ENERGY EFFICIENT AND SECURE WIRELESS COMMUNICATION SYSTEMS

by

Wing Kwan Ng

M. Phil., The Hong Kong University of Science and Technology, 2008
B. Eng., The Hong Kong University of Science and Technology, 2006

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

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

July 2012

Abstract

Orthogonal frequency division multiple access (OFDMA), multiple–input multiple–output (MIMO), and base station (BS) cooperation are the core techniques for the next generation wireless communication systems. As the communication systems evolve, both service providers and users are demanding not only high data rates, but also energy efficiency and data security. As a result, it is necessary to design novel resource allocation algorithms that meet these new needs. This dissertation develops four resource allocation algorithms which are tailored for different design goals and communication environments.

For systems employing the combination of OFDMA and decode-and-forward (D-F) relaying technologies, we propose a novel resource allocation algorithm for secure communication. The proposed algorithm takes into account artificial noise generation to combat a passive multiple antenna eavesdropper and the effects of imperfect channel state information at the transmitter (CSIT).

Subsequently, we investigate the energy efficiency of OFDMA systems which also provide communication security. We formulate the resource allocation algorithm design as a non-convex optimization problem. By exploiting the properties of fractional programming, the considered non-convex optimization problem is transformed to an equivalent convex optimization problem with a tractable solution, which can be obtained with an iterative algorithm.

Thirdly, we study resource allocation for energy efficient communication in OFD-

MA downlink networks with a large number of transmit antennas. Our proposed resource allocation algorithm takes into account the circuit power consumption, imperfect CSIT, a minimum data rate requirement, and a maximum tolerable channel outage probability.

Lastly, we propose a resource allocation algorithm for energy efficient communication in OFDMA downlink networks with cooperative BSs. The resource allocation algorithm design problem is formulated as a non-convex optimization problem which takes into account the circuit power consumption, the limited backhaul capacity, and the minimum required data rate for joint BS zero-forcing beamforming (ZFBF) transmission. By using the concept of perturbation function, we show that the duality gap in the considered system is always zero under some general conditions, despite the non-convexity of the primal problem. Thus, an efficient closed-form power allocation solution for maximization of the energy efficiency of data transmission is derived.

Preface

Chapters 2–5 are based on works under the supervision of Professor Robert Schober and collaboration with Dr. Ernest S. Lo, the chief representative of the Centre Tecnològic de Telecomunicacions de Catalunya - Hong Kong (CTTC-HK).

For all chapters, I conducted the paper survey on related topics, formulated the problems, proposed problem solutions, and performed the analysis and the simulations of the considered communication systems. I also wrote all paper drafts. Note that I use the name "Derrick W. K. Ng" for all international publications.

Four papers related to Chapter 2 have been published:

- Derrick W. K. Ng, Ernest S. Lo, and R. Schober, "Secure Resource Allocation and Scheduling for OFDMA Decode-and-Forward Relay," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 3528-3540, Oct. 2011.

- Derrick W. K. Ng, Ernest S. Lo, and R. Schober, "Resource Allocation for Secure OFDMA Networks with Imperfect CSIT," accepted in *Proc. IEEE Global Telecommun. Conf. (Globecom)*, Houston, Texas, Dec. 2011.

- Derrick W. K. Ng and R. Schober, "Resource Allocation for Secure OFDMA Decode-and-Forward Relaying Networks," in *Proc. 2011 Canadian Workshop on Inform. Theory (CWIT 2011)*, Kelowna, BC, Canada, pp. 202-205, May 2011.

- Derrick W. K. Ng and R. Schober, "Resource Allocation for Secure OFDMA

Communication Systems," in *Proc. 2011 Australian Commun. Theory Workshop (AusCTW)*, Melbourne, Australia, pp. 13-18, Feb. 2011.

One paper related to Chapter 3 has been accepted:

- Derrick W. K. Ng, Ernest S. Lo, and R. Schober, "Energy-Efficient Resource Allocation for Secure OFDMA Systems", accepted in *IEEE Trans. on Veh. Technol.*, May 2012.

Three papers related to Chapter 4 have been accepted:

- Derrick W. K. Ng, Ernest S. Lo, and R. Schober, "Energy-Efficient Resource Allocation in OFDMA Systems with Large Numbers of Base Station Antennas", accepted in *IEEE Trans. Wireless Commun.*, Jun. 2012.

- Derrick W. K. Ng, Ernest S. Lo, and R. Schober, "Energy-Efficient Resource Allocation in SDMA Systems with Unlimited Numbers of Base Station Antennas," accepted in *IEEE Intern. Commun. Conf. (ICC) 2012*, Ottawa, Canada.

- Derrick W. K. Ng and R. Schober, "Energy-Efficient Resource Allocation in OFDMA Systems with Large Numbers of Base Station Antennas," accepted in *IEEE Intern. Commun. Conf. (ICC) 2012 Workshop on Green Commun. and Networking*, Ottawa, Canada.

Two papers related to Chapter 5 have been submitted:

- Derrick W. K. Ng, Ernest S. Lo, and R. Schober, "Energy-Efficient Resource Allocation in Multi-Cell OFDMA Systems with Limited Backhaul Capacity," submitted, Nov. 2011.

- Derrick W. K. Ng, Ernest S. Lo, and R. Schober, "Energy-Efficient Resource Allocation in Multi-Cell OFDMA Systems with Limited Backhaul," accepted in *IEEE Wireless Commun. and Networking Conf. (WCNC) 2012*, Paris, France.

Table of Contents

Abstract ii

Preface iv

Table of Contents vii

List of Tables xiii

List of Figures xiv

List of Abbreviations xx

Notation xxiii

Acknowledgments xxv

Dedication xxvi

1 Introduction 1

1.1 Resource Allocation and Scheduling for Wireless Communication Systems 2

1.2 Resource Allocation Optimization Framework 3

1.2.1 Utility Function-Based Optimization 3

1.2.2 QoS Measure 5

1.2.3 Resource Allocation in OFDMA Systems 6
1.3 Cooperative Communication - Virtual MIMO 6
1.4 Energy Efficient Communication 10
1.5 PHY Layer Security 13
1.6 Contributions of the Thesis 14
1.7 Organization of the Thesis 16

2 Secure Resource Allocation and Scheduling for OFDMA DF Relay Networks 20
2.1 Introduction 20
2.2 OFDMA Downlink Network Model 22
2.2.1 Channel Model 22
2.2.2 Channel State Information 24
2.2.3 Artificial Noise Generation 26
2.3 Resource Allocation and Scheduling 28
2.3.1 Instantaneous Channel Capacity, Channel Outage, and Secrecy Outage 28
2.3.2 Optimization Problem Formulation 31
2.4 Solution of the Optimization Problem 33
2.4.1 Transformation of the Optimization Problem 33
2.4.2 Dual Problem Formulation 37
2.4.3 Dual Decomposition and Sub-Problem Solution 38
2.4.4 Solution of the Master Problem 41
2.5 Results 42
2.5.1 Convergence of Distributed Iterative Algorithm 43

2.5.2 Average Secrecy Outage Capacity versus Transmit Power and ESR . . . 45
2.5.3 Average Secrecy Outage Capacity versus N_E . . . 48
2.5.4 Average Secrecy Outage Capacity versus Number of Users . . 49
2.6 Conclusions . . . 51

3 Energy-Efficient Resource Allocation for Secure OFDMA Systems 52
3.1 Introduction . . . 52
3.2 OFDMA Downlink Network Model . . . 53
3.2.1 Channel Model . . . 53
3.3 Resource Allocation and Scheduling . . . 56
3.3.1 Instantaneous Channel Capacity, Secrecy Outage, and Energy Efficiency . . . 56
3.3.2 Optimization Problem Formulation . . . 60
3.4 Solution of the Optimization Problem . . . 60
3.4.1 Transformation of the Objective Function . . . 61
3.4.2 Iterative Algorithm for Energy Efficiency Maximization . . . 62
3.5 Results . . . 69
3.5.1 Convergence of Iterative Algorithm . . . 70
3.5.2 Energy Efficiency and Average Secrecy Outage Capacity versus Transmit Power . . . 71
3.5.3 Energy Efficiency and Secrecy Outage Capacity versus Number of Users . . . 76
3.5.4 Energy Efficiency and Average Secrecy Outage Capacity versus N_E . . . 78
3.6 Conclusions . . . 80

4 EE Resource Allocation in OFDMA Systems with Large Numbers of BS Antennas 81
4.1 Introduction 81
4.2 OFDMA Downlink Network Model 83
4.2.1 Channel Model 83
4.2.2 Channel State Information 85
4.3 Resource Allocation 86
4.3.1 Instantaneous Channel Capacity and Outage Capacity 86
4.3.2 Optimization Problem Formulation 88
4.4 Solution of the Optimization Problem 89
4.4.1 Problem Transformation 90
4.4.2 Iterative Algorithm for Energy Efficiency Maximization . . . 90
4.5 Results 100
4.5.1 Energy Efficiency versus Maximum Inter-user Interference Temperature I 102
4.5.2 Convergence of Iterative Algorithm 104
4.5.3 Energy Efficiency and Average Outage Capacity versus Transmit Power 105
4.5.4 Energy Efficiency versus Number of Users 109
4.6 Conclusions 111

5 EE Resource Allocation in Multi-Cell OFDMA Systems with Limited Backhaul Capacity 112
5.1 Introduction 112
5.2 Multi-Cell OFDMA Network Model 114
5.2.1 Multi-Cell System Model and Central Unit 114

5.2.2 OFDMA Channel Model 116
5.2.3 Backhaul Model 117
5.3 Resource Allocation and Scheduling 117
5.3.1 Instantaneous Channel Capacity 117
5.3.2 Optimization Problem Formulation 119
5.4 Solution of the Optimization Problem 120
5.4.1 Transformation of the Objective Function 121
5.4.2 Iterative Algorithm for Energy Efficiency Maximization . . . 122
5.5 Results and Discussions 130
5.5.1 Convergence of Iterative Algorithm 5.1 and Duality Gap . . . 131
5.5.2 Energy Efficiency and Average Capacity versus Transmit Power 133
5.5.3 Energy Efficiency and Average System Capacity versus Number of Users 137
5.6 Conclusions 140

6 Summary of Thesis and Future Research Topics 141
6.1 Summary of Results 141
6.2 Future Work 143
6.2.1 Secure Communications in OFDMA Systems with an Active Eavesdropper 144
6.2.2 Energy Efficiency: Optimal Locations for Distributed Antennas 145
6.2.3 Energy Harvesting in Energy Efficient OFDM Systems . . . 145

Bibliography 147

Appendices

A Mathematical Preliminaries 160

A.1 Convex Analysis 160

A.1.1 Definitions and Basic Properties 160

A.1.2 Optimization Problem and Perturbation Function 161

B Proof of Lemma 2.1 164

C Proof of Theorem 3.1 and Algorithm Convergence 169

D Proof of Proposition 3.1 173

E Proof of Proposition 4.1 175

F Proof of the Concavity of the Transformed Problem in (4.14) . . 179

G Proof of Theorem 5.1 181

List of Tables

3.1 Iterative Resource Allocation Algorithm. 63

4.1 Iterative Resource Allocation Algorithm. 91

4.2 Coordinate Ascent Method. 100

4.3 System parameters 101

5.1 Iterative Resource Allocation Algorithm. 122

5.2 Semi-Orthogonal User Selection Algorithm. 124

List of Figures

1.1 Timing diagram for resource allocation and scheduling in an FDD system. 4

1.2 Block diagram of a resource allocator in an OFDMA system. $S_{i,k}$ and $r_{i,k}$ are the subcarrier allocation variable and data rate allocation variable for user $k \in \{1, \ldots, K\}$ on subcarrier $i \in \{1, \ldots, n_F\}$, respectively. $H_{i,k}$ is the instantaneous channel gain between the BS and user k on subcarrier i. 7

1.3 An example of subcarrier allocation in an OFDMA system. 7

1.4 A point-to-point MIMO communication system. 8

1.5 The upper half of the figure represents a one-way DF half-duplex relaying system and the lower half illustrates a system with BS cooperation. 9

1.6 An illustration of energy efficiency (EE) versus signal-to-noise ratio (SNR) in Rayleigh fading channel. $P_C = 10$ dB for the system with circuit power consumption consideration. 12

1.7 A general wire-tap channel 14

2.1 An illustration of a downlink OFDMA DF relay network. There are one BS and $M = 3$ DF relays with $N_T = 4$ antennas, $K = 10$ desired users equipped with a single antenna, and one eavesdropper with $N_E = 2$ antennas. For an effective eavesdropping, the eavesdropper chooses a location closer to either the BS or a relay than all the desired users. 22

2.2 An illustration of the relationship between packet data rate $R_{m,k}^{data}[i]$, secrecy data rate $R_{m,k}^{sec}[i]$, the capacity of the user channel, $C_{m,k}[i]$, and the capacity of the eavesdropper channel, $C_{m,E}[i]$, for four possible cases. 30

2.3 A flow chart of the proposed iterative distributed resource allocation and scheduling algorithm. 42

2.4 Lagrange multiplier λ versus number of iterations with $K = 15$ users and $M = 3$ relays for different transmit power levels. The BS and each relay are equipped with $N_T = 9$ antennas. There are $N_E = 2$ receive antennas at the eavesdropper. 44

2.5 Average secrecy outage capacity versus transmit power for different numbers of transmit antennas N_T. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS. 45

2.6 Average secrecy outage capacity versus ESR $\frac{\sigma_e^2}{\sigma_{h_{R_{m,k}}}^2}$ for different numbers of transmit antennas N_T and eavesdropper antennas N_E. 47

2.7 Average secrecy outage capacity versus the number of antennas N_E employed at the eavesdropper for different ESR $\frac{\sigma_e^2}{\sigma_{h_{R_{m,k}}}^2}$ and different secrecy outage requirements δ. $N_T = 9$ antennas at the BS and relays. 48

2.8 Average secrecy outage capacity versus the number of desired users for different numbers of transmit antennas N_T at the BS with a total transmit power $P_T = 43$ dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m away from the BS. The double arrows indicate the performance gain achieved by an increasing number of transmit antennas N_T. 50

3.1 Illustration of an OFDMA downlink network. There are one BS with $N_T = 4$ antennas, $K = 9$ desired users equipped with a single antenna, and one eavesdropper with $N_E = 2$ antennas. For an effective eavesdropping, the eavesdropper chooses a location closer to the BS compared to the locations of all the desired users. 54

3.2 Energy efficiency (bit-per-milli Joule) versus the number of iterations with $K = 15$ users for different numbers of transmit antennas at the BS. The maximum transmit power at the BS is P_t =43 dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS. 71

3.3 Energy efficiency (bit-per-milli Joule) versus maximum transmit power, P_t, for different numbers of transmit antennas N_T. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS. 72

3.4 Average secrecy outage capacity versus maximum transmit power, P_t, for different numbers of transmit antennas N_T. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS. . . 74

3.5 Average total power consumption versus maximum transmit power, P_t, for different numbers of transmit antennas N_T. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS. . 75

3.6 Energy efficiency (bit-per-milli Joule) versus the number of users K for different numbers of transmit antennas N_T and a maximum transmit power of $P_t = 22$ dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS. 76

3.7 Average secrecy outage capacity versus the number of users K for different numbers of transmit antennas N_T and a maximum transmit power of $P_t = 22$ dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS. 77

3.8 Energy efficiency (bit-per-milli Joule) versus the number of antennas at the eavesdropper for different static circuit powers, P_C, and different values of δ for a maximum transmit power of $P_t = 43$ dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS. 79

3.9 Average secrecy outage capacity versus the number of antennas at the eavesdropper for different static circuit powers, P_C, and different values of δ for a maximum transmit power of $P_t = 43$ dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS. . 79

4.1 Illustration of an OFDMA downlink network. There are one BS with a large number of antennas and $K = 9$ desired users equipped with a single antenna. 84

4.2 The normalized performance of the proposed algorithm versus the maximum interference-temperature-to-noise ratio $\frac{I}{N_0 W}$ for different values of P_T and different numbers of users. The y-axis is normalized by the performance of the optimal algorithm. 102

4.3 Energy efficiency versus the number of iterations with $K = 15$ users for different maximum transmit powers P_T and channel estimation error variance $\sigma_e^2 = 0.1$. 104

4.4 Energy efficiency versus maximum transmit power, P_T, for different resource allocation algorithms with channel estimation error variance $\sigma_e^2 = 0.1$. The minimum required number of antennas is $N_{th} = 24$. . . 106

4.5 Average outage capacity (bit/s/Hz) versus maximum transmit power, P_T, for different resource allocation algorithms, channel estimation error variance $\sigma_e^2 = 0.1$, and $K = 15$ users. 108

4.6 Average total power consumption, $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})\}$, versus maximum transmit power, P_T, for different resource allocation algorithms, channel estimation error variance $\sigma_e^2 = 0.1$, 10 iterations, and $K = 15$ users. 109

4.7 Energy efficiency (bit-per-Joule) versus the number of users K for different resource allocation algorithms, different channel estimation error variances σ_e^2, and a maximum transmit power of $P_T = 46$ dBm. . . . 110

5.1 A multi-cell system with $M = 3$ cells with a fully connected backhaul link topology. There are in total $K = 27$ users in the system. Each transceiver is equipped with a single antenna. 115

5.2 Illustration of the dual decomposition of a large problem into a two-layer problem. 129

5.3 Energy efficiency (bit-per-Joule) versus number of iterations with different maximum transmit power allowances per BS, P_T, and different numbers of users K. The dashed lines represent the maximum achievable energy efficiencies for different cases. 132

5.4 Duality gap versus the maximum transmit power allowance at each BS, P_T, for different backhaul capacities $R_{\max_m}$. 133

5.5 Energy efficiency (bit-per-Joule) versus the maximum transmit power allowance at each BS, P_T, for different resource allocation algorithms and different backhaul capacities with $K = 45$ users. 135

5.6 Average system capacity (bit/s/Hz/cell) versus the maximum transmit power allowance at each BS, P_T, for different resource allocation algorithms and different backhaul capacities with $K = 45$ users. . . . 136

5.7 Average total power consumption, $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{W}, \mathcal{S})\}$, versus the maximum transmit power allowance at each BS, P_T, for different backhaul capacities, 10 iterations, and $K = 45$ users. 138

5.8 Energy efficiency (bit-per-Joule) versus the number of users K for different maximum transmit power allowances at each BS, P_T, and different backhaul capacities. 139

5.9 Average system capacity (bit/s/Hz/cell) versus the number of users K for different maximum transmit power allowances at each BS, P_T, and different backhaul capacities. 139

G.1 Geometric interpretation of duality and perturbation function for concave and non-convex optimization problem with a 1-dimensional perturbation vector $y \in \mathbb{R}^1$ for illustration. The shaded areas represent the set of values of the primal problem under different perturbations $y \in \mathbb{R}^1$. 181

List of Abbreviations

3GPP	3rd Generation Partnership Project
AF	Amplify–and–Forward
ASIC	Application-Specific Integrated Circuit
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BF	Beamforming
BS	Base Station
CCDF	Complementary Cumulative Distribution Function
CDF	Cumulative Distribution Function
CO_2	Carbon Dioxide
CSI	Channel State Information
CSIR	Channel State Information at Receiver
CSIT	Channel State Information at Transmitter
DF	Decode–and–Forward
DPC	Dirty Paper Coding
DSL	Digital Subscriber Line
EARTH	Energy Aware Radio and Network Technologies
EE	Energy Efficiency
ESR	Estimation Error to Signal Ratio
FDD	Frequency Division Duplex

FER	Frame Error Rate
ICT	Information and Communication Technology
i.i.d.	Independent and Identically Distributed
KKT	Karush Kuhn Tucker
LTE	(3GPP) Long Term Evolution
LTE-A	(3GPP) Long Term Evolution - Advanced
MAC	Medium Access Control
MIMO	Multiple–Input Multiple–Output
MMSE	Minimum Mean Square Error
MRC	Maximum Ratio Combining
MRT	Maximum Ratio Transmission
MSE	Mean Square Error
MUD	Multi-User Diversity
NP–hard	Non–Deterministic Polynomial–Time Hard
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PHY	Physical
QoS	Quality of Service
QSI	Queue State Information
RF	Radio Frequency
SIC	Successive Interference Cancellation
SINR	Signal–to–Interference–plus–Noise Ratio
SIR	Signal–to–Interference Ratio
SNR	Signal–to–Noise Ratio
TDD	Time Division Duplex

WDM	Wavelength Division Multiplexing
WiMAX	Worldwide Interoperability for Microwave Access
WRANs	Wireless Regional Area Networks
ZFBF	Zero Forcing Beamforming

Notation

$(\cdot)^T$	Transpose
$(\cdot)^H$	Hermitian transpose
$(\cdot)^*$	Complex conjugate
$\mathbf{0}$	All–zero column vector
$\mathbb{Z}^+$	The set of positive integer
$\mathbb{R}$	The set of real number
$\mathcal{E}_X\{\cdot\}$	Statistical expectation with respect to random variable X
$\det(\cdot)$	Matrix determinant
$\big[\,\cdot\,\big]_{a,b}$	The element in row a and column b of a matrix
$\big[\,\cdot\,\big]^+$	$\max\{0, x\}, \quad x \in \mathbb{R}$
$\Re\{\cdot\}$	Real part of a complex number
$\Pr\{\cdot\}$	Probability of an event
$\mathbb{C}^{N\times M}$	The space of all $N \times M$ matrices with complex entries
$\lvert\,\cdot\,\rvert$	Absolute value of a complex number
$\lVert\,\cdot\,\rVert^2$	Euclidean norm
$1(\cdot)$	An indicator function which is 1 when the event is true and 0 otherwise
$\mathrm{trace}(\cdot)$	Trace of a matrix
$\mathcal{CN}(\mu, \sigma^2)$	A complex Gaussian random variable with mean μ and variance σ^2
$\mathcal{O}(g(x))$	An asymptotic upper bound, i.e., $f(x) = \mathcal{O}(g(x))$ if $\lim_{x\to\infty} \lvert\frac{f(x)}{g(x)}\rvert \leq N$ for $0 < N < \infty$

$\Theta(g(x))$	An asymptotically tight bound. i.e., $f(x) = \Theta(g(x))$ if $\lim_{x\to\infty} c\lvert g(x)\rvert \leq \lim_{x\to\infty} \lvert f(x)\rvert \leq \lim_{x\to\infty} d\lvert g(x)\rvert$ for some constants $c \leq d$

Acknowledgments

First and foremost I would like to express my deep and sincere gratitude to my advisor, Professor Robert Schober, for his support and invaluable advice during my Ph.D. study. As a distinguished professor and excellent teacher, Robert sets an example of being a great researcher. The knowledge and attitude I learned from him benefit me forever. I am much indebted for his patience and encouragement over the years. Without his support and guidance, this thesis would not be possible.

Also, I greatly thank the members of my doctoral committee, Dr. Vijay Bhargava, Dr. Vikram Krishnamurthy, Dr. Lutz Lampe, and Dr. Vincent Wong for the time and effort in evaluating my work and providing valuable feedback and suggestions. Many thanks go to all the people who helped me to settle down when I first came to Canada. Last but not least, I would like to say thank you to Miss Yue Cindy Zhu. I truly cherish our friendship and in June 2012, it was the most memorable time during my Ph.D. journey.

Dedication

To My Family

Chapter 1

Introduction

Energy efficiency, high data rates, and security are the main driving forces for the evolution of wireless communication systems. Traditionally, these requirements have been fulfilled by increasing both the transmit power and the bandwidth. However, nowadays the radio spectrum available for wireless services is extremely scarce and universal frequency reuse is a new trend to accommodate the increasing number of users. In other words, increasing the transmission bandwidth will not always be an option in the future. On the other hand, power consumption in cellular networks is not only a financial burden to the service providers, but also one of the main sources of greenhouse gas emission. Besides, due to the universal frequency reuse, strong co-channel interference puts system designers in a dilemma since increasing the transmit power may not be beneficial to the overall system performance. As a consequence, a better system design for utilizing the limited resources is needed. Orthogonal frequency multiple access (OFDMA), multiple-input multiple-output (MIMO), and base station (BS) cooperation are considered viable solutions to achieve the aforementioned objectives [1]-[10]. Hence, we will provide a brief overview of these techniques in this chapter.

This chapter is organized as follows. In Sections 1.1 and 1.2, we briefly review resource allocation and corresponding optimization frameworks for wireless communication systems, respectively. In Sections 1.3 and 1.4, we introduce the concept of cooperative communication and energy efficiency, respectively. In Section 1.5, we

provide a brief review of physical (PHY) layer security. The contributions made in this thesis are summarized in Section 1.6, and the thesis organization is provided in Section 1.7.

1.1 Resource Allocation and Scheduling for Wireless Communication Systems

Resource allocation is the concept of making the best use of limited resources based on the available information. The four main resources that are limited in the physical layer are: power, bandwidth, time, and space (if multiple antennas are used). For facilitating the resource allocation process, channel state information (CSI) and/or queue state information (QSI) are needed at the transmitter(s) [11, 12]. There are different forms of CSI, which include instantaneous perfect CSI, instantaneous imperfect CSI, and statistical CSI. The CSI in multiuser systems can be interpreted as an indicator which is used to differentiate between users based to their channel qualities. In practice, the CSI is usually obtained via feedback from users in frequency division duplex (FDD) systems or via uplink pilot measurements in time division duplex (TDD) systems. In any case, some implicit costs are incurred in collecting the CSI. On the other hand, the QSI indicates the availability of data and the delay requirements of each user. In each scheduling slot, the resource allocator updates the resource allocation policy[1] to provide some quality of service (QoS) guarantees to the users at the least possible cost. In fact, a QoS requirement usually acts as a constraint in the resource allocation process which generally decreases the degrees of freedom in resource allocation. For instance, a minimum data rate requirement for a particular

[1]The validity of each resource allocation policy depends on the coherence time of the considered channel which is related to the carrier frequency and the speed the of user(s).

user will result in a performance degradation of the overall system. This is because the resource allocator is forced to allocate more resources to that user for satisfying the data rate requirement, regardless of his/her possibly poor channel quality.

1.2 Resource Allocation Optimization Framework

In this section, we introduce some classic problem formulations used in the literature for resource allocation. A typical optimization problem usually consists of a utility function used as an objective function, a set of constraints, and a set of optimization variables. The utility function quantifies an abstract concept and provides a tangible performance metric, i.e., it maps the satisfaction of user(s) into a real number. On the other hand, the set of constraints is used to confine a feasible solution set. In general, the constraints are defined according to some physical limitations or basic QoS requirements in the system. We will discuss the details in the following subsections.

1.2.1 Utility Function-Based Optimization

In each scheduling slot, the resource allocator selects the users for the next transmission frame and determines their power and rate allocation, cf. Figure 1.1, based on the information available at the scheduler such as the CSI [13, 14, 15]. Ideally, a resource allocator should exploit both the information from the physical (PHY) layer and that from the layers above the medium access control (MAC) layer in order to achieve the best possible performance. In the following, we briefly discuss two utility functions commonly used for resource allocation purpose.

Maximum system throughput resource allocation: In most wireless applications, the aggregate data rate of users is the most important figure of merit for evaluation of the system performance from the service provider's point of view

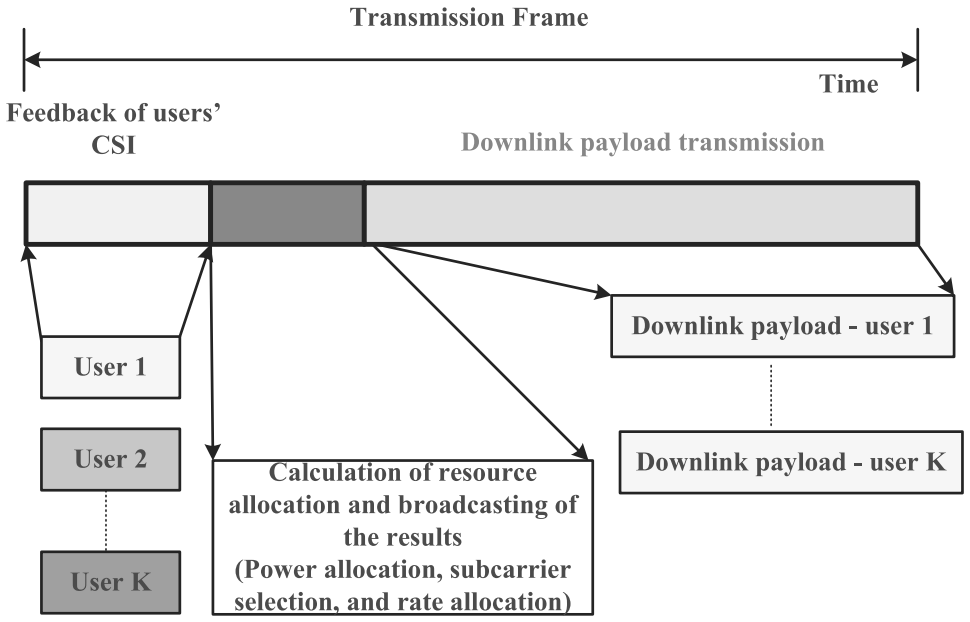


Figure 1.1: Timing diagram for resource allocation and scheduling in an FDD system.

[16, 17]. Considering a system with K users, the corresponding utility function can be expressed as

$$\mathcal{U}_{\text{M-Thp}}(R_1, \ldots, R_K) = \sum_{k=1}^{K} R_k, \tag{1.1}$$

where R_k is the instantaneous data rate of user $k \in \{1, \ldots, K\}$ in the current time slot. It is often assumed that data rate R_k is given by Shannon's channel capacity equation due to its monotonicity and physical meaning.

Weighted aggregate throughput scheduler : Although the maximum throughput scheduler results in the optimal utilization of the system resources from a system throughput point of view, it does not take into account *fairness* in the resource allocation process. Users with poor channel conditions may suffer from starvation since they are rarely selected for transmission which is undesirable from the users' point of view. Therefore, the weighted throughput scheduler [18, 19, 20] was proposed to

resolve the fairness issue. The corresponding utility function can be expressed as

$$\mathcal{U}_{\text{W-Thp}}(R_1, \ldots, R_K) = \sum_{k=1}^{K} w_k R_k, \tag{1.2}$$

where $w_k \geq 0$ are positive constants provided by the upper layers, which allow the resource allocator to give different priorities to different users and to enforce certain notions of fairness.

1.2.2 QoS Measure

The increasing demand for high data rate wireless network services imposes great challenges on resource allocation optimization since operators are required to satisfy diverse QoS requirements for different applications. Different QoS measures have to be incorporated in the problem formulation in order to overcome these challenges. While many different QoS measures have been considered in the literature, we discuss here a few important ones which will be used in the next chapters.

Minimum data rate requirement: The behaviour of the resource allocator depends highly on the objective function. In some situation, such as under the framework of aggregate transmitted power minimization, a low data rate may be the best option from a mathematical point of view. However, such a data rate may be undesirable in practical systems since the basic required services cannot be guaranteed. As a result, a minimum data rate requirement is usually imposed in the problem formulation, whenever there is a trade-off between the objective function and the system throughput [21, 22, 23].

Frame error rate (FER) and outage probability: At the PHY layer, the bit error rate (BER) is usually considered as performance measure. However, the FER is more relevant from the MAC layer perspective, especially if retransmission

is taken into consideration. In general, the FER is hard to calculate analytically and typically results in complicated expressions which are not useful for resource allocation and scheduling design. However, in slow fading channels, if strong forward error correction codes are used for error protection (such as e.g. turbo codes or low density parity check codes), the outage probability is a good approximation for the FER [24, 25]. This connection between the outage probability and the FER can be exploited to arrive at a simple resource allocation algorithm.

1.2.3 Resource Allocation in OFDMA Systems

In this section, we review an advanced wireless communication technology - OFDMA. OFDMA is already employed in several high speed wireless communication network standards including IEEE 802.22 Wireless Regional Area Networks (WRANs), IEEE 802.16 Worldwide Interoperability for Microwave Access (WiMAX), and Long Term Evolution (LTE). In OFDMA, a wide-band frequency spectrum is divided into many orthogonal narrowband subcarriers [26, 27] and data streams from different users are multiplexed on different subcarriers, cf. Figures 1.2 and 1.3, according to some utility function such as (1.1) and (1.2). On the other hand, in an OFDMA system, the fading coefficients of different subcarriers are likely independent for different users. This phenomenon is known as multiuser diversity (MUD) [28]. As a result, maximum system spectral efficiency can be achieved by selecting the best user for each subcarrier and adapting the corresponding transmit power.

1.3 Cooperative Communication - Virtual MIMO

The degrees of freedom introduced by multiple antennas at the transmitters and receivers of wireless communication systems facilitate multiplexing gains and diversity

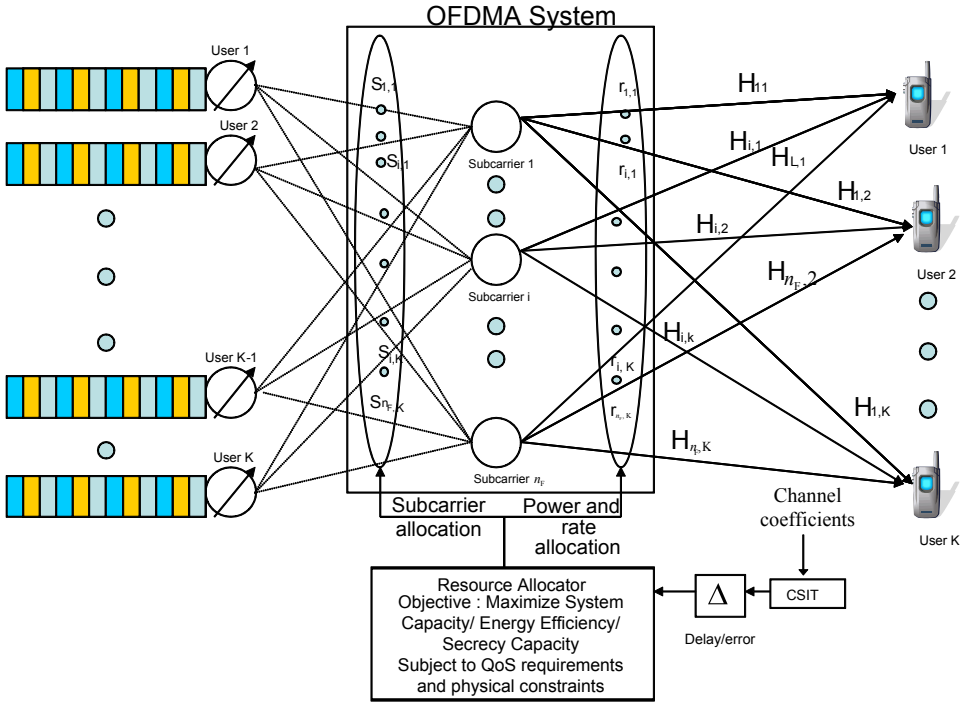


Figure 1.2: Block diagram of a resource allocator in an OFDMA system. $S_{i,k}$ and $r_{i,k}$ are the subcarrier allocation variable and data rate allocation variable for user $k \in \{1, \ldots, K\}$ on subcarrier $i \in \{1, \ldots, n_F\}$, respectively. $H_{i,k}$ is the instantaneous channel gain between the BS and user k on subcarrier i.

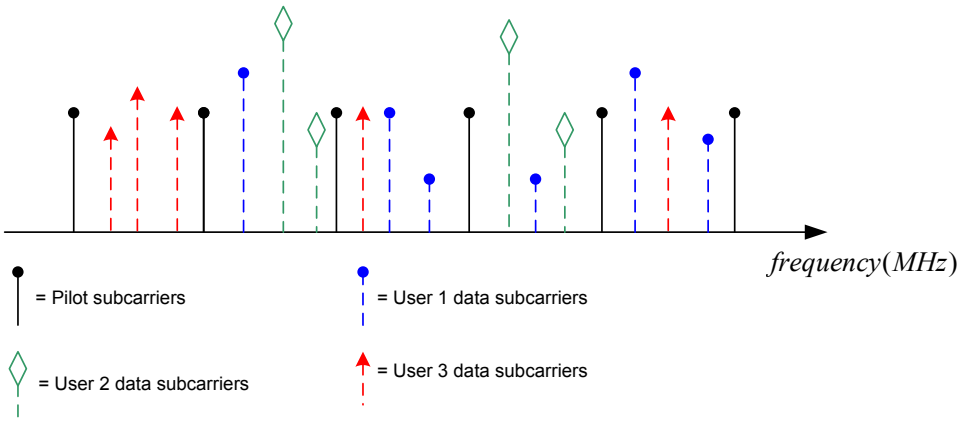


Figure 1.3: An example of subcarrier allocation in an OFDMA system.

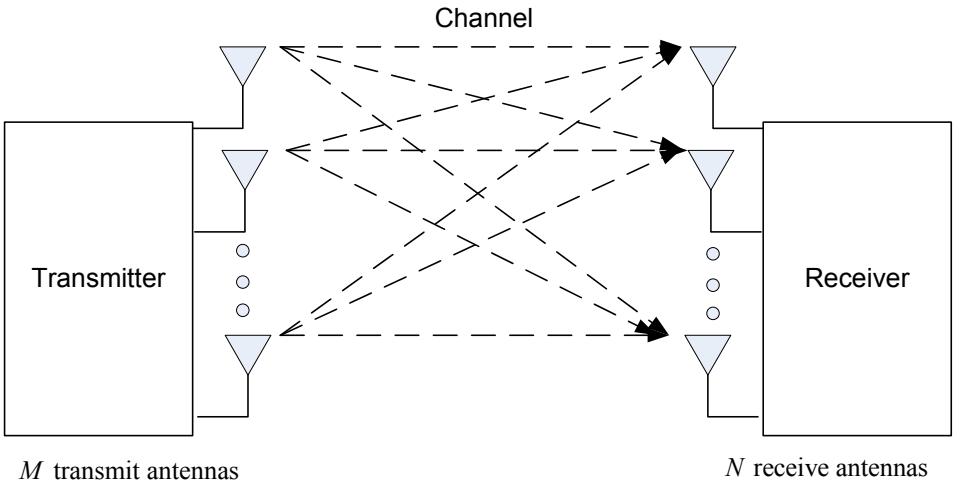


Figure 1.4: A point-to-point MIMO communication system.

gains [28, 29]. A wireless point-to-point link with M transmit and N receive antennas constitutes an M-by-N MIMO communication system, cf. Figure 1.4. The ergodic capacity of an M-by-N MIMO fading channel increases practically linearly with $\min\{M, N\}$ provided that the fading meets certain mild conditions [29, 30]. Hence, it is not surprising that MIMO has attracted a lot of research interest over the past decade since it enables significant performance and throughput gains without requiring extra transmit power and bandwidth. However, limitations on the number of antennas that a wireless device is able to support as well as the significant signal processing power and complexity required in MIMO transceivers limit the gains that can be achieved in practice.

To overcome the limitations of traditional MIMO, the concept of *cooperative communication* has recently been proposed for wireless networks such as fixed infrastructure cellular networks and wireless ad hoc networks. The basic idea of cooperative communication is that the single antenna terminals of a multiuser system can share their antennas and create a *virtual MIMO* communication system. Thereby, three different types of cooperation may be distinguished, namely, user cooperation [31]-

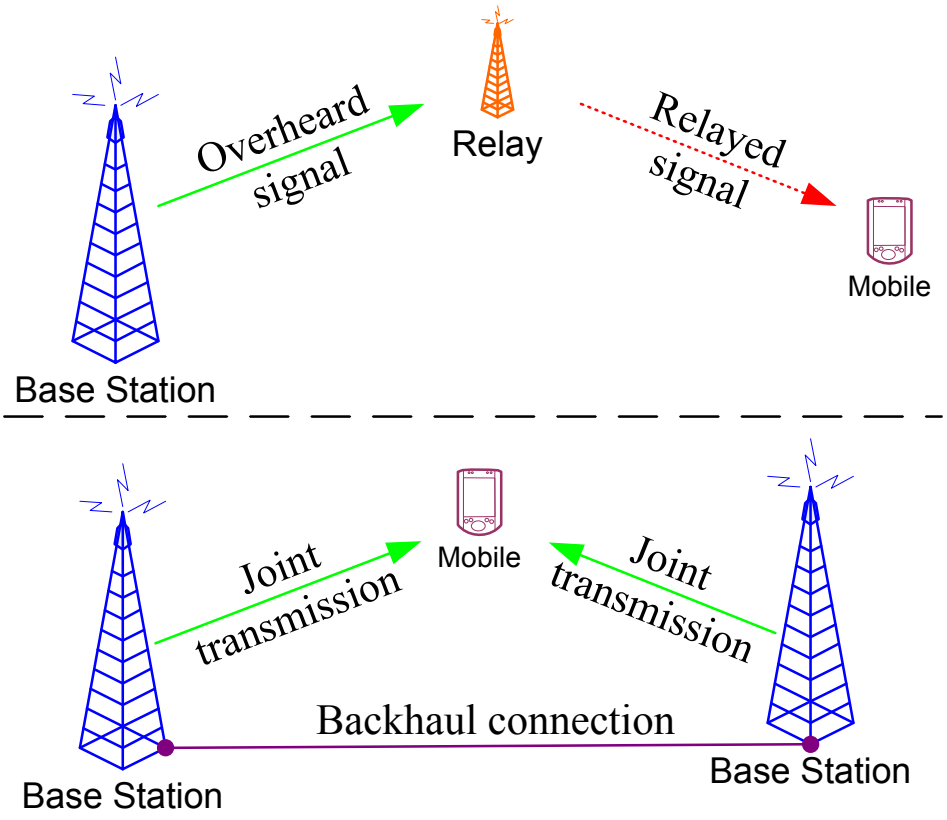


Figure 1.5: The upper half of the figure represents a one-way DF half-duplex relaying system and the lower half illustrates a system with BS cooperation.

[33], BS cooperation [34]-[37], and relaying [38]-[43]. BS cooperation and relaying are illustrated in Figure 1.5. There are different pros and cons for these three cooperation schemes. Theoretically, user cooperation and BS cooperation are able to provide huge performance gains, when compared with non-cooperative networks. However, user cooperation requires sophisticated signaling control algorithms and BS cooperation requires a strong backhaul connection for information exchange. On the other hand, cooperative relaying with dedicated relays requires significantly less signaling overhead and allows for low cost implementation. Nevertheless, a substantial spectral efficiency loss is inherent to relaying systems due to the half-duplex constraint[2].

[2]In general, a transceiver is unable to transmit and receive at the same time and frequency due to strong self-interference. A simple solution is to separate the uplink and downlink channels into

1.4 Energy Efficient Communication

Recently, green communication has received much attention in industry, academia, and government. In fact, information and communication technology (ICT) devices consume roughly 3% of the world wide energy supply and are responsible for 2% of the global carbon dioxide (CO_2) emission [44, 45]. The BSs and backhaul networks of cellular communication networks consume approximately 60 billion kWh per year and produce over a hundred million tons of CO_2 world wide. These figures are projected to double by the year 2020 if no further actions are taken. As a result, different stakeholders are cooperating with each other for reducing the CO_2 emissions of communication networks. For example, the European Commission has initiated a research project, Energy Aware Radio and neTwork tecHnologies (EARTH) [46], which investigates both the theoretical and practical achievable energy efficiency limits; the target is to enhance the energy efficiency of mobile systems by at least a factor of 50% compared to the current systems.

In the literature, a large amount of work has been devoted to resource allocation problems which aim to strike a balance between energy consumption and performance in terms of spectral efficiency (bit/s/Hz) of wireless systems. Some novel approaches developed for the physical layer, such as adaptive modulation and power control, have been successfully implemented in practical systems, e.g. the third generation wideband code division multiple access system. Energy efficiency (bit/Joule) becomes a new system performance measure when environmental issues are taken into considerations. This may result in a dilemma for system designers as energy efficiency and spectral efficiency are conflicting design goals. In the past decades, energy efficiency has been studied from the information-theoretic perspective [47]-[49]. In an additive

orthogonal signaling dimensions which is known as half-duplex.

white Gaussian noise (AWGN) channel, the maximum channel capacity C (bit/s) is given by Shannon's capacity formula:

$$C = W \log_2 \left(1 + \frac{P|h|^2}{N_0 W}\right), \tag{1.3}$$

where P is the transmit power, W is the system bandwidth, N_0 is the noise power spectral density, and $|h|^2$ is the instantaneous channel gain between the transmitter and receiver. The traditional way to define energy efficiency (bit/Joule) is

$$\eta_{EE} = \frac{W \log_2 \left(1 + \frac{P|h|^2}{N_0 W}\right)}{P} = \frac{2C}{N_0(2^{2C} - 1)}. \tag{1.4}$$

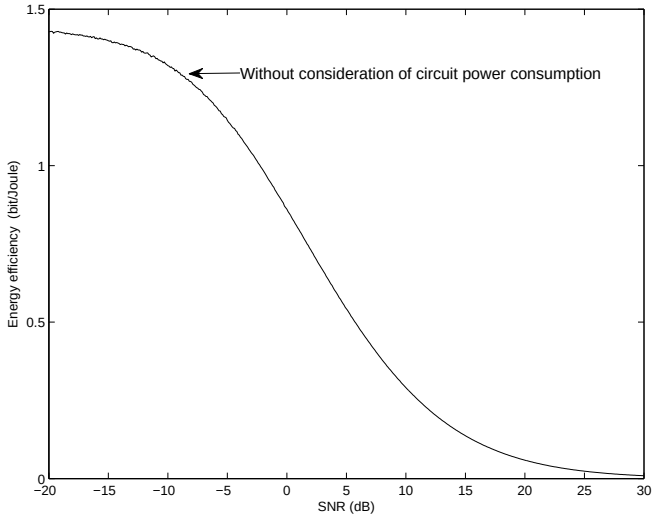
Observe from (1.4) that η_{EE} is a monotonic decreasing function of C, and hence the maximum energy efficiency η_{EE}^* occurs when $C \to 0$ or $P \to 0$ which yields

$$\eta_{EE}^* = \lim_{C \to 0} \frac{2C}{N_0(2^{2C} - 1)} = \frac{|h|^2}{\ln(2) N_0}. \tag{1.5}$$

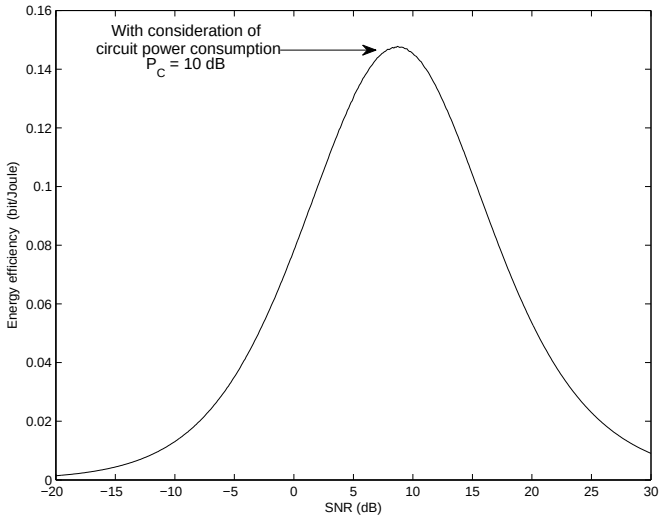
Figure 1.6(a) illustrates the monotonicity of energy efficiency in a single carrier system. However, in practice, there is always a static circuit power consumption, P_C, associated with the radiated power in the radio frequency (RF) [50, 51, 52]. Taking the circuit power consumption into consideration, the energy efficiency can be rewritten as

$$\eta_{EE} = \frac{W \log_2 \left(1 + \frac{P|h|^2}{N_0 W}\right)}{P + P_C}. \tag{1.6}$$

Interestingly, the introduction of P_C transforms the energy efficiency-versus-SNR curve from a monotonic decreasing function to a function with a bell shape with



(a) EE versus SNR without consideration of circuit power consumption.



(b) EE versus SNR with consideration of circuit power consumption.

Figure 1.6: An illustration of energy efficiency (EE) versus signal-to-noise ratio (SNR) in Rayleigh fading channel. $P_C = 10$ dB for the system with circuit power consumption consideration.

respect to SNR, cf. Figure 1.6(b). In other words, transmission with an arbitrarily low power, i.e., $P \to 0$, may no longer be the best option for maximizing the energy

efficiency for the case of $P_C > 0$. As a result, there is a trade-off between the total transmitted power and the maximum energy efficiency which should be taken into account for resource allocation algorithm design.

1.5 PHY Layer Security

Communication via different media has become an indispensable part of our daily life and security is an important issue in all communication applications. Along with the rapid development of wireless communication networks, wireless security becomes a critical concern [53, 54]. In a wireless network, eavesdropping is facilitated by the broadcast nature of the wireless medium. Traditionally, secrecy communication has relied on cryptographic encryption algorithms adopted in the application layer which requires that some form of information (key) has been shared between the legitimate entities [55]. Besides, these algorithms tend to ignore the behavior of the communication channels and assume error free communication between the legitimate entities. However, for many wireless scenarios, cryptographic encryption is impossible due to a lack of trusted third party administrators for key distribution and/or a prohibitively high complexity in messages decryption. Furthermore, error free communication cannot be always guaranteed in non-deterministic wireless channels [56, 57]. The special characteristics of wireless channels require innovative designs to ensure wireless security. In response to the shortcomings of cryptographic encryption, a natural question is: Is it possible to provide perfectly secure communication by taking advantage of the characteristics of PHY communication channels? Indeed, this form of security is referred to as information-theoretic security/PHY layer security [58]-[61] and was first studied by Wyner [62] for the classic wire-tap channel, cf. Figure 1.7. Wyner showed that a source (Alice) and a destination (Bob) can exchange perfectly secure

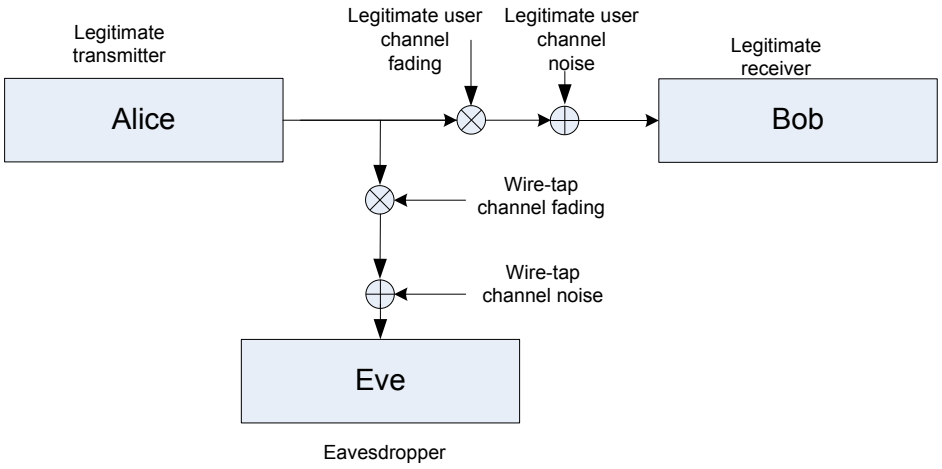


Figure 1.7: A general wire-tap channel

messages with a non-zero rate if the desired receiver enjoys better channel conditions than the passive eavesdropper (Eve). Unfortunately, this condition may not hold for wireless fading channels. In particular, Eve has a better average channel gain than Bob if she is closer to the transmitter than Bob. As a result, perfectly secure communication seems to be impossible under such scenario. Nevertheless, by exploiting the extra degrees of freedom in multiple antennas systems, secure communication is still possible if artificial noise is injected, cf. Chapter 2.

