

QoS-aware Adaptive Resource Management in OFDMA Networks

by
Aini Li

A thesis submitted to the University of London for the
degree of
Doctor of Philosophy

School of Electronic Engineering and Computer Science
Queen Mary University of London
United Kingdom

November 2016

Abstract

One important feature of the future communication network is that users in the network are required to experience a guaranteed high quality of service (QoS) due to the popularity of multimedia applications. This thesis studies QoS-aware radio resource management schemes in different OFDMA network scenarios.

Motivated by the fact that in current 4G networks, the QoS provisioning is severely constrained by the availability of radio resources, especially the scarce spectrum as well as the unbalanced traffic distribution from cell to cell, a joint antenna and subcarrier management scheme is proposed to maximise user satisfaction with load balancing. Antenna pattern update mechanism is further investigated with moving users.

Combining network densification with cloud computing technologies, cloud radio access network (C-RAN) has been proposed as the emerging 5G network architecture consisting of baseband unit (BBU) pool, remote radio heads (RRHs) and fronthaul links. With cloud based information sharing through the BBU pool, a joint resource block and power allocation scheme is proposed to maximise the number of satisfied users whose required QoS is achieved. In this scenario, users are served by high power nodes only. With spatial reuse of system bandwidth by network densification, users' QoS provisioning can be ensured but it introduces energy and operating efficiency issue. Therefore two network energy optimisation schemes with QoS guarantee are further studied for C-RANs: an energy-effective network deployment scheme is designed for C-RAN based small cells; a joint RRH selection and user association scheme is investigated in heterogeneous C-RAN.

Thorough theoretical analysis is conducted in the development of all proposed algorithms, and the effectiveness of all proposed algorithms is validated via comprehensive simulations.

Acknowledgments

I would like to express my sincere gratitude to both of my supervisors, Dr. Yan Sun and Prof. Laurie Cuthbert. My first two years' PhD study was supervised by Prof. Laurie Cuthbert. His wide knowledge, dedicated supervision and generous encouragement have been of great value for me. After his retirement, I was transferred to Dr. Yan Sun, who has provided me with invaluable guidance and persistent help. Her strongest support to both my research and personal life has laid the basis for this thesis. Her enthusiasm towards research and positive attitude towards life will be a lifelong inspiration for me.

I would also like to take this opportunity to thank Prof. Steve Uhlig for his incisive and invaluable comments on my thesis. I would like to thank Prof. Yue Chen, Dr. Yue Gao, Dr. Chris Phillips, Dr. Yasir Alfadhl and Dr. Jesus Requena Carrion in our department for their insightful comments during my research stages. I would like to thank Dr. Xu Yang and Dr. Yapeng Wang from the joint lab in MPI for their valuable advice during my visit in Macao.

I would like to thank Prof. Gerard Parr (University of East Anglia) and Dr. Jonathan Loo (Middlesex University London) for agreeing to be my PhD examiners.

I would like to express my special thanks to Prof. Xiaodong Xu from Beijing University of Posts and Telecommunications for his generous sharing of knowledge and insightful suggestions, especially as he spent lots of time on discussing research ideas and development with me. Also I would like to thank his lab for hosting me

as a research visitor in 2015.

Also I would like to express my gratitude to my colleagues and friends for the great moments and friendship. I wish them all the best of luck in the future. Moreover, warm thanks are given to all the helpful staff at Queen Mary who have made it such a great place to study. I would like to thank China Scholarship Council for their financial support during my PhD study.

Last but not least, I would like to thank my dearest parents. They give me enormous support and courage all the time. They are always there for me and have always been my strength especially during hardships. Without their love, I would not have made it through. Their generous support and love are beyond words.

Table of Contents

Abstract	i
Acknowledgments	iii
Table of Contents	v
List of Figures	ix
List of Tables	xii
List of Abbreviations	xiii
List of Publications	xv
1 Introduction	1
1.1 Research Motivation	1
1.2 Research Scope	4
1.3 Research Contributions	6
1.4 Thesis Outline	8
2 Fundamental Concepts and State-of-the-Art	10
2.1 Resource Management for Wireless Communications	10
2.1.1 Evolution of Wireless Communications	11
2.1.2 Overview of RRM schemes	13

2.1.3	Basics of OFDM and OFDMA	15
2.1.4	OFDMA and the LTE Frame Structure	17
2.2	Quality of Service for Wireless Networks	18
2.2.1	QoS Definition	18
2.2.2	RRM for QoS Provisioning	20
2.2.3	QoS Parameters	20
2.2.4	Basics of QoS in LTE	22
2.3	Resource Allocation in OFDMA Networks	25
2.3.1	Sum-rate Maximisation	26
2.3.2	Max-min Fairness	26
2.3.3	Proportional Fairness	27
2.3.4	QoS Satisfaction	28
2.3.5	Discussion	33
2.4	Network Densification	34
2.4.1	Heterogeneous Network Deployments	35
2.4.2	Advantages and Challenges	36
2.5	Cloud Radio Access Network	38
2.5.1	Energy Efficiency Issue	40
2.6	Energy Efficient RRM Schemes for Greener 5G Systems	42
2.6.1	Base Station Sleep Mode Techniques in Green Cellular Networks	45
2.6.2	Joint Design for BS Switch off and User Association	46
2.6.3	Discussion	48
2.7	Summary	48
3	Joint Antenna and Subcarrier Management Scheme	50
3.1	Introduction	50

3.2	System Model	51
3.2.1	Semi-smart Antennas	52
3.2.2	Channel Model and Achievable Rate	53
3.3	Proposed Joint Optimisation Algorithm	54
3.3.1	Subcarrier Allocation	55
3.3.2	Antenna Pattern Optimisation	58
3.4	Simulation Set-up and Results	64
3.4.1	Simulation Parameters	67
3.4.2	Validation of Subcarrier Allocation	68
3.4.3	Verification of Antenna Pattern Adaptation	69
3.4.4	Performance Evaluation	70
3.5	Optimising QoS-aware OFDMA Networks with User Priority	74
3.5.1	Validation of User Priority	75
3.5.2	Satisfaction Improvement with Antenna Adaptation	77
3.6	Antenna Pattern Update Mechanism	80
3.6.1	User movement	80
3.7	Summary	84
4	Resource Management Schemes in Cloud Radio Access Networks	86
4.1	Introduction	86
4.2	Joint Resource Block and Power Allocation with Cloud based Infor- mation Sharing	87
4.2.1	System Model	87
4.2.2	Problem Formulation	89
4.2.3	The Proposed Resource Allocation Algorithm	91
4.2.4	Simulation Set-up and Results	95
4.3	Energy-Effective Network Deployment Scheme	99

4.3.1	System Model	100
4.3.2	Problem Formulation	103
4.3.3	Proposed Network Deployment Scheme	105
4.3.4	Simulation Parameters and Results	108
4.4	Joint Remote Radio Head Selection and User Association Scheme .	113
4.4.1	System Model	115
4.4.2	JRSUA Scheme	119
4.4.3	Simulation Parameters and Results	130
4.5	Summary	142
5	Conclusions and Further Work	145
5.1	Conclusions	145
5.2	Lesson and Learn	147
5.3	Future Work	148
5.3.1	Extension to Current Work	148
5.3.2	Joint Transmission in C-RAN	149
5.3.3	Delay Performance in C-RAN	149
5.3.4	Resilience Mechanism Design in BBU Pool	149
5.3.5	RRH Selection with Traffic Prediction	149
	References	151
	References	151

List of Figures

Figure 1.1	Radio access network architecture: Distributed BS architecture	5
Figure 1.2	Radio access network architecture: Centralised C-RAN architecture	6
Figure 2.1	Maintaining orthogonality between subcarriers	16
Figure 2.2	Resource block structure	19
Figure 2.3	Different bearers employed in end-to-end service delivery in LTE and LTE-Advanced [ATH13]	23
Figure 2.4	Network densification in 5G	36
Figure 2.5	C-RAN architecture: BBU pool, RRH and fronthaul links .	39
Figure 2.6	Traffic dynamics across five cells in urban locality of Seoul, Korea (Fig. 2 from [SRK16])	42
Figure 3.1	A flowchart of the working principle of a GA including population and three genetic operators	60
Figure 3.2	An illustration of the single-point crossover operator. Two parent solutions to create two new children solutions	63
Figure 3.3	An illustration of the mutation operation. The fifth bit is mutated to create a new string	64
Figure 3.4	Flowchart of the antenna management algorithm	65
Figure 3.5	Flowchart of simulation platform for joint antenna and sub-carrier optimisation	66

Figure 3.6	Satisfaction variation against user density	68
Figure 3.7	Satisfaction and convergence against iteration	69
Figure 3.8	Users only at one boundary of Cell 1	70
Figure 3.9	Users at one boundary of Cell 1 and the facing antenna element coverage of Cell 2	71
Figure 3.10	Satisfaction improvement in adaptive antenna pattern com- pared to fixed antenna pattern in different load networks	72
Figure 3.11	GA working on maximising satisfaction ratio	73
Figure 3.12	Satisfaction improvement and the normalised elapsed time comparisons in different generations	74
Figure 3.13	Satisfaction of UC1 users, UC2 users and all users	76
Figure 3.14	Satisfaction comparison with and without priority-aware	77
Figure 3.15	Satisfaction comparison with 30% UC1 and 40% UC1	78
Figure 3.16	Satisfaction improvement over different objectives (optimis- ing all user satisfaction and optimising UC1 satisfaction) with UC1 percentage=40%	79
Figure 3.17	Adaptive patterns with 40% UC1 users	80
Figure 3.18	30 seconds' user random movement	82
Figure 3.19	User satisfaction when having GA triggered	83
Figure 4.1	Flowchart for USU RB reallocation	95
Figure 4.2	Wrap-around matching up virtual cells with real cells	97
Figure 4.3	Satisfaction and system power consumption over iterations with different loads	98
Figure 4.4	Satisfaction comparison of proposed and MaxSINR algorithms	99
Figure 4.5	User satisfaction for different QoS requirement scenarios	100
Figure 4.6	C-RAN deployment illustration with RRH and TDN	101

Figure 4.7	Power consumption with different load in dense scenario . . .	112
Figure 4.8	Power consumption with different load in light scenario . . .	113
Figure 4.9	Power consumption comparison in dense and light scenarios	114
Figure 4.10	Settings of macro RRH, pico RRHs and users	132
Figure 4.11	UA initial result and RB initial utilisation ratios with 15 users	133
Figure 4.12	UA initial result and RB initial utilisation ratios with 16 users	134
Figure 4.13	UA final result and RB final utilisation ratios with 16 users .	134
Figure 4.14	Network power consumption comparison among JRSUA, SWES and GA schemes with 12 pico RRHs	136
Figure 4.15	Network power consumption comparison among JRSUA, SWES and GA schemes with 24 pico RRHs	137
Figure 4.16	Active pico RRHs number comparison among JRSUA, SWES and GA schemes with 24 pico RRHs	138
Figure 4.17	Performance comparison between optimal solution and JR- SUA algorithm in fibre fronthaul	139
Figure 4.18	Performance comparison between optimal solution and JR- SUA algorithm with 5 Mbps wireless fronthaul	140
Figure 4.19	Performance comparison between optimal solution and JR- SUA algorithm with 10 Mbps wireless fronthaul	141
Figure 4.20	Network power consumption comparison among different fronthaul combinations	142
Figure 4.21	User associations with macro and pico RRHs in mixed-scenario 2	143

List of Tables

Table 1-A	Research problem and proposed scheme	4
Table 2-A	The main features of FDMA, TDMA, CDMA and OFDMA .	14
Table 2-B	State-of-the-art resource allocation techniques	26
Table 3-A	Research objectives and contributions	51
Table 3-B	System parameters	67
Table 3-C	GA parameters	67
Table 3-D	Satisfied user number comparison with 600 users	72
Table 3-E	Satisfied user number comparison with different generations .	74
Table 3-F	Random movement parameters	81
Table 3-G	Movement parameters in Macao scenario	82
Table 4-A	Research objectives and contributions	87
Table 4-B	System parameters for simulation	96
Table 4-C	Simulation parameters	110
Table 4-D	GA parameters	111
Table 4-E	Simulation parameters	131
Table 4-F	Simulation elapsed time comparison	138
Table 4-G	Mixed fronthaul settings	142

List of Abbreviations

3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
AN	Access Node
BBU	Baseband Unit
BS	Base Station
C-RAN	Cloud Radio Access Network
EEND	Energy-effective Network Deployment
GA	Genetic Algorithm
GBR	Guaranteed Bit Rate
HPN	High Power Node
IP	Internet Protocol
JRSUA	Joint RRH Selection and User Association
LTE	Long Term Evolution
MMKP	Multi-choice Multidimensional Knapsack Problem
NPCM	Network Power Consumption Minimisation
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access

QoS	Quality of Service
RB	Resource Block
RF	Radio Frequency
RRH	Remote Radio Head
RRM	Radio Resource Management
SINR	Signal-to-interference-plus-noise Ratio
SU	Satisfied User
TDN	Traffic Demand Node
UA	User Association
UE	User Equipment
USU	Unsatisfied User
VoIP	Voice over Internet Protocol

List of Publications

1. **Aini Li**, Yan Sun, Xiaodong Xu and Chunjing Yuan, “Joint Remote Radio Head Selection and User Association in Cloud Radio Access Networks,” in *2016 IEEE 27th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) Inclusive Radio Communication Networks for 5G and Beyond Workshop*, Valencia, Spain, September 2016.
2. **Aini Li**, Yan Sun, Xiaodong Xu and Chunjing Yuan, “An Energy-Effective Network Deployment Scheme for 5G Cloud Radio Access Networks,” in *2016 IEEE Conference on Computer Communications (INFOCOM) 5G and Beyond - Enabling Technologies and Applications Workshop*, pages 684-689, San Francisco, CA, USA, April 2016.
3. **Aini Li**, Yan Sun, Xiaodong Xu and Xu Yang, “QoS-aware User Satisfaction Oriented Joint Resource Block and Power Allocation in OFDMA Systems with Cloud based Information Sharing,” in *2015 IEEE 16th International Conference on Communication Technology (ICCT)*, pages 219-224, Hangzhou, China, October 2015.
4. **Aini Li**, Laurie Cuthbert and Yapeng Wang, “Dynamic Optimization of QoS for Moving Users in an OFDMA Network with Semi-smart Antennas,” in *2015 IEEE/ACM International Wireless Communications and Mobile Com-*

puting Conference (IWCMC), pages 1068-1072, Dubrovnik, Croatia, August 2015.

5. **Aini Li**, Yue Liu, Laurie Cuthbert, Yue Gao and Yapeng Wang, “Optimizing Radio Resources for Heterogeneous QoS-aware OFDMA Networks using Semi-smart Antennas,” in *2014 IEEE International Conference on Communication Systems (ICCS)*, pages 112-116, Macao, China, November 2014.

Chapter 1

Introduction

1.1 Research Motivation

The skyrocketing proliferation of multimedia infotainment applications and high-end devices (e.g., smartphones, tablets, wearable devices, laptops, machine-to-machine communication devices) exacerbates the stringent demand for high data rate services. For the current bandwidth-constrained fourth generation (4G) networks, how to provide a large number of devices with high data rate services remains a challenge. Radio resource management (RRM) has been a critical topic in wireless communications aiming to utilise the limited frequency spectrum resources and network infrastructure as efficiently as possible [STB15].

To provide quality of service (QoS) required by the bandwidth-intensive multimedia applications, RRM schemes must be QoS-driven. To allocate resources, the resource management system must consider resource availability as well as the QoS requirements of individual users [BS09]. A common assumption in previous resource allocation literature is that a user's utility strictly increases with its received data rate. This is true for elastic applications. However for the increasingly popular video traffic, more specific QoS is required. These inelastic applications

can not work properly when their minimum rate requirement is violated, while do not obtain additional benefits when given more resources than needed [SCH13].

Apart from spectrum scarcity, unevenly loaded traffic results in unbalanced performance over the cells, which leads to degraded overall QoS provisioning [SBL14]. Previous studies have indicated that users are often unevenly distributed therefore the number of associated users may vary from cell to cell. This uneven load distribution can degrade the QoS that users experience, especially in highly congested cells [Lag10]. RRM schemes need to be adaptive that match the demand of resources (“load”) with the supply of resources (“capacity”) [ASY⁺14].

Motivated by the abovementioned facts, in this thesis, QoS-aware user satisfaction oriented resource management schemes with load balancing are investigated. We give emphasis to the fast-growing multimedia traffic with specific QoS requirements. In existing spectrum-constrained 4G cellular networks with unbalanced load distribution, we aim to devise joint load balancing and resource allocation schemes to maximise the number of satisfied user whose required QoS is achieved.

From a long-term perspective, incremental improvement of QoS provisioning based on existing 4G cellular networks with macro base stations only is not enough. Now we are facing the rapid evolution towards the fifth generation (5G) era, the demands for high-speed data applications have been growing exponentially recently. According to the latest visual network index report from Cisco [CIS16], the global mobile data traffic will increase eightfold between 2015 and 2020. The first commercial 5G system planned in 2020 is expected to provide approximately 1000 times higher wireless capacity compared with the current 4G system [ABC⁺14].

To achieve the 5G vision, a paradigm shift is needed in radio access networks. In particular, the network densification overlaying existing macro cell

networks with different sizes of small cells, has been recognized as the key approach [SZL⁺15][BLM⁺14]. Meanwhile, cloud computing technology has emerged as a promising solution for providing high energy efficiency as well as gigabit data rate services across wireless software defined communication networks [PLJ⁺14]. Therefore, cloud radio access network (C-RAN) is proposed from both the wireless and the information technology industries as the promising 5G network architecture [CLRH⁺14].

In C-RAN, the traditional base station is decoupled into two parts: the baseband units (BBUs) clustered as a BBU pool with centralised cloud server and the remote radio heads (RRHs) with antennas located at the cell sites. The BBU pool and the RRHs are connected through fronthaul links. In practical systems, the fronthaul is capacity-constrained [PWLP15]. By conducting most signal processing functionality in the BBU pool, RRHs can be relatively simple and can be densely deployed with minimum cost [KAAR15]. Such a drastic shift in the cellular paradigm leads to an open issue about the energy efficiency. Since RRHs will be deployed in a large scale scenario to serve the peak time traffic. It is inevitable that the RRHs will be underutilised most of other times, but still consumes most of the peak power. In the long run, it may lead to deploying more RRHs than the cell phones served [DDD⁺15].

This motivates us to optimise the energy consumption in C-RAN by selecting appropriate subset of RRHs to adapt to the temporal and spatial traffic dynamics. The centralisation at the BBU pool facilitates the implementation of such dynamic RRH selection strategy. User association is jointly considered to guarantee the QoS delivery to end users.

1.2 Research Scope

This research focuses on QoS-aware adaptive resource management in different OFDMA multi-cell network scenarios, which can be summarised in Table 1-A.

Table 1-A: Research problem and proposed scheme

Network scenario	Research problem	Proposed scheme
4G	QoS provisioning improvement: user satisfaction maximisation with load balancing	QoS-aware cell adaptation
Beyond 4G	QoS provisioning improvement: user satisfaction maximisation	Cloud based information sharing
5G	Optimise the energy consumption with QoS guarantee	Energy-effective network deployment; Joint RRH selection and user association

Currently, 4G LTE-based radio access networks (RANs) are implemented using a distributed architecture where base stations (BSs) perform all RAN processing [BRW⁺15]. In Fig 1.1, every BS is responsible of the resource management with respect to the users within its cell. Due to the popularity of bandwidth-intensive multimedia applications, today's cellular network is likely to be resource-constrained to meet the stringent QoS demands. Apart from the scarceness of spectrum resource, QoS provisioning is also constrained by the unbalanced traffic distribution from cell to cell. For the 4G network which is already deployed, we work on improving QoS provisioning using existing network infrastructure and spectrum resources. Chapter 3 proposes QoS-aware user satisfaction maximisation schemes with load balancing.

A natural evolution of the distributed BS architecture is cloud radio access network (C-RAN) [Mob11]. By centralisation and virtualisation, the baseband resources are pooled at the baseband unit (BBU), suited at the cloud centre. The

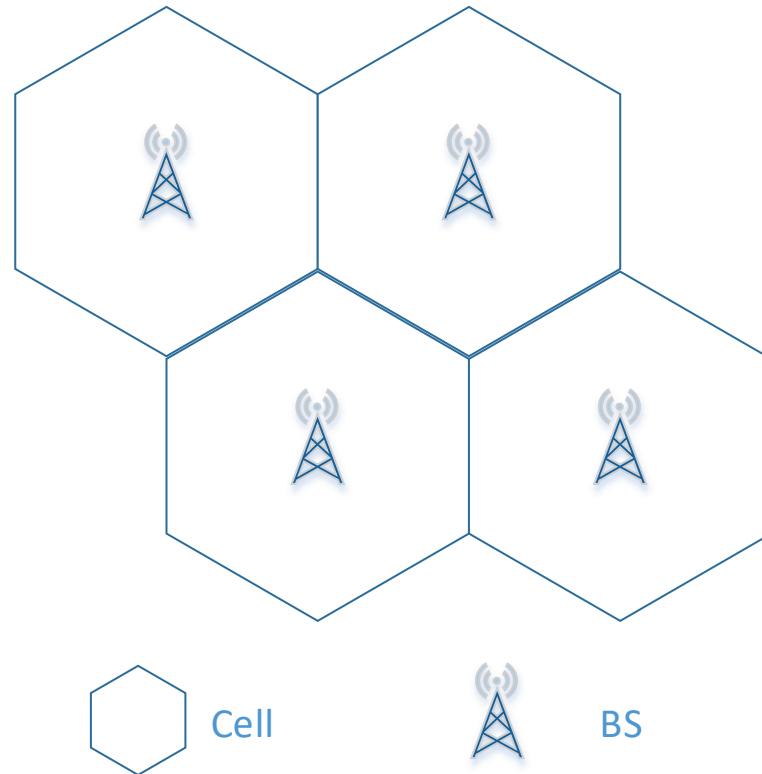


Figure 1.1: Radio access network architecture: Distributed BS architecture

simplified remote radio heads (RRHs) are distributed at the cell sites (Fig. 1.2). BBU pool provides centralised signal processing and resource management across multiple cells. Although C-RAN offers a viable solution to resolve the challenges of rising costs and increasing signalling traffic in dense distributed BS architecture, such a drastic shift in the cellular paradigm leads to an open issue about the energy efficiency and operating efficiency. The primary concern of Chapter 4 is to optimise the energy consumption with QoS guarantee in C-RAN by selecting appropriate subset of RRHs to adapt to the temporal and spatial traffic dynamics.

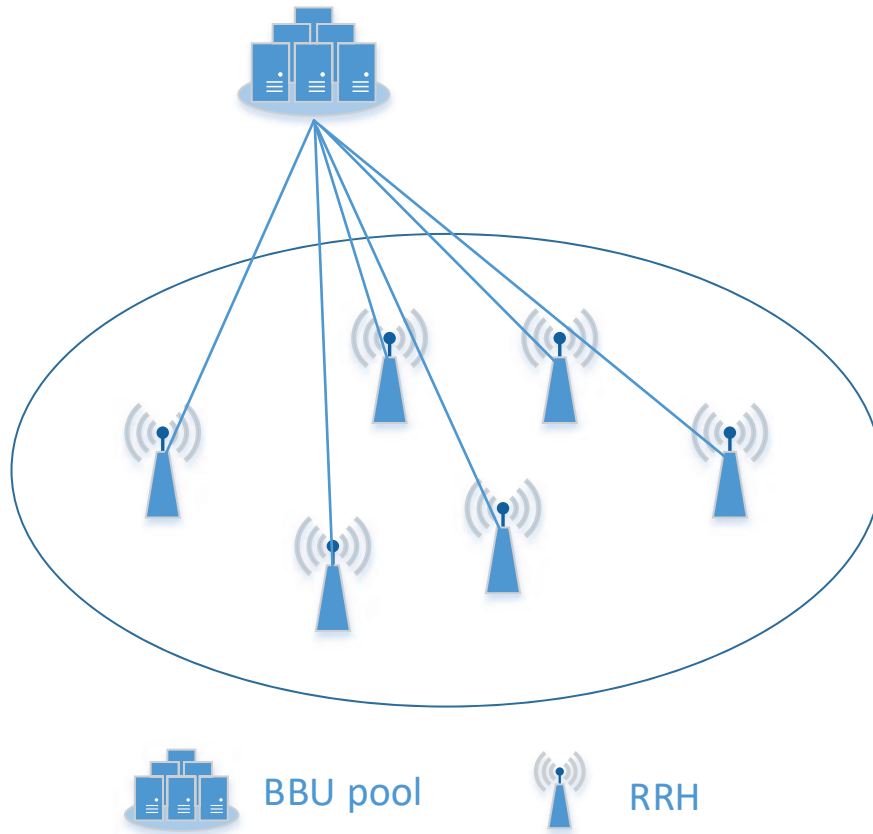


Figure 1.2: Radio access network architecture: Centralised C-RAN architecture

1.3 Research Contributions

This thesis proposes several novel resource management schemes in OFDMA networks. Orthogonal frequency division multiple access (OFDMA) technology has been used in 4G standards and recommended as the default approach in 5G. In resource-constrained 4G scenario, we optimise the user satisfaction. In 5G scenario with network densification, we optimise the energy consumption.

The contributions of this thesis are summarised as follows:

- A joint antenna and subcarrier management scheme is proposed for multi-cell OFDMA downlink. It addresses the problem of QoS-aware user satisfaction maximisation in resource-constrained system with unbalanced load. Semi-smart antennas are utilised to achieve adaptive cell coverage corresponding to the load distribution. The antenna pattern is optimised by genetic algorithm (GA). Subcarrier is allocated in a co-opetition manner according to user's QoS requirements including minimum rate requirement and user/traffic priority.
- Based on the above finding, a dynamic optimisation of QoS with low-overhead for moving users is proposed. Rather than changing the antenna pattern frequently, we monitor the user movement periodically and optimise the semi-smart antennas using GA to update the antenna pattern only when the satisfaction ratio drops below a certain threshold.
- A joint resource block (RB) and power allocation scheme is proposed with cloud based information sharing. The proposed algorithm, based on low complexity heuristics, devises different power-RB polices for satisfied user and unsatisfied user sets respectively, which operates jointly to address the QoS provisioning problem. Cloud based information sharing through the BBU pool is introduced to manage the co-channel assignment in order to mitigate the inter-cell interference.
- An energy-effective network deployment (EEND) scheme is designed to reduce the network energy consumption in dense C-RAN: A new problem formulation scheme is introduced where RRH is treated as a dimension of assignable resource to achieve dynamic network construction; a C-RAN structure is proposed that the system can dynamically select a subset of

RRHs according to the varying traffic demand; an energy-effective deployment is achieved through multi-choice multidimensional knapsack problem based RRH-traffic demand node association and sleeping strategy based active RRH set determination.

- A joint RRH selection and user association (JRSUA) scheme is proposed to minimise the network power consumption in C-RAN. Based on the proposed JRSUA, we formulate an optimisation problem aiming at minimising the total power consumption in the C-RAN while satisfying users traffic demands. Since the RRH selection and user association are mutual dependent, we decouple the network power consumption minimisation problem into two components. First, we solve the user association for given active RRH set using multiple-choice multidimensional knapsack model with consideration of both radio resource and fronthaul capacity constraints. Then, we devise a low complexity heuristic algorithm that selects the best active RRHs by repeatedly solving the user association problem.
- Thorough theoretical analysis is presented in the development of each proposed algorithm. And the effectiveness of each proposed algorithm is confirmed via comprehensive simulation results.

1.4 Thesis Outline

The rest of the thesis is organised as follows.

Chapter 2 provides background information about RRM aspects in wireless communications, OFDMA basics, the concepts of QoS and C-RAN. Furthermore, it presents literature review of state-of-the-art resource allocation approaches in

OFDMA networks and energy efficient RRM schemes, especially sleeping techniques for greener cellular systems and highlights open challenges in this context.

Chapter 3 studies QoS-aware user satisfaction maximisation schemes with load balancing. A joint antenna and subcarrier management algorithm is proposed for multi-cell OFDMA downlink at first. Then a dynamic optimisation of QoS with low-overhead for moving users is proposed. The system model, detailed development of proposed algorithms, simulation set-up and performance evaluation are provided.

Chapter 4 investigates three radio resource management schemes for C-RANs. Firstly, a joint resource block and power allocation algorithm is proposed with cloud based information sharing through BBU pool. Secondly, an energy-effective network deployment scheme is developed to reduce the network energy consumption in C-RAN based small cells. At last, a joint RRH selection and user association scheme is proposed in heterogeneous C-RAN. For each algorithm, system model, problem formulation and performance comparison are presented in details.

Chapter 5 draws the conclusions of this thesis. The ideas for future work based on the research carried out in this thesis are also discussed.

Chapter 2

Fundamental Concepts and State-of-the-Art

2.1 Resource Management for Wireless Communications

Wireless mobile communication has evolved over four generations, with the fifth generation yet to come. As a new generation stepped forward, it accompanied the implementation of new air interface technologies as well as the enhancement of system performances. In the early generations, performance enhancements mainly came from physical layer technologies like modulation and multiple access, but the contribution of radio resource management has increased with each generation. In essence, maximised performance of a wireless system is likely attained by optimised radio resource management implemented on the basis of matured physical layer technologies [LPS09].

2.1.1 Evolution of Wireless Communications

Wireless communication is the transfer of information over a distance without the use of wires, cables or any other electrical conductors [BB10]. In the context of wireless communications, a cellular network or mobile network refers to a communication system divided into small areas, where users are served by a base station with limited range [H⁺07]. Recent decades have witnessed a great evolution of wireless communication systems.

It was in the 1980s when the first generation (1G) analogue systems were widely deployed to support circuit-switched voice telephony [LPS09]. The best known example is the Advanced Mobile Phone System (AMPS), developed by Bell Labs in the 1970s and launched in the United States in 1983. Europe also deployed a similar analogue system to AMPS called the Total Access Communication System (TACS). TACS operated at a higher frequency and with lower bandwidth channels than AMPS [Gol05]. The 1G system used frequency division multiple access (FDMA) technology and analogue frequency modulation [Gar10][SVDK06].

In the early 1990s, the 1G wireless systems evolved to become second generation (2G) systems by virtue of the advances in digital technologies. The 2G wireless systems also were targeted at voice services, but digital modulation and multiple access methods were incorporated as well. Time division multiple access (TDMA) was employed in the Global System for Mobile Communications (GSM), and code division multiple access (CDMA) was used in the IS-95 CDMA system [LPS09]. Once digital cellular became available, operators began supporting data services in addition to voice with packet switching [Gol05].

In the late 1990s, the fragmentation of frequency bands and standards in 2G systems motivated the International Telecommunication Union (ITU) to formulate

a specification for a single global frequency band and standard for the third generation (3G) digital cellular systems. The standard was named the International Mobile Telecommunications 2000 (IMT-2000). In addition to voice services, IMT-2000 was to provide high data rate services such as broadband Internet access, interactive gaming and high quality audio and video entertainment. However, agreement on a single standard did not materialized, with most countries supporting one of two competing standards: wideband CDMA (WCDMA) and CDMA2000. Both standards used CDMA and supported both circuit- and packet-switched services with broader bandwidths [LPS09][Gol05].

In 2008, the ITU issued the International Mobile Telecommunications Advanced (IMT-Advanced) requirements for the fourth generation (4G) systems. As opposed to earlier generations, 4G systems does not support traditional circuit-switched telephony, but based on all-Internet Protocol (IP) packet-switched networks. Two dominant 4G systems are commercially deployed: the IEEE 802.16 based Worldwide Interoperability for Microwave Access (WiMAX) and 3rd Generation Partnership Project (3GPP) based Long Term Evolution (LTE). Both WiMAX standard (both downlink and uplink) and LTE standard (downlink) have selected orthogonal frequency division multiple access (OFDMA) as the physical layer multiple access technology [STB15] [AGM07].

Now, we are facing the advent of fifth generation (5G). 5G will provide an order of magnitude improvement in performance in the areas of more capacity, lower latency, more mobility, more accuracy of terminal location, increased reliability and availability [CZ14]. Several forums and projects such as METIS (Mobile and Wireless Communications Enablers for the Twenty-Twenty Information Society) [FT13][BV13] and IMT-2020 [Mar15][WP514] have been established to shape the 5G vision and study its key enabling technologies. Suggested by the ITU, the

5G deployments will commence from 2020 [Pra14]. OFDM and OFDMA have been recommended as the default modulation and multiple access formats for 5G [ARS16][ZRB14][ABC⁺14]. We will briefly introduce the basics of OFDM and OFDMA in Section 2.1.3.

2.1.2 Overview of RRM schemes

As stated above, the evolution of wireless communications is characterized mainly by a shift in the physical layer technology, i.e., FDMA \rightarrow TDMA \rightarrow CDMA \rightarrow OFDMA [CSSK⁺15], of which the main features [Gar10][STB15] are summarised in Table 2-A. In reality, the enhancements solely due to transmission technology (or physical layer technologies in general) are limited in various aspects. First, transmission technologies are already matured and there is not much scope for further improvements without some unexpected breakthroughs. Second, most transmission technologies are developed to accomplish specific objectives, so each individual technology cannot yield a universally optimal performance solution. This points to the necessity for radio resource management [LPS09].

Radio resource management (RRM) encompasses a wide range of techniques and procedures to manage radio transmission characteristics such as transmit power, frequency channel, antenna beamforming, cell search, cell reselection and so on. The aim of RRM is to utilise the limited frequency spectrum resources and network infrastructure as efficiently as possible [TRV06][STB15].

The ever-increasing size of wireless mobile community coupled with the demands for high-speed multimedia communications stands in clear contrast to the scarceness of spectrum resource that has been allocated in international agreements. Efficient RRM is of paramount importance to make best use of limited

Table 2-A: The main features of FDMA, TDMA, CDMA and OFDMA

FDMA	FDMA is the most common analogue multiple access method. It is a technique whereby spectrum is divided up into frequencies and then assigned to users. With FDMA, only one subscriber at any given time is assigned to a channel.
TDMA	TDMA improves spectrum capacity by splitting each frequency into time slots. TDMA allows each user to access the entire radio frequency channel for the short period of a call. Other users share this same frequency channel at different time slots.
CDMA	CDMA is based on “spread” spectrum technology. CDMA increases spectrum capacity by allowing all users to occupy all channels at the same time. Transmissions are spread over the whole radio band, and each voice or data call are assigned a unique code to differentiate from the other calls carried over the same spectrum.
OFDMA	OFDMA extends the multicarrier technology of OFDM to provide a very flexible multiple access scheme. OFDM subdivides the bandwidth available for signal transmission into a multitude of narrowband subcarriers, arranged to be mutually orthogonal, which can carry independent information streams; in OFDMA, this subdivision of the available bandwidth is exploited in sharing the subcarriers among multiple users.

resource to fulfill the increasing demands [ZKAQ01][PKBV11]. By determining mechanisms to use the available resources optimally, RRM can lead to a significant improvement in transmission rate without using more bandwidth [LPS09].

In general, RRM strategies for wireless communications consists of three basic sets [SA06][CTS06][LPS09][HLN13]:

1. Frequency/time resource allocation schemes such as channel allocation, scheduling, transmission rate control which determine timing, ordering and the amount of spectrum resource to allocate to each user;
2. Power control and allocation schemes which control the transmit power of the base stations and mobile devices and distribute a given amount of transmit power over multiple orthogonal subchannels with the intention to maximise

the efficiency;

3. Call admission control, base station assignment, load management and handover schemes which control access node connection.

2.1.3 Basics of OFDM and OFDMA

Orthogonal frequency division multiplexing (OFDM) is a multicarrier modulation technique. The basic principle of OFDM is dividing a given high-bit-rate data stream into several parallel lower bit-rate streams and modulating each stream on separate subcarriers. Multicarrier modulation schemes eliminate or minimise inter-symbol interference (ISI) by making the symbol time large enough so that the channel-induced delays are an insignificant fraction of the symbol duration. In order to completely eliminate ISI, guard intervals filled with cyclic prefix are used between OFDM symbols. Since each subcarrier has a bandwidth less than the coherence bandwidth of the channel, the subcarriers experience relatively flat fading, thus enabling a low-complexity equalization [AGM07]. While each subcarrier is separately modulated by a data symbol, the overall modulation operation across all the subcarriers results in a frequency-multiplexed signal, so as to resist frequency-selective fading [HAWJ10].

OFDM is a spectrally efficient version of multicarrier modulation, where the subcarriers are selected such that they are all orthogonal to one another and do not interfere with each other. Each of the centre frequencies for the subcarriers is selected from the set that has such a difference in the frequency domain that the neighbouring subcarriers have zero value at the sampling instant of the desired subcarrier [HT09], as shown in Fig. 2.1. An efficient implementation of OFDM transmitter and receiver can be built with the inverse fast Fourier transform (IFFT)

and the FFT, respectively [BJN12].

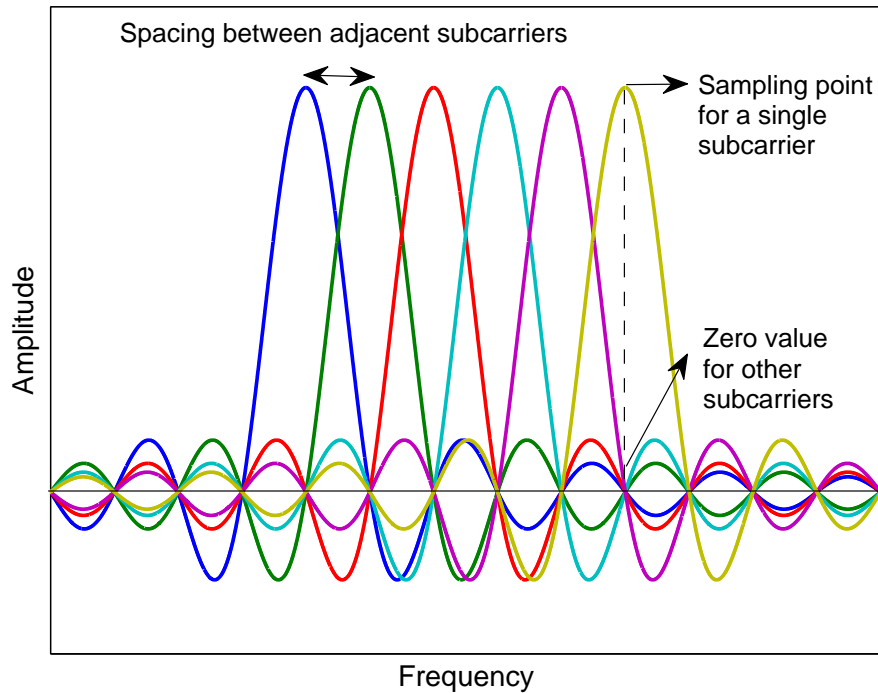


Figure 2.1: Maintaining orthogonality between subcarriers

OFDMA is an extension of OFDM to the implementation of a multi-user communication system. OFDMA distributes subcarriers to different users at the same time, so that multiple users can be scheduled to receive data simultaneously [STB15]. Since the widespread use of digital technology for communications, OFDMA become more feasible and affordable for consumer use. During recent years, OFDMA technology has been widely adopted in many areas. It has been chosen for the cellular telecommunications standard LTE and LTE-Advanced. It is also been used by other standards such as WiMAX and many more. The overall motivation for OFDMA in LTE and in other systems has been due to the following qualities: a natural way to cope with frequency selectivity; computationally efficient implementation of transmitter and receiver; simple frequency domain equaliz-

ers; good spectral properties and handling of multiple bandwidths; link adaptation and frequency domain scheduling, as well as compatibility with advanced antenna technologies [HT09][Dew13].

Given these impressive qualities and the large amount of inertia in its favour, the approach of OFDM and OFDMA is the unquestionable frontrunner for 5G [ABC⁺14][Qua15]. However, there exist some weak points that could possibly become more pronounced in 5G networks. Perhaps the main source of concerns, or at least of open questions, is the applicability of OFDM to millimetre wave spectrum [GTC⁺14]. Second, the spectral efficiency of OFDM could perhaps be further improved upon if the requirements of strict orthogonality were relaxed and if the cyclic prefixes were smaller or discarded [HSL⁺14]. To address the weaknesses, some alternative approaches are being actively investigated, such as time-frequency packing [DRO11], generalized frequency division multiplexing (GFDM) [DPF12] and filterbank multicarrier [FB11]. Most of these can be considered incremental departures from OFDM rather than the step-function changes that took place in previous cellular generations [SGA14].

2.1.4 OFDMA and the LTE Frame Structure

OFDMA is an excellent choice of multiplexing scheme for the 3GPP LTE downlink since it is vastly superior to packet-oriented approaches in terms of efficiency and latency. In OFDMA, users are allocated a specific number of subcarriers for a predetermined amount of time. These are referred to as resource blocks (RBs) in the LTE specifications [Acc16c][Acc16b]. Allocation of RBs is handled by a scheduling function at the 3GPP base station.

LTE frames are 10 ms in duration. They are divided into 10 subframes, each

subframe being 1 ms long. Each subframe is further divided into two slots, each of 0.5 ms duration. Slots consist of either 6 or 7 OFDM symbols, depending on whether the normal or extended cyclic prefix is employed [Acc16b].

The total number of available subcarriers depends on the overall transmission bandwidth of the system. The LTE specifications define parameters for system bandwidths from 1.25 MHz to 20 MHz. An RB is defined as consisting of 12 consecutive subcarriers for one slot (0.5 ms) in duration [Ahm13], which is shown in Fig. 2.2. An RB is the smallest unit of resource allocation assigned by the base station scheduler [Acc16c].

