From Orthogonal to Non-orthogonal Multiple Access: Energy- and Spectrum-Efficient Resource Allocation

Lei Lei

LiU LINKÖPING UNIVERSITY

Department of Science and Technology,
Linköping University, SE-601 74 Norrköping, Sweden

Norrköping 2016

**From Orthogonal to Non-orthogonal Multiple Access:
Energy- and Spectrum-Efficient Resource Allocation**
Lei Lei

Abstract

The rapid pace of innovations in information and communication technology (ICT) industry over the past decade has greatly improved people's mobile communication experience. This, in turn, has escalated exponential growth in the number of connected mobile devices and data traffic volume in wireless networks. Researchers and network service providers have faced many challenges in providing seamless, ubiquitous, reliable, and high-speed data service to mobile users. Mathematical optimization, as a powerful tool, plays an important role in addressing such challenging issues.

This dissertation addresses several radio resource allocation problems in 4G and 5G mobile communication systems, in order to improve network performance in terms of throughput, energy, or fairness. Mathematical optimization is applied as the main approach to analyze and solve the problems. Theoretical analysis and algorithmic solutions are derived. Numerical results are obtained to validate our theoretical findings and demonstrate the algorithms' ability of attaining optimal or near-optimal solutions.

Five research papers are included in the dissertation. In Paper I, we study a set of optimization problems of consecutive-channel allocation in single carrier-frequency division multiple access (SC-FDMA) systems. We provide a unified algorithmic framework to optimize the channel allocation and improve system performance. The next three papers are devoted to studying energy-saving problems in orthogonal frequency division multiple access (OFDMA) systems. In Paper II, we investigate a problem of jointly minimizing energy consumption at both transmitter and receiver sides. An energy-efficient scheduling algorithm is developed to provide optimality bounds and near-optimal solutions. Next in Paper III, we derive fundamental properties for energy minimization in load-coupled OFDMA networks. Our analytical results

suggest that the maximal use of time-frequency resources can lead to the lowest network energy consumption. An iterative power adjustment algorithm is developed to obtain the optimal power solution with guaranteed convergence. In Paper IV, we study an energy minimization problem from the perspective of scheduling activation and deactivation of base station transmissions. We provide mathematical formulations and theoretical insights. For problem solution, a column generation approach, as well as a bounding scheme are developed. Finally, towards to 5G communication systems, joint power and channel allocation in non-orthogonal multiple access (NOMA) is investigated in Paper V in which an algorithmic solution is proposed to improve system throughput and fairness.

Populärvetenskaplig Sammanfattning

Den snabba utvecklingen inom informations- och kommunikations teknikområdet har avsevärt förbättrat människors upplevelser av mobilkommunikation. Detta i sin tur har lett till en exponentiell ökning av antalet anslutna mobila enheter och mängden datatrafik i mobila nätverk. Forskare och nätverksoperatörer har stått inför många utmaningar för att tillhandahålla tjänster som är sömlösa, allestädes närvarande, pålitliga, och använder sig av höghastighetsdata för mobila enheter. För att ta itu med dessa problem kan matematisk optimering tillämpas för att tillhandahålla en generell metod och systematiska riktlinjer för att analysera och lösa dessa problem.

Den här avhandlingen fokuserar på att angripa radioresurs optimeringsproblem för fjärde och femte generationens (4G och 5G) mobila kommunikationssystem i syfte att optimera tilldelningen av den begränsade resursen frekvens/effekt/tid för att uppnå högsta möjliga prestanda. Vinsterna med att optimera resursallokeringen inkluderar förbättrad nätverkskapacitet, uppfyllandet av varierande prestandakrav, samt reducera kapital- och driftutgifter. De i avhandlingen angripna optimeringsproblemen kan kategoriseras i två klasser, energi- respektive spektrumeffektiv resursallokering. Den första klassen syftar till att minimera total energikonsumtion givet vissa prestandakrav, Den senare syftar till att maximera systemets genomströmning givet begränsad effektbudget samt att tillfredsställa krav på servicekvalitet.

Huvudsyftet med den här avhandlingen är att undersöka grundläggande egenskaper av resursallokeringsproblem för olika 4G och 5G kommunikationssystem. Vi studerar en uppsättning energi- och spektrumeffektiva resursallokeringsproblem. Matematisk optimering tillämpas som huvudstrategi för att analysera och lösa dessa problem. Vi tillhandahåller matematiska formuleringar och teoretisk förståelse för hur man kan optimera resursallokering. Baserat på vår teoretiska analys utvecklar vi algoritmer av hög kvalitet för att optimera prestandan. Numeriska resultat erhålls för att validera våra teoretiska resultat och demonstrera algoritmernas förmåga att uppnå optimala eller nästan optimala lösningar.

List of Publications

Included Papers

1. L. Lei, D. Yuan, C. K. Ho, and S. Sun, "A Unified Graph Labeling Algorithm for Consecutive-Block Channel Allocation in SC-FDMA," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5767-5779, Nov. 2013.

2. L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Resource Scheduling to Jointly Minimize Receiving and Transmitting Energy in OFDMA Systems," *Proceedings of IEEE International Symposium on Wireless Communication Systems (ISWCS)*, pp. 187-191, Aug. 2014.

3. C. K. Ho, D. Yuan, L. Lei, and S. Sun. "Power and Load Coupling in Cellular Networks for Energy Optimization," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 509-519, Jan. 2015.

4. L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Optimal Cell Clustering and Activation for Energy Saving in Load-Coupled Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6150-6163, Nov. 2015.

5. L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Power and Channel Allocation for Non-orthogonal Multiple Access in 5G Systems: Tractability and Computation," *IEEE Transactions on Wireless Communications*. (submitted Nov. 2015, revised Feb. 2016)

Additional Related Publications

The author also contributed to the following publications which are not included in this dissertation:

1. Y. Zhao, T. Larsson, D. Yuan, E. Rönnberg, L. Lei, "Power Efficient Uplink Scheduling in SC-FDMA: Benchmarking by Column Generation," *Journal of Optimization and Engineering*, preprint, 2015.

2. L. You, L. Lei, and D. Yuan, "Optimizing Power and User Association for Energy Saving in Load-Coupled Cooperative LTE," *IEEE International Conference on Communications (ICC)*, 2016.

3. M. Lei, X. Zhang, L. Lei, Q. He, and D. Yuan, "Successive Interference Cancellation for Throughput Maximization in Wireless Powered Communication Networks," Submitted to The 11th International Conference on Wireless Algorithms, Systems, and Applications (WASA), 2016.

4. L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Joint Optimization of Power and Channel Allocation with Non-orthogonal Multiple Access for 5G Cellular Systems," *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, 2015.

5. L. You, L. Lei, and D. Yuan, "Load Balancing via Joint Transmission in Heterogeneous LTE: Modeling and Computation," *Proceedings of IEEE Symposium on Personal, Indoor, Mobile and Radio Communications (PIMRC)*, 2015.

6. L. You, L. Lei, and D. Yuan, "A Performance Study of Energy Minimization for Interleaved and Localized FDMA," *Proceedings of IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2014.

7. C. K. Ho, D. Yuan, L. Lei, and S. Sun, "Optimal Energy Minimization in Load-Coupled Wireless Networks: Computation and Properties," *Proceedings of IEEE International Conference on Communications (ICC)*, 2014.

8. L. You, L. Lei, and D. Yuan, "Range Assignment for Power Optimization in Load-Coupled Heterogeneous Networks", *Proceedings of IEEE International Conference on Communication Systems (ICCS)*, 2014.

9. L. Lei, S. Fowler, and D. Yuan, "Improved Resource Allocation Algorithm Based on Partial Solution Estimation for SC-FDMA Systems," *Proceedings of IEEE Vehicular Technology Conference (VTC Fall)*, 2013.

10. H. Zhao, L. Lei, D. Yuan, T. Larsson, and E. Rönnberg, "Power Efficient Uplink Scheduling in SC-FDMA: Bounding Global Optimality by Column Generation," *Proceedings of IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2013.

11. L. Lei, V. Angelakis, and D. Yuan, "Performance Analysis of Chunk-based Resource Allocation in Wireless OFDMA Systems," *Proceedings of IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2012.

12. D. Yuan, V. Angelakis, and L. Lei, "Minimum-length Scheduling in Wireless Networks With Multi-user Detection and Interference Cancellation: Optimization and Performance Evaluation," *Proceedings of IEEE International Conference on Communication Systems (ICCS)*, 2012.

The dissertation is a continuation and an extension of the author's Licentiate thesis.

- L. Lei, "Radio Resource Optimization for OFDM-based Broadband Cellular Systems," Licentiate Thesis No. 1649, Linköping Studies in Science and Technology, Linköping University, 2014.

Acknowledgment

First and foremost, I would like to express my deep and sincere gratitude to my supervisor, Prof. Di Yuan, for giving me the opportunity to conduct my Ph.D. studies at Linköping University, and providing me excellent guidance and continuous support during theses years. I have learned many valuable lessons from such an outstanding researcher who always selflessly shares his research experience and expertise with me. This gratitude also goes to my co-supervisors, Assoc. Prof. Vangelis Angelakis and Dr. Erik Bergfeldt, for their kind support and guidance. The knowledge and the attitude on research I have learned from all of them will benefit me a lot in my future career development.

I would like to thank all the colleagues and former members at the Division of KTS, for creating such a friendly and pleasant working atmosphere. I am grateful to Assoc. Prof. Vangelis Angelakis, Assoc. Prof. Scott Fowler, and Dr. Lei Chen. Without their guidance in my early-stage Ph.D. studies, I may not carry out my first research work, academic presentation, paper, and teaching task smoothly. Thanks to Dr. Nikolaos Pappas for his thorough reading and detailed comments for improving the quality of this dissertation. Thanks to my office-mates Qing and Ioannis. I am so pleasant to share the office with them and have nice technical and non-technical talks. Also, many thanks to Viveka for providing me various administrative assistance.