1.6 Contributions of the Thesis

This thesis considers energy efficient and secure resource allocation algorithm designs for cellular systems that may find application in several current or upcoming wireless communication standards. The main contributions of this thesis are listed in the following.

1. We propose a resource allocation and scheduling algorithm for OFDMA DF relaying systems, where a multiple antenna eavesdropper, artificial noise gener-

ation for secure communication, and the negative effect of imperfect CSIT are taken into consideration. An efficient iterative and distributed resource allocation algorithm with closed-form power, secrecy data rate, packet data rate, and subcarrier allocation is derived by dual decomposition and requires only local CSI at each relay. Simulation results not only show that the performance of the proposed algorithm converges to the optimal performance within a small number of iterations, but also demonstrate the achievable secrecy outage capacity when the eavesdropper is closer to the BS/relay than the desired users.

2. We formulate the resource allocation for energy efficient OFDMA systems as a mixed non-convex and combinatorial optimization problem, in which a multiple antenna eavesdropper, dynamic circuit power consumption, artificial noise injection for secure communication, and secrecy data rate requirements were taken into consideration. By exploiting the properties of fractional programming, the considered problem is transformed to an equivalent convex problem with a tractable solution. An efficient iterative resource allocation algorithm with closed-form power, secrecy data rate, and subcarrier allocation is derived by dual decomposition for maximization of the number of securely delivered bit-per-Joule. Simulation results demonstrate the fast convergence of the proposed algorithm in achieving the maximum energy efficiency of the considered system in the presence of a multiple antenna eavesdropper.

3. We propose an iterative resource allocation algorithm which maximizes the energy efficiency of an OFDMA communication system with a large number of BS antennas. Simulation results not only show that the proposed algorithm converges to the optimal solution within a small number of iterations, but demonstrate also the trade-off between energy efficiency and the number of

transmit antennas: The use of a large number of antennas is always beneficial for the system outage capacity, even if the CSIT is imperfect. However, an exceedingly large number of antennas may not be a cost effective solution for improving the system performance, at least not from an energy efficiency point of view.

4. We propose an iterative resource allocation algorithm for energy efficiency maximization in multi-cell OFDMA networks with joint BS zero-forcing beamforming (ZFBF) transmission. Although the considered problem is a non-convex and combinatorial optimization problem, we demonstrate that when the number of subcarriers is sufficiently large, the duality gap is practically zero despite the non-convexity of the primal problem. As a result, an efficient closed-form power allocation is obtained in each iteration via dual decomposition. Simulation results show that the performance gain due to joint BS transmission can be limited by the backhaul capacity. Yet, an exceeding large backhaul capacity may not enhance the system throughput since the performance bottleneck can also be the wireless link capacity.

1.7 Organization of the Thesis

In the following, we provide a brief overview of the remainder of this thesis.

In Chapter 2, we formulate an optimization problem for secure resource allocation and scheduling in OFDMA half-duplex DF relay-assisted networks. Our problem formulation takes into account artificial noise generation to combat a passive multiple antenna eavesdropper and the effects of imperfect CSIT in slow fading. The packet data rate, secrecy data rate, power, and subcarrier allocation policies are optimized

to maximize the average secrecy outage capacity (bit/s/Hz securely and successfully delivered to the users via relays). Simulation results illustrate that our proposed distributed iterative algorithm guarantees a non-zero secrecy data rate for given target secrecy outage and channel outage probability requirements.

In Chapter 3, resource allocation for energy efficient secure communication in an OFDMA downlink network is studied. The considered problem is modeled as a non-convex optimization problem which takes into account the sum rate dependent circuit power consumption, a multiple antenna eavesdropper, artificial noise generation, and different QoS requirements including a minimum required secrecy sum rate and a maximum tolerable secrecy outage probability. The power, secrecy data rate, and subcarrier allocation policies are optimized for maximization of the energy efficiency of secure data transmission (bit/Joule securely delivered to the users). The considered non-convex optimization problem is transformed to a convex optimization problem by exploiting the properties of fractional programming which results in an efficient iterative resource allocation algorithm. In each iteration, the transformed problem is solved by using dual decomposition. Simulation results illustrate that the proposed iterative resource allocation algorithm maximizes the system energy efficiency and guarantees a non-zero secrecy data rate for the desired users within a small number of iterations.

In Chapter 4, resource allocation for energy efficient communication in an OFDMA downlink network with a large number of transmit antennas is studied. The considered problem is modeled as a non-convex optimization problem which takes into account the circuit power consumption, imperfect CSIT, and different QoS requirements including a minimum required data rate and a maximum tolerable channel outage probability. The power allocation, data rate adaptation, antenna allocation,

and subcarrier allocation policies are optimized for maximization of the energy efficiency of data transmission (bit/Joule delivered to the users) via an iterative algorithm. Simulation results illustrate the fast convergence of the proposed algorithm and demonstrate the trade-off between energy efficiency and the number of transmit antennas.

In Chapter 5, we study resource allocation for energy efficient communication in multi-cell OFDMA downlink networks with cooperative BSs. We formulate the resource allocation problem as a non-convex optimization problem which takes into account the circuit power consumption, the limited backhaul capacity, and the minimum required data rate for joint BS ZFBF transmission. By exploiting the properties of fractional programming, the considered non-convex optimization problem in fractional form is transformed into an equivalent optimization problem in subtractive form, which enables the derivation of an efficient iterative resource allocation algorithm. In each iteration, a low complexity suboptimal semi-orthogonal user selection policy is computed. Besides, by using the concept of perturbation function, we show that the duality gap of the resulting power allocation problem is zero under some general conditions, despite the non-convexity of the primal problem. Thus, dual decomposition can be used in each iteration to derive an efficient closed-form power allocation solution for maximization of the energy efficiency of data transmission (bit/Joule delivered to the users). Simulations are used to unveil the trade-off between energy efficiency, network capacity, and backhaul capacity.

Finally, Chapter 6 summarizes the contributions of this thesis and outlines areas of future research.

Appendix A summarizes some tools and theories from convex optimization which used frequently in the thesis. Appendices B - G contain the proofs of the propositions,

claims, lemmas, and theorems used in this thesis.

Chapter 2

Secure Resource Allocation and Scheduling for OFDMA Decode-and-Forward Relay Networks

2.1 Introduction

Recently, a large amount of work has been devoted to information-theoretic PHY layer security [62]-[72], as a complement to the traditional cryptographic encryption adopted in the application layer. The pioneering work on PHY layer security by Wyner [62] showed that a source and a destination can exchange perfectly secure messages with a non-zero rate if the desired receiver enjoys better channel conditions than the passive eavesdropper(s). In [63], [64], and [65], resource allocation in multi-carrier systems with PHY layer security considerations was studied for the case of a single-user system, a two-user system, and a multi-user system, respectively. On the other hand, power allocation for systems employing cooperative jamming enabled by amplify-and-forward (AF) and DF relays was investigated in [66] and [67], respectively. In these works, the CSI of the eavesdroppers is assumed to be known at the BS such that secure communication can be guaranteed. Yet, eavesdroppers are usually passive and silent in order to hide their existence. Thus, the CSI of the

eavesdroppers cannot be measured at the BS by estimating handshaking signals or be obtained via feedback from the eavesdroppers. On the other hand, secure communication systems employing multiple antenna have been proposed for the case where the eavesdropper's CSI is not available. By exploiting the extra degrees of freedom in a multiple antennas system, artificial noise or interference is injected into the null space of the desired users to degrade the channels of the eavesdroppers. In [68] and [69], the authors studied the power allocation problem for maximizing the ergodic secrecy capacity in single-user single-carrier systems with artificial noise generation assuming the CSI of the eavesdropper is perfectly known at the BS. However, the assumption of ergodic channels cannot be justified for delay sensitive applications in practice since the transmitted packets of these applications experience quasi-static fading. Hence, a secrecy outage [70, Chapter 5] occurs whenever the scheduled secrecy data rate exceeds the secrecy capacity between the BS and the eavesdroppers, which introduces a QoS concern for secrecy. In [71] and [72], under the assumption of perfect CSI of the desired users, the authors proposed resource allocation algorithms with secrecy QoS consideration in multi-carrier single-hop and two-hop systems, respectively. Yet, the CSI of the desired users may be outdated at the transmitter even if the users are moving with pedestrian speeds. The imperfect CSIT introduces two kinds of performance degradation which have not been taken into account in [63]-[72]. First, in quasi-static fading without perfect CSIT, the transmitted packet is corrupted whenever the transmit data rate exceeds the channel capacity between the active legitimate transceivers even of channel capacity achieving codes are used for error protection, i.e., a channel outage occurs [28, Chapter 5.4]. Second, with imperfect user CSIT, the artificial noise not only interferes the eavesdropper but also the desired users since their null space information is inaccurate. Therefore, in

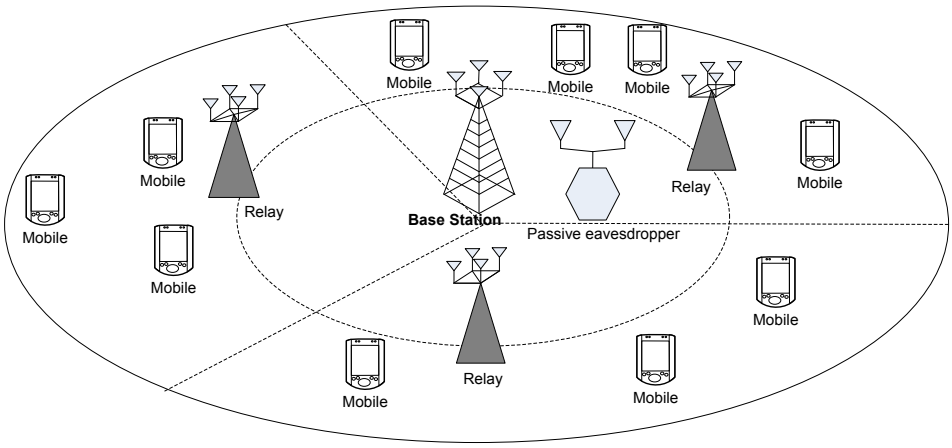


Figure 2.1: An illustration of a downlink OFDMA DF relay network. There are one BS and $M = 3$ DF relays with $N_T = 4$ antennas, $K = 10$ desired users equipped with a single antenna, and one eavesdropper with $N_E = 2$ antennas. For an effective eavesdropping, the eavesdropper chooses a location closer to either the BS or a relay than all the desired users.

this chapter, a distributed resource allocation algorithm for OFDMA DF relay networks is proposed which takes into account secrecy outage, channel outage, and the potentially negative effects of artificial noise generation.

2.2 OFDMA Downlink Network Model

2.2.1 Channel Model

We consider a downlink OFDMA network which consists of a BS with N_T antennas, M DF relays with N_T antennas each, an eavesdropper[3] with N_E antennas, and K mobile users equipped with a single antenna, cf. Figure 2.1. A single cell with two ring-shaped boundary regions is studied. The region between the inner boundary and the outer boundary is divided into M sectors of equal size as shown in Figure 2.1 and each user is assigned to only one relay according to some predefined criteria

[3]An eavesdropper with N_E antennas is equivalent to multiple eavesdroppers with a total of N_E antennas which are connected to a joint processing unit.

such as average SNR. Users in all sectors are competing for resources with each other. We assume that there is no direct transmission between the BS and the mobile users due to heavy blockage and long distance transmission. We also assume that the resource allocation for relay assisted users (located between the inner and the outer boundaries) and non-relay assisted users (located inside the inner boundary) is done separately. Both the BS and the relays adopt MIMO-beamforming (BF) to enhance the system performance. We assume that $N_T > N_E$ to enable secure communication. The eavesdropper is passive and its goal is to decode the information transmitted by the BS without causing interference to the communication channels.

The impulse responses of all channels are assumed to be time-invariant (slow fading). We consider an OFDMA DF relay assisted system with n_F subcarriers. The received symbols in the first time slot at relay $m \in \{1, \ldots, M\}$ for user $k \in \{1, \ldots, K\}$ and the eavesdropper on subcarrier $i \in \{1, \ldots, n_F\}$ are given by, respectively,

$$\mathbf{y}_{BR_m}[i] = \mathbf{H}_{BR_m}[i]\mathbf{x}_k[i] + \mathbf{n}_{R_m}[i] \quad \text{and} \tag{2.1}$$

$$\mathbf{y}_{BE}[i] = \mathbf{G}_{BE}[i]\mathbf{x}_k[i] + \mathbf{e}_1[i], \tag{2.2}$$

where $\mathbf{x}_k[i] \in \mathbb{C}^{N_T \times 1}$ denotes the transmitted symbol vector. $\mathbf{H}_{BR_m}[i] \in \mathbb{C}^{N_T \times N_T}$ denotes the channel matrix between the BS and relay m on subcarrier i and $\mathbf{G}_{BE}[i] \in \mathbb{C}^{N_E \times N_T}$ is the channel matrix between the BS and the eavesdropper on subcarrier i. Both variables, $\mathbf{H}_{BR_m}[i]$ and $\mathbf{G}_{BE}[i]$, include the effects of path loss and multipath fading. $\mathbf{n}_{R_m}[i] \in \mathbb{C}^{N_T \times 1}$ and $\mathbf{e}_1[i] \in \mathbb{C}^{N_E \times 1}$ are the AWGN in subcarrier i at relay m and the eavesdropper in the first time slot, respectively. Each entry in both vectors has distribution $\mathcal{CN}(0, N_0)$, where N_0 is the noise variance. In the second time slot, relay m decodes message $\mathbf{x}_k[i]$ and re-encodes the message as $\mathbf{q}_{R_m,k}[i] \in \mathbb{C}^{N_T \times 1}$. Then, relay m forwards the re-encoded message $\mathbf{q}_{R_m,k}[i]$ to user k. Therefore, the signals

received at user k and the eavesdropper on subcarrier i from relay m are given by, respectively,

$$y_{R_m,k}[i] = \mathbf{h}_{R_m,k}[i]\mathbf{q}_{R_m,k}[i] + n_k[i] \quad \text{and} \tag{2.3}$$

$$\mathbf{y}_{R_m,E}[i] = \mathbf{G}_{R_m,E}[i]\mathbf{q}_{R_m,k}[i] + \mathbf{e}_2[i]. \tag{2.4}$$

$\mathbf{h}_{R_m,k}[i] \in \mathbb{C}^{1\times N_T}$ and $\mathbf{G}_{R_m,E}[i] \in \mathbb{C}^{N_E\times N_T}$ denote the channel matrices from relay m to users k and from relay m to the eavesdropper on subcarrier i, respectively. $n_k[i] \in \mathbb{C}^{1\times 1}$ and $\mathbf{e}_2[i] \in \mathbb{C}^{N_E\times 1}$ are the AWGN in subcarrier i at user k and the eavesdropper in the second time slot, respectively. For the sake of notational simplicity and without loss of generality, a normalized noise variance of $N_0 = 1$ is assumed for all receivers in the following.

2.2.2 Channel State Information

The resource allocation and scheduling problem presented in the next section can be solved either centrally at the BS or in a distributed fashion. For the centralized solution, the BS requires the CSI of all BS-to-relay and relay-to-user links at the beginning of each scheduling slot. In contrast, for the distributed solution the relays only require the CSI of their own BS-to-relay and relay-to-user links, whereas the BS does not need any CSI. We assume a FDD system where the CSI of the relay-to-user links is obtained through feedback from the users to the relays at the beginning of each scheduling slot, while the CSI of the BS-to-relay links can be obtained at the relays either in the handshaking phase or from a previous transmission. In the following, since path loss is a slowly varying random process which changes on the order of seconds, we assume that the path loss can be estimated perfectly. For the

multipath fading, we take into account the different natures of the BS-to-relay and the relay-to-user links. In particular, since both the BS and the relays are static, the BS-to-relay links are assumed to be time-invariant. Thus, the BS-to-relay fading gains $\mathbf{H}_{BR_m}[i]$, $m \in \{1, \ldots, M\}$, $i \in \{1, \ldots, n_F\}$, can be reliably estimated at the relays with negligible estimation error. Therefore, we can assume perfect CSIT for the BS-to-relay links. On the other hand, although we also assume that the users can obtain perfect estimates of the relay-to-user fading gains $\mathbf{h}_{R_m,k}[i]$, $m \in \{1, \ldots, M\}$, $k \in \{1, \ldots, K\}$ for signal detection purpose, the corresponding CSI may be outdated at the relays (for the distributed solution) and at the BS (for the centralized solution) because of the mobility of the users and the feedback delay. To capture this effect, we model the multipath fading CSIT of the link between user k and relay m on subcarrier i as

$$\mathbf{h}_{R_m,k}[i] = \hat{\mathbf{h}}_{R_m,k}[i] + \Delta\mathbf{h}_{R_m,k}[i], \tag{2.5}$$

where $\hat{\mathbf{h}}_{R_m,k}[i]$ and $\Delta\mathbf{h}_{R_m,k}[i]$ denote, respectively, the estimated CSI vector and the CSIT error vector. $\hat{\mathbf{h}}_{R_m,k}[i]$ and $\Delta\mathbf{h}_{R_m,k}[i]$ are Gaussian random vectors and each vector has independent and identically distributed (i.i.d.) elements. Besides, the elements of vectors $\mathbf{h}_{R_m,k}[i]$, $\hat{\mathbf{h}}_{R_m,k}[i]$, and $\Delta\mathbf{h}_{R_m,k}[i]$ have zero means and variance $\sigma^2_{h_{R_m,k}}$, $\sigma^2_{h_{R_m,k}} - \sigma^2_e$, and σ^2_e, respectively. Assuming a minimum mean square error (MMSE) estimator, the CSI error vector and the actual CSI vector are mutually uncorrelated [73, p.177].

On the other hand, the CSI of the eavesdropper is unavailable at both the BS and the relays. Thus, in order to secure the desired wireless communication links, *artificial noise* signals are generated at both the BS and the relays to degrade the channels between the BS/relays and the eavesdropper.

2.2.3 Artificial Noise Generation

The BS and relay m choose $\mathbf{x}_k[i]$ and $\mathbf{q}_{R_m,k}[i]$ as the linear combination of the information bearing signal and an artificial noise signal which can be presented as

$$\mathbf{x}_k[i] = \underbrace{\mathbf{b}_{BR_m,k}[i]u_k[i]\sqrt{\alpha_{BR_m,k}[i]P_{BR_m,k}[i]}}_{\text{Desired Signal}} + \underbrace{\mathbf{V}_{BR_m}[i]\mathbf{v}[i]}_{\text{Artificial Noise}} \quad \text{and} \quad (2.6)$$

$$\mathbf{q}_{R_m,k}[i] = \underbrace{\hat{\mathbf{r}}_{R_m,k}[i]u_k[i]\sqrt{\alpha_{R_m,k}[i]P_{R_m,k}[i]}}_{\text{Desired Signal}} + \underbrace{\mathbf{W}_{R_m,k}[i]\mathbf{w}[i]}_{\text{Artificial Noise}}, \quad (2.7)$$

respectively. $u_k[i] \in \mathbb{C}^{1\times 1}$ is the information bearing signal, $\mathbf{v}[i] \in \mathbb{C}^{N_T-1\times 1}$ and $\mathbf{w}[i] \in \mathbb{C}^{N_T-1\times 1}$ are artificial noise vectors whose elements are i.i.d. complex Gaussian random variables with variance $\sigma_v^2[i]$ and $\sigma_w^2[i]$, respectively. Since $\mathbf{H}_{BR_m}[i]$ and $\hat{\mathbf{h}}_{R_m,k}[i]$ are known at the BS and relay m, respectively, MIMO-BF can be used to maximize the received SNR ratio at the desired receivers. The beamforming vectors adopted at the BS and relay m, i.e, $\mathbf{b}_{BR_m,k}[i] \in \mathbb{C}^{N_T\times 1}$ and $\hat{\mathbf{r}}_{R_m,k}[i] \in \mathbb{C}^{N_T\times 1}$, are chosen to be the eigenvectors corresponding to the maximum eigenvalue of $\mathbf{H}^{\dagger}_{BR_m}[i]\mathbf{H}_{BR_m}[i]$ and $\hat{\mathbf{h}}^{\dagger}_{R_m,k}[i]\hat{\mathbf{h}}_{R_m,k}[i]$, respectively. Furthermore, we define two orthogonal bases, $\mathbf{V}_{BR_m}[i] \in \mathbb{C}^{N_T\times N_T-1}$ and $\mathbf{W}_{R_m,k}[i] \in \mathbb{C}^{N_T\times N_T-1}$, by using the remaining eigenvectors of $\mathbf{H}^{\dagger}_{BR_m}[i]\mathbf{H}_{BR_m}[i]$ and $\hat{\mathbf{h}}^{\dagger}_{R_m,k}[i]\hat{\mathbf{h}}_{R_m,k}[i]$, respectively. $P_{BR_m,k}[i]$ represents the transmit power at the BS on subcarrier i to relay m for serving user k. $P_{R_m,k}[i]$ denotes the transmit power at relay m on subcarrier i to user k. Variables $0 < \alpha_{BR_m,k}[i] \leq 1$ and $0 < \alpha_{R_m,k}[i] \leq 1$ are the fractions of power devoted to the information bearing signal at the BS and relay m on subcarrier i for user k, respectively. Since the CSI of the eavesdropper is unavailable at both the BS and the relays, the remaining powers at the BS and relay m on subcarrier i are equally distributed across $N_T - 1$ dimensions for generating the artificial noises with variances $\sigma_v^2[i] = \frac{(1-\alpha_{BR_m,k}[i])P_{BR_m,k}[i]}{N_T-1}$

and $\sigma_w^2[i] = \frac{(1-\alpha_{R_m,k}[i])P_{R_m,k}[i]}{N_T-1}$, respectively. Hence, the received signals in (2.1) can be rewritten as

$$\mathbf{y}_{BR_m}[i] = \mathbf{H}_{BR_m}[i]\big(\mathbf{b}_{BR_m,k}[i]u_k[i]\sqrt{\alpha_{BR_m,k}[i]P_{BR_m,k}[i]}+\mathbf{V}_{BR_m}[i]\mathbf{v}[i]\big)+\mathbf{n}_{R_m}[i], \quad (2.8)$$

$$\mathbf{y}_{BE}[i] = \mathbf{G}_{BE}[i]\big(\mathbf{b}_{BR_m,k}[i]u_k[i]\sqrt{\alpha_{BR_m,k}[i]P_{BR_m,k}[i]}+\mathbf{V}_{BR_m}[i]\mathbf{v}[i]\big)+\mathbf{e}_1[i]. \quad (2.9)$$

In the second time slot, relay m eliminates the artificial noise by pre-processing the received signal as

$$\begin{aligned}\tilde{\mathbf{y}}_{BR_m}[i] &= (\mathbf{H}_{BR_m}[i]\mathbf{b}_{BR_m,k}[i])^\dagger \mathbf{y}_{BR_m}[i] \\ &= \sqrt{\alpha_{BR_m,k}[i]P_{BR_m,k}[i]}\lambda_{\max_{BR_m}}[i]u_k[i] + \tilde{\mathbf{n}}_{R_m}[i], \end{aligned} \quad (2.10)$$

where $\tilde{\mathbf{n}}_{R_m}[i] = \mathbf{b}_{BR_m,k}^\dagger[i]\mathbf{H}_{BR_m}^\dagger[i]\mathbf{n}_{R_m}[i]$ is AWGN which has the same distribution as $\mathbf{n}_{R_m}[i]$ and $\lambda_{\max_{BR_m}}[i]$ is the maximum eigenvalue of $\mathbf{H}_{BR_m}^\dagger[i]\mathbf{H}_{BR_m}[i]$. It can be observed that the artificial noise signal generated at the BS does not interfere with the desired signal at relay m due to the adopted pre-processing. On the other hand, the signal received at user k and the eavesdropper on subcarrier i from relay m in (2.3) and can be rewritten as

$$y_{R_m,k}[i] = \mathbf{h}_{R_m,k}[i]\Big(\hat{\mathbf{r}}_{R_m,k}[i]u_k[i]\sqrt{\alpha_{R_m,k}[i]P_{R_m,k}[i]} + \mathbf{W}_{R_m,k}[i]\mathbf{w}[i]\Big) + n_k[i] \text{ and} \quad (2.11)$$

$$\mathbf{y}_{R_m,E}[i] = \mathbf{G}_{R_m,E}[i]\Big(\hat{\mathbf{r}}_{R_m,k}[i]u_k[i]\sqrt{\alpha_{R_m,k}[i]P_{R_m,k}[i]} + \mathbf{W}_{R_m,k}[i]\mathbf{w}[i]\Big) + \mathbf{e}_2[i], \quad (2.12)$$

respectively. Note that due to the imperfect CSIT at relay m, there is an artificial noise leakage from the null space to the range space of user k on subcarrier i. The negative effects of artificial noise generation with imperfect CSIT are demonstrated in the next section via channel capacity equations and the concept of outages.

2.3 Resource Allocation and Scheduling

2.3.1 Instantaneous Channel Capacity, Channel Outage, and Secrecy Outage

Since we assume perfect CSIR, the instantaneous channel capacity between the BS and relay m on subcarrier i is given by

$$C_{BR_m,k}[i] = \log_2\Big(1 + \alpha_{BR_m,k}[i]P_{BR_m,k}[i]\lambda_{\max_{BRm}}[i]\Big). \tag{2.13}$$

On the other hand, user k first estimates the effective channel $\mathbf{h}_{R_m,k}[i]\hat{\mathbf{r}}_{R_m,k}[i]$ for coherent detection. Hence, the instantaneous channel capacity between relay m and users k on subcarrier i is obtained as

$$C_{R_m,k}[i] = \log_2\Big(1 + \frac{\alpha_{R_m,k}[i]P_{R_m,k}[i]\|\mathbf{h}_{R_m,k}[i]\hat{\mathbf{r}}_{R_m,k}[i]\|^2}{1 + (1-\alpha_{R_m,k})P_{R_m,k}[i]\sigma_e^2}\Big). \tag{2.14}$$

Thus, the channel capacity between the BS and user k via relay m on subcarrier i is given by

$$C_{m,k}[i] = \frac{1}{2}\min\Big\{C_{BR_m,k}[i], C_{R_m,k}[i]\Big\}, \tag{2.15}$$

where the pre-log factor $\frac{1}{2}$ is due to the two channel uses required for transmitting one message.

In practice, the eavesdropper has to be close to either the BS or one of the relays for effective eavesdropping. Thus, one of the signals received in the two time slots will be much stronger than the other one making selection combining of the two received signals at the eavesdropper near optimal. Besides, since we assume the BS and the

relays do not have any CSI of the eavesdropper, we follow the approach in [68, 69] and consider a capacity upper bound for the eavesdropper for resource allocation purposes assuming the absence of thermal noise at the eavesdropper receiver. Therefore, the capacity of the eavesdropper is upper bounded[4] by

$$C_{m,E}[i] = \frac{1}{2}\log_2\Big(1+\max\{\Gamma_{B,E}[i],\Gamma_{R_m,E}[i]\}\Big), \tag{2.16}$$

$$\Gamma_{B,E}[i] = \frac{\alpha_{BR_m,k}[i](N_T-1)}{1-\alpha_{BR_m,k}[i]}\mathbf{g}_1^{\dagger}[i](\mathbf{G}_1[i]\mathbf{G}_1^{\dagger}[i])^{-1}\mathbf{g}_1[i], \tag{2.17}$$

$$\Gamma_{R_m,E}[i] = \frac{\alpha_{R_m,k}[i](N_T-1)}{1-\alpha_{R_m,k}[i]}\mathbf{g}_2^{\dagger}[i](\mathbf{G}_2[i]\mathbf{G}_2^{\dagger}[i])^{-1}\mathbf{g}_2[i], \tag{2.18}$$

where $\mathbf{g}_1[i]=\mathbf{G}_{BE}[i]\mathbf{b}_{BR_m,k}[i]$, $\mathbf{G}_1[i]=\mathbf{G}_{BE}[i]\mathbf{V}_{BR_m}[i]$, $\mathbf{g}_2[i]=\mathbf{G}_{R_m,E}[i]\mathbf{q}_{R_m,k}[i]$, and $\mathbf{G}_2[i]=\mathbf{G}_{R_m,E}[i]\mathbf{W}_{R_m,k}[i]$. We note that the proposed resource allocation algorithm (see next section) can also be applied if other combining schemes such as optimal maximum ratio combining (MRC) are used. (We substitute $F_{z_c}^{-1}(\cdot,i)$ in (2.22) by the inverse cumulative distribution function (CDF) of the resultant MRC SIR.) The maximum achievable secrecy data rate $R_{m,k}^{sec}[i]$ of a perfectly secure communication between the BS and user k on subcarrier i via relay m with outage consideration can be expressed as

$$R_{m,k}^{sec}[i] \quad \times \quad 1(R_{m,k}^{data}[i] < C_{m,k}[i]) \times \; 1(R_{m,k}^{data}[i] - C_{m,E}[i] > R_{m,k}^{sec}[i]). \tag{2.19}$$

$R_{m,k}^{data}[i]$ denotes the actual packet data rate transmitted from the BS to user k via relay m. The relationships between the variables in (2.19) are illustrated in Figure 2.2. In the considered system, there are two types of outage measures. The first one is known as channel outage [28, Section 5.4] which corresponds to the first indicator function in

[4]The upper bound is referring to the individual equations in (2.17) and (2.18) for which the absence of thermal noise at the eavesdropper is assumed.

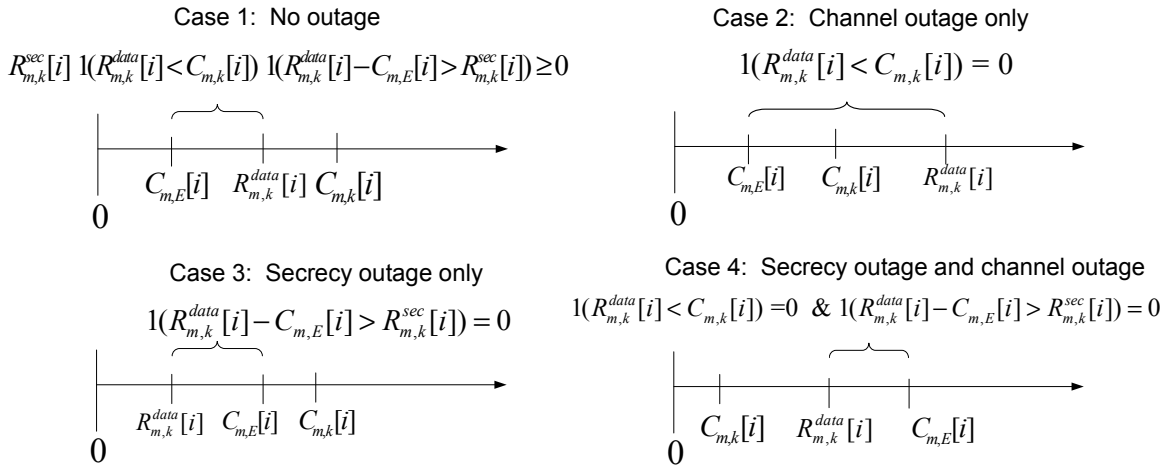


Figure 2.2: An illustration of the relationship between packet data rate $R_{m,k}^{data}[i]$, secrecy data rate $R_{m,k}^{sec}[i]$, the capacity of the user channel, $C_{m,k}[i]$, and the capacity of the eavesdropper channel, $C_{m,E}[i]$, for four possible cases.

(2.19). It occurs whenever the transmit data rate exceeds the instantaneous channel capacity between two desired transceivers, i.e., $R_{m,k}^{data}[i] > C_{m,k}[i]$. If $R_{m,k}^{data}[i] > C_{m,k}[i]$, any transmitted packet is corrupted even if a channel capacity achieving code is applied for error protection. Indeed, channel outage can be avoided by data rate adaptation when the CSIT of the desired user channel can be perfectly obtained. Yet, highly accurate CSIT is difficult to obtain if the users are not static. The second type of outage is secrecy outage [70, Chapter 5] which corresponds to the second indicator function in (2.19). If the CSI of all links (including the links of the eavesdropper) are available at the BS, the resource allocator can set the target secrecy data rate $R_{m,k}^{sec}[i]$ to match the channel conditions [68], i.e., $R_{m,k}^{sec}[i] < R_{m,k}^{data}[i] - C_{m,E}[i]$ and $R_{m,k}^{data}[i] > C_{m,E}[i]$, such that a packet with secrecy data rate $R_{m,k}^{sec}[i]$ and data rate $R_{m,k}^{data}[i]$ can be securely delivered and successfully decoded by the desired user. However, here the eavesdropper is assumed to be passive and its CSI is not available at the BS, i.e., $C_{m,E}[i]$ is a random variable for the BS. Hence, a *secrecy outage* occurs whenever the target secrecy data rate $R_{m,k}^{sec}[i]$ exceeds the secrecy capacity,

i.e., $R_{m,k}^{data}[i] - C_{m,E}[i]$.

In order to model the unreliability and the insecurity due to both *channel outage* and *secrecy outage*, respectively, we consider the performance in terms of the *average secrecy outage capacity*, which is defined as the total average bit/s/Hz securely and successfully delivered to the K mobile users and is given by

$$\begin{aligned} & U_{sec}(\mathcal{P},\mathcal{R},\mathcal{S}) \\ = & \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m} w_k \sum_{i=1}^{n_F} \frac{s_{m,k}[i]}{n_F} \\ & \times\mathcal{E}\Big\{R_{m,k}^{sec}[i]\times 1\Big(R_{m,k}^{data}[i]-C_{m,E}[i]>R_{m,k}^{sec}[i]\Big)\ 1\Big(R_{m,k}^{data}[i]<C_{m,k}[i]\Big)\Big\} \\ = & \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m} w_k \sum_{i=1}^{n_F} \frac{s_{m,k}[i]}{n_F}\Bigg\{R_{m,k}^{sec}[i]\times \Pr\Big[R_{m,k}^{data}[i]-C_{m,E}[i]>R_{m,k}^{sec}[i]\Big|\mathbf{\Delta}_{m,k}[i]\Big] \\ & \times \Pr\Big[R_{m,k}^{data}[i]<C_{m,k}[i]\Big|\mathbf{\Delta}_{m,k}[i]\Big]\Bigg\}. \end{aligned} \tag{2.20}$$

Here, $\mathcal{P},\mathcal{R}$, and $\mathcal{S}$ are the power, data rate (secrecy data rate and packet data rate), and subcarrier allocation policies, respectively. $\mathcal{U}_m$ denotes the set of users associated with relay m. $s_{m,k}[i]\in\{0,1\}$ is the subcarrier allocation indicator. w_k is a positive constant provided by the upper layers, which allows the resource allocator to give different priorities to different users and to enforce certain notions of fairness. $\mathbf{\Delta}_{m,k}[i]$ represents a pair of CSI vectors, namely the perfect CSI vector of the BS-to-relay m link and the imperfect CSI vector of the relay m-to-user k channel on subcarrier i.

2.3.2 Optimization Problem Formulation

The optimal power allocation policy, $\mathcal{P}^*$, data rate (secrecy data rate and packet data rate) allocation policy, $\mathcal{R}^*$, and subcarrier allocation policy, $\mathcal{S}^*$, can be obtained from

Problem 2.1 (Optimization Problem Formulation)

$$
\begin{aligned}
&\arg \max_{\mathcal{P},\mathcal{R},\mathcal{S},\alpha_{BR_m,k}[i],\alpha_{R_m,k}[i]} U_{sec}(\mathcal{P},\mathcal{R},\mathcal{S}) \\
&\text{s.t. C1: } \Pr\left[R_{m,k}^{data}[i] \geq C_{m,k}[i] \middle| \mathbf{\Delta}_{m,k}[i]\right] \leq \varepsilon, \quad \forall k,i, \\
&\quad \text{C2: } \Pr\left[R_{m,k}^{sec}[i] \geq R_{m,k}^{data}[i] - C_{m,E}[i] \middle| \mathbf{\Delta}_{m,k}[i]\right] \leq \delta, \forall k,i, \\
&\quad \text{C3: } \sum_{m=1}^{M} \sum_{k\in\mathcal{U}_m} \sum_{i=1}^{n_F} P_{BR_m,k}[i] s_{m,k}[i] \leq P_{B_T}, \\
&\quad \text{C4: } \sum_{k\in\mathcal{U}_m} \sum_{i=1}^{n_F} P_{R_m,k}[i] s_{m,k}[i] \leq P_{R_T}, \quad \forall m, \\
&\quad \text{C5: } \sum_{m=1}^{M} \sum_{k\in\mathcal{U}_m} s_{m,k}[i] \leq 1, \quad \forall i, \\
&\quad \text{C6: } s_{m,k}[i] = \{0,1\}, \ \forall i,k,m, \\
&\quad \text{C7: } P_{BR_m,k}[i], P_{R_m,k}[i] \geq 0, \quad \forall i,k,m, \\
&\quad \text{C8: } 0 < \alpha_{BR_m,k}[i], \alpha_{R_m,k}[i] \leq 1, \ \forall i,k,m.
\end{aligned} \tag{2.21}
$$

Here, C1 represents the required data rate outage probability due to the imperfect CSI of the relay-to-user channels. In C2, δ denotes the required secrecy outage probability in the system. Note that C1 and C2 represent two QoS metrics for communication reliability and communication security, respectively. C3 (C4) represents the individual power constraint for the BS (relays) with maximum transmit power P_{B_T} (P_{R_T}). Constraints C5 and C6 are imposed to guarantee that each subcarrier will be used by one user only. C7 and C8 are the boundary constraints of the power allocation variables.

Remark 2.1 *The optimal amount of artificial noise strikes a balance between the channel capacity and the secrecy capacity. When there is no power allocated to the artificial noise generation, the channel capacity will be maximized since all the power is allocated to the information bearing signal. However, a certain secrecy outage prob-*

ability cannot be guaranteed and the secrecy capacity decreases dramatically to zero for most channel conditions. On the contrary, when nearly all the power is allocated to the artificial noise generation, although the capacity of the eavesdropper channel approaches zero, because of the imperfect CSIT, the excessive artificial noise will also interfere with the desired user signal which decreases both the channel capacity and the secrecy capacity. Besides, even with perfect CSIT, the channel capacity decreases if we allocate more power to the artificial noise.

2.4 Solution of the Optimization Problem

2.4.1 Transformation of the Optimization Problem

For derivation of an efficient resource allocation algorithm, it is convenient to incorporate the channel outage constraint C1 and the secrecy outage probability constraint C2 in (2.21) into the objective function. This is possible if the constraints in C1 and C2 are fulfilled with equality for the optimal solution. Thus, in the following we replace the "$\leq$"-signs in C1 and C2 by "$=$"-signs and the resulting optimization problem may be viewed as a restricted version of the original problem (2.21) since the latter has a smaller feasible set. We are now ready to introduce the following Lemma.

Lemma 2.1 (Equivalent Objective Function) *For a given channel outage probability ε and a given secrecy outage probability δ in C1 and C2, respectively, the equivalent secrecy data rate in high SNR on subcarrier i for user k via relay m is*

lower bounded by

$$
\begin{aligned}
R_{m,k}^{sec}[i] &> \left[R_{m,k}^{data}[i] - \frac{1}{2}\log_2\left(1+\frac{\alpha_{BR_m,k}^*[i]\Lambda_E[i]}{1-\alpha_{BR_m,k}^*[i]}\right)\right]^+, \\
\text{where } R_{m,k}^{data}[i] &= \frac{\min\left\{\log_2\left(1+\Gamma_{BR_m,k}[i]\right), \log_2\left(1+\Gamma_{R_m,k}[i]\right)\right\}}{2}, \\
\Gamma_{BR_m,k}[i] &= \alpha_{BR_m,k}^*[i]P_{BR_m,k}[i]\lambda_{\max_{BR_m}}[i], \\
\Gamma_{R_m,k}[i] &= \frac{\alpha_{BR_m,k}^*[i]P_{R_m,k}[i]F_{\chi_2}^{-1}(\varepsilon,i)}{1+(1-\alpha_{BR_m,k}^*)P_{R_m,k}[i]\sigma_e^2}, \\
\Lambda_E[i] &= (N_T-1)F_{z_c}^{-1}(\delta,i), \\
\alpha_{BR_m,k}^*[i] &= \alpha_{R_m,k}^*[i] = \frac{1}{\sqrt{\Lambda_E[i]}}, \qquad (2.22)
\end{aligned}
$$

where $F_{z_c}^{-1}(\cdot,i)$ denotes the inverse function of $F_{z_c}(z,i)$ which is defined in the Appendix in (B.8), and $F_{\chi_2}^{-1}(\cdot,i)$ denotes the inverse CDF of a non-central chi-square random variable with two degrees of freedom and non-centrality parameter $\|\hat{\mathbf{h}}_{R_m,k}[i]\|^2$.

Proof: Please refer to the Appendix B.

There are two important observations from the above lemma. First, the asymptotically optimal $\alpha_{BR_m,k}^*[i]$ and $\alpha_{R_m,k}^*[i]$ in (2.22) suggests that in high SNR, the optimal fraction of power devoted to the artificial noise only depends on the channel statistic of the eavesdropper channel and the secrecy outage probability requirement. Second, the signal–to–interference–plus–noise ratio (SINR) of the eavesdropper, $\frac{\alpha_{R_m,k}^*[i]}{1-\alpha_{R_m,k}^*[i]}\Lambda_E[i]$, approaches a constant value at high SNR. More importantly, the SINR of the eavesdropper on each subcarrier is independent of the transmit power variables in both hops, which simplifies the derivation of the optimal resource allocation algorithm. This important observation will be verified in Section 2.5 via simulation.

By substituting (2.22) into (2.20), a modified objective function is obtained and the considered problem becomes an NP-hard mixed combinatorial and convex optimization problem, where the combinatorial nature comes from the binary constraints in the subcarrier assignment. Therefore, we follow the approach in [74] and relax constraint C6 in (2.21). In particular, we allow $s_{m,k}[i]$ to assume any real value between zero and one. Then, $s_{m,k}[i]$ can be interpreted as a time sharing factor for the K users for utilizing subcarrier i. For facilitating the time sharing on each subcarrier, we introduce two new variables and define them as $\tilde{P}_{BR_m,k}[i] = P_{BR_m,k}[i]s_{m,k}[i]$ and $\tilde{P}_{R_m,k}[i] = P_{R_m,k}[i]s_{m,k}[i]$. These two variables are the actual transmit power of the BS and relay m on subcarrier i for user k under the time-sharing assumption. Then, we can transform Problem 2.1 in (2.21) into its epigraph form [75]:

Problem 2.2 (Transformed Optimization Problem)

$$
\begin{aligned}
&\max_{\mathcal{P},\mathcal{R},\mathcal{S},z_{m,k}[i]} \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m}\sum_{i=1}^{n_F} w_k z_{m,k}[i] \\
&\quad \text{s.t.} \quad \text{C5, C7,} \\
&\text{C3:} \sum_{m=1}^{M}\sum_{k\in\mathcal{U}_m}\sum_{i=1}^{n_F} \tilde{P}_{BR_m,k}[i] \le P_{B_T}, \\
&\text{C4:} \sum_{k\in\mathcal{U}_m}\sum_{i=1}^{n_F} \tilde{P}_{R_m,k}[i] \le P_{R_T}, \quad \forall m, \\
&\text{C6: } 0 \le s_{m,k}[i] \le 1, \qquad \forall m,k,i, \\
&\text{C9:} s_{m,k}[i]\widetilde{R}^{1st}_{m,k}[i] \ge z_{m,k}[i], \forall m,k,i, \\
&\text{C10:} s_{m,k}[i]\widetilde{R}^{2nd}_{m,k}[i] \ge z_{m,k}[i], \forall m,k,i,
\end{aligned} \tag{2.23}
$$

where

$$\begin{aligned} & \widetilde{R}^{1st}_{m,k}[i] \qquad (2.24) \\ = \; & \frac{1}{2}\Big[\log_2\Big(1+\alpha^*_{BR_m,k}[i]\tilde{P}_{BR_m,k}[i]\lambda_{\max_{BR_m}}[i]\Big)-\log_2\Big(1+\frac{\alpha^*_{BR_m,k}[i]\Lambda_E[i]}{1-\alpha^*_{BR_m,k}[i]}\Big)\Big]^+ \text{and} \end{aligned}$$

$$\begin{aligned} & \widetilde{R}^{2nd}_{m,k}[i] \\ = \; & \frac{1}{2}\Big[\log_2\Big(1+\frac{\alpha^*_{R_m,k}[i]\tilde{P}_{R_m,k}[i]F^{-1}_{\chi_2}(\varepsilon,i)}{1+(1-\alpha^*_{R_m,k})\tilde{P}_{R_m,k}[i]\sigma_e^2}\Big)-\log_2\Big(1+\frac{\alpha^*_{R_m,k}[i]\Lambda_E[i]}{1-\alpha^*_{R_m,k}[i]}\Big)\Big]^+ \qquad (2.25) \end{aligned}$$

are the achievable secrecy data rate in the first and second hop, respectively. The extra constraints C9 and C10 represent the hypograph [75] of the original optimization problem in (2.21). Mathematically, the operators $[\cdot]^+$ in C9 and C10 in (2.23) destroy the concavity of the problem. Nevertheless, as will be seen in the Karush Kuhn Tucker (KKT) conditions in (2.35), those users with negative secrecy data rate will not be considered in the subcarrier selection process, since, secure communication cannot be guaranteed for those users. Therefore, we can remove the operators $[\cdot]^+$ from $\widetilde{R}^{1st}_{m,k}[i]$ and $\widetilde{R}^{2nd}_{m,k}[i]$ in (2.24), while preserving the concavity of the transformed problem. On the other hand, the constant term $\frac{1}{n_F}$ was removed from the transformed objective function for simplicity as it does not affect the values of the arguments which maximize the objective function. Besides, C8 was also removed from the optimization problem as the asymptotically optimal $\alpha^*_{R_m,k}[i]$ and $\alpha^*_{BR_m,k}[i]$ in (2.22) always satisfy the constraint. The extra constraints C9 and C10 represent the hypograph [75] of the original optimization problem in (2.21). Now, the transformed problem is jointly concave with respect to all optimization variables for $\tilde{P}_{R_m,k} \to \infty$, and under some mild conditions [75], solving the dual problem is equivalent to solving the primal problem.

2.4.2 Dual Problem Formulation

In this subsection, we solve the resource allocation and scheduling optimization problem by solving its dual. For this purpose, we first need the Lagrangian function of the primal problem. Upon rearranging terms, the Lagrangian can be written as

$$\begin{aligned}
&\mathcal{L}(\lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathcal{P}, \mathcal{R}, \mathcal{S}, z_{m,k}[i]) \\
&= \sum_{m=1}^{M} \sum_{k \in \mathcal{U}_m} \sum_{i=1}^{n_F} (w_k - (\mu_{m,k}[i] + \nu_{m,k}[i])) z_{m,k}[i] - \lambda \Big(\sum_{m=1}^{M} \sum_{k \in \mathcal{U}_m} \sum_{i=1}^{n_F} \tilde{P}_{BR_m,k}[i] - P_{B_T} \Big) \\
&- \sum_{m=1}^{M} \gamma_m \Big(\sum_{k \in \mathcal{U}_m} \sum_{i=1}^{n_F} \tilde{P}_{R_m,k}[i] - P_{R_T} \Big) - \sum_{m=1}^{M} \sum_{k \in \mathcal{U}_m} \sum_{i=1}^{n_F} \beta[i](s_{m,k}[i] - 1) \\
&+ \sum_{m=1}^{M} \sum_{k \in \mathcal{U}_m} \sum_{i=1}^{n_F} s_{m,k}[i] \Big(\mu_{m,k}[i] \widetilde{R}_{m,k}^{1st}[i] + \nu_{m,k}[i] \widetilde{R}_{m,k}^{2nd}[i] \Big),
\end{aligned} \tag{2.26}$$

where $\lambda \geq 0$ is the Lagrange multiplier corresponding to the power constraint at the BS. $\boldsymbol{\gamma}$ is the Lagrange multiplier vector corresponding to the individual relay power constraints with elements $\gamma_m \geq 0$, $m \in \{1, \ldots, M\}$. $\boldsymbol{\beta}$ is the Lagrange multiplier vector associated with the subcarrier usage constraints with elements $\beta[i] \geq 0$, $i \in \{1, \ldots, n_F\}$. $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are the Lagrange multiplier vectors for constraints C9 and C10 in (2.23) with elements $\mu_{m,k}[i]$ and $\nu_{m,k}[i]$. The boundary constraints C6 and C7 will be absorbed into the KKT conditions when deriving the optimal solution in Section 2.4.3.

Thus, the dual problem is given by

$$\min_{\lambda,\, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\nu} \geq 0} \; \max_{\mathcal{P}, \mathcal{R}, \mathcal{S}, z_{m,k}[i]} \mathcal{L}(\lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathcal{P}, \mathcal{R}, \mathcal{S}, z_{m,k}[i]). \tag{2.27}$$

In general, the above dual problem can be unbounded if $z_{m,k}[i] \to \infty$. Consider the

parts of the dual function in the inner maximization which are related to $z_{m,k}[i]$:

$$\begin{aligned}&\max_{z_{m,k}[i]} \sum_{m=1}^{M} \sum_{k\in\mathcal{U}_m} \sum_{i=1}^{n_F} (w_k - (\mu_{m,k}[i] + \nu_{m,k}[i]))z_{m,k}[i] \\ = &\begin{cases} 0 & \text{if } \mu_{m,k}[i] + \nu_{m,k}[i] = w_k \\ \infty & \text{otherwise} \end{cases}. \end{aligned} \tag{2.28}$$

In order to have a bounded dual function, the Lagrange multipliers $\mu_{m,k}[i]$ and $\nu_{m,k}[i]$ must satisfy $\mu_{m,k}[i] + \nu_{m,k}[i] = w_k$. Thus, the dual problem is simplified to

$$\min_{\lambda,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\mu},\geq 0} \max_{\mathcal{P},\mathcal{R},\mathcal{S}} \quad \widetilde{\mathcal{L}}(\lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \mathcal{P}, \mathcal{R}, \mathcal{S}, z_{m,k}[i]), \tag{2.29}$$

where $\widetilde{\mathcal{L}}(\lambda,\ \boldsymbol{\beta},\ \boldsymbol{\gamma},\ \boldsymbol{\mu},\ \mathcal{P}, \mathcal{R}, \mathcal{S}) = \mathcal{L}(\lambda, \boldsymbol{\beta},\ \boldsymbol{\gamma},\ \boldsymbol{\mu},\ \boldsymbol{\nu},\ \mathcal{P}, \mathcal{R}, \mathcal{S}, z_{m,k}[i])|_{\nu_{m,k}[i]=w_k-\mu_{m,k}[i]}$. Note that the auxiliary variables $z_{m,k}[i]$ vanish when we set $\nu_{m,k}[i] = w_k - \mu_{m,k}[i]$.