2.2 Quality of Service for Wireless Networks

2.2.1 QoS Definition

Quality of service (QoS) in the field of telecommunications can be defined as a set of specific requirements provided by a network to users, which are necessary in order to achieve the required functionality of an application (service). The users specify their performance requirements in the form of QoS parameters, and the network commits its bandwidth making use of different QoS schemes to satisfy the request. Each service model has its own QoS parameters.

The quality of service can be a differentiator in the business market. Its parameters and measures are necessary to provide an indication of how well a service is, and therefore, it is an important point when selecting services offered by different service providers. If service features or price are similar, quality becomes the differentiator for users, at the same time, service providers can make use of quality

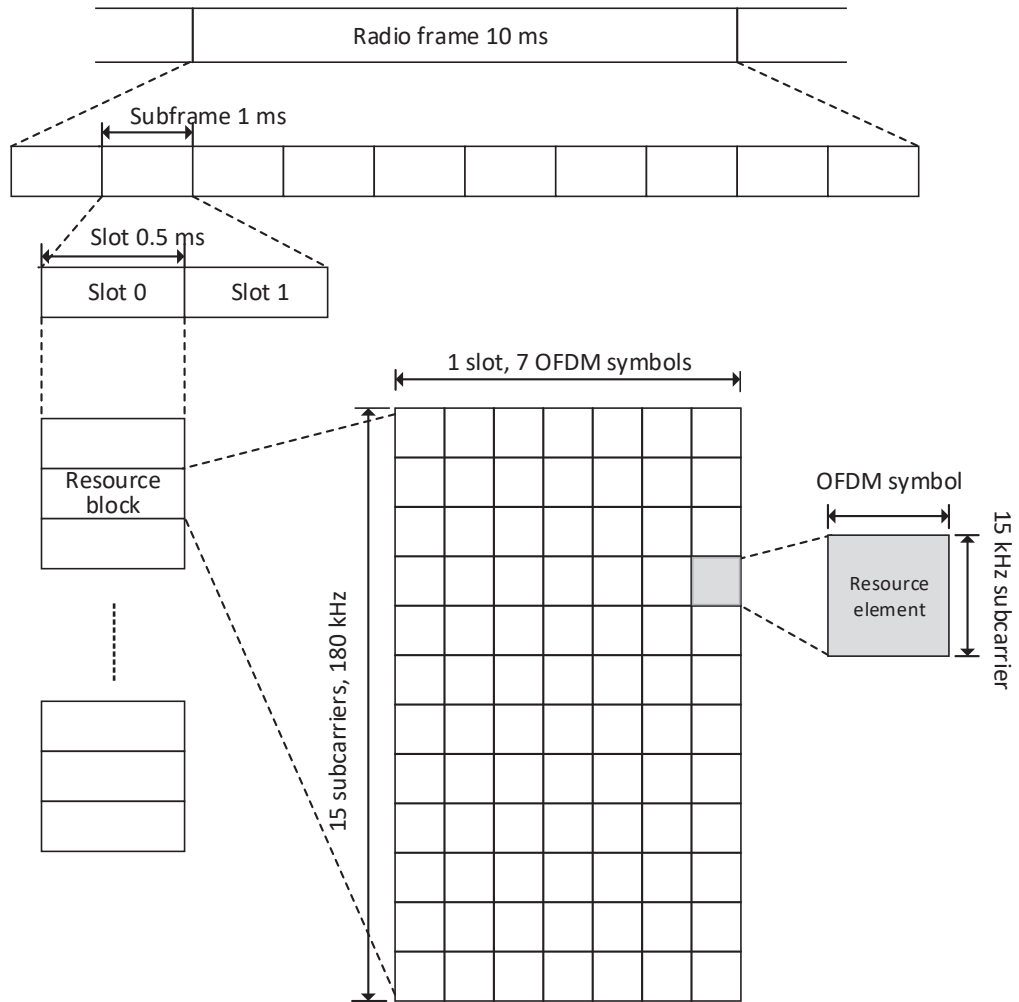


Figure 2.2: Resource block structure

to have an image of a “trusted” provider [BS09].

2.2.2 RRM for QoS Provisioning

Radio resource management (RRM) is considered as one of the most significant and challenging aspects in the provisioning of QoS for wireless networks. Conceptually, RRM policies, in conjunction with network planning and air interface design, determine QoS performance both at the individual user level and at the network level [CTS06].

From the users' point of view, they want the best possible service features such as maximum throughput or lowest block-and-drop rate. Especially, multimedia users are assumed to be selfish and care only about the utility benefits that they can derive from the network [PvdS07]. Each user generally desires to acquire as much of the network bandwidth as possible [ZL11].

Meanwhile, the network wish to serve the maximum number of users possible at a given time, at the agreed quality of service, as cost effectively as possible. The greater the percentage of satisfied users served with good quality and, consequently, the more profitable the network will be.

RRM has an important part to play in helping to match these potentially conflicting sets of requirements [CTS06].

2.2.3 QoS Parameters

To provide and sustain QoS, RRM must be QoS-driven. To allocate resources, the RRM schemes must consider resource availability as well as the QoS requirements of applications, which are quantified by QoS parameters.

QoS parameters are used to identify the quality of service that a certain ap-

plication requires or expects from the network in order to function properly. By identifying these application requirements, intermediate forwarding nodes can assign different priorities and/or resource reservations for different application flows [Adi10].

The most common QoS parameters considered in packet-switched networks are throughput, delay, jitter and packet loss [Lag10][Mel13][BS09].

Throughput, also known as data/bit rate, or capacity is usually denoted in terms of bits per second (bits/s or bps). The throughput represents the amount of data that can be transmitted in a determined time period. Throughput also interferes with delay, since the smaller the throughput, the greater the delay in delivering a packet. As a QoS parameter, the throughput identifies the amount of data per second the application expects to receive in the destination node. This QoS parameter is one of the most critical requirements for real-time applications.

Delay, is also known as latency. The delay shows the time taken for a packet or a set of packets from a flow to leave a sender and arrive at the destination. This is mostly affected by the number of hops between the sender and the receiver as well as the state of the intermediate nodes (e.g., number of packets in their buffers, link and CPU usage, etc.). Delay time can be increased if the packets face long queues in the network, or cross a less direct route to avoid congestion. Delay is related to packet loss due to queue overflow which can happen in situations with very large delays. When used as a QoS parameter, the delay represents the maximum time a packet can take to arrive at the destination and still be valid.

Jitter represents the delay variation of the received packets belonging to the same flow. A small jitter means that most packets have the similar end-to-end delays. On the other hand, a large jitter means that some packets will be received

with relatively small delays while other packets will be received with relative large delays.

Packet loss is the rate in which packets can be lost during the routing process without greatly disrupting the application functions. Wireless and IP networks will fail to deliver, or drop some packets if they arrive when the buffers are already full. As a QoS parameter, the packet loss rate identifies an acceptable loss ratio in which the application can still function.

2.2.4 Basics of QoS in LTE

Since LTE is all-IP based, in LTE network QoS is implemented between user equipment (UE) and Public Data Network Gateway (PGW) and is applied to a set of bearers: radio bearer, S1 bearer and S5/S8 bearer, collectively called as Evolved Packet System (EPS) bearer. A radio bearer is the over-the-air connection. An S1 bearer is the connection between the evolved base station (eNodeB) and the Serving Gateway (SGW). Finally, the EPS bearer is established between the UE and the PGW, as shown in Fig. 2.3. Bearer is basically a virtual concept and is a set of network configuration to provide special treatment to the set of traffic e.g., voice over IP (VoIP) packets are prioritized by network compared to web browser traffic [TAH11] [ATH13].

A bearer can be classified as either a default or a dedicated bearer. When a mobile device first attaches to an LTE network, it is assigned a default bearer, which is associated with the user's IP address. The default bearer is a non-GBR (Guaranteed Bit Rate) bearer which does not have a bit rate guarantee and offers only best-effort service. A dedicated bearer acts as another bearer for QoS differentiation purposes. Dedicated bearers are either non-GBR bearers or GBR

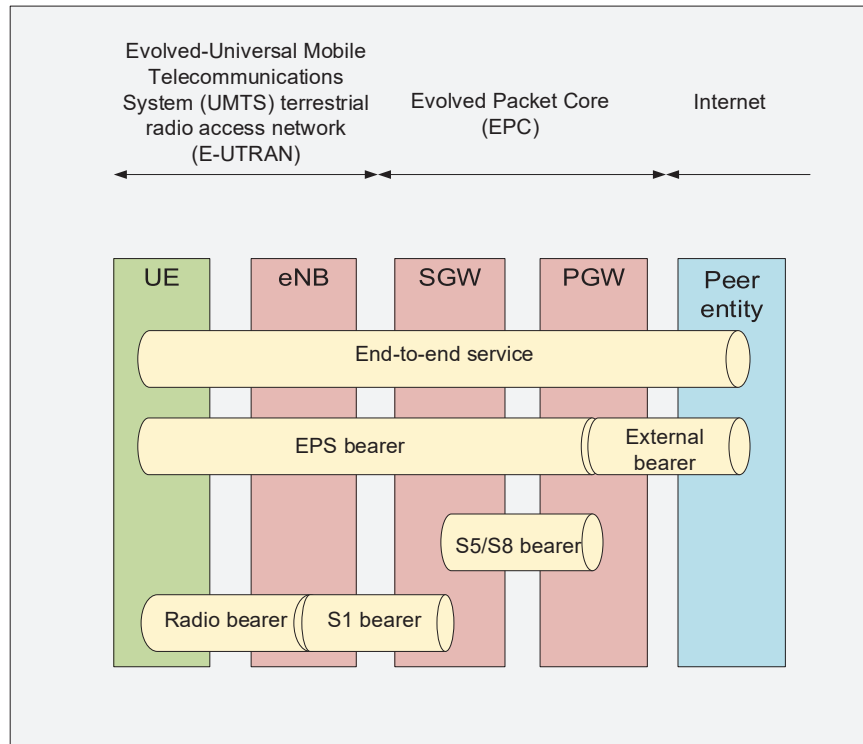


Figure 2.3: Different bearers employed in end-to-end service delivery in LTE and LTE-Advanced [ATH13]

bearers [PHH⁺12].

A GBR bearer is associated with a guaranteed bit rate, which is the minimum bit rate that the mobile is expected to receive. A GBR bearer has dedicated network resources and is suitable for sensitive real-time voice and video applications, in which the guaranteed bit rate might correspond to the minimum bit rate of the user's codec [Cox12]. Take the video codec for example, the utility U_v will be 0 if the minimum required rate is not achieved [PvdS10]. This will result in unacceptable quality to the user according to the peak signal to noise ratio (PSNR), which is a measure of video quality and can be calculated by $PSNR = 10 \log_{10} U_v$ [ZL11][GYZ09][PvdS10]. A non-GBR bearer does not have dedicated resources and is used for best-effort traffic, such as file downloads. Non-GBR bearers do not

guarantee any particular bit rate. These can be used for non-real-time services such as web browsing or email in which the data rate can fall to zero [ETS11][NT16].

In the access network, it is the responsibility of the eNodeB to ensure the necessary QoS for a bearer over the radio interface. Each bearer has an associated QoS class identifier (QCI) by a scalar number [Nak11]. Each QCI is characterized by resource type (GBR and non-GBR), priority, packet delay budget and packet error loss rate. The priority level determines the order in which data packets are handled. Low numbers receive a high priority. The packet delay budget is an upper bound for the delay that a packet receives between the mobile and the PGW [Cox12]. 3GPP defines nine categories for delay, with 50 ms being the tightest and 300 ms the slackest. The packet error rate is an upper bound for the proportion of packets that are lost because of errors in transmission and reception. It has nine categories with 10^{-6} being best and 10^{-2} being the worst [ATH13].

For users associated GBR traffic, they are inelastic in their QoS requirements, a step function can be used to present their utility, which describes user satisfaction:

$$U_G(r) = u(r - R_{min}) = \begin{cases} 1 & r \geq R_{min} \\ 0 & otherwise \end{cases} \quad (2.1)$$

where $u(\cdot)$ is the step function, r is the user's achievable rate and R_{min} is the minimum rate requirement [She95][BD09][KL05].

On the other hand, for non-GBR traffic, the satisfaction utility usually has such feature: it increases with the achievable data rate, however it saturates as achievable data rate increases [SL05][JGL05]. A log based function can be used to grasp such feature:

$$U_n(r) = a + b \log(r + c) \quad (2.2)$$

where a , b and c are properly chosen constants [BD09].

2.3 Resource Allocation in OFDMA Networks

In cellular networks, resource allocation refers to the aforementioned (Section 2.1.2) channel allocation or/and power allocation [HLN13]. In OFDMA-based networks, resource allocation schemes typically includes assigning a subset of subcarriers to the users and distributing the power amount over each used subcarrier [SBL14]. Resource allocation in OFDMA networks has attracted extensive attention and been intensively studied over the last decade.

Efficiency and fairness are two crucial issues in resource allocation for wireless communication systems [RC14][ZL11]. Spectral efficiency is defined as the data rate per unit bandwidth and is calculated by dividing the sum-rate of a system by its total bandwidth [BB99]. Fairness, on the other hand, indicates how equally the resources are distributed among the users. It could be defined in terms of bandwidth where each user is assigned an equal number of subcarriers [OOaDTH05], or it could be in terms of power where each user is allocated equal portion of the power from the budget. It could also be in terms of data rate, where the objective is to allocate the resources to the users so that all the users achieve the same data rate [RC00a].

The major state-of-the-art resource allocation techniques reported in the literature are: 1) sum-rate maximisation; 2) max-min fairness; 3) proportional fairness and 4) user satisfaction maximisation. The main principles of these algorithms are summarised in Table 2-B.

Table 2-B: State-of-the-art resource allocation techniques

Sum-rate maximisation	Allocate most resources to the users with the best channel conditions
Max-min fairness	Assign more resources to users exhibiting poor channel conditions
Proportional fairness	Users with low average rates benefit more from being scheduled than users with high average rates
User satisfaction maximisation	Allocate resources in a co-opetition manner: a satisfied user has no inclination to change the choice of resource

2.3.1 Sum-rate Maximisation

Schemes with sum-rate maximisation approach [KL06b][HRW⁺07] [CTZK09] [HZN10] [ZW13] [TDA11] aim to maximise total sum-rate over all users of the system, or system throughput subject to the constraint on the total power expenditure. These schemes consider the overall throughput of the network rather than each user's achievable data rate. The basic idea is to allocate most resources to the users with the best channel gains, which is unfair to the users far away to the base stations or with bad channel conditions. Besides, it may result in QoS violation to them if they are multimedia users with demanding rate requirements. Although the sum-rate of a system provides a good measurement of the spectral efficiency, it is not a valid indication of each user's QoS satisfaction [SAR09].

2.3.2 Max-min Fairness

One approach to achieve both spectral efficiency and fairness is the max-min rate maximisation fairness. The max-min fairness approach has been described in [RC00b] and studied in [EIE14][Le12][NSSL11][SDYZ10], whereby maximising the minimum data rate across all users. Roughly speaking, the goal of the max-

min fair approach is to optimise the performance of the worst link amongst all users. The idea behind the max-min fair approach is to treat all users as fairly as possible. However it is inappropriate when different users have different priorities [SBL14]. When there are insufficient resources, it may lead to the case that users share the resources so fairly that none of them gets the desired QoS. Apart from the drawback in terms of QoS provisioning, under the max-min fair solution, some users may consume significantly more bandwidth than others, at the cost of a reduction in the overall throughput of the network [ZL14][TK04].

2.3.3 Proportional Fairness

Another approach to accomplish efficiency fairness is proportional fairness. According to Kim and Han [KH05], a proportional fair optimisation problem should maximise the sum of logarithmic average user data rates. Proportional fair allocation schemes have been investigated in [LQZZ11][FWZW10][CHK11]. It should be noted that the proportional fair approach heuristically tries to balance the fairness among users in terms of outcome of throughput, while implicitly maximising the system throughput in a greed manner. The proportional fairness is a pure outcome fairness metric, which is simple to use, but does not guarantee fairness in a strict sense [Adi09]. The main limitation of the proportionally fair allocation is that utility (maximisation) functions are commonly assumed to be concave. Lee et al. [LMS04b] show that, if it is applied to non-concave utility functions (e.g., for multimedia communication), the system can be unstable and cause congestion in the network.

2.3.4 QoS Satisfaction

With the explosive growth of the Internet and the rapid advance of compression techniques, delay-sensitive and bandwidth-intensive multimedia applications get more and more popular. These multimedia services such as videoconferencing and video streaming have specific QoS requirements (such as minimum data rate) to achieve the required functionality [ZL11]. Therefore, resource allocation schemes should be equipped with QoS provisioning.

2.3.4.1 QoS as Constraints

In most previous works, QoS is considered in the form of constraints based on the abovementioned or other utility optimisation approaches.

Sum-rate maximisation approach based schemes with QoS consideration have been investigated in [CTZK09][LCWY12] in multi-cell scenario. A low-complexity multi-cell OFDMA downlink channel assignment method is proposed in [CTZK09] to maximise the total capacity with subchannel demand constraint using graphic framework. Based on the assumption that all subchannels are statistically equal, they use subchannel demand to approximate the throughput requirement. The designed scheme in [LCWY12] combines soft frequency reuse and power control to maximise the system throughput while satisfying the minimum target data rate of each user for OFDMA uplink.

An adaptive power and bandwidth allocation scheme is devised in [GZAE10b] to achieve proportional fair scheduling under delay constraints for real time sessions in single-cell OFDMA-based wireless systems. Delay and packet loss requirements are translated into rate requirements and convex optimisation techniques are used to

find the solution. The authors in [ZTH⁺10] considers the objective of network-wide proportional fairness through Hungarian algorithm based intra-cell fast scheduling as well as inter-cell interference coordination with QoS guarantees.

Based on dynamic programming and branch-and-bound, both near-optimal and optimal channel and power assignment algorithms are proposed in [LCS09] to satisfy multi-user rate requirements in a single-cell OFDMA downlink with the minimum total transmitted power. The power minimisation subject to rate constraint problem has also been studied in [KBCH10a][KBCH10b] in a two-cell downlink OFDMA system impaired by multi-cell interference. [KBCH10a] provides a binary form optimal resource allocation solution and [KBCH10b] gives a asymptotic analysis based practical solution.

In [XLZ⁺12], subcarrier assignment and power allocation algorithms are designed to maximise generalized energy efficiency for the downlink transmission and minimum individual energy efficiency for the uplink case in single-cell OFDMA network, both under certain prescribed per-user traffic-related minimum rate requirements. Based on bisection power adaptation, brute-force search and heuristic subcarrier assignment algorithms are proposed to obtain optimal and suboptimal solutions. The authors in [WZS13] propose the semi-Markov decision process based stochastic optimisation scheme for QoS guaranteed OFDMA multi-cell cooperation networks. Their objective was to maximise the energy efficiency at the base stations while ensuring the targeted QoS-guaranteed transmission rate for mobile users.

One limitation is that all the above literature in this section have analysed the resource allocation problem in the adequate resources system that can meet all the users' data rate requirements. However the traffic demands grow rapidly,

driven largely by wireless and mobile devices, and multimedia services like video streaming [Cis13]. Multimedia applications usually have high rate requirements. Multimedia content, and specially video streaming, requires per-user data rates of hundreds kilobits per second in order to be of useful quality [SHBBJ13]. These translate into a heavy demand for the spectral resources. Therefore a more realistic scenario is that the bandwidth-limited wireless networks cannot provide the QoS to all the users [GZAE10a]. Another drawback of the above maximisation approaches is that user's utility is assumed to be strictly increasing with received data rate. This is true, for example, when users are running elastic applications such as file downloading. However for multimedia applications, such as VoIP and video streaming, they cannot work property when their required data rates are violated, but do not obtain additional benefits when given more resources than needed [LMS04a].

2.3.4.2 User Satisfaction Maximisation

The above two considerations motivate us to study the user satisfaction approach aiming to maximise the number of satisfied users. User satisfaction approach introduces a judicious mixture of competition and cooperation. The idea behind this judicious mixture is co-opetition, a concept from economic [BN97]. To allow co-opetition, a user is called satisfied user if its achieved QoS is above or equal to predefined QoS threshold. Rather than assuming that users wish to increase their data rates whenever possible, we assume that each user has a predefined QoS requirement. If the requirement is satisfied, then the user has no inclination to change the choice of resource [SCH13][GYZ09].

Based on low complexity heuristics, a satisfaction oriented resource allocation

(SORA) is proposed in [SLFC07] for OFDMA systems to maximise the number of satisfied users. Based on average data rate, it divides the users into two sets: satisfied and unsatisfied. Satisfied user does not transmit in this transmission time interval (TTI), and subcarriers are allocated to the unsatisfied users in a round-robin fashion. When a user achieves the required data rate for the current TTI, it is moved to the satisfied user set. It adopts uniform power distribution among subcarriers. The authors of [SLFC07] further extends the SORA algorithm with beam blocking and self-interference avoidance in orthogonal random beamforming multiple antennas scenario to exploit spatial multiplexing in [MJC10].

Research work [WWSZ10] proposes a user satisfaction based subcarrier allocation algorithm for OFDMA relay networks in the resource-constrained system. It divides users into real-time user and non-real-time user and the objective is to maximise total user satisfaction. The proposed iterative solution includes deciding the transmission link: direct or relay link; and then assigning subcarrier to the user which can get the maximum rate; reallocating the subcarriers to meet real-time user's data requirement and adjusting the allocation among the non-real-time users. Equal power allocation is assumed.

However [SLFC07][MJC10][WWSZ10] all consider single-cell scenario. The limitation is twofold. The first is that in a multi-cell environment which is a more realistic scenario, the inter-cell interference may impair user's QoS satisfaction. The other limitation is that the algorithms in [SLFC07][MJC10][WWSZ10] cannot tackle the challenge of mutual dependency of signal-to-interference-plus-noise ratio (SINR). In multi-cell OFDMA networks, the SINR used to calculate the achievable data rate is unknown before the resource allocation.

[LCYW12] proposes a non-cooperative game based uplink subcarrier assign-

ment algorithm to maximise user satisfaction for multi-cell OFDMA-based cognitive radio networks. Each cell acts as a player and individually controls the strategy of subcarrier allocation in order to maximise its own utility. The utility in [LCYW12] is to maximise the total number of users whose QoS requirements is achieved. Based on the interference feedback, an iterative solution is proposed to solve the subcarrier assignment problem. Power allocation is also fixed.

Except for [ZTH⁺10], all above multi-cell QoS satisfaction schemes in Section 2.3.4 deal with uniform distribution of users across the entire network. However several studies have showed that in multi-cell scenarios users are often unevenly distributed in space and, hence, the number of associated users may vary from base station (BS) to BS [Lag10]. This may translate in an uneven load distribution which can degrade the QoS that users experience, especially if some BSs result in highly congested [SBL14]. For the network with non-uniformly distributed users, the performance of those schemes will be limited due to the unbalanced load among different cells. For example, if one cell is under-loaded and its neighbouring cells are overloaded with all subcarriers used, the free subcarriers in the under-loaded cell cannot be utilised to improve the service in neighbouring cells. Besides, using adaptive cell coverage, user's interference can be alleviated by adjusting the coverage cooperatively. Previously cell edge users with high interference could be relieved by amplifying the serving BS's antenna gain and reducing the interference BS's antenna gain [YWZC10]. Load balancing is an effective RRM function (set 3 in Section 2.1.2) that provides dynamic load re-distribution in real time according to current geographic traffic conditions. It can be used to improve the system performance for any distributed systems containing unevenly distributed traffic, especially for resolving the traffic hotspots [SAR⁺10].

In [ZTH⁺10], the unsatisfied cell-edge users in an overloaded cell send to the

neighbouring BSs a request of restraining them from using the same channels. Thus under-loaded cells make concessions to the requesters and shoulder part of the heavy burden of the overloaded cell. It realizes load balancing in a new light. However this scheme depends on significant backhaul communications between users and the dominant interfering BSs. Work [KLL09] presents a minimum data-rate guaranteed allocation algorithm with load balancing in a pseudo cell structure, formed by the adjacent major-interfering sectors. To achieve load balancing, the frequency reuse factor of every subcarrier in neighbouring cells is dynamically determined. The problem with this approach is that it needs a large amount of signalling exchange between neighbouring BSs in pseudo cells. Without increasing the signalling traffic between users and neighbouring BSs or among adjacent BSs, a semi-smart antenna based geographic load balancing scheme is proposed to improve system total capacity in [YWZC10]. It takes transmit power, cell interference and traffic densities and patterns into consideration at the same time. A drawback is the scheme implements simple fair subcarrier allocation in which each user is allocated same number of subcarriers. This subcarrier allocation method fails to fulfill QoS provisioning to individual users.

2.3.5 Discussion

With bandwidth demanding GBR traffic such as real-time voice and video becomes more and more popular, today's mobile network is likely to be resource-constrained to meet the ever-soaring demands. One important feature about GBR service is that user's satisfaction is a step function with received data rate. Therefore, the most common resource allocation optimisation approaches including sum-rate maximisation and proportional fairness which assume user's utility strictly increases

with bit rate are not appropriate especially in resource-constrained system. If a user's required bit rate is achieved, giving it additional resource to further improve its rate would be a waste to other users who want to be served. Rather than optimising rate subject to QoS of all users achieved, user satisfaction maximisation which optimises the total number of satisfied users would be more suitable.

Apart from the contrast of high traffic demand and scarce bandwidth resource, another challenge of QoS provisioning is that the traffic is often unevenly distributed from cell to cell. This may degrade the QoS experience to the users in a congested cell. Therefore, resource allocation scheme should incorporate load balancing functionality to enhance QoS provisioning in multi-cell scenario. To make our own contribution to address the problem of QoS-aware user satisfaction maximisation with load balancing, a joint antenna and subcarrier management scheme is proposed in Chapter 3.

2.4 Network Densification

The aforementioned RRM schemes provide incremental improvements to the already deployed 4G system. However the 5G will need to be a paradigm shift bringing new unique network and service capabilities [ABC⁺14] [TGW⁺14]. Cisco's annual visual network index reports have provided quantitative evidences that the wireless data explosion is real and will continue. The latest report [CIS16] has foretasted that by 2020, global mobile data traffic will increase eightfold; traffic from wireless and mobile devices will account for two-thirds of the total IP traffic; the number of devices connected to IP networks will be three times as high as the global population in 2020. Fuelled by the popularity of smartphones, tablets and video traffic, the forecast makes it plain that an incremental advance

on current mobile networks will not meet the demands that network will face in 2020 [ABC⁺14]. Network densification is one of the key mechanisms for wireless evolution into 5G.

2.4.1 Heterogeneous Network Deployments

The vision set by 5G to achieve remarkable improvements in data capacity pushes the wireless research communities in both academia and industry to increase the density of base station deployments, particularly in urban areas, in order to provide better throughput and coverage performance [DGK⁺13]. Since deployment of additional macro base station involves significant cost and elaborate site planning, low-power nodes (i.e., small cells, which may be employed indoors or outdoors) offer a simpler cost-effective solution. According to the 3GPP [3GP13], a heterogeneous network may consist of different tiers of infrastructure elements/access nodes (ANs), such as macro, micro, pico and femto base stations. Outdoor small cells deployed by an operator, commonly known as pico cells, typically use a transmit power of 30 dBm [BLM⁺14].

In 5G networks, network densification has a fundamentally different and more important role than that in 4G. Network densification in 4G is recommended as a complement for cellular networks by using low-power ANs like pico BSs to enhance the capacity and coverage in partial areas, such as hotspot and indoor scenarios [GTM⁺16]. However the 5G dense network is proposed as the dominant theme to deploy in overall cellular scenarios. It is reasonable to expect constantly decreasing cell size and increasing cell number [DGK⁺13]. Thus, in theory, cells can shrink until nearly every AN serves a single user (Fig. 2.4). This allows each AN to devote its resources to an ever-smaller number of user [ABC⁺14].

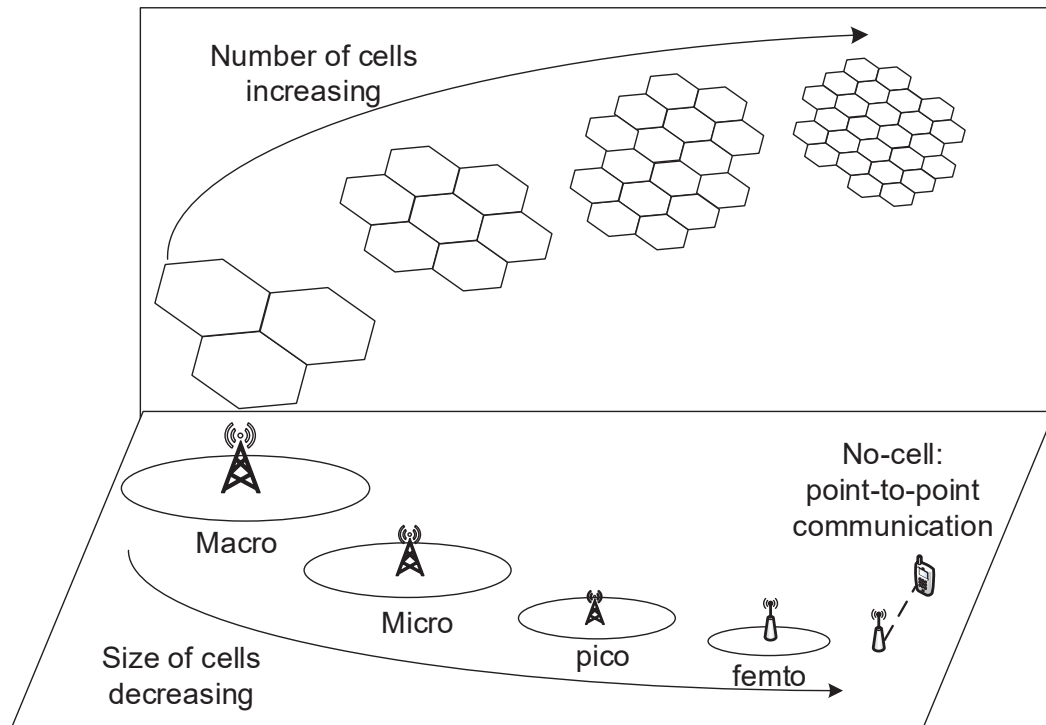


Figure 2.4: Network densification in 5G

2.4.2 Advantages and Challenges

Making the cells smaller is a straightforward and effective way to increase the network capacity [ABC⁺14]. In particular, network densification offers the advantage of proximal communications, which in turn provides the means towards fulfilling the following critical communication principle.

1. Spatial reuse of system bandwidth: from a system perspective, network densification enables extreme spatial reuse of system bandwidth across a geographic area and ensures reduction in the number of users competing for resources at each access node. With an increase of access node density is directly translated into an increase of total system capacity [BLM⁺14].

2. Reduction of path loss effects: by bringing the access node closer to the user, the large-scale fading can be reduced. By decreasing the propagation distance, the energy used for signal transmission will be reduced, which is helpful to increase the power efficiency of the radio access network [DDD⁺15] and extend the battery stand-by time of the user equipment [Mob11].

However as the densification becomes extreme, some challenges arise:

1. Affording the rising costs of installation and maintenance: evolving to ever-smaller cells requires ever-smaller, lower-power and cheaper ANs [And13]. Nevertheless, obtaining permits and paying large monthly site rental fees for operator-controlled small cell placements have proven a major hindrance to the growth of pico cell, distributed antennas, and other enterprise-quality small cell deployments [ABC⁺14].
2. Increasing the signalling traffic for coordination: the process of coordinating the transmission and reception requires significant signalling overhead and message passing between the BSs of different tiers, so as for interference management. Messages and signalling are exchanged between pairs of BSs via connecting links that are typically capacity limited, so such signalling may not always be feasible [DDD⁺15].

Cloud radio access network (C-RAN) offers a viable solution to resolve the above challenges by centralisation and virtualisation [ARS16]. The baseband resources are pooled at the baseband unit (BBU), situated at the central office [Cvi14] and the simplified remote radio heads (RRHs) are distributed at the cell sites. Through the shared use of storage or computing resources, the total cost of ownership, especially capital expenditure (CAPEX) and operating expenditure (OPEX) can be saved [DGK⁺13]. Most of the control signalling takes place at the central cloud

rather than between the connecting links at the cell sites [DDD⁺15]. This top-down architectural change is paving the way for dense 5G deployment by making it affordable, flexible and efficient [AIS⁺14].

2.5 Cloud Radio Access Network

Inspired by the green soft cooperative, cloud and clean access networks in [CLRH⁺14], the cloud radio access network has been proposed from both operators (e.g., NTT, KT, France Telecom/Orange, Telefonica, SoftBank/Sprint, and China Mobile) as well as equipment vendors (e.g., Alcatel-Lucent Light Radio [Luc12], Nokia-Siemens Liquid Radio [Net11]) by incorporating cloud computing into radio access networks (RANs).

The general architecture of C-RAN consists of three main components, namely (i) BBU pool with centralised processors, (ii) RRHs with antennas located at the remote sites, (iii) fronthaul links which connects the RRHs to the BBU pool [PLZW15], as depicted in Fig. 2.5.

In C-RANs, the traditional base station is decoupled into two entities: a base-band unit and a remote radio head. Each virtualised BBU consists of a virtual machine hosting a BBU placed in a metro data center and connected by virtual links to a set of RRHs (i.e., a set of cells) [CCMM16]. A large number of BBUs are clustered together as a BBU pool in one big cloud centre [JMSC15]. A BBU pool can be located at a convenient, easily accessible site, enabling cost savings on site rental and maintenance [CCY⁺15]. BBU pool provides centralised signal processing and resource management across multiple cells, where the cloud computing technology enables flexible spectrum management and advanced network

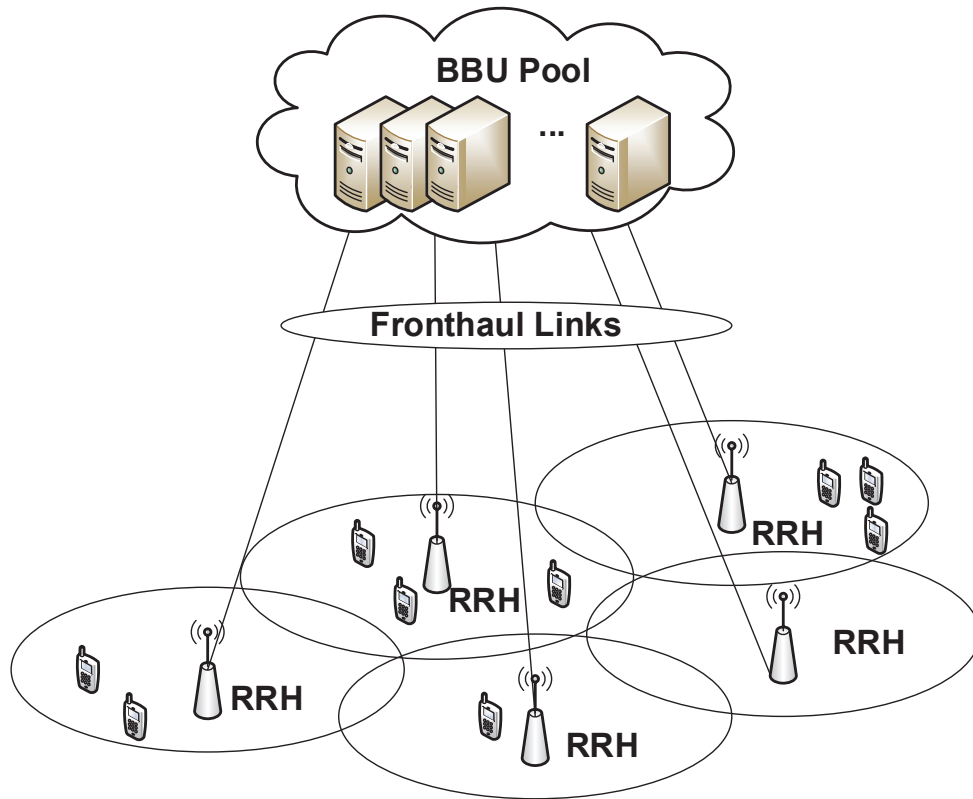


Figure 2.5: C-RAN architecture: BBU pool, RRH and fronthaul links

coordination [PWLP15]. The centralised manner also provides the means to selectively turn RRHs on/off in line with the traffic fluctuations in different scenarios [WHY14]. BBU resiliency against network and processing failures is critical for C-RAN deployments, which can be achieved by dedicated virtual link protection [CAK⁺16] and resilient virtual machine placement [CMDTM16].

The remote radio heads (RRHs) are relatively simple, light-weight radio units with antennas [SAS⁺16]. RRHs are mainly used to provide data rate for user equipments (UEs) with a basic wireless signal coverage, by transmitting radio frequency (RF) signals to UEs in the downlink and forwarding the baseband signals from UEs to the BBU pool for further processing in the uplink. In general, RRHs perform RF amplification, up/down conversion, filtering, analogue-to-digital con-

version, digital-to-analogue conversion, and interface adaptation. By conducting most signal processing functions in the BBU pool, RRHs can be relatively simple and can be placed up on poles or rooftops [CCY⁺15]. RRHs are distributed on the cell sites and can be densely deployed in a cost-efficient manner [PWLP15].

Fronthaul is defined as the link between BBUs and RRHs. Fronthaul can be realized by different technologies, such as optical fibre communication, standard wireless communication, or even millimetre wave communication. Generally, fronthaul falls into two categories: ideal without any constraints, and non-ideal. Optical fibre is considered to be the ideal fronthaul for C-RANs because it can provide a high transmission capacity but at the expense of high cost and inflexible deployment. In contrast with optical fibre, wireless fronthauls employing the cellular transmission or microwave communication technologies are cheaper and more flexible to deploy, therefore are anticipated to be prominent in practical C-RANs. However, the capacity they can provide is limited [WHY14][PWLP15].

2.5.1 Energy Efficiency Issue

By conducting most signal processing functions in the BBU pool, RRHs can be relatively simple and distributed in a large scale with a cost-effective manner. This leads to open issues about energy efficiency and operating efficiency since in most circumstances more RRHs than needed are deployed to meet the estimated highest service demands at all times. Due to the variation of spatial traffic, it would be feasible to switch off some RRHs.

The high density of RRHs results in severe interference and also inefficient energy consumption. First, close proximity of many RRHs results in increased interference, and hence the transmit power of RRHs and/or UEs needs to be

increased to meet any given QoS. Second, the amount of energy consumed by a large number of RRHs [Acc14] as well as by the fronthaul network to support the connections with the BBU pool [TMW⁺11] will also become considerable.

The introduction of dense 5G networks consisting a wide number of small cells also threatens to increase the operating energy costs, thereby aggravating the detrimental greenhouse (CO₂) gas emissions [SRK16]. With the increase of global energy consumption in wireless communication, energy issue has gained a lot of attentions [SCSY14]. Existing work on cellular traffic distribution [PSBD11] has already pointed out time-varying, space-varying cellular traffic patterns. Fig. 2.6 shows a week-long cellular traffic dynamics, captured across five cells in a dense, urban locality of Seoul, Korea. It can be seen that the traffic profile of the night-time is much lower than that of the day-time. It is also observed that there is a difference between the traffic profiles of weekdays and weekends. The figure is consistent with the data presented in [PSBD11]. Since the operators need to deploy their access nodes to support the peak time traffic, it is inevitable that the access nodes will be underutilised most of other times, especially, at night and on weekends. In fact, in densely deployed networks, the number of users associated with each access node is small, which leads to higher traffic dynamics among access nodes [GZNY10]. The higher the deployment density, the higher the chance that access nodes will carry no traffic or only a low traffic load due to spatial and temporal traffic fluctuations [WTN14]. Note, however, that access nodes consume most of their peak power even when they are in little and no activity [CZB⁺10]. Therefore, an effective way to achieve energy saving in mobile communication networks is to dynamically switch off access nodes, especially for scenarios with low traffic load where less access nodes can meet the traffic needs of all users [FJL⁺13].

Such facts motivate us to select appropriate RRHs to adapt to the temporal

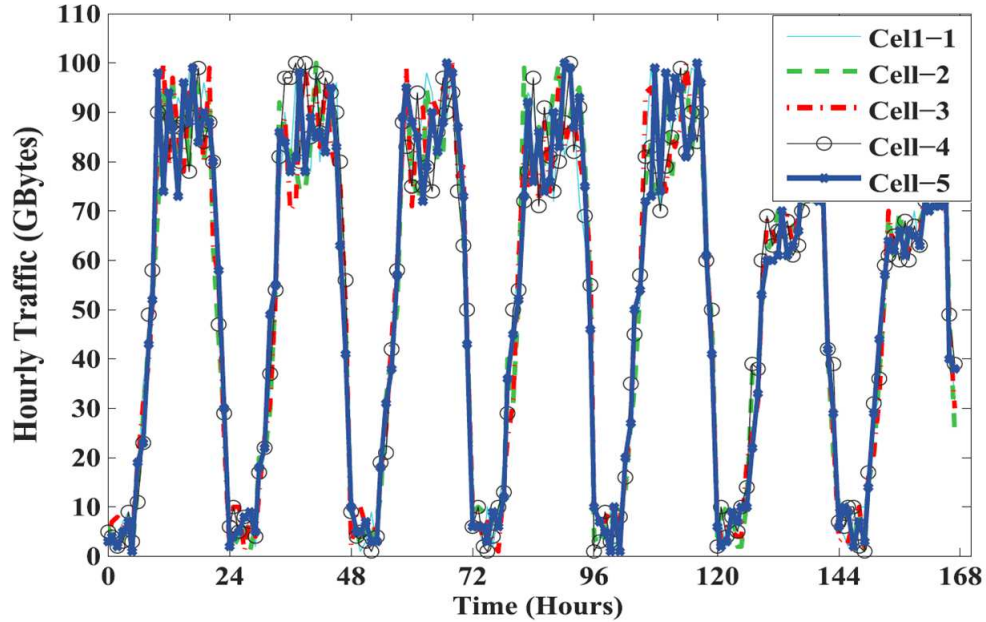


Figure 2.6: Traffic dynamics across five cells in urban locality of Seoul, Korea (Fig. 2 from [SRK16])

and spatial data dynamics, thereby optimising the energy consumption in C-RAN, which is the primary concern of our work in Chapter 4. In addition, such an RRH selection strategy is easy to implement in C-RAN architecture.

