) is true for \(0 < P_{TX,M} \leq P_{\max,M}\) and \(0 < P_{TX,P} \leq P_{\max,P}\), where \(P_{\max,M}\) and \(P_{\max,P}\) are the maximum RF transmit power for MeNBs and pico eNBs, respectively. If \(P_{TX,M} = 0\) (or \(P_{TX,P} = 0\)), then \(P_{\text{Macro}} = N_{\text{TRX},MP_{\text{sleep},M}}\) (or \(P_{\text{Pico}} = N_{\text{TRX},PP_{\text{sleep},P}}\)), where \(P_{\text{sleep},M}\) and \(P_{\text{sleep},P}\) are the power consumption of the sleep modes of MeNBs and pico eNBs, respectively. Table 6.1
presents the corresponding parameter values of the linearized power consumption model for various base station types.

6.3 Energy-Efficient Resource Allocation Problem

In this section, we formulate a non-cooperative resource allocation problem in OFDM systems employing the FFR method. Our objective is to maximize the energy efficiency per sector by determining the optimal RB allocation and power level assignment on each subband. In the sequel, we define the energy efficiency per sector, formulate the problem, and present its complexity analysis. Then, we proceed to propose our algorithm, along with its complexity analysis, optimality conditions, and convergence analysis.

Let \( R_k(\gamma_k) \) denote the throughput of user \( k \) that depends on its SINR \( \gamma_k \). Also, let \( \mathcal{K}_{M,i}^C \), \( \mathcal{K}_{M,i}^X \), \( \mathcal{K}_{P,i}^C \), and \( \mathcal{K}_{P,i}^X \) denote the set of MUEs in sector \( i \) connected to the MeNB in the cell-center and cell-edge regions, and the set of PUEs connected to the pico eNBs in the cell-center and cell-edge regions of this sector, respectively. Note that the subscript \( i \) denotes the sector indices. Then, the energy efficiency per sector \( i \) is given by

\[
\eta_{EE,i} = \frac{\sum_{k \in \mathcal{K}_{M,i}^C \cup \mathcal{K}_{M,i}^X} R_k(\gamma_k) + \sum_{k \in \mathcal{K}_{P,i}^C \cup \mathcal{K}_{P,i}^X} R_k(\gamma_k)}{\psi_i}
\]  

(6.7)

where the total power consumed per sector \( i \) is denoted by \( \psi_i \) which can be expressed as

\[
\psi_i = N_{TRX,M,i} (P_{0,M} + \Delta_M P_{TX,M}) + N_{\text{picos}} N_{TRX,P,i} (P_{0,P} + \Delta_P P_{TX,P})
\]  

(6.8)

and where \( N_{\text{picos}} \) is the number of pico eNBs in sector. The energy efficiency is given in units of bits/Joule.
The energy-efficient resource allocation problem can be formulated as

$$\max_{\mathbf{F}, \mathbf{P}_M, \mathbf{P}_P} \eta_{EE,i}$$

(6.9)

where \( \mathbf{F} \) denotes the RB allocation vector, and \( \mathbf{P}_M \) and \( \mathbf{P}_P \) are the MeNB and pico eNB power assignment vectors. The solution requires a joint search over the frequency and power domains. It is shown in [107] that the optimal solution is the multilevel water-filling solution. However, finding the optimal RB assignments among \( K \) users and \( N_{RB} \) RBs requires \( K^{N_{RB}} \) searches [107]. Therefore, this approach is impractical for real applications and it can lead to large latencies in practice as there are more users in the system. In the next subsection, we present our proposed algorithm that divides the resource allocation problem into two stages decoupling the frequency and power allocation problems.

### 6.3.1 Proposed Solution

Obtaining the instantaneous interference conditions of the complete network is often impractical for real applications due to excessive traffic overhead it would require. Therefore, recent studies have focused on non-cooperative or clustered base station resource allocation algorithms in multicell systems, see, e.g., [85, 86], and Chapter 11 of [103]. In this chapter, we investigate a non-cooperative solution in which each MeNB sector maximizes its own energy efficiency. We assume that there is a fast and reliable information exchange between the MeNB and the pico eNBs in the same sector such that the channel conditions of the PUEs are known at the MeNB. In LTE, this is exchanged over the X2 interface [103].

The proposed algorithm starts with determining cell-center region boundaries. We compare two different algorithms for this purpose, one of which maximizes the energy efficiency, while the other ensures a fair distribution of resources which is presented here as a reference. Then, we decouple the frequency and power allocation problems into two stages. In the first stage,
we solve the frequency assignment problem. Once these are obtained, in the second stage, we assign the power levels that maximize the energy efficiency at each sector.

**Setting The Cell-Center Region Boundaries**

The cell-center region radius per sector, $r_{th}$, is an important design parameter that affects the system performance in OFDMA systems employing the FFR method. It determines the set of MUEs in the cell-center and cell-edge regions. We assume that MeNBs have perfect knowledge about the locations of pico eNBs. Once the cell-center radius is determined, it is easy to identify the pico eNBs in each region. Considering the spatial and temporal variations of the user distribution in each cell, this parameter needs to be dynamically adjusted per sector. We propose an algorithm to determine the cell-center region boundaries and compare its performance to another algorithm where the frequency resources are distributed proportionally among the users in the same sector.

In the first algorithm, the MUE with the highest reference signal received power (RSRP) measurement, that is typically the closest MUE to the MeNB, is selected to be in the cell-center region, while the rest of the MUEs are in the cell-edge region. The SINR of this user is, in general, expected to be greater than the other MUEs. Consequently, the cell-center region subbands are allocated to only one MUE maximizing its throughput. In general, this algorithm achieves the maximum throughput and energy efficiency at the cost of system fairness. We refer to this algorithm as Adaptive $R_{th}$ Algorithm 1.

In the second algorithm, the cell-center region boundary is determined such that the ratio of cell-center and cell-edge MUEs is proportional to the ratio of the subbands allocated in these regions. First, the MUEs are sorted in ascending order based on their path losses. Let $N^C_M$ and $N^X_M$ denote the number of subcarriers that the MeNB uses in the cell-center and cell-edge regions. For $K$ MUEs in sector $i$, we round $(N^C_M/(N^C_M + N^X_M) \cdot K)$ to the nearest
integer, and \([0.5 + \frac{N_C^G}{M} (N_C^G + N_X^G) \cdot K]\) users are considered in the cell-center region. The rest of the users are assigned to the cell-edge subbands. This achieves higher fairness, but it comes at the cost of a decrease in the sector throughput. We refer to this algorithm as Adaptive \(R_{th}\) Algorithm 2. Note that a similar method is proposed in [119] to determine the cell-center boundaries.

In terms of implementation, MUEs are differentiated as cell-center and cell-edge users using the RSRP measurements. If the RSRP of a user is higher than a threshold, the user is considered to be in the cell-center region. Alternatively, the reference signal received quality (RSRQ) measurements can also be used depending on the system design. Especially, when the shadow fading effects are significant, the RSRQ measurements, which can be considered as the wideband SINR, can provide a better correlation with the user-experienced SINR. On the other hand, it is shown in [120] that both schemes perform almost the same. In this chapter, we employ the RSRP measurements to determine cell-center and cell-edge users due to simplicity. For example, with the Adaptive \(R_{th}\) Algorithm 1, the user with the highest RSRP is determined to be in the cell-center.