I would also like to express my gratitude to Dr. Sumei Sun and Dr. Chin Keong Ho from Institute for Infocomm Research (I^2R) in Singapore, for providing me the opportunity and financial support to conduct my research work at I^2R, and for all the stimulating discussions and fruitful collaborations. I am also grateful to Prof. Antonio Capone, Prof. Xingjun Zhang, Prof. Xiaohu Ge, Mr. Stefano Napoli, Mr. George Vasilakis, and Dr. Sofoklis Kyriazakos, for hosting my research visiting in EC FP7 Marie Curie projects. With great thanks, I

In addition, I wish to thank all my dear friends in China, Sweden, and Singapore for their pleasant friendship, and for all the wonderful moments we have experienced together. Also, thanks to all of our football players in ITN and the city. I am enjoyable for all the games we have played during the past four years.

Last and most importantly, I would like to thank my wife Chong, and express my deepest gratitude to our parents, for all their always encouragement, support, and love.

<div style="text-align: right;">Norrköping, January 2016
Lei Lei</div>

Abbreviations

3GPP	The Third Generation Partnership Project
4G	The Fourth Generation
5G	The Fifth Generation
BIP	Binary Integer Programming
BS	Base Station
CG	Column Generation
CSI	Channel State Information
DP	Dynamic Programming
DRX	Discontinuous Reception
DTX	Discontinuous Transmission
ICI	Inter-Cell Interference
ICT	Information and Communication Technology
IFDMA	Interleaved FDMA
IP	Integer Programming
LFDMA	Localized FDMA
LP	Linear Programming
LTE	Long Term Evolution
LTE-A	LTE-Advance

MA	Multiple Access
MIMO	Multiple-Input Multiple-Output
MIP	Mix Integer Programming
MUSA	Multi User Shared Access
MUST	Multiuser Superposition Transmission
NLP	Nonlinear Programming
NOMA	Non-Orthogonal Multiple Access
NP	Non-Deterministic Polynomial
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
PAPR	Peak-to-Average Power Ratio
PDMA	Pattern Division Multiple Access
QoS	Quality of Service
RB	Resource Block
RC	Radio Components
RRO	Radio Resource Optimization
RU	Resource Unit
SC	Superposition Coding
SC-FDMA	Single Carrier Frequency Division Multiple Access
SCMA	Sparse Code Multiple Access
SIC	Successive Interference Cancellation

SINR	Signal-to-Interference-Plus-Noise Ratio
TTI	Transmission Time Interval
UE	User Equipment

Contents

Part I

Introduction and Overview

1   Introduction

1.1   Motivation

Over the past few decades, mobile communication systems have been successively evolved to the fourth generation (4G), i.e., Long Term Evolution (LTE) and LTE-advance (LTE-A). Even for a contemporary communication system, a fundamental issue, i.e., how to serve users' stringent data demands for mobile communication by using limited network resources, still exists. The issue stems from two aspects. On one side, the explosive growth in traffic data volume and number of connected devices will continue. From Cisco's annual visual network index reports, the number of broadband subscribers could reach tens billions by 2020, most of which are mobile devices [1]. Besides, mobile users' demand for high-speed data service is increasing exponentially, mainly driven by the advanced mobile devices and multimedia applications [2]. On the other side, with such tremendous growth, the scarcity of radio resources for cellular networks remains and becomes even more severe. The licensed frequency bands which are the scarce and expensive resource for network service providers, are limited within a narrow spectrum for the wireless communication in cellular networks, typically from several hundred megahertz (MHz) to few gigahertz (GHz) [3]. The physical spectrum has been heavily used and become crowded. It can be foreseen that the tremendous growth of data traffic and number of mobile devices could exhaust the capacity in existing cellular networks [4]. Moreover, this rapid growth has resulted in high energy consumption in cellular networks.

Energy consumption has become a serious concern for the entire information and communication technology (ICT) sector. First, the escalation of consumed energy in mobile communication systems indirectly causes a huge increase of greenhouse gas emission. It has been commonly recognized as a threat to environment and sustainable development. Second, the cost of high energy consumption is a heavy burden of capital expenditure and operational expenditure for network operators. It has been reported in [5] that over 70% of electricity bills for network operators come from the energy consumption in base stations (BSs).

Third, energy-efficient communication is crucial for battery-powered devices, e.g., smart phones, since higher energy consumption will drain the battery faster. Thus, the efforts for exploiting energy-saving potentials and improving energy efficiency for cellular networks are necessary and important.

For managing these issues of capacity and energy, advanced technologies and mobile communication systems have to be continuously developed and evolved, since the existing system will reach its performance limits and may not be able to address these challenges. The 4G LTE/LTE-A mobile communication system has been deployed in many countries [6, 7]. Towards the future, the fifth generation (5G) systems are expected to be deployed in the near future [2]. For each generation of the mobile communication systems, intelligent radio resource management is of importance in improving spectrum efficiency and reducing energy consumption. Mathematical optimization can be applied as a reliable and powerful tool to provide general methodology and systematic guidelines in analyzing and addressing problems.

Motivated by the importance and the arising challenges of spectrum efficiency and energy saving in 4G and 5G systems, this dissertation addresses several radio resource allocation problems for orthogonal frequency division multiple access (OFDMA) and single carrier-frequency division multiple access (SC-FDMA) systems in LTE networks, and non-orthogonal multiple access (NOMA) systems in 5G systems. The main objective of this dissertation is to investigate fundamental characteristics of the resource allocation problems, address the problems by optimization approaches, and provide high-quality algorithmic solutions to optimize system performance. The theoretical results and algorithmic ideas developed in this dissertation will shed light on the resource management for the future networks.

1.2   Dissertation Outline and Organization

The dissertation is divided into two parts. In Part I, we provide a general introduction to the addressed optimization problems, along with the related technologies and mathematical tools. Part II consists of five research papers. Part I is organized as follows. In Chapter 2, we introduce the multiple access technologies OFDMA and SC-FDMA in LTE,

and NOMA in 5G networks. In Chapter 3, we present radio resource optimization problems in OFDMA, SC-FDMA, and NOMA. In Chapter 4, we present the mathematical optimization tools which are used in this dissertation. In Chapter 5, we provide a short description of the contributions for each paper appended in Part II.

2 Multiple Access Technologies in LTE and Beyond

A standardized multiple access (MA) scheme is usually considered as the representative feature for a cellular system in each generation, e.g., code division multiple access in 3G, and OFDMA/SC-FDMA in 4G. An appropriate MA scheme enables massive mobile devices accessing the limited network resources efficiently and achieving supreme system performance. This dissertation has addressed several resource allocation problems for 4G and 5G networks. Next, we briefly introduce the basis of OFDMA, SC-FDMA, and NOMA.

2.1 Orthogonal Multiple Access in LTE

In 4G LTE systems, sometimes also referred to as LTE-A which was standardized in the third generation partnership project (3GPP) Release 10 [7], two advanced orthogonal multiple access (OMA) schemes, OFDMA and SC-FDMA, have been adopted as the standard MA schemes for downlink and uplink transmission, respectively [8]. Both MA schemes are considered as the appropriate technique to support users' diverse quality of service (QoS) requirements, exploit the flexible frequency granularity and achieve high spectral efficiency. The frequency bandwidth can be from 1.25 MHz to 20 MHz [9]. By adopting multiple-input multiple-output (MIMO), LTE-A is able to support a peak data rate in Gbps [3, 10].

2.1.1 OFDMA

In LTE downlink, OFDMA is based on the concept of multi-carrier transmission. In the frequency domain, the spectrum is divided into a large number of narrow-band subcarriers (or subchannels). The subcarrier bandwidth equals 15 kHz in both LTE downlink and uplink. The center frequency of each subcarrier is selected such that all the subcarriers are mathematically orthogonal to each other, and thus eliminates the interference between the adjacent subcarriers. The orthogonality avoids the need of separating the subcarriers by means of guard-bands, i.e.,

placing empty frequency bandwidth between adjacent subcarriers, and therefore saves the bandwidth resource. In the time domain, transmissions are organized into frames with length 10 milliseconds (ms) each. A frame is divided into 10 equally sized subframes of length 1 ms. Each subframe, corresponding to one transmission time interval (TTI), consists of two equal time slots of length 0.5 ms. Each time slot consists of six or seven orthogonal frequency division multiplexing (OFDM) symbols depending on the choice of cyclic prefix [3]. A basic resource unit (RU) in LTE is a resource block (RB) which consists of 12 subcarriers with a total bandwidth 180 KHz in the frequency domain and one 0.5 ms slot in the time domain [11]. Multiple user equipments (UEs) in a cellular network can transmit or receive data by using such time-frequency RBs.

Another advantage of OFDMA is its robustness in the presence of multipath fading. In data transmission, the high-speed data stream is divided into multiple substreams with lower data rate. These bitstreams are modulated into data symbols and transmitted simultaneously over different subcarriers. The bandwidth of each subcarrier is much smaller than the coherence bandwidth. Thus, each narrow-band subcarrier only experiences relatively flat fading with approximately constant channel gain during each TTI. This allows OFDMA to efficiently resist frequency-selective fading. More detailed discussions of OFDMA can be found in [8, 12].

2.1.2 SC-FDMA

In LTE uplink, one of the disadvantages of OFDMA is its high peak-to-average power ratio (PAPR) in transmitted OFDM signals, resulting in a need for a highly linear power amplifier [13]. High PAPR reduces the power efficiency and imposes a burden of power consumption on UE, and therefore shortens the battery life. This limitation is not a serious issue for dowlink transmission because of the availability of power supply at BSs. However, the power consumption is a major concern in uplink transmission since a mobile UE is usually limited by its battery capacity.