2.6 Energy Efficient RRM Schemes for Greener 5G Systems

Recently, certain research efforts have been devoted into developing energy efficient RRM schemes in C-RANs. However in most of prior works, the global network parameters and the network topology are static. For instance, the authors in [PZJ⁺15][CGF⁺15][LCF14][CJL⁺14] propose their energy efficiency (EE) schemes based on fixed network construction and traffic load.

Authors in [PZJ⁺15] investigate the joint resource block and power optimisation problem to maximise EE performance in the OFDMA-based C-RAN system. Research work [CGF⁺15] proposes a self-adaptive power allocation scheme in software defined C-RAN indoor scenarios aiming to save energy consumption. A C-RAN based power allocation scheme is proposed in [LCF14], working under large-scale multiple-input multiple-output (MIMO) scenario with enhanced energy efficiency. C-RAN architecture is implemented into a heterogeneous network structure in [CJL⁺14] and this implementation shows EE advantages by cooperative transmission. A simple but efficient pre-coding antenna beamforming is proposed to reduce the computation complexity of BBU.

However the traffic load fluctuates significantly over time and locations. RRH in different areas will experience different traffic demand fluctuations through the day. For example, RRHs located in office areas will experience the highest traffic demand during daytime, while in the evening they will remain underutilised. On the contrary, RRHs located in residential areas are underutilised in daytime. Such spatial-temporal fluctuations result in large capacity surpluses when system is underutilisation. Thus identifying this leads to an energy savings opportunity via network adaptation to the actual traffic demand.

From the consideration of traffic variations, scanning recent work in this area, most research attentions have been paid to BBU management. In [WZZ14], a graph-based dynamic frequency reuse scheme is proposed in C-RAN to reduce energy consumption. Under this scheme, the number of active BBU is reduced. By allowing flexible mapping between BBU and RRH due to cloud-based centralised baseband processing, the number of BBUs is reduced to acquire more energy gains in [SPH15] using bin packing algorithm. A traffic load balancing based strategy is designed in [KAAR15], to assign a minimum number of active BBUs to the

RRHs so that power saving is maximised. In this strategy, if the resource usage of one BBU reaches the upper limit, partial traffic of this BBU will be offloaded to another light-loaded or sleeping BBU; in contrast, an underutilised BBU will be switched off after its traffic is completely offloaded to another suitable BBU. One problem of these BBU management schemes is that they match the BBU resources to the total traffic demand of each RRH. This “BBU-to-RRH” mapping does not guarantee the QoS provisioning at user level. They investigate the mapping based on given RRH deployment, thus overlook the potential of energy saving using dynamic RRH sleeping techniques.

[PLZW15] and [PLJ⁺14] have pointed out that RRHs should be adaptive to the traffic volume and can fall into sleep mode under administration of the BBU pool to save much energy when the traffic load is low. This presents energy saving opportunities of approximately 60 percent in contrast to non-sleep mode [ABH11]. However, the amount of research works for RRH switch off strategies has not been large.

Power minimisation beamforming design for C-RAN has been addressed previously in the literature [SZL14][LZL15]. In [SZL14], the joint design on selection of active RRHs and coordinated beamforming among active RRHs is done, with the objective of minimising the total power consumption of RRHs and the corresponding fronthaul links. A greedy algorithm and two iterative algorithms based on a three-stage group sparse beamforming framework are developed. The defined problem in [LZL15] has similar formulation except that both downlink and uplink transmissions are considered jointly. A virtual downlink transmission is established to convert the original problem into an equivalent form, which can be solved by relaxed-integer programming. In both schemes, cooperative joint transmission requires significant fronthaul capacity. Both studies assume the fronthaul links are

ideal. However in practical systems the fronthaul is often capacity constrained [TZJ11], which has a significant impact on both spectral and energy efficiency in C-RANs. Therefore, the network planning of equipment deployment and radio resource allocation schemes should account for the fronthaul capacity constraints [HLD15].

2.6.1 Base Station Sleep Mode Techniques in Green Cellular Networks

Although existing literature concerning RRH operation is quite limited. Sleeping strategy which putting some BSs into sleep mode (or switch off/deactivate some BSs) plays an important role to save energy, which has been proven in conventional green cellular networks. It's an important resource management technique from the considerations of traffic dynamics and energy efficiency. Sleep mode, or sleeping technique [ABH11] is regarded as a promising solution to reduce the power consumption dramatically, which allows the hardware components of the base stations to be switched off and makes the power consumption adaptable to the actual traffic [SCSY14].

Sleep mode techniques cover approaches that selectively turn off some resources in the existing network architecture during non-peak traffic hours. These approaches generally try to save energy by monitoring the traffic load in the network and then decide whether to turn off (or switch to sleep mode, also referred as low-power mode or deep idle mode in some literature). The fact that BSs have been designed to serve peak traffic leads to wastage of energy during low traffic hours. Sleep mode operations exploit the opportunity by turning lightly loaded BSs to sleep and to save fixed part of energy consumption [WZZY15].

The network energy minimisation problem is to determine the optimal set of active BSs that are needed to support the required traffic demand. This is a challenging combinatorial optimisation problem [SKYK11]. The search space increases exponentially with the number of BSs. Let \mathcal{L} be the set of BSs located inside the considered geographical area. Theoretically, it requires high computational complexity for finding the optimal active BS set among $2^{|\mathcal{L}|}$ on/off combinations.

A practically implementable switching on/off based energy saving (SWES) algorithm is proposed in [OSK13] that can be operated in a distributed manner with low computational complexity. The key design principle of the SWES algorithm is to turn off a BS one by one that will minimally affect the network by using a newly introduced notion of network-impact, which takes into account the additional load increments brought to its neighbouring BSs. In [CD14], Genetic algorithm (GA) is used to determine the active BS set for reducing overall network energy consumption in OFDMA cellular networks with much lower complexity compared to exhaustive search, especially when the number of BSs is large. However, in both studies, the user association strategy is based on given design of BS switch off strategy. In [OSK13], when a BS is turned off, the users served by it will be handed over to the neighbouring BS which provides the best signal strength. In GA-based scheme [CD14], when evaluating each solution representing the set of active BSs, users are attached to the BS with the strongest signal strength.

2.6.2 Joint Design for BS Switch off and User Association

User association, as the word implies, means associating mobile end users with BSs in an energy efficient way. In order for the sleep mode techniques to function, users originally connecting to BSs that go asleep need to be associated with new active

BSs. This process is required to ensure the QoS does not degrade significantly during BS sleeping operations.

Joint design of BS operation and user association which couples BS switch off tightly with user association offers potential aggregated gains. It has also been noticed that simply associating a user to the closest BS may be sub-optimal when traffic distribution is inhomogeneous, because the closest BS, if it is located in a low traffic area, may be preferable to be turned off [SKYK11]. Therefore, optimal user association, based on locations of users and BSs, average or instantaneous received signal quality as well as traffic load, is an essential condition for sleep mode schemes to be advantageous [SQKS13].

Energy efficient joint design has been studied according to different metrics. When taking the maximisation on energy efficiency (EE) as the optimisation objective, the joint design is investigated in [TSR⁺15] based on providing a heuristic solution. In particular, the small base station (SBS) with the lowest EE value is the most possible candidate for being switched off; if the system EE is improved after this SBS is deactivated and its served users are re-associated, the algorithm is repeated to successively switch off the SBS with lowest EE value at each round, till there is no gain in EE anymore. As stated in its conclusion, one limitation of this proposed approach is that it does not take QoS constraints into account.

The joint design is also studied to minimise the total cost of flow-level performance (file transfer delay) and energy consumption [SKYK11]. For the user association problem, an iterative algorithm implemented by both the user and base station sides is proposed to achieve load balancing, given the problem is feasible. For the BS operation problem, they propose simple greedy-on and greedy-off algorithms. Work [XZZL15] is an extension of [SKYK11], where both on-grid power

and renewable energy are considered.

In [RT14], the joint optimisation problem is formulated to minimise the total energy consumption while maintaining the QoS of users. The strategy is to put high power macro base stations into sleep mode and offload the users to low-power small base stations or neighbouring macro base stations. A many-to-one matching based algorithm is used to solve user association. A voting-based macro base station sleep mechanism is proposed.

2.6.3 Discussion

By identifying the energy saving potentials using sleeping techniques, we make our own contribution to address the network energy minimisation problem in C-RAN architecture. The centralisation in BBU pool facilitates the implementation of RRH selection to be adaptive to traffic dynamics. Different from previous works reviewed above, we jointly consider user association, RRH selection, radio resource constraints and fronthaul capacity constraints. Rather than putting high power node into sleep mode, our strategy is to put pico RRH into sleep mode since high power node in C-RAN needs to be active all the time for coverage in the control plane. This part of work is in Chapter 4.

2.7 Summary

This chapter provides an overview of radio resource management for wireless communications and introduces the basics of quality of service and the concept of cloud radio access networks. Existing resource allocation approaches in OFDMA

networks are summarised and discussed. Also, recent research interests in energy efficient RRM schemes, especially sleeping techniques are reviewed and discussed.

Chapter 3

Joint Antenna and Subcarrier Management Scheme

3.1 Introduction

This chapter focuses on investigating QoS-aware user satisfaction maximisation in resource-constrained distributed 4G system with unbalanced traffic. As discussed in Section 2.3.5, instead of rate maximisation resource allocation algorithms, it is imperative to devise satisfaction maximisation algorithms for the ever-increasing multimedia traffic in a resource-constrained system, as well as to support load balancing. From this consideration, a joint antenna and subcarrier management scheme is proposed for the multi-cell OFDMA downlink. We utilise a semi-smart antenna approach to achieve cell coverage adaptation responding to the unbalanced load distribution. The research objectives and contributions of this chapter can be summarised in Table 3-A.

Specifically, in Section 3.2, general system model is introduced. Then a joint antenna and subcarrier optimisation algorithm is proposed to improve the network satisfaction performance in Section 3.3. Simulation set-up and results are presented

Table 3-A: Research objectives and contributions

Research objectives	Research contributions
QoS provisioning improvement: user satisfaction maximisation with load balancing	A joint antenna and subcarrier management scheme is proposed. Semi-smart antennas are utilised to achieve adaptive cell coverage and optimised by genetic algorithm. The proposed scheme can lead to 10-15% satisfaction improvement compared with that in fixed antenna pattern.
Dynamic optimisation of QoS with low-overhead for moving users	We monitor the user movement periodically and optimise the antenna pattern only when the satisfaction ratio drops below a certain threshold. This method avoids unnecessary handovers. Simulations show that the proposed scheme offers about 15% satisfaction improvements in the overall user satisfaction.

in Section 3.4. Additionally by taking user priorities into account, the joint optimisation algorithm is further developed and evaluated in Section 3.5. Finally a dynamic optimisation of QoS with user movement is proposed and evaluated in Section 3.6.

3.2 System Model

An OFDMA multi-cell network serving I users with L base stations (BSs) is considered in this chapter. In this research, our focus is on downlink transmission, i.e., from BSs to users. The frequency reuse factor is 1 and each cell has N subcarriers. Each cell is divided into three sectors and each sector has one third of the total subcarriers. This thesis takes the minimum rate requirement as a QoS indicator into consideration because data transmission rate is the most important factor to determine a user's satisfaction [SAR09] and is widely adopted in the literature mentioned in Section 2.3.4. Besides, the channel models formed in this thesis is

from the perspective of physical layer, it is difficult to evaluate other QoS supporting abilities such as delay bounds and packet loss ratio [WN03]. Specifically in this thesis, every user has an independent QoS requirement denoted by R_i . In this chapter, for simplicity, we assume power allocation is uniformly distributed to each subcarrier, as in [SLFC07][WWSZ10][LCYW12] and [ZXH⁺10]. Denoting the transmit power of BS j by P_j , the BS transmit power on subcarrier n is $p_j^n = P_j/N$. We define by $\mathbf{A}_j = [a_{ij}^n]$ the allocation matrix, where the subcarrier allocation indicator a_{ij}^n is 1 if subcarrier n is assigned to user i and 0 otherwise. Within a cell, one subcarrier can be assigned to only one user. BSs use semi-smart antennas while users remain omnidirectional antennas.

3.2.1 Semi-smart Antennas

The concept of semi-smart antennas is put forward in [NPP⁺03], which has the functionality to change the radiation pattern in response to some system need. In semi-smart antennas approach, only few antenna elements are needed to create highly shaped beams with low angular accuracy and gain. The theory behind this approach is based on utilising a load balancing scheme to shape cellular coverage according to the traffic needs [DBC04][WBJW08].

Compared to the fully smart antennas, which points an individual beam towards each user, the semi-smart antennas approach produces a flexible broadly-shaped coverage pattern for an area. Both the complexity and the cost are much lower using the semi-smart antennas [ACP06][YWZC10].

In 4G systems, semi-smart antennas consists of three sectors and in each sector, a multi-element array antenna is used to cover 120 degree. The array antenna is controlled by adjusting the excitation of the amplitude and phase of individual an-

tenna elements to produce desired radiation patterns [YWZC10]. In this research, the semi-smart antennas at each BS consist of three sectors with a 2-element array antenna for each sector. Each BS's coverage pattern is controlled by the 6 individual antenna elements with each antenna element covering 60 degree. Denoting the gain of antenna element d of BS j by g_j^d , the antenna pattern of one BS is determined by $\mathbf{G}_j = [g_j^d]$.

It would be possible to use a more complex set of antennas and control the phase as well as the gain, but the approach adopted in this chapter is simple but more realistic in practice.

3.2.2 Channel Model and Achievable Rate

Let h_{ij} denote the channel gain between user i and BS j , which embodies the effects of path loss, log normal shadowing and antenna gains as large scale fading component. Since this research focuses on system capacity investigation, which is only related to the average signal condition, so small scale fading component is not considered here [Xue00] [Liu14]. Then the SINR of user i in cell j , or with respect to BS j can be expressed by

$$\gamma_{ij}^n = \frac{h_{ij} p_j^n}{\sum_{j'=1, j' \neq j}^L h_{ij'} p_{j'}^n (\sum_{i' \in \mathcal{U}_{j'}, i' \neq i} a_{i'j'}^n) + \sigma^2} \quad (3.1)$$

where \mathcal{U}_j denotes the set of users in cell j and σ^2 is the noise power. Then the achievable transmission rate of user i on subcarrier n is given according to the Shannon capacity formula [KdVYV12]

$$r_{ij}^n = \Delta B \log_2(1 + \gamma_{ij}^n) \quad (3.2)$$

where ΔB is the bandwidth/spacing of the subcarrier. The received/achievable total data rate of user i , r_i is determined by

$$r_i = \sum_{n=1}^N a_{ij}^n r_{ij}^n \quad (3.3)$$

3.3 Proposed Joint Optimisation Algorithm

In this research, we give emphasis to the guaranteed bit rate (GBR) traffic whose satisfaction is a step function with the received data rate, as mentioned in Section 2.2.4. Rather than assuming that users wish to increase their data rates whenever possible, we define that user i is called satisfied user if its achievable data rate r_i is above or equal to the predefined QoS requirement R_i . The satisfied users have no inclination to change their subcarrier allocation.

The objective of the joint antenna and subcarrier management scheme is to achieve the maximum number of satisfied users:

$$\max_{\mathbf{A}_j, \mathbf{G}_j, \forall j} \left\{ \sum_{i=1}^I u(r_i - R_i) \mid \sum_{i \in \mathcal{U}_j} a_{ij}^n \leq 1 \right\} \quad (3.4)$$

where $u(\cdot)$ is a step function and the constraint states that in a cell, an occupied subcarrier can only be allocated to one user as this is the way OFDMA operates.

Problem 3.4 involves the joint optimisation of two variables, \mathbf{A}_j for the subcarrier allocation and \mathbf{G}_j for the antenna pattern adaptation. We first solve the subcarrier allocation for given \mathbf{G}_j based on non-cooperative game. Then we optimise the antenna pattern using genetic algorithm.

3.3.1 Subcarrier Allocation

3.3.1.1 Fundamentals on Non-cooperative Game

Non-cooperative game theory is used in subcarrier allocation. As a BS is fully in charge of subcarrier allocation independently in a multi-cell OFDMA network, it can be regarded as a selfish and independent player trying to gain more profit for itself while competing with other BSs that have the same non-cooperative behaviour. For distributed operation, the BS in each cell should be capable of operating without the information about other cells. Also it avoids massive signalling between BSs to reduce both power and time consumption. In this situation, non-cooperative game theory can render a useful and powerful tool for efficient resource management [KL06a].

A non-cooperative game has a strategic form, including player, action space and utility function. A non-cooperative game involves a finite set of players. The action space contains all the players' strategies against the others. The utility function measures the payoff of the player determined by the strategies chosen by all the players. The Nash equilibrium is regarded as the solution of a non-cooperative game. A Nash equilibrium consists of each player's best response against all others' strategies. In other words, it is a steady-state point that none of the players has incentives to change its strategy since none of them can unilaterally increase its utility function given that the other players stick to their current strategies [LW10].

In this chapter, each BS is a player. The strategy is its subcarrier allocation scheme \mathbf{A}_j . \mathcal{U}_j is the user set of player j . It is determined by the given antenna

pattern \mathbf{G}_j . The utility of a player is its satisfied user number:

$$S_j = \sum_{i \in \mathcal{U}_j} u(r_i - R_i) \quad (3.5)$$

3.3.1.2 Subcarrier Allocation Algorithm

We adopt the subcarrier allocation algorithm proposed in [LCYW12] but modify it to downlink transmission.

The subcarrier algorithm has three basic modules: the initial subcarrier allocation; the release of subcarriers occupied by unsatisfied users and the reallocation of those released resources.

A. Initial subcarrier allocation

In the initial round of subcarrier allocation, the three sectors in each cell start allocating subcarriers to their users simultaneously. For any user, a certain number of subcarriers must be allocated to it in order to meet the QoS requirement R_i . However, since there is no interference information available beforehand, the number of subcarriers for a user, denoted by num_{ij} , can be calculated based on $\lceil R_i/r_{ij}^n \rceil$ but without the interference term. $\lceil x \rceil$ means taking the smallest integer greater than x . To predict the effect of inter-cell interference and to allow more users to get their required bit rate, a guard parameter C_1 is used. Therefore, $num_{ij} = \lceil R_i/r_{ij}^n(1 + C_1) \rceil$.

B. Subcarrier release

After one round of subcarrier allocation of all cells, the r_i of every user is calculated and then compared with R_i . The subcarriers occupied by the unsatisfied users,

defined as the users whose r_i is smaller than R_i will be released and put back to the vacant subcarrier pool because those users occupy resources but do not actually have sufficient resources to be able to meet their minimum QoS requirement. Rather than giving some users this resource that does not meet their requirement, we take the resource back and distribute it to the others who can actually meet their requirement by occupying the previously released subcarriers.

C. Subcarrier reallocation

Following the release of subcarriers, the cells start a process of reallocating the vacant subcarriers to serve the unsatisfied users. After each round each BS will have a table that saves information including the interference on each subcarrier. In one cell, the unsatisfied users is reassigned subcarriers from the vacant subcarrier pool according to the interference on the subcarrier and which sector the subcarrier belongs to. Priority is given to the subcarriers with the best SINR perceived by that user taking into account the interference and belongs to its own sector. The received bit rate will be calculated whenever a new subcarrier is assigned to the user. The assignment stops when R_i is reached. As the interference information is calculated from the last round of all-cell allocation, it may not precisely represent the interference for the allocation in the current round. Therefore, we add a correction factor CF_{ij} to the number of subcarriers allocated to one unsatisfied user according to the user's history: $CF_{ij}(t) = CF_{ij}(t-1) + 1$, where t represents the iteration/round.

D. Overall subcarrier allocation algorithm

The overall subcarrier allocation algorithm is as follows:

- i) Initial subcarrier allocation and initialise the correction factor CF_{ij} to 0.

- ii) Subcarrier release: calculate r_i and release the subcarriers allocated to unsatisfied users and then increase their CF_{ij} by 1.
- iii) Subcarrier reallocation for those users still unsatisfied.
- iv) Subcarrier release and increase their CF_{ij} by 1 for unsatisfied users (as in step ii)).
- v) Go back to step iii) and continue the process until the subcarrier allocation result converges or the maximum iteration value is reached. The condition for determining convergence is the subcarrier allocation scheme of the system remains the same for two consecutive iterations.

3.3.2 Antenna Pattern Optimisation

The optimisation of semi-smart antennas is NP-hard [YWZC10] and the optimal solution cannot be obtained easily. However, because of the recent research progress in optimisation techniques, more and more NP-hard optimisation problems can now be solved by using meta-heuristic optimisation methods, for example genetic algorithm (GA) [XWTS13]. GA is a well-known effective search technique to find a optimal or near-optimal solution [Mit98] which does not require complex mathematical functions [CD14]. GA is particularly useful on problems with little domain knowledge because it requires no priori analysis of the hypothesis space or of the problem domain [HG89][WYMC09]. GA deals simultaneously with a set of possible solutions, called population to search different regions of a search space, which allows to find a diverse set of solutions for difficult problems with non-convex and discontinuous solution spaces [YWZC10]. Its basic idea is to update the solution population based on the fitness function until getting a satisfactory

solution [ZZZX12]. Since this chapter concentrates on the use of GA, rather than on research into GA, the reader is referred to [HG89][Mit98] for more information on the GA details.

3.3.2.1 Fundamentals on Genetic Algorithm

GAs are search and optimisation procedures motivated by the principles of natural genetics and natural selection. Some fundamental ideas of genetics are borrowed to construct search algorithms that are robust and require minimal problem information.

Fig. 3.1 shows a flowchart of the GA working principle. It begins its search with an initial population (a random set of solutions). Once a population is initialised, each solution in it is evaluated in the context of the underlying objective and constraint functions. A termination criterion is then checked. If the termination criterion is not satisfied, the population is modified by three main genetic operators (reproduction operator, crossover operator and mutation operator). Hopefully a better population is created. The generation counter is incremented to indicate that one generation of GA is completed.

A. Solution representation

In GA, design parameters are coded into binary strings where each bit is a gene. A solution (chromosome) is represented by the overall binary string including all the design parameters, which can be referred to as chromosomal representation.

B. Fitness assignment

Once a string (or a solution) is created, it will be evaluated by the underlying

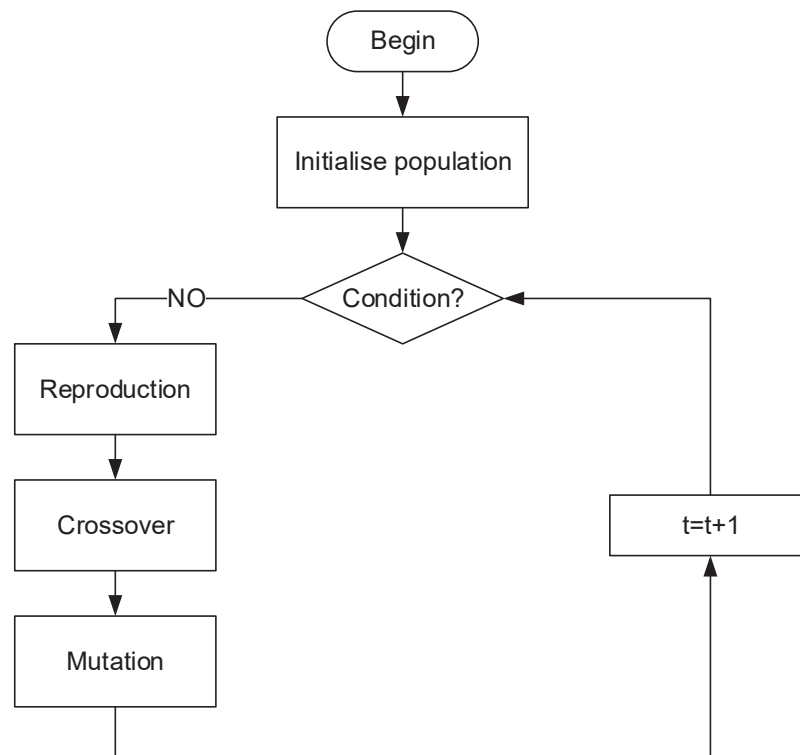


Figure 3.1: A flowchart of the working principle of a GA including population and three genetic operators

objective and constraint functions and then assigned a fitness value. In the absence of constraints, the fitness of a string will be the solution's objective function value. In most cases, however, the fitness is made equal to the objective function value.

C. Genetic operators

Genetic operators are the major part of the working of a GA. Three main genetic operators are reproduction operator, crossover operator and mutation operator. These three operators are simple and straightforward. Reproduction operator selects good strings. Crossover operator recombines good substrings from two good strings together to hopefully form a better string. Mutation operator locally alters

a string to hopefully create a better string. Even though none of these claims are guaranteed and/or tested during a GA generation, it is expected that if bad strings are created they will be eliminated by the reproduction operator in the next generation and if good strings are created, they will be emphasised.

- Reproduction operator

Reproduction operator emphasises good solutions and eliminates bad solutions in a population without changing the population size. This is achieved by: a. Identify good solutions; b. Make multiple copies of good solutions; c. Replace bad solutions with copies of good solutions.

- Crossover Operator

Crossover operator is applied to the strings of the mating pool (new population formed after reproduction operator). In crossover operators, two strings are picked from the mating pool at random and some portion of the strings are exchanged between the strings. In order to preserve some good strings, not all strings in the population are used in crossover. If a crossover probability of p_c is used then $100p_c\%$ strings in the population are used in the crossover operation and the remaining are simply copied to the new population.

- Mutation Operator

Like gene mutation, mutation operator changes a 1 to a 0 and vice versa with a small mutation probability, p_m . Mutation is used to keep diversity in the population. Mutating a string with a small probability is not a random operation since the process has a bias for creating a few solutions in the neighbourhood of the original solution.

As seen from the above description of GAs working principles, GAs work very differently compared to the traditional search and optimisation methods. The fundamental differences are: GAs work with a coding of variables instead of the variables themselves; GAs work with a population of solutions instead of a single solution; GAs do not require any auxiliary information except the objective function values and GAs use probabilistic rules to guide their search [Deb99].

3.3.2.2 Antenna Management Algorithm

The encoding scheme to represent the antenna patterns as chromosomes uses a gain vector $\mathbf{G}_j = [g_j^1, g_j^2, \dots, g_j^D]$, in which each gain value is coded as a gene, representing antenna gains along D directions. We adopt the widely-used binary coding method [Cui08] where each antenna gain g_j^d is encoded to a binary string. \mathbf{G}_j determines one antenna coverage pattern. The multi-cell system has L BSs, so that a chromosome is represented as $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_L]$. With binary encoding, a solution/chromosome is the overall binary string including all antenna gains, represented as the binary equivalence of \mathbf{G} . Based on the precision, assuming one gain value is coded into n_g -bit string, then one chromosome includes DLn_g bits in total.

Each chromosome is associated with a fitness function to evaluate the goodness of its indicated solution. The fitness function here is to maximise the user satisfaction, as expressed in equation (3.4). Every chromosome is fed into the subcarrier allocation algorithm described in Section 3.3.1.2 as the input determining the user set \mathcal{U}_j of every player BS j .

We let the chromosomes go to the next generation through a breeding process consisting of three operations, namely reproduction, crossover, and mutation.

In reproduction operation, the chromosome selection rule is based on a roulette wheel selection, such that the higher of the fitness, the greater opportunity of the chromosome to be selected [CSP11]. The possibility of chromosome C_p to be selected is $F(C_p)/(\sum_{q=1}^{N_{re}} F(C_q))$, where $F(C_p)$ is the fitness value of chromosome C_p . Note that the selected chromosomes are not removed from the population; therefore, it is possible that the same chromosome is selected more than once.

We use the most widely used single-point crossover method [Kay11] to combine the genes of the two selected individuals. A crossover point is first chosen at random somewhere along the length of the string of the chromosome. Then, across the string of the chromosome, any codes after the crossover point are swapped between the parent chromosomes to form two new children. An illustration of the single-point crossover operator is given in Fig. 3.2. By inheriting partial characteristics from the parents in this way, the chromosomes of their offspring are expected to provide better solutions.

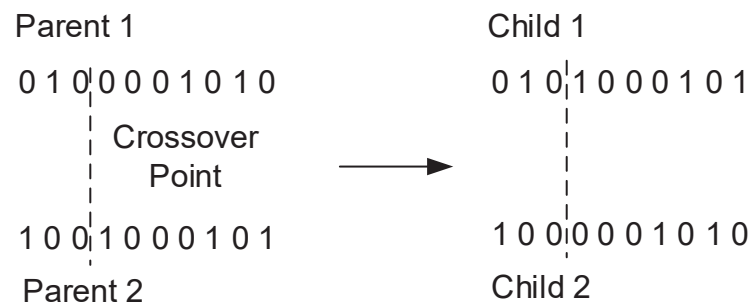


Figure 3.2: An illustration of the single-point crossover operator. Two parent solutions to create two new children solutions

After crossover, all children chromosomes go through a mutation operation. For every gene element in the chromosome, a random number $r \in [0, 1]$, is generated and compared with the mutation probability, p_m . If $r < p_m$, the code will be

replaced as the other possible value; otherwise, it will remain unchanged. In our case, if mutation happens, gene 0 becomes 1 and vice versa, illustrate in Fig. 3.3.

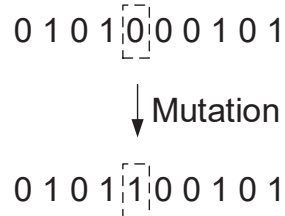


Figure 3.3: An illustration of the mutation operation. The fifth bit is mutated to create a new string

To prevent the best solution of the prior generation from being lost in the breeding process, we use elitism strategy taking the two best individuals of each parent generation to the next generation. All other parents will be replaced by the offspring generation. As we can see, two selected parents can produce two children; therefore, to keep the size of the population unchanged, the breeding process has to be repeated until $N_{po} - 2$ children are generated. Once a new population of N_{po} chromosomes has been formed, it replaces the old generation. The procedure is then repeated for a total of N_{ge} generations. When the algorithm is terminated, the antenna pattern solution is based on the best individual of the current population. The flowchart for the antenna management algorithm is shown in Fig. 3.4.

3.4 Simulation Set-up and Results

To evaluate the performance of proposed algorithm, a multi-cell OFDMA network is simulated. Fig. 3.5 illustrates the flowchart of the simulation platform used in this chapter. The functionality of each module is summarised as follows.

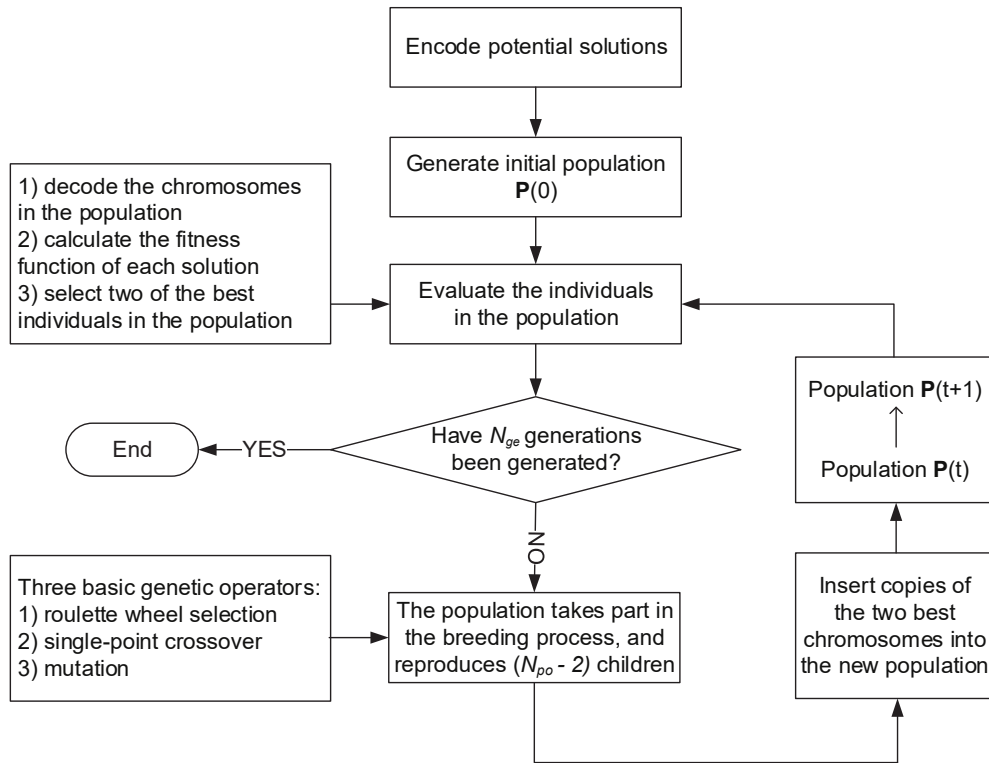


Figure 3.4: Flowchart of the antenna management algorithm

A. Network initialisation

This module generates network topology (e.g., cell and sector); generates the positions of BSs and users, where the BS is located at the centre of every cell and the users are distributed in uniform or unbalanced mode; marks the priority and QoS requirement of users; does wrap-around matching up virtual cells with real cells.

B. Joint optimisation

This module implements the proposed joint optimisation algorithm. In every generation, firstly the channel model considering the semi-smart antenna gain is updated. Also, the user set \mathcal{U}_j for each BS is updated accordingly. Then we implement the subcarrier allocation algorithm in Section 3.3.1.2. For every individual

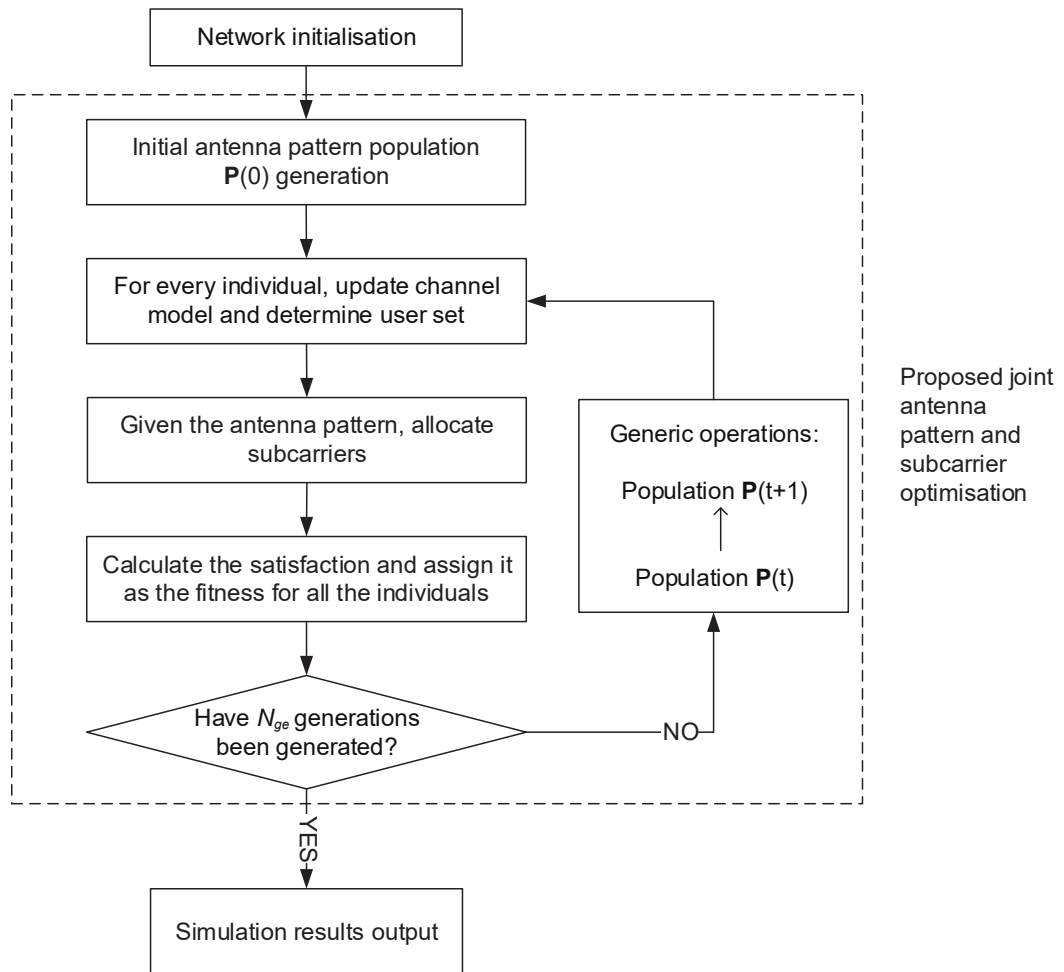


Figure 3.5: Flowchart of simulation platform for joint antenna and sub-carrier optimisation

representing one solution of overall antenna pattern for the seven-cell network, the fitness is evaluated based on the user satisfaction calculated by $\sum_{j=1}^L S_j$, where S_j is given in equation (3.5). At last, the population will evolve by the three main genetic operators until termination.

Table 3-B: System parameters

Parameter	Value
Network topology	7 hexagonal cells with wrap-around
Cell radius	500 m
Carrier frequency	2.0 GHz
Bandwidth	5 MHz for downlink
Subcarrier bandwidth	15 kHz
Downlink subcarrier number	300
Antenna type	3 sectors semi-smart antennas, with 6 controllable elements (0-20 dBi)
BS transmit power	43 dBm
Path loss model (dB)	$128.1 + 37.6 \log_{10}(d)$ (d in km)
Shadowing standard deviation	10 dB
Noise density	-174 dBm/Hz
User minimum rate requirement	512 kbps [CJJ09]

3.4.1 Simulation Parameters

The system parameters and the parameters for the GA simulation are shown in Table 3-B and Table 3-C respectively. The values chosen in Table 3-B follow the 3GPP standards [Acc16a]. The antenna type parameter is according to the latest research works on semi-smart antennas [YWZC10][ZZZX12]. The parameter settings for GA follow the latest literature applying GA in OFDMA resource management [XWTS13][TCYR13][XLmT14].

Table 3-C: GA parameters

Parameter	Value
Population size N_{po}	50
Generation number N_{ge}	100
Crossover probability p_c	0.8
Mutation probability p_m	0.05

3.4.2 Validation of Subcarrier Allocation

Fig. 3.6 shows how user satisfaction varies with increasing user density. The satisfaction represented by the circle marked blue curve is the ratio of total number of satisfied users (star marked green curve) over the total number of users across the entire seven-cell network. Users are uniformly distributed here and the antenna gains are fixed at 12 dBi as in [LCYW12]. All the results are averaged over 100 drops of simulation. The green curve agrees the same trend of number of satisfied users variation as Figure 7 in [LCYW12]: when the system is light loaded, it can ensure 100% satisfaction; as the user density increases, the subcarriers will become insufficient to serve all users.