Frequency Assignment Problem

In order to solve the frequency assignment problem we study two schedulers. First, we study the SRM scheduler discussed in [102]. In this scheduler, the RBs are assigned to users such that the throughput is maximized. This scheduler is investigated to maximize the throughput, although this comes at the cost of a decrease in the system fairness. The second scheduler we study is the EBW scheduler. Consider \(K\) users sharing \(N_{RB}\) RBs. Then, \(K_h = \text{mod} (N_{RB}, K)\) users get \([N_{RB}/K] + 1\) RBs, whereas \(K_i = K - K_h\) users are given \([N_{RB}/K]\) RBs. Although the EBW scheduler does not necessarily maximize the energy efficiency per sector, it is proposed in [11] to calibrate system level simulations. This scheduler serves as a point of reference to a fair distribution of resources, which is used to quantify the scheduling
gain in Section 6.5. Note that although these two schedulers are commonly studied in the literature, they do not guarantee satisfaction of the QoS constraints of every user. These constraints can be addressed in the resource allocation problem using QoS-aware schedulers, see, e.g., [105]. In QoS-aware schedulers, the multi-user frequency assignments are carried out to satisfy constraints such as the user’s guaranteed bit rate requirement, packet error rate of different data traffic classes, latency, etc.

Power Assignment Problem

The second stage of the proposed algorithm solves the power control problem by assigning the optimal power levels to the subbands. The proposed algorithm uses the gradient ascent method to solve this problem. First we observe that by controlling the transmissions for the cell-center MUEs, we also determine how much interference is created for the cell-edge PUEs. Similarly, the downlink transmissions for the cell-edge MUEs determine the interference for the cell-center PUEs. In order to capture these two effects, we introduce two variables into the optimization problem as $\beta$ and $\varepsilon$, as discussed in Section 6.2.

Consider the following function $\eta_i(\varepsilon, \beta)$ that only includes the throughput of users in sector $i$ who are affected by the optimization variables $\varepsilon$ and $\beta$. Those users are the MUEs in both regions and the PUEs in the cell-edge region. While calculating $\eta_i(\varepsilon, \beta)$ only the interference created within each sector is considered. The energy efficiency function in sector $i$ can be modified as

$$
\eta_i(\varepsilon, \beta) = \frac{\sum_{k \in \mathcal{K}_{M,i}} R_k(\gamma_k) + \sum_{k \in \mathcal{K}_{X,i}} R_k(\gamma_k) + \sum_{k \in \mathcal{K}_{P,i}} R_k(\gamma_k)}{\psi_i}
$$

(6.10)

where the power consumed in sector $i$ is denoted by $\psi_i$. We modify (6.8) to account for the
power control parameter $\beta$ such that $\psi_i$ can be expressed as

$$\psi_i = N_{TRX,M,i}(P_{0,M} + \Delta_M \beta P_{\text{max},M}) + N_{\text{picos}}N_{TRX,P,i}(P_{0,P} + \Delta_P P_{\text{max},P}). \quad (6.11)$$

Hence, $\beta$ can be used to introduce energy savings in the total RF transmit power.

The energy efficiency per sector definition in (6.10) can be expanded as

$$\eta_i(\varepsilon, \beta) = \Delta f \sum_{k \in \mathcal{K}_M} \sum_{n \in \mathcal{N}_{Mk}} \log_2 \left( 1 + \frac{\beta P_{\text{max},M} g_{k,m}^{(n)}}{(N_0^C + \varepsilon N_X^C) I_k^{(n)}} \right) + \sum_{k \in \mathcal{K}_P} \sum_{n \in \mathcal{N}_{Pk}} \log_2 \left( 1 + \frac{\beta P_{\text{max},P} g_{k,p}^{(n)}}{(N_0^P + \varepsilon N_X^P) N_0 \Delta f} \right)$$

$$+ \Delta f \frac{\sum_{k \in \mathcal{K}_P} \sum_{n \in \mathcal{N}_{Pk} \cap \mathcal{N}_{Mk}} \log_2 \left( 1 + \frac{P_{\text{max},P} g_{k,p}^{(n)} / N_0^P}{\beta P_{\text{max},M} g_{k,M}^{(n)} / N_0^C + N_0 \Delta f} \right)}{\psi_i} \quad (6.12)$$

where $I_k^{(n)}$ is the interference from the picocells using subband $A$ in cell-edge region plus the thermal noise effective over a subcarrier at the PUE. The expression $n \in \mathcal{N}_{Pk} \cap \mathcal{N}_{Mk}$ in (6.12) denotes the subcarriers that the downlink transmissions of the cell-center MUEs create interference for the cell-edge PUEs, which are the subcarriers in subband $A$ for the FFR scheme in Fig. 6.1. It needs to be emphasized that (6.12) considers only the intra-cell interference and not the inter-cell interference. This enables fast implementation as it does not necessitate information exchange between MeNBs and asynchronous implementation at each MeNB sector. For that reason, this type of formulation is robust against inter-cell backhaul transmission delays.

The optimization problem that maximizes the energy efficiency per sector can be written as

$$\max_{\varepsilon, \beta} \eta_i(\varepsilon, \beta) \quad (6.13)$$

s.t. $\varepsilon \geq 1$ and $0 \leq \beta \leq 1$.

The first constraint is to favor the MUEs in the cell-edge region such that they are transmitted
at least $\varepsilon$ times the power allocated for the MUEs in the cell-center region. This parameter also affects the interference incurred at the cell-edge PUEs. The second constraint scales the total RF transmit power of the MeNB and sets the boundary conditions. Hence, the variable $\beta$ not only determines the interference, but it also introduces energy savings to the system.

### 6.3.2 Optimality Conditions

The Lagrangian of the problem in (6.13) can be written as

$$\mathcal{L}(\varepsilon, \beta, \lambda) = \eta_i(\varepsilon, \beta) + \lambda_1(\varepsilon - 1) + \lambda_2\beta + \lambda_3(1 - \beta)$$

where $\lambda_1$, $\lambda_2$, and $\lambda_3$ are the Lagrange multipliers and $\lambda = (\lambda_1, \lambda_2, \lambda_3)$. Note that if $(\varepsilon^*, \beta^*)$ solves (6.13), then $\eta_i(\varepsilon^*, \beta^*) \geq \eta_i(\varepsilon, \beta)$ for all $\varepsilon \geq 1$ and $0 \leq \beta \leq 1$. Furthermore, there exists $\lambda^* \geq 0$ such that the following optimality conditions are satisfied

$$\begin{align*}
\frac{\partial \mathcal{L}(\varepsilon^*, \beta^*, \lambda^*)}{\partial \varepsilon} &= \frac{\partial \eta_i(\varepsilon^*, \beta^*)}{\partial \varepsilon} + \lambda_1^* = 0, \\
\frac{\partial \mathcal{L}(\varepsilon^*, \beta^*, \lambda^*)}{\partial \beta} &= \frac{\partial \eta_i(\varepsilon^*, \beta^*)}{\partial \beta} + \lambda_2^* - \lambda_3^* = 0,
\end{align*}$$

and the complementary slackness conditions are

$$\begin{align*}
\lambda_1^*(\varepsilon^* - 1) &= 0, \\
\lambda_2^*\beta^* &= 0, \\
\lambda_3^*(1 - \beta^*) &= 0, \quad \text{and} \quad \lambda_1^*, \lambda_2^*, \lambda_3^* \geq 0.
\end{align*}$$

The equations in (6.15)-(6.16) are commonly known as the KKT conditions [64]. It needs to be emphasized that the power control parameters $\varepsilon$ and $\beta$ depend on the number of RBs allocated to the cell-center and cell-edge regions, the channel conditions, the maximum transmit powers, and the bandwidth of each subcarrier.
6.3.3 Gradient Ascent Method

The gradient ascent method starts at an initial \((\varepsilon, \beta)\) value evaluated at time \(t\) and the parameters \(\varepsilon\) and \(\beta\) are updated

\[
\varepsilon_{t+1} = \varepsilon_t + \mu_t \nabla_\varepsilon \eta_i(\varepsilon_t, \beta_t) \quad \text{and} \quad \beta_{t+1} = \beta_t + \mu_t \nabla_\beta \eta_i(\varepsilon_t, \beta_t)
\]