To overcome the disadvantage, SC-FDMA, a modified version of OFDMA, has been adopted as the standard MA scheme for LTE up-

link transmission. SC-FDMA has similar performance as OFDMA but with lower PAPR [14]. As in OFDMA, orthogonal subcarriers are used to transmit information symbols in SC-FDMA, but they are transmitted sequentially rather than in parallel as in OFDMA. This mechanism reduces the envelope fluctuations of the transmitted signal waveform, and thus offers lower PAPR. This property makes SC-FDMA more attractive for uplink transmission especially for low-cost equipments with limited power. Performance comparisons between OFDMA and SC-FDMA have been extensively investigated, see, e.g., [15, 16].
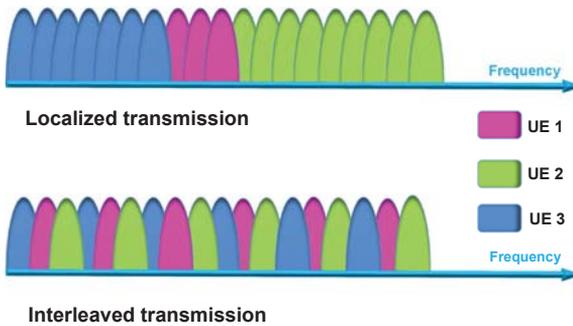

Figure 1: An illustration for localized FDMA and interleaved FDMA.

The subcarrier assignment among multiple UEs in SC-FDMA can be implemented by two subcarrier mapping schemes, i.e., localized FDMA (LFDMA) and interleaved FDMA (IFDMA) [17]. Figure 1 illustrates these two schemes. In LFDMA, each UE selects a set of consecutive subcarriers to transmit data. In IFDMA, the subcarriers occupied by a UE are distributed equidistantly over the entire frequency band. Detailed discussions of performance comparison between LFDMA and IFDMA can be found in [18, 19].

2.2 Non-orthogonal Multiple Access Towards 5G

With the deployment of commercial LTE networks worldwide, 4G is reaching maturity. Looking forward to the future, the rapid growth in

traffic data volume and the number of connected mobile devices, and the emergence of diverse application scenarios are still the main driving force to develop the next generation communication system [20]. In recent years, 5G has attracted extensive research and development efforts from the wireless communication community. The performance requirements of 5G systems have been firstly identified to adequately support wireless communications in future scenarios. It is widely accepted that, in comparison to LTE networks, 5G will be able to support 1000-fold gains in system capacity, peak data rate of fiber-like 10 Gbps and 1 Gbps for low mobility and high mobility, respectively, and at least 100 billion devices connections, ultra low energy consumption and latency [2, 4].

To fulfill these stringent requirements, the design of 5G network architecture will be different from LTE, and the current OMA schemes also need to be evolved. Several *non-orthogonal* MA schemes are under investigation for 5G. Compared to OMA in LTE, the new MA enables considerable performance improvements in system throughput and capacity of connecting mobile devices. Moreover, the non-orthogonal design of MA provides good backward compatibility with OFDMA and SC-FDMA [21]. In Release 13, 3GPP has initiated a study on downlink multiuser superposition transmission (MUST) for LTE [22], aiming at investigating multi-user non-orthogonal transmission, and the design of advanced receivers [23]. The concept of *non-orthogonal multiple access* is that the same frequency resource, e.g., subchannels, RBs, can be shared by multiple-user signals in the code or power domain, resulting in non-orthogonality among user access. By relying on advanced receivers, multi-user detection and successive interference cancellation (SIC) are applied for signal separation at the receiver side [24].

In this dissertation, we focus on a non-orthogonal MA scheme in the power domain. The concept is proposed in [25, 26, 27]. This scheme applies superposition coding (SC) to superpose multiple UEs' signals at the transmitter, and performs SIC at the receiver to separate and decode multi-user signals. Throughout this dissertation, we simply use "NOMA" to denote this power-domain non-orthogonal MA scheme. Figure 2 shows an illustration for single-cell OMA and NOMA in the power (as well as frequency) domain. In OMA, each UE has exclu-
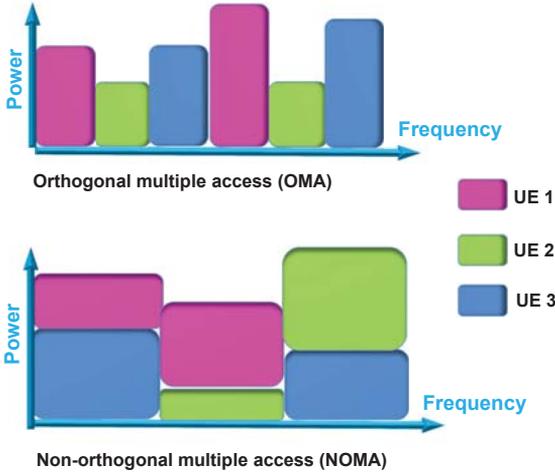
Figure 2: An illustration for OMA and NOMA.

sive access to the radio resource, whereas each subchannel in NOMA can accommodate more UEs. In OMA, the maximum number of UEs who can concurrently access the subchannels is limited by the number of subchannels. Compared to OMA, the number of the simultaneously multiplexed UEs in NOMA can be largely increased [28]. Dynamic switching between OMA and NOMA is considered in some works [25]. In practical scenarios, a hybrid scheme can be designed so that NOMA or OMA is only performed when it enables better performance over the other scheme.

In the following, we take a two-UE case in downlink to present the basics of SC and SIC in NOMA. As shown in Figure 3, a BS serves two UEs by using the same subchannel. UE 1 is geographically much closer to the BS than UE 2, thus we assume that use 1 has a stronger link to the BS, with better channel condition than UE 2. At the transmitter, the BS is supposed to transmit signals x_1 and x_2 for two UEs, respectively. After SC in NOMA, x_1 and x_2 are superposed to a signal x which is broadcasted to both receivers. The received signals at UE 1 and UE 2 are $y_1 = h_1 x + n_1$ and $y_2 = h_2 x + n_2$, respectively, where
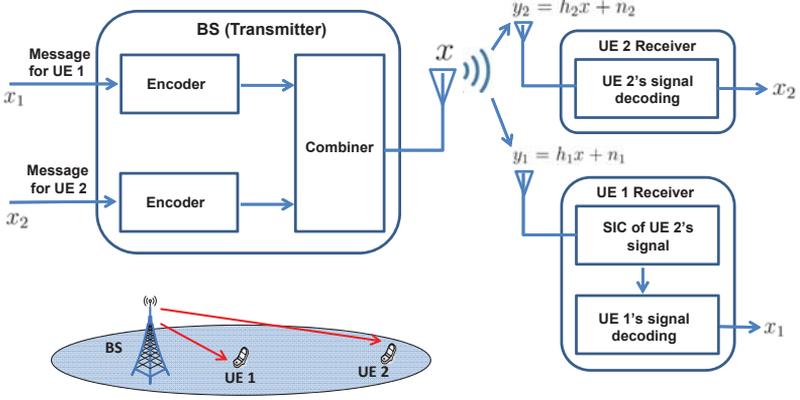
Figure 3: An illustration: superposition coding and SIC receiver.

h_k ($k = 1, 2$) is the complex channel gain for UE k and n_k is additive white Gaussian noise for UE k. Assuming that $|h_1|^2/n_1 > |h_2|^2/n_2$, then the signals which can be decoded at UE 2 can most likely be decoded by UE 1 as well [29]. At the receiver side, UE 1's receiver first decodes the interfering signal x_2 from y_1. After subtracting x_2, the signal x_1 intended for UE 1 is decoded from $h_1 x_1 + n_1$. At UE 2's receiver, no interference cancellation takes place, and the signal of interest, x_2, is directly decoded from y_2 by treating x_1 as noise. Apart from the power-domain NOMA, there are some other candidate non-orthogonal MA schemes are under investigation for 5G, e.g., sparse code multiple access (SCMA), multi-user shared access (MUSA), and pattern division multiple access (PDMA) [30, 31]. Theses non-orthogonal MA schemes share the same idea. That is multiple UEs can simultaneously use the same subchannels [31]. Before the above non-orthogonal MA schemes are incorporated into 5G standards, several key issues must be addressed, e.g., advanced low-complexity receiver for SIC.

11

3 Radio Resource Optimization in Cellular Networks

Radio resource allocation or scheduling is important for performance improvement in cellular networks. The goal of *resource allocation* is to optimize the assignment of the limited frequency/power/time resource to achieve the best performance by taking realistic constraints into account. The benefits of optimization in cellular networks include boosting network performance, satisfying diverse QoS requirements, saving energy, as well as reducing the capital expenditure and operation expenditure. A radio resource optimization (RRO) problem typically consists of a utility function as the objective, a set of constraints, and variables to be optimized. The utility can be chosen from a range of performance metrics. The constraints are usually according to some physical limitations in cellular networks or QoS requirements in practice. All constraints combining with the optimization variables define a feasible solution region for the optimization problem.

In general, we aim to find optimal solutions from the feasible region, or develop near-optimal solutions. For the addressed RRO problems in the dissertation, it is a difficult and challenging task to solve the problems, especially for the large-scale instances. We focus on three classes of optimization problems, i.e., utility optimization in SC-FDMA, energy-efficient scheduling in OFDMA, and joint power and channel allocation in NOMA. In this section, we first present some widely used utility metrics. Then, we give a brief introduction of each optimization problem.

3.1 Performance Metrics

Utility, in general, can be an abstract concept, e.g., fairness and satisfaction, or a real performance measure, e.g., consumed power in Watt. A utility function is used to quantify and provide a tangible performance metric in the objective of an optimization problem. In the following, we summarize some classic utility functions, including those used in the dissertation.

- Throughput maximization

  Throughput, also referred to as sum-rate utility, represents the aggregate data rate of the UEs in a cellular network. The utility function can be expressed by $\sum_{k=1}^{K} R_k$ if we consider K UEs in the system, where R_k is the instantaneous data rate of UE k, usually computed by Shannon's channel capacity equation in bits per second. Besides, spectrum efficiency, in bits per second per Hz, can also be used to quantify throughput in unit bandwidth.