2.4.3 Dual Decomposition and Sub-Problem Solution

By dual decomposition, the dual problem is decomposed into two parts (nested loops): the first part (inner loop) consists of $M \times n_F$ sub-problems with identical structure; the second part (outer loop) is the master dual problem. The dual problem can be solved iteratively where in each iteration each relay solves n_F local sub-problems (inner loop) by utilizing the local CSI and exchanges some information with the BS which solves the master problem (outer loop) with the gradient method.

The sub-problem to be solved by relay m is given by

$$\max_{\mathcal{P},\mathcal{R},\mathcal{S}} \quad \widetilde{\mathcal{L}}_m(\lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \mathcal{P}, \mathcal{R}, \mathcal{S}) \tag{2.30}$$

for a fixed set of Lagrange multipliers where $\widetilde{\mathcal{L}}_m(\lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \mathcal{P}, \mathcal{R}, \mathcal{S}) =$

$$\sum_{k\in\mathcal{U}_m}\sum_{i=1}^{n_F} s_{m,k}[i]\Big(\mu_{m,k}[i]\widetilde{R}^{1st}_{m,k}[i] + \nu_{m,k}[i]\widetilde{R}^{2nd}_{m,k}[i]\Big) + \lambda P_{B_T} + \gamma_{m,k}[i]P_{R_T}$$
$$- \sum_{k\in\mathcal{U}_m}\sum_{i=1}^{n_F}\beta[i]\Big(s_{m,k}[i]-1\Big) - \lambda\Big(\sum_{k\in\mathcal{U}_m}\sum_{i=1}^{n_F}\tilde{P}_{BR_m,k}[i]\Big) - \gamma_m\sum_{k\in\mathcal{U}_m}\sum_{i=1}^{n_F}\tilde{P}_{R_m,k}[i]. \quad (2.31)$$

Note that the above sub-problem is the inner loop optimization in (2.29).

Using standard optimization techniques and the KKT conditions, the optimal power allocation for both hops for user k via relay m on subcarrier i are obtained as

$$P^*_{BR_m,k}[i] = \left[\frac{\mu_{m,k}[i]}{(2\ln(2))\lambda} - \frac{1}{\alpha^*_{BR_m,k}[i]\lambda_{\max_{BR_m}}[i]}\right]^+, \quad (2.32)$$

$$P^*_{R_m,k}[i] = \left[\frac{\sqrt{\Omega_{m,k}[i](\Omega_{m,k}[i]\gamma_m\ln(2) + 2\Xi_{m,k}[i]\nu_{m,k}[i])}}{2\Xi_{m,k}[i]\sqrt{\gamma_m}\sqrt{\ln(2)}} - \frac{2(1-\alpha^*_{R_m,k}[i])\sigma_e^2 + \Omega_{m,k}[i]}{2\Xi_{m,k}[i]}\right]^+, \quad (2.33)$$

where $\Xi_{m,k}[i] = (1-\alpha^*_{R_m,k}[i])\sigma_e^2\Big[\Omega_{m,k}[i] + \sigma_e^2(1-\alpha^*_{R_m,k}[i])\Big]$ and $\Omega_{m,k}[i] = F^{-1}_{\chi_2}(\varepsilon, i)\alpha^*_{R_m,k}[i]$. The optimal power allocations in (2.32) and (2.33) have the form of *multi-level* water-filling. It can be observed that the dual variable $\mu_{m,k}[i]$ affects the power allocation by changing the water-level, $\frac{\mu_{m,k}[i]}{(2\ln(2))\lambda}$, of user k for satisfying constraint C9 in (2.23). On the other hand, the water level of each user in (2.33) depends not only on his/her priority via $\nu_{m,k}[i]$, but also on the CSIT error statistic of the desired channel and the required channel outage probability, i.e., $F^{-1}_{\chi_2}(\varepsilon, i)\alpha^*_{R_m,k}[i]$.

In order to obtain the optimal subcarrier allocation, we take the derivative of the sub-problem with respect to $s_{m,k}[i]$, which yields $\frac{\partial\widetilde{\mathcal{L}}_m(\lambda,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\mu},\mathcal{P},\mathcal{R},\mathcal{S})}{\partial s_{m,k}[i]} = A_{m,k}[i] - \beta[i]$, where $A_{m,k}[i] \geq 0$ is the marginal benefit [76] for allocating subcarrier i to user k via

relay m and is given by $A_{m,k}[i] =$

$$\frac{\nu_{m,k}[i]}{2}\left(\log_2\left(1+\Gamma^*_{R_m,k}[i]\right)-\frac{\Gamma^*_{R_m,k}[i]/\left(\ln(2)(1+\Gamma^*_{R_m,k}[i])\right)}{1+\sigma_e^2 P^*_{R_m,k}[i](1-\alpha^*_{R_m,k}[i])}-\log_2\left(1+\frac{\alpha^*_{R_m,k}[i]\Lambda_E[i]}{1-\alpha^*_{R_m,k}[i]}\right)\right)$$
$$+\frac{\mu_{m,k}[i]}{2}\left(\log_2\left(1+\Gamma^*_{BR_m,k}[i]\right)-\frac{\Gamma^*_{BR_m,k}[i]}{\ln(2)(1+\Gamma^*_{BR_m,k}[i])}-\log_2\left(1+\frac{\alpha^*_{BR_m,k}[i]\Lambda_E[i]}{1-\alpha^*_{BR_m,k}[i]}\right)\right) \tag{2.34}$$

for $\alpha^*_{BR_m,k}[i]=\alpha^*_{R_m,k}[i]$, where $\Gamma^*_{BR_m,k}[i]=\Gamma_{BR_m,k}[i]\Big|_{P_{BR_m,k}[i]=P^*_{BR_m,k}[i]}$ and $\Gamma^*_{R_m,k}[i]=\Gamma_{R_m,k}[i]\Big|_{P_{R_m,k}[i]=P^*_{R_m,k}[i]}$. On the contrary, if a user has good channel conditions with positive secrecy data rate on subcarrier i, he/she can provide a higher marginal benefit to the system. Thus, the optimal subcarrier selection determined by relay m on subcarrier i is given by

$$s^*_{m,k}[i]=\begin{cases}1 & \text{if } A_{m,k}[i]=\max\limits_{a,b} \quad A_{a,b}[i]\geq\beta[i]\geq 0\\ 0 & \text{otherwise}\end{cases}. \tag{2.35}$$

The dual variable $\beta[i]\geq 0$ acts as the global price in using subcarrier i in the system. Only users who can provide large marginal benefits to the system are considered for selection by the resource allocator. $A_{m,k}[i]\geq 0$ has the physical meaning that users with negative secrecy data rate on subcarrier i are not selected as they can only provide a negative marginal benefit to the system. Note that each subcarrier will be used for serving only one user eventually. Finally, the optimal transmitted packet data rate $R^{*data}_{m,k}[i]$ and secrecy data rate $R^{*sec}_{m,k}[i]$ are obtained by substituting (2.32), (2.33) into the equivalent packet data rate and secrecy data rate in Lemma 2.1 for the subcarrier with $s^*_{m,k}[i]=1$.

2.4.4 Solution of the Master Problem

For solving the master problem at the BS, each relay forwards the local resource allocation policies (i.e., $\mathcal{P}, \mathcal{R}$, and $\mathcal{S}$) to the BS. Since the dual function is differentiable, the gradient method can be used to solve the minimization of the master problem in (2.27). The solution is given by

$$\lambda(t+1) = \Big[\lambda(t) - \xi_1(t) \times (P_{B_T} - \sum_{m=1}^{M} \sum_{k \in \mathcal{U}_m} \sum_{i=1}^{n_F} P_{BR_m,k}[i] s_{m,k}[i])\Big]^+, \tag{2.36}$$

$$\gamma_m(t+1) = \Big[\gamma_m(t) - \xi_2(t) \times (P_{R_T} - \sum_{k \in \mathcal{U}_m} \sum_{i=1}^{n_F} P_{R_m,k}[i] s_{m,k}[i])\Big]^+, \forall m, \tag{2.37}$$

$$\mu_{m,k}[i](t+1) = \Big[\mu_{m,k}[i](t) - \xi_3(t) \times s_{m,k}[i](\widetilde{R}^{1st}_{m,k}[i] - \widetilde{R}^{2nd}_{m,k}[i])\Big]^+_{\mathbb{U}_{m,k}[i]}, \forall m, k, i, \tag{2.38}$$

$$\beta[i](t+1) = \Big[\beta[i](t) - \xi_4(t) \times (1 - \sum_{m=1}^{M} \sum_{k \in \mathcal{U}_m} \sum_{i=1}^{n_F} s_{m,k}[i])\Big]^+, \forall i, \tag{2.39}$$

where $t \geq 0$ is the iteration index and $\xi_a(t)$, $a \in \{1, 2, 3, 4\}$, are positive step sizes. $\nu_{m,k}[i]$ can be obtained from $\nu_{m,k}[i] = \Big[w_k - \mu_{m,k}[i]\Big]^+$. $\mathbb{U}_{m,k}[i]$ in (2.38) denotes the projection operator on the feasible set $\mathbb{U}_{m,k}[i] = \{\mu_{m,k}[i] \Big| 0 \leq \mu_{m,k}[i] \leq w_k\}$. Since the transformed problem is convex in nature, it is guaranteed that the algorithm converges to the optimal solution if the chosen step sizes satisfy the general conditions stated in [77, Section 1.2]. In summary, the master problem adjusts the water-levels of (2.32) and (2.33) through the gradient update equations (2.36) and (2.37) until the individual power constraints of the BS and the relays are satisfied, respectively. Finally, (2.38) reduces the difference between the capacity of user k in the first and second hops, which corresponds to the selection of the minimum capacity in (2.15). We note that there is no intra-cell/inter-sector interference in the considered system since the resource allocation algorithm is applied to the entire cell and all users in all sectors are competing for resources. By combining (2.35) and (2.39), it can be

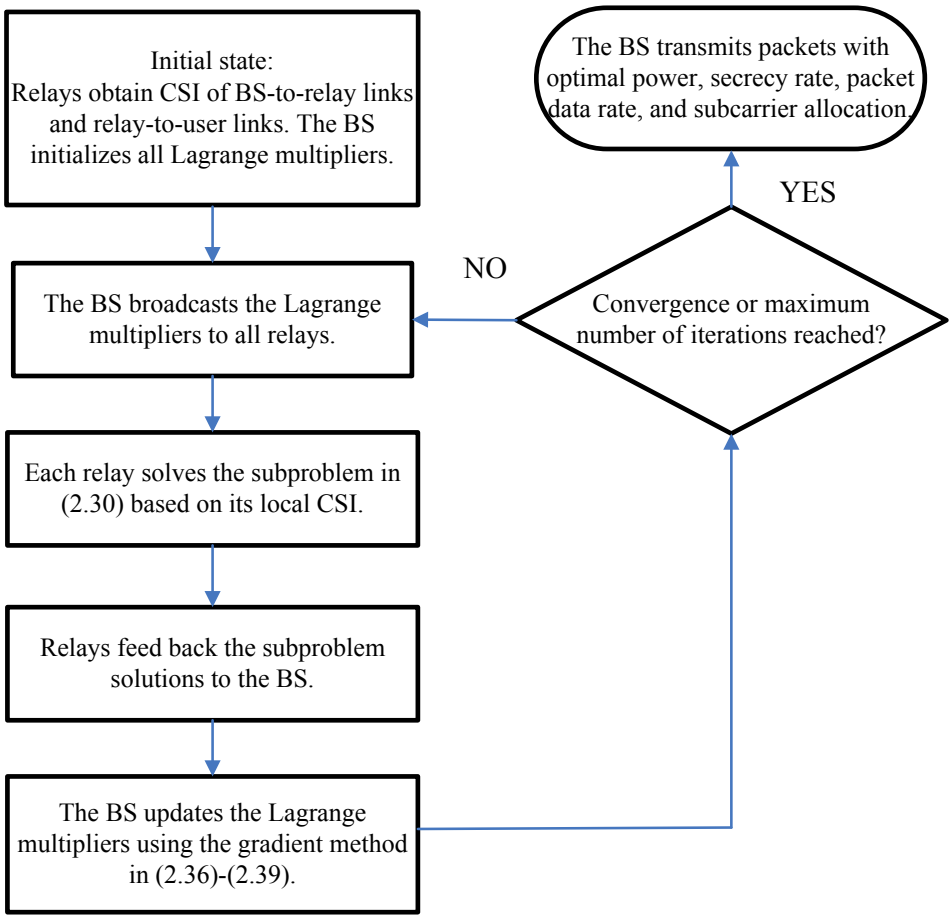


Figure 2.3: A flow chart of the proposed iterative distributed resource allocation and scheduling algorithm.

shown that, for the optimal solution, there is no time-sharing between the assigned subcarriers. The overall distributed algorithm is illustrated in Figure 2.3.

2.5 Results

In this section, we evaluate the system performance using simulations. A cell is modeled as two concentric ring-shaped discs where the outer boundary has a radius of 1 km and the inner boundary a radius of 0.5 km, cf. Figure 2.1. The number of subcarriers is $n_F = 128$ with carrier center frequency 2.5 GHz, bandwidth $\mathcal{B} = 5$ MHz,

and $w_k = 1, \forall k$. Each subcarrier has a bandwidth of 39 kHz and a noise variance of $N_0 = -128$ dBm. The 3rd Generation Partnership Project (3GPP) path loss model is used [78] with a reference distance of $d_0 = 35$ m. There are $M = 3$ relay stations in the cell which are equally distributed at the inner cell boundary for assisting the transmission. The K desired users are uniformly distributed between 0.5 km and the cell boundary at 1 km. We assume that the eavesdropper is located 35 m away from the BS which represents an unfavorable scenario, since all the desired users are farther away from the BS than the eavesdropper. The small scale fading coefficients of the BS-to-user and BS-to-eavesdropper links are modeled as i.i.d. Rayleigh random variables. On the other hand, a strong line of sight communication channel between the BS and the relays is expected since they are placed in relatively high positions in practice and the number of blockages between them are limited. Hence, the small scale fading coefficients of the BS-to-relay links are modelled as i.i.d. Rician random variables with Rician factor $\kappa = 6$ dB. The channel ESR is set to $\frac{\sigma_e^2}{\sigma_{h_{R_m,k}}^2} = 0.05$, unless further specified. The target secrecy outage probability and channel outage probability are set to $\delta = 0.05$ and $\varepsilon = 0.05$, respectively, unless further specified. We assume that the maximum transmit power at each transmission device is P_T, i.e., the BS and the relay have a maximum transmit power of $P_{R_T} = P_{B_T} = P_T$. The average secrecy outage capacity is obtained by counting the number of packets securely delivered to and decoded by the users averaged over both the macroscopic and microscopic fading.

2.5.1 Convergence of Distributed Iterative Algorithm

Figure 2.4 illustrates the evolution of the Lagrange multiplier λ of the distributed iterative algorithm over time for different maximum transmit powers P_T with $K = 15$

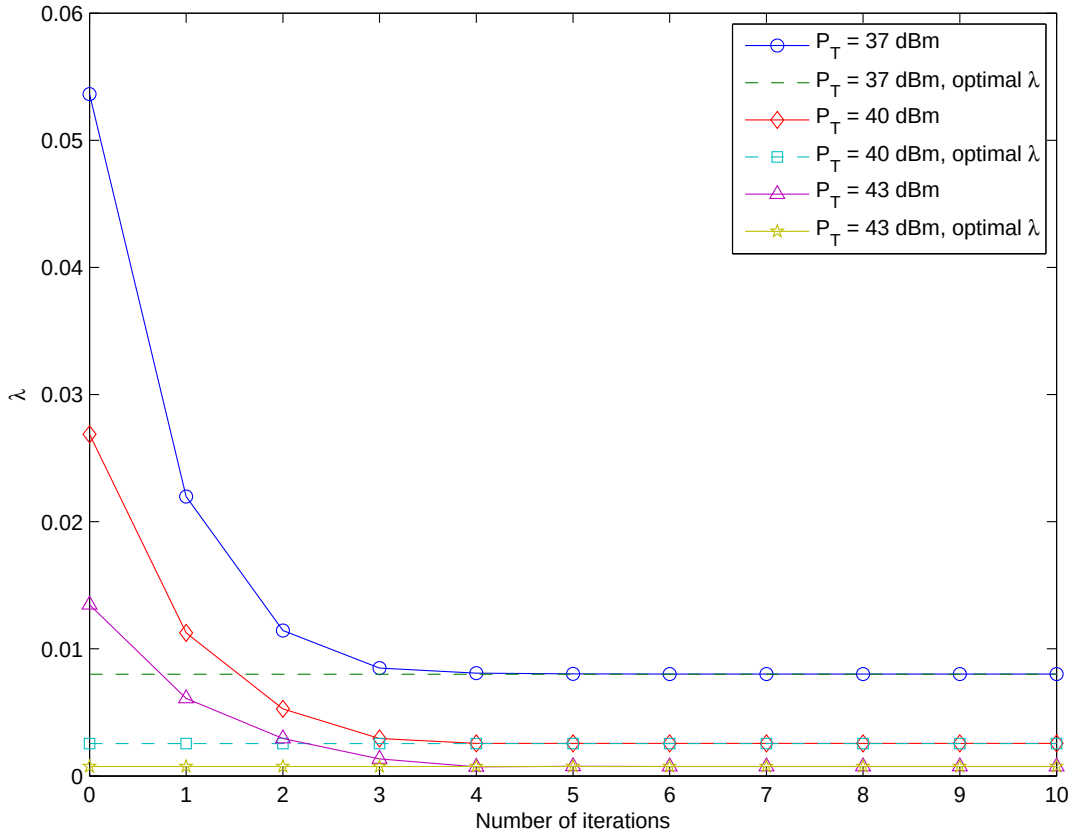


Figure 2.4: Lagrange multiplier λ versus number of iterations with $K = 15$ users and $M = 3$ relays for different transmit power levels. The BS and each relay are equipped with $N_T = 9$ antennas. There are $N_E = 2$ receive antennas at the eavesdropper.

users and $M = 3$ relays. Both the BS and each relay have $N_T = 9$ transmit antennas, while the eavesdropper has $N_E = 2$ receive antennas. Positive constant step sizes $\xi_1(t), \xi_2(t), \xi_3(t)$, and $\xi_4(t)$, which were optimized for fast convergence, were adopted in (2.36)-(2.39). The result in Figure 2.4 was averaged over 10000 independent adaptation processes. For the considered transmit power values, it can be observed that the distributed iterative algorithm converges fast and typically achieves at least 95% of the optimal value within 5 iterations.

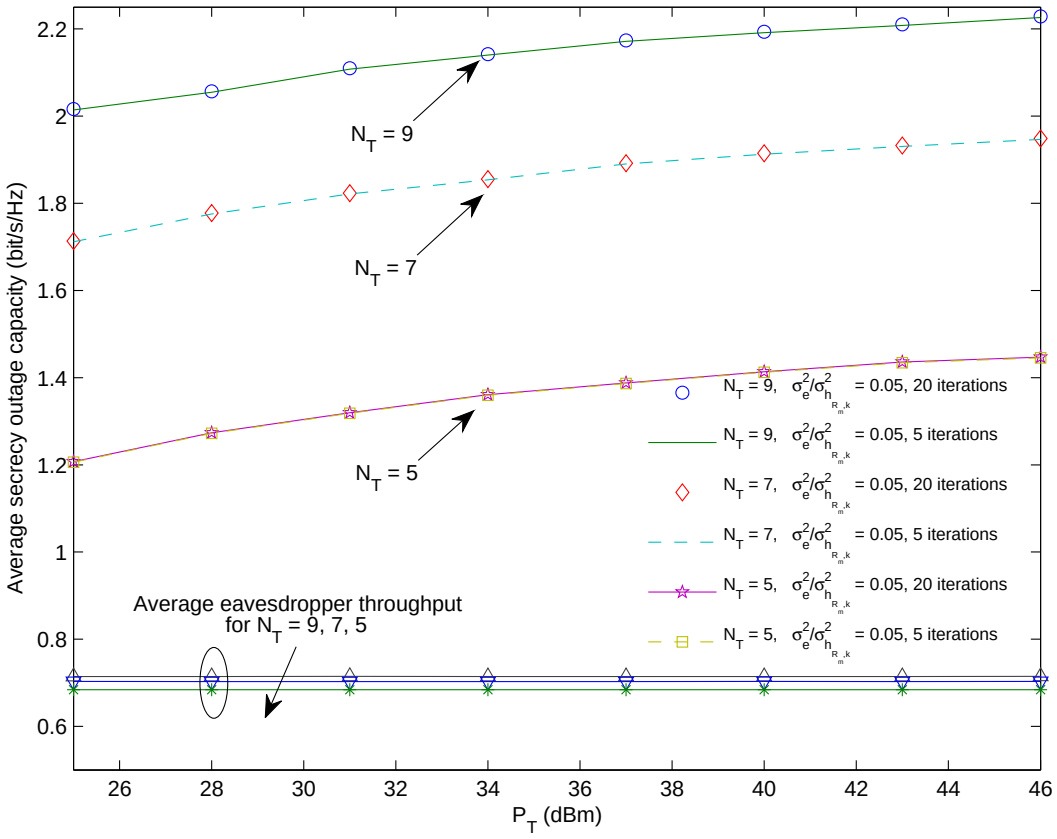


Figure 2.5: Average secrecy outage capacity versus transmit power for different numbers of transmit antennas N_T. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS.

2.5.2 Average Secrecy Outage Capacity versus Transmit Power and ESR

Figure 2.5 illustrates the average secrecy outage capacity and the throughput of the eavesdropper versus the total transmit power for $K = 15$ users for different numbers of transmit antennas N_T at both the BS and the relays. The eavesdropper is equipped with $N_E = 2$ antennas. The number of iterations for the proposed iterative resource allocation algorithm is 5 and 20. It can be seen that the performance difference

between 5 iterations and 20 iterations is negligible which confirms the practicality of our proposed iterative resource allocation algorithm. On the other hand, for a better illustration of the effectiveness of the artificial noise generation, Figure 2.5 also includes the performance of the eavesdropper in terms of average throughput. As observed in Lemma 2.1, the average throughput between the BS and the eavesdropper does not scale with the transmit power in the high transmit power regime due to the artificial noise introduced by the BS, despite the fact that the eavesdropper is located closer to the BS than all the desired users. On the other hand, it can be observed that although the imperfect CSI has a negative impact on the average secrecy outage capacity due to the artificial noise leakage, the system performance scales with the transmit power thanks to the proposed optimization technique. Besides, it can be observed that an increasing number of transmit antennas N_T benefits the desired users in terms of average secrecy outage capacity. Yet, there is a diminishing return when N_T is large due to the *channel hardening* effect [28] in the desired channels. On the contrary, the throughput of the eavesdropper is limited by artificial noise and the performance gain achieved at the eavesdropper due to increasing N_T is marginal.

Figure 2.6 illustrates the average secrecy outage capacity versus ESR $\frac{\sigma_e^2}{\sigma_{h_{R_m,k}}^2}$ for $K = 15$ users with different numbers of receive antennas at the eavesdropper and different numbers of transmit antennas at the BS and relays. The number of iterations is set to 5. It can be observed that as the estimation error increases, the system performance decreases since the CSI available for resource allocation becomes less accurate, and the resource allocation has to be more conservative in order to satisfy the outage requirements of the selected users. Besides, when N_T is not significantly larger than N_E, the average secrecy outage capacity is comparatively small for moderate ESRs values. This is because the resource allocator shuts down some

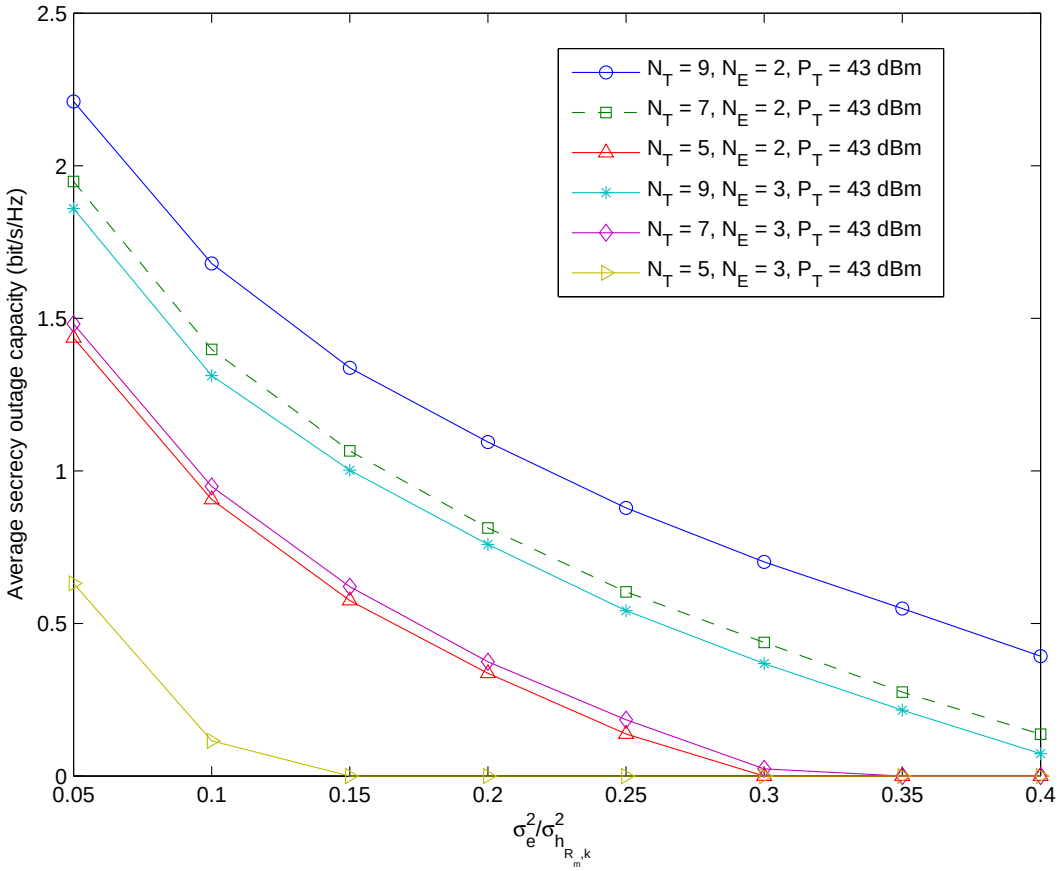


Figure 2.6: Average secrecy outage capacity versus ESR $\frac{\sigma_e^2}{\sigma_{\hat{h}_{Rm,k}}^2}$ for different numbers of transmit antennas N_T and eavesdropper antennas N_E.

subcarriers if the channel conditions of all the users are not good enough to guarantee secure communication, which results in a low average system performance. On the other hand, Figure 2.6 suggests that if the number of transmit antennas N_T is large enough compared to the number of eavesdropper receive antennas N_E, e.g. $N_T = 9$ and $N_E = 2$, the proposed resource allocation scheme is able to guarantee an average secrecy outage capacity of 0.5 bit/s/Hz (corresponding to 2.5 Mbps for a 5 MHz bandwidth) even in high ESR (e.g. $\frac{\sigma_e^2}{\sigma_{\hat{h}_{Rm,k}}^2} = 0.35$, estimation error of 35%), while satisfying both the channel outage and secrecy outage requirements.

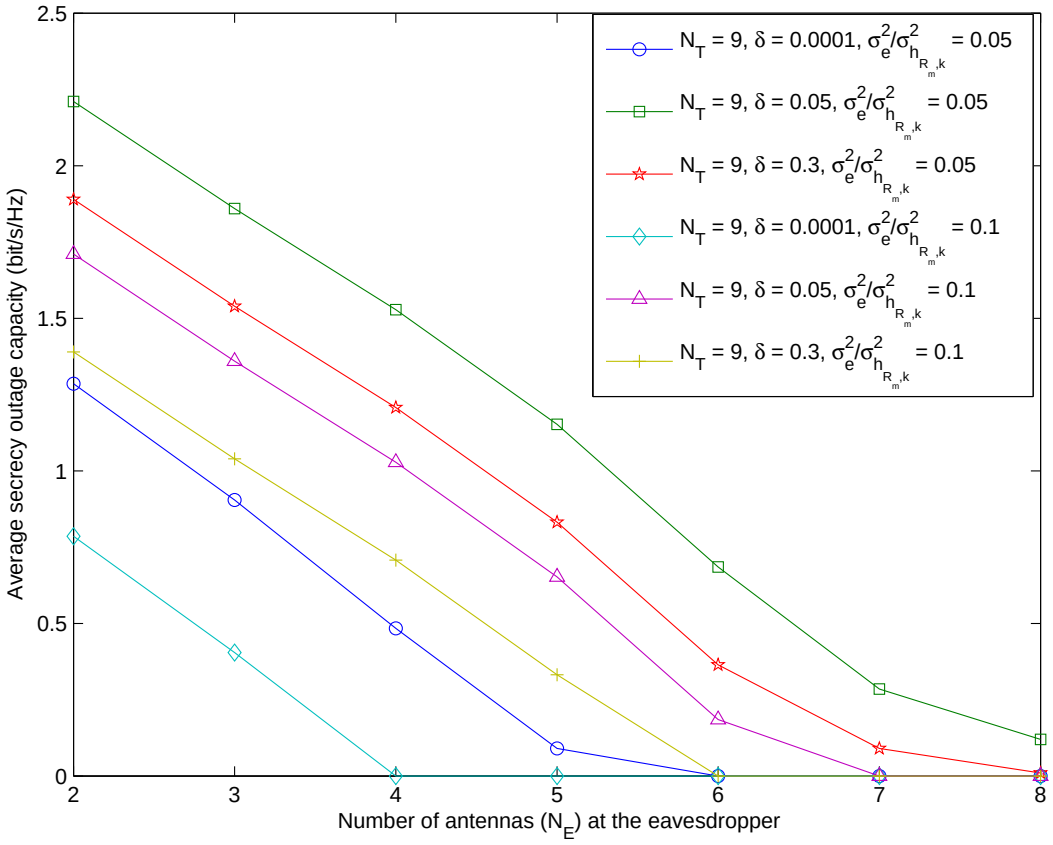


Figure 2.7: Average secrecy outage capacity versus the number of antennas N_E employed at the eavesdropper for different ESR $\frac{\sigma_e^2}{\sigma_{h_{R_m,k}}^2}$ and different secrecy outage requirements δ. $N_T = 9$ antennas at the BS and relays.

2.5.3 Average Secrecy Outage Capacity versus N_E

Figure 2.7 depicts the average secrecy outage capacity versus the number of receive antennas N_E employed at the eavesdropper for different secrecy outage requirements and ESRs. There are $K = 15$ users and $N_T = 9$ transmit antennas at the BS and the relays. The number of iterations for the iterative algorithm is 5. It can be observed that the secrecy outage capacity decreases as N_E increases, since more of the transmitted power has to be devoted to the artificial noise generation for degrading the

channels of the eavesdropper, which results in less transmit power for information transmission. On the other hand, we observe that a more stringent secrecy outage probability requirement does not necessarily lead to a higher average secrecy outage capacity. This is because a larger fraction of power has to be allocated to the artificial noise for degrading the channel of the eavesdropper and less power is available for information transmission. Yet, a less stringent secrecy outage probability requirement may also lead to a unsatisfactory system performance since the eavesdropper has a higher chance in decoding the desired information. As observed in Figure 2.7, there exist an optimal secrecy outage requirement δ for each ESR value, which maximizes the overall system performance. However, optimizing the value of δ in the physical layer may require further information from the application layer (e.g. tolerable information leakage of a particular data type such as video or email), which is beyond the scope of this thesis.

2.5.4 Average Secrecy Outage Capacity versus Number of Users

Figure 2.8 depicts the average secrecy outage capacity versus the number of users for different numbers of transmit antennas for $P_T = 43$ dBm. The number of iterations is 5. It can be observed that the average secrecy outage capacity grows with the number of users since the proposed resource allocation and scheduling algorithm is able to exploit MUD, despite the existence of the eavesdropper. However, for large N_T, the system performance scales with the number users slowly. Indeed, since a large number of transmit antennas reduce channel fluctuations in the desired user channel and cause *channel hardening*, they decrease the potentially achievable MUD gain in the subcarrier allocation process. On the other hand, the performance of the

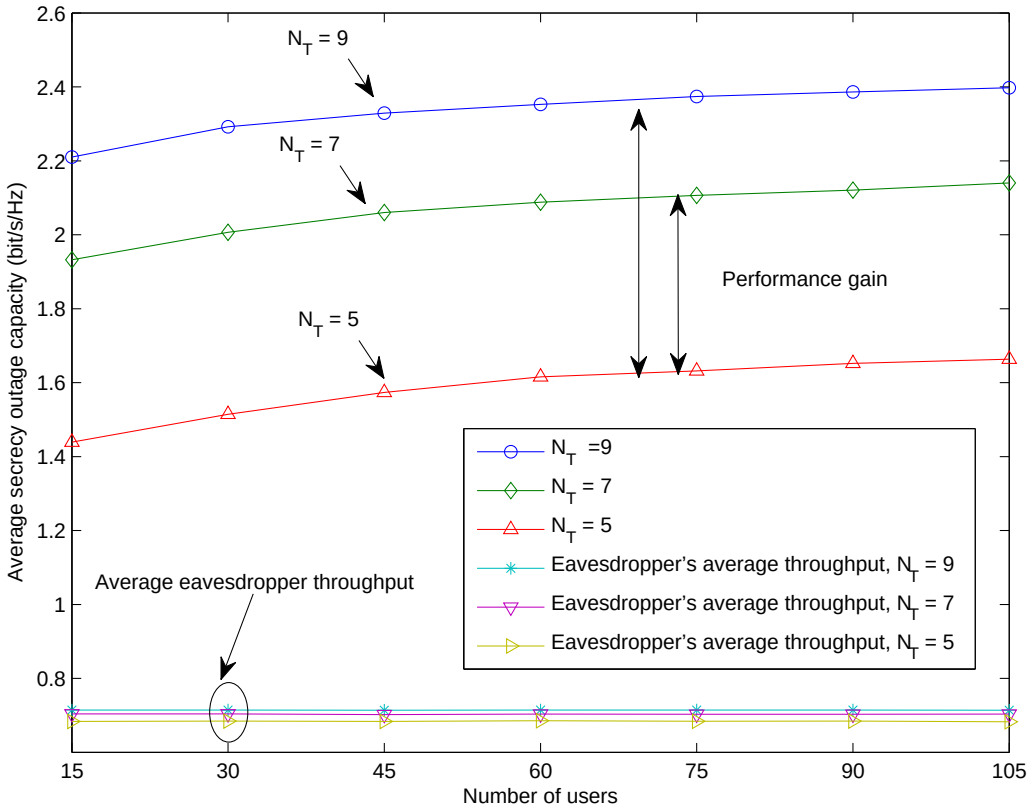


Figure 2.8: Average secrecy outage capacity versus the number of desired users for different numbers of transmit antennas N_T at the BS with a total transmit power $P_T = 43$ dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m away from the BS. The double arrows indicate the performance gain achieved by an increasing number of transmit antennas N_T.

eavesdropper does not scale with the number of users since the channels between the eavesdropper and the desired users are generally uncorrelated.

Remark 2.2 *Simulation results for when the eavesdropper is located close to a relay are not shown since the resulting system performance is close to that of the considered case where the eavesdropper is located close to the BS. This is because when the capacity upper bound of the eavesdropper in (2.16) is adopted for resource allocation, a large amount of artificial noise is generated to combat the eavesdropper which*

saturates the throughput of the eavesdropper, cf. Figure 2.5.

2.6 Conclusions

In this chapter, we formulated the resource allocation and scheduling design for OFDMA DF relaying systems as a non-convex and combinatorial optimization problem, where a multiple antenna eavesdropper, artificial noise generation for secure communication, and the negative effect of imperfect CSIT were taken into consideration. By relaxing the combinatorial subcarrier allocation constraints, the considered problem was transformed into a convex problem. An efficient iterative and distributed resource allocation algorithm with closed-form power, secrecy data rate, packet data rate, and subcarrier allocation requiring only local CSI at each relay was derived by dual decomposition. Simulation results not only showed that the performance of the proposed algorithm converges to the optimal performance within a small number of iterations, but also demonstrated the achievable secrecy outage capacity when the eavesdropper is closer to the BS/relay than the desired users.

Chapter 3

Energy-Efficient Resource Allocation for Secure OFDMA Systems

3.1 Introduction

In the previous chapter, we have introduced a novel resource allocation and scheduling scheme for secure OFDMA cellular networks. However, this does not provide a full picture of an efficient secure communication systems since energy efficiency was not taken into account in the resource allocation process. Indeed, the increasing interest in multimedia services has led to a tremendous demand for high data rate communications with certain guaranteed QoS properties. This demand has significant financial implications for service providers because of the rapidly increasing energy consumption for achieving the required QoS. As a result, energy efficient system designs, which adopt *energy efficiency* (bit-per-Joule) as the performance metric, have recently received much attention in both industry and academia [79]-[83]. In [79] and [80], power allocation algorithms for energy efficient multi-carrier systems were studied assuming a static circuit power consumption. In [81] and [82], energy efficient link adaptation for a sum rate-dependent dynamic circuit power consumption was considered. However, if user selection and link adaptation are jointly optimized, the algorithms proposed in [79]-[82] may no longer be applicable. In [83], a risk-return model was proposed for energy-efficient power allocation in multi-carrier systems.

Yet, the proposed algorithm is suboptimal and does not achieve the maximum energy efficiency. Besides, the energy efficiencies of the systems in [62]-[72] are unknown and the optimization of the amount of power devoted to artificial noise generation for maximization of the energy efficiency remains an unsolved problem.

Motivated by the aforementioned observations, we formulate the resource allocation problem for energy efficient secure communication in OFDMA systems with artificial noise generation as an optimization problem. By exploiting the properties of fractional programming, the considered non-convex optimization problem is transformed to an equivalent convex optimization problem with a tractable solution, which can be obtained with an iterative algorithm. In each iteration, the transformed problem is solved by using dual decomposition, and closed-form power, secrecy data rate, and subcarrier allocation polices maximizing the energy efficiency are provided. The proposed algorithm does not only converge fast to the optimal solution, but also fulfills the secrecy outage tolerance requirements of the users.

3.2 OFDMA Downlink Network Model

In this section, we present the adopted channel and signal models.

3.2.1 Channel Model

We consider an OFDMA downlink network which consists of a BS with N_T antennas, an eavesdropper with N_E antennas, and K mobile users equipped with a single antenna, cf. Figure 3.1. We assume that $N_T > N_E$ to enable secure communication. The eavesdropper is passive and its goal is to decode the information transmitted by the BS without causing interference to the communication channels. The impulse responses of all channels are assumed to be time-invariant (slow fading). We consider

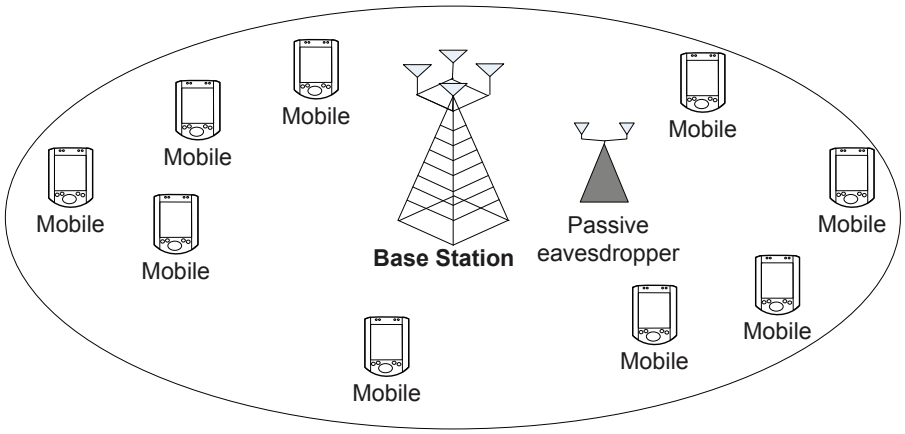


Figure 3.1: Illustration of an OFDMA downlink network. There are one BS with $N_T = 4$ antennas, $K = 9$ desired users equipped with a single antenna, and one eavesdropper with $N_E = 2$ antennas. For an effective eavesdropping, the eavesdropper chooses a location closer to the BS compared to the locations of all the desired users.

an OFDMA system with n_F subcarriers. The received symbols at user k and the eavesdropper on subcarrier $i \in \{1, \ldots, n_F\}$ are given by, respectively,

$$y_k[i] \;=\; \mathbf{h}_k[i]\mathbf{x}_k[i] + n[i] \quad \text{and} \quad \mathbf{y}_E[i] = \mathbf{G}[i]\mathbf{x}_k[i] + \mathbf{e}[i], \tag{3.1}$$

where $\mathbf{x}_k[i] \in \mathbb{C}^{N_T \times 1}$ denotes the transmitted symbol vector. $\mathbf{h}_k[i] \in \mathbb{C}^{1 \times N_T}$ is the channel vector between the BS and user k on subcarrier i and $\mathbf{G}[i] \in \mathbb{C}^{N_E \times N_T}$ is the channel matrix between the BS and the eavesdropper on subcarrier i. Both variables, $\mathbf{h}_k[i]$ and $\mathbf{G}[i]$, include the effects of path loss and multipath fading. $n[i]$ is the AWGN in subcarrier i at user k with distribution $\mathcal{CN}(0, N_0)$, where N_0 is the noise power spectral density. $\mathbf{e}[i] \in \mathbb{C}^{N_E \times 1}$ is the AWGN vector in subcarrier i at the eavesdropper and each entry of $\mathbf{e}[i]$ has distribution $\mathcal{CN}(0, N_0)$. We assume that the CSI (path loss information and multipath fading) of the desired users is perfectly known at the BS due the accurate channel measurements. On the other hand, we assume that the BS knows only the number of antennas N_E employed by

the eavesdropper[5] and the associated channel distribution with an unknown variance. Since the CSI of the eavesdropper is unavailable at the BS, in order to secure the desired wireless communication links, an *artificial noise* signal is generated at the BS to degrade the channels between the BS and the eavesdropper.

Artificial Noise Generation: The BS chooses $\mathbf{x}_k[i]$ as the linear combination of the information bearing signal $u_k[i]$ and the artificial noise signal $\mathbf{v}_k[i]$, i.e.,

$$\mathbf{x}_k[i] = \underbrace{\mathbf{b}_k[i]u_k[i]}_{\text{Desired Signal}} + \underbrace{\mathbf{V}_k[i]\mathbf{v}_k[i]}_{\text{Artificial Noise}}, \tag{3.2}$$

where $\mathbf{v}_k[i]$ is a vector of i.i.d. complex Gaussian random variables with variance $\sigma_v^2[i]$. Since $\mathbf{h}_k[i]$ is known at the BS, without loss of generality, we define an orthogonal basis $\mathbf{V}_k[i]$ for the null space of $\mathbf{h}_k[i]$, such that $\mathbf{h}_k[i]\mathbf{V}_k[i]\mathbf{v}_k[i] = 0$ and $\mathbf{V}_k^\dagger[i]\mathbf{V}_k[i] = \mathbf{I}$, where $\mathbf{I}$ is a $(N_T - 1) \times (N_T - 1)$ identity matrix. In other words, the artificial noise signal does not interfere with the desired users. Without loss of generality, we define the transmit power devoted to the information bearing signal for user k in subcarrier i as $p_k[i]$. Then, the SNR at user k is maximized by choosing $\mathbf{b}_k[i] = \frac{p_k[i]\mathbf{h}_k^\dagger[i]}{\|\mathbf{h}_k[i]\|}$ such that the information bearing signal lies in the range space of $\mathbf{h}_k[i]$. Hence, the received signals in (3.1) can be rewritten as

$$y_k[i] = \mathbf{h}_k[i]\mathbf{b}_k[i]u_k[i] + n[i] = p_k[i]\lambda_{\max_k}[i]u_k[i] + n[i] \quad \text{and} \tag{3.3}$$

$$\mathbf{y}_E[i] = \mathbf{G}[i]\mathbf{b}_k[i]u_k[i] + \mathbf{G}[i]\mathbf{V}_k[i]\mathbf{v}_k[i] + \mathbf{e}[i], \tag{3.4}$$

respectively, where $\lambda_{\max_k}[i]$ is the maximum eigenvalue of $\mathbf{h}_k^\dagger[i]\mathbf{h}_k[i]$. Suppose the

[5]Note that the number of antennas employed by the eavesdropper is equivalent to the eavesdropping capability of the eavesdropper. In practice, the BS may not know the number of eavesdropper antennas. Hence, the BS may assume N_E as $N_E = N_T - 1$ to ensure security by considering the worst-case scenario.

total transmit power on subcarrier i for user k is $P_k[i]$. We establish the following relationships [69]:

$$\begin{aligned} P_k[i] &= p_k[i] + (N_T - 1)\sigma_v^2[i], \quad p_k[i] = \alpha_k[i]P_k[i], \quad \text{and} \\ \sigma_v^2[i] &= \frac{(1-\alpha_k[i])P_k[i]}{N_T - 1}, \end{aligned} \tag{3.5}$$

where $0 < \alpha_k[i] \leq 1$ represents the fraction of power devoted to the information bearing signal on subcarrier i for user k.

3.3 Resource Allocation and Scheduling

In this section, we introduce the adopted system performance metric and formulate the corresponding resource allocation problem. Since the adopted approach is based on information theory, the buffers at the BS are assumed to be always full and there are no empty scheduling slots due to an insufficient number of source packets at the buffers.

3.3.1 Instantaneous Channel Capacity, Secrecy Outage, and Energy Efficiency

In this subsection, we define the adopted system performance measure. Given perfect CSI at the receiver, the maximum channel capacity between the BS and user k on subcarrier i with subcarrier bandwidth W is given by

$$C_k[i] = W \log_2 \Big(1 + \frac{p_k[i]\lambda_{\max_k}[i]}{N_0 W}\Big). \tag{3.6}$$

Without loss of generality, we normalize the received symbols at the eavesdropper by a factor $\|\mathbf{G}[i]\|$. Hence, the received symbols at the eavesdropper can be expressed as

$$\tilde{\mathbf{y}}_E[i] = \frac{\mathbf{y}_E[i]}{\|\mathbf{G}[i]\|} = \tilde{\mathbf{G}}[i]\mathbf{x}_k[i] + \tilde{\mathbf{G}}[i]\mathbf{V}_k[i]\mathbf{v}_k[i] + \tilde{\mathbf{e}}[i], \tag{3.7}$$

where $\tilde{\mathbf{G}}[i] = \frac{\mathbf{G}[i]}{\|\mathbf{G}[i]\|}$ and $\tilde{\mathbf{e}}[i] = \frac{\mathbf{e}[i]}{\|\mathbf{G}[i]\|}$. Note that the effect of the path loss between the BS and the eavesdropper is now modeled as a position dependent noise vector $\tilde{\mathbf{e}}[i]$ with variance $\frac{N_0 W}{\|\mathbf{G}[i]\|^2}$ in each entry instead of position dependent channel gains [68, 69]. We focus on the case of negligible eavesdropper noise, i.e., $\frac{N_0 W}{\|\mathbf{G}[i]\|^2} \to 0$, which physically means that the eavesdropper is much closer to the BS than the desired users. The capacity between the BS and the eavesdropper on subcarrier i under this noiseless worst case scenario is given by

$$\begin{aligned} C_E[i] &= W \log_2 \det\Big(\mathbf{I} + p_k[i]\mathbf{g}_1\mathbf{g}_1^\dagger[i](\sigma_v^2[i]\mathbf{G}_2[i]\mathbf{G}_2^\dagger[i])^{-1}\Big) \\ &= W \log_2 \Big(1 + \frac{\alpha_k[i](N_T - 1)}{1 - \alpha_k[i]}\mathbf{g}_1^\dagger[i](\mathbf{G}_2[i]\mathbf{G}_2^\dagger[i])^{-1}\mathbf{g}_1[i]\Big), \end{aligned} \tag{3.8}$$

where $\mathbf{g}_1[i] = \tilde{\mathbf{G}}[i]\mathbf{b}_k[i]$ and $\mathbf{G}_2[i] = \tilde{\mathbf{G}}[i]\mathbf{V}_k[i]$.

Therefore, the maximum achievable secrecy capacity on subcarrier i is given by the difference of the BS-to-user k channel capacity and the BS-to-eavesdropper channel capacity [68], which can be expressed as

$$C_{sec,k}[i] = (C_k[i] - C_E[i])\ 1(C_k[i] > C_E[i]). \tag{3.9}$$

If the CSI of the BS-to-eavesdropper link is available at the BS, the resource allocator can set the target *secrecy data rate* $R_k[i]$ and control the channel capacity $C_k[i]$ to

match the channel conditions via power adaptation, i.e., $R_k[i] = C_k[i] - C_E[i]$ and $C_k[i] > C_E[i]$, such that secure communication is guaranteed for secrecy data rate $R_k[i]$. However, here the eavesdropper is assumed to be passive and its CSI is not available at the BS, i.e., $C_E[i]$ is a random variable for the BS. Hence, a *secrecy outage* [70, Chapter 5] occurs whenever the target secrecy data rate $R_k[i]$ exceeds the secrecy capacity, despite the fact that we have considered the worst case scenario in (3.8). In order to model the effect of *secrecy outage*, we consider the performance in terms of the secrecy outage capacity rather than the ergodic capacity [28]. The *average secrecy outage capacity* is defined as the total average number of bit/s securely delivered to the K mobile users (averaged over multiple scheduling slots) and is given by

$$\begin{aligned} U_{sec}(\mathcal{P},\mathcal{R},\mathcal{S}) &= \sum_{k=1}^{K} w_k \sum_{i=1}^{n_F} s_k[i] R_k[i] \mathcal{E}_{\tilde{\mathbf{G}}[i]}\Big\{ \mathbf{1}\Big(C_k[i] - C_E[i] > R_k[i]\Big)\Big\} \\ &= \sum_{k=1}^{K} w_k \sum_{i=1}^{n_F} s_k[i] R_k[i] \Pr\Big[R_k[i] < C_k[i] - C_E[i] \Big| \mathbf{h}_k[i]\Big], \end{aligned} \tag{3.10}$$

where vector $\mathbf{h}_k[i]$ represents the CSI between the BS and user k on subcarrier i. $\mathcal{P}, \mathcal{R}$, and $\mathcal{S}$ are the power, secrecy data rate, and subcarrier allocation policies, respectively. $s_k[i] \in \{0, 1\}$ is the subcarrier allocation indicator. w_k is a positive constant provided by the upper layers, which allows the resource allocator to give different priorities to different users and to enforce certain notions of fairness. On the other hand, for designing a resource allocation algorithm for energy efficient communication, it is important to include the total power consumption in the optimization objective function. Thus, we model the power dissipation in the system as the sum of one

static term and two dynamic terms which can be expressed as [81, 82]

$$U_{TP}(\mathcal{P},\mathcal{R},\mathcal{S}) \;=\; P_C + \underbrace{\sum_{k=1}^{K}\sum_{i=1}^{n_F} P_k[i]s_k[i]}_{\text{Power amplifier}} + \underbrace{\delta\sum_{k=1}^{K}\sum_{i=1}^{n_F} s_k[i]R_k[i]}_{\text{Linear sum rate dependent power}} \quad ,(3.11)$$

where P_C is a static circuit power consumption of device electronics such as mixers, filters, and digital-to-analog converters. The middle term in (3.11) denotes the power consumption in the power amplifier. The last term[6] in (3.11) represents a linear sum rate dependent power dissipation, where the value of $\delta \geq 0$ reflects the relative importance of this term.