(6.17)

where \(\varepsilon_{t+1}\) and \(\beta_{t+1}\) are the updated values at time \(t + 1\), respectively. \(
\nabla_\varepsilon \eta_i(\varepsilon, \beta) = \partial \eta_i(\varepsilon, \beta) / \partial \varepsilon \) and \(\nabla_\beta \eta_i(\varepsilon, \beta) = \partial \eta_i(\varepsilon, \beta) / \partial \beta\) are the partial derivatives of \(\eta_i\) with respect to \(\varepsilon\) and \(\beta\), respectively, evaluated at time \(t\). These partial derivatives are multiplied by a sufficiently small and positive step size \(\mu_t\). The step size at each iteration is chosen according to the Armijo rule [64]. In this rule, the step size is chosen as \(\mu_t = \mu_0^m s\), where \(s\) is a constant and \(m\) is the first non-negative integer that satisfies the following inequality

\[
\eta_i(\varepsilon_{t+1}, \beta_{t+1}) - \eta_i(\varepsilon_t, \beta_t) \geq \rho \mu_t \nabla \eta_i(\varepsilon_t, \beta_t)^T d_t
\]

(6.18)

where \(\rho\) is a fixed constant and \(d_t\) is a feasible direction. The gradient is shown by \(\nabla \eta_i(\varepsilon, \beta) = [\nabla_\varepsilon \eta_i(\varepsilon, \beta) \ \nabla_\beta \eta_i(\varepsilon, \beta)]^T\), where \([::]^T\) denotes the transpose operator. Starting from \(m = 0\), it is successively increased until (6.18) is satisfied. Note that the gradient ascent method without the Armijo rule can fail to converge to a stationary point as illustrated in [64, p. 26], but when the step size is determined with the Armijo rule, it is guaranteed that the energy efficiency per sector increases per iteration until the algorithm converges. This enables selecting the increment that sufficiently improves the current objective value.

Typical values of these constants are such that \(\rho \in [10^{-5}, 10^{-1}]\) and \(\mu_0 \in [0.1, 0.5]\) [64]. The directional vector \(d_t\) is an ascent direction if it satisfies \(\nabla \eta_i(\varepsilon_t, \beta_t)^T d_t > 0\) if \(\nabla \eta_i(\varepsilon_t, \beta_t) \neq 0\), and \(\nabla \eta_i(\varepsilon_t, \beta_t)^T d_t = 0\) if \(\nabla \eta_i(\varepsilon_t, \beta_t) = 0\). In this chapter, we consider the steepest descent method, that is \(d_t = \nabla \eta_i(\varepsilon_t, \beta_t)^T\). The expressions of the partial derivatives \(\nabla_\varepsilon \eta_i(\varepsilon, \beta)\) and \(\nabla_\beta \eta_i(\varepsilon, \beta)\) can be found in Chapter 6.7.
6.3.4 Convergence Analysis

In what follows, we investigate the convergence of the proposed algorithm. To this end, we first show the quasiconcavity of the objective function $\eta_i$ with respect to the optimization variables $\varepsilon$ and $\beta$. Then, we study the optimality of the solutions obtained by the gradient ascent method.

**Definition 6.1.** A function $f$ is called quasiconcave if its domain, denoted by $\text{dom} f$, is convex and for any $x, y \in \text{dom} f$,

$$f(\phi x + (1 - \phi)y) \geq \min\{f(x), f(y)\}$$ (6.19)

where $0 \leq \phi \leq 1$ [121]. Similarly, a function $f$ is called strictly quasiconcave if it satisfies (6.19) with strict inequality for $x \neq y$ and $0 < \phi < 1$ [121].

**Proposition 1.** (First-Order Characterization) Let $f(x)$ be a continuously differentiable function on an open and convex set $D \subset \mathbb{R}^n$. Then, $f$ is quasiconcave if and only if $f(y) \geq f(x)$ implies $\nabla f(x)^T(y - x) \geq 0$, for all $x, y \in D$ [121].

**Proposition 2.** (Second-Order Characterization) Let $f(x)$ be a twice differentiable function and its dimensions be denoted by $n$. If $f(x)$ is quasiconcave, then $y^T \nabla^2 f(x)y \leq 0$ holds for all $x \in \text{dom} f$ and $y \in \mathbb{R}^n$ satisfying $y^T \nabla f(x) = 0$ [121].

**Lemma 1.** Let $D$ be a nonempty convex set and $f$ be a strictly quasiconcave function. Then, any local maximum is a global solution of the problem $P = \sup\{f(x)|x \in D\}$ [121].

A proof of Lemma 1 can be found in [121]. In cases where the concavity (or similarly convexity) of the problem cannot be assumed, the results of Lemma 1 are important to determine global extreme points. In fact, these are directly applicable to problems in several research fields such as those in economics. We refer the reader to Appendix C.6 of [121] for an example problem in economics, the standard consumer demand problem in a deterministic
framework. It needs to be emphasized that care needs to be taken in applying Lemma 1. The local extreme point of a quasiconcave (or quasiconvex) problem should not be confused with the stationary points as for determining the global extreme point. Next, using the definitions and lemmas above, we present Lemma 2, whose proof is provided in Chapter 6.7.

**Lemma 2.** The energy efficiency per sector expression $\eta_i$ is strictly quasiconcave in $\varepsilon$ and $\beta$.

**Assumption 2.** The proposed update rules for $\varepsilon$ and $\beta$ in (6.17) converge to the global optimal solution of the problem (6.13) as $t \to \infty$ for a sufficiently small step size $\mu_t$.

**Proof.** It follows from Lemma 1 that if there exists a local solution to the maximization problem, then there is a global maximum. It is straightforward to show that the gradient ascent method with sufficiently small step sizes converges to the globally maximum solution.

---

### 6.4 Energy-Efficient Resource Allocation with Pricing

In this section, we extend the preceding energy-efficient resource allocation and introduce pricing methods to achieve an upper bound to the non-cooperative energy efficiency maximization problem. Pricing methods are shown to be efficient in wireless communications [90–94,115]. In particular, we investigate the ADP framework and employ the interference pricing function. This function is motivated by the fact that without the penalty functions, base stations act selfish and transmit without any inter-cell interference concerns. When the pricing schemes are introduced, resources can be more efficiently shared among selfish players and the inter-cell interference can be reduced so that a better overall solution can be
achieved. Consider that an MeNB $i$ solves the following maximization problem

$$\max_{\epsilon, \beta} \eta_i(\epsilon, \beta) - \theta_i(\epsilon, \beta)$$

s.t. $\epsilon \geq 1$ and $0 \leq \beta \leq 1$.  \hspace{1cm} (6.20)

where the energy efficiency function, $\eta_i(\epsilon, \beta)$, is penalized with the interference pricing cost function, $\theta_i(\epsilon, \beta)$. This problem has the same boundary constraints as in (6.13). Depending on the system design, various penalty functions can be defined, see the survey in [115]. In this chapter, we take the inter-cell interference into consideration. The interference pricing function is defined as [92–94]

$$\theta_i(\epsilon, \beta) = \sum_{n \in \mathcal{N}_M} \sum_{j \in \mathcal{K}^{(n)}, j \neq k} p_k^{(n)} \pi_j^{(n)}/g_{i,j}^{(n)}$$