- Weighted sum-rate and fairness

  In some application scenarios, instead of merely considering maximum throughput, UEs' priority and fairness in resource allocation need to be taken into account. The corresponding utility function is expressed as $\sum_{k=1}^{K} W_k R_k$ by introducing a weight factor W_k for each UE k to maintain fairness among UEs. For example, the UE with poor channel condition could be allocated with higher weight than the UE with good channel condition, to avoid excessive imbalance in resource allocation among UEs.

  In resource scheduling over a time duration, weights can be used to obtain fairness. For example, one can update $W_k = 1/\bar{R}_k$ for each k in each TTI, where $\bar{R}_k$ is the average rate of UE k. There are several measurements to quantify fairness [32]. Jain's fairness index is one of the most common measures to represent fairness, defined as $v = \frac{(\sum_{k=1}^{K} \bar{R}_k)^2}{K \sum_{k=1}^{K} \bar{R}_k^2}$ [33]. A larger value of v, $0 \leq v \leq 1$, represents fairer rate distribution among UEs from a system perspective.

- Power/energy minimization

  Power/energy minimization is another category of objectives. It is an important performance metric especially for battery-limited transmitters in uplink transmission. A typical utility function is $\sum_{k=1}^{K} P_k$, where P_k is the transmit power for UE k. In addition, another performance metric is energy efficiency which is defined as transmitted bits per unit energy consumption [34]. The function is expressed by $\frac{\sum_{k=1}^{K} R_k}{\sum_{k=1}^{K} P_k}$ in bit per Joule.

- Minimum number of allocated channels

  Taking the scarcity of frequency resource into account, another type of utility metric amounts to minimizing the number of allocated subchannels required to meet each UE's data demand [35]. The benefits consist of two aspects. First, the inter-cell interference (ICI) is mitigated by minimizing the used channels in each cell. Second, it is relevant to make as much resource available as possible for elastic traffic, while guaranteeing the rates for real-time applications. The utility function is defined as $\sum_{k=1}^{K} |\mathcal{N}_k|$ for OFDM-based systems due to exclusive channel access, and $|\bigcup_{k=1}^{K} \mathcal{N}_k|$ for NOMA systems, where $\mathcal{N}_k$ is the set of allocated subchannels of UE k.

3.2   Utility Optimization in SC-FDMA Systems

As discussed in Section 2.1, both OFDMA and SC-FDMA can provide fine-granularity and flexibility in channel allocation. In SC-FDMA, there are two types of criteria in assigning subchannels to UEs, LFDMA and IFDMA [3]. For both assignments in uplink as well as OFDMA in LTE downlink, there is a common constraint of ensuring exclusivity in subchannel allocation. That is, one subchannel can be occupied by one UE at most. Moreover, the constraints of consecutive and interleaved allocation are respectively imposed by LFDMA and IFDMA. In terms of the power allocation, the following constraints are usually considered. First, each UE's transmit power and the power allocated on each subchannel should be less than some maximum power levels, since the transmitter in uplink is typically a battery-powered device. Next, power allocation over the allocated subchannels for a UE can be uniform or adaptive in power allocation.

Based on channel state information (CSI), the RRO in SC-FDMA amounts to determining the optimal subchannel and power allocation, such that the objective is maximized or minimized, and all the constraints are satisfied. For SC-FDMA (as well as OFDMA), resource allocation can be categorized into two groups. The first is to minimize utilities, e.g., total transmit power or number of used subchannels, with the constraints of satisfying QoS requirements, e.g., achieving mini-

mum data rate for each UE, see e.g., [36, 37, 38, 39]. The second aims to maximize utilities, e.g., system throughput, with the constraints of limited power as well as QoS requirements, see e.g., [40, 41].

In SC-FDMA, localized or interleaved allocation can lead to system performance improvement, but it also limits the freedom in channel allocation and thus rises differences over OFDMA. Finding global optimum for the consecutive channel allocation problems is generally hard [42, 43], whereas interleaved channel allocation problems are tractable [44]. The channel allocation problems in IFDMA, e.g., throughput maximization [44] or power minimization [45], can be mapped to the classical matching problems which are polynomial-time solvable. The consecutive-channel allocation problems in LFDMA are more challenging than IFDMA. Some works have been devoted for addressing the problems [46, 47, 48]. Global optimal solutions for these problems in LFDMA may not be achievable in practice due to the heavy computational overhead. Many research works have been focused on developing near-optimal solutions with low-complexity [49, 50, 51, 52].

Surveys of SC-FDMA and OFDMA resource allocation can be found in [53, 54, 55]. Overviews of the optimization approaches for dynamic resource allocation are provided in [56, 57]. Lots of research papers have investigated resource allocation approaches for SC-FDMA and OFDMA, see e.g., [58, 59, 60, 61, 62].

3.3   Energy-efficient Scheduling in OFDMA Systems

Energy consumers in a cellular network are typically BSs and UEs. It becomes crucial to address their energy consumption when more and more BSs and UEs are in the network, and high data traffic is demanded [63]. Energy is the product of power and time. The energy-saving issues can be addressed from these two aspects.

From the power domain, radio components (RCs), e.g., power amplifier and radio frequency components, in BSs and UEs dominate the power consumption [64]. To save power, one way is to improve the power efficiency of RCs in hardware design. Another efficient and intuitive way is to deactivate transceivers' RCs whenever there is no data to receive or transmit. This concept is supported and implemented by discontinuous reception (DRX) and discontinuous transmission (DTX)

modes in LTE standards [7]. DRX and DTX enable devices to work at sleep/active states by deactivating/activating the transceivers' RCs momentarily (milliseconds) [65]. When a transceiver's RC is deactivated, it will not emit or receive any data traffic [65]. It has been shown in [5, 34] that the power consumption on circuit of a BS in the sleep mode is much lower than the power consumption in the active mode. This provides significant gains in power saving and reduces ICI.

Combining with the concept of DTX/DRX, energy consumption can be reduced by intelligent scheduling approaches in the time domain. In order to minimize energy consumption while satisfying UEs' QoS requirements, energy-efficient (EE) scheduling and operation in this dissertation is investigated from two perspectives. One focuses on the time-frequency resource optimization in a single-cell OFDMA system, and the other focuses on optimizing BSs operation in a multi-cell OFDMA network.

- EE scheduling in single-cell scenarios

    The adoption of OFDMA gives flexibility in allocating subchannels and time slots to UEs. For an OFDMA single cell, time-frequency units are represented by a resource grid as shown in Figure 4. A small block is a time-frequency RU to be allocated to different UEs.
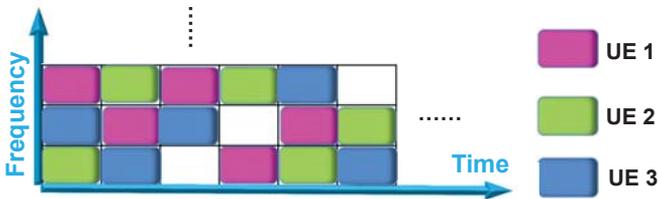

Figure 4: An illustration of the resource grid in OFDMA: 6 time slots and 3 subchannels allocated among 3 UEs.

    From Figure 4, each UE accesses subchannels over five time slots. The transmitter, i.e., BS, should work at the active mode over six time slots to deliver data for UEs, and at the receiver side, each UE must keep active for five slots to receive data [66]. The total

energy consumption of the cellular system has two parts: hardware circuit power consumption at both BS and UEs for being active, and a dynamic part which is transmit power depending on how many RUs are used for data transmission. To reduce the energy, each UE as well as BS can be jointly scheduled to use fewer time slots to receive or transmit data. In addition, given the fact that a subchannel with poor channel condition for a UE may be in a good state for other UEs, the overall energy performance can be improved if the RU-UE assignment is optimized. One example is shown in Figure 5. For delivering the same UE data demand as in Figure 4, the allocation of subchannels and UEs is optimized by EE scheduling at the RU level. The span of the active slots for BS and UEs are "squeezed". After optimization, BS can deliver each UE's data by using five time slots in total, and be in the sleep mode in the last time slot to reduce energy. Each UE only needs to be active over three time slots for receiving data, and be sleeping for the remaining slots to save energy. Moreover, fewer RUs are used to deliver UEs' demand thus enables less transmit power.
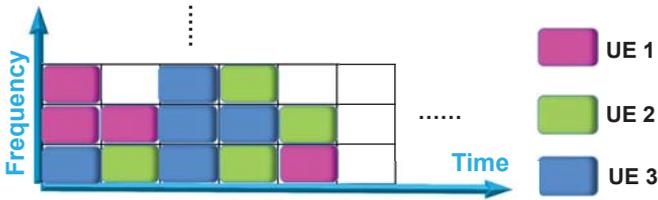
Figure 5: Optimized resource allocation: scheduling UEs to fewer time slots to reduce energy consumption.

Some algorithmic solutions are proposed to separately reduce the energy consumption at the transmitter [37] or at the receiver [67] by optimizing resource allocation. Also, many research works investigate resource allocation approaches for a single-cell OFDMA system to improve energy efficiency, e.g., see [39, 68].

• EE scheduling and operation in multi-cell scenarios

In a multi-cell network, due to the scarcity of the spectrum, co-channel frequency deployment is usually considered [69]. Multiple BSs use the same frequency band, and thus interfere with each other. The system performance is thus limited by ICI which may result in low system throughput and high energy consumption.