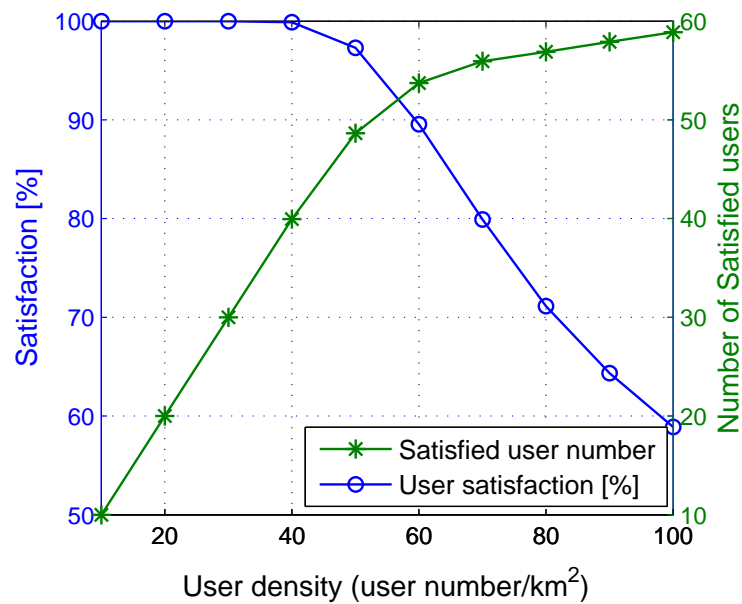


Figure 3.6: Satisfaction variation against user density

According to the non-cooperative game theory, the sign of convergence is that all players keep the current strategy. In this case, when all the BSs maintain

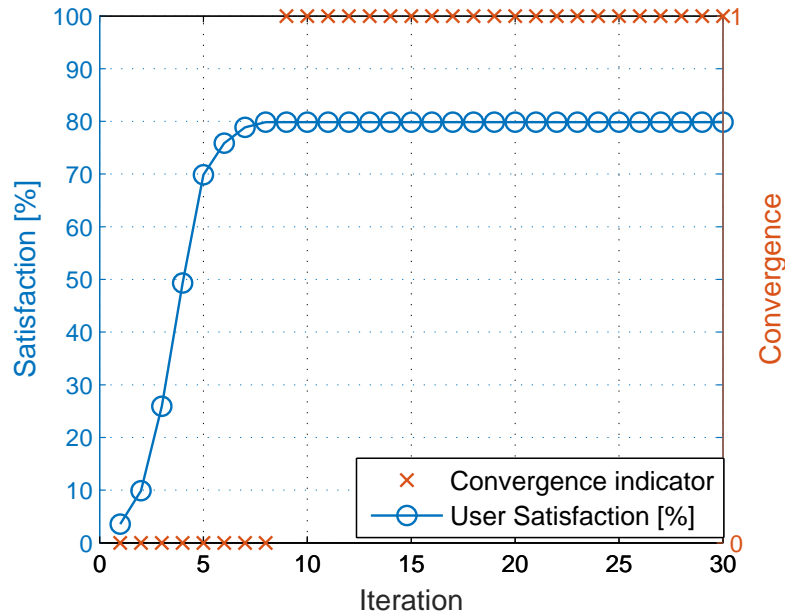


Figure 3.7: Satisfaction and convergence against iteration

their current subcarrier allocation schemes for the following iteration it means the algorithm has converged. The scheme maintained is the final subcarrier allocation to be used in the transmission. Fig. 3.7 shows the satisfaction and convergence against the iteration in one drop of simulation with 70 users/km². It indicates the satisfaction increases in early iterations and keeps steady as reaching subcarrier allocation convergence in the 9th iteration. This is in-line with the Figure 9's result in [LCYW12].

3.4.3 Verification of Antenna Pattern Adaptation

Drop the users only at one boundary of Cell 1, shown in Fig. 3.8(a). The pink dots represent satisfied user while the red dots represent unsatisfied user. According to Table 3-C, after 100 generations evolution, the GA adapts the semi-smart antenna pattern like that in Fig. 3.8(b). In this scenario, GA learns the antenna gain set

(chromosome) which has the biggest g_1^1 value to maximise the received power in order to increase the number of satisfied user. The impact of generation number will be investigated later.

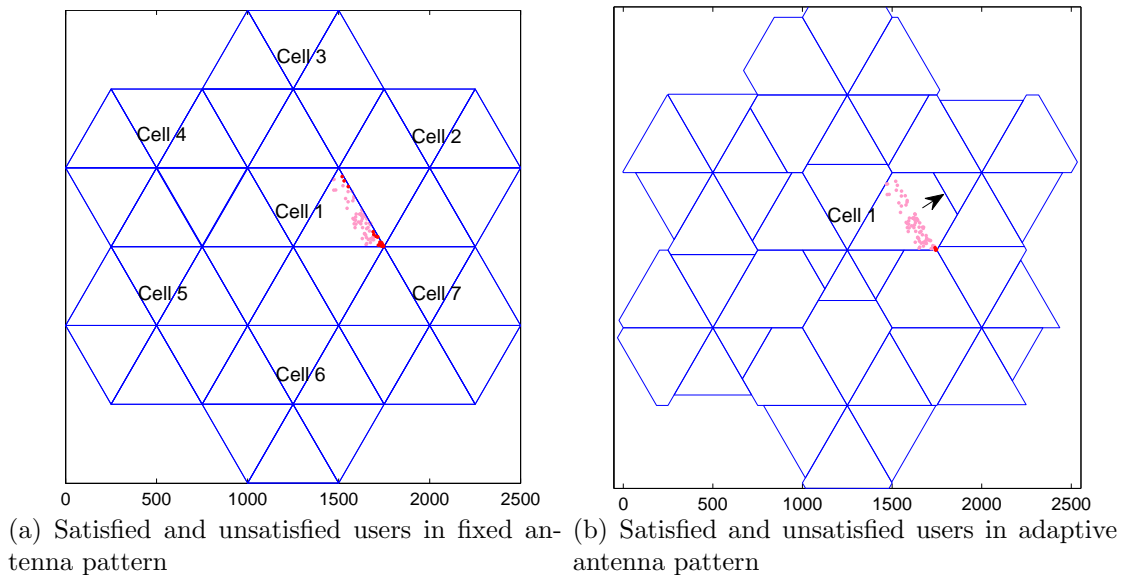


Figure 3.8: Users only at one boundary of Cell 1

Drop users at one boundary of Cell 1 and the facing antenna element coverage of Cell 2, shown in Fig. 3.9(a). After 100 generations evolution, the GA learns the boundary like that in Fig. 3.9(b). In this scenario, GA learns the chromosome which makes all the users served by Cell 2 to eliminate interference. From the comparison we can see, there are more satisfied users in Fig. 3.9(b).

3.4.4 Performance Evaluation

The overall approach of performance evaluation is to compare the user satisfaction over the entire network with and without the antenna patterns being optimised. Fig. 3.10 illustrates the average 10% satisfaction improvement when using the

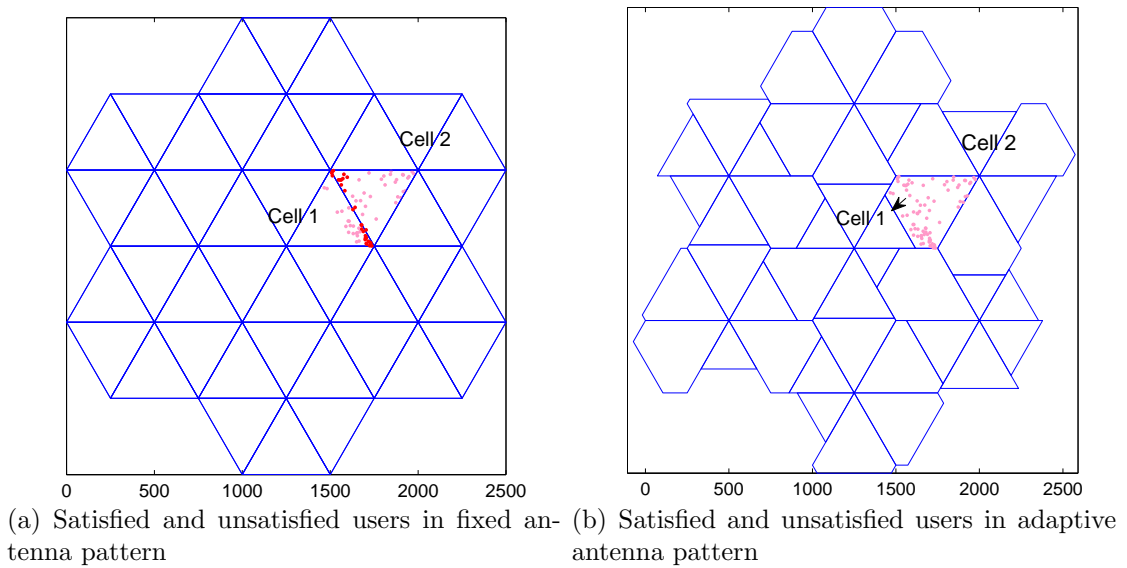


Figure 3.9: Users at one boundary of Cell 1 and the facing antenna element coverage of Cell 2

GA based adaptive pattern compared with a uniform fixed pattern. The user number within the seven-cell network varies from 375 to 675. Instead of uniform user distribution, unbalanced user distribution between the facing antenna element coverages is used in this chapter. Satisfaction means the ratio of total number of satisfied users over the total number of users over the entire network. The results shown in Fig. 3.10 give the average, maximum and minimum values from 10 drops of simulation.

When the number of user is 600, Table 3-D compares the satisfied user number in fixed antenna pattern and adaptive antenna pattern in the 10 drops and summarises that adaptive antenna gives approximate 10% satisfaction improvement.

One example of the GA evolution on maximising the satisfaction ratio over 100 generations is shown in Fig. 3.11. It can be seen that the GA functions properly to increase the fitness value as the number of generation increases. It shows similar probabilistic increase of fitness value against generations with Figure

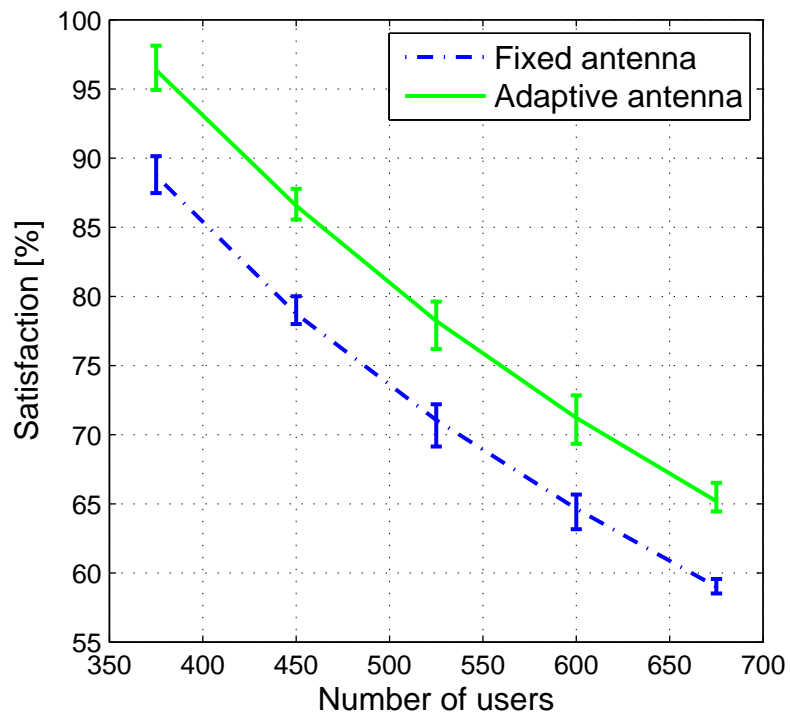


Figure 3.10: Satisfaction improvement in adaptive antenna pattern compared to fixed antenna pattern in different load networks

3(b) in [XWTS13].

Table 3-D: Satisfied user number comparison with 600 users

Drop	Fixed antenna	Adaptive antenna	Number increase	Satisfaction improvement
1	387	430	43	11.11%
2	379	416	37	9.76%
3	394	435	41	10.41%
4	389	428	39	10.03%
5	389	426	37	9.51%
6	385	417	32	8.31%
7	392	437	45	11.48%
8	390	432	42	10.77%
9	385	427	42	10.91%
10	388	425	37	9.54%

Table 3-E lists 10 drops of satisfied user number in fixed antenna and GA

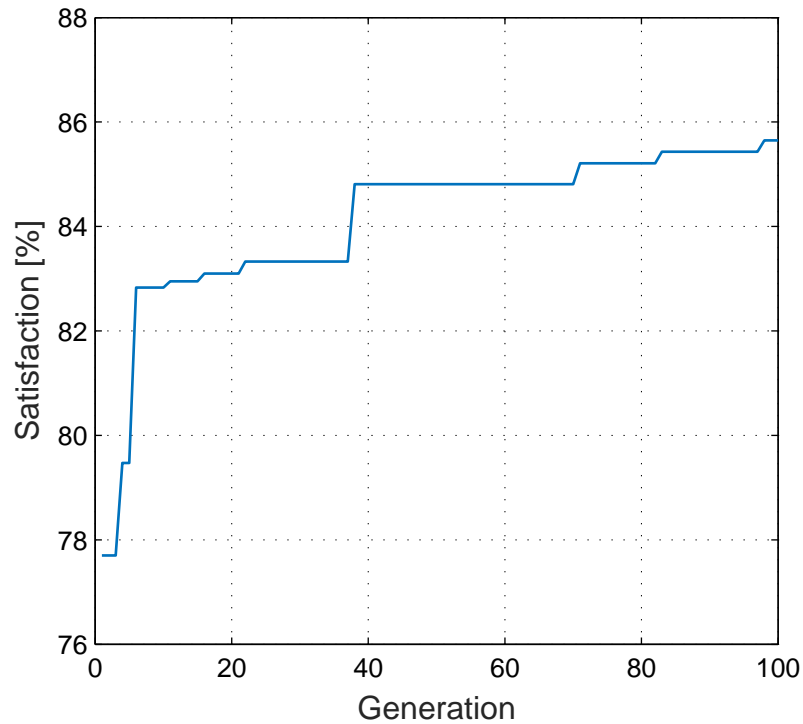


Figure 3.11: GA working on maximising satisfaction ratio

based adaptive antenna with generation number N_{ge} equals to 50, 100, 200 and 500 respectively. Based on the data in Table 3-E, Fig. 3.12 compares the average satisfaction improvement together with the normalised elapsed time in the simulations. We normalise the elapsed time of each drop simulation with 50 generations to 1. From the comparison, 200 generations and 500 generations offer slightly higher satisfaction improvement but consume considerably longer elapsed time. It's reasonable to set the generation number N_{ge} to 100.

Table 3-E: Satisfied user number comparison with different generations

Drop	Fixed antenna	$N_{ge}=50$	$N_{ge}=100$	$N_{ge}=200$	$N_{ge}=500$
1	387	425	430	430	430
2	379	411	416	416	420
3	394	427	435	436	436
4	389	424	428	428	428
5	389	420	426	426	429
6	385	416	417	419	421
7	392	433	437	437	437
8	390	430	432	432	435
9	385	421	427	427	427
10	388	416	425	425	431

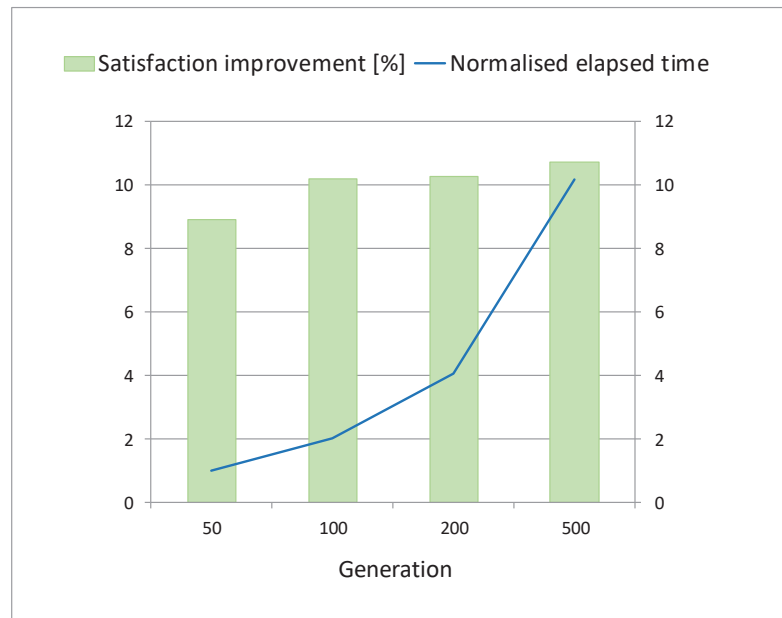


Figure 3.12: Satisfaction improvement and the normalised elapsed time comparisons in different generations

3.5 Optimising QoS-aware OFDMA Networks with User Priority

There are premium users who always want to have better user experience on their 4G LTE device. These users are willing to pay more for high bandwidth and better

network access on their devices. Not only the users but some services themselves need better priority handling in the network (e.g., VoIP call). To be able to fulfill this, QoS plays the key role. QoS defines priorities for certain users/services during the time of high congestion in the network. In this section, we differentiate the users into two classes: UC1 (user class 1)-users with high priority to access the resources; UC2 (user class 2)-users with low priority to access the resources. Resource allocation needs to guarantee, where possible, that UC1 are all satisfied.

The UC1 priority module is added between subcarrier release and subcarrier reallocation modules described in Section 3.3.1.2. For every cell, each unsatisfied UC1 user in turn chooses vacant subcarriers to achieve its minimum QoS requirement. If there are not enough vacant subcarriers to satisfy all the unsatisfied UC1 users, satisfied UC2 users are forced to release their subcarriers and put them back in the vacant subcarrier pool for the unsatisfied UC1 users to be reallocated.

3.5.1 Validation of User Priority

Fig. 3.13 shows one drop of simulation with uniform user density 75 users/km² and 20% high priority UC1 users. The square marked red curve represents the satisfaction of UC1 users, which is the ratio of satisfied UC1 user number over the total UC1 user number; the triangle marked yellow curve is the satisfaction for UC2 users accordingly. The circle marked blue curve indicates the ratio of the number of satisfied users including UC1 and UC2 over the total number of users in the system. As shown in Fig. 3.13, the subcarrier allocation algorithm converges in the 11th iteration.

Fig. 3.14 validates the module of UC1 priority. The satisfaction of UC1 and UC2 with the same experiment setting but without priority-aware is shown in

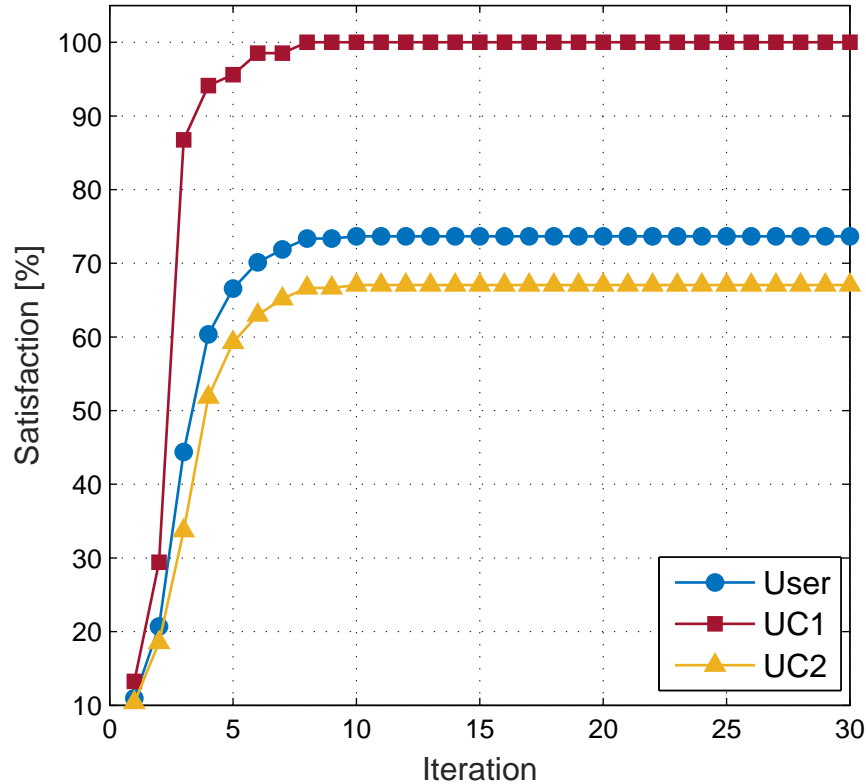


Figure 3.13: Satisfaction of UC1 users, UC2 users and all users

the dashed curves. As a comparison, priority-aware subcarrier allocation algorithm improves the satisfaction of UC1 from around 75% to 100%, but at the expense of a slightly lower proportion of UC2 users getting their required QoS, as would be expected in the algorithm design. It matches with the result of Figure 9 in [LCYW12].

Fig. 3.15 compares the satisfaction of UC1 users, UC2 users and all users with different UC1 percentage against different number of users. The solid curves show the satisfaction results with 40% high priority UC1 users while the dashed curves show those with 30% UC1 users. The results reveal that higher percentage UC1 results in lower satisfaction performance which aligns with the result in Figure 11

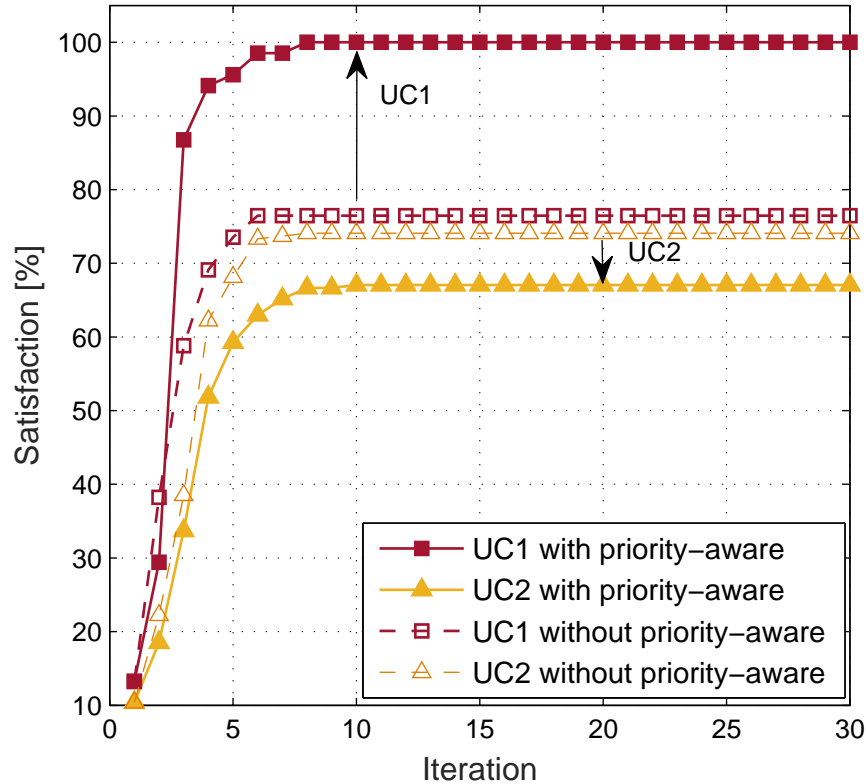


Figure 3.14: Satisfaction comparison with and without priority-aware

from [LCYW12].

3.5.2 Satisfaction Improvement with Antenna Adaptation

In this section we investigate the scenario with a fixed UC1 percentage=40%, Fig. 3.16 shows the antenna pattern optimisation with different objectives: A. Maximise the user's satisfaction; B. Maximise the UC1's satisfaction.

The black dotted curves are the satisfaction (circle for all-user, cross for UC1, square for UC2) with a fixed antenna pattern. They act as references. The green solid curves are the satisfaction results with GA optimising for all users. The blue

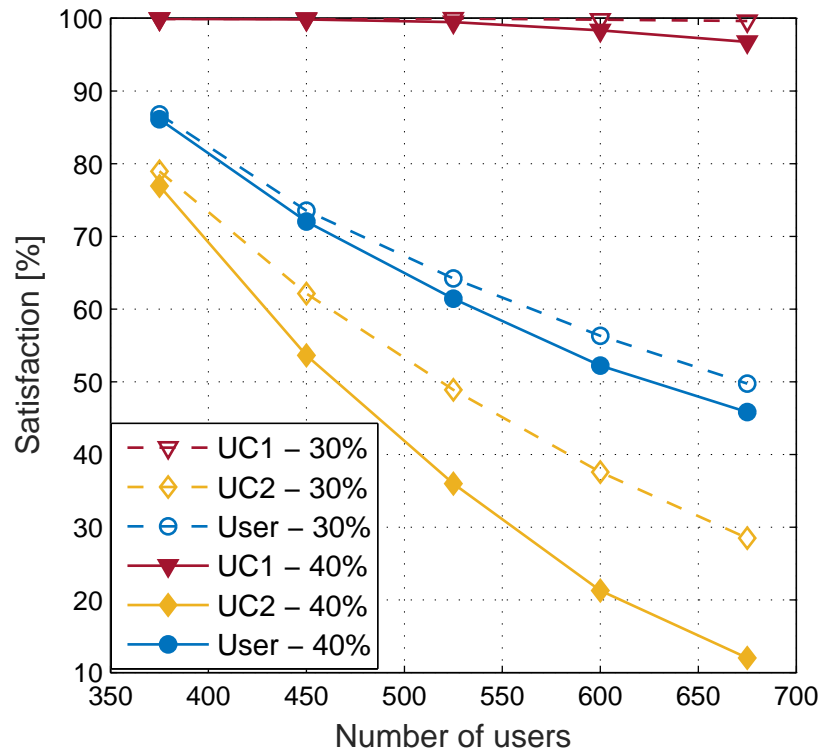


Figure 3.15: Satisfaction comparison with 30% UC1 and 40% UC1

dashed curves are the satisfaction with the GA optimising for UC1 users.

With GA optimising the all-user satisfaction, the satisfaction improves by approximately 15% in average, the UC2 satisfaction improves similarly, while the UC1 satisfaction suffers slightly drop when the network is heavily loaded and there are not sufficient resources.

With the GA optimising the UC1 satisfaction, when the system is not overloaded, the GA will ensure the UC1 satisfaction reaches 1; when the system has excessive users, the GA is able to slightly enhance UC1 satisfaction. However using GA to optimise UC1 satisfaction may lead to lower all-user and UC2 satisfactions in some drops of simulation.

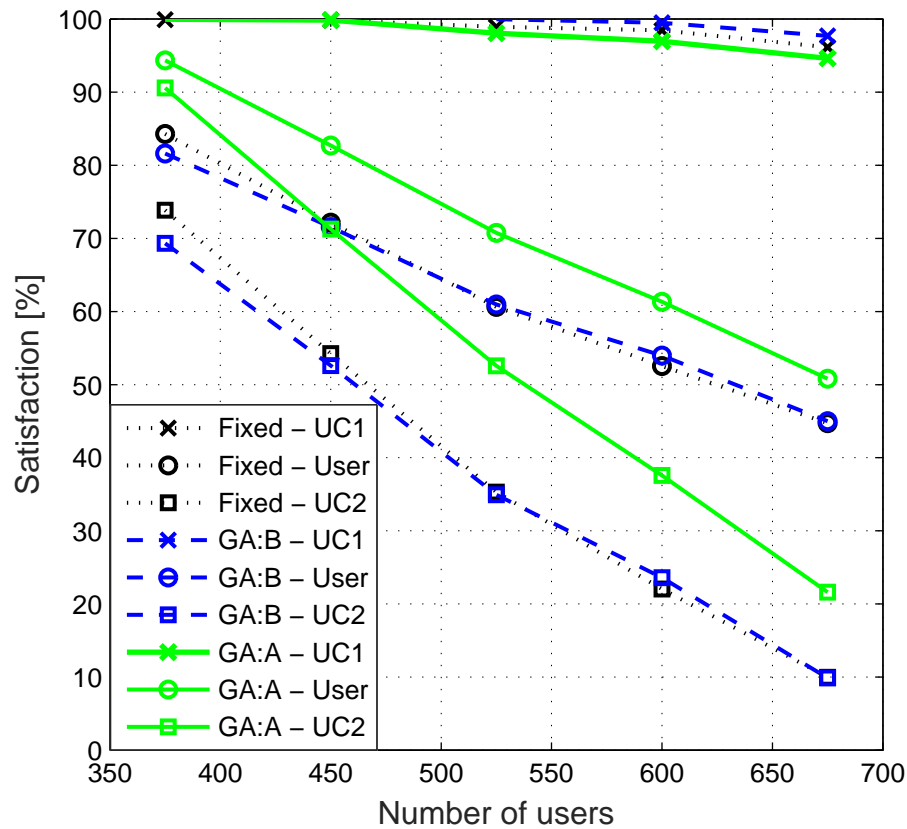


Figure 3.16: Satisfaction improvement over different objectives (optimising all user satisfaction and optimising UC1 satisfaction) with UC1 percentage=40%

Due to the trade-off between all-user satisfaction improvement and UC1 satisfaction improvement, the resource management system should select appropriate optimisation goal according to different needs.

Fig.3.17(a) and Fig. 3.17(b) show two example adaptive coverage patterns after GA evolution with 600 users. The blue dots are the UC1 users and the green dots are the UC2 users. The only different condition is that the optimisation objective of Fig. 3.17(a) is maximising the satisfaction of all users (objective A) while that of Fig. 3.17(b) is maximising the satisfaction of UC1 users (objective B). From the comparison, GA learns the cell boundaries to different objectives in different

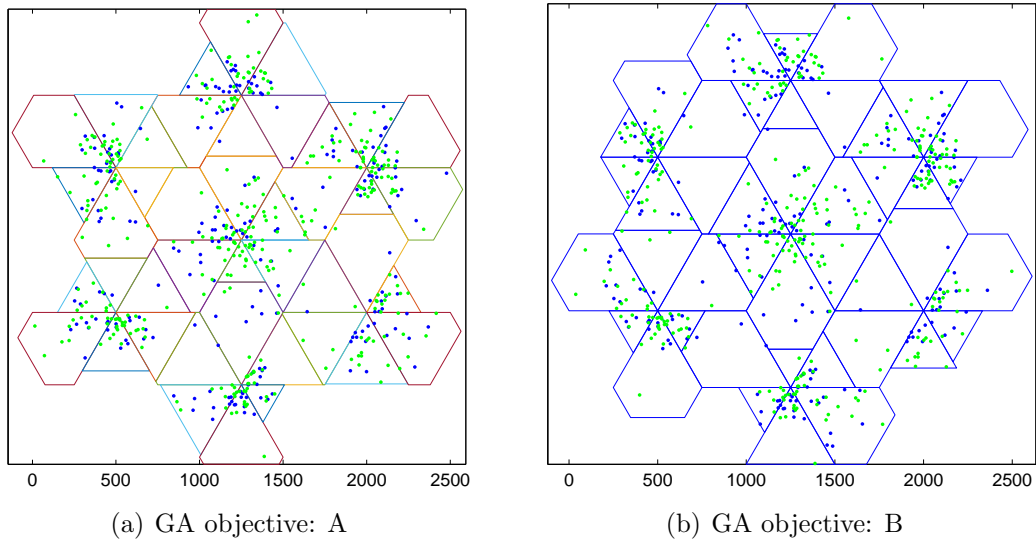


Figure 3.17: Adaptive patterns with 40% UC1 users

ways.

3.6 Antenna Pattern Update Mechanism

In a real-world communication network, it is not sensible to change the cell boundaries and the antenna coverage rapidly as this would lead to more handovers and signalling. Adding user movement is one way to test when to change antenna pattern and in this chapter a threshold-triggered optimisation scheme is proposed and is found to bring approximate 15% improvement in the user satisfaction.

3.6.1 User movement

Our approach is adding user movement into the simulation and setting some threshold to determine when to trigger the antenna pattern optimisation.

The whole threshold-triggered optimisation procedure is:

- i) Apply GA at the start point, record the optimal gain set. Set threshold as half the GA satisfaction improvement over that in fixed antenna pattern.
- ii) Keep that gain set in following time sampling points unless the actual satisfaction improvement is lower than the threshold. Trigger GA optimisation to get a new optimal gain set then calculate the new threshold.
- iii) Go back to step ii) until the end point of the simulation.

3.6.1.1 Random Moving Users

Here the movement model is set based on [TAH11][GZAM10] and the parameters are listed in Table 3-F.

Table 3-F: Random movement parameters

Parameter	Value
Vehicular speed	30 km/h
Pedestrian speed	3 km/h
Pedestrian percentage	30%
Sampling time interval	1 s [EODB13]
Initial moving direction	random
Subsequent moving direction	within 60 degree area based on the initial direction

The simulation result during 30 seconds is shown in Fig. 3.18 with a total of 500 users. The red trigger stars explain that at these time points the GA is triggered to change the antenna pattern. During the 30 seconds' simulation, the adaptive antenna gives average 12.85% satisfaction improvement.

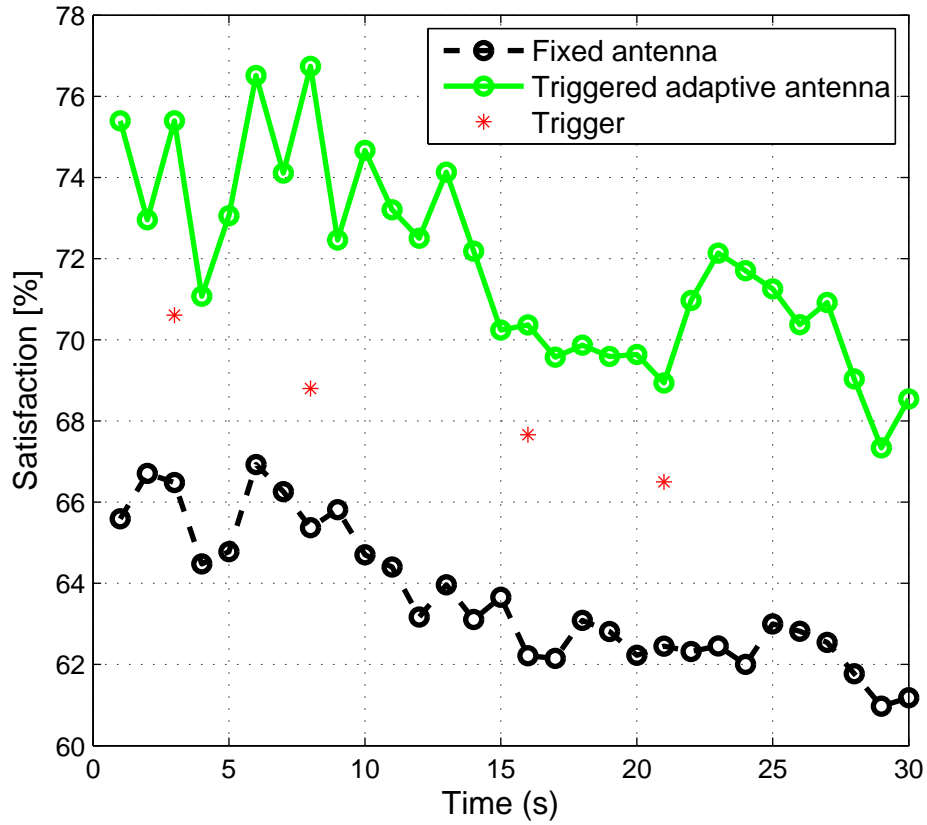


Figure 3.18: 30 seconds' user random movement

3.6.1.2 Macao Scenario

Take a realistic user movement example for visitors in Macao taking bus or taxis to their hotels. The settings in the Macao scenario are as follows in Table 3-G.

Table 3-G: Movement parameters in Macao scenario

Parameter	Value
Vehicular speed	20, 40 km/h
Sampling time interval	3 s
Moving direction	towards its hotel
Number of hotels	3
Movement stop condition	all the users are within 50 m distance to their hotels

One simulation with 485 users in system is shown here to illustrate the GA effect in this scenario. The threshold-triggered antenna pattern optimisation process is the same as that explained before. The comparison of user satisfaction achieved in a fixed antenna pattern and that applying GA optimisation when triggered is shown in Fig. 3.19.

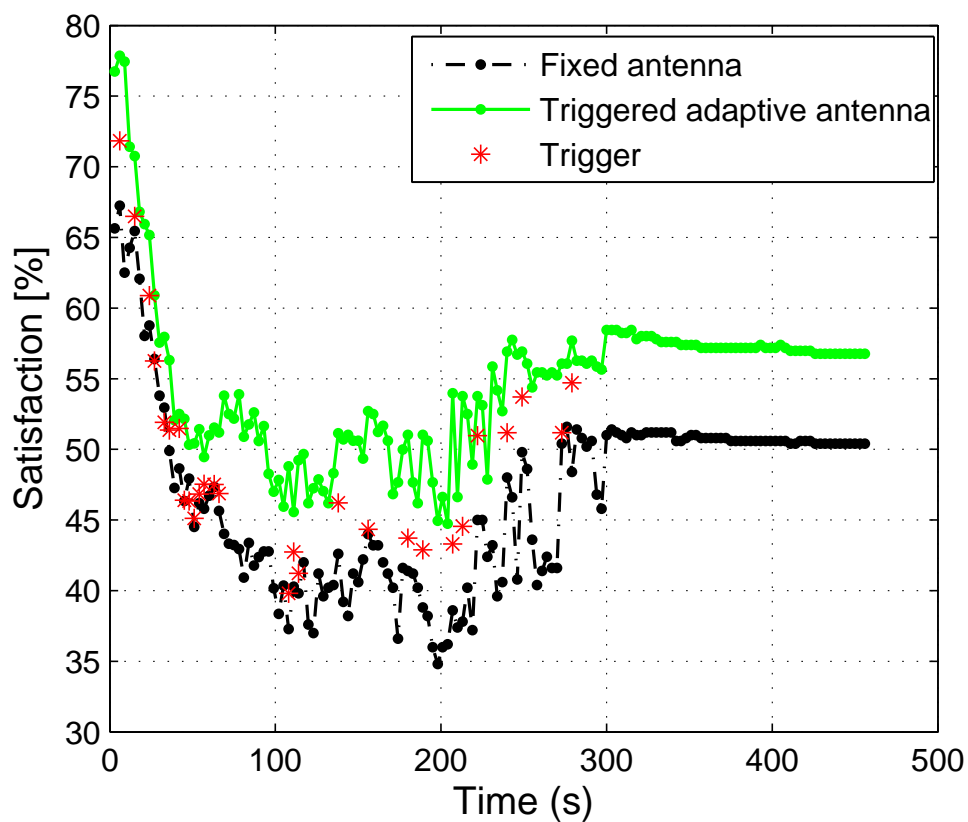


Figure 3.19: User satisfaction when having GA triggered

Movement of a cluster of users will tend to result in worse performance as users move in and out of cell boundaries. So in and out of cell edge conditions, the clustering of users leads to local shortages of subcarriers. Under these conditions, using triggered optimised semi-smart antenna patterns does improve the user satisfaction, on average 17.72% over the time scale.

It is also worth noticing that optimisation is only required intermittently so saving on computational load and continual changes in antenna patterns would lead to more handover and signalling load. In the later time intervals, triggering does not occur. This does not mean that optimisation would not improve the performance; it's just because the trigger threshold is not met. This is a limitation of the triggered approach as optimisation may lead to improvements so a combination of trigger plus timed re-runs of the optimisation may lead to better performance. This depends whether the main criterion is to maximise performance or minimise the number of pattern changes.

3.7 Summary

This chapter studies QoS-aware user satisfaction maximisation with load balancing in OFDMA downlink. Section 3.2 specifies the general system model considered in this chapter. In Section 3.3, a joint semi-smart antenna and subcarrier resource management scheme is proposed for QoS-aware multi-cell OFDMA networks. It is demonstrated in Section 3.4 that using a GA to optimise the BS's coverage pattern in order to maximise the number of satisfied users does lead to a 10% satisfaction improvement compared with that in fixed antenna pattern. With the use of a user class differentiated model in Section 3.5, the GA with the maximising all user satisfaction objective brings about an improvement of 15%. At last, Section 3.6 proposes a low-overhead dynamic QoS-aware optimisation with adding user movement. The traffic distribution is monitored periodically and the antenna pattern optimisation is only run when the satisfaction improvement in the system drops below a certain level. This method avoids unnecessary management like handovers. Simulations of two different scenarios show that the proposed scheme

offers 12.85% and 17.72% satisfaction improvements in the overall user satisfaction, without excessive triggering.

This chapter has proposed QoS-aware cell adaptation schemes in distributed BS architecture. Each BS makes its own decision on its antenna pattern and subcarrier allocation to improve the QoS satisfaction. However, incremental improvement of QoS provisioning on current mobile networks will not meet the explosive traffic demands in 5G systems. Instead of distributed manner, cloud radio access network (C-RAN) has been proposed as a promising 5G network architecture with centralisation. The next chapter will investigate resource management schemes in C-RAN.

Chapter 4

Resource Management Schemes in Cloud Radio Access Networks

4.1 Introduction

By utilising cloud based information sharing through the BBU pool, a joint resource block and power allocation scheme is proposed at the beginning of this chapter to maximise the number of satisfied users whose required QoS is provided in multi-cell OFDMA downlink served with high power nodes only.

With network densification by the deployment of low power nodes, Section 4.3 and Section 4.4 focus on optimising energy consumption of C-RAN downlink with QoS guarantee, as discussed in previous Section 2.6.3. Multiple-choice multidimensional knapsack model is used to formulate the RRH assignment problem and sleep mode techniques are used to determine the active RRH set.

To summarise, the research objectives and contributions of this chapter are listed in Table 4-A.

Table 4-A: Research objectives and contributions

Research objectives	Research contributions
QoS provisioning improvement: user satisfaction maximisation with centralised control	A joint RB and power allocation with cloud based information sharing is proposed. The proposed scheme is capable to achieve 20-30% satisfaction improvement compared to the conventional resource allocation approach.
Optimise the energy consumption with QoS guarantee	An energy-effective network deployment scheme is proposed in C-RAN based small cells, which saves 40% energy compared to genetic algorithm. A joint RRH selection and user association is proposed in heterogeneous C-RAN, which outperforms the other counterparts and provides near-optimal performance with reduced complexity.

4.2 Joint Resource Block and Power Allocation with Cloud based Information Sharing

This section proposes a QoS-aware user satisfaction oriented joint resource block (RB) and power allocation algorithm for an OFDMA multi-cell system. The objective is to optimise user satisfaction over the entire network by maximising the number of satisfied users whose QoS target is met. By taking the advantage of the cloud based information sharing via BBU pool, a two-dimensional power-RB strategy is utilised to mitigate the inter-cell interference in order to increase the user satisfaction.

4.2.1 System Model

This section considers a downlink resource allocation in an OFDMA multi-cell system with L high power nodes (HPNs) (e.g., macro or micro base stations) serving I users with variety of QoS requests. Each HPN is geographically located

at the centre of each cell. The BBU pool is interfaced with HPNs by X2/S1, whose definitions are inherited from the standardization definitions of the 3GPP [PLJ⁺14]. It is assumed that all the cells in the network share the same frequency band, which can be divided into N resource blocks (RBs). One RB comprises twelve consecutive subcarriers.