(6.21)

where $\mathcal{N}_M$ denotes set of subcarriers used in the cell-center and cell-edge regions, that is $\mathcal{N}_M^C \cup \mathcal{N}_M^X \subseteq \mathcal{N}_M$. The transmit power of an MeNB $i$ to user $k$ on subcarrier $n$ is represented by $p_k^{(n)}$, i.e., $P_M$ on the subcarriers used in the cell-center region and $\epsilon P_M$ for those in the cell-edge region. The set of users assigned to subcarrier $n$ in the system is denoted by $\mathcal{K}^{(n)}$ and the notation $j \in \mathcal{K}^{(n)}, j \neq k$ denotes any user that MeNB $i$ interferes on subcarrier $n$. The interference caused by MeNB $i$ to user $j$ on subcarrier $n$ is denoted by $g_{i,j}^{(n)}$. The interference prices are given by $\pi_j^{(n)} = -\partial u_j^{(n)}(\gamma_j^{(n)})/\partial I_j^{(n)}$, where $u_j^{(n)}(\gamma_j^{(n)})$ and $I_j^{(n)}$ denote the utility and the incurred interference of user $j$ on subcarrier $n$, respectively. The KKT conditions of the problem in (6.20) are similar to those in (6.15)-(6.16) except for the interference terms of each user also includes the inter-cell interference. Note that using the results in [92–94], it can be shown that the solution of the problem in (6.20) is guaranteed to converge to a point which satisfies the KKT conditions, although global optimality is not guaranteed. Despite the fact that interference pricing can be implemented in an asynchronous and distributive manner, it requires each base station to acquire the interference prices of all the other base stations on all subcarriers. This means that, based on the FFR scheme in Fig. 6.1,
\(57(N_X^M + N_Y^M)\) prices need to be distributed at each subframe through the backhaul, which is hard to implement in practice. Therefore, this is presented as an upper bound for energy efficiency in the non-cooperative energy efficiency maximization problem without pricing. Note that in terms of implementation, the pricing algorithm requires both the intra-cell and inter-cell information exchange among base stations, whereas the proposed algorithm without pricing only necessitates the intra-cell information. Both of these can be conveyed through the X2 interface for LTE systems.

6.5 Numerical Results

In this section, the performance of the proposed algorithm is evaluated. We quantify the individual contributions of power control, frequency scheduling, cell-center region radius, spectrum allocation, and pricing. We compare the performance of the proposed algorithm to the orthogonal and cochannel spectrum allocation, and demonstrate the achievable gains. In addition, we investigate three algorithms for determining the cell-center region boundaries. We investigate the Adaptive \(R_{th}\) Algorithms 1 and 2, described in Section 6.3.1. Also, we consider the case with the fixed FFR boundaries, in which the cell radius that maximizes the average energy efficiency is chosen through enumeration methods. Note that this analysis is presented in Figs. 6.5(a)-(b). Based on this analysis, we find that when the cell-center region radius is taken as 0.3 times the cell radius, it maximizes the average energy efficiency. In terms of spectrum allocation, MeNBs and pico eNBs transmit over all subcarriers in the cochannel allocation, whereas the spectrum is divided into 32 and 18 non-overlapping RBs for the MeNBs and pico eNBs, respectively, in the orthogonal channel allocation. For the FFR method, there are 32 RBs in the subband \(A\) and 6 RBs are allocated to subbands \(B\), \(C\), and \(D\) [111]. The same spectrum allocation is employed for the no power control FFR algorithm case, abbreviated as No PC in Figs. 6.3 and 6.4. We compare our proposed algorithm
Table 6.2: Simulation Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel bandwidth</td>
<td>10 MHz</td>
</tr>
<tr>
<td>Total number of data RBs</td>
<td>50 RBs</td>
</tr>
<tr>
<td>Freq. selective channel model (CM)</td>
<td>Extended Typical Urban CM</td>
</tr>
<tr>
<td>UE to MeNB PL model</td>
<td>$128.1 + 37.6 \log_{10}(d)$</td>
</tr>
<tr>
<td>UE to Pico eNB PL model</td>
<td>$140.7 + 36.7 \log_{10}(d)$</td>
</tr>
<tr>
<td>Effective thermal noise power, $N_0$</td>
<td>$-174 \text{dBm/Hz}$</td>
</tr>
<tr>
<td>UE noise figures</td>
<td>9 dB</td>
</tr>
<tr>
<td>MeNB and Pico eNB antenna gain</td>
<td>14 dB and 5 dB</td>
</tr>
<tr>
<td>UE antenna gain</td>
<td>0 dBi</td>
</tr>
<tr>
<td>Antenna horizontal pattern, $A(\theta)$</td>
<td>$-\min(12(\theta/\theta_{3\text{dB}})^2, A_m)$</td>
</tr>
<tr>
<td>$A_m$ and $\theta_{3\text{dB}}$</td>
<td>20 dB and 70°</td>
</tr>
<tr>
<td>Penetration loss</td>
<td>20 dB</td>
</tr>
<tr>
<td>Macro- and picocell shadowing</td>
<td>8 dB and 10 dB</td>
</tr>
<tr>
<td>Inter-site distance</td>
<td>500 m</td>
</tr>
<tr>
<td>Minimum macro- to user distance</td>
<td>50 m</td>
</tr>
<tr>
<td>Minimum pico- to user distance</td>
<td>10 m</td>
</tr>
<tr>
<td>Minimum pico- to macro- distance</td>
<td>75 m</td>
</tr>
<tr>
<td>Minimum pico- to pico- distance</td>
<td>40 m</td>
</tr>
<tr>
<td>Traffic model</td>
<td>Full buffer</td>
</tr>
</tbody>
</table>

with different power control algorithms and spectrum allocations. In particular, we compare the proposed algorithm with the power control algorithms proposed in [108] and [112], and spectrum allocation schemes of cochannel, orthogonal channel, and FR schemes. First, we have implemented the constant power allocation algorithm proposed in [108] using the FFR scheme in Fig. 6.1. This algorithm maximizes the sector sum rate and achieves the smallest duality gap. However, as we will demonstrate shortly, it does not perform well in an interference-dominated multicell environment. Second, we evaluate the performance of the algorithm proposed in [112] which employs the power control parameters as $(\varepsilon, \beta) = (4, 1)$ and the FFR scheme in Fig. 6.1. For different spectrum allocation schemes, we implement the cochannel and orthogonal frequency allocation in which base stations transmit at maximum power. To demonstrate the gain of FFR over FR spectrum allocation, we have also implemented the FR scheme using the proposed power control algorithm. In the FR scheme, the system bandwidth is divided into three subbands in which each MeNB sector uses one of the three subbands, while the remaining two subbands are allocated to the picocell tier.

The simulation layout is shown in Fig. 6.1. It assumes a HetNet deployment with 19 hexagonal cells in which MeNBs are employed with 3-sector antennas. In each sector, there is
Figure 6.2: Illustration of the energy efficiency of a sector and the proposed algorithm solutions using the gradient ascent method.

A single antenna, i.e., $N_{TRX,M,i} = 1, \forall i$. For the pico eNBs, omnidirectional antennas are employed, i.e., $N_{TRX,P,i} = 1, \forall i$. There are 4 randomly placed pico eNBs in each sector. In order to observe the clustering effects, we consider nonuniform user distribution and generate 30 users per sector. First, we place one user per pico eNB within a 40 meter radius, while the rest of the users are randomly generated within the sector area. The users are associated to the base stations with the highest reference signal received power (RSRP) method [12]. While generating the pico eNBs and users, several minimum distance constraints are considered and these are presented in Table 6.2, along with the other parameters and simulation models used in our numerical results. These parameters and channel models are in accordance with [11] for the baseline simulation of HetNets. Also, we consider $\rho = 10^{-2}$ and $\mu_0 = 0.05$ for the gradient ascent method.

Fig. 6.2 illustrates the energy efficiency of a sector for different $(\varepsilon, \beta)$ pairs. The improvement of the proposed algorithm at each iteration is denoted by the red circles. Armijo rule is implemented to select the step sizes. As studied in Section 6.3.1, this rule guarantees that energy efficiency increases at every iteration until it converges. Fig. 6.2 shows that the energy efficiency of the network is mostly affected by $\beta$. On the other hand, the effects of $\varepsilon$ on energy efficiency are minor for the same $\beta$. This is expected since $\beta$ determines
the macrocell transmit power level which directly affects the total consumed power in the network, whereas $\varepsilon$ does not change the total consumed power, but its effects are mostly observed on the total throughput.