One way to reduce energy consumption as well as ICI is to optimize transmit power and resource usage [70, 71]. Based on the data from real measurements, transmit power is usually considered as a linear function of the utilization level of RUs in a cell. The work in [64] shows that linear models can provide reasonably good approximation for transmit power in BSs with respect to the resource usage. Observing this, for fixed transmit power per RU, the total power consumption increases when more RUs are used to transmit data in a cell. As a consequence, this cell will radiate more interference to other cells. If we look at the re-
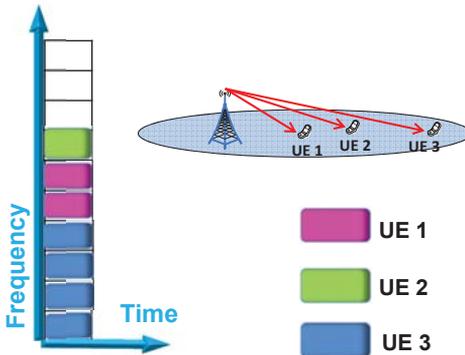
Figure 6: An illustration for cell's load: 10 available RUs and load = 0.7.

source usage in a particular time slot, a cell's *load* is defined as the fractional usage of RUs of the cell. For example in Figure 6, seven RUs out of ten RUs available in total are used to deliver UEs' data demand, then the load in this cell is 0.7. The sum of the power on these seven RUs is the overall transmit power in this cell. Given the transmit power per RU, higher load value means stronger interference to other cells. An analytical model, called

load-coupling model, widely used in literature [72, 73, 74], is developed to capture the characteristic between resource usage and interference. This model takes into account the load level to estimate ICI [75]. For example, the interference generated by a cell 1 to other cells depends on the load level and the transmit power per RU in cell 1, and the interference, in its turn, has impact on the load levels of other cells. For satisfying UEs' data demand by consuming minimum energy, the optimal load level of each cell and the transmit power per RU should be decided. This forms an optimization problem. Some relevant analytical results about energy savings, load balancing, and offloading can be found in [72, 73, 74, 75].

Another efficient solution to reduce energy consumption is to switch some BSs off or to the sleep mode with low power consumption. For the former, some works focus on switching on/off BSs based on the analysis of traffic profile, e.g., night and day. Some BSs can be switched off for a long duration when the traffic in these cells is low [5], and the availability of UEs' access is guaranteed in the service area. For the latter, some research efforts focus on activating/deactivating BSs' RC for short period to reduce energy consumption and ICI. In order to satisfy data demand for all UEs within a strict transmission time requirement, how to select the best combination of activating/deactivating BSs in the network, meanwhile consuming minimum energy are addressed in some works, see, e.g., [76]. Some optimization approaches and algorithms for EE scheduling are proposed, see, e.g., [77].

Besides, some other factors can affect energy consumption in multi-cell networks, e.g., load imbalance among cells [78]. The term load balancing describes any mechanism that transfers some traffic from the cells with heavy load to the neighbors with less load in order to balance the load in the entire network, while improving the network performance, e.g. energy consumption [79, 80]. Inappropriate BS-UE association can lead to large disparity in cells' load, that is, some cells, e.g., macro BSs, may be overloaded, whereas some neighboring low-power BSs, e.g., micro or pico BSs, are underloaded. This is undesirable, and

possibly results in high energy consumption, since the overloaded cells may need to greatly increase power to satisfy the large amount of data demand for some UEs. To deal with this problem, load balancing can make the use of the radio resource more efficient across the network. From the analysis in some research works [79], energy consumption can be reduced if the underloaded cells offload some traffic from neighboring overloaded cells, e.g., by applying LTE cell range expansion [80]. Several algorithmic solutions have been proposed to improve the performance of load balancing and energy savings [72, 73].

3.4   Power and Channel Allocation in NOMA

New MA technologies in 5G, e.g., NOMA, are expected to significantly improve system performance. Compared to OMA, new MA technologies introduce new challenges in RRO. In the following, we outline the representative optimization problem of NOMA.

Unlike OMA, multiple UEs in NOMA are allowed to simultaneously use the same subchannel, as shown in Figure 2. Then intra-cell interference in NOMA is non-negligible, since even after SIC processing, some co-channel interference exists among the UEs. The system performance and RRO are influenced by this. The benefits of using NOMA and the impact in RRO can be illustrated by the following example. For instance, a cell-center UE 1 and a cell-edge UE 2 are deployed in NOMA downlink. UE 1 has much better channel condition than UE 2 mainly due to the geographical distances from the BS. In OMA, UE 2 is allocated with a subchannel with poor channel condition, then UE 1 cannot access this channel resource, whereas in NOMA, UE 1 can reuse the same subchannel with good channel condition to improve the overall performance. According to the principle of SIC in NOMA, UE 1 can remove the co-channel interference from UE 2, but the interference to UE 2 due to co-channel multiplexed UE 1 remains. Observing this, the power allocation to UE 1 not only affects UE 1's performance, e.g., throughput, but also affects the performance of UE 2.

The overall performance in NOMA is very dependent on which UEs are grouped together and allocated to which subchannel, as well as how much power should be allocated to the UEs. In addition, due to practical limitations, it may not be realistic to have a large number of UEs

to be allocated on each subchannel [26]. The number of multiplexed UEs on each subchannel is typically bounded by a number. All these are considered in RRO problems in NOMA. Solving the problems is challenging in general. Some algorithms and schemes are proposed to optimize the channel and power allocation for NOMA dowlink and uplink [26, 27, 81].

In summary, most of the proposed approaches address the RRO problems in SC-FDMA, OFDMA and NOMA by applying the following methods:

- Simplifying and making assumptions to reduce the complexity of the optimization process, e.g., assuming uniform power allocation, predefining fixed groups of BS, UE, or channel, before the optimization process;

- Splitting the difficult optimization procedure into several (preferably independent) easier-to-solve problems to make the overall problem tractable but it may sacrifice optimality, e.g., separating the joint channel and power allocation into two separate steps: channel allocation and power allocation;

- Relaxing some "complicating" constraints or variables, e.g., linear relaxation for integer optimization problems;

- Developing exact, approximation, or heuristic algorithms, as well as providing upper and lower bounds for global optimum;

4   Mathematical Optimization

Mathematical optimization is the main approach for addressing the RRO problems in this dissertation. In this section, we provide an introduction to the optimization methods from three aspects: mathematical modeling, problem complexity, and algorithms. For details of mathematical optimization, the reader is referred to [82, 83].

4.1   Mathematical Modeling

Mathematical modeling amounts to constructing a mathematical formulation to represent the considered problem. A general optimization problem can be expressed as:

$$\min \ f_0(\boldsymbol{x}) \tag{1a}$$

$$\text{subject to} \ \ f_i(\boldsymbol{x}) \leq 0, \ i = 1, \dots, p \tag{1b}$$

$$h_j(\boldsymbol{x}) = 0, \ j = 1, \dots, q \tag{1c}$$

where the n-dimension vector $\boldsymbol{x}$ is the set of optimization variables of the problem, the function f_0 is the objective function, $f_i(\boldsymbol{x}) \leq 0$ and $h_j(\boldsymbol{x}) = 0$ define p inequality and q equality constraints, respectively. The model describes the problem of finding an optimal solution $\boldsymbol{x}^*$ that minimizes f_0 among all possible $\boldsymbol{x}$. Optimization problems can be classified according to the particular forms of the constraint and objective function (linear, nonlinear, convex), and variables (continuous, discrete).

- Linear programming

  The optimization problem in (1) is called a linear programming (LP) formulation if the objective and constraint functions are linear and all variables are continuous. The problem is nonlinear programming (NLP) if the objective or some constraint functions in (1) are nonlinear. An LP problem is usually formulated in the following standard form:

$$\min_{\boldsymbol{x}} \ \boldsymbol{c}^T \boldsymbol{x}, \ \ \text{subject to} \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0 \tag{2}$$

where c^T is a transposed n-dimension vector of coefficients, b is a column vector in m-dimension, and A is a matrix with m rows and n columns. The linear constraints and continuous variables define a feasible region of an LP problem as a polyhedron [84]. The problem is said *feasible* if there exists at least one point in this feasible region. The problem is *infeasible* if the feasible region is empty, and the problem is *unbounded* if the optimal objective value $c^T x^*$ is $-\infty$. LP problems can be efficiently solved by the simplex algorithm and interior-point algorithms in practice. More information of both algorithmic solutions, and more discussions of LP can be found in [82, 84].

- Integer programming

  If all the variables are restricted to be integral values, then the problem is referred to as integer programming (IP) problem, or combinatorial optimization problem. The standard form of an IP problem is shown below:

$$\min_{x} \ c^T x, \ \text{subject to } Ax = b, x \geq 0 \text{ and integer} \qquad (3)$$

  As a special case of IP, if all variables are constrained to be in $\{0, 1\}$, then it is a binary integer programming (BIP) problem. The problem (3) is called mixed integer programming (MIP) if only some variables, not all, have to be integral.

  Despite the resemblance of formulation (3) to LP formulation (2), solving MIP and IP problems is much harder than LP in general. Some algorithms, e.g., branch-and-bound and branch-and-cut algorithms, can guarantee to solve the general MIP or IP exactly to global optimum, but it is typically time-consuming in particular for the large-scale instances. The execution time is in general exponential with the number of integer variables [82]. One may need to develop sub-optimal but low-complexity algorithms, as well as optimality bounds. Details of algorithmic approaches for solving IP and MIP can be found in [82].

- Convex programming

23

The term convex programming is used to represent a class of the general optimization problem (1) in which the objective and inequality constraint functions, i.e., f_0 and $f_1, \ldots, f_p$, are convex, and the equality constraint functions $h_1, \ldots, h_q$ are affine [85]. By definition, solving the problem amounts to optimizing a convex objective over a convex set. An LP problem is also a convex problem. Any local optimum of a convex problem is a global optimum. For a convex problem, strong duality holds, i.e., zero duality gap between the original problem and the dual problem holds. It means an original convex problem can be optimally solved from the dual domain by constructing its Lagrange dual. In general, convex problems can be solved efficiently by some well-known algorithms. e.g., interior-point algorithm. More detailed information of convex optimization can be found in [85].