The *weighted energy efficiency* of the considered secure system is defined as the total average number of securely delivered bits/Joule (averaged over multiple scheduling slots)

$$U_{eff}(\mathcal{P},\mathcal{R},\mathcal{S}) \;=\; \frac{U_{sec}(\mathcal{P},\mathcal{R},\mathcal{S})}{U_{TP}(\mathcal{P},\mathcal{R},\mathcal{S})}. \quad (3.12)$$

[6]Depending on the definition of energy efficiency, the last term in (3.11) represents the back-end processing power of the transmitter only, the receivers only, or both the transmitter and receivers.

3.3.2 Optimization Problem Formulation

The optimal power allocation policy, $\mathcal{P}^*$, secrecy data rate allocation policy, $\mathcal{R}^*$, and subcarrier allocation policy, $\mathcal{S}^*$, can be obtained by solving

$$\begin{aligned}
&\max_{\mathcal{P},\mathcal{R},\mathcal{S},\alpha_k[i]} \; U_{eff}(\mathcal{P},\mathcal{R},\mathcal{S}) \\
\text{s.t. } &\text{C1: } \Pr\left[R_k[i] \geq C_k[i] - C_E[i] \middle| \mathbf{h}_k[i]\right] \leq \varepsilon, \; \forall k, i, \\
&\text{C2: } \sum_{k=1}^{K}\sum_{i=1}^{n_F} P_k[i]s_k[i] \leq P_t, \quad \text{C3: } \sum_{k=1}^{K}\sum_{i=1}^{n_F} s_k[i]R_k[i] \geq r, \\
&\text{C4: } \sum_{k=1}^{K} s_k[i] \leq 1, \quad \forall i, \qquad \text{C5: } P_k[i] \geq 0, \; \forall i, k, \\
&\text{C6: } s_k[i] = \{0,1\}, \quad \forall i, k, \quad \text{C7: } 0 < \alpha_k[i] \leq 1, \; \forall i, k.
\end{aligned} \tag{3.13}$$

In C1, ε denotes the maximum tolerable secrecy outage probability, i.e., C1 is a QoS metric for communication security. C2 is a transmit power constraint for the BS. The value of P_t puts a limit on the power consumption of the power amplifier to limit the amount of out-of-cell interference. C3 specifies the minimum system secrecy outage capacity requirement r. Note that although variable r in C3 is not an optimization variable in the formulation, a balance between energy efficiency and aggregate system secrecy outage capacity can be struck by varying r. C4 and C6 are imposed to guarantee that each subcarrier is used by one user only. C5 and C7 are the boundary constraints for the power allocation variables.

3.4 Solution of the Optimization Problem

The objective function in (3.13) is a ratio of two concave functions which is a non-convex function. In general, a brute force approach is required for obtaining a global

optimal solution. However, such a method has exponential complexity with respect to the numbers of subcarriers which is computationally infeasible even for small size systems. In order to derive an efficient resource allocation algorithm, we introduce the following transformation.

3.4.1 Transformation of the Objective Function

The fractional objective function in (3.12) can be classified as nonlinear fractional program [84]. For the sake of notational simplicity, we define $\mathcal{F}$ as the set of feasible solutions of the optimization problem in (3.13). Without loss of generality, we define the maximum energy efficiency q^* of the considered system as

$$q^* = \frac{U_{sec}(\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*)}{U_{TP}(\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*)} = \max_{\mathcal{P},\mathcal{R},\mathcal{S},\alpha_k[i]} \frac{U_{sec}(\mathcal{P}, \mathcal{R}, \mathcal{S})}{U_{TP}(\mathcal{P}, \mathcal{R}, \mathcal{S})}. \tag{3.14}$$

We are now ready to introduce the following Theorem.

Theorem 3.1 *The optimal resource allocation policies $\{\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*\} \in \mathcal{F}$ achieve the maximum energy efficiency q^* if and only if*

$$\begin{aligned} \max_{\mathcal{P},\mathcal{R},\mathcal{S},\alpha_k[i]} \quad & U_{sec}(\mathcal{P}, \mathcal{R}, \mathcal{S}) - q^* U_{TP}(\mathcal{P}, \mathcal{R}, \mathcal{S}) \\ = \quad & U_{sec}(\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*) - q^* U_{TP}(\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*) = 0, \end{aligned} \tag{3.15}$$

for $U_{sec}(\mathcal{P}, \mathcal{R}, \mathcal{S}) \geq 0$ and $U_{TP}(\mathcal{P}, \mathcal{R}, \mathcal{S}) > 0$.

Proof: Please refer to Appendix C.

Theorem 3.1 reveals that for an optimization problem with an objective function in fractional form, there exists an equivalent[7] objective function in subtractive form,

[7]Here, "equivalent" means both problem formulations will lead to the same resource allocation policies.

e.g. $U_{sec}(\mathcal{P},\mathcal{R},\mathcal{S}) - q^* U_{TP}(\mathcal{P},\mathcal{R},\mathcal{S})$ in the considered case. As a result, we can focus on the equivalent objective function in the rest of the chapter.

3.4.2 Iterative Algorithm for Energy Efficiency Maximization

In the next section, we propose an iterative algorithm (known as the Dinkelbach method [84]) for solving (3.13) with an equivalent objective function. The proposed algorithm is summarized in Table 3.1 and the convergence to optimal energy efficiency is guaranteed.

Proof: Please refer to the second part of Appendix C for the proof of convergence.

Note that the algorithm converges to the optimal solution with a superlinear convergence rate and please refer to [85] for a detailed proof. As shown in Table 3.1, in each iteration in the main loop, we solve the following optimization problem for a given parameter q:

$$\begin{aligned} \max_{\mathcal{P},\mathcal{R},\mathcal{S}} \quad & U_{sec}(\mathcal{P},\mathcal{R},\mathcal{S}) - qU_{TP}(\mathcal{P},\mathcal{R},\mathcal{S}) \\ & \text{s.t. C1, C2, C3, C4, C5, C6, C7.} \end{aligned} \tag{3.16}$$

In the following, we derive the solution of the main loop problem (3.16) by dual decomposition.

Solution of the Main Loop Problem

The main loop optimization problem in (3.16) is a mixed combinatorial and non-convex problem. The combinatorial nature comes from the Boolean subcarrier as-

Table 3.1: Iterative Resource Allocation Algorithm.

Algorithm 3.1 Iterative Resource Allocation Algorithm

Initialize the maximum number of iterations L_{max} and the maximum tolerance ϵ
Set maximum energy efficiency $q = 0$ and iteration index $n = 0$
repeat {Main Loop}
  Solve the inner loop problem in (3.16) for a given q and obtain resource allocation policies $\{\mathcal{P}', \mathcal{R}', \mathcal{S}'\}$
  if $U_{sec}(\mathcal{P}', \mathcal{R}', \mathcal{S}') - qU_{TP}(\mathcal{P}', \mathcal{R}', \mathcal{S}') < \epsilon$ **then**
    Convergence = **true**
    return $\{\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*\} = \{\mathcal{P}', \mathcal{R}', \mathcal{S}'\}$ and $q^* = \frac{U_{sec}(\mathcal{P}', \mathcal{R}', \mathcal{S}')}{U_{TP}(\mathcal{P}', \mathcal{R}', \mathcal{S}')}$
  else
    Set $q = \frac{U_{sec}(\mathcal{P}', \mathcal{R}', \mathcal{S}')}{U_{TP}(\mathcal{P}', \mathcal{R}', \mathcal{S}')}$ and $n = n + 1$
    Convergence = **false**
  end if
until Convergence = **true** or $n = L_{max}$

signment constraint C6, while the non-convexity comes from the secrecy outage constraint C1, since it is neither convex nor concave with respect to the optimization variables. It is convenient to incorporate the outage requirement constraint C1 in (3.13) into the objective function. This is possible if the constraint in C1 is fulfilled with equality for the optimal solution. Thus, in the following we replace the "$\leq$"-sign in C1 by a "="-sign and the resulting optimization problem may be viewed as a restricted version of the original problem (3.13) since it has a smaller feasible set[8]. We are now ready to introduce the following proposition.

Proposition 3.1 (Equivalent Secrecy Data Rate) *For a given outage probability ε in C1, the equivalent secrecy data rate which incorporates the secrecy outage*

[8]We can also adopt the chance constrained programming transformation in [86]. However, this transformation introduces an additional search algorithm which may result in an unacceptable complexity for the problem at hand.

probability on subcarrier i for user k with optimal $\alpha_k^*[i]$ is given by

$$R_k[i] = W\Big[\log_2\Big(1+P_k[i]\Upsilon_k[i]\Big)-\log_2\Big(1+\frac{\alpha_k^*[i]\Lambda_E[i]}{1-\alpha_k^*[i]}\Big)\Big]^+, \tag{3.17}$$

$$\Upsilon_k[i] = \frac{\alpha_k^*[i]\lambda_{\max_k}[i]}{N_0W}, \quad \Lambda_E[i]=(N_T-1)F_{z_c}^{-1}(\varepsilon), \quad \alpha_k^*[i]=\frac{1}{\sqrt{\Lambda_E[i]}}, \tag{3.18}$$

where $F_{z_c}^{-1}(\varepsilon)$ denotes the inverse function of $F_{z_c}(z)=\frac{\sum_{n=0}^{N_E-1}\binom{N_T-1}{n}z^n}{(1+z)^{N_T-1}}=\varepsilon$.

Proof: Please refer to Appendix D.

From the above proposition, it can be observed that the SINR of the eavesdropper, $\Phi_E[i]=\frac{\alpha_k^*[i]\Lambda_E[i]}{1-\alpha_k^*[i]}$, approaches a constant value at high SNR. More importantly, the SINR of the eavesdropper on each subcarrier is independent of the optimization variables, which simplifies the derivation of the optimal resource allocation algorithm.

By substituting (3.17) into (3.16), a modified objective function, which incorporates the secrecy outage requirement, can be obtained for the main loop problem in (3.16). To handle the combinatorial constraint C6, cf. (3.13), we follow the approach in [74] and relax constraint C6. In particular, we allow $s_k[i]$ to be a real value between zero and one instead of a Boolean. Then, $s_k[i]$ can be interpreted as a time sharing factor for the K users for utilizing subcarrier i. Although the relaxation of the subcarrier allocation constraint is generally suboptimal, the authors in [87] analytically show that the duality gap due to the relaxation becomes zero when the number of subcarriers goes to infinity. Therefore, using the equivalent secrecy data rate in Proposition 3.1, the auxiliary powers $\tilde{P}_k[i]=P_k[i]s_k[i]$, and the continuous

relaxation of C6, we can rewrite the problem in (3.16) for a given parameter q as

$$\begin{aligned}\max_{\mathcal{P},\mathcal{R},\mathcal{S}} \quad & \tilde{U}_{sec}(\mathcal{P},\mathcal{R},\mathcal{S}) - q\tilde{U}_{TP}(\mathcal{P},\mathcal{R},\mathcal{S}) \\ & \text{s.t. C4,} \quad \text{C5,} \\ & \text{C2:} \sum_{k=1}^{K}\sum_{i=1}^{n_F} \tilde{P}_k[i] \le P_t, \\ & \text{C3:} \sum_{k=1}^{K}\sum_{i=1}^{n_F} s_k[i]\tilde{R}_k[i] \ge r, \\ & \text{C6: } 0 \le s_k[i] \le 1, \qquad \forall i,k, \end{aligned} \tag{3.19}$$

where $\tilde{U}_{sec}(\mathcal{P},\mathcal{R},\mathcal{S}) = U_{sec}(\mathcal{P},\mathcal{R},\mathcal{S})\Big|_{P_k[i]=\frac{\tilde{P}_k[i]}{s_k[i]}}$, $\tilde{U}_{TP}(\mathcal{P},\mathcal{R},\mathcal{S}) = U_{TP}(\mathcal{P},\mathcal{R},\mathcal{S})\Big|_{P_k[i]=\frac{\tilde{P}_k[i]}{s_k[i]}}$, and $\tilde{R}_k[i] = R_k[i]\Big|_{P_k[i]=\frac{\tilde{P}_k[i]}{s_k[i]}}$. Mathematically, the $[\cdot]^+$ operator in (3.17) destroys the concavity of the objective function. Nevertheless, as will be seen in the KKT conditions in (3.24), users with negative secrecy data rate will not be considered in the subcarrier selection process. Therefore, we can safely remove the $[\cdot]^+$ operator from variable $\tilde{R}_k[i]$ and preserve the concavity of the transformed problem. Besides, C7 is removed from the optimization problem as the asymptotically optimal $\alpha_k^*[i]$ in (3.18) always satisfies C7 for $\Lambda_E[i] \gg 1$. Besides, $\max_{\mathcal{P},\mathcal{R},\mathcal{S}} \quad \tilde{U}_{sec}(\mathcal{P},\mathcal{R},\mathcal{S}) - q\tilde{U}_{TP}(\mathcal{P},\mathcal{R},\mathcal{S}) \ge 0$ holds for any values of q updated by Dinkelbach method, cf. Proposition C.2 in Appendix C. Therefore, the transformed problem (3.19) is jointly concave with respect to all optimization variables and under some mild conditions [75], it can be shown that strong duality holds and the duality gap is equal to zero. In other words, solving the dual problem is equivalent to solving the primal problem. Therefore, numerical methods such as the interior-point method and the ellipsoid method can be used to solve the transformed main loop problem in (3.19) and convergence to the optimal solution in polynomial time is guaranteed. However, these numerical methods do

not provide any useful system design insight such as the role of energy efficiency q in the resource allocation process. Hence, in the following subsections, an iterative algorithm for the transformed main loop problem in (3.19) will be derived based on dual decomposition.

Dual Problem

In this subsection, we solve the main loop problem in (3.19) by solving its dual. For this purpose, we first need the Lagrangian function of the primal problem. Upon rearranging terms, the Lagrangian can be written as

$$\mathcal{L}(\mu,\gamma,\boldsymbol{\beta},\mathcal{P},\mathcal{R},\mathcal{S}) = \sum_{k=1}^{K}(w_k+\gamma)\sum_{i=1}^{n_F} s_k[i]\tilde{R}_k[i] - \mu\sum_{k=1}^{K}\sum_{i=1}^{n_F}\tilde{P}_k[i] + \mu P_t + \sum_{i=1}^{n_F}\beta[i]$$
$$-q\Big(P_C + \sum_{k=1}^{K}\sum_{i=1}^{n_F} s_k[i]\delta\tilde{R}_k[i] + \sum_{k=1}^{K}\sum_{i=1}^{n_F}\tilde{P}_k[i]\Big) - \gamma r - \sum_{k=1}^{K}\sum_{i=1}^{n_F}\beta[i]s_k[i], \tag{3.20}$$

where $\mu \geq 0$ and $\gamma \geq 0$ are the Lagrange multipliers corresponding to the power constraint and the secrecy outage capacity constraint, respectively. $\boldsymbol{\beta}$ is the Lagrange multiplier vector associated with the subcarrier usage constraints with elements $\beta[i] \geq 0$, $i \in \{1, \ldots, n_F\}$. The boundary constraints C5 and C6 will be absorbed into the KKT conditions when deriving the optimal solution in the following.

Thus, the dual problem of (3.19) is given by

$$\min_{\mu,\gamma,\boldsymbol{\beta}\geq 0}\ \max_{\mathcal{P},\mathcal{R},\mathcal{S}}\quad \mathcal{L}(\mu,\gamma,\boldsymbol{\beta},\mathcal{P},\mathcal{R},\mathcal{S}). \tag{3.21}$$

In the following, we solve the above dual problem iteratively by decomposing it into two layers: Layer 1 consists of n_F subproblems with identical structure; Layer 2 is the master dual problem to be solved with the gradient method.

Dual Decomposition and Layer 1 Solution

By dual decomposition, the BS first solves the following Layer 1 subproblem

$$\max_{\mathcal{P},\mathcal{R},\mathcal{S}} \quad \mathcal{L}(\mu,\gamma,\boldsymbol{\beta},\mathcal{P},\mathcal{R},\mathcal{S}) \tag{3.22}$$

for a fixed set of Lagrange multipliers and parameter q. Using standard optimization techniques and the KKT conditions, the optimal power allocation for user k on subcarrier i is obtained as

$$\tilde{P}_k^*[i] \quad = \quad s_k[i]P_k^*[i] = s_k[i]\left[\frac{W(w_k+\gamma-\delta q)}{(\ln(2))(\mu+q)} - \frac{N_0 W}{\lambda_{\max_k}[i]\alpha_k^*[i]}\right]^+. \tag{3.23}$$

The optimal power allocation has the form of *multi-level* water-filling. It can be observed that the energy efficiency variable $q \geq 0$ prevents energy inefficient transmission by truncating the water-levels. There is also another interesting observation in (3.23). Let us focus on the case of equal priority users without secrecy data rate constraint, i.e., $w_k = 1$ and $\gamma = 0$. If we require a certain energy efficiency $q = q_{req}$, then (3.23) reveals a simple *necessary condition*[9] for a non-zero feasible solution: $\delta q_{req} < 1$.

In order to obtain the optimal subcarrier allocation, we take the derivative of the subproblem with respect to $s_k[i]$ and set it to zero, which yields

$$\left.\frac{\partial \mathcal{L}(\mu,\gamma,\boldsymbol{\beta},\mathcal{P},\mathcal{R},\mathcal{S})}{\partial s_k[i]}\right|_{P_k[i]=P_k^*[i]} \quad = \quad A_k[i] - \beta[i] = 0, \tag{3.24}$$

where $A_k[i] \geq 0$ can be interpreted as the marginal benefit [76] for allocating subcar-

[9]Note that the KKT conditions provide both the necessary and sufficient conditions for the "optimality" of a solution of the considered optimization problem. In contrast, $\delta q_{req} < 1$ provides a necessary condition for a non-zero transmit power solution.

rier i to user k and is given by

$$A_k[i] = W(w_k+\gamma-\delta q)\Bigg(\log_2\Big(1+P_k^*[i]\Upsilon_k[i]\Big)-\log_2\Big(1+\frac{\alpha_k^*[i]\Lambda_E[i]}{1-\alpha_k^*[i]}\Big) - \frac{P_k^*[i]\Upsilon_k[i]}{(\ln(2))(1+P_k^*[i]\Upsilon_k[i])}\Bigg). \tag{3.25}$$

$A_k[i] \geq 0$ has the physical meaning that users with negative data rate on subcarrier i are not selected as they can only provide a negative marginal benefit to the system. On the contrary, if a user has a larger weighting w_k and enjoys good channel conditions with positive data rate on subcarrier i, he/she can provide a higher marginal benefit to the system. Thus, the optimal allocation of subcarrier i at the BS to user k is given by

$$s_k^*[i] = \begin{cases} 1 & \text{if } A_k[i] = \max\limits_j \quad A_j[i] \quad \text{and } A_j[i] \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.26}$$

Note that each subcarrier will be used for serving only one user eventually. Finally, the optimal secrecy data rate $R_k^*[i]$ is obtained by substituting (3.23) into the equivalent secrecy data rate in (3.17) for the subcarrier with $s_k^*[i]=1$.

Solution of Layer 2 Master Problem

The dual function is differentiable and, hence, the gradient method can be used to solve the Layer 2 master problem in (3.21) which leads to

$$\mu(t+1) = \Big[\mu(t)-\xi_1(t)\times(P_t-\sum_{k=1}^{K}\sum_{i=1}^{n_F}\tilde{P}_k[i])\Big]^+, \tag{3.27}$$

$$\gamma(t+1) = \Big[\gamma(t)-\xi_2(t)\times(\sum_{k=1}^{K}\sum_{i=1}^{n_F}s_k[i]\tilde{R}_k[i]-r)\Big]^+, \tag{3.28}$$

where index $t \geq 0$ is the iteration index and $\xi_u(t)$, $u \in \{1, 2\}$, are positive step sizes. Updating $\beta[i]$ is not necessary as it has the same value for all users and does not affect the subcarrier allocation in (3.26). Therefore, we can simply set $\beta[i] = 0$ in each iteration. Since the transformed problem for a given parameter q is convex in nature, it is guaranteed that the iteration between Layer 1 the Layer 2 converges to the optimal solution of (3.19) in the main loop, if the chosen step sizes satisfy the infinite travel condition [75, 88]. Then, the updated Lagrange multipliers in (3.27) and (3.28) are used for solving the subproblems in (3.22) via updating the resource allocation policies. This procedure between Layer 1 and Layer 2 is repeated until convergence is achieved in each iteration of the main loop.

A summary of the overall algorithm is given in Table 3.1. In each iteration of the main loop, we solve the main loop problem in (3.19) for a given parameter q, cf. (3.17)-(3.28). Then, we update parameter q and use it for solving the main loop problem in the next iteration. This procedure is repeated until the proposed algorithm converges.

3.5 Results

In this section, we evaluate the system performance through simulations. A single cell with a radius of 1 km is considered, cf. Figure 3.1. The number of subcarriers is $n_F = 128$ with carrier center frequency 2.5 GHz, bandwidth $\mathcal{B} = 3$ MHz, and $w_k = 1, \forall k$. Each subcarrier has a bandwidth of 23.4 kHz and a noise variance of $N_0 = -130$ dBm. The 3GPP path loss model is used [78] with a reference distance of $d_0 = 35$ m. The K desired users are uniformly distributed between the reference distance and the cell boundary at 1 km. We assume that the eavesdropper is located 35 m away from the BS which represents an unfavourable scenario, since all the desired users are farther

away from the BS than the eavesdropper. The small scale fading coefficients of the BS-to-user and BS-to-eavesdropper links are modeled as i.i.d. Rayleigh random variables. The target secrecy outage probability is set to $\varepsilon = 0.01$. The average secrecy outage capacity is obtained by counting the number of packets securely delivered to and decoded by the users averaged over both the macroscopic and microscopic fading. Unless specified otherwise, we assume a static circuit power consumption of $P_C = 40$ dBm [89], a sum rate dependent power consumption parameter $\delta = 0.1$ mJ/bit, and a secrecy data rate requirement of $r = 2$ of bit/s/Hz. Note that if the resource allocator is unable to guarantee the required secrecy data rate in a time slot, we set the energy efficiency in that particular time slot to zero to account for the corresponding failure.

3.5.1 Convergence of Iterative Algorithm

Figure 3.2 illustrates the evolution of the proposed iterative algorithm for different numbers of transmit antennas N_T and a maximum transmit power of $P_t = 43$ dBm at the BS. The eavesdropper is equipped with $N_E = 2$ receive antennas and the result in Figure 3.2 was averaged over 10000 independent adaptation processes where each adaptation process involves different realizations for the path loss and the multipath fading. It can be observed that the iterative algorithm converges to the optimal value within 5 iterations for all considered numbers of transmit antennas. In other words, the maximum system energy efficiency can be achieved within a few iterations on average with a superlinear convergence rate [85].

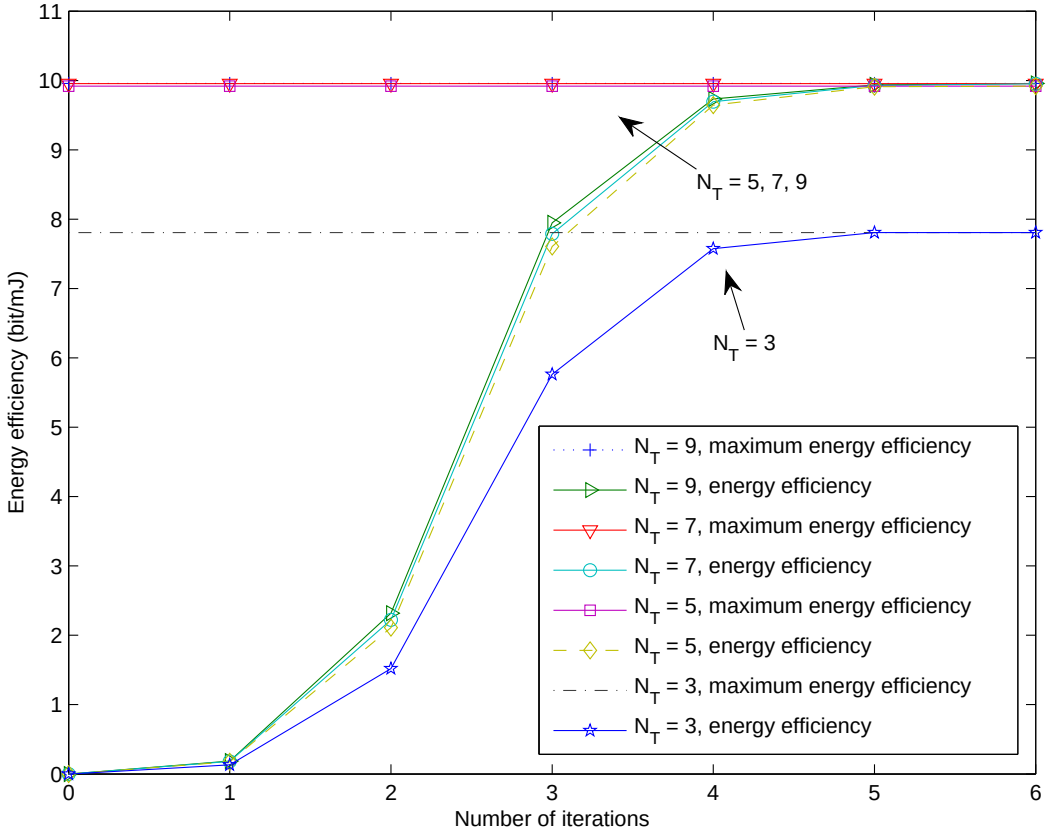


Figure 3.2: Energy efficiency (bit-per-milli Joule) versus the number of iterations with $K = 15$ users for different numbers of transmit antennas at the BS. The maximum transmit power at the BS is P_t =43 dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS.

3.5.2 Energy Efficiency and Average Secrecy Outage Capacity versus Transmit Power

Figure 3.3 illustrates the energy efficiency versus the total transmit power for $K = 15$ users for different numbers of transmit antennas N_T at the BS. The eavesdropper is equipped with $N_E = 2$ antennas. The number of iterations for the proposed iterative resource allocation algorithm is 5 and 10. It can be seen that the performance difference between 5 iterations and 10 iterations is negligible which confirms the

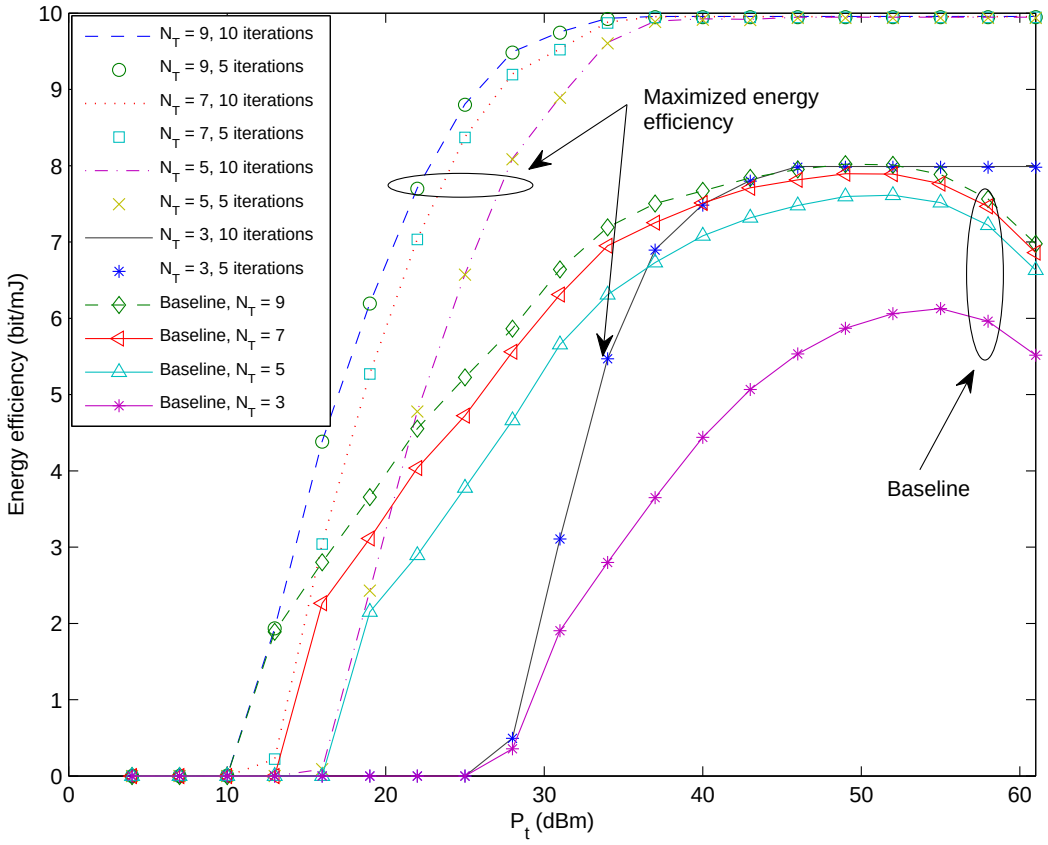


Figure 3.3: Energy efficiency (bit-per-milli Joule) versus maximum transmit power, P_t, for different numbers of transmit antennas N_T. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS.

practicality of our proposed iterative resource allocation algorithm. On the other hand, it can be observed that an increasing number of transmit antennas N_T benefits the system in terms of energy efficiency. This is because less power is required for maintaining a high receive SNR at the desired users, which results in energy savings. Besides, when both the number of transmit antennas and the maximum transmit power at the power amplifier are large enough, e.g. $N_T = 7$ and $P_t = 43$ dBm, the energy efficiency approaches a constant value, $1/\delta$, for $\delta > 0$, since the dynamic power consumption dominates the denominator in the energy efficiency equation in (3.12).

Figure 3.3 also contains the energy efficiency of a baseline resource allocation scheme. For the baseline scheme, we maximize the secrecy outage capacity (bit/s/Hz) with constraints C1-C7 in (3.13), instead of the energy efficiency. The optimal resource allocation polices for the baseline scheme can be obtained by using a similar approach as in [90]. It can be observed that the proposed algorithm provides a significant performance gain in terms of energy efficiency over the baseline scheme. This is because the latter scheme uses excess power to increase the secrecy outage capacity by sacrificing the system is energy efficiency, especially in the high transmit power regime.

Figure 3.4 shows the average secrecy outage capacity versus maximum transmit power P_t for $K = 15$ users and different numbers of transmit antennas at the B-S. We compare the system performance of the proposed algorithm again with the baseline scheme. The number of iterations in the proposed algorithm is set to 5. It can be observed that the average secrecy outage capacity of the proposed algorithm approaches a constant in the high transmit power regime, the value of which depends on the number of transmit antennas. This is because the proposed algorithm clips the transmit power at the BS in order to maximize the system energy efficiency. As will be shown in Figure 3.5, the average transmit power of the proposed algorithm remains static in the high transmit power regime. We note that, as expected, the baseline scheme achieves a higher average secrecy outage capacity than the proposed algorithm since the former scheme consumes all the available transmit power in all s-cenarios. However, the superior secrecy outage capacity of the baseline scheme comes at the expense of low energy efficiency. On the other hand, an increasing number of antennas benefits the secrecy outage capacity because of an improved beamforming gain. Yet, there is a diminishing return when N_T is large due to the *channel hardening*

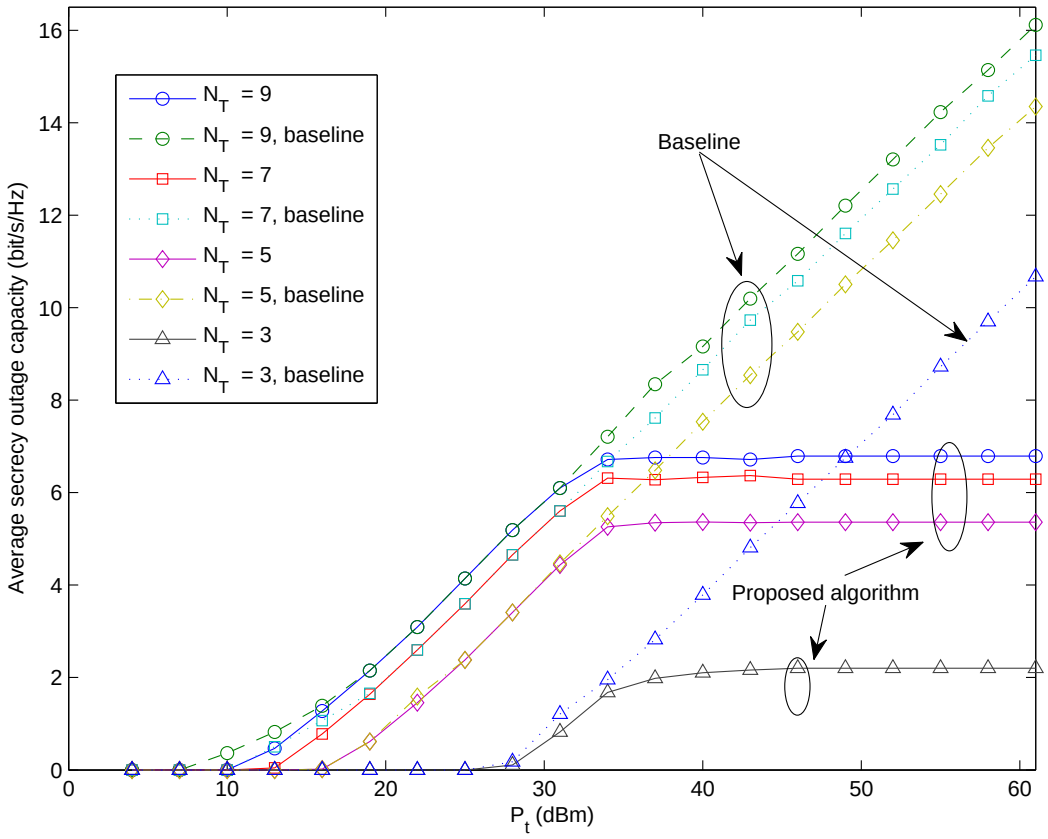


Figure 3.4: Average secrecy outage capacity versus maximum transmit power, P_t, for different numbers of transmit antennas N_T. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS.

effect [28] in the desired channels.

Figure 3.5 depicts the average total power consumption, i.e., $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{R}, \mathcal{S})\}$, versus maximum transmit power P_t for the proposed algorithm and the baseline scheme. As can be observed, the proposed algorithm consumes much less power than the baseline scheme, especially in the high transmit power regime. Besides, an increasing number of transmit antennas results in less power consumption due to a larger beamforming gain. Note that for $P_t < 37$ dBm, the proposed algorithm with $N_T = 3$ consumes the smallest power among all considered cases. This is because

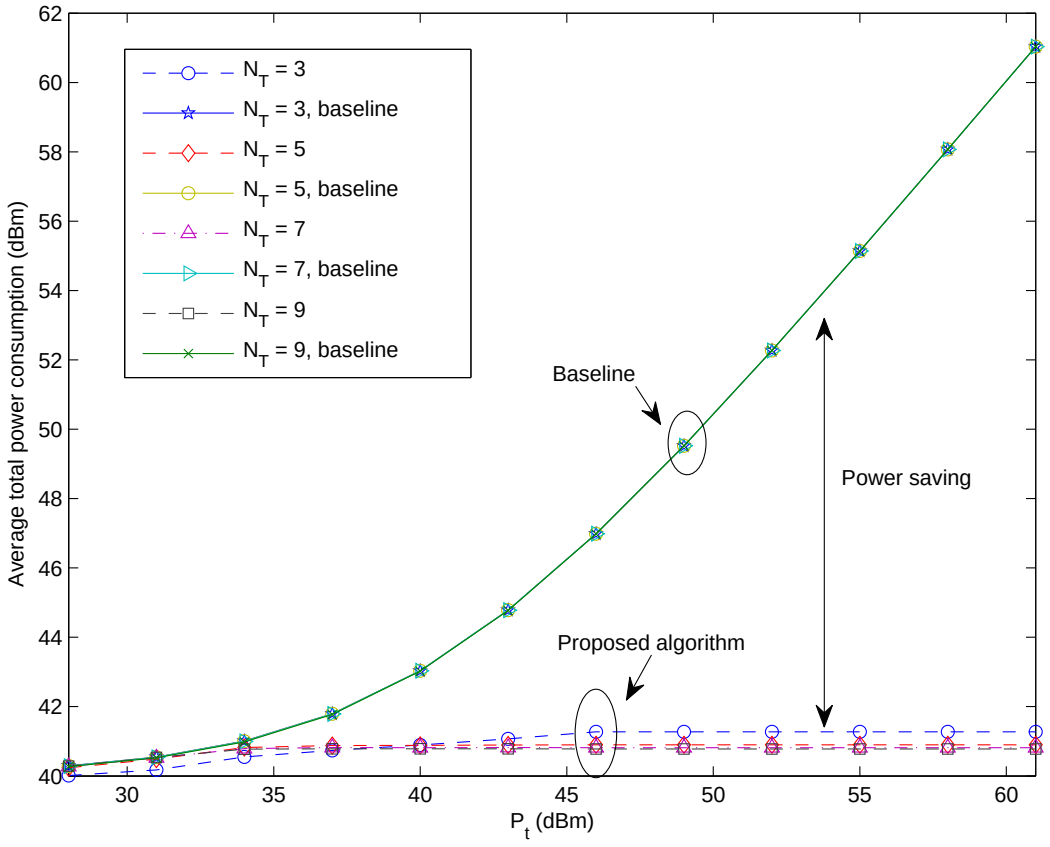


Figure 3.5: Average total power consumption versus maximum transmit power, P_t, for different numbers of transmit antennas N_T. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS.

with fewer antennas the probability that the secrecy data rate requirement is met is lower. Therefore, an extra energy saving is achieved when the transmitter is shut down. However, this leads to both low energy efficiency and low secrecy data rate.

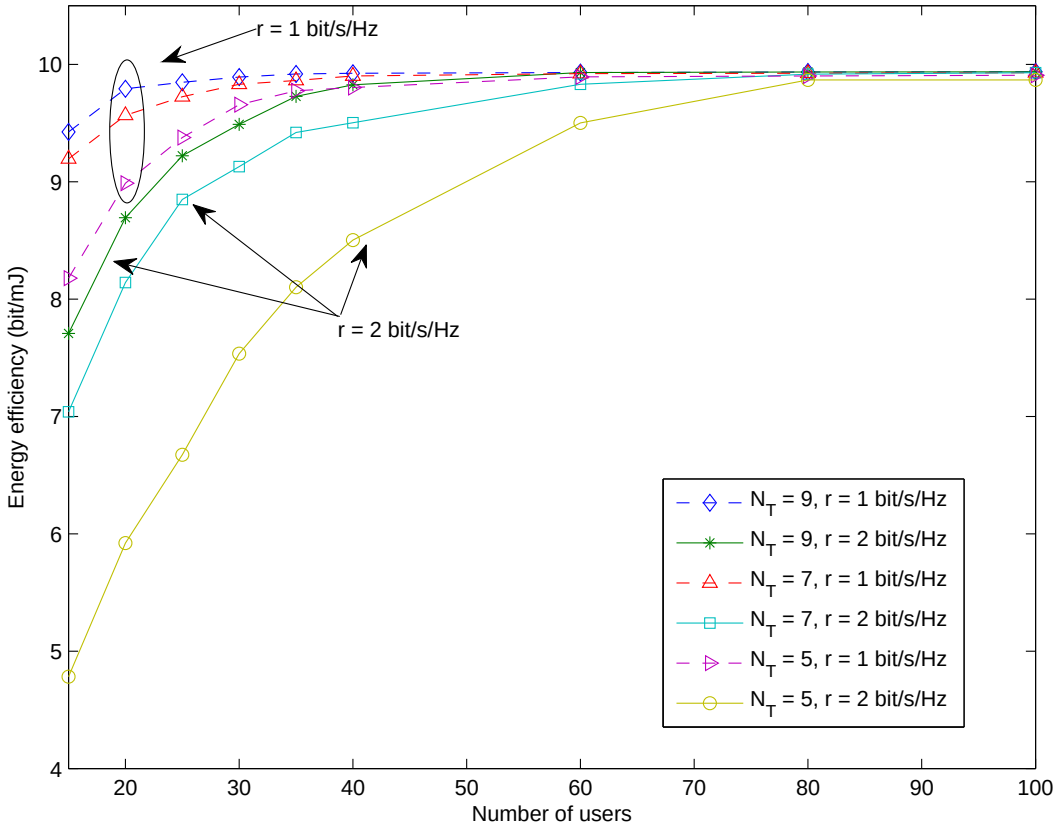


Figure 3.6: Energy efficiency (bit-per-milli Joule) versus the number of users K for different numbers of transmit antennas N_T and a maximum transmit power of $P_t = 22$ dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS.

3.5.3 Energy Efficiency and Secrecy Outage Capacity versus Number of Users

Figures 3.6 and 3.7 depict the energy efficiency and the average secrecy outage capacity versus the number of users, respectively. Different numbers of transmit antennas, different secrecy data rate requirements r, $P_T = 22$ dBm, and 5 iterations are considered. It can be observed that both the energy efficiency and the average secrecy outage capacity grow with the number of users since the proposed resource allocation and scheduling algorithm is able to exploit MUD, despite the existence of the eaves-

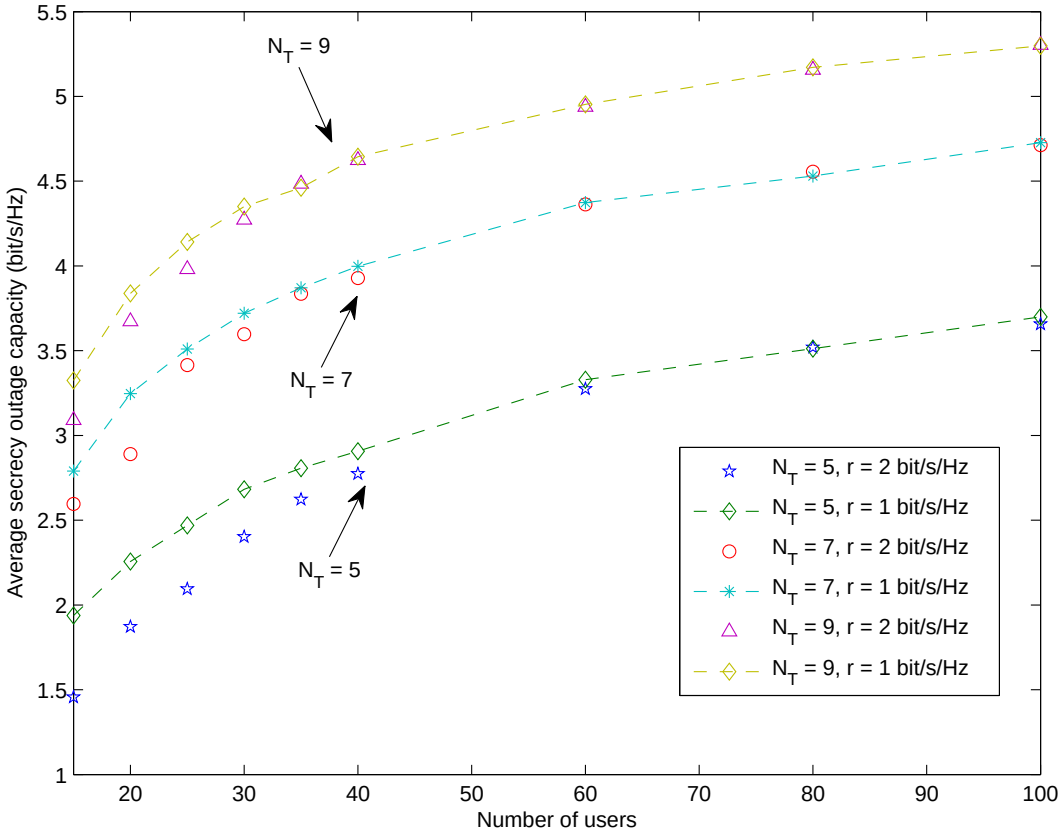


Figure 3.7: Average secrecy outage capacity versus the number of users K for different numbers of transmit antennas N_T and a maximum transmit power of $P_t = 22$ dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS.

dropper. Besides, when the number of users is large, the energy efficiency eventually approaches a constant which is similar to the case of high transmit power. Indeed, the MUD introduces an extra power gain [28, Section 6.6] to the system which provides further energy savings. On the contrary, the average secrecy outage capacity scales with the number of users without an upper limit. Yet, for large N_T, both the average secrecy outage capacity and the energy efficiency scale with the number users slowly. Indeed, since a large number of transmit antennas reduce channel fluctuations in the desired user channel and cause *channel hardening*, the potentially achievable MUD

gain in the subcarrier allocation process is decreased.

3.5.4 Energy Efficiency and Average Secrecy Outage Capacity versus N_E

Figures 3.8 and 3.9 illustrate, respectively, the energy efficiency and average secrecy outage capacity versus the number of receive antennas N_E employed at the eavesdropper for different dynamic circuit power constants δ and different static circuit powers P_C. There are $K = 15$ users and $N_T = 9$ transmit antennas at the BS. The number of iterations for the iterative algorithm is 5. It can be observed that both the energy efficiency and secrecy outage capacity decrease as N_E increases, since more of the transmitted power has to be devoted to artificial noise generation for degrading the channels of the eavesdropper, which leaves less power for information transmission. In addition, the average secrecy outage capacity is insensitive to the value of δ which suggests a constant secrecy data transmission rate when dynamic power consumption is taken consideration. On the other hand, we observe that larger values of δ and P_C lead to a lower energy efficiency since more energy is consumed in the circuit. Yet, a non-zero energy efficiency and average secrecy outage capacity can still be achieved as long as $N_T > N_E$, despite the fact that the eavesdropper is closer to the BS than the desired users. Interestingly, although a higher value of P_C results in a low energy efficiency, it increases the average secrecy outage capacity by allowing a higher transmit power.

Remark 3.1 *Note that in Figure 3.8, the energy efficiencies for the case of* $P_C = 50$ *dBm and* $P_C = 40$ *dBm cross at* $N_E = 6$. *This is because we shut down the power amplifier of the transmitter when the system cannot fulfill the secrecy data rate requirement which impacts the energy efficiency curves.*

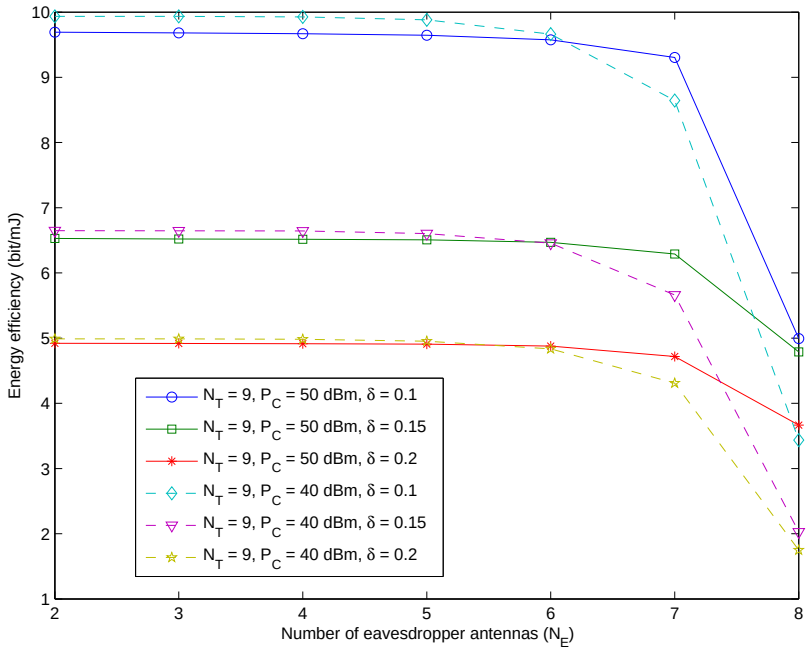


Figure 3.8: Energy efficiency (bit-per-milli Joule) versus the number of antennas at the eavesdropper for different static circuit powers, P_C, and different values of δ for a maximum transmit power of $P_t = 43$ dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS.

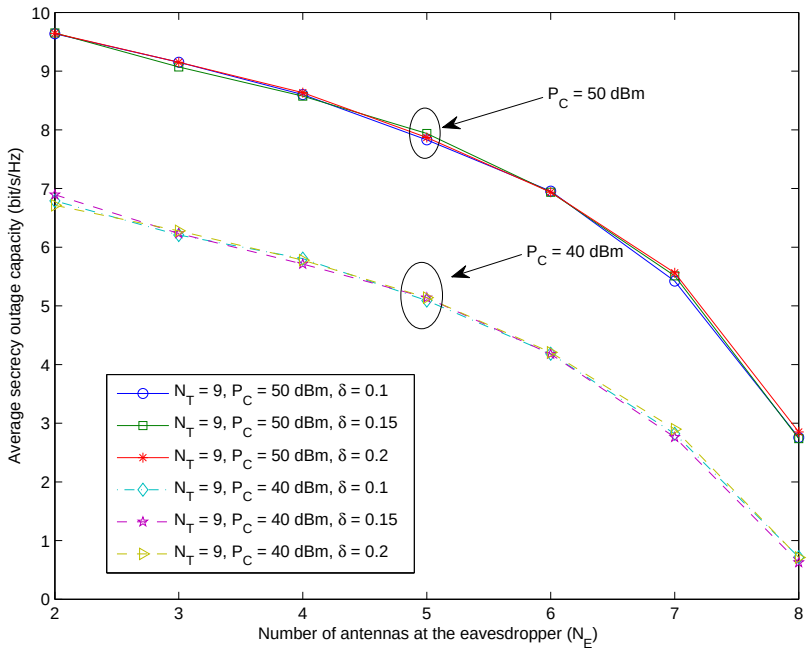


Figure 3.9: Average secrecy outage capacity versus the number of antennas at the eavesdropper for different static circuit powers, P_C, and different values of δ for a maximum transmit power of $P_t = 43$ dBm. The eavesdropper is equipped with $N_E = 2$ antennas and is located 35 m from the BS.

3.6 Conclusions

In this chapter, we formulated the resource allocation for energy efficient OFDMA systems as a mixed non-convex and combinatorial optimization problem, in which a multiple antenna eavesdropper, dynamic circuit power consumption, artificial noise injection for secure communication, and secrecy data rate requirements were taken into consideration. By exploiting the properties of fractional programming, the considered problem was transformed to an equivalent problem with a tractable solution. An efficient iterative resource allocation algorithm with closed-form power, secrecy data rate, and subcarrier allocation was derived by dual decomposition for maximization of the number of securely delivered bit-per-Joule. Simulation results not only showed that the proposed algorithm converges to the optimal solution within a small number of iterations, but also demonstrated the achievable maximum energy efficiency in the presence of a multiple antenna eavesdropper.