In Figs. 6.3(a)-(b) and Figs. 6.4(a)-(b), we investigate the individual contributions of the methods to determine the cell-center boundaries, frequency scheduling, and power control gains for different schedulers. In Figs. 6.3(a)-(b) the EBW scheduler is studied and Figs. 6.4(a)-(b) depict the results for the SRM scheduler. The average energy efficiency of 57 sectors is plotted in Fig. 6.3(a) and Fig. 6.4(a), and the average sum throughput of sectors

Figure 6.3: Average energy efficiency per sector (a) and average sector throughput (b) are depicted for the EBW scheduler.

Figure 6.4: Average energy efficiency per sector (a) and average sector throughput (b) are depicted for the SRM scheduler.
is depicted in Fig. 6.3(b) and Fig. 6.4(b). The proposed algorithm starts at full transmit power and iteratively updates $\varepsilon$ and $\beta$ along the derivative using the update rule in (6.17). The initial values are chosen as $(\varepsilon_0, \beta_0) = (2, 1)$. It can be observed that both the energy efficiency and throughput increase monotonically per iteration.

Let us first identify the power control gain. In Fig. 6.3(a), when we compare the power control and no power control cases of the FFR spectrum allocation with EBW scheduler, it can be observed that the energy efficiency of the FFR spectrum allocation increases 2.08 times with the power control. In addition, we observe from Fig. 6.3(b) that power control brings a 36% throughput increase. Similar gains are observed with the SRM scheduler in Figs. 6.4(a)-(b) such that the energy efficiency gain is 2.12 times and the throughput gain is 38%. These gains are due to adjusting the downlink transmit power and reducing both the intra- and inter-cell interference in the network. When we evaluate the performance of the algorithm in [108], we observe that it provides negligible gains compared to the case without power control in both metrics. This is due to the fact that the selfish behavior of base stations impairs the network energy efficiency and throughput in an interference dominated region. Notice that in Figs. 6.3(a)-(b) and Figs. 6.4(a)-(b) both the power control and no power control curves start almost at the same points and at each iteration we observe the power control curves monotonically increase their values indicating the power control gain.

Second, we analyze the effects of frequency scheduling on the energy efficiency. Comparing Fig. 6.3(a) and Fig. 6.4(a), we observe that the energy efficiency of the FFR scheme with the SRM scheduler is 11% better than that of the EBW scheduler. The throughput gain of these two schedulers is similar. This shows that the energy efficiency gain is mostly related to the throughput gain of the scheduler and the scheduler type has a small effect on the power consumption.

Third, we analyze the performance of spectrum allocation schemes in two-tiers. We observe that employing the EBW scheduler, the fixed FFR scheme outperforms the cochannel allo-
cation by 25% and the orthogonal channel allocation by 59% in terms of energy efficiency. Similar gains are observed in the throughput performance as well. For the SRM scheduler, the cochannel allocation method has 12% and 29% better performance in both metrics compared to the fixed FFR method and orthogonal channel allocation, respectively. This result highlights the importance of the scheduler in use for different spectrum allocations. We note that with power control, FFR performs the best.

Fourth, we compare the FR scheme to the FFR spectrum allocation. We employ the proposed power control method and the SRM scheduler in both cases. As previously mentioned, two types of FR schemes in a two-tier network are investigated. From Figs. 6.4(a)-(b), we observe that the energy efficiency of FR1 and FR2 are 403 and 262 kbits/Joule, respectively, whereas it is 506 kbits/Joule for the FFR allocation with adaptive cell-center regions. The average sector throughput of FR1, FR2, and FFR spectrum allocations are 142.2, 92.8, and 176.6 Mbits/sec, respectively. These results show that the FFR spectrum allocation provides 26 – 93% more energy efficient transmissions and 24 – 90% more throughput per sector compared to the FR spectrum allocation. These results also show that FFR is an effective spectrum allocation method, which effectively utilizes the available spectrum by reducing the interference and increasing the achievable throughput. Similar conclusions can be drawn based on Figs. 6.3(a)-(b).

Fifth, we study the effects of the cell-center region boundaries. The proposed Adaptive $R_{th}$ Algorithm 1 has similar energy efficiency and throughput performance compared to the fixed radius FFR method with $r_{th} = 0.3R$ in both schedulers. However its performance is significantly higher compared to other constant radii values as depicted in Figs. 6.5(a)-(b). In addition, it can be expected that Adaptive $R_{th}$ Algorithm 1 significantly outperforms the fixed FFR solution in a dynamic scenario. When compared to the algorithm in [112], Adaptive $R_{th}$ Algorithm 1 outperforms the one in [112] by factors of 2.9 and 2.5 in terms of energy efficiency for the EBW and SRM schedulers, respectively. These gains are due
to implementing adaptive cell-center boundaries and better power allocation. Despite the poor performance of Adaptive $R_{th}$ Algorithm 2 in terms of energy efficiency, it provides a fair distribution of resources among MUEs. For example, Jain’s fairness index of Adaptive $R_{th}$ Algorithm 2 using the EBW scheduler is 0.5, whereas it is 0.025 with the Adaptive $R_{th}$ Algorithm 1 using the same scheduler.

Finally, we compare the single-tier and two-tier deployments using the FFR scheme with the proposed power control method and the SRM scheduler. From Figs. 6.4(a)-(b), we observe that the energy efficiency of the single-layer network is 380 bits/Joule and it increases to 506 bits/Joule when picocells are deployed, indicating a 1.3x gain. Moreover, the sector throughput increases from 55 to 175 Mbits/sec from a single-layer to two-tier HetNet deployment, respectively, corresponding to a 3.2x gain. These results demonstrate the substantial gains that can be achieved with the picocell deployment. Again, similar conclusions can be drawn based on Figs. 6.3(a)-(b), in fact with even larger gains.

Fig. 6.5(a) shows the average energy efficiency per sector for different cell-center radii using the EBW scheduler. The performance of the proposed adaptive $R_{th}$ algorithms is also
presented. We observe that the performance of the FFR system strictly depends on the cell-center region boundaries. When the fixed radius FFR methods are considered and different radii values are enumerated, the system energy efficiency varies between 139 and 439 kbits/Joule. Our simulation results show that the proposed Adaptive $R_{th}$ Algorithm 1 outperforms all other FFR methods in both metrics. On the other hand, Adaptive $R_{th}$ Algorithm 2 performs significantly worse as it prioritizes fairness rather than throughput. Note that there is a 3x gain between the two algorithms in terms of energy efficiency which illustrates the energy efficiency and fairness trade-off.

Fig. 6.5(b) compares the average sector throughput per sector of the fixed fractional cell-center radii to the adaptive $R_{th}$ algorithms for the EBW scheduler. Similar to the energy efficiency, throughput performance strictly depends on the cell-center region boundaries such that the average sector throughput varies between 56 and 153 Mbps for different constant cell-center radii values. An important result is that the cell-center radius that maximizes the throughput is the same as that maximizes the energy efficiency. Note that, with another power consumption model, these two radii can take different values.

Figs. 6.5(a)-(b) also illustrate the effects of the pricing on the energy efficiency and average sector throughput, respectively. Despite the extra information required, the pricing provides only 0.7% and 0.4% improvements in terms of energy efficiency with the Adaptive $R_{th}$ Algorithm 1 using the EBW and SRM schedulers, respectively. The reason for this is as follows. Due to the nature of the FFR method, the inter-cell interference on subbands $B$, $C$, and $D$ are significantly suppressed and the pricing method on these bands does not improve the performance of the algorithm remarkably. Note that with the Adaptive $R_{th}$ Algorithm 1, when the number of users in the sector increases, the cell-center radius gets closer to the MeNB, and the inter-cell interference to this user is very limited. On the other hand, we observe that with the Adaptive $R_{th}$ Algorithm 2, pricing provides higher gains. The energy efficiency improvements are 13% and 8% for EBW and SRM schedulers with pricing, respec-
tively. This is due to the fact that since the cell-center radius is closer to cell edge than it is for the Adaptive $R_{th}$ Algorithm 1, the inter-cell interference becomes more critical and the pricing gain is higher.