It is worth noting that having an LP, IP, MIP, or a convex formulation in hand is always advantageous than NLP and nonconvex formulations in general, since the former can be directly treated by standard solvers, e.g., CPLEX [86] and GUROBI [87] for solving LP, IP, and MIP problems, and CVX and YALMIP [88] for convex problems. Even though the state-of-art solvers may not be very efficient for solving some large instances, but at least for small instances, the global optimum can be expected. The obtained global optimum can be used for benchmarking and evaluating the developed heuristic solutions or bounding schemes. Convex problems are usually solved efficiently, but in practice recognizing a convex function is much difficult than identifying LP and IP problems. In addition, some original problems may not have a convex form due to inappropriate formulations, but for some of them it is possible to transform them to convex problems [85]. In general, recognizing a convex problem, or possibly transforming a noncovex problem to a convex optimization problem is challenging. Solving NLP and nonconvex problems can be difficult even for small instances. So far, there are no reliable and effective approaches for solving general NLP and nonconvex problems to global optimum.

4.2   Problem Complexity

The computational complexity theory helps algorithm developers to identify how difficult for solving a problem is. A problem is said to be *tractable* if it can be optimally solved by polynomial-time complexity algorithms, and *intractable* otherwise. By the theory, the class P (Polynomial) is defined as all decision problems which are tractable. A decision problem, also referred to as the decision or recognition version of an optimization problem, has only a yes-or-no solution. The class NP (non-deterministic polynomial) contains the problems that might have polynomial-time solutions [82]. The class of NP-complete consists of the most difficult problems in NP. More precisely, a decision problem is said to be NP-complete if it belongs to NP, and all other problems in NP can reduce to this problem polynomially [89]. The class of NP-hard includes the problems that are at least as difficult as NP-complete, not necessarily in NP. If the decision version of an optimization problem is NP-complete, then the optimization problem is NP-hard. This is because solving an optimization problem is no easier than solving its decision version, since the former requires to find the optimal values, whereas the latter only needs to provide a yes-or-no answer.

It is widely accepted that NP-complete problems are intractable. If a problem is NP-complete (or NP-hard), one cannot expect a polynomial-time algorithm with global optimality guarantee, unless NP = P. Thus, problem's tractability or intractability is of significance in developing algorithmic solutions. Once an optimization problem is proved to be NP-hard, it means the problem is intractable in general. Instead of obtaining global optimum, we may need to develop suboptimal algorithms with polynomial-time complexity. If we need to prove the NP-completeness for a new decision problem, say P_{new}, the following steps can be applied [90]:

1. Selecting a suitable and already known NP-complete problem P_{npc}

2. Constructing a special instance of P_{new}

3. Establishing a polynomial-time transformation from P_{npc} to the special instance of P_{new}

4. Proving that any instance in P_{npc} is yes if and only if the constructed instance in P_{new} is yes

If the above steps are successfully performed, the problem P_{npc} is said to be reducible to P_{new}, then the problem P_{new} is also NP-complete. Note that the theory of NP-completeness always focuses on the worst case. Solving a constructed instance in P_{new} is hard, then the worst case of P_{new} is therefore intractable. Details of the computational complexity theory and examples of hardness proofs can be found in [82, 89].

4.3 Algorithmic Solutions

Algorithmic approaches for solving optimization problems are roughly categorized into two broad types: exact and heuristic. An exact algorithm, e.g., branch-and-bound, guarantees to find the global optimal solution, but it may take exponential time. Heuristic algorithms are developed for solving hard problems, and used to find suboptimal solutions with polynomial-time complexity. The algorithmic approaches used in this dissertation for tackling RRO problems are presented below.

- Dynamic programming

  Dynamic programming (DP) is a stagewise and recursive method for solving discrete problems. The idea of DP is to divide the entire solution process into multiple stages, and systematically search all possibilities to guarantee optimality. The global optimum can be recursively obtained by solving subproblems at each stage. DP guarantees to find global optimum only if the problem has optimal substructure [83]. That is, at any stage, the partial (or local) optimal solution obtained so far can be reused without any change by the later stages for finding the global optimum.

  The main procedure of DP is to: 1) divide the solution process into N stages, 2) process from the first stage to N, one by one, 3) obtain and store the partial optimal solution at each stage, 4) move from one stage to the next stage by following the constructed recursion formula, 5) at stage N, the global optimum is the accumulation of the partial optimum obtained from the previous stages.

- Column generation

  Column generation (CG) is an efficient decomposition method for tackling large-scale problems, e.g., LP problems [91]. If we consider LP formulation (2) for example, a variable x_j, $j \in \{1, \ldots, n\}$, is associated with a column vector $\boldsymbol{a_j}$ in the $m \times n$ matrix $\boldsymbol{A}$, where $\boldsymbol{A} = [\boldsymbol{a_1}, \ldots, \boldsymbol{a_n}]$ and $\boldsymbol{a_j} = [a_{1j}, \ldots, a_{mj}]^T$. The term *column* can refer to a column vector in $\boldsymbol{A}$. A large-scale optimization problem contains a huge number of variables along with columns in matrix $\boldsymbol{A}$. Some algorithms, e.g., the simplex method, require to explicitly consider all variables and columns in the algorithm execution. This may result in a prohibitive amount of time to obtain optimal solution in practice. CG provides an efficient way to address this issue. The idea is that variables and columns are not enumerated explicitly, instead, they are generated only when needed.

  By applying CG, the original problem is decomposed into a master problem and a subproblem (or pricing problem). The algorithm starts from solving a small-scale master problem with only a few columns and variables in initial $\boldsymbol{A}_0$ and $\boldsymbol{x}_0$. When the master problem (an LP problem) is solved, we obtain the optimal dual value for each constraint. These dual values are incorporated to the objective of the subproblem. Then the subproblem is solved to determine a new variable x_{new} and a column $\boldsymbol{a}_{new}$ to add to $\boldsymbol{x}_0$ and $\boldsymbol{A}_0$, respectively, for the master problem. The above iterative process is repeated until no new column is able to be generated by the subproblem. In iterations, the solution quality of the master problem is successively improved by adding the new column and variable. One advantage of CG is that the size of initial columns is small, and they may retain small until the optimum is obtained [83, 91].

- Lagrangian relaxation

  Relaxation is one of the approaches to address the hard optimization problems. It is motivated by the fact that if we relax or remove some "complicating" constraints for a hard problem [83], the relaxed problem may become easy to solve. Moreover, the

27

relaxation for an original problem usually leads to a lower-bound or an upper-bound solution for a minimization or a maximization problem, respectively. If we just simply remove some constraints, the resulting bounds may be weak, i.e., far away from the primal optimum. To address this issue, Lagrangian relaxation transfers some constraints associated with Lagrange multipliers to the objective function, and finds the optimal multipliers which results in the best possible bound.

For example in formulation (1), if we relax the equality constraints $h_j(\boldsymbol{x}) = 0, \ j = 1, \ldots, q$, we refer to the following problem

$$\min \ f_0(\boldsymbol{x}) + \sum_{j=1}^{q} \lambda_j h_j(\boldsymbol{x}) \tag{4a}$$

$$\text{subject to} \ f_i(\boldsymbol{x}) \leq 0, \ i = 1, \ldots, p \tag{4b}$$

as the Lagrangian relaxation or Lagrangian subproblem of the original problem (1), where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)$ is the vector of multipliers (also called dual variables). We refer to function $L(\boldsymbol{\lambda}) = \min\{f_0(\boldsymbol{x}) + \sum_{j=1}^{q} \lambda_j h_j(\boldsymbol{x}) \mid f_i(\boldsymbol{x}) \leq 0, \ i = 1, \ldots, p\}$ as the Lagrangian function. In order to obtain the best lower bound, we need to solve the following Lagrangian dual problem.

$$L^* = \max_{\boldsymbol{\lambda}} L(\boldsymbol{\lambda}) \tag{5}$$

For each $\boldsymbol{\lambda}$, we can solve $L(\boldsymbol{\lambda})$ and update $\boldsymbol{\lambda}$. This procedure is repeated until the stop criteria is satisfied. L^* is obtained as the best possible lower bound of the optimal value Z^* of the original problem (1), i.e., $L(\boldsymbol{\lambda}) \leq L^* \leq Z^*$.

- Heuristics

  Heuristic algorithms aim at finding suboptimal solutions for difficult optimization problems, e.g., NP-hard problems, with a reasonable running time. Unlike exact algorithms, heuristic algorithms have no global optimality guarantee in general. Greedy algorithm is one of the intuitive and commonly used heuristics. At each step or iteration, the algorithm only makes the choice

that seems best at the moment, i.e., locally optimal choice. It is typically a one-pass algorithm. This means the algorithm stops once a suboptimal (and feasible) solution is found. For some easy problems, greedy algorithms can guarantee to find global optimum, e.g., the minimum spanning tree problem [83]. However, for solving some hard problems, the solution quality of greedy algorithms may not be very satisfactory, since the problems may not have optimal substructure, and a local optimum may not necessarily be the global optimum. To overcome the disadvantage, many other heuristic algorithms such as simulated annealing, tabu search, genetic algorithms, are proposed to trade off computational efficiency and solution quality.

5   Contributions

This dissertation aims at investigating spectrum- and energy-efficient resource allocation to optimize system performance in 4G and 5G communication systems. The research topics cover resource allocation and system performance optimization in SC-FDMA, energy-efficient scheduling in OFDMA, energy minimization in load-coupled OFDMA networks, and throughput and fairness optimization in NOMA resource allocation. The scope of the dissertation is formed by mathematical modeling for the RRO problems, analysis of problems' complexity, algorithm development, as well as theoretical and numerical results analysis.

The dissertation consists of five research papers. In these papers, the main ideas, the core concept of the proposed algorithms, and the major theoretical results are generated from the discussions among all the authors. The author of this dissertation has contributed to Paper I, II, IV, and V as the first author, mainly taking the works of the development of optimization models, algorithm design and implementation, theoretical analysis and part of the theorem and lemma proofs, all the simulation works and numerical results analysis, as well as writing. The author has contributed to Paper III as a co-author, focused on algorithm design, development and implementation, performance evaluation to verify the theoretical findings, numerical results analysis, as well as the writing of these parts. The papers and the main scientific contributions are summarized as follows:

- **Paper I: A Unified Graph Labeling Algorithm for Consecutive-Block Channel Allocation in SC-FDMA**

  Paper I deals with three localized SC-FDMA resource allocation problems, utility maximization, power minimization, and channel minimization. For solving these optimization problems, we provide the structural insight that allocating consecutive channels optimally can be mapped to finding an optimal path in an acyclic graph.