The HPN transmit power in cell j can be denoted by $\mathbf{p}_j = [p_j^1, p_j^2 \dots p_j^N]$, where p_j^n is the transmit power over the RB n in cell j . The system transmit power can be represented by a power matrix $\mathbf{P} = [\mathbf{p}_1 \dots \mathbf{p}_j \dots \mathbf{p}_L]$. For cell j , the RB assignment is denoted by a $N \times I$ boolean assignment matrix $\mathbf{a}_j = [a_{ij}^n]$, where a_{ij}^n equals 1 if RB n is assigned to user i , subject to each RB can only be assigned to one user in each cell. Similarly, the network RB assignment set can be described as $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_j \dots \mathbf{a}_L]$.

Each HPN is assumed to have perfect knowledge of the channel gain h_{ij} between user i and the HPN in cell j . Given the system transmit power matrix \mathbf{P} and RB assignment set \mathbf{A} , the SINR of user i in cell j using RB n can be expressed as

$$\gamma_{ij}^n = \frac{h_{ij} p_j^n}{\sum_{j'=1, j' \neq j}^L h_{ij'} p_{j'}^n (\sum_{i' \in \mathcal{U}_j, i' \neq i} a_{i'j'}^n) + \sigma^2} \quad (4.1)$$

where \mathcal{U}_j denotes the set of users in cell j and σ^2 is the noise power. The achievable data rate on RB n is given by Shannon capacity formula

$$r_{ij}^n = \Delta f \log_2(1 + \gamma_{ij}^n) \quad (4.2)$$

where Δf is the bandwidth of one RB. The transmission data rate of user i in cell j can be calculated by

$$r_{ij} = \sum_{n=1}^N a_{ij}^n r_{ij}^n \quad (4.3)$$

To be specific, the data rate of user i is determined by the power and RB allocation in both its own cell and the interfering cells. The inter-cell interference, however, is unknown before the resource allocation. To address this problem, cloud based information sharing of the power matrix \mathbf{P} and RB assignment set \mathbf{A} is introduced. In addition, the BBU pool is utilised to manage the co-channel RB allocation. The BBU pool is interfaced with each HPN to exchange the resource allocation and user satisfaction parameters.

4.2.2 Problem Formulation

This section takes the minimum data rate requirement as a QoS indicator into consideration. Specifically, every user has an independent QoS requirement denoted by R_i . The objective of the proposed scheme is to maximise the number of satisfied users whose required data rate is achieved:

$$\max_{\mathbf{P}, \mathbf{A}} \sum_{i=1}^I u(r_{ij} - R_i), \quad \forall j \quad s.t. \quad 1 \leq j \leq L \quad (4.4)$$

where $u(a)$ is the step function defined as:

$$u(a) = \begin{cases} 1 & a \geq 0 \\ 0 & otherwise \end{cases} \quad (4.5)$$

subject to:

$$0 \leq \sum_{n=1}^N p_j^n \leq P^{max}, \quad \forall j \quad (4.6)$$

$$a_{ij}^n = \begin{cases} 1 & \text{if RB } n \text{ is assigned to user } i \\ 0 & otherwise \end{cases} \quad (4.7)$$

$$\sum_{i \in \mathcal{U}_j} a_{ij}^n \leq 1, \quad \forall n, j \quad (4.8)$$

The objective function (4.4) formulates the joint RB and power allocation problem aimed at optimising the system in the view of the user satisfaction. Given the QoS requirement, whether a user is satisfied or not depends on its transmission data rate in equation (4.3). This is related to the power matrix \mathbf{P} and RB assignment set \mathbf{A} . It compares every individual user's actual data rate against its minimum data rate requirement. If the QoS requirement is met, this user is counted as a satisfied user who is satisfied with the provided service. Otherwise, this user is identified as an unsatisfied user who is unsatisfied with the current resource allocation, which is regarded as not good enough resulting in a degraded experience.

Constraint (4.6) states that the total transmit power of every HPN is constrained to be less than P^{max} . Equation (4.7) indicates the RB assignment matrix for each cell. What is noteworthy is the RB assignment constraint condition (4.8). On the one hand, it captures the character of an OFDMA system that within one cell each RB can be allocated to no more than one user. On the other hand, it reveals that this work chooses a resource-economical approach that does not exhaust all the RBs.

Referring to (4.1) for the inter-cell interference in particular, the interfering cell set for cell j becomes:

$$N_j^{int} = \{j' | 1 \leq j' \leq L \text{ and } j \neq j' \text{ and } \sum_{i' \in \mathcal{U}_{j'}, i' \neq i} a_{i'j'}^n = 1\} \quad (4.9)$$

It indicates that although all the cells in the network share the same bandwidth, the interfering cell set is no longer fixed $L - 1$ cells (excluding itself). Thus the co-channel assignment on every RB becomes more important.

4.2.3 The Proposed Resource Allocation Algorithm

In the resource allocation problem formulated in the previous section, each HPN uses a two-dimensional power-RB strategy to decide the power allocation over all RBs, and the RB assignment to its total serving users with the goal of maximising the satisfied user ratio in the corresponding cell. The challenging part is that each HPN is only responsible for the resource allocation within its own cell. However, the SINR in (4.1) also depends on the power and RB allocation in interfering cells, which remains unknown before the resource allocation. Therefore, the mutual dependency of SINR complicates the problem. The proposed QoS-aware user satisfaction oriented joint resource block and power allocation algorithm with cloud based information sharing in this section is designed to solve the problem.

The users in the network are divided into two sets: a satisfied user (SU) set and an unsatisfied user (USU) set. The definitions of SU and USU can be found in the “Problem Formulation” section. The main policies involved in the proposed allocation algorithm are summarised as below:

Policy 1: for the SU set, the HPN remains the current RB assignment and conducts the power control in order to reduce inter-cell interference caused by SU set.

Policy 2: for the USU set, the HPN will allocate the available RBs to them in the expectation that the new RB may have less inter-cell interference on it for a particular USU. This policy helps USU obtain RB resource in better condition and then become SU in an opportunistic way. However, this policy brings the risk that incurs new inter-cell interference to some existing SU and results in turning it unsatisfied.

Policy 3: to solve the dilemma in Policy 2, information concerning the current power matrix \mathbf{P} , RB assignment set \mathbf{A} , SU set and USU set is shared over the entire network through the BBU pool. However, each HPN still makes its own decision about the next resource allocation. In addition, the BBU pool is used to disable specific RB reassignment to USU if this would cause other SU to become unsatisfied.

The detailed procedure about the proposed algorithm is described by following steps:

A. Initial resource allocation

To every serving user within the cell, each HPN estimates the data rate that can be achieved per RB, employing a simplified form of (4.1) with interference omitted since no interference is available at the initial stage. According to individual user's minimum data rate requirement, R_i , the HPN determines the number of RBs assigned to it. This number, denoted by num_{ij} is assumed to be the ideal RB number one user needs to achieve its QoS requirement and will not change in the following steps. Thus the average data requirement can be calculated by R_i/num_{ij} . By using (4.2) and (4.1), an interference threshold Th_{ij}^n can be calculated indicating the maximum inter-cell interference that user i in cell j can tolerate to maintain it satisfied. At this stage, HPN equally distributes the total power over all RBs:

$$p_j^n = P^{max}/N, \forall j \quad (4.10)$$

B. Unsatisfying RB release

After the initial allocation, every user will calculate its actual data rate using (4.3) with interference considered and compare it with its QoS requirement then

categorized as SU or USU.

Then for every USU, the actual interference is compared with Th_{ij}^n . If the interference value is higher than the threshold, this RB is regarded as unsatisfying RB and is released immediately.

C. SU-power strategy

For the SU set, denoted by $\{SU\}$:

$$a_{ij}^n(t) = a_{ij}^n(t-1), \forall i \in \{SU\} \quad (4.11)$$

$$p_{ij}^n(t) = \begin{cases} p_{ij}^n(t-1) - p_{step} & \text{if } p_{ij}^n(t-1) > p_{min} \\ p_{ij}^n(t-1) & \text{otherwise} \end{cases} \quad (4.12)$$

where $p_{ij}^n(t)$ is the transmit power over the RB n in cell j assigned to user i in the current iteration t .

Equation (4.11) indicates that for every SU, the HPN assumes the RB assigned to it in the last allocation iteration is good enough to maintain it as SU and will not change it in the current iteration t . Equation (4.12) reveals that the HPN will reduce the power on all the RBs occupied by SU by the same level p_{step} if the previous value is higher than the SU protection threshold p_{min} , which is defined as the minimum signal power that can maintain this user as satisfied using the previous iteration's interference value.

D. USU-RB strategy

For the USU set, each HPN will first construct an RB reassignment pool consisting of the released RBs and idle RBs from last iteration; this is also the complement of the RB set occupied by SU. Assuming in cell j , the RB set occupied by SU is

RB_j^{SU} , then the pool can be denoted by:

$$RB_j^{pool} = \{n | 1 \leq n \leq N \text{ and } n \notin RB_j^{SU}\} \quad (4.13)$$

The power distribution over the RB pool set remains the same, namely:

$$p_{ij}^n(t) = p_{ij}^n(t-1), \forall n \in RB_j^{pool} \quad (4.14)$$

The basic idea of USU RB reallocation is that: for user i in cell j , the HPN randomly chooses some RBs from RB_j^{pool} which adds up num_{ij} and expects that this reassignment will have less inter-cell interference to make user i satisfied.

However, an undesirable situation for the SU could arise due to USU RB reallocation (Policy 2). To protect SU and prevent the undesirable situation arising, the approach employed is to utilise the SU information over the entire network in the previous iteration to restrict the random RB reallocation for USU from RB_j^{pool} . Thus, for USU i , the RB reallocation procedure is shown in the following flowchart:

E. Overall resource allocation algorithm

The overall resource allocation algorithm is as follows:

- i) Initial resource allocation and interference threshold calculation.
- ii) SU/USU categorization and unsatisfying RB release.
- iii) SU resource reallocation-power strategy.
- iv) USU resource reallocation-RB strategy using cloud based information sharing.

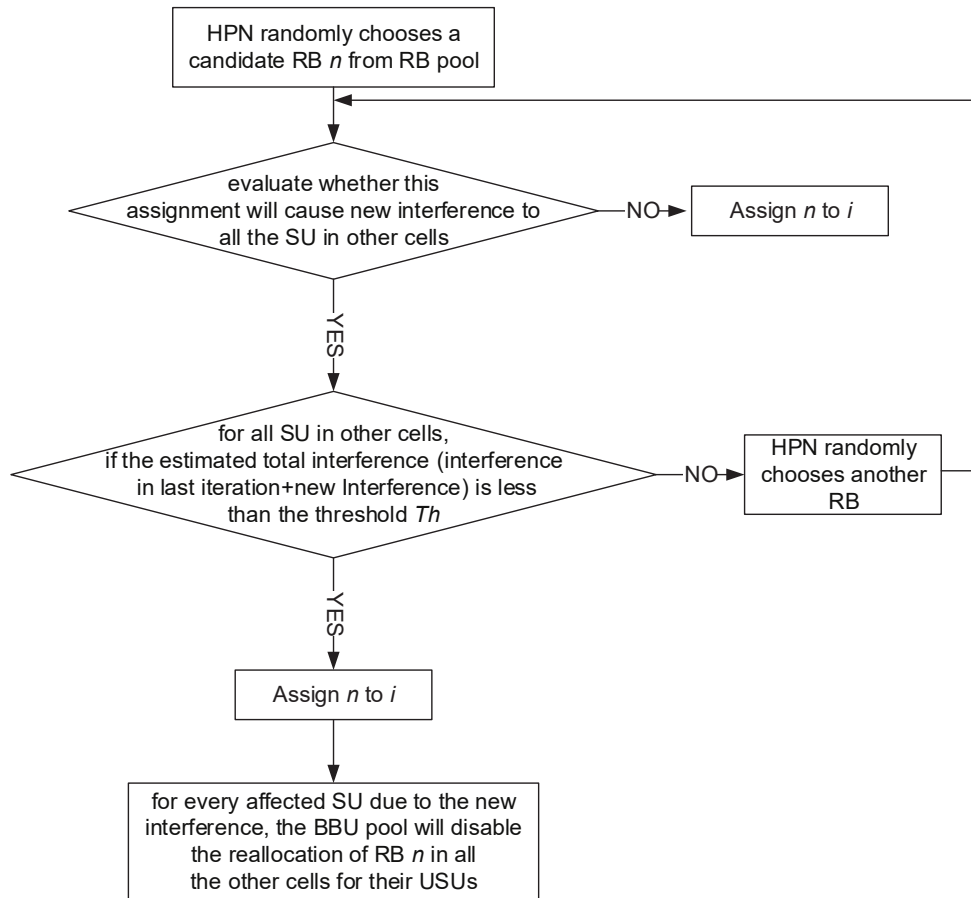


Figure 4.1: Flowchart for USU RB reallocation

- v) Update interference threshold information and go back to step ii) and then continue the process until the allocation ends.

4.2.4 Simulation Set-up and Results

The algorithm is evaluated using a system level simulation and the main parameters are shown in Table 4-B. The values chosen follow the 3GPP standards [Acc16a]. The wrap-around diagram is illustrated in Fig. 4.2. The classic resource allocation

Table 4-B: System parameters for simulation

Parameter	Value
Network topology	7 cells with wrap-around
Cell radius	500 m
Carrier frequency	2.0 GHz
Downlink bandwidth	5 MHz
RB bandwidth	180 kHz
Downlink RB number	25
Maximum HPN transmit power	43 dBm
User rate requirements	User rate-class 1(UR1): 256 kbps User rate-class 2(UR2): 512 kbps User rate-class 3(UR3): 1 Mbps
Path loss model (dB)	$128.1 + 37.6 \log_{10}(d)$ (d in km)
Shadowing standard deviation	10 dB
Noise density	-174 dBm/Hz

algorithm MaxSINR is used as the baseline for comparison purposes [XWTS13]. The objective of MaxSINR is to maximise system throughput. The optimal solution can be obtained by using a greedy approach. Specifically, each RB is assigned to the user who tends to yield the maximum SINR. In our implementation of MaxSINR algorithm, system throughput with increasing number of users is validated against the result of Figure 3 in [XWTS13].

For each load presented in following results, the result is averaged over 500 drops of simulation. In every drop of simulation, users are uniformly and randomly distributed over the seven-cell network. The number of users in each cell does not have to be the same. Moreover, the major performance indicators: satisfaction and power consumption are measured at the system level over the entire network.

How the satisfaction increases and system total transmit power consumption decreases over allocation iterations are illustrated in Fig. 4.3 with the number of users varying from 40 to 160. Satisfaction is measured by the total number of satisfied user over the total number of users in the entire network. Generally, the

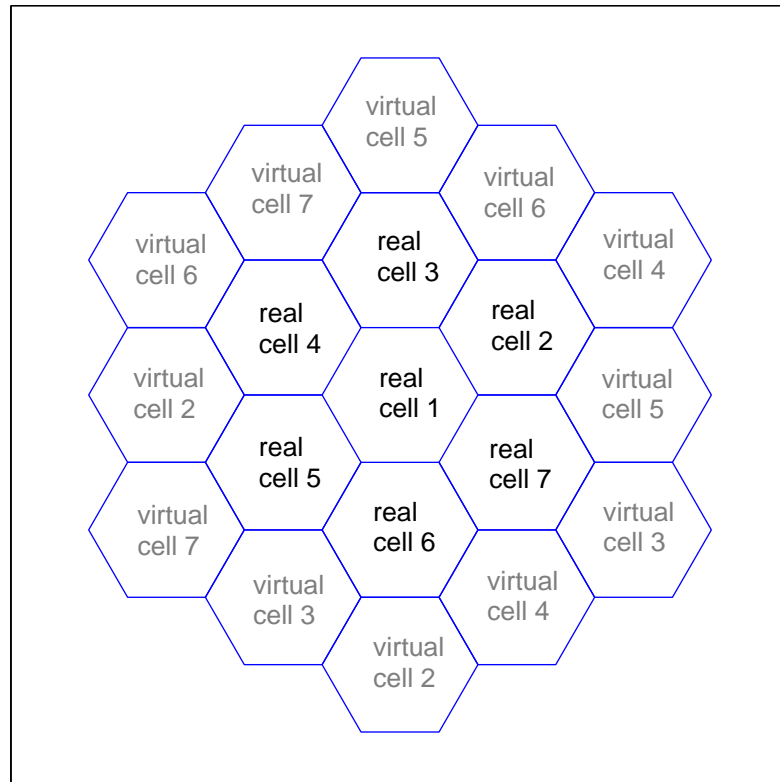


Figure 4.2: Wrap-around matching up virtual cells with real cells

proposed algorithm achieves the objective of maximising the satisfaction over a wide range of load scenarios with an underlying decreasing power consumption, which saves energy and mitigates the inter-cell interference.

Fig. 4.4 compares the proposed algorithm with the MaxSINR algorithm and shows significant superior effectiveness in terms of satisfying the users' QoS requirements. When the system is lightly loaded, the proposed algorithm is able to achieve 100% satisfaction. Compared with the MaxSINR algorithm, the proposed algorithm can sustainably provide around 20%-30% greater satisfaction regardless

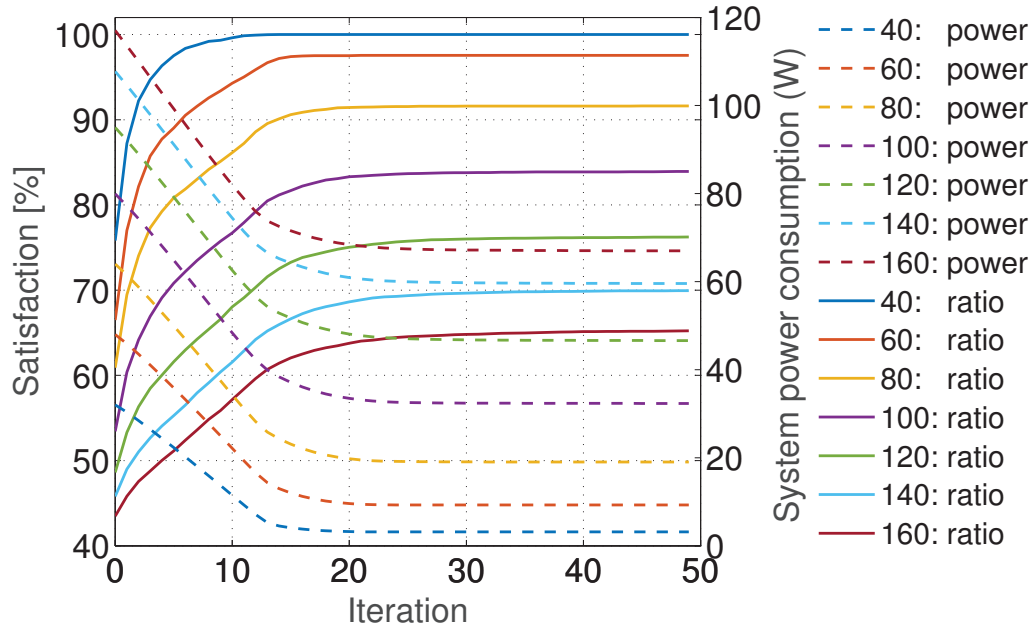


Figure 4.3: Satisfaction and system power consumption over iterations with different loads

of the load situation.

Further investigation of the QoS provisioning across different user rate classes is shown in Fig. 4.5. The proposed algorithm outperforms the other scheme for all three user rate classes (Table 4-B). For 256 kbps and 512 kbps requirements, the proposed algorithm can always achieve over 75% satisfaction in all load scenarios considered. When the service demand is relatively low (load less than 100), the proposed algorithm can achieve over 95% and even 100% satisfaction. However, for user rate-class 3 with a 1 Mbps rate requirement, MaxSINR is not able to provide very high satisfaction in all load situations. In addition, it is difficult for the proposed algorithm to maintain a high user satisfaction, especially at heavy loads. This is because no extra protection is given to the users with very high data rate requirements.

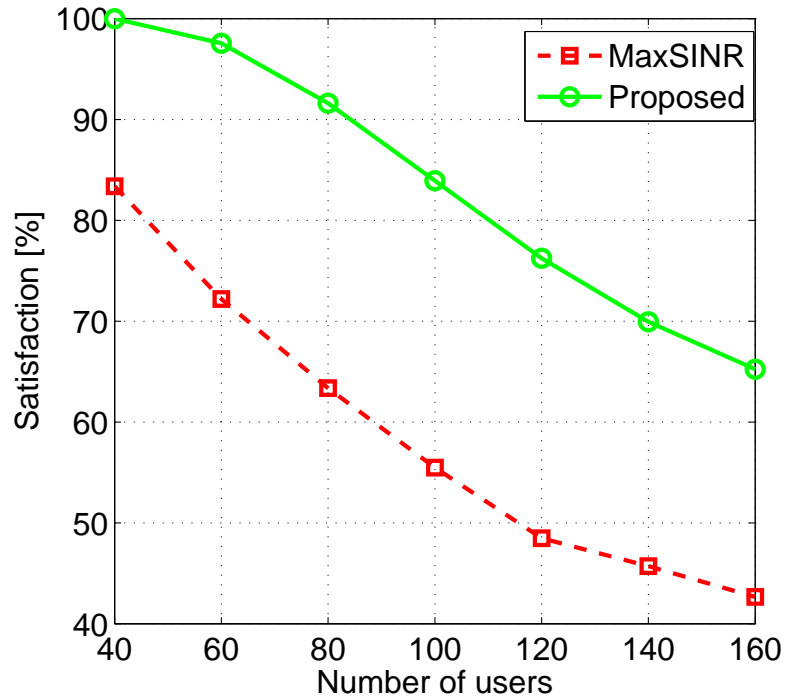


Figure 4.4: Satisfaction comparison of proposed and MaxSINR algorithms

4.3 Energy-Effective Network Deployment Scheme

In C-RAN, the light remote radio heads (RRHs) installed with antennas are densely deployed and connected to the baseband unit (BBU) pool through fronthaul links. Under dense C-RAN architecture with a large number of RRHs, a critical issue is introduced which is how to select appropriate RRHs to adapt to the temporal and spatial data dynamics in order to optimise the energy consumption. In this section, we propose an energy-effective network deployment (EEND) scheme with traffic demand satisfaction. The BBU is empowered with the ability to respond to the varying traffic demand by selecting a certain subset of RRHs. To the best of our knowledge, this is a pioneering work investigating the RRH deployment scheme for dense C-RAN. The network deployment problem is decomposed into

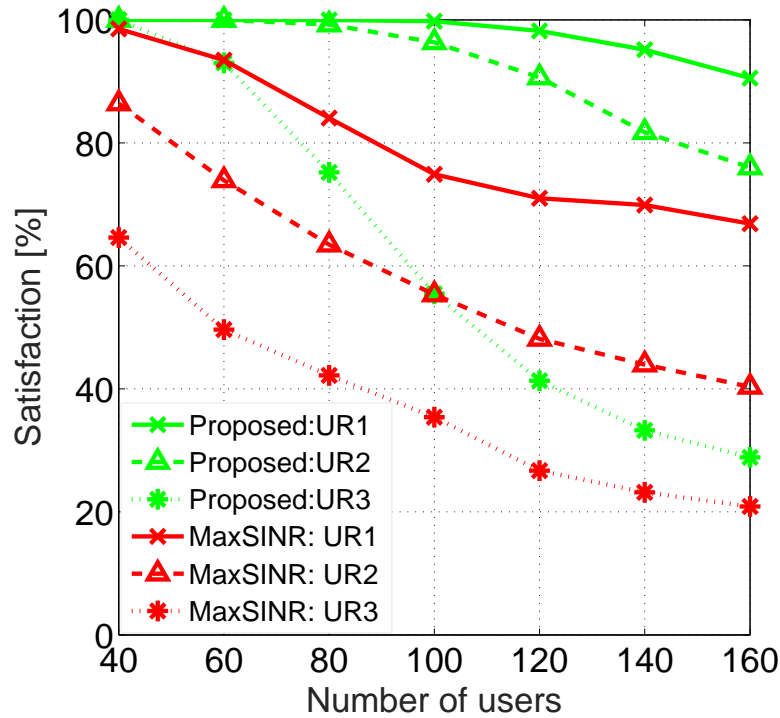


Figure 4.5: User satisfaction for different QoS requirement scenarios

two sub-optimal problems: RRH-traffic demand node association and active RRH set determination. The first sub-optimal problem is modelled as a multiple-choice multidimensional knapsack problem and solved by Lagrange multipliers. In order to solve the second sub-optimal problem, we deactivate the underutilised RRHs based on sleeping techniques.

4.3.1 System Model

We consider a C-RAN where the RRHs are densely deployed to provide seamless coverage. The transmit power budget of RRH j is P_j^{max} . Let \mathcal{L} denote the set of RRHs. To ensure full coverage in the defined area, we adopt the concept of traffic demand node (TDN) [ZWWW14a]. A TDN can be seen as an abstraction of the

average traffic demand in a small area. It is located at a centre of a predefined sub-area that represents an aggregated QoS requirements resulting from individual QoS demands of all potential users in this sub-area. We use minimum data rate requirement as the QoS indicator. The QoS requirement for a TDN corresponds to the aggregated expected traffic over the respective area (Fig. 4.6). We assume in total there are I TDNs in the given area. Each TDN has a rate requirement R_i .

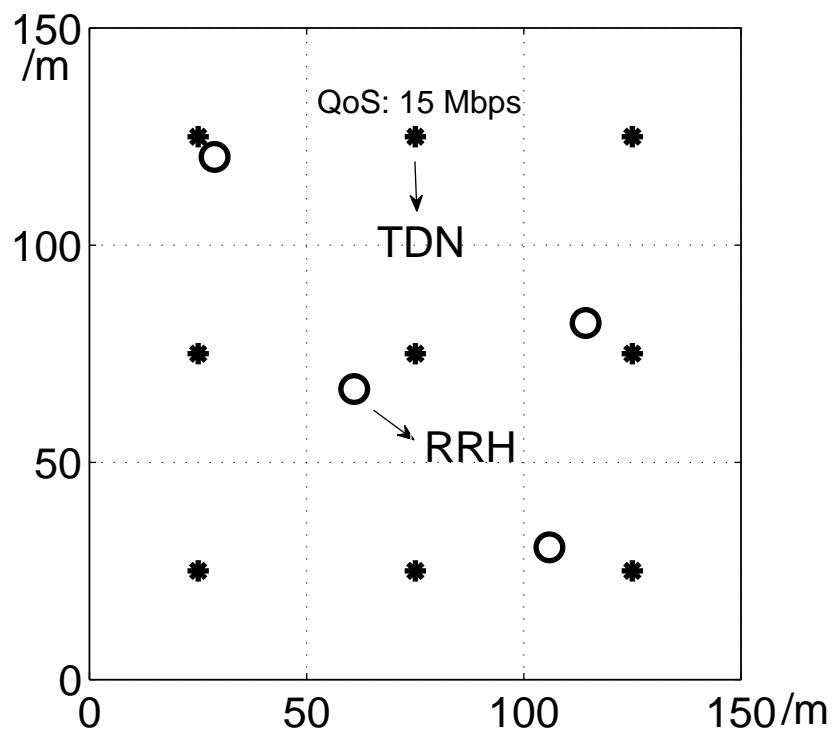


Figure 4.6: C-RAN deployment illustration with RRH and TDN

Denote h_{ij} as the channel gain between RRH j and TDN i , which is a function of the distance between the RRH and the TDN with a predefined path loss propagation model. Assuming the power budget is equally distributed over resource blocks (RBs) for all the RRHs, then for RRH j , the power allocated to each of its RB is P_j^{max}/N , where N is the total RB number for every RRH. Then the

achievable rate of RRH j to TDN i on each RB can be calculated as

$$r_{ij} = \Delta f \log_2 \left(1 + \frac{h_{ij} P_j^{max} / N}{\sigma^2} \right) \quad (4.15)$$

where Δf is the bandwidth of an RB. σ^2 is the noise power. The amount of the traffic load of TDN i in RRH j is

$$\epsilon_{ij} = \frac{\lceil R_i / r_{ij} \rceil}{N} \quad (4.16)$$

where $\lceil x \rceil$ is the minimum integer larger than x . It is only determined by its rate requirement and achievable spectrum efficiency.

The total power consumption of RRH j is the sum of the radio frequency (RF) energy and the circuit energy, denoted by P_{in} . We adopt the parameters from EARTH project [AGD⁺11]

$$P_{in} = \begin{cases} P_j^0 + \delta_j y_j P_j^{max} & y_j > 0 \\ P_j^s & y_j = 0 \end{cases} \quad (4.17)$$

where P_j^0 is the static power consumption as long as RRH j is active, P_j^s is the power consumption when RRH j is inactive. δ_j represents the slope of the load-dependent power consumption, and y_j is the traffic load of RRH j . As we can see from equation (4.17), P_{in} is a discontinuous function of the cell load.

Let $\mathbf{x} = [x_{ij}]_{|\mathcal{L}| \times I}$ be the assignment matrix where

$$x_{ij} = \begin{cases} 1 & \text{if TDN } i \text{ is assigned to RRH } j \\ 0 & \text{otherwise} \end{cases} \quad (4.18)$$

with $|\mathcal{L}|$ denotes the number of elements in set \mathcal{L} . Assume each TDN can only be

connected to one RRH. Then the cell load

$$y_j = \sum_{i=1}^I \epsilon_{ij} x_{ij}. \quad (4.19)$$

We define a boolean variable β_j to indicate the operation mode of RRH j

$$\beta_j = \begin{cases} 1 & y_j > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.20)$$

Then the total energy consumption of the network

$$\begin{aligned} P &= \sum_{j \in \mathcal{L}} P_{in} \\ &= \sum_{j \in \mathcal{L}} \sum_{i=1}^I \epsilon_{ij} x_{ij} \delta_j P_j^{max} + \sum_{j \in \mathcal{L}} (\beta_j P_j^0 + (1 - \beta_j) P_j^s) \end{aligned} \quad (4.21)$$

where $\sum_{j \in \mathcal{L}} \sum_{i=1}^I \epsilon_{ij} x_{ij} \delta_j P_j^{max}$ is the RF energy consumption of the network, and $\sum_{j \in \mathcal{L}} (\beta_j P_j^0 + (1 - \beta_j) P_j^s)$ is the static circuit energy consumption of the network.

4.3.2 Problem Formulation

Spatial-temporal redundancies resulting from traffic demand fluctuations present great opportunities for energy savings. Selecting a subset of RRHs corresponding to the most energy-effective network deployment is one appealing approach. Indeed, if the traffic demand decreases, partial or all rate requirements will become relatively small. This can be utilised to reduce the total network energy consumption by minimise the objective function in (4.21) subject to load constraints.

The objective of the network deployment is to minimise the overall power con-

sumption of the system while satisfying the traffic requirements and the allocated bandwidth of each RRH cannot exceed its budget. Then the problem can be written as the following optimisation problem with radio resource cost constraint:

$$\min P \quad (4.22a)$$

$$\text{s.t.} \quad \sum_{i=1}^I \epsilon_{ij} x_{ij} \leq 1 \quad \forall j \in \mathcal{L} \quad (4.22b)$$

$$\sum_{j \in \mathcal{L}} x_{ij} = 1 \quad i \in \{1, 2 \dots I\} \quad (4.22c)$$

$$x_{ij} \in \{0, 1\} \quad (4.22d)$$

The constraint in (4.22b) ensures no more radio resources than available is used in each RRH. (4.22c) states that each TDN should be connected to one RRH. The final constraint avoids partial assignment of TDNs to RRHs, which in turn leads to the combinatorial nature of the problem.

The above problem is very similar to the multi-choice multidimensional knapsack problem (MMKP). Due to the circuit power component considered in objective function (4.22a), we cannot decide the profit of each TDN-RRH association since it depends on the previous states of the packing process. This makes the above formulated problem not strict MMKP. We cannot directly use existing techniques for MMKP. Moreover, the fact that P_j^s is usually much smaller than P_j^0 according to the EARTH project makes the network tends to reduce the number of active RRHs to save energy. So the network topology is no longer static which makes the problem more difficult.

To tackle the challenges, we decompose problem (4.22) into two sub-optimal

problems:

Sub-optimal problem 1: Find the optimal TDN-RRH association for given active RRH set.

Sub-optimal problem 2: Determine the active RRH set for given TDN-RRH association matrix.

4.3.3 Proposed Network Deployment Scheme

To solve sub-optimal problem 1: we map it to the multi-choice multidimensional knapsack problem (MMKP) and use Lagrange multipliers to solve it.

To solve sub-optimal problem 2: we adopt sleep mode technique to deactivate the underutilised RRHs.

4.3.3.1 MMKP Mapping

The C-RAN with L RRHs is considered as an L dimensional knapsack. The available resources of the knapsack can be represented by $W = \{W_1, W_2 \dots W_L\}$. Since we consider the radio resource, W_j is the total number of resource blocks of RRH j and equals to N . Each TDN i is regarded as a class and the RRHs are the items in the class. Each item j of class i has a profit v_{ij} and requires resources from each dimension of the knapsack $w_{ij} = \{w_{ij}^1, w_{ij}^2, \dots, w_{ij}^L\}$. In terms of energy efficiency, the profit can be defined as

$$v_{ij} = r_{ij}/P_j^{max} \quad (4.23)$$

Since the allocation of TDN i only requires radio resources in the associated RRH j , we have

$$w_{ij}/W = \{\epsilon_{ij}^1, \dots, \epsilon_{ij}^s, \dots, \epsilon_{ij}^L\} \quad (4.24)$$

where

$$\epsilon_{ij}^s = \begin{cases} \epsilon_{ij} & s = j \\ 0 & otherwise \end{cases} \quad (4.25)$$

Thus, the first sub-optimal problem can be written as

$$\max_{x_{ij}} \sum_{i=1}^I \sum_{j=1}^L v_{ij} x_{ij} \quad (4.26a)$$

$$\text{s.t.} \quad \sum_{i=1}^I \sum_{j=1}^L \left(\frac{w_{ij}^s}{W^s}\right) x_{ij} \leq 1 \quad \forall s \in \mathcal{L} \quad (4.26b)$$

$$\sum_{j \in \mathcal{L}} x_{ij} = 1 \quad i \in \{1, 2 \dots I\} \quad (4.26c)$$

$$x_{ij} \in \{0, 1\} \quad (4.26d)$$

4.3.3.2 Lagrange Multipliers

By introducing the Lagrange multiplier λ_j associated with the radio resource constraint of each RRH, we can transform the constrained maximisation problem into unconstrained problem

$$\max_{x_{ij}} \left(\sum_{i=1}^I \sum_{j=1}^L v_{ij} x_{ij} - \sum_{i=1}^I \sum_{j=1}^L \lambda_j \epsilon_{ij} x_{ij} \right) \quad (4.27)$$

From expression (4.27), it is noted that if the Lagrange multiplier λ_j is known,

the optimisation problem can be solved. In fact, rewrite (4.27) as

$$\max_{x_{ij}} \left(\sum_{i=1}^I \sum_{j=1}^L (v_{ij} - \lambda_j \epsilon_{ij}) x_{ij} \right) \quad (4.28)$$

the optimal solution is given by

$$x_{ij}^* = \begin{cases} 1 & \text{if } \varrho_{ij} = v_{ij} - \lambda_j \epsilon_{ij} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.29)$$

where ϱ_{ij} is defined as the utility, a metric that combines the profit, radio resource cost and the Lagrange multiplier. Since one TDN can only associate to one RRH, we can choose one assignment from (4.29) with the maximum utility. Hence, the optimisation problem can be solved by computing the L Lagrange multipliers. The main difficulty is how to efficiently compute the λ_j set. Based on the concept of graceful degradation of the most valuable choices in [MJS97][GZF10], an initial solution x_{ij} is derived from (4.29) by setting all λ_j equal to 0. In this way, the utility equals to the profit. So each TDN will choose the best RRH regardless of the radio resource cost.

Then the total radio resource cost of each RRH π_j is calculated, denoted by a cost set $\psi = \{\pi_1, \pi_2, \dots, \pi_L\}$. If the initial assignment is feasible, which means none of π_j is larger than 1, then that is an optimal solution. Otherwise, the Lagrange multiplier associated to the most violated constraint is iteratively increased to force the reassignment till a feasible solution is found. In each iteration, the RRH j^* with the most constraint violation is found at first. Then for every TDN currently allocated to the RRH j^* , the Lagrange multiplier increase of the most violated

constraint required to move TDN i from RRH j^* to another RRH j is defined as

$$\Delta \lambda_{ij^* \rightarrow j} = \frac{v_{ij^*} - v_{ij} - (\lambda_{j^*} \epsilon_{ij^*} - \lambda_j \epsilon_{ij})}{\epsilon_{ij^*}} \quad (4.30)$$

The TDN I^* and another RRH J^* causing the least $\Delta \lambda_{ij^* \rightarrow j}$ is chosen for exchange. The process is repeated until a feasible solution not exceeding the resource constraints is found.

4.3.3.3 Deactivate Underutilised RRHs

The following Algorithm 1 describes the RRH deactivation process. It assumes initially all the RRHs are active (line 01) and obtains the initial TDN-RRH association solution (line 02). Then it calculates the resource utilisation ratio for each RRH (line 04) and sets the median value of the resource utilisation ratio set as the reference deactivation threshold (line 05). Then the algorithm would repeatedly deactivate the RRHs whose resource utilisation ratios are below the threshold (line 06) and solve the TDN-RRH association (line 07) until there are not enough RRHs to satisfy all the TDNs' QoS requirements. Lines 08-13 explain the threshold update process.

4.3.4 Simulation Parameters and Results

Based on 3GPP standard [Acc16a] and the EARTH project [AGD⁺11], the setting of system parameters and power parameters are listed in Table 4-C.

Due to the lack of existing comparison work, we adopt a widely-used heuristic search and optimisation method genetic algorithm (GA) as the comparison scheme.

Algorithm 1: RRH deactivation	
01	Initialise the active RRH set \mathcal{L}
02	Run MMKP-Lagrange algorithm to get the initial TDN-RRH association $[x_{ij}]$
03	if the RRHs are enough for the traffic demands
04	Calculate the RRH utilisation ratios then get the ψ
05	Set the threshold the median of ψ
06	Deactivate the RRHs whose utilisation ratios are below the threshold and get a new active RRH set \mathcal{L}'
07	Run MMKP-Lagrange algorithm to get a new TDN-RRH association $[x_{ij}']$
08	if the RRHs are enough for the traffic demands
09	Update the threshold as the minimum value larger than the old threshold
10	Go to line 06
11	else
12	if iteration==1 threshold values descend
13	Update the threshold as the maximum value smaller than the old threshold
14	Go to line 06
15	else
16	End the deactivation process
17	end
18	end
19	else
20	End the deactivation process
21	end

Value encoding is used to represent the solution [REBTH12]. A chromosome is defined having I genes corresponding to I groups of items, the i^{th} gene takes an integer number from the set $\mathcal{L} = \{1, 2, \dots, L\}$. This value means the item/RRH selected from the group. This encoding scheme assures that only one item is chosen from each group.

To handle the constraints, the following criteria are used to evaluate the fitness of a solution [Deb00]: any feasible solution is preferred to an infeasible solution; among two feasible solutions, the one having better objective function is preferred; among two infeasible solutions, the one having small constraint violation

Table 4-C: Simulation parameters

Parameter	Value
Scenario	Dense: 250 m × 250 m region with 9 RRHs Light: 500 m × 500 m region with 16 RRHs [LPDCJ15]
TDN grid	50 m × 50 m [ZWWW14b]
RRH power budget	30 dBm
Circuit power consumption (active)	84.0 W
Circuit power consumption (sleep)	56.0 W
Load-dependent power slope	2.8
RB bandwidth	180 kHz
Downlink RB number	25
Path loss model (dB)	$140.7 + 36.7 \log_{10}(d)$ (km)
Thermal noise	-174 dBm/Hz

is preferred. Then the fitness function considering both the objective function and constraints is

$$F(x) = \begin{cases} f(x) & \text{if } x \text{ is a feasible solution} \\ f_{worst} + v(x) & \text{otherwise} \end{cases} \quad (4.31)$$

where $f(x)$ is the objective function and $v(x)$ is the violation function. f_{worst} is the objective function of the worst feasible solution in the population. Thus the fitness of a feasible solution equals to its objective, while the fitness of an infeasible solution not only depends on the amount of constraint violation, but also on the population of solutions in the current generation. The parameters for genetic algorithm comparison are listed in Table 4-D.

Fig. 4.7 shows the comparison of active power consumption over different load in dense scenario. For each traffic load, 10 drops of simulation are conducted to get the average values. The total traffic demand is uniformly distributed across

Table 4-D: GA parameters

Parameter	Value
Population size	50
Generation number	200
Crossover probability	0.8
Mutation probability	0.05

the TDNs in considered region. Due to the assumption that all the RRHs are originally in sleep mode, the active power consumption means the extra power needed to activate the selected RRHs ($P^0 - P^s = 84.0 - 56.0 = 28 W$) plus the RF power. The blue-stripe bars show the proposed EEND algorithm. The red-dash bars are the results from constraints based genetic algorithm aiming to minimise the total energy consumption including RF power and static power in one step. The gray-grid bars represent the power consumption under the same traffic distribution when system is full-on as the baseline (BASE).