Chapter 4

Energy-Efficient Resource Allocation in OFDMA Systems with Large Numbers of Base Station Antennas

4.1 Introduction

In the previous two chapters, we have proposed two novel resource allocation algorithms for secure OFDMA communications under different scenarios. As mentioned in the introductory chapter, resource allocation tries to make the best use of the available degrees of freedom in a system. Starting from this chapter, we exploit the spatial degrees of freedom via MIMO. It is well known that MIMO technology provides extra degrees of freedom which facilitate multiplexing gains and diversity gains. It can be shown that the ergodic capacity of a MIMO fading channel increases practically linearly with the minimum of the number of transmit and receive antennas [28, 30]. Hence, it is not surprising that MIMO has attracted a lot of research interest over the past decade since it enables significant performance enhancement without requiring additional transmit power and bandwidth resources. However, the complexity of MIMO receivers limits the gains that can be achieved in practice, especially for handheld devices. An alternative is multiuser MIMO [91, 92] where a

transmitter with a large number of antennas serves multiple single antenna users. In [91], the authors investigated the uplink sum capacity (bit-per-second-per-Hertz) of cellular networks assuming unlimited numbers of antennas at both the BS and the users. In [92], high throughputs for both the uplink and the downlink were shown for a time-division duplex multi-cell system which employed multiple BSs equipped with large numbers of antennas. In [91, 92], substantial capacity gains and better interference management capabilities were observed for MIMO, compared to single antenna systems.

On the other hand, the combination of MIMO and OFDMA is considered a viable solution for achieving very high data rates communication [93]-[95]. In fact, the data rate improvement due to multiple antennas is unlimited if we allow the numbers of antennas employed at both the transmitter and the receiver to grow[10]. Yet, the advantages of MIMO and OFDMA do not come for free. They have significant financial implications for service providers due to the rapidly increasing cost for energy consumption in circuitries, which is often overlooked in the literature. Besides, all of these works [93]-[95] assume that perfect global CSI of all links is available at the BS. Hence, the power allocation can be done optimally and channel outage [28] can be avoided by data rate adaptation. However, in practice, CSIT is hardly perfect due to the mobility of users and/or estimation errors. Thus, channel outages occur with a non-zero probability and maximum tolerable outage probability requirements should be taken into consideration. Furthermore, if user selection and link adaptation are jointly optimized in MIMO-OFDMA systems, the energy efficient resource allocation algorithms proposed in [79]-[83], which were designed for perfect CSIT and a single user, are no longer applicable. In addition, a fixed number of active antennas has

[10]This statement is true if the number of scatters is large enough and all the antennas are uncorrelated.

been assumed to be used for transmission in the existing literature, e.g. [79]-[83]. In other words, the optimal number of active antennas used for transmission has not been investigated, at least not from an energy efficiency point of view.

Motivated by the aforementioned observations, in this chapter, we formulate the resource allocation problem for energy efficient communication in OFDMA systems with a large number of antennas and imperfect CSIT as an optimization problem. By exploiting the properties of fractional programming, the considered non-convex optimization problem in fractional form is transformed into an equivalent optimization problem in subtractive form with a tractable solution, which can be computed with an iterative algorithm. Because of the large numbers of antennas, the iterative algorithm requires only path loss and shadowing information. In other words, the BS updates the resource allocation policies based on the realizations of path loss and shadowing, which only change in the order of seconds. In each iteration, the transformed objective function is further lower bounded by a concave function which can be maximized by using dual decomposition. As a result, closed-form power, data rate, antenna, and subcarrier allocation policies are obtained for maximizing the energy efficiency.

4.2 OFDMA Downlink Network Model

In this section, we present the adopted OFDMA downlink network channel models.

4.2.1 Channel Model

We consider an OFDMA network which consists of a BS with multiple antennas and K mobile users equipped with a single antenna, cf. Figure 4.1. The impulse responses of all channels are assumed to be time-invariant (slow fading). There

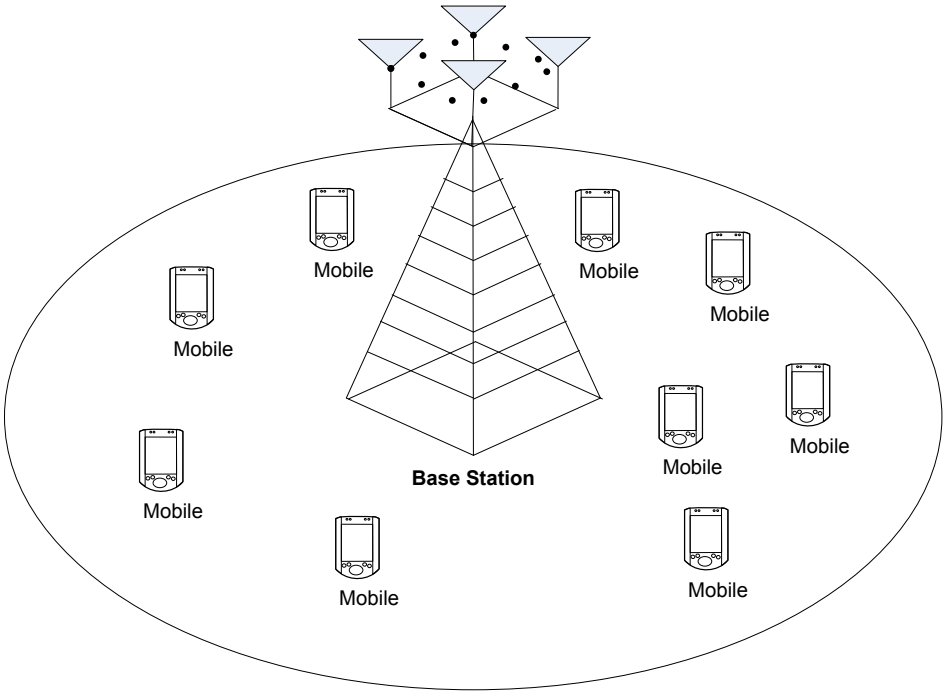


Figure 4.1: Illustration of an OFDMA downlink network. There are one BS with a large number of antennas and $K = 9$ desired users equipped with a single antenna.

are n_F subcarriers in each OFDM symbol. The downlink received symbol at user $k \in \{1, \ldots, K\}$ on subcarrier $i \in \{1, \ldots, n_F\}$ is given by

$$y_{i,k} = \sqrt{P_{i,k} l_k g_k} \mathbf{h}_{i,k}^T \hat{\mathbf{f}}_{i,k} x_{i,k} + \underbrace{\sum_{j \neq k} \mathbf{h}_{i,k}^T \hat{\mathbf{f}}_{i,j} x_{i,j} \sqrt{P_{i,j} l_k g_k s_{i,j}}}_{\text{Subcarrier reuse interference}} + z_{i,k}, \quad (4.1)$$

where $x_{i,k}$ and $\hat{\mathbf{f}}_{i,k} \in \mathbb{C}^{N_{T_{i,k}} \times 1}$ are the transmitted data symbol and the precoding vector used by the BS to transmit to user k on subcarrier i, respectively. $N_{T_{i,k}}$ is the number of active antennas allocated to user k on subcarrier i for transmission. $P_{i,k}$ is the transmit power for the link from the BS to user k in subcarrier i. $s_{i,j} \in \{0, 1\}$ is the subcarrier allocation indicator in subcarrier i for user j. $\mathbf{h}_{i,k} \in \mathbb{C}^{N_{T_{i,k}} \times 1}$ contains the small scale fading coefficients between the BS and user k on subcarrier i. l_k and

g_k represent the path loss and the shadowing between the BS and user k, respectively. $z_{i,k}$ is the AWGN in subcarrier i at user k with distribution $\mathcal{CN}(0, N_0)$, where N_0 is the noise power spectral density.

4.2.2 Channel State Information

In the following, since path loss and shadowing are slowly varying random processes which both change on the order of seconds for low mobility users, we assume that the path loss and shadowing coefficients can be estimated perfectly. For the multipath fading, we assume that the users can obtain perfect estimates of the BS-to-user fading gains $\mathbf{h}_{i,k}^{\dagger}\hat{\mathbf{f}}_{i,k} \in \mathbb{C}^{1\times 1}$, $i \in \{1, \ldots, n_F\}$, $k \in \{1, \ldots, K\}$ for signal detection purpose. However, the corresponding CSIT, i.e., $\mathbf{h}_{i,k} \in \mathbb{C}^{N_{T_{i,k}}\times 1}$ may be outdated/inaccurate at the BS because of the mobility of the users or errors in uplink channel estimation. To capture this effect, we model the multipath fading CSIT of the link between the BS and user k on subcarrier i as

$$\mathbf{h}_{i,k} = \hat{\mathbf{h}}_{i,k} + \Delta\mathbf{h}_{i,k}, \tag{4.2}$$

where $\hat{\mathbf{h}}_{i,k}$ and $\Delta\mathbf{h}_{i,k}$ denote the estimated CSIT vector and the CSIT error vector, respectively. $\hat{\mathbf{h}}_{i,k}$ and $\Delta\mathbf{h}_{i,k}$ are Gaussian random vectors and each vector has i.i.d. elements. Besides, the elements of vectors $\mathbf{h}_{i,k}$, $\hat{\mathbf{h}}_{i,k}$, and $\Delta\mathbf{h}_{i,k}$ have zero means and normalized variances of 1, $1-\sigma_e^2$, and σ_e^2, respectively. Assuming an MMSE estimator, the CSIT error vector and the actual CSIT vector are mutually uncorrelated.

4.3 Resource Allocation

In this section, we introduce the adopted system performance metric and formulate the corresponding resource allocation problem.

4.3.1 Instantaneous Channel Capacity and Outage Capacity

In this subsection, we define the adopted system performance measure. Given perfect CSI at the receiver, the maximum channel capacity between the BS and user k on subcarrier i with subcarrier bandwidth W is given by

$$C_{i,k} = W \log_2 \left(1 + \Gamma_{i,k}\right) \quad \text{and} \quad \Gamma_{i,k} = \frac{P_{i,k} l_k g_k \hat{\mathbf{f}}_{i,k}^{\dagger} \mathbf{h}_{i,k} \mathbf{h}_{i,k}^{\dagger} \hat{\mathbf{f}}_{i,k}}{W N_0 + \sum_{j \neq k} \hat{\mathbf{f}}_{i,j}^{\dagger} \mathbf{h}_{i,k} \mathbf{h}_{i,k}^{\dagger} \hat{\mathbf{f}}_{i,j} P_{i,j} s_{i,j} l_k g_k}, \quad (4.3)$$

where $\Gamma_{i,k}$ is the received SINR at user k on subcarrier i. The beamforming vector adopted at the BS is chosen to be the eigenvector corresponding to the maximum eigenvalue of $\hat{\mathbf{h}}_{i,k} \hat{\mathbf{h}}_{i,k}^{\dagger}$, i.e, $\hat{\mathbf{f}}_{i,k} = \frac{\hat{\mathbf{h}}_{i,k}}{\|\hat{\mathbf{h}}_{i,k}\|}$, which is known as maximum ratio transmission (MRT). Note that ZFBF is not considered in this system model since it requires the inversion of an $N_{T_{i,k}} \times N_{T_{i,k}}$ matrix on each subcarrier for each user, which is computational expensive for large $N_{T_{i,k}}$, n_F, and K.

In most existing resource allocation designs, the system performance is measured in terms of ergodic capacity. This is a meaningful measure when the resource allocator has perfect CSIT or the channels are fast fading (ergodic realizations of CSI within the encoding frame) such that an arbitrarily small decoding error probability can be achieved as long as the channel error correction code is strong enough. Yet, when the resource allocator has imperfect CSIT in slow fading, a packet outage occurs whenever the transmit data rate exceeds the instantaneous channel capacity despite the use of channel capacity achieving codes for error protection. In order to model

the effect of packet errors, we adopt the outage capacity [28] as performance measure. The *average weighted system outage capacity* is defined as the total average number of bit/s successfully delivered to the K mobile users and is given by $U(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}) =$

$$\sum_{k=1}^{K} w_k \sum_{i=1}^{n_F} s_{i,k} \mathcal{E}\Big\{ R_{i,k} \times \ 1\Big(R_{i,k} \leq C_{i,k} \Big) \Big\} = \sum_{k=1}^{K} w_k \sum_{i=1}^{n_F} s_{i,k} R_{i,k} \Pr\Big[R_{i,k} \leq C_{i,k} \Big], \quad (4.4)$$

where $\mathcal{P}, \mathcal{A}, \mathcal{R}$, and $\mathcal{S}$ are the power, antenna, data rate, and subcarrier allocation policies, respectively. $R_{i,k}$ is the scheduled data rate for user k on subcarrier i. $0 \leq w_k \leq 1$ is a positive constant provided by the upper layers, which allows the resource allocator to give different priorities to different users and to enforce certain notions of fairness. On the other hand, for designing an energy efficient resource allocation algorithm, the total power consumption has to be included in the optimization objective function. Thus, we model the power dissipation, $U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})$, in the system as the sum of two dynamic terms and one static term [96]:

$$\begin{aligned} U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}) \ = \ & \underbrace{\max_{i,k} \{ s_{i,k} \times N_{T_{i,k}} \} \times P_C}_{\substack{\text{Circuit power consumption of all antennas at the BS}}} \\ + \ & \underbrace{\sum_{k=1}^{K} \sum_{i=1}^{n_F} \rho P_{i,k} s_{i,k}}_{\text{BS power amplifier}} + P_0, \end{aligned} \quad (4.5)$$

where P_C is the constant circuit *power consumption per antenna* which includes the power dissipations in the transmit filter, mixer, frequency synthesizer, and digital-to-analog converter which is independent of the actual transmitted power. In the considered system, we assume that there are a maximum number antennas, $N_{\max}$, at the BS. However, we only activate some of them for the sake of energy efficient

communication[11]. Note that the physical meaning of the term $\max_{i,k}\{s_{i,k} \times N_{T_{i,k}}\}$ is that an antenna is activated and consumes power even it is used only by some of the users on some of the subcarriers. $\rho \geq 1$ is a constant which accounts for the inefficiency of the power amplifier. For example, if $\rho = 5$, for every 10 Watts of radiated power in the RF, 50 Watts are consumed in the power amplifier and the power efficiency is $\frac{1}{\rho} = \frac{1}{5} = 20\%$. P_0 is the basic power consumed at the BS independent of the number of transmit antennas. Unlike the power dissipation model used in (3.11) in Chapter 3, the model adopted in here includes the power efficiency of the power amplifier and also the power consumption per antenna.

Hence, the *weighted energy efficiency* of the considered system is defined as the total average number of bit/Joule successfully delivered to the users which is given by

$$U_{eff}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}) \;=\; \frac{U(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})}{U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})}. \tag{4.6}$$

4.3.2 Optimization Problem Formulation

The optimal power allocation policy, $\mathcal{P}^*$, antenna allocation policy, $\mathcal{A}^*$, data rate adaption policy, $\mathcal{R}^*$, and subcarrier allocation policy, $\mathcal{S}^*$, can be obtained by solving

$$\max_{\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S},} \; U_{eff}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}) \tag{4.7}$$

$$\text{s.t.} \;\; \text{C1:} \sum_{k=1}^{K}\sum_{i=1}^{n_F} s_{i,k}R_{i,k} \geq r, \qquad \text{C2:} \sum_{k=1}^{K}\sum_{i=1}^{n_F} P_{i,k}s_{i,k} \leq P_T,$$

$$\text{C3:} \Pr\Big(C_{i,k} < R_{i,k}\Big) \leq \varepsilon, \; \forall i,k, \quad \text{C4:} \; P_{i,k} \geq 0, \; \forall i,k,$$

$$\text{C5:} \; s_{i,k} = \{0,1\}, \; \forall i,k, \qquad \text{C6:} \; N_{T_{i,k}} = \{1,2,3,\ldots,N_{\max}\}, N_{T_{i,k}} \in \mathbb{Z}^+ \; \forall i,k.$$

[11]The optimal number of active antennas will be found in next section based on optimization.

C1 specifies the minimum system data rate requirement r. C2 is a transmit power constraint for the BS in the downlink. The value of P_T in C2 puts a limit on the amount of out-of-cell interference in the downlink. C3 specifies the channel outage probability requirement ε. Note that the number of active antennas is an optimization variable in the considered problem. Hence, the imperfect CSI of the multipath fading can only be acquired by the BS after the resource allocator has decided on the number of active antennas. Therefore, the outage probability conditional on the multipath fading, which is commonly used in the literature, cannot be adopted in C3. C5 is a combinatorial constraint on the subcarrier assignment. Furthermore, C5 implicitly imposes a fairness constraint, since no user can dominate the subcarrier reuse process. In other words, selected users are not allowed to multiplex different messages on the same subcarrier, since a sophisticated receiver would be required at each user, such as a successive interference cancellation (SIC) receiver, to recover more than one messages. Besides, the weaker users have a higher chance of being selected for reusing a subcarrier. C4 is the boundary constraint for the power allocation variables. C6 is the combinatorial constraint on the number of antennas.

4.4 Solution of the Optimization Problem

The objective function in (4.7) is a non-convex function. In general, a brute force approach is required for obtaining a global optimal solution. However, such a method has exponential complexity with respect to the number of subcarriers which is computationally infeasible even for small size systems. In order to obtain an efficient resource allocation algorithm, we introduce the following transformation.

4.4.1 Problem Transformation

The fractional objective function in (4.6) can be classified as a nonlinear fractional program [84]. For the sake of notational simplicity, we define $\mathcal{F}$ as the set of feasible solutions of the optimization problem in (4.7). Without loss of generality, we define the maximum energy efficiency q^* of the considered system as

$$q^* = \frac{U(\mathcal{P}^*, \mathcal{A}^*, \mathcal{R}^*, \mathcal{S}^*)}{U_{TP}(\mathcal{P}^*, \mathcal{A}^*, \mathcal{R}^*, \mathcal{S}^*)} = \max_{\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}} \frac{U(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})}{U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})}, \quad \forall\{\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}\} \in \mathcal{F}. \quad (4.8)$$

By Theorem 3.1 in Chapter 3, the maximum energy efficiency q^* is achieved if and only if

$$\begin{aligned} \max_{\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}} \quad & U(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}) - q^* U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}) \\ = \quad & U(\mathcal{P}^*, \mathcal{A}^*, \mathcal{R}^*, \mathcal{S}^*) - q^* U_{TP}(\mathcal{P}^*, \mathcal{A}^*, \mathcal{R}^*, \mathcal{S}^*) = 0, \end{aligned} \quad (4.9)$$

for $U(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}) \geq 0$ and $U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}) > 0$.

Theorem 3.1 reveals that for an optimization problem with an objective function in fractional form, there exists an equivalent[12] objective function in subtractive form, e.g. $U(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}) - q^* U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})$ in the considered case. As a result, we can focus on the equivalent objective function in the rest of the chapter.

4.4.2 Iterative Algorithm for Energy Efficiency Maximization

In this section, we propose an iterative algorithm (known as the Dinkelbach method [84]) for solving (4.7) with an equivalent objective function. The proposed algorithm

[12]Here, "equivalent" means both problem formulations will lead to the same resource allocation policies.

Table 4.1: Iterative Resource Allocation Algorithm.

Algorithm 4.1 Iterative Resource Allocation Algorithm

1: Initialize the maximum number of iterations L_{max} and the maximum tolerance ϵ
2: Set maximum energy efficiency $q = 0$ and iteration index $n = 0$
3: **repeat** {Main Loop}
4: Solve the inner loop problem in (4.10) for a given q and obtain resource allocation policies $\{\mathcal{P}', \mathcal{A}', \mathcal{R}', \mathcal{S}'\}$
5: **if** $U(\mathcal{P}', \mathcal{A}', \mathcal{R}', \mathcal{S}') - qU_{TP}(\mathcal{P}', \mathcal{A}', \mathcal{R}', \mathcal{S}') < \epsilon$ **then**
6: Convergence = **true**
7: **return** $\{\mathcal{P}^*, \mathcal{A}^*, \mathcal{R}^*, \mathcal{S}^*\} = \{\mathcal{P}', \mathcal{A}', \mathcal{R}', \mathcal{S}'\}$ and $q^* = \frac{U(\mathcal{P}', \mathcal{A}', \mathcal{R}', \mathcal{S}')}{U_{TP}(\mathcal{P}', \mathcal{A}', \mathcal{R}', \mathcal{S}')}$
8: **else**
9: Set $q = \frac{U(\mathcal{P}', \mathcal{A}', \mathcal{R}', \mathcal{S}')}{U_{TP}(\mathcal{P}', \mathcal{A}', \mathcal{R}', \mathcal{S}')}$ and $n = n + 1$
10: Convergence = **false**
11: **end if**
12: **until** Convergence = **true** or $n = L_{max}$

is summarized in Table 4.1 and the convergence to the optimal energy efficiency is guaranteed.

Note that the algorithm converges to the optimal solution with a superlinear convergence rate and please refer to [85] for a detailed proof of the rate of convergence. As shown in Table 4.1, in each iteration in the main loop (line 4 in Table 4.1), we solve the following optimization problem for a given parameter q:

$$\max_{\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}} \quad U(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}) - qU_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})$$
$$\text{s.t. C1, C2, C3, C4, C5, C6.} \tag{4.10}$$

In the following, we derive the solution of the main loop problem (4.10) by dual decomposition.

Sub-Optimal Solution of the Main Loop Problem

The transformed problem is a mixed combinatorial and non-convex optimization problem. The combinatorial nature comes from the integer constraint for both subcarrier allocation and antenna allocation while the non-convexity is due to the following properties: First, the multiuser interference due to subcarrier reuse appears in the denominator of the capacity equation in (4.3) which couples the power allocation variables. Second, the outage probability requirement in C3 is neither concave nor convex with respect to the optimization variables. Furthermore, the probability distribution of the SINR in (4.3) is coupled with the optimization variables which makes the resource allocation algorithm design untractable. In order to derive an efficient resource allocation algorithm, we introduce the following proposition by taking advantage of the large numbers of antennas.

Proposition 4.1 (Equivalent Data Rate) *For a given outage probability ε in C3, the equivalent data rate which incorporates the outage probability on subcarrier i for user k is given by*

$$R_{i,k} = (1-\varepsilon)W\log_2\Big(1+\frac{P_{i,k}l_k g_k N_{T_{i,k}}(1-\sigma_e^2)(1-\delta)}{WN_0+\sum_{j\neq k}(\frac{2}{\varepsilon})P_{i,j}s_{i,j}l_k g_k}\Big),, \tag{4.11}$$

where $0<\delta<1$ is a constant back-off factor. Note that $N_{T_{i,k}} \geq \lceil N_{th} \rceil$ and $\lceil N_{th} \rceil$ is the solution of (E.6) in Appendix E which indicates the minimum number of antennas required for Proposition 4.1 to hold.

Proof: Please refer to Appendix E for a proof of Proposition 4.1 and the meaning of δ.

The next step in solving the considered problem is to handle the inter-user interference on each subcarrier. To this end, we introduce an additional constraint C7 to

the original problem which is given by

$$\text{C7:} \sum_{j \neq k} (\frac{2}{\varepsilon}) P_{i,j} s_{i,j} l_k g_k \leq I, \quad \forall k, i. \tag{4.12}$$

C7 can be interpreted as the maximum inter-user interference temperature [97] (tolerable interference level) in each subcarrier. By varying[13] the value of I, the resource allocator is able to control the amount of interference in each subcarrier to improve the system performance. Furthermore, by substituting $\sum_{j \neq k} (\frac{2}{\varepsilon}) P_{i,j} s_{i,j} l_k g_k$ in (4.11) by I, the inter-user interference can be decoupled from the objective function, which facilitates the design of an efficient resource allocation algorithm. Then, the scheduled data rate between the BS and user k on subcarrier i can be lower bounded by

$$\begin{aligned} R_{i,k} &= W \log_2 \Big(1 + \frac{P_{i,k} l_k g_k N_{T_{i,k}} (1-\sigma_e^2)(1-\delta)}{W N_0 + I}\Big) \\ &> W \log_2 \Big(\frac{P_{i,k} l_k g_k}{W N_0 + I}\Big) + W \log_2 \Big(N_{T_{i,k}} (1-\sigma_e^2)(1-\delta)\Big). \end{aligned} \tag{4.13}$$

By substituting (4.13) into (4.10), a modified objective function, which incorporates the channel outage requirement, can be obtained for the main loop problem in (4.10). Indeed, it can be observed that the scheduled data rate for user k on subcarrier i in (4.13) depends only on the *path loss* and *shadowing* information of user k due to the large numbers of antennas. In other words, we expect that the optimal resource allocation policy will be the same for all subcarriers of user k.

To handle the combinatorial constraints C5 and C6, cf. (4.7), we follow the approach in [74] and relax constraints C5 and C6. In particular, we allow $s_{i,k}$ to be

[13]The maximum inter-user interference temperature variable I is not an optimization variable in the proposed framework. However, a suitable value of I can be found via simulation in an off-line manner.

a real value between zero and one instead of a Boolean, while $N_{T_{i,k}}$ can be a positive real value. Then, $s_{i,k}$ can be interpreted as a time sharing factor for the K users for utilizing subcarrier i. Although the relaxations of $N_{T_{i,k}}$ and $s_{i,k}$ are generally sub-optimal, they facilitate the design of an efficient resource allocation algorithm. Therefore, using the equivalent data rate in Proposition 4.1, the auxiliary time-shared powers $\tilde{P}_{i,k} = P_{i,k}s_{i,k}$, the auxiliary time-shared antenna, $\tilde{N}_{T_{i,k}} = N_{T_{i,k}}s_{i,k}$, and the continuous relaxation of both C5 and C6, we can rewrite the problem in (4.10) for a given parameter q as

$$\begin{aligned}
&\max_{\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}} \quad \tilde{U}(\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}) - q\tilde{U}_{TP}(\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}) \\
&\qquad\qquad \text{s.t. C4,} \\
&\text{C1:} \sum_{k=1}^{K}\sum_{i=1}^{n_F} s_{i,k}\tilde{R}_{i,k} \geq r, \qquad\qquad \text{C2:} \sum_{k=1}^{K}\sum_{i=1}^{n_F} \tilde{P}_{i,k} \leq P_T, \\
&\text{C5: } 0 \leq s_{i,k} \leq 1, \quad \forall i,k, \\
&\text{C6: } N_{\max} \geq \tilde{N}_{T_{i,k}} \geq \left\lceil \tilde{N}_{th} \right\rceil, \forall i,k, \quad \text{C7:} \sum_{j\neq k} (\frac{2}{\varepsilon})\tilde{P}_{i,j}s_{i,j}l_k g_k \leq I, \forall i,k,
\end{aligned} \tag{4.14}$$

where $\tilde{U}(\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}) = U(\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S})\Big|_{P_{i,k}=\frac{\tilde{P}_{i,k}}{s_{i,k}}, N_{T_{i,k}}=\frac{\tilde{N}_{T_{i,k}}}{s_{i,k}}}$, $\tilde{U}_{TP}(\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S})$ $= U_{TP}(\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S})\Big|_{P_{i,k}=\frac{\tilde{P}_{i,k}}{s_{i,k}}, N_{T_{i,k}}=\frac{\tilde{N}_{T_{i,k}}}{s_{i,k}}}$, and $\tilde{R}_{i,k} = R_{i,k}\Big|_{P_{i,k}=\frac{\tilde{P}_{i,k}}{s_{i,k}}, N_{T_{i,k}}=\frac{\tilde{N}_{T_{i,k}}}{s_{i,k}}}$.

The transformed problem (4.14) is now jointly concave with respect to all optimization variables, cf. Appendix F. Thus, under some mild conditions [75], it can be shown that strong duality holds and the duality gap is equal to zero. In other words, solving the dual problem is equivalent to solving the primal problem[14].

[14]Note that if we solve (4.14) instead of (4.10) in each main loop iteration of Algorithm 4.1 in Table 4.1, the algorithm converges to a lower bound of the maximum energy efficiency of (4.7).

Dual Problem

In this subsection, we solve the main loop problem in (4.14) by solving its dual. For this purpose, we first need the Lagrangian function of the primal problem. Upon rearranging terms, the Lagrangian can be written as

$$\mathcal{L}(\mu,\gamma,\boldsymbol{\beta},\boldsymbol{\theta},\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}) = \sum_{k=1}^{K}(w_k+\gamma)\sum_{i=1}^{n_F} s_{i,k}\tilde{R}_{i,k} - \mu\sum_{k=1}^{K}\sum_{i=1}^{n_F}\tilde{P}_{i,k} + \mu P_T - \gamma r$$
$$-q\Big(\max_{i,k}\{\tilde{N}_{T_{i,k}}\}\times P_C + \sum_{k=1}^{K}\sum_{i=1}^{n_F}\rho\tilde{P}_{i,k} + P_0\Big) - \sum_{k=1}^{K}\sum_{i=1}^{n_F}\theta_{i,k}\Big(\sum_{j\neq k}(\frac{2}{\varepsilon})\tilde{P}_{i,j}l_k g_k - I\Big), \quad (4.15)$$

where $\mu \geq 0$ and $\gamma \geq 0$ are the Lagrange multipliers corresponding to the power constraint and the required minimum outage capacity constraint, respectively. $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are the Lagrange multiplier vectors associated with the inter-user interference temperature constraint C7 and subcarrier usage constraint C4 with elements $\theta_{i,k} \geq 0$ and $\beta_i \geq 0$, respectively. The boundary constraints C4, C5, and C6 will be absorbed into the KKT conditions when deriving the optimal solution in the following. Thus, the dual problem of (4.14) is given by

$$\min_{\mu,\gamma,\boldsymbol{\beta},\boldsymbol{\theta}\geq 0}\ \max_{\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}}\quad \mathcal{L}(\mu,\gamma,\boldsymbol{\beta},\boldsymbol{\theta},\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}). \quad (4.16)$$

In the following, we solve the above dual problem iteratively by decomposing it into two layers: Layer 1 consists of n_F subproblems with identical structure; Layer 2 is the master dual problem to be solved with the gradient method.

Dual Decomposition and Layer 1 Solution

By dual decomposition, the BS first solves the following Layer 1 subproblem

$$\max_{\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}} \quad \mathcal{L}(\mu,\gamma,\boldsymbol{\beta},\boldsymbol{\theta},\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S}) \tag{4.17}$$

for a fixed set of Lagrange multipliers and a given parameter q. Using standard optimization techniques and the KKT conditions, the optimal power allocation for user k on subcarrier i is obtained as

$$\tilde{P}^*_{i,k} = s_{i,k}P^*_{i,k} = s_{i,k}\left[\frac{W(w_k+\gamma)}{(\ln(2))(\mu+q\rho+\Omega_{i,k})}\right], \quad \text{where} \quad \Omega_{i,k} = \sum_{j\neq k}\theta_{i,k}(\frac{2}{\varepsilon})l_j g_j \tag{4.18}$$

represents the interference to the other users created by this power allocation. The optimal power allocation has the form of *multi-level* water-filling. It can be observed that the energy efficiency variable $q \geq 0$ prevents energy inefficient transmission by truncating the water-levels. On the contrary, a large value of $\Omega_{i,k}$ results in a lower water-level in the power allocation to reduce the interference caused to the other users such that the constraint C7 in (4.14) is satisfied.

Similarly, the close-to-optimal[15] number of activated antennas for user k on subcarrier i is given by

$$\tilde{N}^*_{T_{i,k}} = N^*_{T_{i,k}}s_{i,k} = \left[\left[\frac{W(\max\limits_{k\in\Psi_i} w_k+\gamma)}{P_C(\frac{q}{\Phi_i})\ln(2)}\right]\right]^{N_{\max}}_{\lfloor N_{th}\rfloor} s_{i,k}, \tag{4.19}$$

where Ψ_i denotes a selected user set for using subcarrier i and $\Phi_i = \sum_{b\in\Psi_i} 1(\max\limits_{k\in\Psi_i} w_k = w_b)$ counts the number of w_k which have a value equal to $\max\limits_{k\in\Psi_i} w_k$ for all selected users.

[15]Here, the sub-optimality is due to the floor and ceiling functions in (4.19) which are required for fulfilling the combinatorial constraint in practice.

If the data rate requirement constraint C1 in (4.7) is stringent, the dual variable γ is large and forces the resource allocator to assign more antennas to all selected users, cf. (4.19), for satisfying constraint C1. Besides, (4.19) reveals that all users will eventually use the same number of antennas. This behavior can be explained by the following example: Suppose user 1 and user 2 are using N_1 and N_2 antennas such that $N_1 > N_2$. Yet, from user 2's point of view, the cost for $N_1 - N_2$ extra antennas has been paid by user 1 already. Therefore, since no extra cost has to be paid, user 2 is willing to use extra antennas until $N_2 = N_1$, since this will benefit the system performance.

In order to obtain the optimal subcarrier allocation, we take the derivative of the subproblem with respect to $s_{i,k}$, which yields $\frac{\partial \mathcal{L}(\mu,\gamma,\boldsymbol{\beta},\boldsymbol{\theta},\mathcal{P},\mathcal{A},\mathcal{R},\mathcal{S})}{\partial s_{i,k}}\Big|_{P_{i,k}=P^*_{i,k},N_{T_{i,k}}=N^*_{T_{i,k}}} = M_{i,k}$, where $M_{i,k} \geq 0$ can be interpreted as the marginal benefit [76] for allocating subcarrier i to user k and is given by

$$\begin{aligned} M_{i,k} &= (1-\varepsilon)W(w_k+\gamma)\Bigg(\log_2\Big(\frac{P^*_{i,k}l_k g_k}{WN_0+I}\Big) \\ &\quad + \log_2\Big(N^*_{T_{i,k}}(1-\sigma_e^2)(1-\delta)\Big) - 2/\ln(2)\Bigg). \end{aligned} \tag{4.20}$$

$M_{i,k} \geq 0$ has the physical meaning that users with negative scheduled data rate on subcarrier i are not selected as they can only provide a negative marginal benefit to the system.

On the contrary, if a user has a larger weight w_k and enjoys good channel conditions with positive data rate on subcarrier i, he/she can provide a higher marginal benefit to the system. Thus, the optimal allocation of subcarrier i at the BS is based

on the following criterion:

$$s^*_{i,k} = 1 \text{ if } M_{i,k} \geq 0 \quad \text{and} \quad s^*_{i,k} = 0 \quad \text{otherwise.} \tag{4.21}$$

As explained earlier, since the multipath fading has vanished because of the beamforming with a large number of antennas, all the subcarriers of user k experience the same channel gain. Hence, the resource allocation policy for user k on subcarrier i, i.e., (4.18)-(4.21), is identical to that of the other $n_F - 1$ subcarriers of user k. Indeed, (4.21) can be interpreted as a chunk-based subcarrier allocation. In other words, if subcarrier i is allocated to user k, the other $n_F - 1$ subcarriers should also be allocated to user k since they provide the same marginal benefit. As a result, the complexity in solving the Layer 1 problem is reduced by a factor of n_F.

Finally, the optimal data rate allocation $R^*_{i,k}$ is obtained by substituting (4.18) and (4.19) into the equivalent data rate in (4.11) in Proposition 1 for the subcarriers with $s_{i,k} = 1$.

Solution of Layer 2 Master Problem

The dual function is differentiable and, hence, the gradient method can be used to solve the Layer 2 master problem in (4.16) which leads to

$$\mu(m+1) = \Big[\mu(m) - \xi_1(m) \times (P_T - \sum_{k=1}^{K}\sum_{i=1}^{n_F} \tilde{P}_{i,k})\Big]^+, \tag{4.22}$$

$$\gamma(m+1) = \Big[\gamma(m) - \xi_2(m) \times (\sum_{k=1}^{K}\sum_{i=1}^{n_F} s_{i,k}\tilde{R}_{i,k} - r)\Big]^+, \tag{4.23}$$

$$\theta_{i,k}(m+1) = \Big[\theta_{i,k}(m) - \xi_3(m) \times (I - \sum_{j\neq k}(\frac{2}{\varepsilon})P_{i,j}s_{i,j}l_k g_k)\Big]^+ \quad \forall i,k, \tag{4.24}$$

where index $m \geq 0$ is the iteration index and $\xi_u(m)$, $u \in \{1, 2, 3\}$, are positive step sizes. Since the transformed problem for a given parameter q is concave in nature, it is guaranteed that the iteration between Layer 1 and Layer 2 converges to the optimal solution of (4.14) in the main loop, if the chosen step sizes satisfy the infinite travel condition [75, 88]. Then, the updated Lagrange multipliers in (4.22)-(4.24) are used for solving the subproblems in (4.17) via updating the resource allocation policies.

Although equations (4.18)-(4.24) provide a solution for solving the main loop problem (line 4, Table 4.1), (4.19) involves non-causal knowledge of the subcarrier allocation process for all users. This can be easily resolved by the coordinate ascent method [87, 77] with the previous derived solutions for Layer 1 and Layer 2. The coordinate ascent method is outlined in Table 4.2. The coordinate ascent method for solving (4.14) is implemented as two nested layers in Algorithm 4.2. In Table 4.2, $t_{\max}$ and $m_{\max}$ are the maximum number of iterations for the two nested loops. $\mathcal{P}_t$, $\mathcal{A}_t$, $\mathcal{R}_t$, and $\mathcal{S}_t$ are the power allocation, antenna allocation, data rate adaptation, and subcarrier allocation policies in the t-th iteration, respectively. Layer 1, i.e., line 4 to line 7, is solving the maximization in (4.16) by using the coordinate ascent method for a given set of Lagrange multipliers. In particular, in line 5, we first keep the subcarrier allocation fixed and optimize the power allocation policy and antenna allocation policy. Then, in line 6, we use the optimized policies $\mathcal{P}_{t+1}$, $\mathcal{A}_{t+1}$, and $\mathcal{R}_{t+1}$ from line 5 to optimize the subcarrier allocation policy. Convergence of Layer 1 to the optimum point for a given set of Lagrange multipliers is ensured for convex optimization problems [77]. On the other hand, Layer 2, i.e., line 3 to line 9, solves the minimization in (4.16) by updating the Lagrange multipliers.

A summary of the overall algorithm is given in Table 4.1. In each iteration of the main loop (line 3 in Table 4.1), we solve the main loop problem in (4.14) for

Table 4.2: Coordinate Ascent Method.

Algorithm 4.2 Coordinate Ascent Method for Solving (4.14)

1: Set the iteration indices $t = 0, m = 0$, maximum number of iterations $t_{\max}, m_{\max}$
2: Initialize the Lagrange multipliers $\mu, \gamma, \theta_{i,k}$ and resource allocation policies $\{\mathcal{P}_t, \mathcal{A}_t, \mathcal{R}_t, \mathcal{S}_t\}$ for $t = 0$
3: **repeat** {Layer 2}
4: **repeat** {Layer 1}
5: Solve the power allocation, antenna allocation, and data rate allocation by using (4.18) and (4.19) for all subcarriers with subcarrier allocation policy $\mathcal{S}_t$. Assign the solutions to $\mathcal{P}_{t+1}$, $\mathcal{A}_{t+1}$, and $\mathcal{R}_{t+1}$
6: Solve the subcarrier allocation for all subcarriers by using (4.21) together with $\mathcal{P}_{t+1}$, $\mathcal{A}_{t+1}$, and $\mathcal{R}_{t+1}$. Assign the solution to $\mathcal{S}_{t+1}$; $t = t + 1$
7: **until** Convergence= **true** or $t = t_{\max}$
8: Update μ, γ, and $\theta_{i,k}$ by gradient method according to (4.22)-(4.24); $m = m+1$
9: **until** Convergence= **true** or $m = m_{\max}$
10: **return** $\{\mathcal{P}_t, \mathcal{A}_t, \mathcal{R}_t, \mathcal{S}_t\}$ as $\{\mathcal{P}', \mathcal{A}', \mathcal{R}', \mathcal{S}'\}$ to line 4 in Algorithm 4.1

a given parameter q by dual decomposition and the coordinate ascent method with Algorithm 4.2. After obtaining the solution in the main loop, we update parameter q and use it for solving the main loop problem in the next iteration. This procedure is repeated until the proposed algorithm converges.

Note that the algorithm in Table 4.1 is only based on path loss and shadowing information[16], thanks to the large numbers of antennas. In other words, we only execute the algorithm according to the coherence time of shadowing and path loss which is in the order of seconds for low mobilities users.

4.5 Results

In this section, we evaluate the system performance through simulations. A single cell with a radius of 1 km is considered, cf. Figure 4.1. The simulation parameters can be

[16]The calculation of power, data rate, antenna, and subcarrier allocations are based on the path loss information and shadowing information. However, the computation of precoding vector $\hat{\mathbf{f}}_{i,k} = \frac{\hat{\mathbf{h}}_{i,k}}{\|\hat{\mathbf{h}}_{i,k}\|}$ requires multipath information.

Table 4.3: System parameters

Cell radius	1 km
Reference distance d_0	35 m
Users distribution	Uniformly distributed between d_0 and cell boundary
Small scale fading distribution	Rayleigh fading with zero mean and unit variance
Carrier center frequency	2.5 GHz
Number of subcarriers n_F	256
Total bandwidth	5 MHz
Subcarrier bandwidth	19.5 kHz
Noise power per subcarrier N_0W	-131 dBm
Channel path loss model	3GPP- Urban Micro
Lognormal shadowing	Standard deviation of 8 dB
Circuit power per antenna P_C	30 dBm [98]
Static circuit power consumption P_0	40 dBm [89]
Minimum data rate requirement r	7 bit/s/Hz
Power amplifier (PA) power efficiency	$1/\rho = 0.2$
Constant back-off factor δ	0.3
CSIT error variance σ_e^2 (unless specified)	0.1
Outage probability requirement ε	0.1
N_{th}	33
$N_{\max}$	100

found in Table 4.3. In practice, the values of P_C and P_0 depend on the application-specific integrated circuits (ASIC) and the implementation. The values of P_C and P_0 adopted in this chapter are for illustration purpose and are based on [98] and [89], respectively. Note that if the resource allocator is unable to guarantee the minimum data rate in a time slot, we set the energy efficiency and outage capacity in that particular time slot to zero to account for the corresponding failure. On the other hand, in the following results, the "number of iterations" is referring to the number of iterations of Algorithm 4.1 in Table 4.1.

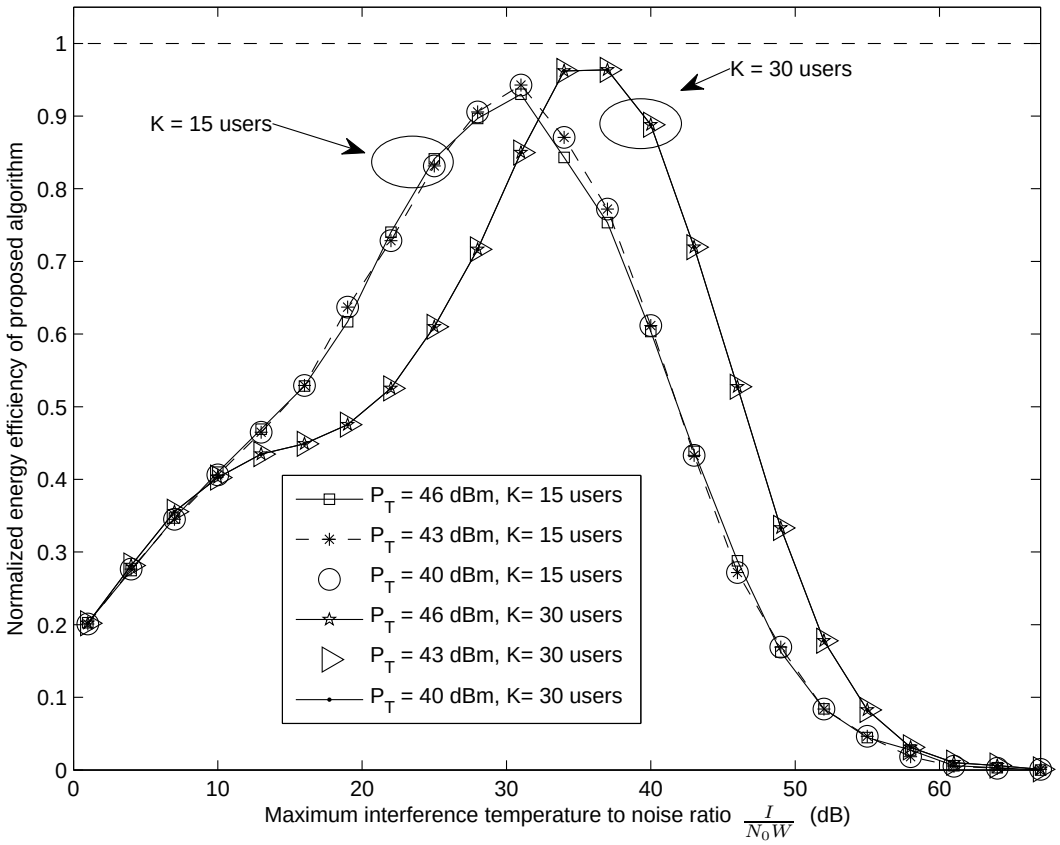


Figure 4.2: The normalized performance of the proposed algorithm versus the maximum interference-temperature-to-noise ratio $\frac{I}{N_0 W}$ for different values of P_T and different numbers of users. The y-axis is normalized by the performance of the optimal algorithm.

4.5.1 Energy Efficiency versus Maximum Inter-user Interference Temperature I

In this section, we focus on the impact of the value of I on the system performance. As can be seen from (4.12) and (4.13), the multi-user interference temperature I, which is the key for transforming the main loop problem in (4.14) into a convex optimization problem, plays an important role in the proposed resource allocation

algorithm. The value of I puts a limit on the subcarrier reuse by controlling the amount of interference temperature[17]. For instance, by setting $I = 0$, each subcarrier can be used by one user only. On the contrary, $I \gg 1$ allows all users to transmit simultaneously on the same subcarrier. Figure 4.2 shows the performance of the proposed algorithm versus the value of I for different P_T and different numbers of users K. The y-axis is normalized by the close-to-optimal performance obtained by solving the original non-convex problem (4.14) using iterative spectrum balancing together with antenna constraint relaxation[18], such that it illustrates the achievable percentage of the performance of the reference scheme. The x-axis is the interference temperature-to-noise ratio, i.e., $\frac{I}{N_0 W}$. It can be seen that for a wide range of $\frac{I}{N_0 W}$, we can achieve around 95% of the optimal performance and enjoy the convexity of the transformed problem. Furthermore, the choice of I is highly dependent on the number of users. This is because a higher value of $\frac{I}{N_0 W}$ can be tolerated for a larger number of users as the selected users can better cope with the co-channel interference in each subcarrier due to MUD. On the other hand, as expected, the optimal value of I is not sensitive to P_T when P_T is large, since the resource allocator clips the total transmit power for energy efficiency maximization, cf. (4.18).

In the following simulations, a fixed value of I is chosen for the proposed algorithm in each simulation point, such that we always achieve more than 95% of the average performance of the optimal resource allocation algorithm.

[17] In practice, suitable values for I for implementing the proposed algorithm can be found in an off-line manner.

[18] The upper bound is obtained by assuming perfect channel state information is available at the base station. In addition, we remove constraints C3, C5, C6, and C7 from the optimization problem in (4.14) for obtaining the upper bound performance. The resulting optimization problem can be solved by using the Dinkelbach method and the spectrum balancing algorithm from [87]. Note that the spectrum balancing algorithm is a close-to-optimal numerical method for solving non-convex optimization problems in multicarrier systems. However, it converges slowly and is computational infeasible for large size systems.

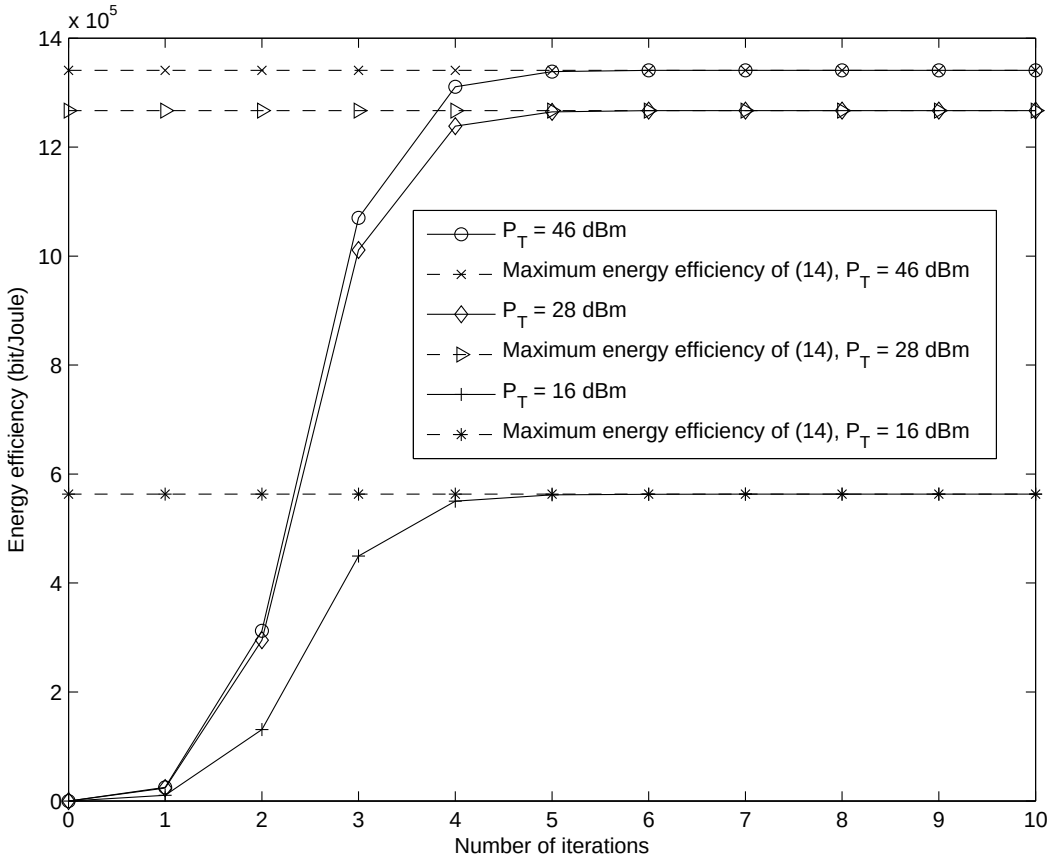


Figure 4.3: Energy efficiency versus the number of iterations with $K = 15$ users for different maximum transmit powers P_T and channel estimation error variance $\sigma_e^2 = 0.1$.

4.5.2 Convergence of Iterative Algorithm

Figure 4.3 illustrates the evolution of the proposed iterative algorithm for different values of the maximum transmit power, P_T, at the BS and $K = 15$ users. The results in Figure 4.3 were averaged over 100000 independent adaptation processes where each adaptation process involves different realizations of path loss, shadowing, and multipath fading. It can be observed that the iterative algorithm converges to the optimal value within 10 iterations for all considered numbers of transmit antennas.

In other words, the maximum system energy efficiency can be achieved within a few iterations.