  First, the complexity of the three problems has been analyzed. We

30

prove their NP-hardness. Next, a unified algorithmic framework is proposed for solving the problems by applying and developing the concept of graph labeling. The advantage of the proposed algorithm is that the solution procedures for tackling three resource allocation problems are unified under a common algorithmic framework. Also, the algorithm allows a trade-off between computational efforts and optimality by adjusting a algorithmic parameter. The proposed algorithm guarantees global optimality for some special classes of the three problems. Numerical results show that the proposed algorithmic framework is competitive in attaining near-optimal solutions.

The paper has been published in *IEEE Transactions on Wireless Communications*. Parts of the results have been published in the following conference:

L. Lei, S. Fowler, and D. Yuan, "Improved Resource Allocation Algorithm Based on Partial Solution Estimation for SC-FDMA Systems," *Proceedings of IEEE Vehicular Technology Conference (VTC Fall)*, 2013.

- **Paper II: Resource Scheduling to Jointly Minimize Receiving and Transmitting Energy in OFDMA Systems**

This paper addresses an energy-efficient scheduling problem for OFDMA downlink. We jointly minimize receiving and transmitting energy instead of considering energy reduction only at the transmitter or the receiver. The energy-saving gains are from two sides. At the transmitter side, we minimize the transmit energy as well as the circuit energy consumption. For the receiver side, we minimize the number of time slots for receiving data to reduce the receivers' energy consumption.

We formulate the optimization problem by means of integer programming. To alleviate the high computational complexity for obtaining the global optimal solution, an energy-efficient scheduling algorithm based on column generation is developed to provide a tight lower bound and a feasible near-optimal solution. Performance evaluation shows that the proposed algorithm is promising

in solving the problem efficiently. Also, the algorithm is capable of providing close-to-optimum bounds for the global optimum.

The paper has been published in *Proceedings of IEEE International Symposium on Wireless Communication Systems (ISWCS)*.

- **Paper III: Power and Load Coupling in Cellular Networks for Energy Optimization**

Paper III focuses on investigating fundamental properties and algorithmic solutions to optimally minimize energy in load-coupled OFDMA networks. The coupling relation of mutual interference among multiple cells is characterized by a non-linear load-coupling model. Both cell load and transmit power interact via the load-coupling model. Our theoretical and algorithmic investigations provide the answers for the following fundamental questions: what the optimal operating load of each cell should be, and how to compute the transmit power for achieving the optimal load.

We formulate the energy minimization problem for load-coupled networks. Due to the inherently non-linearity, solving the problem is challenging. First, we prove that operating at full load is optimal in minimizing sum energy. Second, to achieve the target operating load for each cell, we propose an iterative power adjustment algorithm to obtain the corresponding optimal transmit power with guaranteed convergence. We present numerical results to corroborate the theoretical findings in a large-scale cellular network, showing the advantage of our solution compared to the conventional solution.

The paper has been published in *IEEE Transactions on Wireless Communications*. Parts of the paper have been published in the following conference:

C. K. Ho, D. Yuan, L. Lei, and S. Sun, "Optimal Energy Minimization in Load-Coupled Wireless Networks: Computation and Properties," *Proceedings of IEEE International Conference on Communications (ICC)*, 2014.

- **Paper IV: Optimal Cell Clustering and Activation for Energy Saving in Load-Coupled Wireless Networks**

We study a problem of energy-efficient scheduling by performing cell activating/deactivating for a load-coupled OFDMA network. The cells are required to serve a target amount of data for the UEs within a time limit to maintain an appropriate level of QoS, while considering the coupling relation among cells due to mutual interference.

First, we provide mathematical formulations, and prove the problem's NP hardness. Second, for solving the problem, we propose a column generation based approach, with the capability of approaching global optimum. Then, for dealing with large-scale networks, we derive a bounding scheme to trade off optimality and the computational complexity. We provide numerical results to demonstrate that our solutions achieve significant energy saving over existing schemes.

The paper has been published in *IEEE Transactions on Wireless Communications*.

- **Paper V: Power and Channel Allocation for Non-orthogonal Multiple Access in 5G Systems: Tractability and Computation**

We investigate optimization problems of jointly optimizing power and channel allocation in downlink NOMA, for maximizing utilities and improving fairness. We provide theoretical insights on complexity and optimality. We investigate how different utility functions and power constraints influence the tractability and optimal strategies in resource allocation.

We formulate the optimization problems for NOMA, taking into account practical considerations of fairness and SIC. To solve the considered problems, we propose an algorithm framework based on Lagrangian dual optimization and dynamic programming. The proposed algorithm is capable of providing near-optimal solutions as well as bounding the global optimum tightly. We use nu-

33

merical results to illustrate the significant performance improvement of the proposed algorithm over existing NOMA and OFDMA schemes.

The paper has been submitted to *IEEE Transactions on Wireless Communications*. Parts of the paper have been published in the following conference:

L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Joint Optimization of Power and Channel Allocation with Non-orthogonal Multiple Access for 5G Cellular Systems," *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, 2015.

Bibliography

[1] "Cisco visual networking index," Feb. 2014 White Paper. [Online]. Available: http://tinyurl.com/b9berc

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

[3] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. Academic Press, 2011.

[4] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, and F. Wiedmann, "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 97–105, Feb. 2014.

[5] H. Holtkamp, G. Auer, S. Bazzi, and H. Haas, "Minimizing base station power consumption," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 2, pp. 297–306, Feb. 2014.

[6] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 10–22, June 2010.

[7] ETSI, "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; stage 2 (3GPP TS 36.300 version 10.5.0 Release 10)," ETSI TS 136 300 V10.5.0, Nov. 2011.

[8] H. Holma and A. Toskala, *LTE for UMTS - OFDMA and SC-FDMA Based Radio Access*.   Wiley, 2009.

[9] S. Sesia, I. Toufik, and M. Baker, *LTE – The UMTS Long Term Evolution: From Theory to Practice*.   Wiley, 2009.

[10] Q. Li, G. Li, W. Lee, M. il Lee, D. Mazzarese, B. Clerckx, and Z. Li, "MIMO techniques in WiMAX and LTE: a feature overview," *IEEE Communications Magazine*, vol. 48, no. 5, pp. 86–92, May 2010.

[11] D. Astely, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, "LTE: the evolution of mobile broadband," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 44–51, Apr. 2009.

[12] R. V. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*.   Artech House, 2000.

[13] T. Jiang and Y. Wu, "An overview: Peak-to-average power ratio reduction techniques for OFDM signals," *IEEE Transactions on Broadcasting*, vol. 54, no. 2, pp. 257–268, June 2008.

[14] M. Rana, M. Islam, and A. Kouzani, "Peak to average power ratio analysis for LTE systems," in *IEEE International Conference on Communication Software and Networks (ICCSN)*, Feb. 2010, pp. 516–520.

[15] G. Berardinelli, L. Ruiz de Temino, S. Frattasi, M. Rahman, and P. Mogensen, "OFDMA vs. SC-FDMA: performance comparison in local area IMT-A scenarios," *IEEE Wireless Communications*, vol. 15, no. 5, pp. 64–72, Oct. 2008.

[16] E. Yaacoub and Z. Dawy, "A comparison of uplink scheduling in OFDMA and SCFDMA," in *IEEE International Conference on Telecommunications (ICT)*, Apr. 2010, pp. 466–470.

[17] H. Myung, J. Lim, and D. Goodman, "Single carrier FDMA for uplink wireless transmission," *IEEE Vehicular Technology Magazine*, vol. 1, no. 3, pp. 30–38, Sept. 2006.

[18] C. Ciochina and H. Sari, "A review of OFDMA and single-carrier FDMA," in *IEEE European Wireless Conference (EW)*, Apr. 2010, pp. 706–710.

[19] S. Song, G. Chen, and K. Letaief, "Localized or interleaved? a tradeoff between diversity and CFO interference in multipath channels," *IEEE Transactions on Wireless Communications*, vol. 10, no. 9, pp. 2829–2834, Sept. 2011.

[20] W. Nam, D. Bai, J. Lee, and I. Kang, "Advanced interference management for 5G cellular networks," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 52–60, May 2014.

[21] METIS, "Requirements and general design principles for new air interface," Deliverable D2.1, Aug. 2013. [Online]. Available: https://www.metis2020.com/documents/deliverables

[22] A. Benjebbour, A. Li, K. Saito, Y. Saito, Y. Kishiyama, and T. Nakamura, "NOMA: From concept to standardization," in *IEEE Conference on Standards for Communications and Networking (CSCN)*, Oct. 2015, pp. 18–23.

[23] 3GPP, "Study on downlink multiuser superposition transmission (MUST) for LTE," ETSI 3GPP TR 36.859, 2015.

[24] G. Wunder, M. Kasparick, S. ten Brink, F. Schaich, T. Wild, Y. Chen, I. Gaspar, N. Michailow, G. Fettweis, D. Ktenas, N. Cassiau, M. Dryjanski, K. Sorokosz, S. Pietrzyk, and B. Eged, "System-level interfaces and performance evaluation methodology for 5G physical layer based on non-orthogonal waveforms," in *Asilomar Conference on Signals, Systems and Computers*, Nov. 2013, pp. 1659–163.

[25] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Nov. 2013, pp. 770–774.

[26] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *IEEE Vehicular Technology Conference (VTC Spring)*, June 2013, pp. 1–5.

[27] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept. 2013, pp. 611–615.

[28] Q. Li, H. Niu, A. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 71–78, Mar. 2014.