The proposed EEND scheme performs superior than that of GA in all load scenarios. This is because the EEND scheme works on the two factors of TDN-RRH association and the number of RRH repeatedly to obtain the final result. However the one-step GA only deals with TDN-RRH associations as chromosomes which is not as effective as the EEND scheme. In the BASE scheme, all the RRHs are active regardless of the varying traffic demand. The comparison of EEND and BASE scheme proves EEND achieves energy saving especially in light load scenarios. In average, EEND gives 25.20% power saving compared to GA and achieves even 45.77% power saving compared with BASE scheme.

When expanding the area to 500 m \times 500 m (light scenario), the result is presented in Fig. 4.8. In light scenario, EEND achieves even greater energy saving benefits, specifically, it saves 49.93% power consumption compared to GA scheme in average of all traffic demand scenarios. As to the BASE scheme, EEND saves

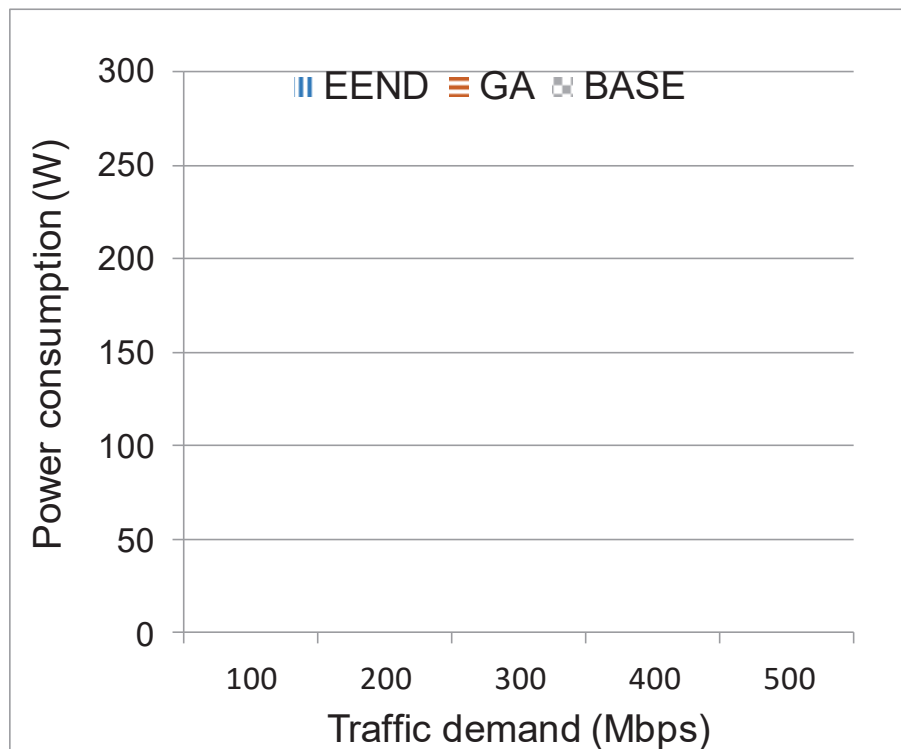


Figure 4.7: Power consumption with different load in dense scenario

averagely 57.97% energy.

The performances of the EEND are compared in two scenarios (Fig. 4.9). The configurations of the two scenarios are shown in Table 4-C. Although there are more RRHs in light scenario than that of dense scenario, the density is actually looser in light scenario as the area is four times of dense scenario. Although the total traffic demand is the same in two scenarios, the TDN distribution is sparser in light scenario, which leads to more RRHs needed to be active to ensure the traffic demand. That explains higher power consumption in light scenario than that of dense scenario.

Since EEND and GA work on very different principles, we use iteration number to compare the complexity. In the proposed scheme, the iteration number is de-

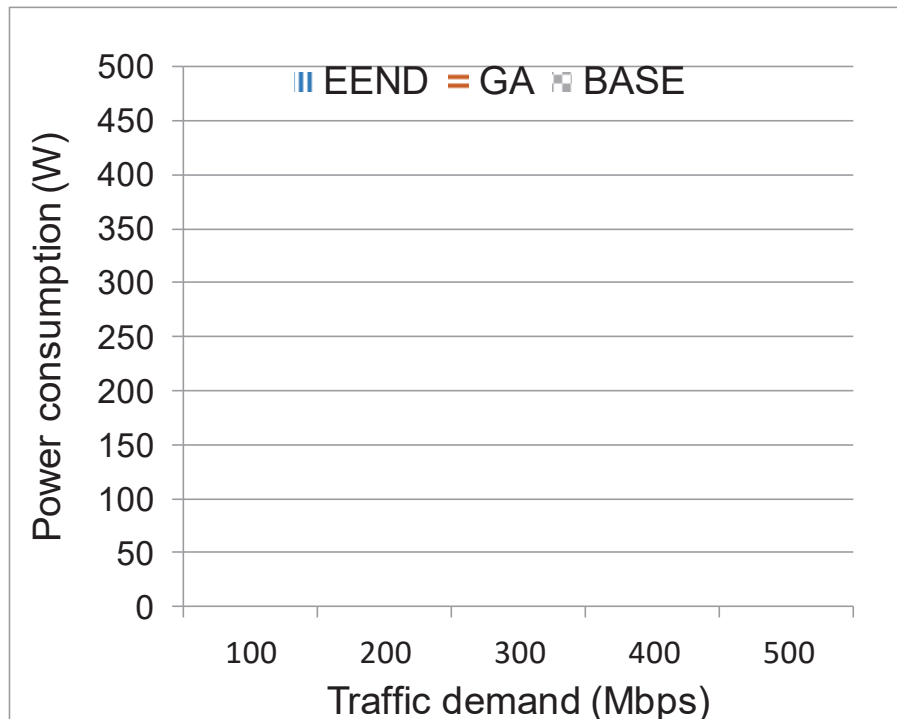


Figure 4.8: Power consumption with different load in light scenario

pendent on the number of RRHs. In the worst case, the iteration number is $L - 1$. However in GA, the iteration number is related to the population size N_{po} and the generation number N_{ge} and equals to $N_{po} \cdot N_{ge}$. In both scenarios in the previous simulations, the complexity of the proposed scheme is much lower than GA.

4.4 Joint Remote Radio Head Selection and User Association Scheme

The cloud radio access network (C-RAN) has been proposed recently as a promising network architecture to meet the explosive data traffic growth in 5G wireless

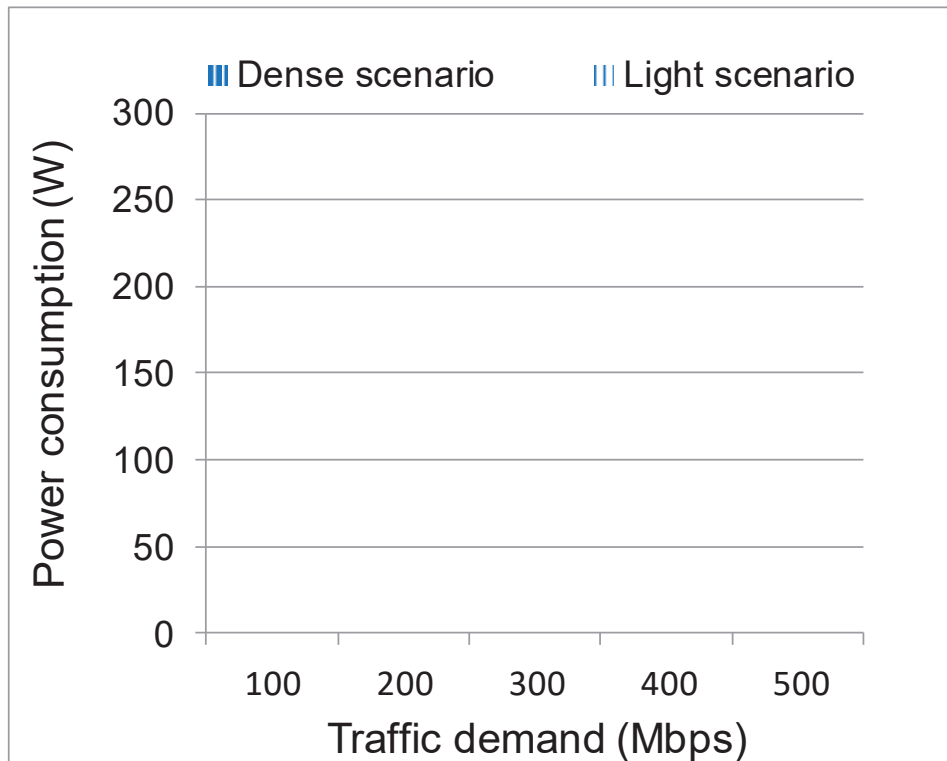


Figure 4.9: Power consumption comparison in dense and light scenarios

communication systems. In C-RAN, all baseband signal processing is centralised at one baseband unit (BBU) pool powered by cloud computing technologies. Whilst the remote radio heads (RRHs), left off on the cell sites, are connected to the BBU pool through fronthaul networks and can be densely deployed with low cost. However, a large number of active RRHs located close to each other may result in severe interference and inefficient energy consumption. To tackle the above challenges, we formulate a network power consumption minimisation (NPCM) problem, which selects a set of active RRHs and constructs the user association with the active RRHs. The capacity limitation of the fronthaul network is considered in the problem. Since the NPCM problem, formulated as an integer programming problem, is NP-hard, we propose a low complexity approximation algorithm that yields

the performance guarantees: joint RRH selection and user association (JRSUA) scheme. A simulation platform is developed to evaluate the performance of JRSUA under three fronthaul network scenarios: fibre, wireless and mixed. Simulation results show that the proposed JRSUA algorithm is able to provide near-optimal performance with reduced complexity and also outperforms the other counterparts.

4.4.1 System Model

We consider the deployment of a C-RAN in an urban area $\mathcal{A} \in \mathbb{R}^2$. There is a macro RRH located at the centre of this area to ensure the coverage and deliver signalings. A number of pico RRHs are deployed in \mathcal{A} to support the peak traffic load. The BBU pool, working as a centralised controller, manages the macro RRH and all the pico RRHs. In this thesis, we assume the BBU pool is resilient to processing and network failures. The RRHs are connected to the BBU pool via fronthaul links.

The index set of the RRHs in \mathcal{A} is denoted by $\mathcal{L} = \{0, 1, 2, \dots, L\}$, where index zero represents the macro RRH. The cells have the same indices corresponding to their RRHs. The fronthaul link connecting RRH j to the BBU pool is assumed to have finite capacity C_j . Let \mathcal{U} denote the set of users in the area \mathcal{A} . The QoS requirement of an arbitrary user i corresponds to the traffic demand at this location per time unit, which can be expressed in terms of minimum rate requirement R_i . It is assumed that all users can receive signals from both macro RRH and any pico RRH in \mathcal{A} .

The pico RRHs and the macro RRH work on different frequency band to avoid inter-tier interference. So a user served by the macro RRH will not receive interference from any of the pico RRHs. This guarantees that the macro RRH can

reach all users within area \mathcal{A} . We assume that all the pico RRHs reuse the same frequency band.

The macro RRH is assumed to work in active mode all the time since it is responsible to provide seamless coverage. As to the pico RRH, there are two operation modes: active or sleep. Let \mathcal{L}_{on} denote the set of active RRHs. If a pico RRH does not belong to \mathcal{L}_{on} , then it's been deactivated into sleep mode by the BBU pool for energy saving consideration.

Since the downlink traffic is much higher than that of uplink in wireless multimedia services [FSAA15], in designing the JRSUA scheme, we focus on downlink transmission.

The SINR of user i with respect to RRH j can be yielded as

$$\gamma_{ij} = \begin{cases} \frac{P_j h_{ij}}{\sum_{j' \in \mathcal{L}_{on} \setminus (\{j\} \cup \{0\})} P_{j'} h_{ij'} + \sigma^2} & j \in \{1, 2, \dots, L\} \\ \frac{P_j h_{ij}}{\sigma^2} & j = 0 \end{cases} \quad (4.32)$$

where P_j is the transmit power of RRH j . h_{ij} is the channel gain of user i with respect to RRH j , which includes path loss and shadowing. Note that fast fading is not considered here since the time scale of user association is much larger than the time scale of fast fading [SKYK11]. σ^2 denotes the thermal noise power. $\mathcal{L}_{on} \setminus (\{j\} \cup \{0\})$ is the set of all pico RRHs in \mathcal{A} excluding RRH j . (4.32) is the SINR measured by user i from a pico RRH. (4.33) is the SINR from the macro RRH to the user.

Accordingly, the spectral efficiency for user i , if it is served by RRH j , denoted

by c_{ij} , can be expected as [BB15],

$$c_{ij} = \log_2(1 + \gamma_{ij}) \quad (4.34)$$

Without loss of generality, the spectral efficiency c_{ij} can be regarded as the achievable rate on an RB by multiplying the RB bandwidth.

Given that n_{ij} RBs are given to user i by RRH j , the data rate seen by user i is

$$r_{ij} = n_{ij}c_{ij} \quad (4.35)$$

In the case of rate QoS constraint, each user intends to keep its rate above its requested rate threshold. In the beginning phase of RRH association, each user i requests a certain rate QoS class in terms of minimum required data rate R_i . Therefore, if user i is associated with RRH j , it is the duty of RRH j to satisfy the following the rate QoS constraint

$$r_{ij} \geq R_i \quad (4.36)$$

By substituting (4.35) in the above inequality, we have

$$n_{ij} \geq \frac{R_i}{c_{ij}} \quad (4.37)$$

We indicate the smallest integer greater than the right hand side of the above inequality by \bar{n}_{ij} as follows

$$\bar{n}_{ij} = \left\lceil \frac{R_i}{c_{ij}} \right\rceil \quad (4.38)$$

where $\lceil \cdot \rceil$ represents the ceil function. Inequalities (4.36) and (4.37) and equation

(4.38) indicate that if user i requires rate QoS class of minimum rate R_i , RRH j will be obliged to allocate at least \bar{n}_{ij} RBs to that user

$$n_{ij} \geq \bar{n}_{ij} \quad (4.39)$$

Assume that RRH j has N_j RBs in total available to allocate to all the users associated with it, we have

$$\epsilon_{ij} = \frac{\bar{n}_{ij}}{N_j} \quad (4.40)$$

where ϵ_{ij} is defined as the radio resource cost, meaning the number of RBs required to satisfy the rate requirement of user i from RRH j , divided by the total available RBs in cell j .

According to the frequency reuse mentioned above in our system model, we can further assume that the macro RRH has N_M RBs available. Since the frequency reuse factor of pico RRHs in \mathcal{A} is 1, let each pico RRH has N_P RBs in total, then we can rewrite the above equation as

$$\epsilon_{ij} = \begin{cases} \frac{\bar{n}_{ij}}{N_P} & j \in \{1, 2, \dots, L\} \\ \frac{\bar{n}_{ij}}{N_M} & j = 0 \end{cases} \quad (4.41)$$

$$(4.42)$$

(4.41) is the radio resource cost for pico RRH-user association and (4.42) is that of the macro RRH-user association.

If user i is served by RRH j , the BBU needs to send the baseband signal to RRH j over its fronthaul link at the estimated rate r_{ij} [LZ16]. For the fronthaul links, the fronthaul resource cost, denoted by η_{ij} , is defined as the ratio of the data

rate of user i , to the available fronthaul capacity of RRH j [HLo16][DY14], that is

$$\eta_{ij} = \frac{r_{ij}}{C_j} \quad (4.43)$$

4.4.2 JRSUA Scheme

4.4.2.1 Problem Formulation and Analysis

With the goal of minimising the network power consumption while satisfying the QoS requirements of the users in area \mathcal{A} , the proposed JRSUA scheme determines the active set of the pico RRHs and the user association with the active RRHs. Let \mathbf{x} be the user association indicator of which value is 1 if user i is associated with RRH j , and 0 otherwise. The user association matrix can be represented by $\mathbf{x} = [x_{ij}]_{|\mathcal{L}| \times |\mathcal{U}|}$, with $|\mathcal{S}|$ denotes the number of elements in set \mathcal{S} .

A. Power model

We adopt the following empirical linear model [AGD⁺11] for the power consumption of an RRH, denoted by P_{in} :

$$P_{in} = \begin{cases} P_j^0 + \delta_j y_j P_j & y_j > 0 \\ P_j^s & y_j = 0 \end{cases} \quad (4.44)$$

where P_j^0 is the static circuit power consumption as long as RRH j is active, P_j^s is the static circuit power consumption when RRH j is in sleep mode. δ_j is the slope of the load-dependent power consumption. y_j is the traffic load of RRH j , which can be calculated by

$$y_j = \sum_{i \in \mathcal{U}} \epsilon_{ij} x_{ij}. \quad (4.45)$$

Then the network power consumption can be expressed as

$$\begin{aligned}
 P_{total} &= \sum_{j \in \mathcal{L}} P_{in} & (4.46) \\
 &= \sum_{j \in \mathcal{L}_{on}} (P_j^0 + \delta_j y_j P_j) + \sum_{j \in \mathcal{L} \setminus \mathcal{L}_{on}} P_j^s \\
 &= \sum_{j \in \mathcal{L}_{on}} (P_j^0 + \sum_{i \in \mathcal{U}} \delta_j \epsilon_{ij} x_{ij} P_j) + \sum_{j \in \mathcal{L} \setminus \mathcal{L}_{on}} P_j^s \\
 &= \sum_{j \in \mathcal{L}_{on}} \sum_{i \in \mathcal{U}} \delta_j \epsilon_{ij} x_{ij} P_j + (P_j^0 - P_j^s) |\mathcal{L}_{on}| + P_j^s |\mathcal{L}|
 \end{aligned}$$

Since the last term is constant, we can re-define the NPCM problem by omitting the constant term $P_j^s |\mathcal{L}|$ as follows

$$\mathcal{P}(\mathbf{x}, \mathcal{L}_{on}) = \sum_{j \in \mathcal{L}_{on}} \sum_{i \in \mathcal{U}} \delta_j \epsilon_{ij} x_{ij} P_j + (P_j^0 - P_j^s) |\mathcal{L}_{on}| \quad (4.47)$$

B. The network power consumption minimisation problem

The network power consumption minimisation (NPCM) problem, which aims to minimise the network power consumption while satisfying user QoS requirement

in the considered area, can be formulated as:

$$\min_{\mathbf{x}, \mathcal{L}_{on}} \mathcal{P}(\mathbf{x}, \mathcal{L}_{on}) \quad (4.48a)$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{U}} \epsilon_{ij} x_{ij} \leq 1 \quad \forall j \in \mathcal{L}_{on} \quad (4.48b)$$

$$\sum_{i \in \mathcal{U}} \eta_{ij} x_{ij} \leq 1 \quad \forall j \in \mathcal{L}_{on} \quad (4.48c)$$

$$\sum_{j \in \mathcal{L}_{on}} x_{ij} = 1 \quad \forall i \in \mathcal{U} \quad (4.48d)$$

$$x_{ij} \in \{0, 1\} \quad (4.48e)$$

$$r_i \geq R_i \quad (4.48f)$$

$$\gamma_i \geq \Lambda_{th} \quad (4.48g)$$

Constraint (4.48b) indicates that the total bandwidth allocated to the associated users by each RRH cannot exceed the bandwidth budget. Constraint (4.48c) assures no more fronthaul capacity than available is used in each RRH connected to BBU pool. Constraint (4.48d) indicates that each user in \mathcal{A} is associated with only one RRH, either one of the pico RRHs or the macro RRH. To avoid partial assignment of users to RRHs, constraint (4.48e) is enforced, which in turn leads to the combinatorial nature of the problem. Constraint (4.48f) ensures that the expected bit rate of user i , denoted as r_i , meets the minimum data rate requirement of each user. The last constraint sets the association rule that user i should associate to the RRH with received SINR γ_i exceeding the minimum threshold Λ_{th} .

4.4.2.2 Proposed JRSUA solution

Solving the above problem is quite challenging since it has the set variables as well as integer variables. To tackle the difficulties, we first solve the user association problem (P-UA) which minimises the network power consumption for given \mathcal{L}_{on} . Then we propose an iterative algorithm for solving the NPCM problem, which finds the best set of active RRHs (P-RS) by solving the P-UA iteratively with different \mathcal{L}_{on} at each iteration.

A. P-UA

When \mathcal{L}_{on} is given, the P-UA can be formulated as

$$\min_{\mathbf{x}} \mathcal{P}(\mathbf{x}, \mathcal{L}_{on}) \quad (4.49a)$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{U}} \epsilon_{ij} x_{ij} \leq 1 \quad \forall j \in \mathcal{L}_{on} \quad (4.49b)$$

$$\sum_{i \in \mathcal{U}} \eta_{ij} x_{ij} \leq 1 \quad \forall j \in \mathcal{L}_{on} \quad (4.49c)$$

$$\sum_{j \in \mathcal{L}_{on}} x_{ij} = 1 \quad \forall i \in \mathcal{U} \quad (4.49d)$$

$$x_{ij} \in \{0, 1\} \quad (4.49e)$$

$$r_i \geq R_i \quad (4.49f)$$

$$\gamma_i \geq \Lambda_{th} \quad (4.49g)$$

As for the given \mathcal{L}_{on} , the last term in (4.47) is fixed, so we can rewrite the above

P-UA as

$$\zeta(\mathcal{L}_{on}) = \min_{\mathbf{x}} \sum_{j \in \mathcal{L}_{on}} \sum_{i \in \mathcal{U}} \delta_j \epsilon_{ij} x_{ij} P_j \quad (4.50a)$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{U}} \epsilon_{ij} x_{ij} \leq 1 \quad \forall j \in \mathcal{L}_{on} \quad (4.50b)$$

$$\sum_{i \in \mathcal{U}} \eta_{ij} x_{ij} \leq 1 \quad \forall j \in \mathcal{L}_{on} \quad (4.50c)$$

$$\sum_{j \in \mathcal{L}_{on}} x_{ij} = 1 \quad \forall i \in \mathcal{U} \quad (4.50d)$$

$$x_{ij} \in \{0, 1\} \quad (4.50e)$$

$$r_i \geq R_i \quad (4.50f)$$

$$\gamma_i \geq \Lambda_{th} \quad (4.50g)$$

The P-UA can be mapped into a multiple-choice multidimensional knapsack problem (MMKP). In the MMKP, we are given I classes of items, where each class i has J_i items. Each item j of class i has a profit, and requires multidimensional resources $w_{ij} = \{w_{ij}^1, w_{ij}^2, \dots, w_{ij}^m\}$. The amounts of available resources of the knapsack are given by $W = \{W_1, W_2, \dots, W_m\}$. The aim of the MMKP is to pack exactly one item from each class in order to maximise or minimise the total profit value, subject to the resource constraints. In our work, the number of I maps to the number of users $|\mathcal{U}|$, the set of J_i corresponds to the active RRH set \mathcal{L}_{on} . The dimension of the resources is $2|\mathcal{L}_{on}|$ including $|\mathcal{L}_{on}|$ dimensional radio resources and $|\mathcal{L}_{on}|$ dimensional fronthaul resources. The portion of resources required to associate user i to RRH j is given by

$$w_{ij}/W = \{\epsilon_{ij}^1, \epsilon_{ij}^2, \dots, \epsilon_{ij}^{|\mathcal{L}_{on}|}, \eta_{ij}^1, \eta_{ij}^2, \dots, \eta_{ij}^{|\mathcal{L}_{on}|}\} \quad (4.51)$$

Since user i only needs resources from the associated RRH j , we have

$$\epsilon_{ij}^s = \begin{cases} \epsilon_{ij} & s = j \\ 0 & \text{otherwise} \end{cases} \quad (4.52)$$

$$\eta_{ij}^s = \begin{cases} \eta_{ij} & s = j \\ 0 & \text{otherwise} \end{cases} \quad (4.53)$$

Since the user association problem (P-UA) is transformed into an MMKP, any technique available to solve the MMKP can be used. There exist two different types of algorithms to solve the MMKP, namely: exact and heuristic algorithms. Due to its high computational complexity, exact algorithms are not suitable for most real-time decision-making applications [HMS04]. So the alternative is to use approximate heuristic approaches with polynomial time complexity. In this work, we develop a heuristic user association algorithm based on [MJS97]. The algorithm of Moser et al. [MJS97] relies on a theorem proven by Everett [EI63] that makes Lagrange multipliers applicable to discrete optimisation problems, such as the MMKP. In this regard, algorithm in [MJS97] has already been considered as a useful tool in some works [KNY08][GZF10] to solve resource management problems in OFDMA wireless networks. Therefore, we have adapted the algorithm of [GZF10] to our specific user association problem.

We can transform the constrained minimisation problem into unconstrained problem by introducing the Lagrange multipliers λ and ξ related to the radio resource constraint and fronthaul resource constraint of each RRH

$$\min_{\mathbf{x}} \left(\sum_{j \in \mathcal{L}_{on}} \sum_{i \in \mathcal{U}} (\delta_j \epsilon_{ij} P_j + \lambda_j \epsilon_{ij} + \xi_j \eta_{ij}) x_{ij} \right) \quad (4.54)$$

If denoting the utility as

$$\varrho_{ij} = \delta_j \epsilon_{ij} P_j + \lambda_j \epsilon_{ij} + \xi_j \eta_{ij} \quad (4.55)$$

then the optimal user association solution is given by

$$x_{ij}^* = \begin{cases} 1 & j = \operatorname{argmin}(\varrho_{ij}) \\ 0 & \text{otherwise} \end{cases} \quad (4.56)$$

The main difficulty in solving the problem is how to efficiently compute the Lagrange multipliers. In this regard, [MJS97] uses an approach based upon the concept of graceful degradation of the most valuable choices. First, an initial temporary solution is derived from (4.54) by considering all Lagrange multipliers equal to zero (i.e., the utility equals to the profit, so that each user is assigned to the “best” RRH irrespective of its radio or fronthaul resource cost). Then, Lagrange multipliers associated to RRHs that would exceed available resources are iteratively increased until a feasible solution is found. That is, the increase of Lagrange multiplier associated to an RRH causes an increase in the utility of its served users, so that some of them could be reassigned to other RRHs providing lower utility.

The P-UA algorithm, shown in Algorithm 2, consists of two phases, namely: initialisation (line 01-08) and drop (line 09-28). Firstly, Lagrange multipliers are set to zero (line 01), and then user profits and resource costs are computed (lines 02-04) for each user. In order to reduce the computational complexity, not all RRHs are viewed as potential choices. Instead, each user i is assumed to have a candidate set, denoted as z_i , composed by the RRHs providing the $\gamma_{ij} \geq \Lambda_{th}$. Then, an initial association is obtained by selecting the most valuable RRH for

Algorithm 2: P-UA algorithm	
01	Set Lagrange multipliers to zero: $\lambda_j \leftarrow 0, \xi_j \leftarrow 0$
02	for each user $i = 1, 2, \dots, \mathcal{U} $
03	for each RRH $j = 1, 2, \dots, z_i $ on the candidate set for i compute profit and costs
04	$\delta_j \epsilon_{ij} P_j$ (4.50a), ϵ_{ij} (4.40), η_{ij} (4.43)
05	Find the most valuable RRH $j = \operatorname{argmin}_j \{\delta_j \epsilon_{ij} P_j\}$ for each user i and update its association accordingly $x_{ij} \leftarrow 1$
06	for each RRH j compute the total radio/fronthaul resource costs
07	$\pi_j = \sum_{i \in \mathcal{U}} \epsilon_{ij} x_{ij}, \tau_j = \sum_{i \in \mathcal{U}} \eta_{ij} x_{ij}$
08	Conform total resource cost set $\psi = \{\pi_1, \pi_2, \dots, \pi_{ \mathcal{L}_{on} }, \tau_1, \tau_2, \dots, \tau_{ \mathcal{L}_{on} }\}$
09	while $\psi_j > 1$ for any j do
10	Find the RRH j^* holding the most offending constraint violation $s = \operatorname{argmax}_j \{\psi_j\}$, where $j^* = s$ if $s = 1, 2, \dots, \mathcal{L}_{on} $ and $j^* = s - \mathcal{L}_{on} + 1$ otherwise
11	Compute the increase of the multiplier associated to constraint s (radio/fronthaul) of RRH j^*
12	for $\{i x_{ij^*} = 1\}$
13	for $\{j = 1 : z_i \}$
14	Compute the increase of the utility
15	$\Delta \varrho_{ij^* \rightarrow j} = \delta_j \epsilon_{ij} P_j - \delta_{j^*} \epsilon_{ij^*} P_{j^*} + \lambda_j \epsilon_{ij} - \lambda_{j^*} \epsilon_{ij^*} + \xi_j \eta_{ij} - \xi_{j^*} \eta_{ij^*}$
16	if $j^* = s$
17	$\Delta \lambda_{ij^* \rightarrow j} = \Delta \varrho_{ij^* \rightarrow j} / \epsilon_{ij^*}$
18	else
19	$\Delta \xi_{ij^* \rightarrow j} = \Delta \varrho_{ij^* \rightarrow j} / \eta_{ij^*}$
20	Find the user to change its association and re-evaluate the corresponding Lagrange multiplier
21	if $j^* = s$
22	$I^* J^* = \operatorname{argmin}_{ij} \{\Delta \lambda_{ij^* \rightarrow j}\}, I' J' = \operatorname{argmin}_{ij} \{\Delta \lambda_{ij^* \rightarrow j}\}, I' \neq I^*$
23	$\lambda_{j^*} \leftarrow \lambda_{j^*} + (\Delta \lambda_{I^* j^* \rightarrow J^*} + \Delta \lambda_{I' j^* \rightarrow J'}) / 2$
24	else
25	$I^* J^* = \operatorname{argmin}_{ij} \{\Delta \xi_{ij^* \rightarrow j}\}, I' J' = \operatorname{argmin}_{ij} \{\Delta \xi_{ij^* \rightarrow j}\}, I' \neq I^*$
26	$\xi_{j^*} \leftarrow \xi_{j^*} + (\Delta \xi_{I^* j^* \rightarrow J^*} + \Delta \xi_{I' j^* \rightarrow J'}) / 2$
27	$x_{I^* j^*} \leftarrow 0, x_{I' j^*} \leftarrow 1$
28	$\pi_{j^*} \leftarrow \pi_{j^*} - \epsilon_{I^* j^*}, \tau_{j^*} \leftarrow \tau_{j^*} - \eta_{I^* j^*}$
29	$\pi_{J^*} \leftarrow \pi_{J^*} + \epsilon_{I' j^*}, \tau_{J^*} \leftarrow \tau_{J^*} + \eta_{I' j^*}$
30	Final user association $\mathbf{x} = [x_{ij}]_{ \mathcal{L}_{on} \times \mathcal{U} }$

each user (line 05). The total radio and fronthaul costs at each RRH j , denoted by π_j and τ_j , respectively, are computed (lines 06-07) and the resource cost set $\psi = \{\pi_1, \pi_2, \dots, \pi_{|\mathcal{L}_{on}|}, \tau_1, \tau_2, \dots, \tau_{|\mathcal{L}_{on}|}\}$ is conformed (line 08).

If the initial assignment is feasible (i.e., none of the elements of ψ is greater than 1), that is an optimal solution. Otherwise, the algorithm continues in the drop phase. Within the drop phase, Lagrange multiplier associated to the most offending constraint violation is repeatedly increased to force user reassignments till a solution not exceeding resource constraints is found. In each iteration of this phase, the RRH j^* with the most offending constraint violation s is determined (line 10), where

$$j^* = \begin{cases} s & s = 1, 2, \dots, |\mathcal{L}_{on}| \\ s - |\mathcal{L}_{on}| + 1 & otherwise \end{cases} \quad (4.57)$$

For each user i currently allocated to the RRH j^* (line 12), the Lagrange multiplier increase of the most offending constraint s required to move user i from RRH j^* to another RRH j in its candidate set is computed (lines 12-18). This is done so that the utility of user i at the overloaded RRH j^* , ϱ_{ij^*} , is increased to a value larger than or equal to the utility on the candidate RRH j , ϱ_{ij} .

Thus, if the most offending constraint violation at RRH j^* is on radio resources, namely RB, the increment to Lagrange multiplier λ_{j^*} should be that:

$$(\delta_{j^*} \epsilon_{ij^*} P_{j^*} + (\lambda_{j^*} + \Delta \lambda_{ij^* \rightarrow j}) \epsilon_{ij^*} + \xi_{j^*} \eta_{ij^*}) \geq (\delta_j \epsilon_{ij} P_j + \lambda_j \epsilon_{ij} + \xi_j \eta_{ij}) \quad (4.58)$$

So the increment $\Delta \lambda_{ij^* \rightarrow j}$ to the radio Lagrange multiplier can be computed as:

$$\Delta \lambda_{ij^* \rightarrow j} \geq \frac{\delta_j \epsilon_{ij} P_j - \delta_{j^*} \epsilon_{ij^*} P_{j^*} + \lambda_j \epsilon_{ij} - \lambda_{j^*} \epsilon_{ij^*} + \xi_j \eta_{ij} - \xi_{j^*} \eta_{ij^*}}{\epsilon_{ij^*}} \quad (4.59)$$

Similarly, if the most offending constraint violation is on fronthaul resources,

the increment to Lagrange multiplier ξ_{j^*} should be that:

$$(\delta_{j^*}\epsilon_{ij^*}P_{j^*} + \lambda_{j^*}\epsilon_{ij^*} + (\xi_{j^*} + \Delta\xi_{ij^*\rightarrow j})\eta_{ij^*}) \geq (\delta_j\epsilon_{ij}P_j + \lambda_j\epsilon_{ij} + \xi_j\eta_{ij}) \quad (4.60)$$

So the increment $\Delta\xi_{ij^*\rightarrow j}$ to the fronthaul Lagrange multiplier can be computed as:

$$\Delta \xi_{ij^*\rightarrow j} \geq \frac{\delta_j\epsilon_{ij}P_j - \delta_{j^*}\epsilon_{ij^*}P_{j^*} + \lambda_j\epsilon_{ij} - \lambda_{j^*}\epsilon_{ij^*} + \xi_j\eta_{ij} - \xi_{j^*}\eta_{ij^*}}{\eta_{ij^*}} \quad (4.61)$$

where the numerator in (4.59) and (4.61) is the increase of the utility of user i , denoted as $\varrho_{ij^*\rightarrow j}$. For each RRH j in the candidate set z_i of users currently allocated to RRH j^* , the increase of the corresponding Lagrange multiplier is computed in lines 12-18. Then, the user I^* and candidate RRH J^* causing the least increase of the corresponding multiplier is chosen for exchange (lines 19-25) as this choice minimises the gap between the optimal solution characterized by (4.56) and the new P-UA solution obtained at this point. However, if the multiplier increase is just computed as the equality, users may have the same utility towards multiple RRHs. To avoid this problem, we compute the increment to be added to the corresponding multiplier as the average between the least increase, corresponding to user I^* and RRH J^* , and the second least increase obtained with user I' and RRH J' . This choice guarantees that only one user is reassigned at each iteration and the next P-UA solution is stable (equal utilities due to the update of the multipliers are avoided). The reassignment of the selected user is performed (line 26), and radio and transport resource costs are updated accordingly (lines 27-28). The process is repeated until a solution not exceeding resource constraints is determined, which is the final user association (line 29).

B. P-RS

Since the RRHs are typically deployed on the basis of peak traffic volume and stayed turned-on irrespective of traffic load, it is possible to save huge energy by switching off some underutilised RRHs during off-peak times. In this section, we shall start by discussing the effect of switching off one RRH. Based on the study of this simple case, we propose a sequential (continuous) algorithm, in which RRHs get switched off one by one while ensuring users' QoS.

Let us consider a simple case in which one RRH is turned off. Apparently, this would result in an increase in the system load of neighbouring RRHs. This is not only because those users originally associated with the switched off RRHs need to be transferred to the neighbour RRHs, but also because this will result lower service rates due to farther distances between the users and their new serving RRHs. However, on the other hand, turning off a RRH may bring positive impact on the system load due to reduced inter-cell interference, in particular, some users originally associated with neighbouring RRHs will see potentially higher service rates.

Now we focus on determining the active RRH set such that $\mathcal{L}_{on}^* = \text{argmin}\zeta(\mathcal{L}_{on})$. The P-RS problem is a challenging combinatorial problem with $O(2^L)$ possible cases, which makes it very difficult to find an optimal solution through exhaustive search, especially when L is large. Thus, we propose a heuristic algorithm to solve the NPCM problem iteratively.

Our proposed algorithm starts from the point where all RRHs are set active (line 01 in Algorithm 3). Then it finds the best RRH to deactivate into sleep mode for energy conservation in each iteration. Specifically, at the beginning of each iteration, it calculates the dynamic power consumption ζ when excluding the pico

Algorithm 3: proposed JRSUA algorithm

```

01 Initialise  $\mathcal{L}_{on} \leftarrow \mathcal{L}$ 
02  $\text{minNPC} \leftarrow \zeta(\mathcal{L}_{on})$ 
03 do
04   for  $\forall j \in \mathcal{L}_{on}$ 
05     calculate  $\zeta(\mathcal{L}_{on} \setminus \{j\})$ 
06   end
07    $l \leftarrow \text{argmin} \zeta(\mathcal{L}_{on} \setminus \{j\})$ 
08   if  $\zeta(\mathcal{L}_{on} \setminus \{l\}) - \zeta(\mathcal{L}_{on}) < P_j^0 - P_j^s$  where  $j \neq 0$ 
09      $\mathcal{L}_{on} \leftarrow \mathcal{L}_{on} \setminus \{l\}$ 
10      $\text{minNPC} \leftarrow \zeta(\mathcal{L}_{on})$ 
11   else
12     break
13   end
14 end

```

RRH from current \mathcal{L}_{on} one by one. Then it selects the pico RRH l with lowest ζ . In line 08, the algorithm checks whether there is a net energy saving. If so, we put RRH l into sleep and update the network power consumption value (minNPC in line 10). Otherwise, we stop the algorithm.

4.4.3 Simulation Parameters and Results

4.4.3.1 Simulation Parameters

In order to validate and evaluate the performance of the proposed scheme, we present numerical and simulation results in this section. Simulation set-up mainly follows the guidelines of 3GPP technical reports [Acc16a][Acc16c]. We adopt the power consumption model from [AGD⁺11] and set the value of power consumption for macro RRH and pico RRH accordingly. The primary system parameters are summarised in Table 4-E.

Table 4-E: Simulation parameters

Parameter	Value
Area	1 km \times 1 km [EYF ⁺ 15]
Number of macro RRH	1
Number of pico RRH	12 [EYF ⁺ 15]
Number of RBs	25
RB bandwidth	180 kHz
Maximum transmit power of macro RRH	43 dBm
Maximum transmit power of pico RRH	30 dBm
Load-dependent power slope of macro RRH	4.7
Load-dependent power slope of pico RRH	4.0
Circuit power consumption of pico RRH (active)	6.8 W
Circuit power consumption of pico RRH (sleep)	4.3 W
Path loss for macro RRH	$128.1 + 37.6 \log_{10}(d)$
Path loss for pico RRH	$140.7 + 36.7 \log_{10}(d)$
Shadowing standard deviation	10 dB
Thermal noise	-174 dBm/Hz
SINR threshold	-5 dB
User data rate requirement	512 kbps

4.4.3.2 User Association Validation

Since the key module of the proposed JRSUA algorithm (Algorithm 3) is user association (UA), we first validate the user association module (Algorithm 2) with reduced network scale. We consider 500 m \times 500 m area with one macro RRH and 4 pico RRHs. The channel bandwidth is 1.4 MHz and configured with 6 RBs [Acc16c]. Here we validate the user association algorithm with radio resource constraint by setting all the C_j into infinity.

Fig. 4.10 illustrates the basic settings before running any algorithms. As we can interpret from Fig. 4.10, there is one macro RRH located at the centre of \mathcal{A} . Four pico RRHs are installed in predefined locations. Here the users representing traffic demands are randomly distributed. And we can adopt other kinds of distributions mapping to different traffic dynamics.