4.5.3 Energy Efficiency and Average Outage Capacity versus Transmit Power

Figure 4.4 illustrates the energy efficiency versus the total transmit power for $K = 15$ users. The number of iterations for the proposed iterative resource allocation algorithm is 5 and 10. The performance difference between 5 iterations and 10 iterations is negligible which confirms the practicality of our proposed iterative resource allocation algorithm. It can be observed that when the maximum transmit power at the power amplifier is large enough, e.g., $P_T \geq 40$ dBm, the energy efficiency of the proposed algorithm approaches a constant value since the resource allocator is not willing to consume more power or activate more antennas, when the maximum energy efficiency is achieved. For comparison, Figure 4.4also contains the energy efficiency of a baseline resource allocation scheme in which resource allocation is performed in the same manner as in the proposed scheme, except that the number of transmit antennas is fixed to $N_{T_{i,k}} = N_{th}, 40, 60, 80, \forall i, k$, respectively. In other words, the baseline scheme optimizes energy efficiency only in terms of resource allocation policies $\{\mathcal{P}, \mathcal{R}, \mathcal{S}\}$, while the proposed algorithm optimizes energy efficiency in terms of resource allocation policies $\{\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S}\}$. It can be observed that activating a fixed number of transmit antennas $N_{T_{i,k}}$ degrades the system performance in terms of energy efficiency. This is because in the baseline scheme, either more power is consumed by the circuitries for operating the antennas or the number of antennas is not large enough for satisfying the minimum data rate requirement. On the other hand, in the high transmit power regime, the performance gain of the proposed algorithm over

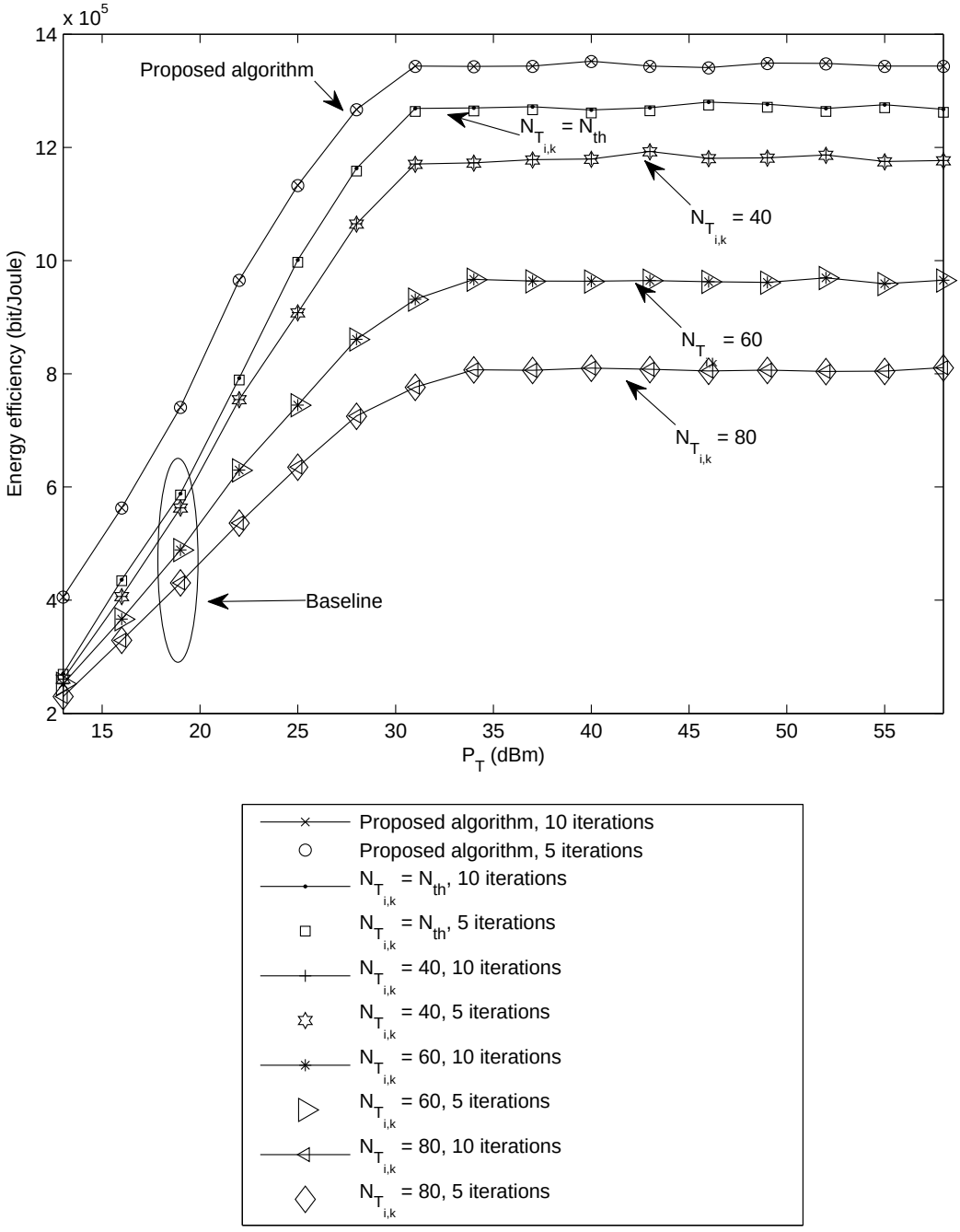


Figure 4.4: Energy efficiency versus maximum transmit power, P_T, for different resource allocation algorithms with channel estimation error variance $\sigma_e^2 = 0.1$. The minimum required number of antennas is $N_{th} = 24$.

the baseline scheme with small $N_{T_{i,k}}$ is reduced. This is due to the fact that in the high transmit power regime, the data rate requirement is satisfied because of the high transmit power and the proposed algorithm tends to use the minimum number of antennas. In fact, the circuit power required for activating an extra antenna is relatively high, compared to the power consumed in the RF. Therefore, the proposed algorithm activates a relatively small number of antennas in the high transmit power regime and thus the performance gain due to antenna allocation becomes less significant.

Figure 4.5 shows the average outage capacity versus maximum transmit power P_T for $K = 15$ users. We compare the system performance of the proposed algorithm again with the baseline resource allocator. The number of iterations in the proposed algorithm is set to 5 and 10. It can be observed that the average outage capacity of the proposed algorithm approaches a constant in the high transmit power regime. This is because the proposed algorithm clips the transmit power at the BS in order to maximize the system energy efficiency. We note that, as expected, the baseline scheme resource allocator achieves a higher average outage capacity than the proposed algorithm in the high transmit power regime for most cases (except for $N_{T_{i,k}} = N_{th}$), since the proposed algorithm tends to use a smaller number of antennas. However, the superior average outage capacity of the baseline scheme comes at the expense of low energy efficiencies. On the contrary, in the low transmit power regime, i.e., $P_T \leq 25$ dBm, the proposed algorithm has a higher average outage capacity than the baseline scheme with $N_{T_{i,k}} \leq 60$ since the baseline scheme is not able to meet the data rate constraint due to insufficient numbers of antennas. On the other hand, an increasing number of antennas in the baseline scheme benefits the average outage capacity due to an improved beamforming gain. However, there is a diminishing return when $N_{T_{i,k}}$ is large due to the *channel hardening* effect [28] in the desired

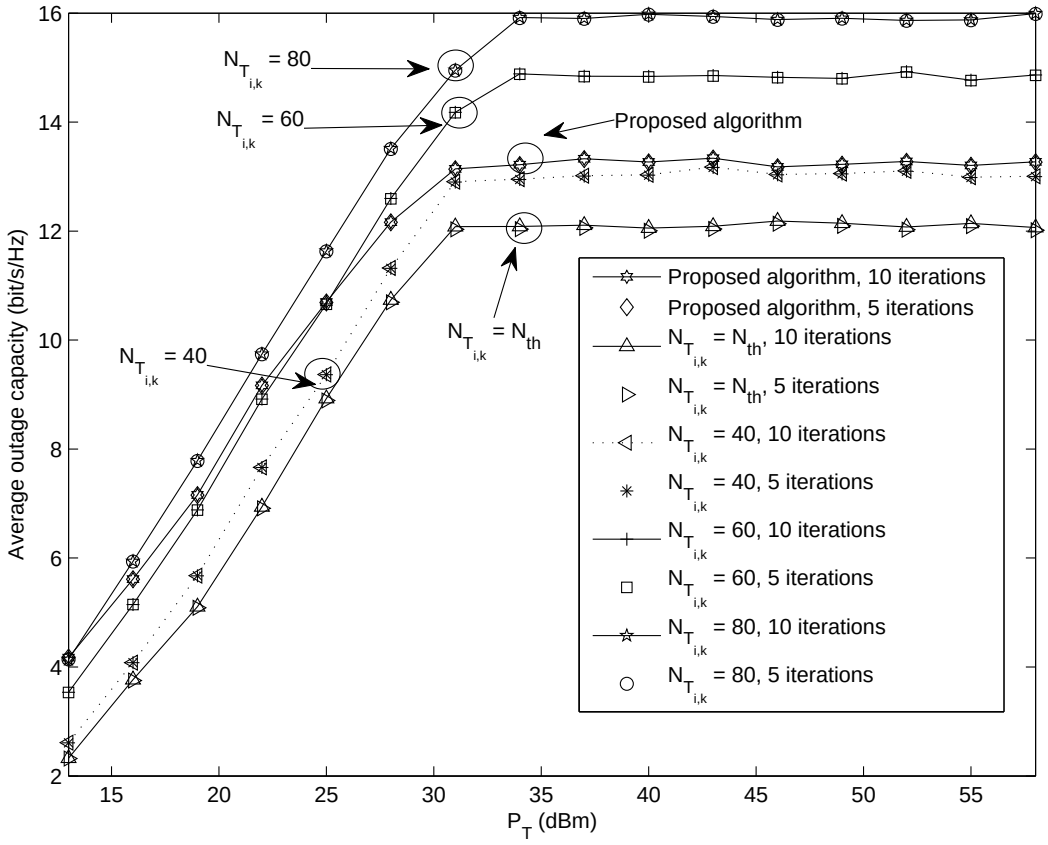


Figure 4.5: Average outage capacity (bit/s/Hz) versus maximum transmit power, P_T, for different resource allocation algorithms, channel estimation error variance $\sigma_e^2 = 0.1$, and $K = 15$ users.

channels.

Figure 4.6 depicts the average total power consumption, i.e., $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})\}$, versus maximum transmit power P_T for the proposed algorithm and the baseline scheme for 10 iterations. In the regime of $P_T \leq 30$ dBm, the proposed algorithm consumes more power than the baseline scheme with $N_{T_{i,k}} \leq 40$. This is because more antennas have to be activated for satisfying the data rate requirement. However, as the maximum transmit power allowance P_T increases, the proposed algorithm gradually approaches a constant power consumption since neither further increasing

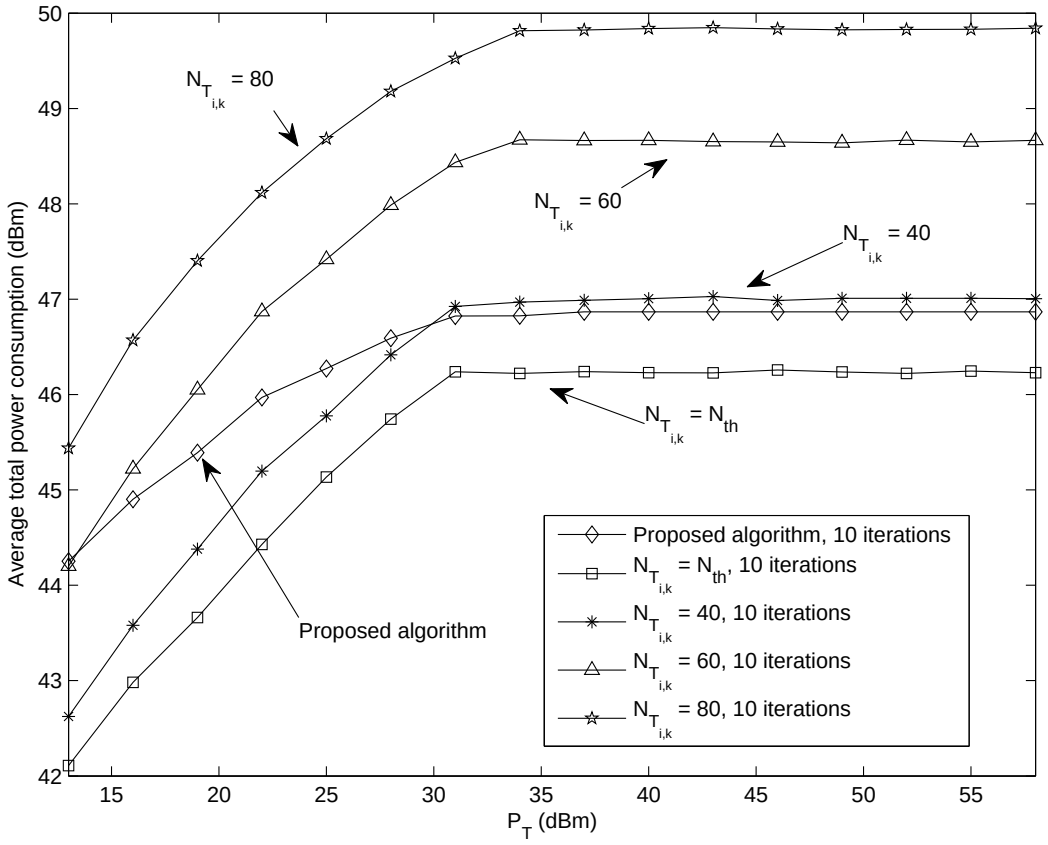


Figure 4.6: Average total power consumption, $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})\}$, versus maximum transmit power, P_T, for different resource allocation algorithms, channel estimation error variance $\sigma_e^2 = 0.1$, 10 iterations, and $K = 15$ users.

the transmit power nor activating more antennas benefits the system energy efficiency.

4.5.4 Energy Efficiency versus Number of Users

Figure 4.7 depicts the energy efficiency versus the number of users. Different CSIT error variances σ_e^2, $P_T = 46$ dBm, and 10 iterations of the proposed algorithm are considered. It can be observed that the energy efficiency grows with the number

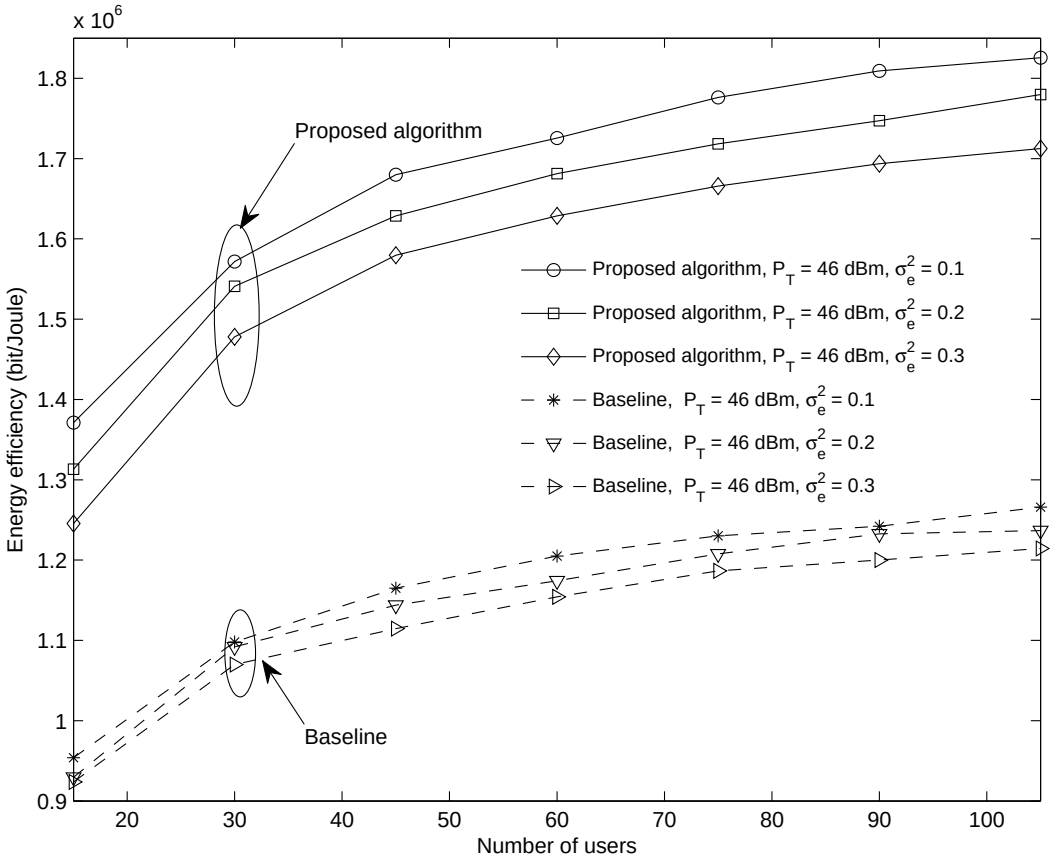


Figure 4.7: Energy efficiency (bit-per-Joule) versus the number of users K for different resource allocation algorithms, different channel estimation error variances σ_e^2, and a maximum transmit power of $P_T = 46$ dBm.

of users since the proposed resource allocation and scheduling algorithm is able to exploit MUD. In general, MUD introduces an extra power gain [28, Chapter 6.6] to the system which provides further energy savings. Indeed, since a large number of transmit antennas reduces the multipath propagation fluctuations in each channel and causes *channel hardening*, the potentially achievable MUD gain due to the multipath channel vanishes. Yet, the MUD gain obtained from path loss and shadowing is still beneficial for the system performance in terms of energy efficiency. For comparison, Figure 4.7 also contains the energy efficiency of the baseline scheme mentioned in

Section 4.5.3 with $N_{T_{i,k}} = 60, \forall i, k$. Figure 4.7 shows that although the baseline scheme is able to exploit MUD, the performance of the proposed resource allocation algorithm is superior to the baseline scheme in all considered scenarios, due to the optimization of the number of antennas.

4.6 Conclusions

In this chapter, we formulated the resource allocation for energy efficient OFDMA systems with a large number of antennas as a mixed non-convex and combinatorial optimization problem, in which circuit power consumption, minimum data rate requirements, and outage probability constraints were taken into consideration. An efficient iterative resource allocation algorithm with closed-form power adaption, antenna allocation, data rate adaption, and subcarrier allocation was derived for maximization of the number of received bit-per-Joule at the users. Simulation results did not only show that the proposed algorithm converges to the optimal solution within a small number of iterations, but demonstrated also the trade-off between energy efficiency and the number of transmit antennas: The use of a large number of antennas is always beneficial for the system outage capacity, even if the CSIT is imperfect. However, an exceedingly large number of antennas may not be a cost effective solution for improving the system performance, at least not from an energy efficiency point of view.

Chapter 5

Energy-Efficient Resource Allocation in Multi-Cell OFDMA Systems with Limited Backhaul Capacity

5.1 Introduction

In the previous chapter, we have proposed a resource allocation algorithm for maximizing the energy efficiency of OFDMA systems with a large number of BS antennas. The inherent degrees of freedom of multiple antennas provide a substantial performance gain compared to single antenna systems. In practice, instead of using co-located antennas at one BS, cooperative communication can also be used to exploit the degrees of freedom in the spatial domain by using distributed single antennas [99]-[104]. A particularly interesting cooperative communication approach is BS cooperation for mitigation of strong multi-cell interference due to aggressive/universal frequency reuse in the network [105]. Over the past decade, a number of interference mitigation techniques have been proposed in the literature, including SIC and interference nulling through multiple antennas, for alleviating the negative side-effects of aggressive/universal frequency reuse [106]-[108]. Unfortunately, these techniques may be too complex for low-power battery driven mobile receiver units. On the con-

trary, BS cooperation, which is known as network MIMO, shifts the signal processing burden to the BSs and provides a promising system performance [109]-[114]. In particular, all BSs share the CSI and data of the users through backhaul communication links, which enables coordinated transmission. In [109], the sum-rate of multi-cell ZFBF systems was studied under the assumption of the Wyner interference model for a large number of users. In [110] and [111], the authors derived the optimal block diagonalization precoding matrix and the optimal max-min beamformer in multi-cell environments, respectively. However, the results in [109]-[111] are based on the ideal backhaul assumption such that an unlimited amount of control signals, user channel information, and precoding data can be exchanged. In practice, the backhaul capacity can be limited due to the deployment costs of the backhaul links. Besides, if a multi-carrier system is considered, the results in [109]-[111] which are valid for single-carrier transmission, may no longer be applicable. Furthermore, numerous resource allocation algorithms were designed for different system configurations by utilizing only the local CSI in multicell systems, e.g. [112]-[114]. Yet, this kind of cooperation may not be able to fully exploit the potential performance gains achievable by BS cooperation, since the backhaul capacity is not fully utilized. On the other hand, user assignment and BS assignment in multi-cell OFDMA systems with limited backhaul capacity constraints were studied in [115] and [116], respectively. In [117], the authors proposed a dynamic frequency allocation with fractional frequency reuse and equal power allocation across all cooperating BSs. In all studies [109]-[117], a substantial capacity gain and better interference management are reported, compared to non-cooperative systems. Yet, the advantages of BS cooperation do not come for free. They have significant financial implications for service providers due to the high power consumption in electronic circuitries, RF transmission, and data exchange via

backhaul links. Thus, *energy efficiency* (bit-per-Joule) may be a better performance metric compared to *system capacity* (bit-per-second-per-Hz) in evaluating the utilization of resources [118]-[120]. However, energy efficiency in multi-cell systems with limited backhaul has not been considered in the literature, e.g. [109]-[117], at least not from resource allocation point of view.

In this chapter, we address the above issues. For this purpose, we formulate the resource allocation problem for energy efficient communication in multi-cell OFDMA systems with limited backhaul capacity as an optimization problem. By exploiting the properties of fractional programming, the considered non-convex optimization problem[19] in fractional form is transformed into an equivalent optimization problem in subtractive form with a tractable solution, which can be found with an iterative algorithm. In each iteration, a low complexity user selection policy is computed. Besides, we show that the duality gap for the resulting power allocation problem is zero when the number of subcarriers is sufficiently large, despite the non-convexity of the problem. As a result, dual decomposition is used to derive a closed-form power allocation solution for maximization of the network energy efficiency.

5.2 Multi-Cell OFDMA Network Model

5.2.1 Multi-Cell System Model and Central Unit

We consider a multi-cell OFDMA network which consists of a total of M BSs and K mobile users. All transceivers are equipped with a single antenna, cf. Figure 5.1. We assume universal frequency reuse and the M BSs share a total bandwidth of $\mathcal{B}$

[19]Non-convex optimization is a general terminology referring to an optimization problem neither minimizing a convex function over convex sets, nor maximizing a concave function over concave sets.

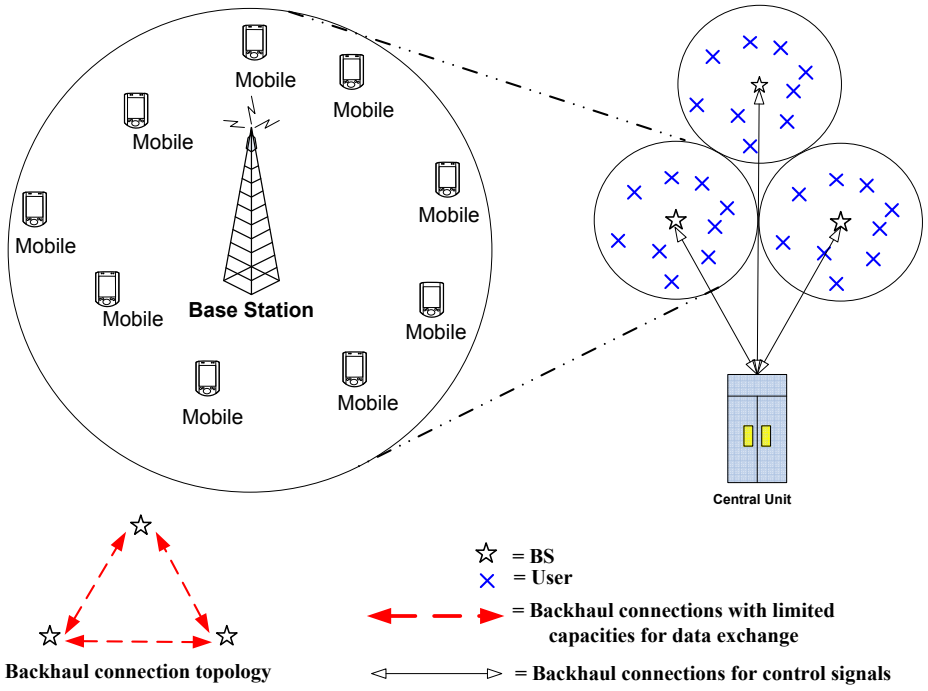


Figure 5.1: A multi-cell system with $M = 3$ cells with a fully connected backhaul link topology. There are in total $K = 27$ users in the system. Each transceiver is equipped with a single antenna.

Hertz. The global CSI is assumed to be perfectly known at a *central unit* and all computations are performed in this unit. Based on the available CSI, the central unit decides the resource allocation policy and broadcasts it to all BSs via backhaul connections which are dedicated to control signals only. On the other hand, all BSs are cooperating with each other by sharing the CSI and the data symbols of all selected users via capacity limited backhaul communication links. Note that the energy consumptions incurred by exchanging CSI and other overheads such as control signals are not considered here since they are relatively insignificant compared to the resources used for data exchange.

5.2.2 OFDMA Channel Model

We consider an OFDMA system with n_F subcarriers. The channel impulse response is assumed to be time-invariant within each frame. Suppose user $k \in \{1, \ldots, K\}$ is associated with BS $m \in \{1, \ldots, M\}$. Let $w^k_{B_m}(i)$ be the precoding coefficient used by BS m to in subcarrier $i \in \{1, \ldots, n_F\}$ for user k. Then, the transmitted signal from BS m to all selected users on subcarrier i is given by

$$\sum_{k\in\mathcal{S}(i)} x^k_m(i) \;=\; \sum_{k\in\mathcal{S}(i)} w^k_{B_m}(i)\sqrt{P^k_{B_m}(i)}u^k(i), \tag{5.1}$$

where $x^k_m(i) = w^k_{B_m}(i)\sqrt{P^k_{B_m}(i)}u^k(i)$ is the pre-coded signal transmitted from BS m for user k on subcarrier i, $P^k_{B_m}(i)$ is the transmit power for the link between BS m and user k in subcarrier i, $u^k(i)$ is the transmitted information symbol for user k on subcarrier i, and $\mathcal{S}(i)$ is the user set selected for using subcarrier i and the cardinality of the set is $|\mathcal{S}(i)| \leq M, \forall i$.

The received signal from M BSs at user k on subcarrier i is given by

$$\begin{aligned} Y^k(i) \;&=\; \left(\sum_{c=1}^{M} H^k_{B_c}(i)w^k_{B_c}(i)\sqrt{P^k_{B_c}(i)l^k_{B_c}}\right)u^k(i) \\ &+\; \underbrace{\sum_{m=1}^{M}\sum_{\substack{j\in\mathcal{S}(i)\\ j\neq k}} \sqrt{P^j_{B_m}(i)l^k_{B_m}}H^k_{B_m}(i)w^j_{B_m}(i)u^j(i)}_{\text{Multiple Access Interference}} + z^k(i), \end{aligned} \tag{5.2}$$

where $l^k_{B_m}$ represents the path loss between BS m and user k, $z^k(i)$ is the AWGN in subcarrier i at user k with zero mean and variance σ^2_z, and $H^k_{B_m}(i)$ is the small scale fading coefficient between BS m and user k in subcarrier i.

5.2.3 Backhaul Model

In practice, the backhaul signal model depends on the specific implementation. For instance, a digital subscriber line (DSL) and an optical fiber are able to deliver high data rates by using orthogonal frequency division multiplexing (OFDM) and wavelength division multiplexing (WDM), respectively. Yet, the media over which is transmitted with both techniques are different. In order to provide a general model for the backhaul, we do not assume a particular type/medium for the backhaul. Instead, we focus on the backhaul capacity of the N_m outgoing backhaul connections of BS m, i.e., $R_{B_{m_{N_m}}}$. The value of N_m depends on the backhaul connection topology. For instance, a fully connected topology in a 3-cell system, cf. Figure 5.1, requires $N_m = 2$ outgoing connections for each BS. Furthermore, we assume that each backhaul has a fixed average power consumption of P_{BH}.

5.3 Resource Allocation and Scheduling

5.3.1 Instantaneous Channel Capacity

In this subsection, we define the adopted system performance measure. Given perfect CSI at the receiver, the maximum channel capacity between all the cooperating BSs and user k on subcarrier i with subcarrier bandwidth $\frac{\mathcal{B}}{n_F}$ is given by

$$C^k(i) = \frac{\mathcal{B}}{n_F} \log_2 \Big(1 + \Gamma^k(i)\Big), \tag{5.3}$$

$$\Gamma^k(i) = \frac{\Big| \sum_{c=1}^{M} H^k_{B_c}(i) w^k_{B_c}(i) \sqrt{P^k_{B_c}(i) l^k_{B_c}} \Big|^2}{\sigma_z^2 + I^k(i)}, \tag{5.4}$$

$$I^k(i) = \sum_{\substack{j \in \mathcal{S}(i) \\ j \neq k}} \Big| \sum_{m=1}^{M} \sqrt{P^j_{B_m}(i)} w^j_{B_m}(i) \sqrt{l^k_{B_m}} H^k_{B_m}(i) \Big|^2, \tag{5.5}$$

where $\Gamma^k(i)$ and $I^k(i)$ are the received SINR and the received interference power at user k on subcarrier i, respectively.

The *weighted system capacity* is defined as the total number of bits successfully delivered to the K mobile users and is given by

$$U(\mathcal{P}, \mathcal{W}, \mathcal{S}) = \sum_{m=1}^{M} \sum_{k \in \mathcal{A}_m} \alpha_k \sum_{i=1}^{n_F} s^k(i) C^k(i), \tag{5.6}$$

where $\mathcal{P}$, $\mathcal{W}$, and $\mathcal{S}$ are the power, precoding coefficient, and subcarrier allocation policies, respectively. $\mathcal{A}_m$ is the user admission set of BS m and $s^k(i) \in \{0, 1\}$ is the subcarrier allocation indicator. $0 < \alpha_k \leq 1$ is a positive constant provided by the upper layers, which allows the resource allocator to give different priorities to different users and to enforce certain notions of fairness. On the other hand, for designing a resource allocation algorithm for energy efficient communication, the total power consumption should be included in the optimization objective function. Thus, we model the power dissipation in the system as the sum of two static terms and one dynamic term which can be expressed as

$$U_{TP}(\mathcal{P}, \mathcal{W}, \mathcal{S}) = P_C \times M + \delta \times P_{BH} + \sum_{m=1}^{M} \sum_{k=1}^{K} \sum_{i=1}^{n_F} \varepsilon P^k_{B_m}(i) |w^k_{B_m}(i)|^2 s^k(i), \tag{5.7}$$

where P_C is the constant signal processing power required at each BS which includes the power dissipations in the transmit filter, mixer, frequency synthesizer, and digital-to-analog converter, etc. $P_C \times M$ represents the total signal processing power consumed by the M BSs. The second term in (5.7) denotes the total power dissipation in the backhaul links where δ is an integer variable which indicates the number of backhaul links in the system. For instance, the topology considered in Figure 5.1 requires $\delta = \sum_{m=1}^{M} N_m = 6$ backhaul connections. The last term in (5.7)

represents the total power consumption in the power amplifiers of the M BSs. $\varepsilon \geq 1$ is a constant which accounts for the inefficiency of the power amplifier. For example, if $\varepsilon = 5$, it means that for every 10 Watts of power radiated in the RF, 50 Watts are consumed in the power amplifier and the power efficiency is $\frac{1}{\varepsilon} = \frac{1}{5} = 20\%$.

Hence, the *weighted energy efficiency* of the considered system is defined as the total average number of bits/Joule

$$U_{eff}(\mathcal{P}, \mathcal{W}, \mathcal{S}) \;\; = \;\; \frac{U(\mathcal{P}, \mathcal{W}, \mathcal{S})}{U_{TP}(\mathcal{P}, \mathcal{W}, \mathcal{S})}. \tag{5.8}$$

5.3.2 Optimization Problem Formulation

The optimal power allocation policy, $\mathcal{P}^*$, precoding policy, $\mathcal{W}^*$, and subcarrier allocation policy, $\mathcal{S}^*$, can be obtained by solving

$$\begin{aligned}
&\max_{\mathcal{P},\mathcal{W},\mathcal{S}} \; U_{eff}(\mathcal{P}, \mathcal{W}, \mathcal{S}) \\
\text{s.t. } &\text{C1:} \sum_{k=1}^{K} \sum_{i=1}^{n_F} |w_{B_m}^k(i)|^2 P_{B_m}^k(i) s^k(i) \leq P_{T_m}, \quad \forall m, \\
&\text{C2:} \sum_{m=1}^{M} \sum_{k \in \mathcal{A}_m}^{K} \sum_{i=1}^{n_F} s^k(i) C^k(i) \geq R_{\min}, \\
&\text{C3:} \sum_{k \in \mathcal{A}_m}^{K} \sum_{i=1}^{n_F} s^k(i) C^k(i) \leq R_{\max_m} = \min\{R_{B_{m_1}}, R_{B_{m_2}}, \ldots, R_{B_{m_{N_m}}}\}, \quad \forall m, \\
&\text{C4:} \sum_{k=1}^{K} s^k(i) \leq M, \quad \forall i, \quad \text{C5: } s^k(i) = \{0, 1\}, \; \forall i, k, \\
&\text{C6: } P_{B_m}^k(i) \geq 0, \; \forall i, k, m,
\end{aligned} \tag{5.9}$$

where C1 is an individual power constraint of BS m. The value of P_{T_m} in C1 puts a limit on the amount of interference generated to the non-cooperative cells in the downlink. C2 specifies the minimum system data rate requirement $R_{\min}$. Note that

although variable $R_{\min}$ in C2 is not an optimization variable in this chapter, a balance between energy efficiency and aggregate system capacity can be struck by varying $R_{\min}$. Operator $\min\{\cdot\}$ in C3 accounts for the fact that the system capacity contributed by a BS is limited by the bottleneck backhaul capacity of that BS. Indeed, C3 is a generalized constraint on the backhaul capacities which is applicable to different topologies such as *star connection topology* [121, Chapter 1] and *fully connected topology*[20] [121, Chapter 4]. Besides, C3 puts a limit on the maximum data rate transmission at each BS due to the limited backhaul capacities. If $R_{B_{\max_m}} \to \infty \, \forall m$, then C3 is always satisfied automatically. i.e., the backhaul capacity is much larger than the wireless link capacity. C4 is the subcarrier reuse constraint. C4 and C5 are imposed to guarantee that each subcarrier can be shared by M users, but each user can only use a subcarrier once. In other words, selected users are not allowed to multiplex different messages on the same subcarrier, since a sophisticated receiver would be required at each user, such as an SIC receiver, to recover more than one messages.

5.4 Solution of the Optimization Problem

The objective function in (5.9) is a ratio of two functions which generally results in a non-convex function. Thus, a brute force approach is required for obtaining a global optimal solution. However, such a method has exponential complexity with respect to the numbers of subcarriers and the number of users which makes it computationally infeasible even for small size systems. In order to derive an efficient resource allocation

[20] In a star connection topology, the network has a hub to convey messages. In other words, all the message exchanges between BSs have to first pass through this hub. In a fully connected topology, all BSs are connected to each other. As a result, a fully connected network does not require the usage of switches/hubs.

algorithm, we introduce the following transformation.

5.4.1 Transformation of the Objective Function

The objective function in (5.9) can be classified as nonlinear fractional program [84]. For the sake of notational simplicity, we define $\mathcal{F}$ as the set of feasible solutions of the optimization problem in (5.9). Without loss of generality, we define the maximum energy efficiency q^* of the considered system as

$$q^* = \frac{U(\mathcal{P}^*, \mathcal{W}^*, \mathcal{S}^*)}{U_{TP}(\mathcal{P}^*, \mathcal{W}^*, \mathcal{S}^*)} = \max_{\mathcal{P},\mathcal{W},\mathcal{S}} \frac{U(\mathcal{P}, \mathcal{W}, \mathcal{S})}{U_{TP}(\mathcal{P}, \mathcal{W}, \mathcal{S})}, \quad \forall \{\mathcal{P}, \mathcal{W}, \mathcal{S}\} \in \mathcal{F}. \tag{5.10}$$

By Theorem 3.1 in Chapter 3, the maximum energy efficiency q^* is achieved if and only if

$$\begin{aligned} \max_{\mathcal{P},\mathcal{W},\mathcal{S}} \quad & U(\mathcal{P}, \mathcal{W}, \mathcal{S}) - q^* U_{TP}(\mathcal{P}, \mathcal{W}, \mathcal{S}) \\ = \quad & U(\mathcal{P}^*, \mathcal{W}^*, \mathcal{S}^*) - q^* U_{TP}(\mathcal{P}^*, \mathcal{W}^*, \mathcal{S}^*) = 0, \end{aligned} \tag{5.11}$$

for $U(\mathcal{P}, \mathcal{W}, \mathcal{S}) \geq 0$ and $U_{TP}(\mathcal{P}, \mathcal{W}, \mathcal{S}) > 0$.

By Theorem 3.1, for any optimization problem with an objective function in fractional form, there exists an equivalent[21] objective function in subtractive form, e.g. $U(\mathcal{P}, \mathcal{W}, \mathcal{S}) - q^* U_{TP}(\mathcal{P}, \mathcal{W}, \mathcal{S})$, in the considered case. As a result, we can focus on the equivalent objective function in the rest of the chapter.

[21] Here, "equivalent" means that both problem formulations will lead to the same resource allocation policies.

Table 5.1: Iterative Resource Allocation Algorithm.

Algorithm 5.1 Iterative Resource Allocation Algorithm

Initialize the maximum number of iterations L_{max} and the maximum tolerance ϵ
Set maximum energy efficiency $q = 0$ and iteration index $n = 0$
repeat {Main Loop}
  Solve the inner loop problem in (5.12) for a given q and obtain resource allocation policies $\{\mathcal{P}', \mathcal{W}', \mathcal{S}'\}$
  if $U(\mathcal{P}', \mathcal{W}', \mathcal{S}') - qU_{TP}(\mathcal{P}', \mathcal{W}', \mathcal{S}') < \epsilon$ **then**
    Convergence = **true**
    return $\{\mathcal{P}^*, \mathcal{W}^*, \mathcal{S}^*\} = \{\mathcal{P}', \mathcal{W}', \mathcal{S}'\}$ and $q^* = \frac{U(\mathcal{P}', \mathcal{W}', \mathcal{S}')}{U_{TP}(\mathcal{P}', \mathcal{W}', \mathcal{S}')}$
  else
    Set $q = \frac{U(\mathcal{P}', \mathcal{W}', \mathcal{S}')}{U_{TP}(\mathcal{P}', \mathcal{W}', \mathcal{S}')}$ and $n = n + 1$
    Convergence = **false**
  end if
until Convergence = **true** or $n = L_{max}$

5.4.2 Iterative Algorithm for Energy Efficiency Maximization

In this section, we propose an iterative algorithm (known as the Dinkelbach method [84]) for solving (5.9) with an equivalent objective function. The proposed algorithm is summarized in Table 5.1 and the convergence to the optimal energy efficiency is guaranteed.

Proof: Please refer to [84] for a proof of convergence.

As shown in Table 5.1, in each iteration in the main loop, we solve the following optimization problem for a given parameter q:

$$\max_{\mathcal{P},\mathcal{W},\mathcal{S}} \quad U(\mathcal{P}, \mathcal{W}, \mathcal{S}) - qU_{TP}(\mathcal{P}, \mathcal{W}, \mathcal{S})$$
$$\text{s.t. C1, C2, C3, C4, C5, C6.} \tag{5.12}$$

Solution of the Main Loop Problem

The transformed problem is a mixed combinatorial and non-convex optimization problem. The non-convex nature comes from the power allocation variables and precoding coefficients. The multiuser interference appears in the denominator of the SINR expression in (5.4) which couples the power allocation variables. On the other hand, the combinatorial nature comes from the integer constraint for subcarrier allocation. To obtain an optimal solution, an exhaustive search is needed with complexity $\sum_{g=1}^{M} \binom{K}{g}^{n_F}$, which is computationally infeasible for $K \gg M$. In order to derive an efficient resource allocation algorithm, we solve the above problem in three steps. In the first step, we employ a low complexity sub-optimal user selection scheme. Then, in the second step, we calculate the ZFBF coefficients for a given selected user set of $\mathcal{S}$. In the final step, we optimize the transmit power at each BS for energy efficiency maximization. We note that by fixing resource allocation policies $\{\mathcal{W}, \mathcal{S}\}$, Algorithm 5.1 in Table 5.1 converges to a sub-optimal solution since only the power allocation is optimized for energy efficiency maximization.

Step 1 (Semi-Orthogonal User Selection [122])

We propose an efficient user selection algorithm. Without loss of generality, we define a row vector $\vec{H}_{BS}^{k}(i) = \left[H_{B_1}^{k}(i)\sqrt{l_{B_1}^{k}} \; H_{B_2}^{k}(i)\sqrt{l_{B_2}^{k}} \; \ldots H_{B_M}^{k}(i)\sqrt{l_{B_M}^{k}} \right]$ which represents a super-channel vector between all BSs and user k with elements $H_{B_m}^{k}(i)\sqrt{l_{B_m}^{k}}$, $k \in \{1, \ldots, K\}$, $m \in \{1, \ldots, M\}$, representing the channel coefficient between BS m and user k on subcarrier i. Let $\mathcal{S}_{\perp}(i)$ be a semi-orthogonal user set for subcarrier i. Then, the adopted semi-orthogonal user selection procedure for each subcarrier is summarized in Table 5.2. η in line 7 in Table 5.2 represents a threshold for measuring orthogonality. Note that a user with a higher value of α_k (priority) has a higher

Table 5.2: Semi-Orthogonal User Selection Algorithm.

Algorithm 5.2 Semi-Orthogonal User Selection Algorithm

1: Initialize $\mathcal{T}_t = \{1, \ldots, K\}$, orthogonality parameter η, vector subspace $\Phi = \{\vec{\phi}_{(1)}, \ldots, \vec{\phi}_{(t)}\}$, iteration index $t = 1$, and $\mathcal{S}_\perp(i) = \emptyset$, where $\vec{\phi}_{(t)} \in \mathbb{C}^{1\times M}$.

2: Update $\mathcal{S}_\perp(i) \rightarrow \mathcal{S}_\perp(i) \bigcup \pi(t), \pi(t) = \arg\max_{a\in\mathcal{T}_t} \|\vec{H}^a_{BS}(i)\|^2$, $\vec{\phi}_{(1)} = \vec{H}^{\pi(t)}_{BS}(i)$, $\mathcal{T}_{t+1} = \mathcal{T}_t/\{\pi(t)\}$.

3: **repeat**

4: For each users $k \in \mathcal{T}_t$, calculate a vector $\vec{H}^k_\perp(i) \in \mathbb{C}^{1\times M}$ which is orthogonal to Φ as

$$\vec{H}^k_\perp(i) = \vec{H}^k_{BS}(i) - \sum_{r=1}^{t-1} \frac{\vec{H}^k_{BS}(i)\vec{\phi}^\dagger_{(r)}}{\|\vec{\phi}_{(r)}\|^2}\vec{\phi}_{(r)}.$$

5: Update $\mathcal{S}_\perp(i) \rightarrow \mathcal{S}_\perp(i) \bigcup \pi(t), \pi(t) = \arg\max_{a\in\mathcal{T}_t} \|\vec{H}^a_\perp(i)\|^2$, $\vec{\phi}_{(t)} = \vec{H}^{\pi(t)}_\perp(i)$.

6: **if** $|\mathcal{S}_\perp(i)| \leq M$, **then**

7: Calculate $\mathcal{T}_{t+1}$ as

$$\mathcal{T}_{t+1} = \left\{ k \in \mathcal{T}_t, k \neq \pi(t), \frac{|\vec{H}^k_{BS}(i)\vec{\phi}^\dagger_{(t)}|}{\|\vec{H}^k_{BS}(i)\|\|\vec{\phi}_{(t)}\|} < \eta \times \alpha_k \right\}, \quad t = t + 1.$$

8: **end if**

9: **until** $\mathcal{T}_t = \emptyset$ or $|\mathcal{S}_\perp(i)| = M$

chance of being selected. On the other hand, as $\eta \rightarrow 0$, the selected users in the set are increasingly orthogonal to each other. In other words, users associated with set $\mathcal{S}_\perp(i)$ cause less interference to other users in the set. Note that with the proposed user selection scheme, the search space for each subcarrier decreases from $\sum_{a=1}^{M}\binom{K}{a}$ to $2KM$ and $\frac{2KM}{\sum_{a=1}^{M}\binom{K}{a}} \ll 1$ for $K \gg M$. Note that although the proposed algorithm is suboptimal, it has been shown in [122] that the proposed scheme performs well in combination with zero-forcing beamforming.

Step 2 (Zero-Forcing Beamformimg)

A multi-cell network with full BS cooperation can be interpreted as a MIMO broadcast channel. It can be shown that dirty paper coding (DPC) is optimal in achieving the multiuser broadcast capacity region. However, DPC requires a very high complexity which is considered impractical. On the contrary, although ZFBF is a suboptimal precoding scheme, it is considered a practical solution, due to its linear complexity and promising performance. Besides, it can be shown that the proposed semi-orthogonal user selection algorithm together with ZFBF can achieve the same asymptotic sum capacity performance as DPC [122]. Therefore, we focus on ZFBF in the rest of the chapter.

If ZFBF is used for transmission, the capacity equation in (5.3) can be rewritten as

$$\begin{aligned} C^k(i) &= \frac{\mathcal{B}}{n_F}\log_2\left(1+\Gamma^k(i)\right) \quad \text{where} \\ \Gamma^k(i) &= \frac{\left|\sum_{c=1}^{M}\sqrt{l^k_{B_c}}H^k_{B_c}(i)w^k_{B_c}(i)\right|^2 P^k_{B_m}(i)}{\sigma_z^2} \end{aligned} \tag{5.13}$$

$P^k_{B_m}(i) = P^k_{B_1}(i) = P^k_{B_2}(i) = \ldots = P^k_{B_M}(i)$ is an imposed constraint together with ZFBF[22]. There are two reasons for imposing this constraint. First, it allows us to separate the power allocation variables from the precoding coefficients. Second, it simplifies the design of power control[23]. Let us consider the above scenario and ZFBF precoding with the above assumption, then the total transmit power of the

[22]Indeed, the transmit power from BS m to user k on subcarrier i is $P^k_{B_m}(i)|w^k_{B_m}(i)|^2$ instead of $P^k_{B_m}(i)$. So even if we enforce $P^k_{B_m}(i) = P^k_{B_1}(i) = P^k_{B_2}(i) = \ldots = P^k_{B_M}(i)$, the actual transmit powers to user k from M base stations (BSs) are not identical since in general $|w^k_{B_1}(i)|^2 \neq |w^k_{B_2}(i)|^2 \neq \ldots \neq |w^k_{B_M}(i)|^2$.

[23]The above approach is commonly used in literature for decoupling the power allocation variables from the precoding coefficients and for the design of simple yet efficient resource allocation algorithms for multiple antenna systems.

M BSs to user k on subcarrier i is given by $\sum_{c=1}^{M} P_{B_c}^k(i)|w_{B_c}^k(i)|^2 = P_{B_m}^k(i)$, since $\sum_{c=1}^{M} |w_{B_c}^k(i)|^2 = 1$. In other words, the precoding coefficients are decoupled from the power allocation variables and the precoding coefficients do not increase the total power consumption. Besides, we can directly control the total amount of transmit power from the M BSs to user k on subcarrier i via optimizing $P_{B_m}^k(i)$. Without loss of generality, we assume that user 1 to user k are selected for using subcarrier i, i.e., $\{1,\ldots,k\} \in \mathcal{S}_{\perp}(i)$. Let us define a super channel matrix $\mathbf{H}_B(i) \in \mathbb{C}^{|\mathcal{S}_{\perp}(i)|\times M}$ such that

$$\mathbf{H}_B^T(i) = \left[\left(\vec{H}_{BS}^1(i)\right)^T \left(\vec{H}_{BS}^2(i)\right)^T \ldots \left(\vec{H}_{BS}^k(i)\right)^T\right]. \tag{5.14}$$

Then, the corresponding ZFBF super matrix $\mathbf{B}(i) \in \mathbb{C}^{M\times|\mathcal{S}_{\perp}(i)|}$ can be calculated in the centralized unit and is given by

$$\mathbf{B}(i) = \mathbf{H}_B^{\dagger}(i)\Big(\mathbf{H}_B(i)\mathbf{H}_B^{\dagger}(i)\Big)^{-1}\mathbf{D}(i), \tag{5.15}$$

where $\mathbf{D}(i) \in \mathbb{C}^{|\mathcal{S}_{\perp}(i)|\times|\mathcal{S}_{\perp}(i)|}$ is a diagonal matrix with diagonal elements $\gamma^k(i) = 1/\sqrt{\left[\Big(\mathbf{H}_B(i)\mathbf{H}_B^{\dagger}(i)\Big)^{-1}\right]_{k,k}} = \left|\sum_{c=1}^{M}\sqrt{l_{B_c}^k}H_{B_c}^k(i)w_{B_c}^k(i)\right|$. Note that $\gamma^k(i)$ represents the equivalent channel gain between all BSs and user k on subcarrier i for ZFBF transmission. Hence, the ZFBF coefficient $w_{B_c}^k(i)$ is given by

$$w_{B_c}^k(i) = \Big[\mathbf{B}(i)\Big]_{c,k} \ \forall k \in \mathcal{S}_{\perp}(i) \tag{5.16}$$

and the central unit delivers the relevant ZFBF coefficients to each BS via additional backhaul connections which are dedicated to control signals.

Dual Problem

The final step in solving the main loop problem is to optimize the power allocation. For a given set of selected users and ZFBF transmission, the problem in (5.12) is still non-convex due to constraint C3. In general, a non-zero duality gap exists if we solve (5.12) by solving its dual. However, we will demonstrate that the duality gap is always zero when the number of subcarriers is sufficiently large. This result is summarized in the following theorem.

Theorem 5.1 *Let P and D denote the optimal values of the primal and the dual problem in (5.12), respectively. For a given selected user set and ZFBF transmission, if the number of subcarriers is sufficiently large, then strong duality holds and the duality gap is zero, i.e., $P = D$.*

Proof: Please refer to Appendix G for a proof of Theorem 5.1.