[29] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[30] K. Au, L. Zhang, H. Nikopour, E. Yi, A. Bayesteh, U. Vilaipornsawai, J. Ma, and P. Zhu, "Uplink contention based SCMA for 5G radio access," in *IEEE Global Communications Conference Workshops (GLOBECOM), 2014*, Dec. 2014, pp. 900–905.

[31] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, Sept. 2015.

[32] T. Girici, C. Zhu, J. R. Agre, and A. Ephremides, "Proportional fair scheduling algorithm in OFDMA-based wireless systems with QoS constraints," *Journal of Communications and Networks*, vol. 12, no. 1, pp. 30–42, Feb. 2010.

[33] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," DEC Technical Report 301, Sept. 1984.

[34] G. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: tutorial, survey, and

open issues," *IEEE Wireless Communications*, vol. 18, no. 6, pp. 28–35, Dec. 2011.

[35] F. Sokmen and T. Girici, "Uplink resource allocation algorithms for single-carrier FDMA systems," in *IEEE European Wireless Conference (EW)*, Apr. 2010, pp. 339–345.

[36] G. Miao, N. Himayat, Y. Li, and D. Bormann, "Energy efficient design in wireless OFDMA," in *IEEE International Conference on Communications (ICC)*, May 2008, pp. 3307–3312.

[37] ——, "Energy efficient design in wireless OFDMA," in *IEEE International Conference on Communications (ICC)*, May 2008, pp. 3307–3312.

[38] Z. Zheng, Q. Ji, L. Dan, and S. Li, "Energy-efficient power allocation for uplink SC-FDMA," in *IEEE International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, Sept. 2012, pp. 1–4.

[39] C. Xiong, G. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in OFDMA networks," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2011, pp. 1–5.

[40] K. Seong, M. Mohseni, and J. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *IEEE International Symposium on Information Theory (ISIT)*, July 2006, pp. 1394–1398.

[41] L. Lei, S. Fowler, and D. Yuan, "Improved resource allocation algorithm based on partial solution estimation for SC-FDMA systems," in *IEEE Vehicular Technology Conference (VTC Fall)*, Sept. 2013, pp. 1–5.

[42] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, "Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink," in *IEEE Conference on Computer Communications (INFOCOM)*, Apr. 2009, pp. 2611–2615.

[43] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "A unified graph labeling algorithm for consecutive-block channel allocation in SC-FDMA,"

IEEE Transactions on Wireless Communications, vol. 12, no. 11, pp. 5767–5779, Nov. 2013.

[44] T. Mert, O. Kaya, and H. Cirpan, "Jointly optimal chunk and power allocation in uplink SC-FDMA," in *IEEE International Conference on Communications (ICC)*, June 2013, pp. 3393–3397.

[45] L. You, L. Lei, and D. Yuan, "A performance study of energy minimization for interleaved and localized FDMA," in *IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Dec. 2014, pp. 16–20.

[46] I. Wong, O. Oteri, and W. Mccoy, "Optimal resource allocation in uplink SC-FDMA systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2161–2165, May 2009.

[47] H. Safa and K. Tohme, "LTE uplink scheduling algorithms: Performance and challenges," in *International Conference on Telecommunications (ICT)*, Apr. 2012, pp. 1–6.

[48] A. Ahmad and M. Assaad, "Polynomial-complexity optimal resource allocation framework for uplink SC-FDMA systems," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2011, pp. 1–5.

[49] D. Kim, J. Kim, H. Kim, K. Kim, and Y. Han, "An efficient scheduler for uplink single carrier FDMA system," in *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept. 2010, pp. 1348–1353.

[50] C. Y. Ng and C.-W. Sung, "Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems," *IEEE Transactions on Wireless Communications,*, vol. 7, no. 5, pp. 1667–1675, May 2008.

[51] W. Xu, K. Niu, J. Lin, and Z. He, "Resource allocation in multicast OFDM systems: Lower/upper bounds and suboptimal algorithm,"

IEEE Communications Letters, vol. 15, no. 7, pp. 722–724, July 2011.

[52] Y.-B. Lin, W.-H. Wu, and Y. Su, "Optimal and suboptimal power allocations for MIMO based multi-hop OFDM systems," in *IEEE Vehicular Technology Conference (VTC Spring)*, May 2012, pp. 1–5.

[53] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems," *IEEE Communications Surveys Tutorials*, vol. 11, no. 3, pp. 92–106, 2009.

[54] E. Yaacoub and Z. Dawy, "A survey on uplink resource allocation in OFDMA wireless networks," *IEEE Communications Surveys Tutorials*, vol. 14, no. 2, pp. 322–337, 2012.

[55] J. Zhang, L.-L. Yang, L. Hanzo, and H. Gharavi, "Advances in cooperative single-carrier FDMA communications: Beyond LTE-advanced," *IEEE Communications Surveys Tutorials*, vol. 17, no. 2, pp. 730–756, 2015.

[56] X. Wang and G. Giannakis, "Resource allocation for wireless multiuser OFDM networks," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4359–4372, July 2011.

[57] E. Yaacoub and Z. Dawy, "A survey on uplink resource allocation in OFDMA wireless networks," *IEEE Communications Surveys Tutorials*, vol. 14, no. 2, pp. 322–337, 2012.

[58] L. Lei, V. Angelakis, and D. Yuan, "Performance analysis of chunk-based resource allocation in wireless OFDMA systems," in *IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Sept. 2012, pp. 90–94.

[59] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - part I: chunk allocation," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2734–2744, Sept. 2009.

[60] ——, "Chunk-based resource allocation in OFDMA systems part II: Joint chunk, power and bit allocation," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 499–509, Feb. 2012.

[61] N. Gao and X. Wang, "Optimal subcarrier-chunk scheduling for wireless OFDMA systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 7, pp. 2116–2123, July 2011.

[62] H. Zhao, L. Lei, D. Yuan, T. Larsson, and E. Ronnberg, "Power efficient uplink scheduling in SC-FDMA: Bounding global optimality by column generation," in *IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Sept. 2013., pp. 119–123.

[63] C.-L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: a 5G perspective," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 66–73, February 2014.

[64] C. Peng, S.-B. Lee, S. Lu, H. Luo, and H. Li, "Traffic-driven power saving in operational 3G cellular networks," in *ACM International Conference on Mobile Computing and Networking (MobiCom)*, Sept. 2011, pp. 121–132.

[65] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Godor, "Reducing energy consumption in LTE with cell DTX," in *IEEE Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–5.

[66] S. Luo, R. Zhang, and T. J. Lim, "Joint transmitter and receiver energy minimization in multiuser ofdm systems," *IEEE Transactions on Communications,*, vol. 62, no. 10, pp. 3504–3516, Oct. 2014.

[67] F.-S. Chu, K.-C. Chen, and G. Fettweis, "Green resource allocation to minimize receiving energy in OFDMA cellular systems," *IEEE Communications Letters*, vol. 16, no. 3, pp. 372–374, Mar. 2012.

[68] C. Xiong, G. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE*

Transactions on Wireless Communications, vol. 10, no. 11, pp. 3874–3886, Nov. 2011.

[69] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. Thomas, J. Andrews, P. Xia, H. Jo, H. Dhillon, and T. Novlan, "Heterogeneous cellular networks: From theory to practice," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 54–64, June 2012.

[70] M. Chiang, P. Hande, T. Lan, and C. W. Tan, "Power control in wireless cellular networks," *Foundations and Trends in Networking*, vol. 2, no. 4, pp. 381–533, 2008.

[71] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, Sept. 1995.

[72] C. K. Ho, D. Yuan, and S. Sun, "Data offloading in load coupled networks: Solution characterization and convexity analysis," in *IEEE International Conference on Communications (ICC)*, June 2013, pp. 1161–1165.

[73] ——, "Data offloading in load coupled networks: A utility maximization framework," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 1921–1931, Apr. 2014.

[74] I. Siomina, A. Furuskar, and G. Fodor, "A mathematical framework for statistical QoS and capacity studies in OFDM networks," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sept. 2009, pp. 2772–2776.

[75] I. Siomina and D. Yuan, "Analysis of cell load coupling for LTE network planning and optimization," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2287–2297, June 2012.

[76] K. Adachi, J. Joung, S. Sun, and P. H. Tan, "Adaptive coordinated napping (CoNap) for energy saving in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5656–5667, Nov. 2013.

[77] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Optimal cell clustering and activation for energy saving in load-coupled wireless networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6150–6163, Nov. 2015.

[78] W. Song, W. Zhuang, and Y. Cheng, "Load balancing for cellular/WLAN integrated networks," *IEEE Network*, vol. 21, no. 1, pp. 27–33, Jan. 2007.

[79] J. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in hetnets: old myths and open problems," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 18–25, Apr. 2014.

[80] H. Wang, L. Ding, P. Wu, Z. Pan, N. Liu, and X. You, "Dynamic load balancing and throughput optimization in 3GPP LTE networks," in *ACM International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2010, pp. 939–943.

[81] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Joint optimization of power and channel allocation with non-orthogonal multiple access for 5G cellular systems," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2015.

[82] M. Pióro and D. Medhi, *Routing, Flow, and Capacity Design in Communication and Computer Networks*. Morgan Kaufmann Publishers, 2004.

[83] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, 1993.

[84] K. Murty, *Linear programming*. Wiley, 1983.

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

[86] IBM, "ILOG CPLEX Optimizer," http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/, 2012.

[87] Gurobi, "Gurobi optimizer reference manual," 2015. [Online]. Available: http://www.gurobi.com

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

[89] M. R. Gary and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*.   W. H. Freeman, 1979.

[90] Y.-F. Liu and Y.-H. Dai, "On the complexity of joint subcarrier and power allocation for multi-user OFDMA systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 583–596, Feb. 2014.

[91] M. Lübecke and J. Desrosiers, "Selected topics in column generation," *Operations Research*, vol. 53, pp. 1007–1023, 2004.

Part II

Included Papers

Included Papers

The articles associated with this thesis have been removed for copyright reasons. For more details about these see:
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-126937