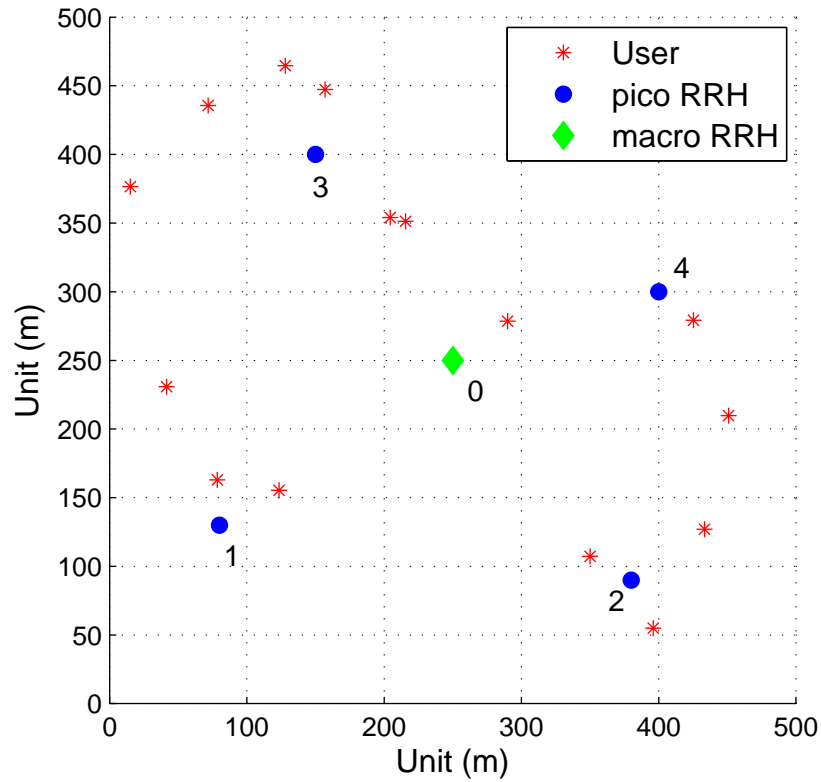


Figure 4.10: Settings of macro RRH, pico RRHs and users

Fig. 4.11(a) and Fig. 4.11(b) shows the initial user association result with 15 users. In Fig. 4.11(a), the solid lines connecting the RRH to the user show the initial association. Fig. 4.11(b) indicates the RB utilisation ratios of all the RRHs. RRH with ID 0 is the macro RRH. Since none of the RB utilisation ratios are larger than 1, the initial association solution is feasible. The algorithm executes line 08 and then directly goes to line 29 (Algorithm 2). We can verify that in initial phase, all the users associate with the RRHs consumes the least power to provide the rate QoS: since we only consider path loss model here, the user association result at initialisation phase is to connect user with the nearest pico RRH to attain highest data rate, thus consumes least load-dependent power (referring to equations (4.38)(4.40) and (4.50a)). We can observe that all the users prefer pico

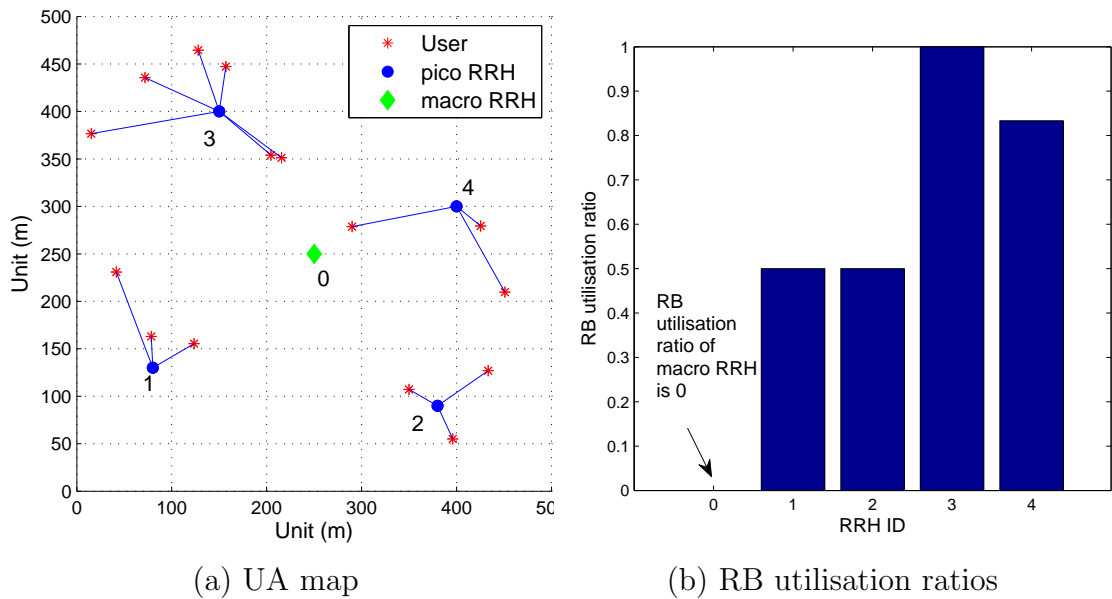


Figure 4.11: UA initial result and RB initial utilisation ratios with 15 users

RRH rather than macro RRH even macro RRH is nearer since the transmit power of macro RRH is much larger than that of pico RRH.

Then we add 1 user in the system to make RRH 3 overloaded. In Fig. 4.12(a), the red stars represent the original users and the magenta cross represents the added user. The solid lines shows the initial association (line 05 in Algorithm 2). All users still tend to connect to the best pico RRHs. The problem is that RRH 3 uses more than 100% RBs, which is shown in Fig. 4.12(b). This violates the radio resource constraint (lines 06-08).

Then the algorithm enters into the drop phase (lines 09-28) to resolve the constraints violation problem. After drop phase, the final user association and RB utilisation ratios are shown in Fig. 4.13(a) and Fig. 4.13(b). From the comparison of initial user association in Fig. 4.12(a) and the final user association in Fig. 4.13(a), the added user originally connected to RRH 3 (the solid line in Fig. 4.12(a)) is changed to connect to the macro RRH (the dotted line in Fig. 4.13(a)).

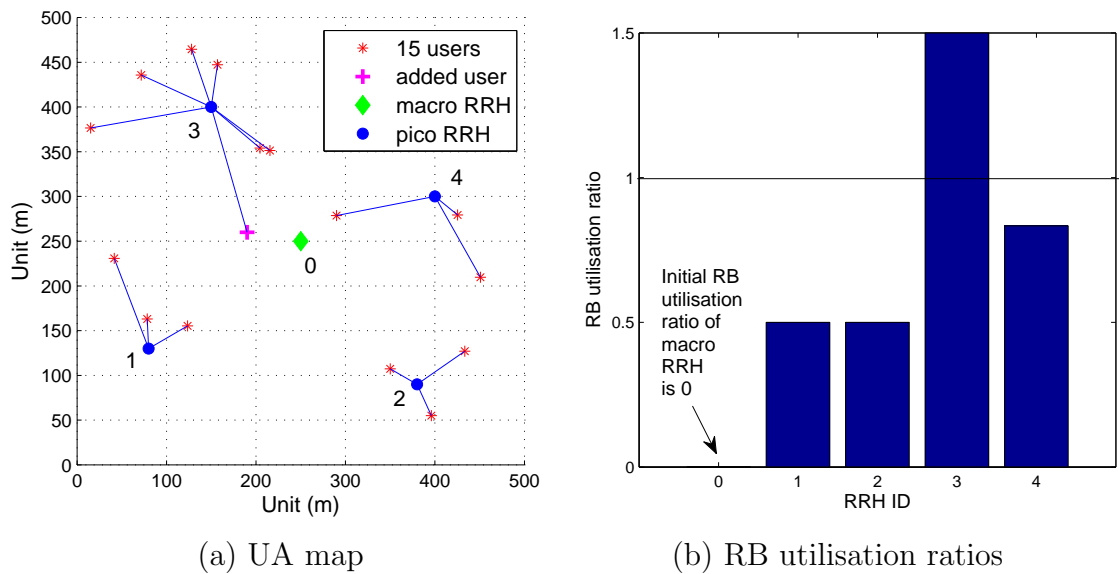


Figure 4.12: UA initial result and RB initial utilisation ratios with 16 users

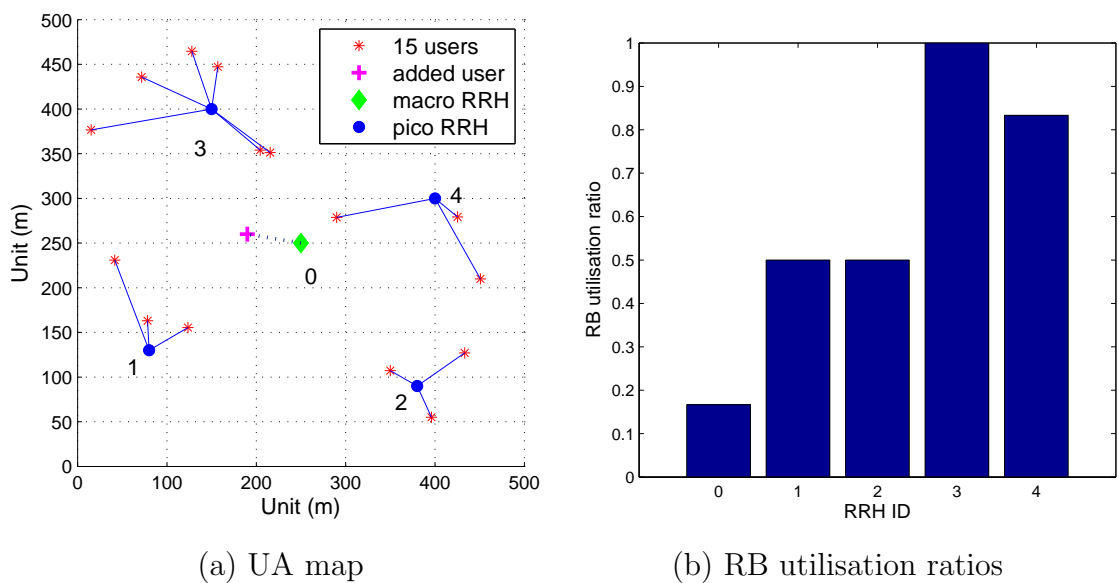


Figure 4.13: UA final result and RB final utilisation ratios with 16 users

By moving the user to the less preferred RRH, the RB utilisation ratios of all RRHs are able to stay no larger than 1, shown in Fig. 4.13(b).

4.4.3.3 Performance Evaluation

To show the effectiveness of the JRSUA, we first compare it against other counterparts using the parameters in Table 4-E. To the best of our knowledge, there is no similar work investigating joint RRH selection and user association in C-RAN scenario, so we compare it with two methods adopted in conventional green cellular networks: SWES [OSK13] and GA [CD14]. Here we set the fronthaul capacity to infinity to eliminate the effect of fronthaul constraint. If not mentioned specifically, all the results are the average values under 100 realizations.

The SWES [OSK13] associates users to the cell which provides the highest signal strength and puts the base stations (BSs) into sleep which will cause the neighbour cells the least load increasing. In [CD14], genetic algorithm (GA) is used to determine the active BS set for reducing overall network energy consumption in OFDMA cellular networks. And the performance of GA is up to the search domain, related to the population and generation number. In our simulation, we set the population size as 50, the generation number as 50, the crossover probability as 0.8 and the mutation probability as 0.05, suggested by [CD14]. To calculate the fitness and handle the constraints, we use

$$F(x) = \begin{cases} f(x) & \text{if } x \text{ is a feasible solution} \\ f_{max} + v(x) & \text{otherwise} \end{cases} \quad (4.62)$$

where $f(x)$ is the objective function, equals to (4.48a). $v(x)$ is the violation function. f_{max} is the objective of the worst feasible solution in the population. Thus the fitness of a feasible solution equals to its objective, while the fitness of an infeasible solution not only depends on the amount of constraint violation, but also on the population of solutions in the current generation. It can be interpreted into

the following criteria to evaluate the fitness of a solution:

- Any feasible solution is preferred to an infeasible solution;
- Among two feasible solutions, the one having better objective function is preferred;
- Among two infeasible solutions, the one having small constraint violation is preferred [Deb00].

In our simulations, SWES scheme is validated against the result of Fig. 5 in [OSK13] and GA scheme is validated against the result of Fig. 4 in [CD14].

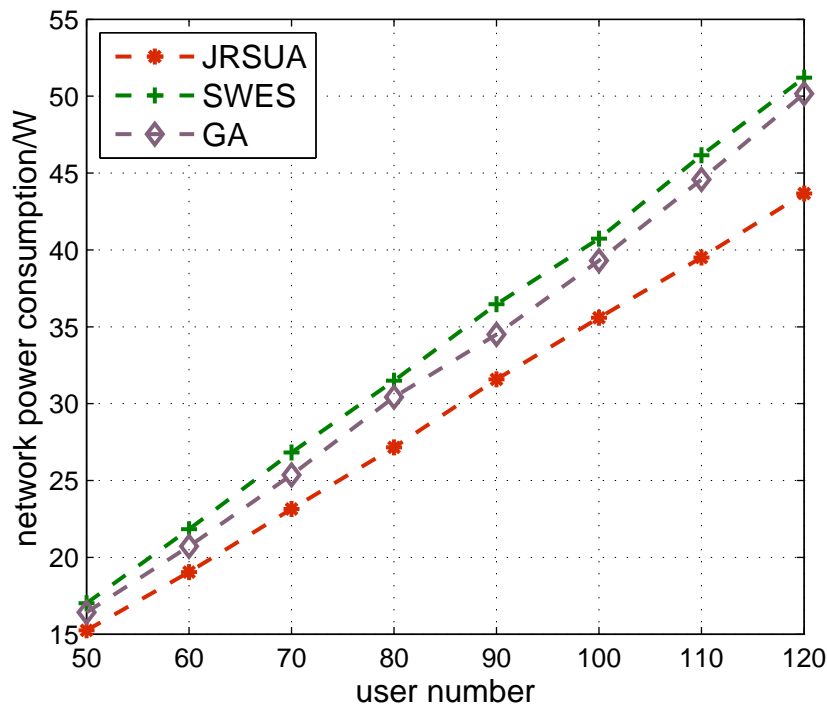


Figure 4.14: Network power consumption comparison among JRSUA, SWES and GA schemes with 12 pico RRHs

The comparison in Fig. 4.14 shows the JRSUA algorithm achieves averagely 10.67% energy saving than GA and 15.25% energy saving than SWES. We also

investigate the performance in a larger network scale: we consider the simulation scenario with one macro RRH and 24 pico RRHs in a 1.6×1.6 km² area [ZHZ⁺15]. The comparison in Fig. 4.15 shows the JRSUA algorithm is able to save even more energy than the other two counterparts with user number range from 100 to 200 since it stands at a global view and the two components (P-UA and P-RS) are coupled properly. However, the user association and BS deactivation in SWES are both based on local view. On average, JRSUA saves 13.81% energy than GA and 18.37% than SWES. The trend also reveals that JRSUA can achieve larger scale of energy saving with heavier traffic load. Less active pico RRHs is required than the other two counterparts which is shown in Fig. 4.16.

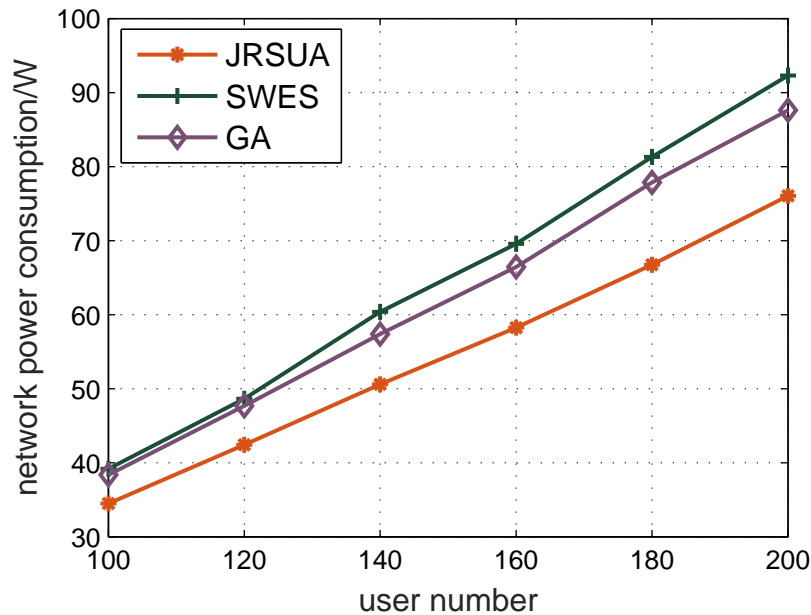


Figure 4.15: Network power consumption comparison among JRSUA, SWES and GA schemes with 24 pico RRHs

The simulation elapsed time of the three algorithms is compared in Table 4-F. The value in the first column is the user number. We use the normalised elapsed time to show the elapsed time difference of the three algorithms: the elapsed time of JRSUA is set as one unit time. From the table we can see, SWES consumes less

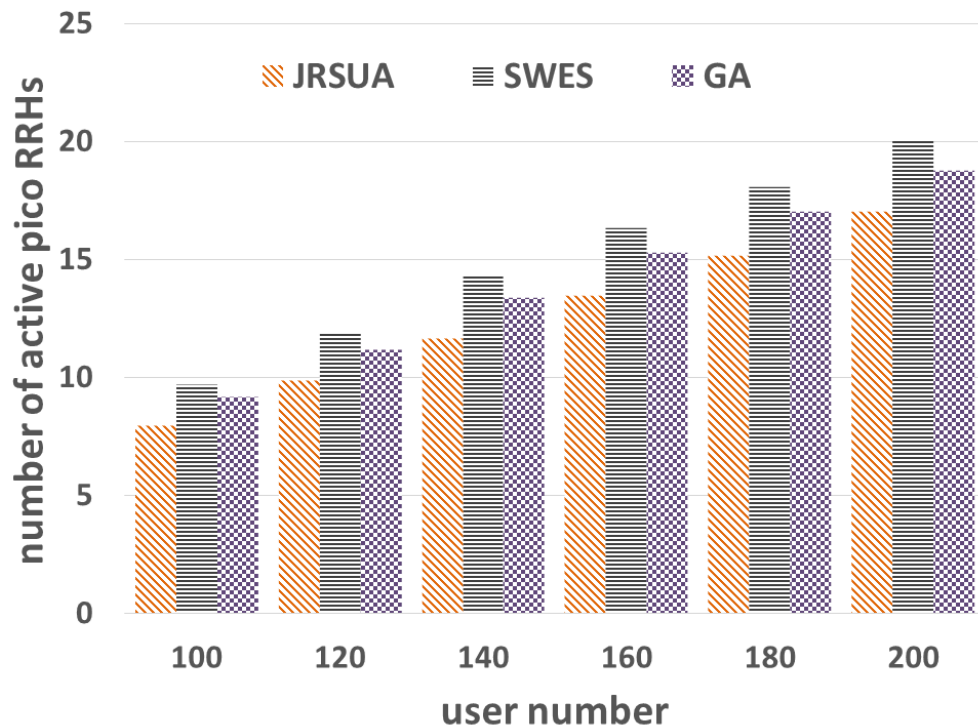


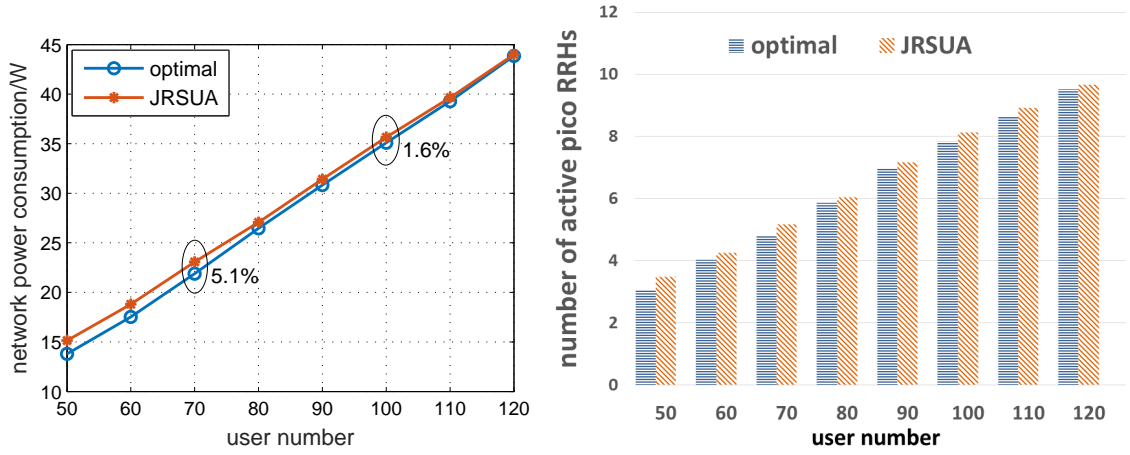
Figure 4.16: Active pico RRHs number comparison among JRSUA, SWES and GA schemes with 24 pico RRHs

time due to the local view. However GA on the other hand consumes several-fold time than JRSUA because of its searching nature. It mainly depends on the values of population size and generation number.

Table 4-F: Simulation elapsed time comparison

User number	JRSUA	SWES	GA
100	1	0.92	7.59
120	1	0.82	7.64
140	1	0.81	9.08
160	1	0.75	9.75
180	1	0.76	11.6
200	1	0.71	13.4

Then we investigate the performance of JRSUA using the network power consumption and the number of active pico RRHs indicators under three fronthaul network scenarios using simulation parameters in Table 4-E.



(a) Network power consumption

(b) Active pico RRHs number

Figure 4.17: Performance comparison between optimal solution and JRSUA algorithm in fibre fronthaul

A. Fibre fronthaul

The results in Fig. 4.17(a) and Fig. 4.17(b) evaluate the proposed JRSUA performance in terms of network power consumption and the number of active pico RRHs against optimal solution from exhaustive search in fibre scenario. The capacity of the fibre fronthaul link is set to 10 Gbps [PSSSS14]. Fig. 4.17(a) confirms the JRSUA is able to perform closely to the optimal solution. However, the complexity of JRSUA is $O(L^2)$ (Algorithm 3), which is greatly reduced from $O(2^L)$ of the exhaustive search solution, where each pico RRH can be either active or sleep, especially when L is large in practical scenes. As the user number increases, it shows the closer trend of performance against the optimal solution. This is because in heavier load situation, Algorithm 3 switches off less pico RRHs which reduces the difference with the optimal solution. We also investigate the number of active pico RRHs needed correspondingly in Fig. 4.17(b). Regardless of the increasing traffic, the number difference of active pico RRHs between JRSUA and optimal solutions is less than 1.

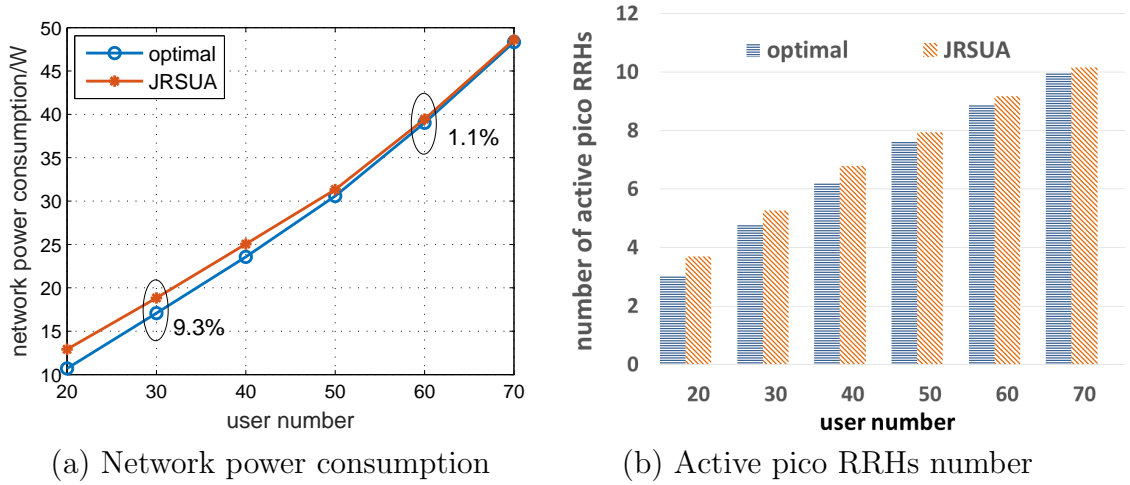


Figure 4.18: Performance comparison between optimal solution and JRSUA algorithm with 5 Mbps wireless fronthaul

B. Wireless fronthaul

With wireless fronthaul links, we investigate two cases where in the first case, all the pico RRHs are assumed to have limited fronthaul capacity as 5 Mbps [HA15]. The network power consumption and number of active pico RRHs comparisons in JRSUA against those in exhaustive search are given in Fig. 4.18(a) and 4.18(b). When doubling the fronthaul capacity limitation to 10 Mbps [DDSK13], the corresponding results are illustrated in Fig. 4.19(a) and 4.19(b). Similarly with the result in fibre scenario, the JRSUA achieves proximity to the optimal solution, especially in heavy loads.

Take a vertical comparison between the results in Fig. 4.18(a) and 4.19(a), we find that within the network power consumption range from 10 W - 50 W and the active pico RRHs number range from three to eleven, the number of served users is different. When the power consumption is within relatively low range, say 10 W - 20 W power consumption and three to five active pico RRHs, the C-RAN with 10 Mbps fronthaul can serve about 20 more users than that with 5 Mbps fronthaul links. However when the network is heavily loaded, the 10 Mbps fronthaul C-RAN

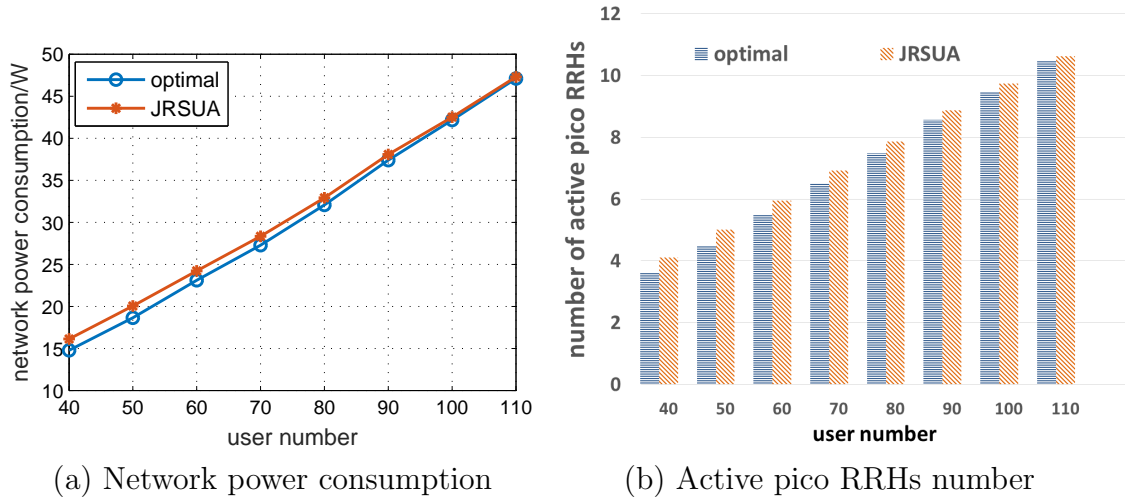


Figure 4.19: Performance comparison between optimal solution and JRSUA algorithm with 10 Mbps wireless fronthaul

can support 40 more users consuming the same level power (e.g., 40 W - 50 W) and using nine to eleven active pico RRHs.

C. Mixed fronthaul

In this section, we investigate a more complicated scenario with mixed fronthaul configuration (Table 4-G). Fig. 4.20 compares the power consumption with three different mixed fronthaul link combinations. The C-RAN can support around 90 users in scenario 1 and 100 users in scenario 3. For comparison reason, the user range in Fig. 4.20 is set from 20 to 80. From the comparison of the JRSUA in three scenarios, we can see that using more fibre fronthaul links consumes less energy since the fibre fronthaul have very large capacity allowing more user associations. And the fronthaul capacity have greater influence in heavy loaded C-RAN.

Fig. 4.21 indicates how many users are associated with the macro RRH and the pico RRHs respectively in mixed-scenario 2. The big gaps imply that users prefer to associate with pico RRHs due to lower dynamic power consumption.

Table 4-G: Mixed fronthaul settings

Scenario	Number of fibre / wireless fronthaul
mixed-scenario 1	6 / 6
mixed-scenario 2	3 / 9
mixed-scenario 3	9 / 3

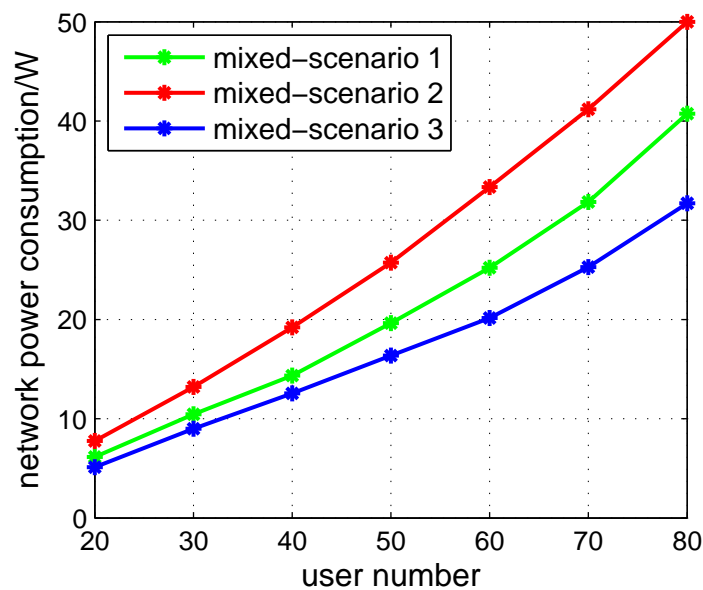


Figure 4.20: Network power consumption comparison among different fronthaul combinations

4.5 Summary

This chapter studies radio resource management schemes in C-RAN architecture. In Section 4.2, a joint power and resource block allocation algorithm is proposed aimed at maximising the user satisfaction. It takes into account diverse QoS requirements and inter-cell interference. Users are firstly divided into satisfied and unsatisfied sets. Then different policies are devised using power-RB strategies to mitigate inter-cell interference and optimise user satisfaction. Through BBU pool, a cloud based information sharing mechanism is used to direct co-channel RB reassignment. Simulation results show that the proposed algorithm achieves

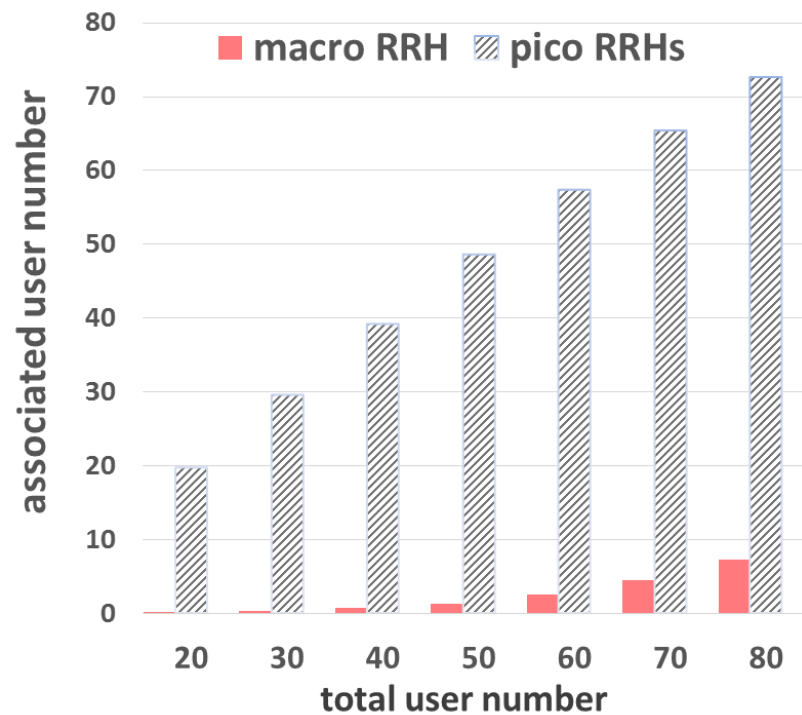


Figure 4.21: User associations with macro and pico RRHs in mixed-scenario 2

higher user satisfaction with QoS requirements.

In Section 4.3, an energy-effective network deployment (EEND) scheme is proposed to optimise the network energy consumption for dense C-RAN. Under multiple constraints, EEND achieves saving energy consumption by selecting only a subset of RRHs without harming traffic demand. A combination of MMKP Lagrange multipliers and sleeping technique is designed in EEND to meet the target. The simulation results illustrate 40% energy reduction compared to GA scheme.

To address the energy efficiency issue in downlink C-RAN, in Section 4.4, by understanding the fronthaul limitation, we have proposed a joint RRH selection and user association (JRSUA) scheme with the goal of network energy minimisation given that all the users within this network are served with required data rate. We first formulate the NPCM problem considering the RRH selection, user

association and fronthaul capacity constraint, which have not been considered all together before. Based on the analysis of the formulated NPCM problem, an iterative JRSUA algorithm including two coupled components P-UA and P-RS is proposed to solve the optimisation problem. From the simulation, we can conclude that the proposed JRSUA algorithm not only outperforms other counterparts with about 15% average network energy saving, but also performs in close proximity within only around 10% difference to the optimal exhaustive search approach in average yet with greatly reduced complexity in various fronthaul scenarios.

Chapter 3 and 4 harness the idea of dynamic network topology to achieve adaptive resource management. Chapter 3 utilises semi-smart antennas at BS level to achieve adaptive cell coverage responding to the load distribution. In Chapter 4, RRH selection is adaptive to the traffic dynamics on the network scale.

Chapter 5

Conclusions and Further Work

5.1 Conclusions

This thesis focuses on QoS-aware radio resource management schemes adaptive to traffic dynamics in different OFDMA multi-cell network scenarios.

By identifying the utility of GBR traffic and the unbalanced traffic among cells, a user satisfaction maximisation resource management scheme is proposed in Chapter 3 working on antenna pattern adaptation and subcarrier allocation jointly. Based on the idea of non-cooperative game, subcarrier is allocated in a distributed way taking into QoS requirements including minimum rate requirement and user/traffic priorities into account. Genetic algorithm is applied to optimise the semi-smart antennas. A dynamic optimisation of QoS for moving users in an OFDMA network with semi-smart antennas is further developed. The traffic distribution is monitored periodically and semi-smart antennas based optimisation is only triggered when the satisfaction improvement drops below a threshold. Simulation results show that the proposed algorithms are able to achieve 10% to 20%

greater user satisfaction compared to fixed antenna gain scenarios.

Benefiting from cloud based information sharing through the BBU pool, a joint resource block (RB) and power allocation scheme is proposed at the beginning of Chapter 4 to maximise the number of satisfied users whose required QoS is provided in multi-cell OFDMA downlink served with high power nodes only. Simulation results show that the proposed joint RB and power allocation algorithm is capable of achieving 20%-30% satisfaction improvement compared to the conventional MaxSINR approach. By utilising network densification, system bandwidth can be reused across a geographic area to reduce the number of users competing the resources at each access node and ensure the QoS provisioning. Two network energy optimisation algorithms with QoS guarantee are proposed for C-RANs in Chapter 4. An energy-effective network deployment (EEND) scheme is devised to cover the traffic demand nodes with appropriate subset of small cell RRHs in dense C-RAN. The proposed network deployment algorithm achieves 40% energy saving compared with genetic algorithm based traffic demand node association scheme and 50% energy saving respect to the “full-on” baseline algorithm. At the end, a joint RRH selection and user association (JRSUA) scheme is proposed with the goal of network power consumption minimisation (NPCM) given that all the users within this network are served with required data rate. Firstly, the NPCM problem is decoupled into user association sub-problem and RRH selection sub-problem. Secondly, we solve the user association for given active RRH set using multiple-choice multidimensional knapsack model with consideration of both radio resource and fronthaul capacity constraints. Thirdly, we devise a low complexity heuristic algorithm that selects the best active RRHs by repeatedly solving the user association problem. The effectiveness of the proposed JRSUA is validated by simulation, which outperforms the other counterparts with 10%-20% network

energy saving and provides near-optimal performance with reduced complexity.

5.2 Lesson and Learn

- The most important thing is to find a research topic which you are interested in. Four years' PhD study is never easy. Especially during the hardships, real interest will be the biggest motivation rather than any deadline, paper or certificate.
- Always do thorough background research before any implementation work. Spending time on background research especially in early PhD stages is never too much.
- Thinking is as important as doing in the research world. Keep your research up to date. You should have a clear mind of every connection from your research to the latest progress in both industry and academia.
- You may have several pieces of works, but make them a complete story in your thesis. Have both horizontal and vertical views on your research. Horizontal means having a comprehensive perspective of every possible related work. Only knowing your work is not enough. Vertical means finding the “hole” from broad literature review as your research interest and digging deep into details.
- Communicate with your supervisor, especially when you are frustrated. Rely on yourself doing the research including figuring out the topic and finding the solutions. Also communicate and discuss with your supervisor on every step. Feeling blue is normal and talk it with your supervisor. They will give

you the strongest support.

- Stay motivated, keep learning. Be curious and insightful. Build your own way to understand the world and solve problems. PhD is a once a lifetime thing and make it worthwhile.

5.3 Future Work

In this section, extensions to current work and some future research directions are proposed.

5.3.1 Extension to Current Work

In this thesis, power allocation is either equal over subcarriers or using a predefined policy. As one possible extension, power allocation can be jointly formulated in the user satisfaction maximisation problem and network power consumption minimisation problem as well.

Since in almost all wireless communications the most important factor to determine a user's satisfaction is its data transmission rate, the main QoS indicator considered in this thesis is the minimum rate requirement. However in future work, cross layer optimisation can be redesigned taking other QoS indicators such as delay bound and loss probability into account with queueing model for real-time flows.

5.3.2 Joint Transmission in C-RAN

This thesis assumes that each user is only associated with one RRH which is simple to implement. In order to moving the networks toward user-centric architectures, joint transmission should be considered which allows users to communicate via several RRHs [DDD⁺15].

5.3.3 Delay Performance in C-RAN

Apart from energy efficient design considering fronthaul capacity constraints, delay performance of C-RANs should also be optimised especially for real-time applications [WLP15]. Latency can be jointly considered with capacity on the fronthaul constraints.

5.3.4 Resilience Mechanism Design in BBU Pool

The centralisation nature of C-RAN imposes higher requirement on the resiliency of cloud BBU hotelling [WTT⁺16] since a single network and/or processing unit failure might produce large disruptions [CMDTM16]. Dedicated virtual link protection like redundant virtual links can be used to deal with connectivity failure. Resilient virtual machine placement, for example backup BBU can survive processing failure [CCMM16].

5.3.5 RRH Selection with Traffic Prediction

In this thesis, RRH selection is based on historical traffic demand information. In certain typical traffic patterns, by using machine learning based prediction tech-

niques [SRK16], we can further optimise long-term network energy consumption considering traffic prediction.

References

- [3GP13] 3GPP. Study of heterogeneous networks management. Technical report, TR 32.835 Release 12, Mar. 2013.
- [ABC⁺14] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Angel Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang. What will 5g be? *IEEE Journal on Selected Areas in Communications*, 32(6):1065–1082, 2014.
- [ABH11] Imran Ashraf, Federico Boccardi, and Lester Ho. Sleep mode techniques for small cell deployments. *IEEE Communications Magazine*, 49(8):72–79, 2011.
- [Acc14] Evolved Universal Terrestrial Radio Access. Study on energy saving enhancement for e-utran. Technical report, 3GPP TR 36.887, 2014.
- [Acc16a] Evolved Universal Terrestrial Radio Access. Further advancements for e-utra physical layer aspects. Technical report, 3GPP TR 36.814, 2016.
- [Acc16b] Evolved Universal Terrestrial Radio Access. Lte physical layer; general description. Technical report, 3GPP TR 36.201, 2016.
- [Acc16c] Evolved Universal Terrestrial Radio Access. User equipment (ue) radio transmission and reception. Technical report, 3GPP TR 36.101, 2016.
- [ACP06] Y. Alfadhil, X. Chen, and C. G. Parini. Design and testing of a cylindrically-conformal array for semi-smart antenna base stations. In *2006 European Conference on Wireless Technology*, pages 306–309, Sept 2006.
- [Adi09] Sasan Adibi. *Fourth-Generation Wireless Networks: Applications and Innovations: Applications and Innovations*. IGI Global, 2009.