6.6 Conclusion

Energy consumption of a wireless network is a serious concern for the next generation cellular networks. To address this problem, we have proposed an energy-efficient resource allocation algorithm for HetNets with the FFR scheme. The proposed algorithm decouples the frequency and power allocation problems and successively solves each of them. It employs the gradient ascent method to solve the power allocation problem. Based on our simulations, we show that the proposed algorithm significantly improves both the energy efficiency and throughput. We also quantify the individual contributions of the effects of cell-center region, power control, and frequency scheduling gains in order to provide design guidelines. It is demonstrated that the proposed power control algorithm provides the most significant gains, while moderate gains can be achieved with the SRM scheduler and Adaptive $R_{th}$ Algorithm 1. An upper bound based on the interference pricing method is also investigated and the performance gap is shown to be only marginal. Finally, we show that significant energy savings are possible with the proposed algorithm which reduces the operational expenditures for the network operators.
6.7 Derivations

In what follows, we prove that $\eta_i(\varepsilon, \beta)$ is quasiconcave in $\varepsilon$ and $\beta$. It follows from Proposition 2 that $\eta_i(\varepsilon, \beta)$ is a quasiconcave function if and only if the following holds

$$y^T \nabla \eta_i(\varepsilon, \beta) = 0 \text{ and } y^T \nabla^2 \eta_i(\varepsilon, \beta) y \leq 0$$

(6.22)

where $y = [y_1 \ y_2]^T$. First, we introduce new definitions for the proof, and then express the first and second order derivatives. Let $R_i(\varepsilon, \beta)$ denote the aggregate throughput of sector $i$ as

$$R_i(\varepsilon, \beta) = \sum_{k \in K^C_M} \sum_{n \in N^C_Mk} R_{1}^{(k,n)}(\varepsilon, \beta) + \sum_{k \in K^X_M} \sum_{n \in N^X_Mk} R_{2}^{(k,n)}(\varepsilon, \beta) + \sum_{k \in K^P_M} \sum_{n \in N^P_k \cap N^X_M} R_{3}^{(k,n)}(\varepsilon, \beta)$$

(6.23)

where

$$R_{1}^{(k,n)}(\varepsilon, \beta) = \log \left(1 + \frac{\beta a}{b + \varepsilon c}\right), \quad R_{2}^{(k,n)}(\varepsilon, \beta) = \log \left(1 + \frac{\varepsilon \beta d}{b + \varepsilon c}\right)$$

$$R_{3}^{(k,n)}(\varepsilon, \beta) = \log \left(1 + \frac{f}{b + \varepsilon c} + h\right)$$

(6.24)

where $a = P_{\max,n} g_{k,m}^{(n)} I_k^{(n)}$, $b = N^C_M$, $c = N^X_M$, $d = P_{\max,n} g_{k,m}^{(n)} (N_0 \Delta f)$, $f = P_{\max,p} / N^P_M g_{k,m}^{(n)}$, $g = P_{\max,M} g_{k,m}^{(n)}$, and $h = N_0 \Delta f$. We denote $R_i(\varepsilon, \beta)$, $R_{1}^{(k,n)}(\varepsilon, \beta)$, $R_{2}^{(k,n)}(\varepsilon, \beta)$, and $R_{3}^{(k,n)}(\varepsilon, \beta)$ by $R_i$, $R_{1}^{(k,n)}$, $R_{2}^{(k,n)}$, and $R_{3}^{(k,n)}$, respectively. Using these definitions, the first derivative of $R_i$ with respect to $\varepsilon$ can be expressed as

$$\frac{\partial R_i}{\partial \varepsilon} = \sum_{k \in K^C_M} \sum_{n \in N^C_Mk} \frac{\partial R_{1}^{(k,n)}}{\partial \varepsilon} + \sum_{k \in K^X_M} \sum_{n \in N^X_Mk} \frac{\partial R_{2}^{(k,n)}}{\partial \varepsilon} + \sum_{k \in K^P_M} \sum_{n \in N^P_k \cap N^X_M} \frac{\partial R_{3}^{(k,n)}}{\partial \varepsilon}$$

(6.25)
where

$$\frac{\partial R_{1}^{(k,n)}}{\partial \varepsilon} = -\frac{ac\beta}{(b + c\varepsilon)(a\beta + b + c\varepsilon)}, \quad \frac{\partial R_{2}^{(k,n)}}{\partial \varepsilon} = \frac{bd\beta}{(b + c\varepsilon)(b + \varepsilon(c + d\beta))},$$

$$\frac{\partial R_{3}^{(k,n)}}{\partial \varepsilon} = \frac{c\beta}{(h(b + c\varepsilon) + g\beta)((f + h)(b + c\varepsilon) + g\beta)}.$$

The second derivative of $R_i$ with respect to $\varepsilon$ is given by

$$\frac{\partial^2 R_i}{\partial \varepsilon^2} = \sum_{k \in K} \sum_{n \in N_{M_k}} \frac{\partial^2 R_{1}^{(k,n)}}{\partial \varepsilon^2} + \sum_{k \in K} \sum_{n \in N_{\Delta M_k}} \frac{\partial^2 R_{2}^{(k,n)}}{\partial \varepsilon^2} + \sum_{k \in K} \sum_{n \in N_{P_k} \cap N_{M_k}} \frac{\partial^2 R_{3}^{(k,n)}}{\partial \varepsilon^2} (6.27)$$

where

$$\frac{\partial^2 R_{1}^{(k,n)}}{\partial \varepsilon^2} = \frac{\beta ac^2 (a\beta + 2(b + c\varepsilon))}{(b + c\varepsilon)^2(a\beta + b + c\varepsilon)^2},$$

$$\frac{\partial^2 R_{2}^{(k,n)}}{\partial \varepsilon^2} = -\frac{bd\beta}{(b + c\varepsilon)^2(b + \varepsilon(c + d\beta))^2}((f + h)(b + c\varepsilon) + g\beta),$$

$$\frac{\partial^2 R_{3}^{(k,n)}}{\partial \varepsilon^2} = \frac{c^2fg\beta}{(h(b + c\varepsilon) + g\beta)^2((f + h)(b + c\varepsilon) + g\beta)}.$$

Similarly, the first derivative of $R_i$ with respect to $\beta$ is

$$\frac{\partial R_i}{\partial \beta} = \sum_{k \in K_{\Delta M}} \sum_{n \in N_{\Delta M_k}} \frac{\partial R_{1}^{(k,n)}}{\partial \beta} + \sum_{k \in K_{\Delta M}} \sum_{n \in N_{\Delta M_k}} \frac{\partial R_{2}^{(k,n)}}{\partial \beta} + \sum_{k \in K_{\Delta M}} \sum_{n \in N_{\Delta M_k}} \frac{\partial R_{3}^{(k,n)}}{\partial \beta} (6.29)$$

where

$$\frac{\partial R_{1}^{(k,n)}}{\partial \beta} = \frac{a}{a\beta + b + c\varepsilon}, \quad \frac{\partial R_{2}^{(k,n)}}{\partial \beta} = \frac{d\varepsilon}{b + \varepsilon(c + d\beta)},$$

$$\frac{\partial R_{3}^{(k,n)}}{\partial \beta} = -\frac{fg(b + c\varepsilon)}{(h(b + c\varepsilon) + g\beta)((f + h)(b + c\varepsilon) + g\beta)}. (6.30)$$
The second derivative of \( R_i \) with respect to \( \beta \) is given by

\[
\frac{\partial^2 R_i}{\partial \beta^2} = \sum_{k \in K} \sum_{n \in N_{M_k}} \frac{\partial^2 R_i^{(k,n)}}{\partial \beta^2} + \sum_{k \in K} \sum_{n \in N_{M_k}^X} \frac{\partial^2 R_i^{(k,n)}}{\partial \beta^2} + \sum_{k \in K} \sum_{n \in N_{P_k} \cap N_{M_k}^C} \frac{\partial^2 R_i^{(k,n)}}{\partial \beta^2} \tag{6.31}
\]

where

\[
\frac{\partial^2 R_i^{(k,n)}}{\partial \beta^2} = -\frac{a^2}{(a \beta + b + c \varepsilon)^2}, \quad \frac{\partial^2 R_i^{(k,n)}}{\partial \beta^2} = -\frac{d^2 \varepsilon^2}{(b + \varepsilon (c + d \beta))^2}, \\
\frac{\partial^2 R_i^{(k,n)}}{\partial \beta^2} = \frac{fg^2 (b + c \varepsilon)((f + 2h) (b + c \varepsilon) + 2g \beta)}{(h (b + c \varepsilon) + g \beta)^2((f + h) (b + c \varepsilon) + g \beta)^2}. \tag{6.32}
\]