By Theorem 5.1, we solve the main loop problem in (5.12) by solving its dual. For this purpose, we first need the Lagrangian function of the primal problem. Upon rearranging terms, the Lagrangian can be written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\lambda},\boldsymbol{\beta},\gamma,\mathcal{P}) &= \sum_{m=1}^{M}\sum_{i=1}^{n_F}\sum_{k\in\mathcal{A}_m\cap\mathcal{S}_\perp(i)}^{K}(\alpha_k+\gamma-\beta_m)C^k(i)-\gamma R_{\min}+\sum_{m=1}^{M}\beta_m R_{\max_m}\\ &- \sum_{m=1}^{M}\lambda_m\Big(\sum_{i=1}^{n_F}\sum_{k\in\mathcal{S}_\perp(i)}^{K}|w_{B_m}^k(i)|^2P_{B_m}^k(i)-P_{T_m}\Big)\\ &- q\Big(\sum_{m=1}^{M}\sum_{k\in\mathcal{A}_m\cap\mathcal{S}_\perp(i)}^{K}\sum_{i=1}^{n_F}\varepsilon P_{B_m}^k(i)|w_{B_m}^k(i)|^2+\delta P_{BH}+P_C\times M\Big), \quad (5.17)\end{aligned}$$

where $\gamma \geq 0$ is the Lagrange multiplier corresponding to the required minimum capacity constraint C2. $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ are the Lagrange multiplier vectors associated with individual power constraint C1 and maximum backhaul capacity constraint C3 with

elements $\lambda_m \geq 0$ and $\beta_m \geq 0$, $m \in \{1, \ldots, M\}$, respectively. Boundary constraint C6 will be absorbed into the KKT conditions when deriving the optimal power allocation in the following. Note that $(\alpha_k + \gamma - \beta_m) \geq 0$ always holds for allocating a non-zero capacity to user k, cf. (5.19). Thus, the dual problem of (5.12), for a given selected user set and ZFBF transmission, is given by

$$\mathrm{D} = \min_{\boldsymbol{\lambda}, \boldsymbol{\beta}, \gamma \geq 0} \max_{\mathcal{P}} \quad \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\beta}, \gamma, \mathcal{P}). \tag{5.18}$$

In the following, we solve the above dual problem iteratively by decomposing it into two layers via *dual decomposition*: Layer 1, the maximization over $\mathcal{P}$ in (5.18), consists of $M \times n_F$ subproblems with identical structure; Layer 2, the minimization over $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, and γ in (5.18), is the master dual problem to be solved by the gradient method, cf. Figure 5.2.

Layer 1 Solution (Power Allocation)

By Theorem 5.1, the KKT conditions are the necessary and sufficient conditions for the optimal solution. Thus, the closed-form power allocation for the BSs serving user k in subcarrier i for a given parameter q is obtained as

$$P_{B_m}^k(i) = \left[\frac{\mathcal{B}/n_F(\alpha_k + \gamma - \beta_m)}{\ln(2)\Omega^k(i)} - \frac{\sigma_z^2}{|\gamma^k(i)|^2} \right]^+ \tag{5.19}$$

$$\text{and } P_{B_a}^k(i) = P_{B_m}^k(i) \, \forall a \neq m, \text{ where } \Omega^k(i) = \Big(\sum_{c=1}^{M} (\lambda_c + q\varepsilon) |w_{B_c}^k(i)|^2 \Big). \tag{5.20}$$

The optimal power allocation solution in (5.19) is in the form of *multi-level water filling*. Note that if a user has a higher value of α_k (higher priority), a higher power will be allocated to the user since she has a higher water level $\frac{\mathcal{B}/n_F(\alpha_k+\gamma-\beta_m)}{\ln(2)\Omega^k(i)}$ compared

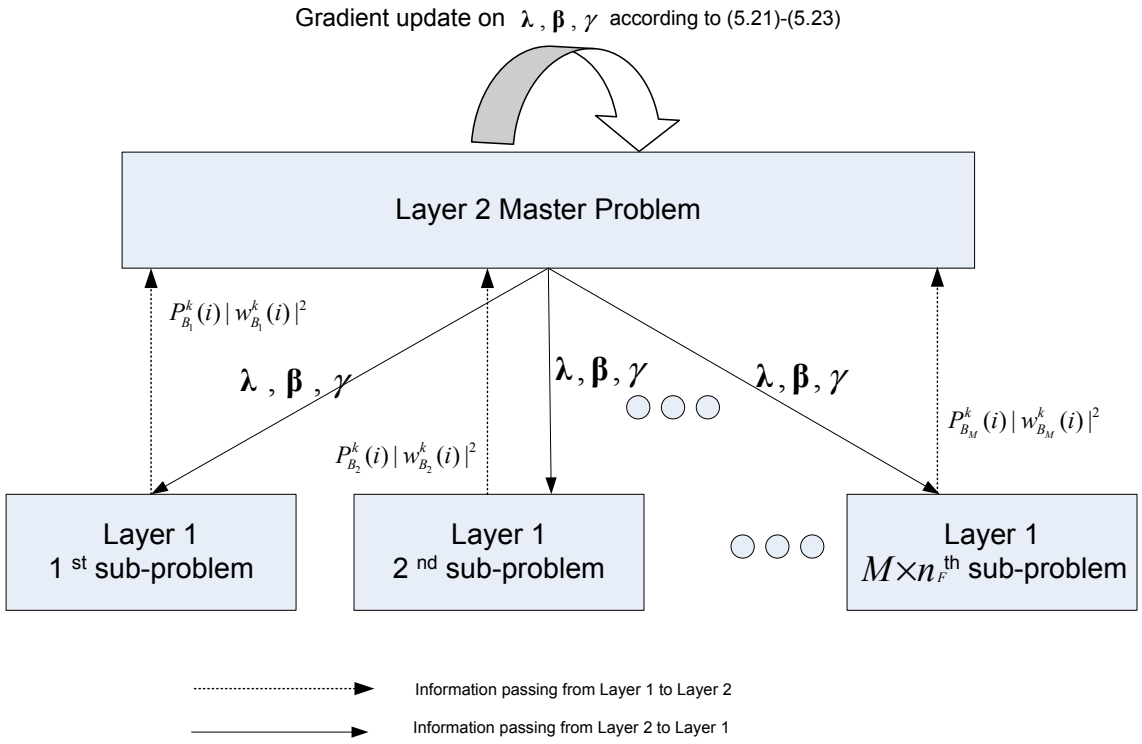


Figure 5.2: Illustration of the dual decomposition of a large problem into a two-layer problem.

to other users. $\beta_m \geq 0$ controls the scheduled data rate via adjusting the water level of the power allocation in (5.19), such that the scheduled data rate will not exceed the backhaul capacity limit. $\Omega^k(i)$ represents the influence of the power consumption of other BSs in the joint transmission on subcarrier i for user k.

Solution of Layer 2 (Master Problem)

To solve the Layer 2 master minimization problem in (5.18), i.e, to find $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, and γ for a given $\mathcal{P}$, the gradient method can be used since the dual function is differentiable.

The gradient update equations are given by:

$$\lambda_m(c+1) = \Big[\lambda_m(c) - \xi_1(c) \times \Big(P_{T_m} - \sum_{i=1}^{n_F} \sum_{k \in \mathcal{S}_\perp(i)}^{K} |w_{B_m}^k(i)|^2 P_{B_m}^k(i)\Big)\Big]^+, \forall m, \quad (5.21)$$

$$\gamma(c+1) = \Big[\gamma(c) - \xi_2(c) \times \Big(\sum_{m=1}^{M} \sum_{i=1}^{n_F} \sum_{k \in \mathcal{A}_m \cap \mathcal{S}_\perp(i)}^{K} C^k(i) - R_{\min}\Big)\Big]^+, \quad (5.22)$$

$$\beta_m(c+1) = \Big[\beta_m(c) - \xi_3(c) \times \Big(R_{\max_m} - \sum_{i=1}^{n_F} \sum_{k \in \mathcal{A}_m \cap \mathcal{S}_\perp(i)}^{K} C^k(i)\Big)\Big]^+, \forall m, \quad (5.23)$$

where index $c \geq 0$ is the iteration index and $\xi_u(c)$, $u \in \{1, 2, 3\}$, are positive step sizes. Then, the updated Lagrange multipliers in (5.21)-(5.23) are used for solving the Layer 1 subproblems in (5.18) via updating the resource allocation policies, cf. Figure 5.2. By Theorem 5.1, the duality gap is zero and it is guaranteed that the iteration between Layer 1 and Layer 2 converges to the optimal solution of (5.18) in the main loop, if the chosen step sizes satisfy the infinite travel condition [75, 88].

5.5 Results and Discussions

In this section, we evaluate the system performance for the proposed resource allocation and scheduling algorithm using simulations. A multi-cell system with 3 cells is considered. The inter-site distance between each pair of BSs is 500 meters as suggested in the 3GPP specification [123]. The number of subcarriers is $n_F = 128$ with carrier center frequency 2.5 GHz, system bandwidth $\mathcal{B} = 1.25$ MHz, and $\alpha_k = 1, \forall k$. Each subcarrier for RF transmission has a bandwidth of 9.7656 kHz and the noise variance is $\sigma_z^2 = -134$ dBm. The 3GPP path loss model is used. The small scale fading coefficients of the BS-to-user links are generated as i.i.d. Rayleigh random variables with zero means and unit variances. We assume that all BSs have the same

maximum transmit power, i.e., $P_{T_m} = P_T$, $\forall m$. Besides, a fully connected backhaul connection topology is considered for simulation purpose, i.e., there are $\delta = 6$ connections, cf. Figure 5.1. For the backhaul connections, we adopt the specifications of a commercial optical fiber modem [124] which supports three types of data rates for backhaul within a distance of 2.5 km: $R_1 = 11.184$ Mbit/s, $R_2 = 34.368$ Mbit/s, and $R_3 = 44.736$ Mbit/s[24]. The maximum power consumption of each backhaul link is $P_{BH} = 15$ Watts. The average system energy efficiency is obtained by counting the amount of data which are successfully decoded by the users and dividing it by the total power consumption averaged over both macroscopic and microscopic fading. We assume a static circuit power consumption of $P_C = 40$ dBm [89], a data rate requirement of $R_{\min} = 4$ bit/s/Hz/cell, and an orthogonality parameter of $\eta = 0.1$. On the other hand, we assume a power efficiency of 20% for the power amplifiers used in the RF, i.e., $\varepsilon = \frac{1}{0.2} = 5$. In the following results, the "number of iterations" refers to the number of iterations of Algorithm 5.1 in Table 5.1.

5.5.1 Convergence of Iterative Algorithm 5.1 and Duality Gap

Figure 5.3 illustrates the evolution of the proposed iterative algorithm for different numbers of users and different maximum transmit powers at each BS. The results in Figure 5.3 were averaged over 100000 independent adaptation processes where each adaptation process involves different realizations for the path loss and the multipath fading. It can be observed that the iterative algorithm converges to the optimal value within 10 iterations for all considered numbers of transmit antennas. In other words,

[24]The values of the backhaul capacities used in the chapter are for illustration purpose. In practice, the choice of backhaul capacities should scale with the bandwidth and the number of subcarriers used in the RF transmission.

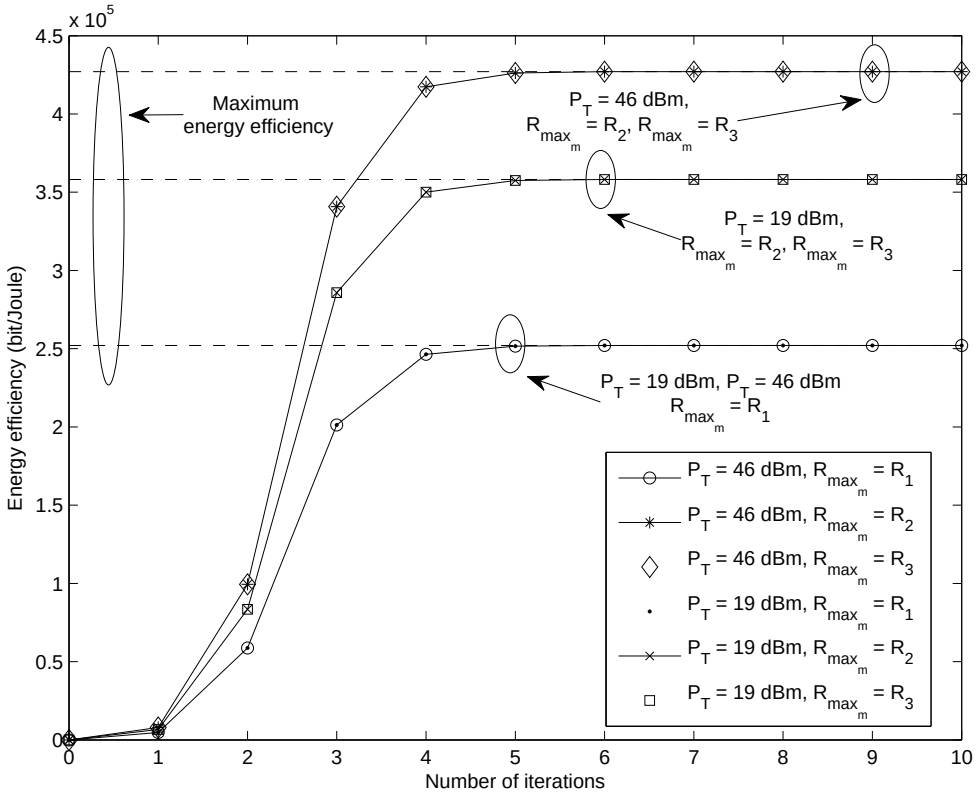


Figure 5.3: Energy efficiency (bit-per-Joule) versus number of iterations with different maximum transmit power allowances per BS, P_T, and different numbers of users K. The dashed lines represent the maximum achievable energy efficiencies for different cases.

the maximum system energy efficiency can be achieved within a few iterations for a given set of selected users and ZFBF transmission.

Figure 5.4 shows the duality gap, D − P, versus the maximum transmit power allowance at each BS, P_T, for different maximum backhaul capacities. The primal problem is solved by using the power allocation solution obtained in the dual decomposition. It can be seen that the duality gap is practically zero for the considered cases, despite the non-convexity of the primal problem. The few small non-zero spikes in the duality gap (in the order of 10^{-7}) are mainly due to a finite-precision arithmetic of computation and a finite number of iterations in solving the dual problem.

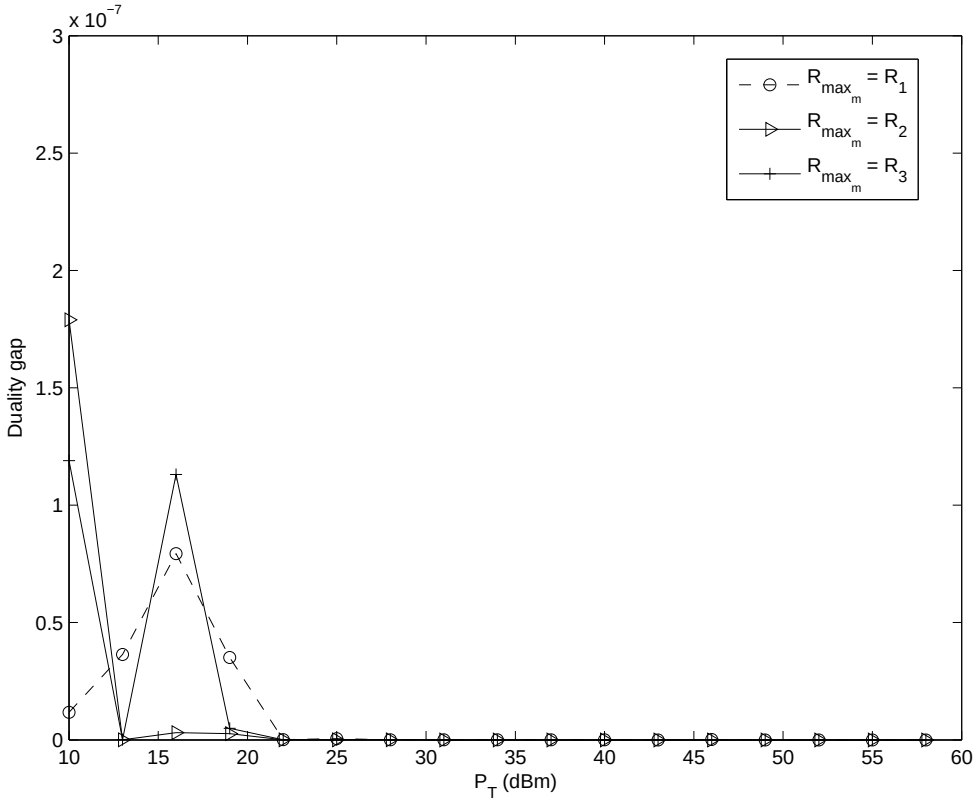


Figure 5.4: Duality gap versus the maximum transmit power allowance at each BS, P_T, for different backhaul capacities $R_{\max_m}$.

5.5.2 Energy Efficiency and Average Capacity versus Transmit Power

Figure 5.5 illustrates the energy efficiency versus the maximum transmit power allowance at each BS, P_T, for $K = 45$ users. The number of iterations for the proposed iterative resource allocation algorithm is 5 and 10. It can be seen that the performance difference between 5 iterations and 10 iterations is negligible which confirms the practicality of our proposed iterative resource allocation algorithm. On the other hand, when both the maximum transmit power allowance at the power amplifier and the capacities of the backhaul links are large enough, e.g., $P_T \geq 30$ dBm and

$R_{\max_m} \geq R_2 \ \forall m$, the energy efficiency of the proposed algorithm approaches a constant value since the resource allocator is not willing to consume more power, when the maximum energy efficiency is achieved. Besides, further increasing the backhaul capacities from $R_{\max_m} = R_2 \ \forall m$ to $R_{\max_m} = R_3 \ \forall m$ is not beneficial for energy efficiency as the system performance is now confined by the capacity of the radio links. However, for the case of backhaul capacity $R_{\max_m} = R_1 \ \forall m$, the energy efficiency is quickly saturated even if the transmit powers at the BSs are low since the system capacity is always limited by the bottleneck of the backhaul connections. For comparison, Figure 5.5 also contains the energy efficiency of a baseline resource allocation scheme in which we maximize the weighted system capacity (bit/s/Hz) with constraints C1-C6 in (5.9) for a given selected users set and ZFBF transmission, instead of the energy efficiency. It can be observed that in the low transmit power regime with high backhaul capacity, i.e., $P_T < 30$ dBm and $R_{\max_m} \geq R_2 \, \forall m$, the baseline scheme has virtually the same performance as the proposed algorithm. In other words, this result suggests that in the low transmit power regime, transmitting with the maximum available power is the most energy efficient option. However, the energy efficiency of the baseline scheme decreases dramatically in the high transmit power regime. This is because there is a diminishing return in the system capacity with respect to the increment of transmit power. Meanwhile, the total power consumption scales linearly with respect to the transmit power. Hence, the capacity gain is unable to compensate for the negative impact of the total power consumption in the RF amplifiers and results in a low energy efficiency. On the other hand, for $R_{\max_m} = R_1 \, \forall m$, the proposed algorithm and the baseline scheme achieve the same energy efficiencies as the degrees of freedom in the resource allocation are limited by the small backhaul capacities.

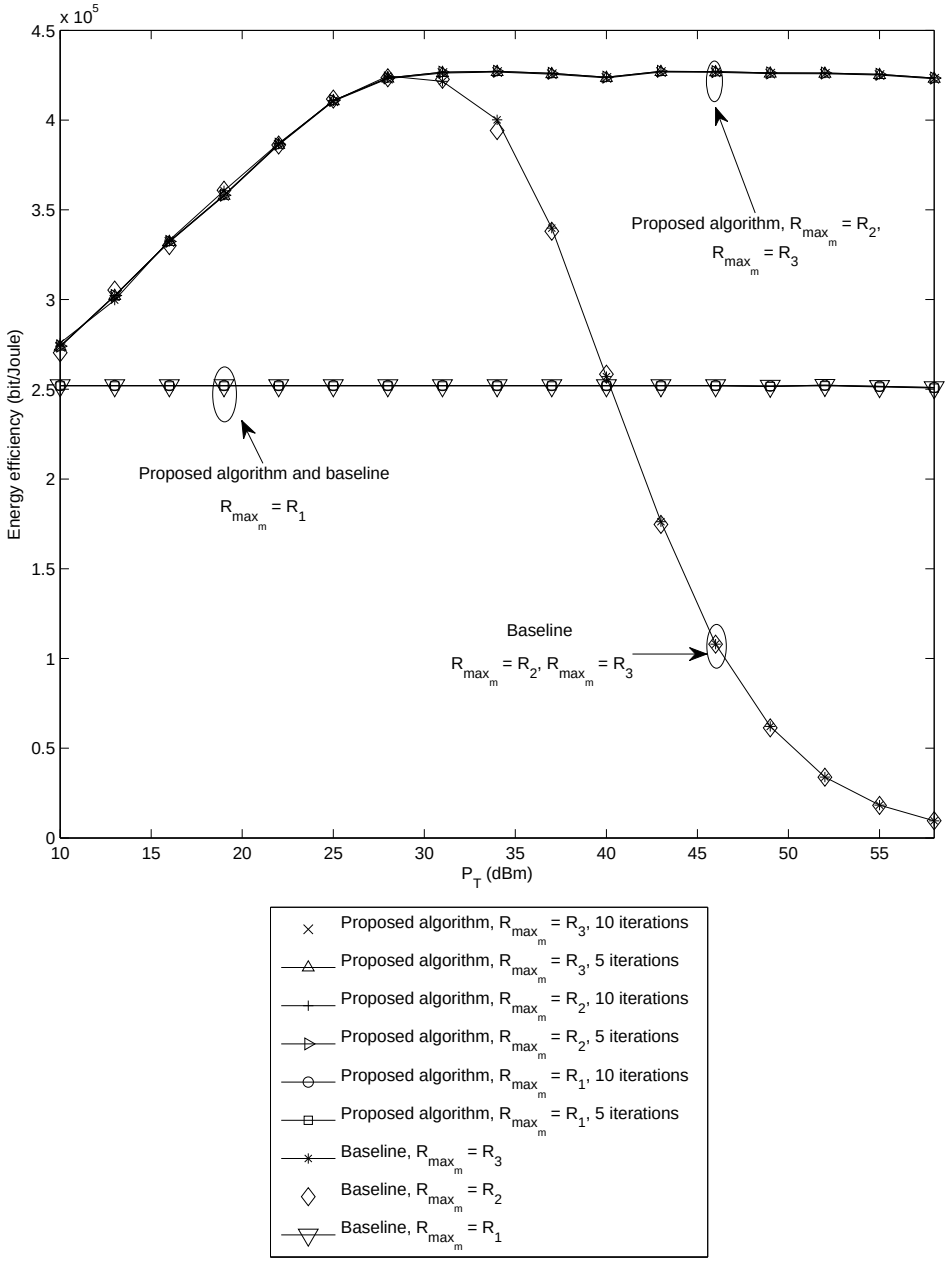


Figure 5.5: Energy efficiency (bit-per-Joule) versus the maximum transmit power allowance at each BS, P_T, for different resource allocation algorithms and different backhaul capacities with $K = 45$ users.

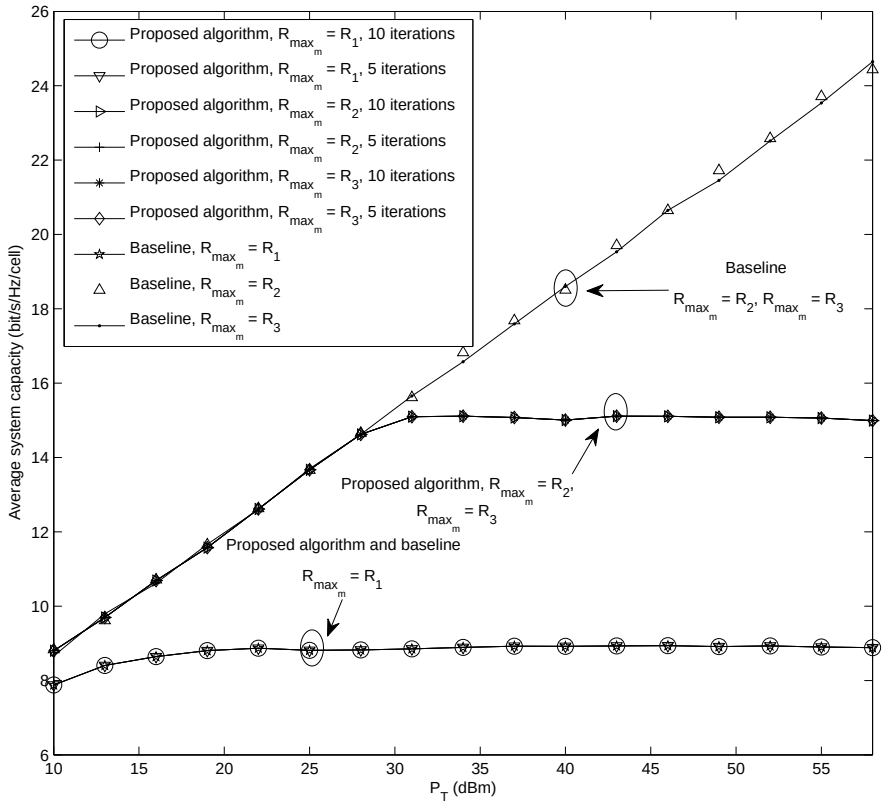


Figure 5.6: Average system capacity (bit/s/Hz/cell) versus the maximum transmit power allowance at each BS, P_T, for different resource allocation algorithms and different backhaul capacities with $K = 45$ users.

Figure 5.6 shows the average system capacity (bit/s/Hz/cell) versus the maximum transmit power P_T for $K = 45$ users and different backhaul capacities. We compare the system performance of the proposed algorithm again with the baseline scheme. The number of iterations in the proposed algorithm is set to 5 and 10. It can be observed that the average system capacity of the proposed algorithm approaches a constant in the high transmit power and high backhaul capacity regimes, i.e, $P_T \geq 30$ dBm and $R_{\max_m} \geq R_2 \ \forall m$. This is because the proposed algorithm clips the transmit power at the BSs in order to maximize the system energy efficiency. However, when

the backhaul capacity is small, i.e., $R_{\max_m} = R_1 \, \forall m$, the maximum achievable average system capacities of both the proposed algorithm and the baseline scheme do not scale with the transmit power. We note that, as expected, the baseline scheme achieves a higher average system capacity than the proposed algorithm in the high transmit power regime for $R_{\max_m} \geq R_2 \ \forall m$ since the former scheme consumes all the available transmit power in all scenarios. However, the superior average system capacity of the baseline scheme comes at the expense of low energy efficiency as shown in Figure 5.5. On the other hand, increasing the backhaul capacity beyond $R_{\max_m} = R_2 \, \forall m$ is not beneficial for the average system capacity as the wireless communication links are the bottleneck links.

Figure 5.7 depicts the average total power consumption, i.e., $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{W}, \mathcal{S})\}$, versus the maximum transmit power P_T for the proposed algorithm and the baseline scheme for 10 iterations. In the considered transmit power regimes, the proposed algorithm consumes less power than the baseline scheme for the case of $R_{\max_m} \geq R_2 \, \forall m$. This is because the proposed algorithm clips the transmit power for energy efficiency maximization. However, when the backhaul capacity is the limiting factor, i.e., $R_{\max_m} = R_1 \, \forall m$, both the baseline scheme and the proposed algorithm consume almost the same amount of power since the transmit power usage is confined by the backhaul capacity instead of energy efficiency maximization.

5.5.3 Energy Efficiency and Average System Capacity versus Number of Users

Figures 5.8 and 5.9 depict the energy efficiency and the average system capacity versus the number of users, respectively. Different backhaul capacities, different maximum transmit power allowances P_T at the BSs, and 10 iterations of the proposed

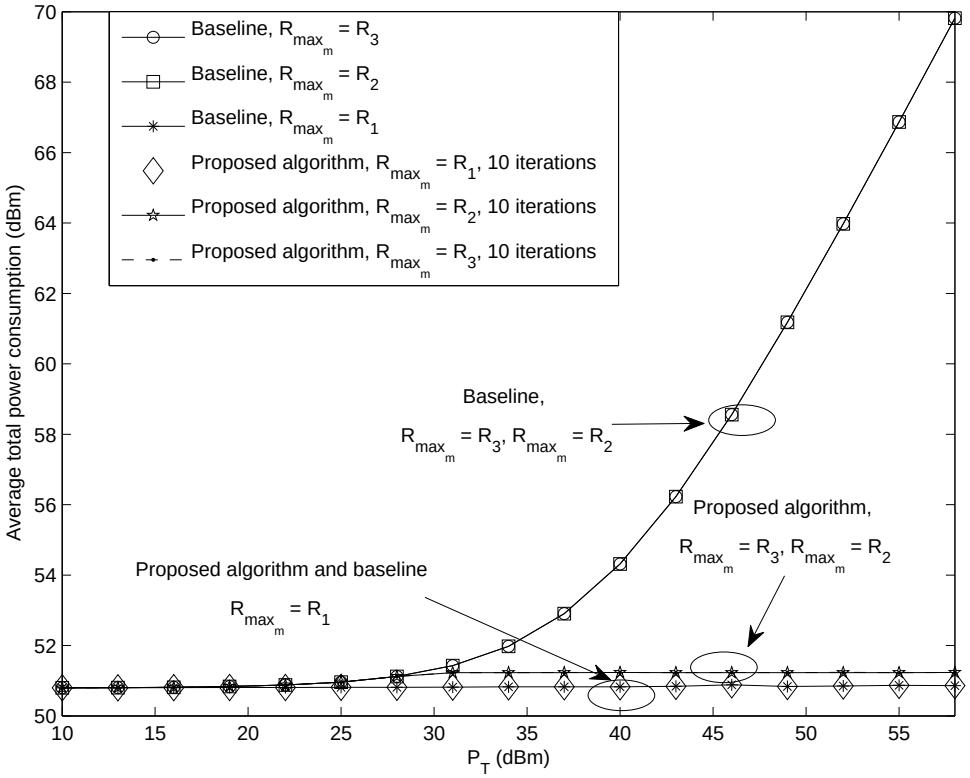


Figure 5.7: Average total power consumption, $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{W}, \mathcal{S})\}$, versus the maximum transmit power allowance at each BS, P_T, for different backhaul capacities, 10 iterations, and $K = 45$ users.

algorithm are considered. It can be observed that for $R_{\max_m} \geq R_2 \, \forall m$, both the energy efficiency and the average system capacity grow with the number of users since the proposed resource allocation and scheduling algorithm is able to exploit multiuser diversity (MUD) due to the semi-orthogonal user selection algorithm. In general, MUD introduces an extra power gain [28, Section 6.6] to the system which provides further energy savings. Yet, when the backhaul capacity is the performance limiting factor, i.e., $R_{\max_m} = R_1 \, \forall m$, the proposed algorithm is unable to take advantage of MUD since the performance gain due to joint BS transmission is limited by the small backhaul capacities. As a result, both the average system capacity and the energy

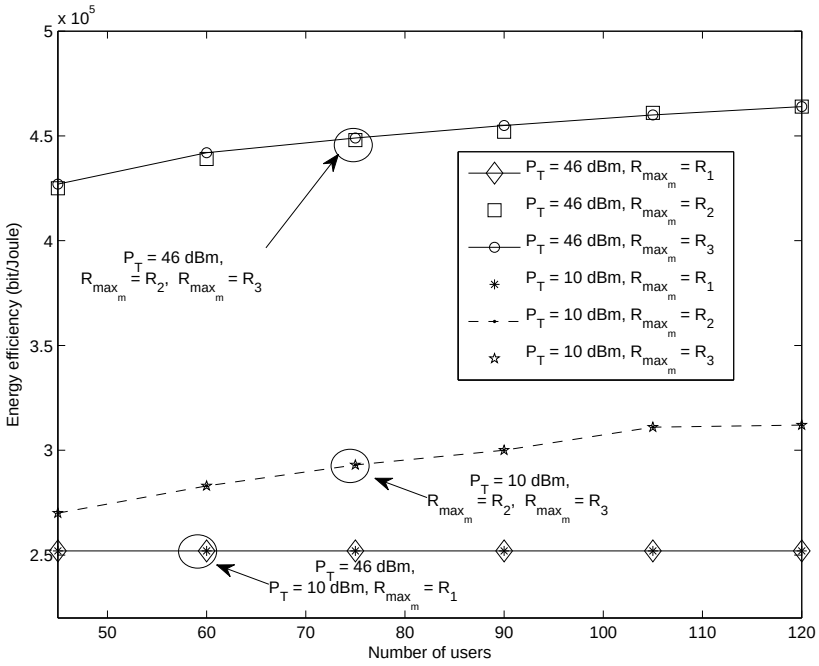


Figure 5.8: Energy efficiency (bit-per-Joule) versus the number of users K for different maximum transmit power allowances at each BS, P_T, and different backhaul capacities.

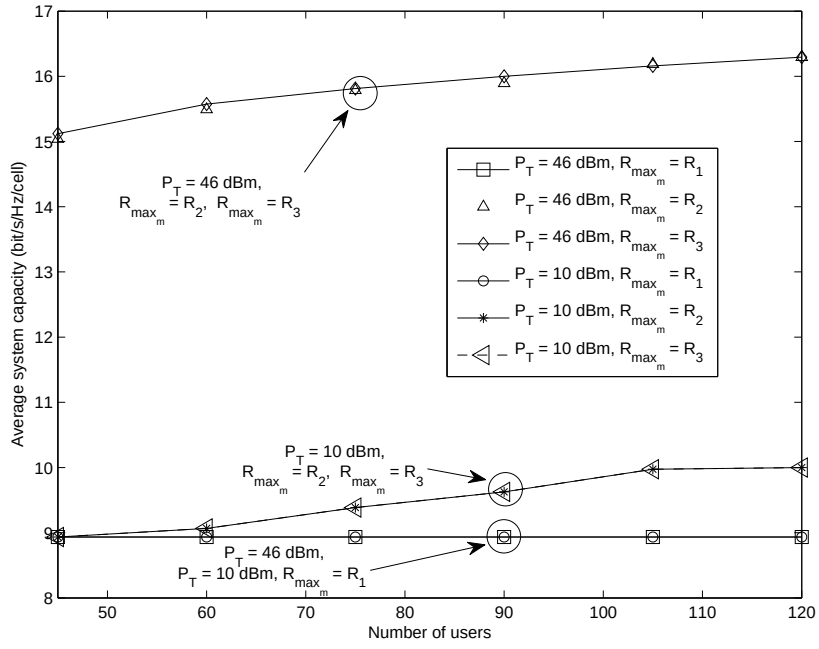


Figure 5.9: Average system capacity (bit/s/Hz/cell) versus the number of users K for different maximum transmit power allowances at each BS, P_T, and different backhaul capacities.

efficiency remain constant when the backhaul is the bottleneck.

5.6 Conclusions

In this chapter, we formulated the resource allocation and scheduling design for multi-cell OFDMA networks with joint BS ZFBF transmission as a non-convex and combinatorial optimization problem, in which the circuit power dissipation, the limited backhaul capacity, and the system data rate requirement were taken into consideration. By exploiting the properties of fractional programming, the considered problem was transformed into an equivalent problem with a tractable iterative solution. In each iteration, a low complexity user selection is performed for maximization of the energy efficiency. Furthermore, we demonstrated that when the number of subcarriers is sufficiently large, the duality gap is practically zero despite the non-convexity of the primal problem. As a result, an efficient closed-form power allocation can be obtained in each iteration via dual decomposition. Simulation results showed that the proposed algorithm converges to the solution within a small number of iterations and unveiled a trade-off between energy efficiency, network capacity, and backhaul capacity.

Chapter 6

Summary of Thesis and Future Research Topics

In this final chapter, in Section 6.1, we summarize our results and highlight the contributions of this thesis. In Section 6.2, we also propose ideas for future related research.

6.1 Summary of Results

This thesis as a whole has focused on resource allocation algorithm design for next generation wireless communication systems, namely: (1) secure OFDMA DF relaying systems; (2) energy efficient secure OFDMA systems; (3) energy efficient OFDMA systems with a large number of BS antennas; (4) energy efficient multi-cell OFDMA systems with BS cooperation and limited backhaul capacity. In the following, we briefly review the main results of each chapter.

In Chapter 2, we proposed a novel resource allocation algorithm for secure OFDMA DF relaying systems. Distributed resource allocation policies, i.e., power, data rate, and subcarrier allocation, were optimized for maximization of the average system secrecy outage capacity. The proposed algorithm takes into account of the artificial noise generation for combating an eavesdropper with multiple antennas. Simulation results showed that a non-zero secrecy outage capacity is achievable, independent of

the location of the multiple antennas eavesdropper.

In Chapter 3, we investigated the energy efficiency in secure OFDMA system by taking the dynamic circuit power consumption into account. By exploiting the fractional programming, a non-convex optimization problem in fractional form was transformed into subtractive form which facilitated the design of an efficient resource allocation for secrecy capacity maximization. Simulation results demonstrated that the energy efficiency approaches a constant in the high transmit power regime which is related to the dynamic power consumption. On the other hand, a non-zero energy efficiency and average secrecy outage capacity can still be achieved as long as the number of antennas at the BS is larger than the number of eavesdropper antennas, despite the fact that the eavesdropper is closer to the BS than the desired users.

In Chapter 4, we have considered the energy efficiency in OFDMA systems with a large number of antennas. We formulated the resource allocation design as an optimization problem which took into account the power consumption of each antenna and provided a high flexibility in resource allocation. In particular, we allowed the BS to only activate a portion of the available antennas for energy saving. Simulation results showed that increasing the number of activated antennas at the BSs is always beneficial to the average system outage capacity. Yet, this is not necessary the case from an energy efficiency point of view. Under the adapted simulation parameters, we showed that an extra number of antennas is activated only whenever we were unable in fulfilling the data rate requirement by power allocation.

In Chapter 5, we have investigated the energy efficiency in multi-cell OFDMA systems with BS cooperation. The resource allocation design was formulated as a non-convex optimization problem which took into account limited backhaul capacity, a minimum system data rate requirement, and individual BS power constraints. By

exploiting the properties of fractional programming, the original problem in fractional form was transformed into a subtractive form which facilitates the design of an iterative efficient design. In each iteration, we solved a non-convex power allocation problem. In particular, we showed that the duality gap of the resulting power allocation problem is zero when the number of subcarrier is sufficiently large, despite the non-convexity of the primal problem. As a result, a novel resource allocation algorithm was derived by using dual decomposition.

6.2 Future Work

Future wireless communication networks will have to strive for higher data rates and more reliable communication to serve multiusers, and at the same time, guarantee other diverse QoS requirements. This brings about several technical problems such as a higher multi-cell interference level as well as heavy energy consumption per user. The above issues have raised serious concerns on whether existing technologies are able to cope with the challenges introduced by future applications. In Chapters 2–5, we have considered four novel resource allocation algorithms for energy efficient and secure communication systems. However, energy efficient and secure system designs are a vast research area and many problems are still unsolved. In the following, we propose some ideas for further research that are similar to or can be based on the work in this thesis.

6.2.1 Secure Communications in OFDMA Systems with an Active Eavesdropper

One extension of Chapters 2 and 3 is to focus on an active eavesdropper under a half duplex constraint [125, 126, 127]. In contrast to a passive eavesdropper, an active eavesdropper can choose between eavesdropping the transmission between the legitimate parties and jamming the communication channels. As a result, the action of the transmitter depends on the behavior of the eavesdropper. If the active eavesdropper chooses to eavesdrop the channel, the transmitter should generate artificial noise along with the information signal to selectively interfere with the eavesdropper, as suggested in Chapters 2 and 3. However, if the eavesdropper decides to jam the communication channels, the transmitter should allocate all the power to the information bearing signal in order to maximize the SINR. On the other hand, the action of the eavesdropper depends on the action of the transmitter. The action dependencies between the transmitter and eavesdropper results in an interesting trade-off in the system performance as well as the resource allocation algorithm design. A possible direction is to study the resource allocation and scheduling problem for secure communications in OFDMA systems with active eavesdroppers. In particular, the problem may be modelled as a two-person zero-sum game with the secrecy capacity as the objective function. The results of this work will provide useful insights for designing resource allocation algorithms for secure communication in mission critical and military networks.

6.2.2 Energy Efficiency: Optimal Locations for Distributed Antennas

Next generation mobile communication systems are required to provide reliable communication and heterogeneous services for a large number of users in the system. Unfortunately, the users located at the cell edges are suffering from strong multi-cell interference and weak desired signal strengths which becomes the bottleneck of the system performance. In order to enhance the system performance of the cell edge users, distributed antenna technologies have been proposed and provide a substantial gain in network downlink throughput by jointly exploiting the degrees of freedom in frequency, time, and space [128]-[133]. Although some preliminary studies have been already conducted for the power allocation and frequency assignment in distributed antenna systems [131, 132, 133], the optimization of the location of the antennas has been overlooked so far, especially in multi-cell interference limited environments. An interesting problem for future work is the resource allocation and antenna location optimization for maximization of the average system capacity or energy efficiency in multi-cell OFDMA systems, under the constraints of maximum transmit power and minimum data rate. The results of this work are useful to service providers for network planning.

6.2.3 Energy Harvesting in Energy Efficient OFDM Systems

Energy harvesting is the process of collecting energy from external sources such as solar energy and thermal energy for charging the battery of the energy source so as to extend the network's lifetime [134]-[137]. In contrast to traditional resource allocation problems where the energy supplied by the energy source is static, the time varying nature of the recharging process presents new challenges in the design of

resource allocation algorithms. In particular, the transmitters must always maintain positive energy storage levels to avoid service outages due to energy depletion. Several contributions in the literature have considered using energy harvesters as the energy source [134]-[137]. In practice, the fluctuations in the battery recharging process can happen at a faster rate than the coherence time of the communication channel. In other words, there are two dynamic processes in the system which should be taken into account for the design of resource allocation algorithms. As a result, the design of resource allocation and scheduling algorithms for maximizing the energy efficiency in energy harvesting systems with consideration of the dynamic differences between the states of the energy buffer and data buffer is needed.

Bibliography

[1] E. Hossain, D. I. Kim, and V. K. Bhargava, *Cooperative Cellular Wireless Networks*, 1st ed. Cambridge University Press, 2011.

[2] V. Lau, W. K. Ng, and D. Hui, "Asymptotic Tradeoff between Cross-Layer Goodput Gain and Outage Diversity in OFDMA Systems with Slow Fading and Delayed CSIT," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2732–2739, Jul. 2008.

[3] D. Lopez-Perez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA Femtocells: A Roadmap on Interference Avoidance," *IEEE Commun. Mag.*, vol. 47, pp. 41–48, Sep. 2009.

[4] T. Jiang, W. Xiang, H.-H. Chen, and Q. Ni, "Multicast Broadcast Services Support in OFDMA-Based WiMAX Systems [Advances in Mobile Multimedia]," *IEEE Commun. Mag.*, vol. 45, pp. 78–86, Aug. 2007.

[5] Q. Li, G. Li, W. Lee, M. il Lee, D. Mazzarese, B. Clerckx, and Z. Li, "MIMO Techniques in WiMAX and LTE: A Feature Overview," *IEEE Commun. Mag.*, vol. 48, pp. 86–92, May 2010.

[6] C.-X. Wang, X. Hong, X. Ge, X. Cheng, G. Zhang, and J. Thompson, "Cooperative MIMO Channel Models: A Survey," *IEEE Commun. Mag.*, vol. 48, pp. 80–87, Feb. 2010.

[7] A. Maltsev, A. Khoryaev, A. Lomayev, R. Maslennikov, C. Antonopoulos, K. Avgeropoulos, A. Alexiou, F. Boccardi, Y. Hou, and K. Leung, "MIMO and Multihop Cross-Layer Design for Wireless Backhaul: A Testbed Implementation," *IEEE Commun. Mag.*, vol. 48, pp. 172–179, Mar. 2010.

[8] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated Multipoint: Concepts, Performance, and Field Trial Results," *IEEE Commun. Mag.*, vol. 49, pp. 102–111, Feb. 2011.

[9] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference Coordination and Cancellation for 4G Networks," *IEEE Commun. Mag.*, vol. 47, pp. 74–81, Apr. 2009.

[10] S. Ramprashad, H. Papadopoulos, A. Benjebbour, Y. Kishiyama, N. Jindal, and G. Caire, "Cooperative Cellular Networks using Multi-user MIMO: Trade-offs, Overheads, and Interference Control Across Architectures," *IEEE Commun. Mag.*, vol. 49, pp. 70–77, May 2011.

[11] C. Mohanram and S. Bhashyam, "Joint Subcarrier and Power Allocation in Channel-Aware Queue-Aware Scheduling for Multiuser OFDM," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 3208–3213, Sep. 2007.

[12] Z. Shen, J. Andrews, and B. Evans, "Adaptive Resource Allocation in Multiuser OFDM Systems with Proportional Rate Constraints," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2726–2737, Nov. 2005.

[13] G. Miao and N. Himayat, "Low Complexity Utility Based Resource Allocation for 802.16 OFDMA Systems," in *Proc. IEEE Wireless Commun. and Netw. Conf.*, Apr. 2008, pp. 1465–1470.

[14] J. Shi and A. Hu, "Maximum Utility-Based Resource Allocation Algorithm in the IEEE 802.16 OFDMA System," in *Proc. IEEE Intern. Commun. Conf.*, May 2008, pp. 311–316.

[15] Cho Yiu Ng and Chi Wan Sung, "Low complexity Subcarrier and Power Allocation for Utility Maximization in Uplink OFDMA Systems," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1667–1675, May 2008.

[16] G. Song and Y. Li, "Cross-Layer Optimization for OFDM Wireless Networks-Part I: Theoretical Framework," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 614–624, Mar. 2005.

[17] M. Tao, Y.-C. Liang, and F. Zhang, "Resource Allocation for Delay Differentiated Traffic in Multiuser OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2190–2201, Jun. 2008.

[18] G. Song and Y. Li, "Cross-layer Optimization for OFDM Wireless Networks-Part II: Algorithm Development," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 625–634, Mar. 2005.

[19] I. Wong and B. Evans, "Optimal Downlink OFDMA Resource Allocation with Linear Complexity to Maximize Ergodic Rates," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 962–971, Mar. 2008.

[20] C. Guthy, W. Utschick, R. Hunger, and M. Joham, "Efficient Weighted Sum Rate Maximization with Linear Precoding," *IEEE Trans. Signal Process.*, vol. 58, pp. 2284–2297, Apr. 2010.

[21] I. G. Fraimis and S. A. Kotsopoulos, "QoS-Based Proportional Fair Allocation Algorithm for OFDMA Wireless Cellular Systems," *IEEE Commun. Lett.*, vol. 15, pp. 1091–1093, Oct. 2011.

[22] H. Kwon and B. G. Lee, "Cooperative Power Allocation for Broadcast/Multicast Services in Cellular OFDM Systems," *IEEE Trans. Commun.*, vol. 57, pp. 3092–3102, Oct. 2009.

[23] M. S. Maw and I. Sasase, "Resource Allocation Scheme in MIMO-OFDMA System for User's Different Data Throughput Requirements," in *Proc. IEEE Wireless Commun. and Netw. Conf.*, Mar. 2007, pp. 1706–1710.

[24] G. Caire, G. Taricco, and E. Biglieri, "Optimum Power Control Over Fading Channels," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1468–1489, Jul. 1999.

[25] R. Narasimhan, "Finite-SNR Diversity-Multiplexing Tradeoff for Correlated Rayleigh and Rician MIMO Channels," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3965–3979, Sep. 2006.

[26] D. Kivanc, G. Li, and H. Liu, "Computationally Efficient Bandwidth Allocation and Power Control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1150–1158, Nov. 2003.

[27] K. Kim, Y. Han, and S.-L. Kim, "Joint Subcarrier and Power Allocation in Uplink OFDMA Systems," *IEEE Commun. Lett.*, vol. 9, pp. 526–528, Jun. 2005.

[28] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, 1st ed. Cambridge University Pres, 2005.

[29] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[30] V. K. N. Lau and Y. K. Kwok, *Channel Adaptation Technologies and Cross Layer Design for Wireless Systems with Multiple Antennas - Theory and Applications*, 1st ed. Wiley John Proakis Telecom Series, 2005.

[31] Z. Han, T. Himsoon, W. Siriwongpairat, and K. Liu, "Resource Allocation for Multiuser Cooperative OFDM Networks: Who Helps Whom and How to Cooperate," *IEEE Trans. Veh. Technol.*, vol. 58, pp. 2378–2391, Jun. 2009.

[32] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity. Part I. System Description," *IEEE Trans. Commun.*, vol. 51, pp. 1927–1938, Nov. 2003.

[33] L. Liu, P. Parag, and J.-F. Chamberland, "Quality of Service Analysis for Wireless User-Cooperation Networks," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3833–3842, Oct. 2007.

[34] W. Yu, T. Kwon, and C. Shin, "Multicell Coordination via Joint Scheduling, Beamforming and Power Spectrum Adaptation," in *Proc. IEEE Conf. on Comput. Commun.*, Apr. 2011, pp. 2570–2578.

[35] A. Tolli, H. Pennanen, and P. Komulainen, "Decentralized Minimum Power Multi-Cell Beamforming with Limited Backhaul Signaling," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 570–580, Feb. 2011.

[36] H. Dahrouj and W. Yu, "Coordinated Beamforming for the Multicell Multi-Antenna Wireless System," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1748–1759, May 2010.

[37] Y. Hua, Q. Zhang, and Z. Niu, "Resource Allocation in Multi-cell OFDMA-based Relay Networks," in *Proc. IEEE Conf. on Computer Commun.*, Mar. 2010, pp. 1–9.

[38] D. W. K. Ng and R. Schober, "Cross-Layer Scheduling for OFDMA Amplify-and-Forward Relay Networks," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 1443–1458, Mar. 2010.

[39] Y. Cui, V. Lau, and R. Wang, "Distributive Subband Allocation, Power and Rate Control for Relay-Assisted OFDMA Cellular System with Imperfect System State Knowledge," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 5096–5102, Oct. 2009.

[40] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[41] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative Strategies and Capacity Theorems for Relay Networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3037–3063, Apr. 2005.

[42] I. Hammerstrom and A. Wittneben, "Power Allocation Schemes for Amplify-and-Forward MIMO-OFDM Relay Links," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2798–2802, Aug. 2007.

[43] S. W. Peters and R. W. Heath, "The Future of WiMAX: Multihop Relaying with IEEE 802.16j," *IEEE Commun. Mag.*, vol. 47, pp. 104–111, Jan. 2009.

[44] G. P. Fettweis and E. Zimmermann, "ICT Energy Consumption - Trends and Challenges," in *Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Process.*, Sep. 2008.

[45] J. Zeng and H. Minn, "Green Cellular Networks: A Survey, Some Research Issues and Challenges," *IEEE Communications Surveys and Tutorials*, Jan. 2012.

[46] "Energy Aware Radio and neTwork tecHnologies (EARTH)." [Online]. Available: https://www.ict-earth.eu/default.html

[47] C. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, pp. 10–21, Jan. 1949.

[48] R. G. Gallager, *Information Theory and Reliable Communication*, 1st ed. Wiley, 1968.

[49] S. Verdu, "Spectral Efficiency in the Wideband Regime," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1319–1343, Jun. 2002.

[50] S. Jayaweera, "Virtual MIMO-Based Cooperative Communication for Energy-constrained Wireless Sensor Networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 984–989, May 2006.

[51] A. Wang and C. Sodini, "On the Energy Efficiency of Wireless Transceivers," in *Proc. IEEE Intern. Commun. Conf.*, vol. 8, Jun. 2006, pp. 3783–3788.