- [Adi10] Sasan Adibi. *Quality of Service Architectures for Wireless Networks: Performance Metrics and Management: Performance Metrics and Management*. IGI Global, 2010.
- [AGD⁺11] Gunther Auer, Vito Giannini, Claude Desset, Istvan Godor, Per Skillermark, Magnus Olsson, Muhammad Ali Imran, Dario Sabella, Manuel J Gonzalez, Oliver Blume, et al. How much energy is needed to run a wireless network? *Wireless Communications, IEEE*, 18(5):40–49, 2011.
- [AGM07] Jeffrey G Andrews, Arunabha Ghosh, and Rias Muhamed. *Fundamentals of WiMAX: understanding broadband wireless networking*. Pearson Education, 2007.
- [Ahm13] Sassan Ahmadi. *LTE-Advanced: a practical systems approach to understanding 3GPP LTE releases 10 and 11 radio access technologies*. Academic Press, 2013.
- [AIS⁺14] Patrick Kwadwo Agyapong, Mikio Iwamura, Dirk Staehle, Wolfgang Kiess, and Anass Benjebbour. Design considerations for a 5g network architecture. *IEEE Communications Magazine*, 52(11):65–75, 2014.
- [And13] Jeffrey G Andrews. Seven ways that hetnets are a cellular paradigm shift. *IEEE Communications Magazine*, 51(3):136–144, 2013.
- [ARS16] M. Agiwal, A. Roy, and N. Saxena. Next generation 5g wireless networks: A comprehensive survey. *IEEE Communications Surveys Tutorials*, 18(3):1617–1655, thirdquarter 2016.
- [ASY⁺14] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon. An overview of load balancing in hetnets: old myths and open problems. *IEEE Wireless Communications*, 21(2):18–25, April 2014.
- [ATH13] N. A. Ali, A. E. M. Taha, and H. S. Hassanein. Quality of ser-

- vice in 3gpp r12 lte-advanced. *IEEE Communications Magazine*, 51(8):103–109, August 2013.
- [BB99] Ezio Biglieri and S Benedetto. Principles of digital transmission with wireless applications. 1999.
- [BB10] Mudit Ratana Bhalla and Anand Vardhan Bhalla. Generations of mobile wireless technology: A survey. *International Journal of Computer Applications*, 5(4), 2010.
- [BB15] H. Boostanimehr and V. K. Bhargava. Unified and distributed qos-driven cell association algorithms in heterogeneous networks. *IEEE Transactions on Wireless Communications*, 14(3):1650–1662, March 2015.
- [BD09] S. Bashar and Z. Ding. Admission control and resource allocation in a heterogeneous ofdma wireless network. *IEEE Transactions on Wireless Communications*, 8(8):4200–4210, August 2009.
- [BJN12] Rohit Bodhe, Shirish Joshi, and Satish Narkhede. Performance comparison of fft and dwt based ofdm and selection of mother wavelet for ofdm. *International Journal of Computer Science and Information Technologies*, 3(3):3993–97, 2012.
- [BLM⁺14] Naga Bhushan, Junyi Li, Durga Malladi, Rob Gilmore, Dean Brenner, Aleksandar Damnjanovic, Ravi Teja Sukhavasi, Chirag Patel, and Stefan Geirhofer. Network densification: the dominant theme for wireless evolution into 5g. *IEEE Communications Magazine*, 52(2):82–89, 2014.
- [BN97] Adam M Brandenburger and Barry J Nalebuff. Co-opetition: A revolution mindset that combines competition and cooperation: the game theory strategy that’s changing the game of business, 1997.
- [BRW⁺15] Jens Bartelt, Peter Rost, Dirk Wubben, Johannes Lessmann, Bruno

- Melis, and Gerhard Fettweis. Fronthaul and backhaul requirements of flexibly centralized radio access networks. *IEEE Wireless Communications*, 22(5):105–111, 2015.
- [BS09] Paolo Bellavista and Roberto Saracco. *Telecommunication Systems and Technologies*. Eolss Publishers Company Limited, 2009.
- [BV13] Nadia Brahmi and Venkatkumar Venkatasubramanian. Summary on preliminary trade-off investigations and first set of potential network-level solutions. *Proc. Eur. 7th Framework Res. Project METIS*, 2013.
- [CAK⁺16] Shihabur Rahman Chowdhury, Reaz Ahmed, Md Mashrur Alam Khan, Nashid Shahriar, Raouf Boutaba, Jeebak Mitra, and Feng Zeng. Protecting virtual networks with drone. In *Network Operations and Management Symposium (NOMS), 2016 IEEE/IFIP*, pages 78–86. IEEE, 2016.
- [CCMM16] M. Fiorani C. C. Meixner, G. B. J. M. Figueiredo and B. Mukherjee. Resilient cloud network mapping with virtualized bbu placement for cloud-ran. In *10th IEEE International Conference on Advanced Networks and Telecommunications Systems (IEEE ANTS 2016)*, 2016.
- [CCY⁺15] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann. Cloud ran for mobile networks - a technology overview. *IEEE Communications Surveys Tutorials*, 17(1):405–426, Firstquarter 2015.
- [CD14] Prabhu Chandhar and Suvra Sekhar Das. Energy saving in ofdma cellular networks with multi-objective optimization. In *Communications (ICC), 2014 IEEE International Conference on*, pages 3951–3956. IEEE, 2014.

- [CGF⁺15] Lei Cheng, Yuehong Gao, Jinxi Fu, Xin Zhang, Zaixue Wei, and Dacheng Yang. Energy efficient control for software defined cloud radio access network based on small cell. In *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, pages 1–5. IEEE, 2015.
- [CHK11] Kae Won Choi, Ekram Hossain, and Dong In Kim. Downlink subchannel and power allocation in multi-cell ofdma cognitive radio networks. *IEEE Transactions on Wireless Communications*, 10(7):2259–2271, 2011.
- [Cis13] Visual Networking Index Cisco. Forecast and methodology, 2012–2017. *White Paper*, 2013.
- [CIS16] VNI CISCO. Cisco visual networking index: Forecast and methodology, 2015-2020: Visual networking index (vni), 2016.
- [CJJ09] Kae Won Choi, Wha Sook Jeon, and Dong Geun Jeong. Resource allocation in ofdma wireless communications systems supporting multimedia services. *IEEE/ACM transactions on networking*, 17(3):926–935, 2009.
- [CJL⁺14] Liming Chen, Hu Jin, Haoming Li, Jun-Bae Seo, Qing Guo, and Victor Leung. An energy efficient implementation of c-ran in hetnet. In *Vehicular Technology Conference (VTC Fall), 2014 IEEE 80th*, pages 1–5. IEEE, 2014.
- [CLRH⁺14] I Chih-Lin, Corbett Rowell, Shuangfeng Han, Zhikun Xu, Gang Li, and Zhengang Pan. Toward green and soft: a 5g perspective. *Communications Magazine, IEEE*, 52(2):66–73, 2014.
- [CMDTM16] Carlos Colman-Meixner, Chris Develder, Massimo Tornatore, and Biswanath Mukherjee. A survey on resiliency techniques in cloud computing infrastructures and applications. *IEEE Communications Surveys & Tutorials*, 18(3):2244–2281, 2016.

- [Cox12] Christopher Cox. *An introduction to LTE: LTE, LTE-advanced, SAE and 4G mobile communications*. John Wiley & Sons, 2012.
- [CSP11] Chetan Chudasama, SM Shah, and Mahesh Panchal. Comparison of parents selection methods of genetic algorithm for tsp. In *International Conference on Computer Communication and Networks, published by International Journal of Computer Applications (IJCA)*, 2011.
- [CSSK⁺15] Raúl Chávez-Santiago, Michał Szydełko, Adrian Kliks, Fotis Foukalas, Yoram Haddad, Keith E Nolan, Mark Y Kelly, Moshe T Masonta, and Ilangko Balasingham. 5g: The convergence of wireless communications. *Wireless Personal Communications*, 83(3):1617–1642, 2015.
- [CTS06] CM Chin, CM Tan, and ML Sim. Future trends in radio resource management for wireless communications. *BT Technology Journal*, 24(2):103–110, 2006.
- [CTZK09] Ronald Y Chang, Zhifeng Tao, Jinyun Zhang, and C-C Jay Kuo. Multicell ofdma downlink resource allocation using a graphic framework. *IEEE Transactions on Vehicular Technology*, 58(7):3494–3507, 2009.
- [Cui08] Mingyi Cui. An improved float-coded genetic algorithm based on wavelet denoising mutation. In *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on*, pages 4018–4023, June 2008.
- [Cvi14] Neda Cvijetic. Optical network evolution for 5g mobile applications and sdn-based control. In *Telecommunications Network Strategy and Planning Symposium (Networks), 2014 16th International*, pages 1–5. IEEE, 2014.

- [CZ14] S. Chen and J. Zhao. The requirements, challenges, and technologies for 5g of terrestrial mobile telecommunication. *IEEE Communications Magazine*, 52(5):36–43, May 2014.
- [CZB⁺10] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gdor, G. Auer, and L. V. Der Perre. Challenges and enabling technologies for energy aware mobile radio networks. *IEEE Communications Magazine*, 48(11):66–72, November 2010.
- [DBC04] Lin Du, J Biahm, and Laurie Cuthbert. A bubble oscillation algorithm for distributed geographic load balancing in mobile networks. In *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 1. IEEE, 2004.
- [DDD⁺15] Hayssam Dahrouj, Ahmed Douik, Oussama Dhifallah, Tareq Y Al-Naffouri, and Mohamed-Slim Alouini. Resource allocation in heterogeneous cloud radio access networks: advances and challenges. *IEEE Wireless Communications*, 22(3):66–73, 2015.
- [DDSK13] Antonio De Domenico, Valentin Savin, and Dimitri Ktenas. A backhaul-aware cell selection algorithm for heterogeneous cellular networks. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, pages 1688–1693. IEEE, 2013.
- [Deb99] Kalyanmoy Deb. An introduction to genetic algorithms. *Sadhana*, 24(4-5):293–315, 1999.
- [Deb00] Kalyanmoy Deb. An efficient constraint handling method for genetic algorithms. *Computer methods in applied mechanics and engineering*, 186(2):311–338, 2000.
- [Dew13] Neelam Dewangan. *A detailed Study of 4G in Wireless Commu-*

- nication: Looking insight in issues in OFDM*. Anchor Academic Publishing (aap-verlag), 2013.
- [DGK⁺13] Panagiotis Demestichas, Andreas Georgakopoulos, Dimitrios Karvounas, Kostas Tsagkaris, Vera Stavroulaki, Jianmin Lu, Chunshan Xiong, and Jing Yao. 5g on the horizon: key challenges for the radio-access network. *IEEE Vehicular Technology Magazine*, 8(3):47–53, 2013.
- [DPF12] R. Datta, D. Panaitopol, and G. Fettweis. Analysis of cyclostationary gfdm signal properties in flexible cognitive radio. In *Communications and Information Technologies (ISCIT), 2012 International Symposium on*, pages 663–667, Oct 2012.
- [DRO11] D. Dasalukunte, F. Rusek, and V. Owall. Multicarrier faster-than-nyquist transceivers: Hardware architecture and performance analysis. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 58(4):827–838, April 2011.
- [DY14] B. Dai and W. Yu. Sparse beamforming and user-centric clustering for downlink cloud radio access network. *IEEE Access*, 2:1326–1339, 2014.
- [EI63] Hugh Everett III. Generalized lagrange multiplier method for solving problems of optimum allocation of resources. *Operations research*, 11(3):399–417, 1963.
- [EIE14] Omar A Elgendy, Mahmoud H Ismail, and Khaled Elsayed. Max-min fair resource allocation for lte-advanced relay-enhanced cells. In *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1432–1437. IEEE, 2014.
- [EODB13] R. Estrada, H. Otrok, Z. Dziong, and H. Barada. Joint bs selection and resource allocation model for ofdma macro-femtocell networks

- incorporating mobility. In *2013 International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT)*, pages 42–47, Aug 2013.
- [ETS11] TR ETSI. 103 122:” speech and multimedia transmission quality (stq). *QoS of connections from current technologies to LTE for delay sensitive applications*, Nov. 2011.
- [EYF⁺15] T. En, W. Ye, D. Fei, P. Zhiwen, and Y. Xiaohu. A practical enb off/on based energy saving scheme for real lte networks. In *2015 17th International Conference on Advanced Communication Technology (ICACT)*, pages 12–17, July 2015.
- [FB11] B. Farhang-Boroujeny. Ofdm versus filter bank multicarrier. *IEEE Signal Processing Magazine*, 28(3):92–112, May 2011.
- [FJL⁺13] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li. A survey of energy-efficient wireless communications. *IEEE Communications Surveys Tutorials*, 15(1):167–178, First 2013.
- [FSAA15] S. Fowler, J. Sarfraz, M. M. Abbas, and V. Angelakis. Gaussian semi-markov model based on real video multimedia traffic. In *2015 IEEE International Conference on Communications (ICC)*, pages 6971–6976, June 2015.
- [FT13] Mikael Fallgren and Bogdan Timus. Scenarios, requirements and kpis for 5g mobile and wireless system. *METIS deliverable D*, 1:1, 2013.
- [FWZW10] B. Fan, W. Wu, K. Zheng, and W. Wang. Proportional fair-based joint subcarrier and power allocation in relay-enhanced orthogonal frequency division multiplexing systems. *IET Communications*, 4(10):1143–1152, July 2010.
- [Gar10] Vijay Garg. *Wireless communications & networking*. Morgan

- Kaufmann, 2010.
- [Gol05] Andrea Goldsmith. *Wireless communications*. Cambridge university press, 2005.
- [GTC⁺14] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie. Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks. *IEEE Journal on Selected Areas in Communications*, 32(6):1152–1163, June 2014.
- [GTM⁺16] Xiaohu Ge, Song Tu, Guoqiang Mao, Cheng-Xiang Wang, and Tao Han. 5g ultra-dense cellular networks. *IEEE Wireless Communications*, 23(1):72–79, 2016.
- [GYZ09] Zhangyu Guan, Dongfeng Yuan, and Haixia Zhang. Optimal and fair resource allocation for multiuser wireless multimedia transmissions. *EURASIP Journal on Wireless Communications and Networking*, 2009(1):1, 2009.
- [GZAE10a] 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*, 12(1):30–42, Feb 2010.
- [GZAE10b] Tolga Girici, Chenxi Zhu, Jonathan R Agre, and Anthony Ephremides. Proportional fair scheduling algorithm in ofdma-based wireless systems with qos constraints. *Journal of communications and networks*, 12(1):30–42, 2010.
- [GZAM10] Arunabha Ghosh, Jun Zhang, Jeffrey G Andrews, and Rias Muhamed. *Fundamentals of LTE*. Pearson Education, 2010.
- [GZF10] H. Galeana-Zapien and R. Ferrus. Design and evaluation of a backhaul-aware base station assignment algorithm for ofdma-based

- cellular networks. *IEEE Transactions on Wireless Communications*, 9(10):3226–3237, October 2010.
- [GZNY10] Jie Gong, Sheng Zhou, Zhisheng Niu, and Peng Yang. Traffic-aware base station sleeping in dense cellular networks. In *Quality of Service (IWQoS), 2010 18th International Workshop on*, pages 1–2. IEEE, 2010.
- [H⁺07] Vesa Hasu et al. *Radio resource management in wireless communication: Beamforming, transmission power control, and rate allocation*. Helsinki University of Technology, 2007.
- [HA15] Tao Han and Nirwan Ansari. User association in backhaul constrained small cell networks. In *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*, pages 1637–1642. IEEE, 2015.
- [HAWJ10] Lajos L Hanzo, Yosef Akhtman, Li Wang, and Ming Jiang. *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus non-coherent and cooperative turbo transceivers*, volume 9. John Wiley & Sons, 2010.
- [HG89] JH Holland and DE Goldberg. *Genetic algorithms in search, optimization and machine learning*, 1989.
- [HLD15] Vu Ha, Long Le, and Ngoc-Dung Dao. Coordinated multipoint (comp) transmission design for cloud-rans with limited fronthaul capacity constraints. 2015.
- [HLN13] Ekram Hossain, Long Bao Le, and Dusit Niyato. *Radio resource management in multi-tier cellular wireless networks*. John Wiley & Sons, 2013.
- [HLo16] V. N. Ha, L. B. Le, and N. D. o. Coordinated multipoint transmission design for cloud-rans with limited fronthaul capacity con-

- straints. *IEEE Transactions on Vehicular Technology*, 65(9):7432–7447, Sept 2016.
- [HMS04] Mhand Hifi, Mustapha Michrafy, and Abdelkader Sbihi. Heuristic algorithms for the multiple-choice multidimensional knapsack problem. *Journal of the Operational Research Society*, 55(12):1323–1332, 2004.
- [HRW⁺07] Lei Huang, Mengtian Rong, Lan Wang, Yisheng Xue, and Egon Schulz. Resource allocation for ofdma based relay enhanced cellular networks. In *2007 IEEE 65th Vehicular Technology Conference-VTC2007-Spring*, pages 3160–3164. IEEE, 2007.
- [HSL⁺14] S. Hong, M. Sagong, C. Lim, S. Cho, K. Cheun, and K. Yang. Frequency and quadrature-amplitude modulation for downlink cellular ofdma networks. *IEEE Journal on Selected Areas in Communications*, 32(6):1256–1267, June 2014.
- [HT09] Harri Holma and Antti Toskala. *LTE for UMTS-OFDMA and SC-FDMA based radio access*. John Wiley & Sons, 2009.
- [HZN10] Yao Hua, Qian Zhang, and Zhisheng Niu. Resource allocation in multi-cell ofdma-based relay networks. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.
- [JGL05] Z. Jiang, Y. Ge, and Y. Li. Max-utility wireless resource management for best-effort traffic. *IEEE Transactions on Wireless Communications*, 4(1):100–111, Jan 2005.
- [JMSC15] M. Jo, T. Maksymyuk, B. Strykhalyuk, and C. H. Cho. Device-to-device-based heterogeneous radio access network architecture for mobile cloud computing. *IEEE Wireless Communications*, 22(3):50–58, June 2015.
- [KAAR15] M Khan, RS Alhumaima, and HS Al-Raweshidy. Reducing en-

- ergy consumption by dynamic resource allocation in c-ran. In *Networks and Communications (EuCNC), 2015 European Conference on*, pages 169–174. IEEE, 2015.
- [Kay11] Mustafa Kaya. The effects of two new crossover operators on genetic algorithm performance. *Applied Soft Computing*, 11(1):881–890, 2011.
- [KBCH10a] Nassar Ksairi, Pascal Bianchi, Philippe Ciblat, and Walid Hachem. Resource allocation for downlink cellular ofdma systems part i: optimal allocation. *IEEE Transactions on Signal Processing*, 58(2):720–734, 2010.
- [KBCH10b] Nassar Ksairi, Pascal Bianchi, Philippe Ciblat, and Walid Hachem. Resource allocation for downlink cellular ofdma systems part ii: practical algorithms and optimal reuse factor. *IEEE Transactions on Signal Processing*, 58(2):735–749, 2010.
- [KdVYV12] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam. Distributed alpha-optimal user association and cell load balancing in wireless networks. *IEEE/ACM Transactions on Networking*, 20(1):177–190, Feb 2012.
- [KH05] Hoon Kim and Youngnam Han. A proportional fair scheduling for multicarrier transmission systems. *IEEE Communications Letters*, 9(3):210–212, March 2005.
- [KL05] Wen-Hsing Kuo and Wanjiun Liao. Utility-based optimal resource allocation in wireless networks. In *GLOBECOM '05. IEEE Global Telecommunications Conference, 2005.*, volume 6, pages 5 pp.–3512, Dec 2005.
- [KL06a] H. Kwon and B. G. Lee. Distributed resource allocation through noncooperative game approach in multi-cell ofdma systems. In

- 2006 IEEE International Conference on Communications*, volume 9, pages 4345–4350, June 2006.
- [KL06b] Hojoong Kwon and Byeong Gi Lee. Distributed resource allocation through noncooperative game approach in multi-cell ofdma systems. In *2006 IEEE International Conference on Communications*, volume 9, pages 4345–4350. IEEE, 2006.
- [KLL09] H. Kwon, W. I. Lee, and B. G. Lee. A minimum data-rate guaranteed resource allocation with low signaling overhead in multi-cell ofdma systems. *Journal of Communications and Networks*, 11(1):26–35, Feb 2009.
- [KNY08] Mohammad Katoozian, Keivan Navaie, and Halim Yanikomeroglu. Optimal utility-based resource allocation for ofdm networks with multiple types of traffic. In *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pages 2223–2227. IEEE, 2008.
- [Lag10] Thomas D Lagkas. *Wireless Network Traffic and Quality of Service Support: Trends and Standards: Trends and Standards*. IGI Global, 2010.
- [LCF14] P. R. Li, T. S. Chang, and K. T. Feng. Energy-efficient power allocation for distributed large-scale mimo cloud radio access networks. In *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1856–1861, April 2014.
- [LCS09] Yuan-Bin Lin, Tai-Hsiang Chiu, and Yu T Su. Optimal and near-optimal resource allocation algorithms for ofdma networks. *IEEE Transactions on Wireless Communications*, 8(8):4066–4077, 2009.
- [LCWY12] Shuhui Liu, Yongyu Chang, Guangde Wang, and Dacheng Yang. Distributed resource allocation with inter-cell interference coordination in ofdma uplink. In *Vehicular Technology Conference (VTC*

- Fall*), 2012 *IEEE*, pages 1–5. IEEE, 2012.
- [LCYW12] Yue Liu, Laurie Cuthbert, Xu Yang, and Yapeng Wang. Qos-aware resource allocation for multimedia users in a multi-cell spectrum sharing radio network. In *Proceedings of the 7th ACM Workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Networks*, PM2HW2N '12, pages 45–52. ACM, 2012.
- [Le12] Long Bao Le. Fair resource allocation for device-to-device communications in wireless cellular networks. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 5451–5456. IEEE, 2012.
- [Liu14] Yue Liu. *QoS-aware Radio Resource Management for Spectrum Sharing Radio Networks*. PhD thesis, Queen Mary University of London, 2014.
- [LMS04a] J. W. Lee, R. R. Mazumdar, and N. B. Shroff. Nonconvexity issues for internet rate control with multiclass services: stability and optimality. In *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 1, page 34, March 2004.
- [LMS04b] J-W Lee, Ravi R Mazumdar, and Ness B Shroff. Nonconvexity issues for internet rate control with multiclass services: stability and optimality. In *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 1. IEEE, 2004.
- [LPDCJ15] D. Lopez-Prez, M. Ding, H. Claussen, and A. H. Jafari. Towards 1 gbps/ue in cellular systems: Understanding ultra-dense small cell deployments. *IEEE Communications Surveys Tutorials*, 17(4):2078–2101, Fourthquarter 2015.

- [LPS09] Byeong Gi Lee, Daeyoung Park, and Hanbyul Seo. *Wireless communications resource management*. John Wiley & Sons, 2009.
- [LQZZ11] C. Liu, X. Qin, S. Zhang, and W. Zhou. Proportional-fair downlink resource allocation in ofdma-based relay networks. *Journal of Communications and Networks*, 13(6):633–638, Dec 2011.
- [Luc12] Alcatel Lucent. lightradio network: A new wireless experience. *White paper*, 2012.
- [LW10] KJ Ray Liu and Beibei Wang. *Cognitive radio networking and security: A game-theoretic view*. Cambridge University Press, 2010.
- [LZ16] L. Liu and R. Zhang. Downlink sinr balancing in c-ran under limited fronthaul capacity. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3506–3510, March 2016.
- [LZL15] S. Luo, R. Zhang, and T. J. Lim. Downlink and uplink energy minimization through user association and beamforming in c-ran. *IEEE Transactions on Wireless Communications*, 14(1):494–508, Jan 2015.
- [Mar15] M. J. Marcus. 5g and "imt for 2020 and beyond" [spectrum policy and regulatory issues]. *IEEE Wireless Communications*, 22(4):2–3, August 2015.
- [Mel13] Abdelhamid Mellouk. *End-to-End Quality of Service Mechanisms in Next Generation Heterogeneous Networks*. John Wiley & Sons, 2013.
- [Mit98] Melanie Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.
- [MJC10] T. F. Maciel, W. C. Freitas Jr., and F. R. P. Cavalcanti. Joint opportunistic beamforming and subcarrier assignment for maximiza-

- tion of user satisfaction in ofdma systems. In *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, pages 1–5, Sept 2010.
- [MJS97] Martin Moser, Dusan P Jokanovic, and Norio Shiratori. An algorithm for the multidimensional multiple-choice knapsack problem. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 80(3):582–589, 1997.
- [Mob11] China Mobile. C-ran: the road towards green ran. *White Paper, ver, 2*, 2011.
- [Nak11] T Nakamura. Feasibility study for further advancement for e-utra (lte-advanced). *3rd Gen. Partner. Proj., Sophia Antipolis, France, TS36*, 912:V9, 2011.
- [Net11] Nokia Siemens Networks. Liquid radio: Let traffic waves flow most efficiently. *White paper*, 2011.
- [NPP+03] P Nahi, CG Parini, S Papadopoulos, Lin Du, John Biggam, and Laurie Cuthbert. A semi-smart antenna concept using real-time synthesis for use in a distributed load balancing scheme for cellular networks. In *Antennas and Propagation, 2003.(ICAP 2003). Twelfth International Conference on (Conf. Publ. No. 491)*, volume 1, pages 168–171. IET, 2003.
- [NSSL11] Cho Yiu Ng, Kenneth W Shum, Chi Wan Sung, and Tat Ming Lok. A layered decomposition framework for resource allocation in multiuser communications. *IEEE Transactions on Vehicular Technology*, 60(2):729–733, 2011.
- [NT16] 3GPP Technical Specification Group Core Network and Terminals. Policy and charging control signalling flows and quality of service (qos) parameter mapping. Technical report, TS 29.213 Release 14,

June 2016.

- [OOaDTH05] Y. Otani, S. Ohno, K. a. Donny Teo, and T. Hinamoto. Subcarrier allocation for multi-user ofdm system. In *2005 Asia-Pacific Conference on Communications*, pages 1073–1077, Oct 2005.
- [OSK13] Eunsung Oh, Kyuho Son, and Bhuma Krishnamachari. Dynamic base station switching-on/off strategies for green cellular networks. *Wireless Communications, IEEE Transactions on*, 12(5):2126–2136, May 2013.
- [PHH⁺12] Miikka Poikselkä, Harri Holma, Jukka Hongisto, Juha Kallio, and Antti Toskala. *Voice over LTE (VoLTE)*. John Wiley & Sons, 2012.
- [PKBV11] Kandaraj Piamrat, Adlen Ksentini, Jean-Marie Bonnin, and César Viho. Radio resource management in emerging heterogeneous wireless networks. *Computer Communications*, 34(9):1066–1076, 2011.
- [PLJ⁺14] Mugen Peng, Yuan Li, Jiamo Jiang, Jian Li, and Chonggang Wang. Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies. *Wireless Communications, IEEE*, 21(6):126–135, 2014.
- [PLZW15] Mugen Peng, Yong Li, Zhongyuan Zhao, and Chonggang Wang. System architecture and key technologies for 5g heterogeneous cloud radio access networks. *Network, IEEE*, 29(2):6–14, 2015.
- [Pra14] Ramjee Prasad. *5G: 2020 and Beyond*. River Publishers, 2014.
- [PSBD11] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das. Understanding traffic dynamics in cellular data networks. In *INFOCOM, 2011 Proceedings IEEE*, pages 882–890, April 2011.
- [PSSSS14] Seok-Hwan Park, Osvaldo Simeone, Ozge Sahin, and Shlomo Shamai Shitz. Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory. *Signal*

- Processing Magazine, IEEE*, 31(6):69–79, 2014.
- [PvdS07] Hyunggon Park and Mihaela van der Schaar. Fairness strategies for multi-user multimedia applications in competitive environments using the kalai-smorodinsky bargaining solution. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 2, pages II–713. IEEE, 2007.
- [PvdS10] H. Park and M. van der Schaar. Fairness strategies for wireless resource allocation among autonomous multimedia users. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(2):297–309, Feb 2010.
- [PWL15] Mugen Peng, Chonggang Wang, Vincent Lau, and H Vincent Poor. Fronthaul-constrained cloud radio access networks: insights and challenges. *Wireless Communications, IEEE*, 22(2):152–160, 2015.
- [PZJ⁺15] Mugen Peng, Kecheng Zhang, Jiamo Jiang, Jiaheng Wang, and Wenbo Wang. Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks. *Vehicular Technology, IEEE Transactions on*, 64(11):5275–5287, 2015.
- [Qua15] Qualcomm. 5g waveform and multiple access techniques, 2015.
- [RC00a] W. Rhee and J. M. Cioffi. Increase in capacity of multiuser ofdm system using dynamic subchannel allocation. In *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, volume 2, pages 1085–1089 vol.2, 2000.
- [RC00b] Wonjong Rhee and John M Cioffi. Increase in capacity of multiuser ofdm system using dynamic subchannel allocation. In *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, volume 2, pages 1085–1089. IEEE, 2000.
- [RC14] Emanuel B Rodrigues and Fernando Casadevall. Fairness-based

- dynamic sub-carrier assignment for real time services in ofdma networks. In *European Wireless 2014; 20th European Wireless Conference; Proceedings of*, pages 1–6. VDE, 2014.
- [REBTH12] M. H. Rasmy, M. A. El-Beltagy, A. A. Tharwat, and A. F. Heikal. A comparative study on the performance of genetic algorithm, artificial immune system and hybrid intelligent approach to multiple-choice multidimensional knapsack problem. In *Informatics and Systems (INFOS), 2012 8th International Conference on*, pages BIO–20–BIO–26, May 2012.
- [RT14] P. Ren and M. Tao. A decentralized sleep mechanism in heterogeneous cellular networks with qos constraints. *IEEE Wireless Communications Letters*, 3(5):509–512, Oct 2014.
- [SA06] A AlQahtani Salman and S Mahmoud Ashraf. Radio resource management issues for 3g and beyond cellular wireless networks. In *GCC Conference (GCC), 2006 IEEE*, pages 1–5. IEEE, 2006.
- [SAR09] Sanam Sadr, Alagan Anpalagan, and Kaamran Raahemifar. Radio resource allocation algorithms for the downlink of multiuser ofdm communication systems. *IEEE communications surveys & tutorials*, 11(3):92–106, 2009.
- [SAR⁺10] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroğlu, D. Falconer, and Y. D. Kim. Fairness-aware radio resource management in downlink ofdma cellular relay networks. *IEEE Transactions on Wireless Communications*, 9(5):1628–1639, May 2010.
- [SAS⁺16] K. Sundaresan, M. Y. Arslan, S. Singh, S. Rangarajan, and S. V. Krishnamurthy. Fluidnet: A flexible cloud-based radio access network for small cells. *IEEE/ACM Transactions on Networking*, 24(2):915–928, April 2016.

- [SBL14] Farshad Shams, Giacomo Bacci, and Marco Luise. A survey on resource allocation techniques in ofdm (a) networks. *Computer Networks*, 65:129–150, 2014.
- [SCH13] Richard Southwell, Xu Chen, and Jianwei Huang. Qos satisfaction games for spectrum sharing. In *INFOCOM, 2013 Proceedings IEEE*, pages 570–574. IEEE, 2013.
- [SCSY14] Yang Sun, Yongyu Chang, Sida Song, and Dacheng Yang. An energy-efficiency aware sleeping strategy for dense multi-tier het-nets. In *2014 IEEE Globecom Workshops (GC Wkshps)*, pages 1180–1185. IEEE, 2014.
- [SDYZ10] Alireza Sharifian, Petar Djukic, Halim Yanikomeroglu, and Jietao Zhang. Max-min fair resource allocation for multiuser amplify-and-forward relay networks. In *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, pages 1–5. IEEE, 2010.
- [SGA14] A. Sahin, I. Guvenc, and H. Arslan. A survey on multicarrier communications: Prototype filters, lattice structures, and implementation aspects. *IEEE Communications Surveys Tutorials*, 16(3):1312–1338, Third 2014.
- [SHBBJ13] Tim Szigeti, Christina Hattingh, Robert Barton, and Kenneth Briley Jr. *End-to-End QoS Network Design: Quality of Service for Rich-Media & Cloud Networks*. Cisco Press, 2013.
- [She95] S. Shenker. Fundamental design issues for the future internet. *IEEE Journal on Selected Areas in Communications*, 13(7):1176–1188, Sept 1995.
- [SKYK11] K. Son, H. Kim, Y. Yi, and B. Krishnamachari. Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks. *IEEE Journal on Selected Areas in Com-*

- munications*, 29(8):1525–1536, September 2011.
- [SL05] Guocong Song and Ye Li. Cross-layer optimization for ofdm wireless networks-part ii: algorithm development. *IEEE Transactions on Wireless Communications*, 4(2):625–634, March 2005.
- [SLFC07] R. B. Santos, F. R. M. Lima, W. C. Freitas, and F. R. P. Cavalcanti. Qos based radio resource allocation and scheduling with different user data rate requirements for ofdma systems. In *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–5, Sept 2007.
- [SPH15] Tshiamo Sigwele, Prashant Pillai, and Yim Fun Hu. itree: Intelligent traffic and resource elastic energy scheme for cloud-ran. In *Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on*, pages 282–288. IEEE, 2015.
- [SQKS13] Y. S. Soh, T. Q. S. Quek, M. Kountouris, and H. Shin. Energy efficient heterogeneous cellular networks. *IEEE Journal on Selected Areas in Communications*, 31(5):840–850, May 2013.
- [SRK16] N. Saxena, A. Roy, and H. Kim. Traffic-aware cloud ran: A key for green 5g networks. *IEEE Journal on Selected Areas in Communications*, 34(4):1010–1021, April 2016.
- [STB15] Stefania Sesia, Issam Toufik, and Matthew Baker. *LTE-the UMTS long term evolution*. Wiley Online Library, 2015.
- [SVDK06] JP Shim, Upkar Varshney, Sasha Dekleva, and Geoffrey Knoerzer. Mobile and wireless networks: services, evolution and issues. *International Journal of Mobile Communications*, 4(4):405–417, 2006.
- [SZL14] Y. Shi, J. Zhang, and K. B. Letaief. Group sparse beamforming for green cloud-ran. *IEEE Transactions on Wireless Communications*, 13(5):2809–2823, May 2014.

- [SZL⁺15] Yuanming Shi, Jun Zhang, Khaled B Letaief, Bo Bai, and Wei Chen. Large-scale convex optimization for ultra-dense cloud-ran. *Wireless Communications, IEEE*, 22(3):84–91, 2015.
- [TAH11] Abd-Elhamid M Taha, Najah Abu Ali, and Hossam S Hassanein. *LTE, LTE-advanced and WiMAX: Towards IMT-advanced Networks*. John Wiley & Sons, 2011.
- [TCYR13] T. O. Ting, Su Fong Chien, Xin-She Yang, and N. Ramli. Resource allocation schemes in energy efficient ofdma system via genetic algorithm. In *2013 19th Asia-Pacific Conference on Communications (APCC)*, pages 723–727, Aug 2013.
- [TDA11] Hina Tabassum, Zaher Dawy, and Mohamed Slim Alouini. Resource allocation via sum-rate maximization in the uplink of multi-cell ofdma networks. *Wireless Communications and Mobile Computing*, 11(12):1528–1539, 2011.
- [TGW⁺14] John Thompson, Xiaohu Ge, Hsiao-Chun Wu, Ralf Irmer, Hong Jiang, Gerhard Fettweis, and Siavash Alamouti. 5g wireless communication systems: prospects and challenges [guest editorial]. *IEEE Communications Magazine*, 52(2):62–64, 2014.
- [TK04] Issam Toufik and Raymond Knopp. Channel allocation algorithms for multi-carrier systems. In *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*, volume 2, pages 1129–1133. IEEE, 2004.
- [TMW⁺11] S. Tombaz, P. Monti, K. Wang, A. Vastberg, M. Forzati, and J. Zander. Impact of backhauling power consumption on the deployment of heterogeneous mobile networks. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1–5, Dec 2011.
- [TRV06] Nishith D Tripathi, Jeffrey H Reed, and Hugh F VanLandingham.

- Radio resource management in cellular systems*, volume 618. Springer Science & Business Media, 2006.
- [TSR⁺15] G. K. Tran, H. Shimodaira, R. E. Rezagah, K. Sakaguchi, and K. Araki. Dynamic cell activation and user association for green 5g heterogeneous cellular networks. In *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2015 IEEE 26th Annual International Symposium on*, pages 2364–2368, Aug 2015.
- [TZJ11] O. Tipmongkolsilp, S. Zaghoul, and A. Jukan. The evolution of cellular backhaul technologies: Current issues and future trends. *IEEE Communications Surveys Tutorials*, 13(1):97–113, First 2011.
- [WBJW08] J. Wu, J. Bigham, P. Jiang, and Y. Wang. Intelligent control of cellular network coverage using semi-smart antennas for load balancing. In *Signal Processing, Communications and Networking, 2008. ICSCN '08. International Conference on*, pages 295–301, Jan 2008.
- [WHY14] R. Wang, H. Hu, and X. Yang. Potentials and challenges of c-ran supporting multi-rats toward 5g mobile networks. *IEEE Access*, 2:1187–1195, 2014.
- [WLP15] W. Wang, V. K. N. Lau, and M. Peng. Delay-optimal fronthaul allocation via perturbation analysis in cloud radio access networks. In *2015 IEEE International Conference on Communications (ICC)*, pages 3999–4004, June 2015.
- [WN03] Dapeng Wu and R. Negi. Effective capacity: a wireless link model for support of quality of service. *IEEE Transactions on Wireless Communications*, 2(4):630–643, July 2003.
- [WP514] ITUR WP5D. Work plan, timeline, process and deliverables for. *IMT-2020 development*, Feb, 2014.
- [WTN14] Meng Wang, Hui Tian, and Gaofeng Nie. Energy efficient power

- and subchannel allocation in dense ofdma small cell networks. In *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, pages 1–5. IEEE, 2014.
- [WTT⁺16] X. Wang, S. Thota, M. Tornatore, H. S. Chung, H. H. Lee, S. Park, and B. Mukherjee. Energy-efficient virtual base station formation in optical-access-enabled cloud-ran. *IEEE Journal on Selected Areas in Communications*, 34(5):1130–1139, May 2016.
- [WWSZ10] B. Wang, X. Wen, D. Su, and W. Zheng. User satisfaction based resource allocation for ofdma relay networks in the resource-constrained system. In *Future Networks, 2010. ICFN '10. Second International Conference on*, pages 304–308, Jan 2010.
- [WYMC09] Yapeng Wang, Xu Yang, A. Ma, and L. Cuthbert. Intelligent resource optimisation using semi-smart antennas in lte ofdma systems. In *Communications Technology and Applications, 2009. ICCTA '09. IEEE International Conference on*, pages 173–179, Oct 2009.
- [WZS13] Ping Wang, Xi Zhang, and Mei Song. Optimal stochastic subcarrier and power allocations for qos-guaranteed services in ofdma multicell cooperation networks. In *2013 IEEE International Conference on Communications (ICC)*, pages 6449–6453. IEEE, 2013.
- [WZZ14] Kaiwei Wang, Ming Zhao, and Wuyang Zhou. Graph-based dynamic frequency reuse in cloud-ran. In *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, pages 105–110. IEEE, 2014.
- [WZZY15] J. Wu, Y. Zhang, M. Zukerman, and E. K. N. Yung. Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey. *IEEE Communications Surveys Tutorials*, 17(2):803–826,

- Secondquarter 2015.
- [XLmT14] L. Xu, Y. Li, and Z. m. Tang. Hybrid-genetic-algorithm-based resource allocation for slow adaptive ofdma system under channel uncertainty. *Electronics Letters*, 50(1):30–32, January 2014.
- [XLZ⁺12] Cong Xiong, Geoffrey Ye Li, Shunqing Zhang, Yan Chen, and Shugong Xu. Energy-efficient resource allocation in ofdma networks. *IEEE Transactions on Communications*, 60(12):3767–3778, 2012.
- [Xue00] Tong Xuejian. *The principle and application of OFDM mobile communication technology*. People’s Posts and Telecommunications Press, 2000.
- [XWTS13] XiaoDong Xu, Da Wang, XiaoFeng Tao, and Tommy Svensson. Resource pooling for frameless network architecture with adaptive resource allocation. *Science China Information Sciences*, 56(2):1–12, 2013.
- [XZZL15] H. Xu, T. Zhang, Z. Zeng, and D. Liu. Joint base station operation and user association in cloud based hcns with hybrid energy sources. In *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2015 IEEE 26th Annual International Symposium on*, pages 2369–2373, Aug 2015.
- [YWZC10] Xu Yang, Yapeng Wang, Dapeng Zhang, and Laurie Cuthbert. Resource allocation in lte ofdma systems using genetic algorithm and semi-smart antennas. In *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, pages 1–6. IEEE, 2010.
- [ZHZ⁺15] L. Zhou, X. Hu, C. Zhu, E. C. H. Ngai, S. Wang, J. Wei, and V. C. M. Leung. Green small cell planning in smart cities under dynamic traffic demand. In *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages

- 618–623, April 2015.
- [ZKAQ01] Jens Zander, Seong-Lyun Kim, Magnus Almgren, and Olav Que-
seth. *Radio resource management for wireless networks*. Artech
House, Inc., 2001.
- [ZL11] Qing Zhang and Guizhong Liu. Rate allocation games in multiuser
multimedia communications. *IET communications*, 5(3):396–407,
2011.
- [ZL14] Yimin Zhang and Guizhong Liu. Proportional fair resource al-
location algorithm for video transmission in ofdma relay system.
In *Wireless Communications and Signal Processing (WCSP), 2014
Sixth International Conference on*, pages 1–6. IEEE, 2014.
- [ZRB14] A. Zakrzewska, S. Ruepp, and M. S. Berger. Towards converged
5g mobile networks-challenges and current trends. In *ITU Kaleido-
scope Academic Conference: Living in a converged world - Impossi-
ble without standards?*, *Proceedings of the 2014*, pages 39–45, June
2014.
- [ZTH⁺10] S. Zheng, H. Tian, Z. Hu, L. Chen, and J. Zhu. Qos-guaranteed
radio resource allocation with distributed inter-cell interference co-
ordination for multi-cell ofdma systems. In *Vehicular Technology
Conference (VTC 2010-Spring), 2010 IEEE 71st*, pages 1–5, May
2010.
- [ZW13] Huiling Zhu and Jiangzhou Wang. Resource allocation in ofdma-
based distributed antenna systems. In *2013 IEEE/CIC Interna-
tional Conference on Communications in China (ICCC)*, pages 565–
570. IEEE, 2013.
- [ZWWW14a] W. Zhao, S. Wang, C. Wang, and X. Wu. Cell planning for het-
erogeneous networks: An approximation algorithm. In *IEEE IN-*

- FOCOM 2014 - IEEE Conference on Computer Communications*, pages 1087–1095, April 2014.
- [ZWWW14b] Wentao Zhao, Shaowei Wang, Chonggang Wang, and Xiaobing Wu. Cell planning for heterogeneous networks: An approximation algorithm. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 1087–1095. IEEE, 2014.
- [ZXH⁺10] Y. Zhao, X. Xu, Z. Hao, X. Tao, and P. Zhang. Resource allocation in multiuser ofdm system based on ant colony optimization. In *2010 IEEE Wireless Communication and Networking Conference*, pages 1–6, April 2010.
- [ZZZX12] Yutao Zhu, Zhimin Zeng, T. Zhang, and L. Xiao. Transmission mode selection for energy saving in the comp system with semi-smart antenna. In *Communications and Information Technologies (ISCIT), 2012 International Symposium on*, pages 855–860, Oct 2012.