The derivative of \( R_i \) with respect to \( \varepsilon \) and \( \beta \) is

\[
\frac{\partial^2 R_i}{\partial \varepsilon \partial \beta} = \sum_{k \in K} \sum_{n \in N_{M_k}^X} \frac{\partial^2 R_i^{(k,n)}}{\partial \varepsilon \partial \beta} + \sum_{k \in K} \sum_{n \in N_{M_k}^X} \frac{\partial^2 R_i^{(k,n)}}{\partial \varepsilon \partial \beta} + \sum_{k \in K} \sum_{n \in N_{P_k} \cap N_{M_k}^C} \frac{\partial^2 R_i^{(k,n)}}{\partial \varepsilon \partial \beta} \tag{6.33}
\]

where

\[
\frac{\partial^2 R_i^{(k,n)}}{\partial \varepsilon \partial \beta} = -\frac{ac}{(a \beta + b + c \varepsilon)^2}, \quad \frac{\partial^2 R_i^{(k,n)}}{\partial \varepsilon \partial \beta} = \frac{bd}{(b + \varepsilon (c + d \beta))^2}, \\
\frac{\partial^2 R_i^{(k,n)}}{\partial \varepsilon \partial \beta} = \frac{cg (f + h) h (b + c \varepsilon)^2 - g^2 \beta^2)}{(h (b + c \varepsilon) + g \beta)^2((f + h) (b + c \varepsilon) + g \beta)^2}. \tag{6.34}
\]

The gradient of \( \eta(\varepsilon, \beta) \) can be expressed as

\[
\nabla \eta(\varepsilon, \beta) = \left( \begin{array}{c}
\frac{\partial \eta(\varepsilon, \beta)}{\partial \varepsilon} \\
\frac{\partial \eta(\varepsilon, \beta)}{\partial \beta}
\end{array} \right) = \left( \begin{array}{c}
\frac{\partial R_i}{\partial \varepsilon} \psi \\
\frac{\partial R_i}{\partial \beta} \frac{1}{\psi} - \frac{R_i}{\psi^2} \frac{\partial \psi}{\partial \beta}
\end{array} \right) \tag{6.35}
\]

where \((\partial R_i / \partial \varepsilon) = (\partial R_1 / \partial \varepsilon) + (\partial R_2 / \partial \varepsilon) + (\partial R_3 / \partial \varepsilon) \) and \((\partial R_i / \partial \beta) = (\partial R_1 / \partial \beta) + (\partial R_2 / \partial \beta) + (\partial R_3 / \partial \beta)\).

Consider that \( y^T \nabla \eta = 0 \) is satisfied and use (6.35) to rearrange terms, we have \((\partial R_i / \partial \varepsilon) y_1 + \)
\[(\partial R_i/\partial \beta) y_2 = 1/\psi (\partial \psi/\partial \beta) R_i y_2.\] The Hessian of \(\eta_i\) is given by

\[
\nabla^2 \eta_i(\varepsilon, \beta) = \frac{1}{\psi} \left( \begin{array}{ccc}
\frac{\partial^2 R_i}{\partial \varepsilon^2} & \frac{\partial^2 R_i}{\partial \varepsilon \partial \beta} & \frac{\partial^2 R_i}{\partial \varepsilon \partial \beta} \\
\frac{\partial^2 R_i}{\partial \varepsilon \partial \beta} & \frac{\partial R_i \partial \psi}{\partial \varepsilon \partial \beta} & \frac{\partial R_i \partial \psi}{\partial \varepsilon \partial \beta} \\
\frac{\partial R_i}{\partial \varepsilon} & \frac{\partial R_i}{\partial \beta} & \Phi
\end{array} \right)
\] (6.36)

where

\[
\Phi = \frac{\partial^2 R_i}{\partial \beta^2} - \frac{2}{\psi} \frac{\partial R_i \partial \psi}{\partial \beta} - \frac{R_i}{\psi} \left( \frac{\partial^2 \psi}{\partial \beta^2} \right)^2 + 2 \frac{R_i}{\psi^2} \left( \frac{\partial \psi}{\partial \beta} \right)^2.
\] (6.37)

For the power consumption model in (6.11), \(\partial^2 \psi/\partial \beta^2 = 0\). When we expand the terms, we get

\[
y^T \nabla^2 \eta_i(\varepsilon, \beta)y = \frac{\partial^2 R_i}{\partial \varepsilon^2} y_1^2 + 2 \frac{\partial^2 R_i}{\partial \varepsilon \partial \beta} y_1 y_2 + \frac{\partial^2 R_i}{\partial \beta^2} y_2^2.
\] (6.38)

Substituting the condition \(y^T \nabla \eta_i = 0\) in (6.38) and rearranging terms, we have

\[
y^T \nabla^2 \eta_i y = \frac{1}{\psi} \left( \frac{\partial^2 R_i}{\partial \varepsilon^2} y_1^2 + 2 \frac{\partial^2 R_i}{\partial \varepsilon \partial \beta} y_1 y_2 + \frac{\partial^2 R_i}{\partial \beta^2} y_2^2 \right) = \frac{1}{\psi} y^T \nabla^2 R_i y.
\] (6.39)

where \(\nabla^2 R_i\) denotes the Hessian of \(R_i\). This means that \(\eta_i\) is quasiconcave if and only if \(R_i\) is quasiconcave. Next, we analyze the quasiconcavity of \(R_i\). Using in (6.28), (6.32), and (6.34) in (6.39) to obtain \(y^T \nabla^2 R_i y \leq 0\), we need to show that

\[
\sum_{k \in K^C \cap N^C_k} \sum_{n \in N^C_k} \left( \frac{\partial^2 R_1^{(n)}}{\partial \varepsilon^2} y_1^2 + 2 \frac{\partial^2 R_1^{(n)}}{\partial \varepsilon \partial \beta} y_1 y_2 + \frac{\partial^2 R_1^{(n)}}{\partial \beta^2} y_2^2 \right) \\
+ \sum_{k \in K^X \cap N^X_k} \sum_{n \in N^X_k} \left( \frac{\partial^2 R_2^{(n)}}{\partial \varepsilon^2} y_1^2 + 2 \frac{\partial^2 R_2^{(n)}}{\partial \varepsilon \partial \beta} y_1 y_2 + \frac{\partial^2 R_2^{(n)}}{\partial \beta^2} y_2^2 \right) \\
+ \sum_{k \in K^X \cap N^X_k \cap N^C_k} \left( \frac{\partial^2 R_3^{(n)}}{\partial \varepsilon^2} y_1^2 + 2 \frac{\partial^2 R_3^{(n)}}{\partial \varepsilon \partial \beta} y_1 y_2 + \frac{\partial^2 R_3^{(n)}}{\partial \beta^2} y_2^2 \right) \leq 0.
\] (6.40)

Notice that the summations in (6.40) are grouped into subcarriers based on their locations within the cell. The contributions of the cell-center and cell-edge MUEs are captured in the
first two summations in (6.40), while the third is of the cell-edge PUEs. For the proof, it is necessary to show that the sum of these summations is non-positive.

First, we identify the condition for the first summation in (6.40), that is the terms for the cell-center MUEs, as

\[
\frac{\beta ac^2 (a\beta + 2(b + c\epsilon))}{(b + c\epsilon)^2} y_1^2 - 2acy_1 y_2 - a^2 y_2^2 \leq 0. \tag{6.41}
\]

Notice that this is the condition that relates to the cell-center MUEs.

Next, we investigate the condition for the second summation in (6.40) to hold, that is

\[
-\frac{bd\beta ((b + c\epsilon) (2c + d\beta) + cd\epsilon\beta)}{(b + c\epsilon)^2} y_1^2 + 2bdy_1 y_2 - d^2\epsilon^2 y_2^2 \leq 0 \tag{6.42}
\]

When we rearrange terms, we have the necessary condition for the cell-edge MUEs such as

\[
2bd (b + c\epsilon)^2 y_1 y_2 \leq bd\beta ((b + c\epsilon) (2c + d\beta) + cd\epsilon\beta) y_1^2 + d^2\epsilon^2 (b + c\epsilon)^2 y_2^2. \tag{6.43}
\]

Finally, we see that the condition for the cell-edge PUEs is

\[
-c^2 fg\beta (2h (f + h) (b + c\epsilon) + g\beta (f + 2h)) y_1^2 + 2c fg ((f + h) h (b + c\epsilon)^2 - g^2 \beta^2) y_1 y_2 + fg^2 (b + c\epsilon) ((f + 2h) (b + c\epsilon) + 2g\beta) y_2^2 \leq 0. \tag{6.44}
\]

In our extensive simulations, we have observed that (6.41)-(6.44) are always satisfied. This makes us conjecture that \( R_i \) and \( \eta_i \) are negative semi-definite, as we have observed numerically. Therefore, we conclude that the function \( \eta_i \) should be quasiconcave in \( \epsilon \) and \( \beta \).
Chapter 7

Conclusions and Future Work

In general, radio resource management for the next-generation wireless networks is a fertile research area with many open research problems. Next-generation networks are expected to be more densely deployed, there will be more users with high data rate and strict quality of service demands, and to make facts even more critical, emerging technologies such as Internet of Things will increase the number of nodes that needs to be served in the system. Considering these aspects, it can be expected that this resource allocation problem will maintain its importance. In this dissertation, we proposed several methods and algorithms towards achieving these goals. In what follows, we itemize the future directions that can stem from this dissertation and several open problems that still need to be solved:

- In Chapter 1, we have discussed several research projects that address the network energy efficiency problem such as EARTH and GreenTouch. When we talk about energy efficiency, in this dissertation, we have used the units of J/bit or J/bit/sec. This metric reflects the capacity and data rate aspects of the network. When the data rates increase, so does the energy efficiency metric. However, this metric does not ensure coverage or service area aspects of the network. To include this constraint,
it was identified as a result of the EARTH project that the units of $W/m^2$ can be used [4]. This means that we can cast an alternative problem to find the optimal strategy maximizing the network energy efficiency, or minimizing the network power consumption, for a given coverage or service area. It is straightforward to extend the framework we have developed in Chapters 5-6 to this energy efficiency metric. It is clear that this metric would also provide valuable information to the network operators and this is a problem that is worth looking at.

- In Chapters 5-6, we have proposed several methods to increase the energy-efficiency of wireless cellular networks through better power control algorithms. Especially, we have identified that when base stations do not selfishly optimize their own utilities and rather cooperate with each other, they can easily reach to a higher “social” equilibrium. In these processes, we have assumed some information exchange through a fast and reliable backhaul connecting the base stations. However, one may raise two implementation related questions: First, what happens when the information transferred on the backhaul fails to make it on time? The answer for this question essentially depends how fast the channel conditions in the system vary and this is related to the user mobility conditions. If there are many highly mobile users in the system, the outdated channel state information may incur high penalties reducing the achievable gains. Under this scenario, multi-user water-filling algorithm is more prone to channel estimation errors due to different power levels on every subcarrier. It is likely that the tolerance to the outdated channel state information is higher with the constant-power allocation algorithm, which makes it more robust against channel estimation errors compared to the water-filling algorithm. However, in both cases, the outdated channel state information will degrade the performance. Second problem one needs to account for is how much gain loss is incurred through quantized information bits. The answer for this question obviously depends on the answers for the following items: How many bits are used for quantization? What is the capacity of the backhaul? If the base
station power consumption model captures the power consumed on the backhaul, let us assume in the units of W/bits, then there would be a trade-off between the power consumption and energy-efficiency performance regarding the quantization error. It is clear that when we transmit with more bits on the backhaul, thereby consuming more power for information exchange, there will be less quantization error. For this case, we can formulate the problem to find the optimal number of quantization bits for interference pricing such that the energy efficiency is maximized considering the backhaul capacity.

- Throughout this dissertation, we considered an idealized and uniform 19-cell hexagonal cellular structure to evaluate our theoretical findings. However, in real-life deployments, this idealized and somewhat symmetric interference conditions for simulations may fail to provide all the insights that network operators need. The base station deployments can be highly irregular and user distributions can be non-uniform due the characteristics and terrain of the area of interest. A recently proposed alternative approach is to use stochastic geometry methods which model the generation of base stations and users based on probabilistic distributions. Some preliminary works modeling the user generation and interference models can be found in [122–127]. An even better solution is to use software tools that a network operator can input the real data (such as the terrain, the possible candidate locations for deployment, the user traffic patterns based on prior information, etc.). Despite some early works on this area such as [122, 123], it is still an open problem to identify which model, hexagonal grid or stochastic geometry, characterizes the network operators requirements the best. It is more likely that one model, which may fit best for some cities, can fail for others.

- Based on the simulation tools developed for this dissertation, it is possible to extend the full buffer traffic model to other traffic models capturing traffic models for VoIP, ftp, email, etc. In general, it is a common practice to include a mixture of traffic
models in a network. In LTE, several models have been proposed in the literature, see, e.g., [128, 129]. Through including these models, new constraints can be captured in the problem such as satisfying latency constraints and minimizing packet loss. This will clearly resemble a more realistic network traffic model. Compared to the problems studied in this dissertation, adding the traffic models will constrain the feasibility region and these will enable us to develop intelligent schedulers whose performance can possibly be improved through cooperating base stations over time, frequency, and even spatial domains. Thus, the joint problem of energy-efficient scheduling and power control in cooperating base stations is a very important open problem that offer many advantages for the network operators.

- Due to the scarcity of available frequency spectrum in the 2-5 GHz spectrum band and the increasing demand for higher data rates, The Federal Communications Commission of the United States (or, simply FCC) has considered using the millimeter-wave spectrum for next-generation wireless networks where higher frequency bands with larger the system bandwidths can be employed. In the millimeter-wave, the propagation loss is much higher and blockage due to objects is more severe. Also, the Doppler and multipath effects are also more severe. Considering the fact that next generation wireless systems will employ more antennas, importance of the energy-efficiency problem will prevail. Our problem formulations and proposed algorithms can be applied to the new developed channel characteristics modeling the millimeter-wave propagation and provide energy-efficient solutions.

- Lastly, one possible future work to investigate is to add the costs associated with deployment and operational expenses of base stations into the problem formulation, which will lead us to combine the deployment and resource allocation problems. Combining these two problems is important due to the following observation. In [130], the authors show that the small cell deployments can reduce the energy consumption
by 30%, but increase the network cost by a factor of 14%. They emphasize that out
of $22 billion that is spent to operational and maintenance bills annually, the global
annual electricity bills makes up approximately $10 billion, corresponding to its 45%.
These expenses exclude the rental and spectrum payments. Using models that include
operation, maintenance, and electricity costs, one can study new metrics such as cost
energy efficiency with the units of bits/Joule/$ and try to maximize this utility for
network operators. Thus, studying the energy savings along with cost levels to satisfy
the traffic load is an important, timely, and novel problem.
Bibliography