[52] F. Costa and H. Ochiai, "Energy-Efficient Physical Layer Design for Wireless Sensor Network Links," in *Proc. IEEE Intern. Commun. Conf.*, Jun. 2011, pp. 1–5.

[53] D. Welch and S. Lathrop, "Wireless Security Threat Taxonomy," in *Proc. IEEE Syst., Man and Cybern. Society Inf. Assurance Workshop*, Jun. 2003, pp. 76–83.

[54] S. Miller, "Facing the Challenge of Wireless Security," *Comput.*, vol. 34, pp. 16–18, Jul. 2001.

[55] Y.-S. Shiu, S. Y. Chang, H.-C. Wu, S.-H. Huang, and H.-H. Chen, "Physical Layer Security in Wireless Networks: A Tutorial," *IEEE Wireless Commun.*, vol. 18, pp. 66–74, Apr. 2011.

[56] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd ed. Wiley-IEEE Press, 2004.

[57] J. G. Proakis, *Digital Communications*, 4th ed. McGraw Hill Higher Education, 2000.

[58] X. Wang, M. Tao, J. Mo, and Y. Xu, "Power and Subcarrier Allocation for Physical-Layer Security in OFDMA-Based Broadband Wireless Networks," *IEEE Trans. Inf. Forensics Security*, vol. 6, pp. 693–702, Sep. 2011.

[59] D. W. K. Ng, E. S. Lo, and R. Schober, "Secure Resource Allocation and Scheduling for OFDMA Decode-and-Forward Relay Networks," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 3528–3540, Oct. 2011.

[60] L. Dong, Z. Han, A. Petropulu, and H. Poor, "Improving Wireless Physical Layer Security via Cooperating Relays," *IEEE Trans. Signal Process.*, vol. 58, pp. 1875–1888, Mar. 2010.

[61] M. Jorgensen, B. Yanakiev, G. Kirkelund, P. Popovski, H. Yomo, and T. Larsen, "Shout to Secure: Physical-Layer Wireless Security with Known Interference," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2007, pp. 33–38.

[62] A. D. Wyner, "The Wire-Tap Channel," Tech. Rep., Oct. 1975.

[63] E. A. Jorswieck and A. Wolf, "Resource Allocation for the Wire-Tap Multi-Carrier Broadcast Channel," in *Proc. International Conf. on Telecommun.*, Jun. 2008, pp. 1–6.

[64] Z. Li, R. Yates, and W. Trappe, "Secrecy Capacity of Independent Parallel Channels," in *Proc. 44th Annu. Allerton Conf. Commun., Control and Computing*, Sep. 2006, pp. 841–848.

[65] X. Wang, M. Tao, J. Mo, and Y. Xu, "Power and Subcarrier Allocation for Physical-Layer Security in OFDMA Networks," in *Proc. IEEE Intern. Commun. Conf.*, Jun. 2011, pp. 1–5.

[66] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Amplify-and-Forward Based Cooperation for Secure Wireless Communications," in *Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Process.*, Apr. 2009, pp. 2613–2616.

[67] L. Jiangyuan, A. P. Petropulu, and S. Weber, "Secrecy Rate Optimization under Cooperation with Perfect Channel State Information," in *Proc. the Forty-Third Asilomar Conf. on Signals, Syst. and Comput.*, Nov. 2009, pp. 824–828.

[68] S. Goel and R. Negi, "Guaranteeing Secrecy using Artificial Noise," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2180–2189, Jun. 2008.

[69] X. Zhou and M. R. McKay, "Secure Transmission with Artificial Noise over Fading Channels: Achievable Rate and Optimal Power Allocation," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 3831–3842, Jul. 2010.

[70] R. Liu and W. Trappe, *Securing Wireless Communications at the Physical Layer*, 1st ed. Springer, 2009.

[71] D. W. K. Ng and R. Schober, "Resource Allocation for Secure OFDMA Communication Systems," in *Proc. 2011 Australia Commun. Theory Workshop*, Feb. 2011, pp. 13–18.

[72] ——, "Resource Allocation for Secure OFDMA Decode-and-Forward Relaying Network," in *Proc. 2011 Canadian Workshop on Inf. Theory*, May 2011, pp. 202–205.

[73] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[74] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1747–1758, Oct. 1999.

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

[76] W. Yu and J. M. Cioffi, "FDMA Capacity of Gaussian Multiple-Access Channels with ISI," *IEEE Trans. Commun.*, vol. 50, pp. 102–111, Jan. 2002.

[77] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, 1999.

[78] "Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations," 3GPP TR 25.996 V7.0.0 (2007-06), Tech. Rep.

[79] A. Akbari, R. Hoshyar, and R. Tafazolli, "Energy-Efficient Resource Allocation in Wireless OFDMA Systems," in *Proc. IEEE Personal, Indoor and Mobile Radio Commun. Sympos.*, Sep. 2010, pp. 1731–1735.

[80] G. Miao, N. Himayat, and G. Li, "Energy-Efficient Link Adaptation in Frequency-Selective Channels," *IEEE Trans. Commun.*, vol. 58, pp. 545–554, Feb. 2010.

[81] C. Isheden and G. P. Fettweis, "Energy-Efficient Multi-Carrier Link Adaptation with Sum Rate-Dependent Circuit Power," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2010, pp. 1–6.

[82] ——, "Energy-Efficient Link Adaptation with Transmitter CSI," in *Proc. IEEE Wireless Commun. and Netw. Conf.*, Apr. 2011, pp. 1–6.

[83] Z. Hasan, G. Bansal, E. Hossain, and V. Bhargava, "Energy-Efficient Power Allocation in OFDM-Based Cognitive Radio Systems: A Risk-Return Model," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 6078–6088, Dec. 2009.

[84] W. Dinkelbach, "On Nonlinear Fractional Programming," *Management Science*, vol. 13, pp. 492–498, Mar. 1967. [Online]. Available: http://www.jstor.org/stable/2627691

[85] S. Schaible, "Fractional Programming. II, On Dinkelbach's Algorithm," *Management Science*, vol. 22, pp. 868–873, 1976. [Online]. Available: http://www.jstor.org/stable/2630018

[86] W.-L. Li, Y. J. Zhang, A.-C. So, and M. Win, "Slow Adaptive OFDMA Systems Through Chance Constrained Programming," *IEEE Trans. Signal Process.*, vol. 58, pp. 3858–3869, Jul. 2010.

[87] W. Yu and R. Lui, "Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems," *IEEE Trans. Commun.*, vol. 54, pp. 1310–1321, Jul. 2006.

[88] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient Methods," *Notes for EE392o Stanford University Autumn*, 2003-2004.

[89] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power Consumption Modeling of Different Base Station Types in Heterogeneous Cellular Networks," in *Proc. Future Network and Mobile Summit*, Jun. 2010, pp. 1–8.

[90] D. W. K. Ng and R. Schober, "Resource Allocation for Secure OFDMA Networks with Imperfect CSIT," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2011, pp. 3947–3952.

[91] D. Aktas, M. Bacha, J. Evans, and S. Hanly, "Scaling Results on the Sum Capacity of Cellular Networks with MIMO Links," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3264–3274, Jul. 2006.

[92] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590–3600, Nov. 2010.

[93] T. Maciel and A. Klein, "On the Performance, Complexity, and Fairness of Suboptimal Resource Allocation for Multiuser MIMO-OFDMA Systems," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 406–419, Jan. 2010.

[94] V. Papoutsis, I. Fraimis, and S. Kotsopoulos, "User Selection and Resource Allocation Algorithm with Fairness in MISO-OFDMA," *IEEE Commun. Lett.*, vol. 14, pp. 411–413, May 2010.

[95] E. Lo, P. Chan, V. Lau, R. Cheng, K. Letaief, R. Murch, and W. Mow, "Adaptive Resource Allocation and Capacity Comparison of Downlink Multiuser MIMO-MC-CDMA and MIMO-OFDMA," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1083–1093, Mar. 2007.

[96] R. Prabhu and B. Daneshrad, "Energy-Efficient Power Loading for a MIMO-SVD System and Its Performance in Flat Fading," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2010, pp. 1–5.

[97] "Report of the Spectrum Efficiency Working," FCC Spectrum Policy Task Force, Tech. Rep., Nov. 2002, [Online] http://www.fcc.gov/sptf/reports.html.

[98] R. Kumar and J. Gurugubelli, "How Green the LTE Technology Can be?" in *Proc. Intern. Conf. on Wireless Commun., Veh. Techn., Inform. Theory and Aerosp. Electron. Syst. Techn.*, Mar. 2011, pp. 1–5.

[99] J. Wang, Y. Zhao, and T. Korhonen, "Cross Layer Optimization with Complete Fairness Constraints in OFDMA Relay Networks," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2008, pp. 1–5.

[100] P. Liu, Z. Tao, Z. Lin, E. Erkip, and S. Panwar, "Cooperative Wireless Communications: A Cross-Layer Approach," *IEEE Wireless Commun.*, vol. 13, pp. 84–92, Aug. 2006.

[101] Y. Yuan, Z. He, and M. Chen, "Virtual MIMO-based Cross-Layer Design for Wireless Sensor Networks," *IEEE Trans. Veh. Technol.*, vol. 55, pp. 856–864, May 2006.

[102] S.-J. Kim, X. Wang, and M. Madihian, "Optimal Resource Allocation in Multi-Hop OFDMA Wireless Networks with Cooperative Relay," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1833–1838, May 2008.

[103] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative Communication in Wireless Networks," *IEEE Commun. Mag.*, vol. 42, pp. 74–80, Oct. 2004.

[104] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. Falconer, and G. P. Fettweis, "Relay-Based Deployment Concepts for Wireless and Mobile Broadband Radio," *IEEE Commun. Mag.*, vol. 42, pp. 80–89, Sep. 2004.

[105] "Downlink Inter-Cell Interference Co-ordination/Avoidance - Evaluation of Frequency Reuse," 3GPP TSG-RAN WG1 Contribution R1-061374, Tech. Rep., 2006.

[106] J. G. Andrews, "Interference Cancellation for Cellular Systems: A Contemporary Overview," *IEEE Wireless Commun.*, vol. 12, pp. 19–29, Apr. 2005.

[107] J. G. Andrews, W. Choi, and R. W. Heath, "Overcoming Interference in Spatial Multiplexing MIMO Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 95–104, Dec. 2007.

[108] D. Ng and V. Lau, "Power Control and Performance Analysis of Outage-Limited Cellular Network with MUD-SIC and Macro-Diversity," *IEEE Trans. Commun.*, vol. 58, pp. 2734–2740, Sep. 2010.

[109] O. Somekh, O. Simeone, Y. Bar-Ness, A. Haimovich, and S. Shamai, "Cooperative Multicell Zero-Forcing Beamforming in Cellular Downlink Channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3206–3219, Jul. 2009.

[110] R. Zhang, "Cooperative Multi-Cell Block Diagonalization with Per-Base-Station Power Constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, pp. 1435–1445, Dec. 2010.

[111] G. Dartmann, W. Afzal, X. Gong, and G. Ascheid, "Joint Optimization of Beamforming, User Scheduling, and Multiple Base Station Assignment in a Multicell Network," in *Proc. IEEE Wireless Commun. and Netw. Conf.*, Mar. 2011, pp. 209–214.

[112] D. W. K. Ng and R. Schober, "Resource Allocation and Scheduling in Multi-Cell OFDMA Systems with Decode-and-Forward Relaying," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2246–2258, Jul. 2011.

[113] L. Venturino, N. Prasad, and X. Wang, "Coordinated Scheduling and Power Allocation in Downlink Multicell OFDMA Networks," *IEEE Trans. Veh. Technol.*, vol. 58, pp. 2835–2848, Jul. 2009.

[114] H. Zhang, L. Venturino, N. Prasad, P. Li, S. Rangarajan, and X. Wang, "Weighted Sum-Rate Maximization in Multi-Cell Networks via Coordinated Scheduling and Discrete Power Control," *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 1214–1224, Jun. 2011.

[115] A. Chowdhery, W. Yu, and J. M. Cioffi, "Cooperative Wireless Multicell OFDMA Network with Backhaul Capacity Constraints," in *Proc. IEEE Intern. Commun. Conf.*, Jun. 2011, pp. 1–6.

[116] H. Galeana-Zapien and R. Ferrus, "Design and Evaluation of a Backhaul-Aware Base Station Assignment Algorithm for OFDMA-Based Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3226–3237, Oct. 2010.

[117] S. Ali and V. Leung, "Dynamic Frequency Allocation in Fractional Frequency Reused OFDMA Networks," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 4286–4295, Aug. 2009.

[118] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental Trade-offs on Green Wireless Networks," *IEEE Commun. Mag.*, vol. 49, pp. 30–37, Jun. 2011.

[119] C. Han, T. Harrold, S. Armour, I. Krikidis, S. Videv, P. Grant, H. Haas, J. Thompson, I. Ku, C.-X. Wang, T. A. Le, M. Nakhai, J. Zhang, and L. Hanzo, "Green Radio: Radio Techniques to Enable Energy-Efficient Wireless Networks," *IEEE Commun. Mag.*, vol. 49, pp. 46–54, Jun. 2011.

[120] H. Bogucka and A. Conti, "Degrees of Freedom for Energy Savings in Practical Adaptive Wireless Systems," *IEEE Commun. Mag.*, vol. 49, pp. 38–45, Jun. 2011.

[121] A. Leon-Garcia and I. Widjaja, *Communication Networks: Fundamentals Concepts and Key Architectures*, 2nd ed. McGraw-Hill Science/Engineering/Math, 2003.

[122] T. Yoo and A. Goldsmith, "On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 528–541, Mar. 2006.

[123] "3rd Generation Partnership Project;Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects (Release 9)," 3GPP TR 36.814 V9.0.0 (2010-03), Tech. Rep.

[124] "E3, T3 and HSSI Manageable Fiber Optic Modems," RADirect, Tech. Rep., [Online] http://www.rad-direct.com/datasheet/fomie3t3.pdf.

[125] A. Mukherjee and A. L. Swindlehurst, "Equilibrium Outcomes of Dynamic Games in MIMO Channels with Active Eavesdroppers," in *Proc. IEEE Intern. Commun. Conf.*, May 2010, pp. 1–5.

[126] G. Amariucai and S. Wei, "Active Eavesdropping in Fast Fading Channels," in *Proc. IEEE Military Commun. Conf.*, Oct. 2009, pp. 1–7.

[127] ——, "Active Eavesdropping in Fast Fading Channels: A Block-Markov Wyner Secrecy Encoding Scheme," in *Proc. IEEE Intern. Sympos. on Inf. Theory*, Jun. 2010, pp. 2508–2512.

[128] H. Dai, "Distributed Versus Co-Located Mimo Systems with Correlated Fading and Shadowing," in *Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Process.*, May 2006, pp. 1–4.

[129] W. Choi and J. Andrews, "Downlink Performance and Capacity of Distributed Antenna Systems in a Multicell Environment," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 69–73, Jan. 2007.

[130] J. Zhang and J. Andrews, "Distributed Antenna Systems with Randomness," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3636–3646, Sep. 2008.

[131] M. Dohler, A. Gkelias, and H. Aghvami, "A Resource Allocation Strategy for Distributed MIMO Multi-Hop Communication Systems," *IEEE Commun. Lett.*, vol. 8, pp. 99–101, Feb. 2004.

[132] H. Zhu, S. Karachontzitis, and D. Toumpakaris, "Low-Complexity Resource Allocation and its Application to Distributed Antenna Systems [Coordinated and Distributed MIMO]," *IEEE Wireless Commun.*, vol. 17, pp. 44–50, Jun. 2010.

[133] Y. Lang, D. Wubben, and K.-D. Kammeyer, "Power Allocations for Adaptive Distributed MIMO Multi-Hop Networks," in *Proc. IEEE Intern. Commun. Conf.*, Jun. 2009, pp. 1–5.

[134] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal Energy Management Policies for Energy Harvesting Sensor Nodes," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1326–1336, Apr. 2010.

[135] V. Shenoy and C. R. Murthy, "Throughput Maximization of Delay-Constrained Traffic in Wireless Energy Harvesting Sensors," in *Proc. IEEE Intern. Commun. Conf.*, May 2010, pp. 1–5.

[136] H. Li, N. Jaggi, and B. Sikdar, "Relay Scheduling for Cooperative Communications in Sensor Networks with Energy Harvesting," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2918–2928, Sep. 2011.

[137] D. Niyato, E. Hossain, M. Rashid, and V. Bhargava, "Wireless Sensor Networks with Energy Harvesting Technologies: A Game-Theoretic Approach to Optimal Energy Management," *IEEE Wireless Commun.*, vol. 14, pp. 90–96, Aug. 2007.

[138] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3rd ed. Wiley-Interscience, 2006.

[139] V. Annapureddy, D. Marathe, T. Ramya, and S. Bhashyam, "Outage Probability of Multiple-Input Single-Output (MISO) Systems with Delayed Feedback," *IEEE Trans. Commun.*, vol. 57, pp. 319–326, Feb. 2009.

[140] H. Gao, P. J. Smith, and M. V. Clark, "Theoretical Reliability of MMSE Linear Diversity Combining in Rayleigh-Fading Additive Interference Channels," *IEEE Trans. Commun.*, vol. 46, pp. 666–672, May 1998.

[141] R. Ganti and M. Haenggi, "Interference and Outage in Clustered Wireless Ad Hoc Networks," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4067–4086, Sep. 2009.

[142] N. Jindal, J. Andrews, and S. Weber, "Rethinking MIMO for Wireless Networks: Linear Throughput Increases with Multiple Receive Antennas," in *Proc. IEEE Intern. Commun. Conf.*, Jun. 2009, pp. 1–6.

[143] C. Fischione, M. D'Angelo, and M. Butussi, "Utility Maximization via Power and Rate Allocation with Outage Constraints in Nakagami-Lognormal Channels," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 1108–1120, Apr. 2011.

[144] C. Goh and X. Q. Yang, *Duality in Optimization and Variational Inequalities*, 1st ed. Taylor & Francis, 2002.

Appendix A

Mathematical Preliminaries

A.1 Convex Analysis

A.1.1 Definitions and Basic Properties

In this section, we introduce some basic properties of concave functions which are useful for understanding the content of this thesis. The three main mathematical references on convex analysis are [75, 77, 138]. For the sake of notational simplicity and to avoid ambiguity, we define the following notations. We use $f\langle\cdot\rangle$ to denote a function f and use $(\cdot)$ to denote a bracket in this section.

Definition A.1 *Let* $f\langle\cdot\rangle : \mathbb{R}^n \to \mathbb{R}$. *Function* f *is said to be concave in* $\mathbb{R}^n$ *if* $\forall\lambda \in [0,1]$ *such that*

$$f\langle\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2\rangle \geq \lambda f\langle\mathbf{x}_1\rangle + (1-\lambda)f\langle\mathbf{x}_2\rangle \tag{A.1}$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$.

Theorem A.1 *Let* $f\langle\cdot\rangle : \mathbb{R}^n \to \mathbb{R}$. *Function* f *is concave in* $\mathbb{R}^n$ *if and only if for any* $\tilde{\mathbf{x}} \in \mathbb{R}^n$,

$$f\langle\mathbf{x}\rangle \leq f\langle\tilde{\mathbf{x}}\rangle + \left[\xi_{\mathbf{x}}\right]^T(\mathbf{x} - \tilde{\mathbf{x}}) \tag{A.2}$$

for each $\mathbf{x} \in \mathbb{R}^n$*, where* $\xi_{\mathbf{x}}$ *is the sub-gradient vector with respect to* $\mathbf{x}$ *at* $\tilde{\mathbf{x}}$*.*

Geometrically, $f\langle\tilde{\mathbf{x}}\rangle + \left[\xi_{\mathbf{x}}\right]^T(\mathbf{x} - \tilde{\mathbf{x}})$ represents a hyperplane which supports the function f from above.

A.1.2 Optimization Problem and Perturbation Function

Without loss of generality, an optimization problem can be written in general form as

$$\begin{aligned} \mathrm{P} = \max_{\mathbf{x} \succeq \mathbf{0}} \quad & f_0\langle\mathbf{x}\rangle & \\ \text{s.t.} \quad & g_i\langle\mathbf{x}\rangle \leq 0, & 1 \leq i \leq L, \\ & h_l\langle\mathbf{x}\rangle = 0, & 1 \leq l \leq M, \end{aligned} \tag{A.3}$$

where $f_0\langle\cdot\rangle : \mathbb{R}^N \to \mathbb{R}$ is the objective function and $\mathbf{x} \in \mathbb{R}^N$ is the vector of optimization variables. $g_i\langle\cdot\rangle : \mathbb{R}^N \to \mathbb{R}$ are L arbitrary functions associated with L inequality constraints. $h_l\langle\cdot\rangle : \mathbb{R}^N \to \mathbb{R}$ are M arbitrary functions associated with M equality constraints. Note that we do not make any assumptions regarding the concavity of functions $f_0\langle\cdot\rangle$, $g_i\langle\cdot\rangle$, and $h_l\langle\cdot\rangle$. Then, the perturbation function is defined as

$$\begin{aligned} v\langle\mathbf{y}\rangle = \max_{\mathbf{x} \succeq \mathbf{0}} \quad & f_0\langle\mathbf{x}\rangle & \\ \text{s.t.} \quad & g_i\langle\mathbf{x}\rangle \leq y_i, & 1 \leq i \leq L, \\ & h_l\langle\mathbf{x}\rangle = y_{L+l}, & 1 \leq l \leq M, \end{aligned} \tag{A.4}$$

where $\mathbf{y} = [y_1, \ldots, y_L, y_{L+1}, \ldots, y_{L+M}]^T \in \mathbb{R}^{L+M}$ is a *perturbation vector*. A geometrical interpretation of the perturbation function is given Figure G.1. Note that $v\langle\mathbf{y}\rangle$ is a *non-decreasing* function of $[y_1, \ldots, y_L]$ since a larger value of each element in

$[y_1, \ldots, y_L]$ results in a larger feasible set.

Duality Theory and KKT Conditions

The basic idea in Lagrangian duality is to take the constraints in (A.3) into account by augmenting the objective function with a weighted sum of the constraint functions. We define the Lagrangian associated with (A.3) as

$$\mathcal{L}\langle \mathbf{x}, \mathbf{u}, \mathbf{w} \rangle = f_0\langle \mathbf{x} \rangle - \sum_{i=1}^{L} u_i \Big(g_i\langle \mathbf{x} \rangle \Big) - \sum_{l=1}^{M} w_l \Big(h_l\langle \mathbf{x} \rangle \Big) \tag{A.5}$$

where $\mathbf{u} \in \mathbb{R}^L$ is the Language multiplier (dual variable) vector with elements $u_i \geq 0$, $i \in \{1, \ldots, L\}$ associated with the i-th inequality constraint $g_i\langle \mathbf{x} \rangle \geq 0$ and $\mathbf{w} \in \mathbb{R}^M$ is the Language multiplier vector with elements w_l, $l \in \{1, \ldots, M\}$ associated with the l-th equality constraint.

Thus, the corresponding dual problem is given by

$$\mathrm{D} = \min_{\mathbf{u} \succeq \mathbf{0}, \mathbf{w}} \max_{\mathbf{x}} \mathcal{L}\langle \mathbf{x}, \mathbf{u}, \mathbf{w} \rangle, \tag{A.6}$$

which is always a convex optimization problem even if the primal problem is not convex. This is because D is a point-wise maximum of a family of affine function with respect to $(\mathbf{u}, \mathbf{w})$.

Theorem A.2 (Weak duality Theorem) *Let* $\mathbf{x}$ *be the a feasible solution to the primal problem P and* $(\mathbf{u}, \mathbf{w})$ *be a feasible solution to the dual problem D. Then,*

$$D \geq P \tag{A.7}$$

holds regardless the convexity of the primal problem.

Definition A.2 (Duality Gap) *The duality gap is defined as the difference between the optimal dual value and the optimal primal value, i.e., $D - P$.*

If the duality gap is zero, i.e., $\mathrm{D}-\mathrm{P}=0$, then strong duality holds and the primal optimum is equal to the dual optimum. In other words, solving the dual problem is equivalent to solving the primal problem.

In the following, we introduce the KKT conditions. Let $\mathbf{x}^*$ and $(\mathbf{u}^*, \mathbf{w}^*)$ be the primal and dual variables at the optimum point. The KKT conditions are given as

$$
\begin{aligned}
g_i\langle \mathbf{x}^* \rangle &\leq 0,\ 1 \leq i \leq L, \\
h_l\langle \mathbf{x}^* \rangle &= 0,\ 1 \leq i \leq M, \\
u_i^* &\geq 0, \text{ (Dual feasibility)} \\
u_i^* g_i\langle \mathbf{x}^* \rangle &= 0, \text{ (Complementary slackness)} \\
\nabla_{\mathbf{x}} f\langle \mathbf{x}^* \rangle - \sum_{i=1}^{L} {u_i}^* \nabla_{\mathbf{x}} g_i\langle \mathbf{x}^* \rangle - \sum_{l=1}^{M} w_l^* \nabla_{\mathbf{x}} h_l\langle \mathbf{x}^* \rangle &= \mathbf{0}, \qquad \text{(A.8)}
\end{aligned}
$$

where u_i^* and w_l^* are the elements of Lagrangian vectors $\mathbf{u}^*$ and $\mathbf{w}^*$, respectively.

Under some technical conditions, (Slater's condition [75]), the KKT conditions are necessary and sufficient for optimality of convex optimization problem. However, if the problem is non-convex, the KKT conditions are only the necessary conditions for optimality.

Appendix B

Proof of Lemma 2.1

The proof of Lemma 2.1 involves three steps. We first derive the channel outage data rate between the BS and user k via relay m on subcarrier i by considering the channel outage probability requirement C1 in (2.21), i.e.,

$$\Pr\left[R_{m,k}^{data}[i] > \frac{1}{2}\min\left\{C_{BR_m,k}[i], C_{R_m,k}[i]\right\}\middle|\boldsymbol{\Delta}_{m,k}[i]\right] = \varepsilon. \tag{B.1}$$

Note that $C_{R_m,k}[i]$ is the only random variable in (B.1) and both $R_{m,k}^{data}[i]$ and $C_{BR_m,k}[i]$ can be controlled via power and packet data rate adaptations. In other words, $R_{m,k}^{data}[i] \leq \frac{1}{2}C_{BR_m,k}[i]$ is guaranteed. Therefore, (B.1) can be written as

$$\begin{aligned}&\Pr\left[R_{m,k}^{data}[i] > \frac{1}{2}C_{R_m,k}[i]\middle|\boldsymbol{\Delta}_{m,k}[i], C_{BR_m,k}[i] > C_{R_m,k}[i]\right]\\ &\times \Pr\left[C_{BR_m,k}[i] > C_{R_m,k}[i]\middle|\boldsymbol{\Delta}_{m,k}[i]\right].\end{aligned} \tag{B.2}$$

On the other hand, it can be observed that the outage capacity $R_{m,k}^{data}[i](1-\varepsilon)$ is linearly increasing with $R_{m,k}^{data}[i]$ for a fixed target outage requirement ε and is upper bounded by $\frac{1}{2}C_{BR_m,k}[i]$. Therefore, the outage capacity is maximized if we control $\frac{1}{2}C_{BR_m,k}[i]$ such that it is equal to $R_{m,k}^{data}[i]$. Therefore, (B.2) can be further simplified

as

$$
\begin{aligned}
& \Pr\left[C_{BR_m,k}[i] > C_{R_m,k}[i] \middle| \mathbf{\Delta}_{m,k}[i]\right] \qquad \text{(B.3)}\\
= \quad & \Pr\left[\frac{(2^{2R^{data}_{m,k}[i]}-1)(1+(1-\alpha_{R_m,k}[i])P_{R_m,k}[i]\sigma_e^2)}{\alpha_{R_m,k}[i]P_{R_m,k}[i]}\right.\\
& \left. > \hat{\mathbf{r}}^{\dagger}_{m,k}[i]\mathbf{h}^{\dagger}_{R_m,k}[i]\mathbf{h}_{R_m,k}[i]\hat{\mathbf{r}}_{R_m,k}[i] \middle| \mathbf{\Delta}_{m,k}[i]\right]\\
= \quad & F_{\chi_2}\Big(\frac{(2^{2R^{data}_{m,k}[i]}-1)(1+(1-\alpha_{R_m,k}[i])P_{R_m,k}[i]\sigma_e^2)}{\alpha_{R_m,k}[i]P_{R_m,k}[i]}, i\Big) = \varepsilon\\
\Rightarrow \quad & R^{data}_{m,k}[i] = \min\Big\{\frac{1}{2}\log_2\Big(1+\frac{\alpha_{R_m,k}[i]P_{R_m,k}[i]F^{-1}_{\chi_2}(\varepsilon,i)}{1+(1-\alpha_{R_m,k})P_{R_m,k}[i]\sigma_e^2}\Big), \frac{1}{2}C_{BR_m,k}[i]\Big\}
\end{aligned}
$$

where $F_{\chi_2}(\cdot, i)$ denotes the CDF of a non-central chi-square random variable with 2 degrees of freedom and non-centrality parameter $\hat{\mathbf{h}}_{R_m,k}[i]\hat{\mathbf{h}}^{\dagger}_{R_m,k}[i]$ [139]. Then, we can derive the outage secrecy data rate by calculating the secrecy outage probability in C2 in (2.21). Without loss of generality, we define the secrecy data rate and outage data rate as $R^{sec}_{m,k}[i] = \frac{1}{2}\log_2(r^{sec}_{m,k}[i])$ and $R^{data}_{m,k}[i] = \frac{1}{2}\log_2(r^{data}_{m,k}[i])$, respectively. We assume that $\alpha_{R_m,k}[i] = \max\{\alpha_{R_m,k}[i], \alpha_{BR_m,k}[i]\}$. This assumption is necessary for deriving an efficient resource allocation algorithm. It results in an upper bound on the secrecy outage probability and a lower bound on the secrecy data rate. Then, the secrecy outage probability can be expressed as

$$
\begin{aligned}
& \Pr\left[R^{data}_{m,k}[i] - C_{m,E}[i] \le R^{sec}_{m,k}[i] \middle| \mathbf{\Delta}_{m,k}[i]\right] = \Pr\left[(\frac{r^{data}_{m,k}[i]}{r^{sec}_{m,k}[i]} - 1)\frac{1}{(N_T-1)}\right.\\
& \left. \le \max\Big\{\frac{\alpha_{BR_m,k}[i]}{1-\alpha_{BR_m,k}[i]}\Omega_1[i], \frac{\alpha_{R_m,k}[i]}{1-\alpha_{R_m,k}[i]}\Omega_2[i]\Big\} \middle| \mathbf{\Delta}_{m,k}[i]\right]\\
& \le \Pr\left[(\frac{r^{data}_{m,k}[i]}{r^{sec}_{m,k}[i]} - 1)\frac{1-\alpha_{R_m,k}[i]}{\alpha_{R_m,k}[i](N_T-1)} \le \max\Big\{\Omega_1[i], \Omega_2[i]\Big\} \middle| \mathbf{\Delta}_{m,k}[i]\right], \qquad \text{(B.4)}
\end{aligned}
$$

where $\Omega_1[i] = \mathbf{g}_1^{\dagger}[i](\mathbf{G}_1[i]\mathbf{G}_1^{\dagger}[i])^{-1}\mathbf{g}_1[i]$ and $\Omega_2[i] = \mathbf{g}_2^{\dagger}[i](\mathbf{G}_2[i]\mathbf{G}_2^{\dagger}[i])^{-1}\mathbf{g}_2[i]$. Note that the upper bound on the secrecy outage probability in (B.4) is due to the assumption of $\alpha_{R_m,k}[i] = \max\{\alpha_{R_m,k}[i], \alpha_{BR_m,k}[i]\}$. If $\alpha_{BR_m,k}[i] = \max\{\alpha_{R_m,k}[i], \alpha_{BR_m,k}[i]\}$, the inequality is also valid by replacing $\alpha_{R_m,k}[i]$ by $\alpha_{BR_m,k}[i]$ in (B.4). On the other hand, since $\Omega_1[i]$ and $\Omega_2[i]$ are i.i.d. random variables, we have the follow identity

$$\Pr\left[z \leq \Omega_1[i]\middle|\mathbf{\Delta}_{m,k}[i]\right] = \Pr\left[z \leq \Omega_2[i]\middle|\mathbf{\Delta}_{m,k}[i]\right], \tag{B.5}$$

where $z = \frac{(r_{m,k}^{data}[i] - r_{m,k}^{sec}[i])(1-\alpha_{R_m,k}[i])}{r_{m,k}^{sec}[i]\alpha_{R_m,k}[i](N_T-1)}$. Hence, the secrecy outage probability in (B.4) can be written as

$$\begin{aligned} F_{z_c}(z,i) &= \Pr\left[z \leq \max\left\{\Omega_1[i], \Omega_2[i]\right\}\middle|\mathbf{\Delta}_{m,k}[i]\right] \\ &= \Pr\left[z \leq \Omega_1[i]\middle|\mathbf{\Delta}_{m,k}[i]\right] + \Pr\left[z \leq \Omega_2[i]\middle|\mathbf{\Delta}_{m,k}[i]\right] \\ &- \Pr\left[z \leq \Omega_1[i]\middle|\mathbf{\Delta}_{m,k}[i]\right] \times \Pr\left[z \leq \Omega_2[i]\middle|\mathbf{\Delta}_{m,k}[i]\right]. \end{aligned} \tag{B.6}$$

On the other hand, it can be observed that $\Omega_1[i]$ is equivalent to the SIR of an N_E-branch MMSE diversity combiner for $N_T - 1$ interferers. The corresponding complementary CDF (CCDF) is given by [69, 140]

$$\Pr\left[z \leq \Omega_1[i]\middle|\mathbf{\Delta}_{m,k}[i]\right] = F_{\Omega}(z) = \frac{\sum_{n=0}^{N_E-1}\binom{N_T-1}{n}z^n}{(1+z)^{N_T-1}}. \tag{B.7}$$

Therefore, the target secrecy outage probability $F_{z_c}(z,i)$ can be obtained by substituting (B.7) into (B.6), which yields

$$\begin{aligned} F_{z_c}(z,i) &= F_{\Omega}(z) + F_{\Omega}(z) - F_{\Omega}(z) \times F_{\Omega}(z) \\ &= \frac{\sum_{n=0}^{N_E-1}\binom{N_T-1}{n}2z^n}{(1+z)^{N_T-1}} - \frac{\sum_{n=0}^{N_E-1}\sum_{m=0}^{N_E-1}\binom{N_T-1}{n}\binom{N_T-1}{m}z^{m+n}}{(1+z)^{2N_T-2}}. \end{aligned} \tag{B.8}$$

For a target secrecy outage probability of δ, z can be expressed as

$$\begin{aligned} z &= F_{z_c}^{-1}(\delta, i) \\ \Longrightarrow \qquad & R_{m,k}^{sec}[i] = \Big[R_{m,k}^{data}[i] - \frac{1}{2}\log_2\Big(1 + \frac{\alpha_{R_m,k}[i](N_T-1)F_{z_c}^{-1}(\delta,i)}{1-\alpha_{R_m,k}[i]}\Big)\Big]^+, \end{aligned} \tag{B.9}$$

where $F_{z_c}^{-1}(\delta, i)$ is the inverse CCDF of random variable $\max\Big\{\Omega_1[i], \Omega_2[i]\Big\}$, which can be computed efficiently by numerical solvers or implemented as a look-up table in practice. The final step in deriving the lemma is to calculate the asymptotically optimal $\alpha_{R_m,k}^*[i]$ and $\alpha_{BR_m,k}^*[i]$ in high SNR. Let $\Phi_{BR_m,k}[i] = P_{BR_m,k}[i]\lambda_{\max_{BR_m}}[i]$, $\Phi_{R_m,k}[i] = F_{\chi_2}^{-1}(\varepsilon, i)/\sigma_e^2$, and $\Lambda_E[i] = (N_T-1)F_{z_c}^{-1}(\delta, i)$. The expression for the secrecy data rate of user k on subcarrier i depends on the link qualities of the BS-to-relay link and the relay-to-user link, cf. (B.4), (B.9). If the BS-to-relay link is weaker than the relay-to-user link, then the secrecy data rate can be expressed as

$$R_k^{sec}[i] = \frac{1}{2}C_{BR_m,k}[i] - \frac{1}{2}\log_2\Big(1 + \frac{\alpha_{BR_m,k}[i]F_{z_c}^{-1}(\delta,i)(N_T-1)}{1-\alpha_{BR_m,k}[i]}\Big). \tag{B.10}$$

On the other hand, if the relay-to-user link is weaker than the BS-to-relay link, the secrecy data rate of user k on subcarrier i in high SNR is lower bounded by

$$\begin{aligned} R_k^{sec}[i] > \frac{1}{2}\Big\{ & \log_2\Big(1 + \frac{P_{R_m,k}[i]F_{\chi_2}^{-1}(\varepsilon,i)\alpha_{B,k}^*[i]}{1+P_{R_m,k}[i]\sigma_e^2}\Big) \\ & - \log_2\Big(1 + \frac{\alpha_{R_m,k}[i]F_{z_c}^{-1}(\delta,i)(N_T-1)}{1-\alpha_{R_m,k}[i]}\Big)\Big\}. \end{aligned} \tag{B.11}$$

In fact, the term $\frac{P_{R_m,k}[i]F_{\chi_2}^{-1}(\varepsilon,i)\alpha_{B,k}^*[i]}{P_{R_m,k}[i]\sigma_e^2}$ in (B.11) can be interpreted as an SINR under a virtual interferer with interference power $P_{R_m,k}[i]\sigma_e^2$. By standard optimization techniques, it can be shown that the optimal $\alpha_{BR_m,k}^*[i]$ and $\alpha_{R_m,k}^*[i]$ maximizing (B.10)

and (B.11) have the same asymptotic expression in high SNR ($P_{BR_m,k}[i], P_{R_m,k} \to \infty$):

$$\begin{aligned} \alpha^*_{BR_m,k}[i] &= \frac{-\Phi_{BR_m,k}[i] + \sqrt{\Phi_{BR_m,k}[i]\Lambda_E[i](\Phi_{BR_m,k}[i] - \Lambda_E[i] + 1)}}{\Phi_{BR_m,k}[i](\Lambda_E[i] - 1)} \\ &\overset{(a)}{\approx} \frac{1}{\sqrt{\Lambda_E[i]}} \quad \text{and} \end{aligned} \tag{B.12}$$

$$\begin{aligned} \alpha^*_{R_m,k}[i] &= \frac{-\Phi_{R_m,k}[i] + \sqrt{\Phi_{R_m,k}[i]\Lambda_E[i](\Phi_{R_m,k}[i] - \Lambda_E[i] + 1)}}{\Phi_{R_m,k}[i](\Lambda_E[i] - 1)} \\ &\overset{(b)}{\approx} \frac{1}{\sqrt{\Lambda_E[i]}}, \end{aligned} \tag{B.13}$$

respectively. (a) is due to the high SNR assumption, i.e., $\Phi_{BR_m,k}[i] \gg \Lambda_E[i] \gg 1$. The assumption of high SNR is necessary for arriving at an efficient resource allocation algorithm. Note that $\Phi_{BR_m,k}[i] \gg \Lambda_E[i]$ is always valid in the high transmit power regime as $\Phi_{BR_m,k}[i]$ increases with the total transmit power while $\Lambda_E[i]$ remains constant. On the other hand, (b) is due to $\Phi_{R_m,k}[i] \gg \Lambda_E[i] \gg 1$, which holds for reasonably small channel estimation error variance σ_e^2 (e.g. $\sigma_e^2 \ll \sigma^2_{h_{R_m,k}}$) and secrecy outage requirement δ (e.g. $\delta \ll 1$).

Appendix C

Proof of Theorem 3.1 and Algorithm Convergence

Proof of Theorem 3.1 We now prove the forward implication of Theorem 3.1 by following a similar approach as in [84]. Without loss of generality, we define q^* and $\{\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*\} \in \mathcal{F}$ as the optimal energy efficiency and the optimal resource allocation policies of the *original objective function* in (3.13), respectively. Then, the optimal energy efficiency can be expressed as

$$q^* = \frac{U_{sec}(\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*)}{U_{TP}(\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*)} \geq \frac{U_{sec}(\mathcal{P}, \mathcal{R}, \mathcal{S})}{U_{TP}(\mathcal{P}, \mathcal{R}, \mathcal{S})}, \qquad \forall \{\mathcal{P}, \mathcal{R}, \mathcal{S}\} \in \mathcal{F}, \tag{C.1}$$

$$\implies U_{sec}(\mathcal{P}, \mathcal{R}, \mathcal{S}) - q^* U_{TP}(\mathcal{P}, \mathcal{R}, \mathcal{S}) \leq 0$$

$$\text{and } U_{sec}(\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*) - q^* U_{TP}(\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*) = 0. \tag{C.2}$$

Therefore, we conclude that $\max\limits_{\mathcal{P}, \mathcal{R}, \mathcal{S}, \alpha_k[i]} U_{sec}(\mathcal{P}, \mathcal{R}, \mathcal{S}) - q^* U_{TP}(\mathcal{P}, \mathcal{R}, \mathcal{S}) = 0$ and it is achievable by resource allocation policies $\{\mathcal{P}^*, \mathcal{R}^*, \mathcal{S}^*\}$. This completes the forward implication.

Next, we prove the converse implication of Theorem 3.1. Suppose $\{\mathcal{P}_e^*, \mathcal{R}_e^*, \mathcal{S}_e^*\}$ is the optimal resource allocation policies of the *equivalent objective function* such that $U_{sec}(\mathcal{P}_e^*, \mathcal{R}_e^*, \mathcal{S}_e^*) - q^* U_{TP}(\mathcal{P}_e^*, \mathcal{R}_e^*, \mathcal{S}_e^*) = 0$. Then, for any feasible resource allocation

policies $\{\mathcal{P}, \mathcal{R}, \mathcal{S}\} \in \mathcal{F}$, we can obtain the following inequality

$$U_{sec}(\mathcal{P}, \mathcal{R}, \mathcal{S}) - q^* U_{TP}(\mathcal{P}, \mathcal{R}, \mathcal{S}) \leq U_{sec}(\mathcal{P}_e^*, \mathcal{R}_e^*, \mathcal{S}_e^*) - q^* U_{TP}(\mathcal{P}_e^*, \mathcal{R}_e^*, \mathcal{S}_e^*) = 0. \quad \text{(C.3)}$$

The above inequality implies

$$\frac{U_{sec}(\mathcal{P}, \mathcal{R}, \mathcal{S})}{U_{TP}(\mathcal{P}, \mathcal{R}, \mathcal{S})} \leq q^* \quad \forall \{\mathcal{P}, \mathcal{R}, \mathcal{S}\} \in \mathcal{F} \quad \text{and} \quad \frac{U_{sec}(\mathcal{P}_e^*, \mathcal{R}_e^*, \mathcal{S}_e^*)}{U_{TP}(\mathcal{P}_e^*, \mathcal{R}_e^*, \mathcal{S}_e^*)} = q^*. \quad \text{(C.4)}$$

In other words, the optimal resource allocation policies $\{\mathcal{P}_e^*, \mathcal{R}_e^*, \mathcal{S}_e^*\}$ for the equivalent objective function is also the optimal resource allocation policies for the original objective function.

This complete the proof of the converse implication of Theorem 3.1. In summary, the optimization of the original objective function and the optimization of the equivalent objective function result in the same resource allocation policies. □

Proof of Algorithm Convergence

We follow a similar approach as in [84] for proving the convergence of Algorithm 3.1 in Table 3.1. We first introduce the following two propositions. For the sake of notational simplicity, we define the equivalent objective function in (3.16) as $F(q') = \max_{\mathcal{P}, \mathcal{R}, \mathcal{S}, \alpha_k[i]} \{U_{sec}(\mathcal{P}, \mathcal{R}, \mathcal{S}) - q' U_{TP}(\mathcal{P}, \mathcal{R}, \mathcal{S})\}$.

Proposition C.1 *$F(q')$ is a strictly monotonic decreasing function in q', i.e., $F(q'') > F(q')$ if $q' > q''$.*

Proof: Let $\{\mathcal{P}', \mathcal{R}', \mathcal{S}'\} \in \mathcal{F}$ and $\{\mathcal{P}'', \mathcal{R}'', \mathcal{S}''\} \in \mathcal{F}$ be the two distinct optimal resource allocation polices for $F(q')$ and $F(q'')$, respectively.

$$\begin{aligned} F(q'') &= \max_{\mathcal{P},\mathcal{R},\mathcal{S},\alpha_k[i]} \{U_{sec}(\mathcal{P},\mathcal{R},\mathcal{S}) - q'' U_{TP}(\mathcal{P},\mathcal{R},\mathcal{S})\} \\ &= U_{sec}(\mathcal{P}'',\mathcal{R}'',\mathcal{S}'') - q'' U_{TP}(\mathcal{P}'',\mathcal{R}'',\mathcal{S}'') \\ &> U_{sec}(\mathcal{P}',\mathcal{R}',\mathcal{S}') - q'' U_{TP}(\mathcal{P}',\mathcal{R}',\mathcal{S}') \geq U_{sec}(\mathcal{P}',\mathcal{R}',\mathcal{S}') - q' U_{TP}(\mathcal{P}',\mathcal{R}',\mathcal{S}') \\ &= F(q'). \end{aligned} \tag{C.5}$$

□

Proposition C.2 *Let $\{\mathcal{P}', \mathcal{R}', \mathcal{S}'\} \in \mathcal{F}$ be an arbitrary feasible solution and $q' = \frac{U_{sec}(\mathcal{P}',\mathcal{R}',\mathcal{S}')}{U_{TP}(\mathcal{P}',\mathcal{R}',\mathcal{S}')}$, then $F(q') \geq 0$.*

Proof:
$$\begin{aligned} F(q') &= \max_{\mathcal{P},\mathcal{R},\mathcal{S},\alpha_k[i]} \{U_{sec}(\mathcal{P},\mathcal{R},\mathcal{S}) - q' U_{TP}(\mathcal{P},\mathcal{R},\mathcal{S})\} \\ &\geq U_{sec}(\mathcal{P}',\mathcal{R}',\mathcal{S}') - q' U_{TP}(\mathcal{P}',\mathcal{R}',\mathcal{S}') = 0. \end{aligned} \tag{C.6}$$

□

We are now ready to prove the convergence of Algorithm 3.1 in Table 3.1.

Proof of Convergence: We first prove that the energy efficiency q increases in each iteration. Then, we prove that if the number of iterations is large enough, the energy efficiency q converges to the optimal q^* such that it satisfy the optimality condition in Theorem 3.1, i.e., $F(q^*) = 0$.

Let $\{\mathcal{P}_n, \mathcal{R}_n, \mathcal{S}_n\}$ be the optimal resource allocation polices in the n-th iteration. Suppose $q_n \neq q^*$ and $q_{n+1} \neq q^*$ represent the energy efficiency of the considered system in iterations n and $n+1$, respectively. By Theorem 3.1 and Proposition C.2, $F(q_n) > 0$ and $F(q_{n+1}) > 0$ must be true. On the other hand, in the proposed

algorithm, we calculate q_{n+1} as $q_{n+1} = \frac{U_{sec}(\mathcal{P}_n, \mathcal{R}_n, \mathcal{S}_n)}{U_{TP}(\mathcal{P}_n, \mathcal{R}_n, \mathcal{S}_n)}$. Thus, we can express $F(q_n)$ as

$$\begin{aligned} F(q_n) &= U_{sec}(\mathcal{P}_n, \mathcal{R}_n, \mathcal{S}_n) - q_n U_{TP}(\mathcal{P}_n, \mathcal{R}_n, \mathcal{S}_n) \\ &= U_{TP}(\mathcal{P}_n, \mathcal{R}_n, \mathcal{S}_n)(q_{n+1} - q_n) > 0 \qquad \text{(C.7)} \\ \Longrightarrow q_{n+1} &> q_n, \quad \because U_{TP}(\mathcal{P}_n, \mathcal{R}_n, \mathcal{S}_n) > 0. \end{aligned}$$

By combining $q_{n+1} > q_n$, Proposition C.1, and Proposition C.2, we can show that as long as the number of iterations is large enough, $F(q_n)$ will eventually approach zero and satisfy the optimality condition as stated in Theorem 3.1. □

Appendix D

Proof of Proposition 3.1

Without loss of generality, we define the secrecy data rate as $R_k[i] = W\log_2(r_k[i])$. Now, the secrecy outage probability can be expressed as

$$\begin{aligned} &\Pr\Big[C_k[i] - C_E[i] \leq R_k[i]\Big|\mathbf{h}_k[i]\Big] = \varepsilon \qquad \text{(D.1)}\\ \Longrightarrow &\Pr\Big[\underbrace{\Big(\frac{1}{r_k[i]}(1+\Gamma_k[i])-1\Big)\frac{1-\alpha_k[i]}{(N_T-1)\alpha_k[i]}}_{\Theta_k[i]} \leq \underbrace{\mathbf{g}_1^\dagger[i](\mathbf{G}_2[i]\mathbf{G}_2^\dagger[i])^{-1}\mathbf{g}_1[i]}_{Z_k[i]}\Big|\mathbf{h}_k[i]\Big] = \varepsilon, \end{aligned}$$

where $\Gamma_k[i] = \frac{\alpha_k[i]P_k[i]\lambda_{\max_k}[i]}{N_0W}$ and $Z_k[i]$ is an unknown random variable for the BS. Since the supermatrix $\mathbf{B}_k[i] = \Big[\mathbf{b}_k[i]\ \mathbf{V}_k[i]\Big]$ is an unitary matrix, $\mathbf{B}_k[i]\tilde{\mathbf{G}}[i]$ has i.i.d. complex Gaussian entries. As a result, $Z_k[i]$ is equivalent to the SIR of an N_E-branch MMSE diversity combiner for N_T-1 interferers. Hence, the corresponding CCDF is given by [69, 140]

$$F_{z_c}(z) = \frac{\sum_{n=0}^{N_E-1}\binom{N_T-1}{n}z^n}{(1+z)^{N_T-1}}, \quad \forall z \geq 0. \qquad \text{(D.2)}$$

Therefore, for a target secrecy outage probability of ε, $\Theta_k[i]$ can be expressed as

$$\begin{aligned} &\Theta_k[i] = F_{z_c}^{-1}(\varepsilon) \Longrightarrow \qquad \text{(D.3)}\\ &R_k[i] = W\Big[\log_2\Big(1+\frac{\alpha_k[i]P_k[i]\lambda_{\max_k}[i]}{N_0W}\Big) - \log_2\Big(1+\frac{\alpha_k[i]}{1-\alpha_k[i]}(N_T-1)F_{z_c}^{-1}(\varepsilon)\Big)\Big]^+, \end{aligned}$$

where $F_{z_c}^{-1}(\varepsilon)$ is the inverse CCDF of random variable $Z_k[i]$, which can be computed efficiently by numerical solvers or implemented as a look-up table for practical implementation. The second step in solving the optimization problem in (3.13) is to calculate the fraction of power allocated to each subcarrier for generating the artificial noise. By standard optimization techniques, the asymptotically optimal $\alpha_k^*[i]$ maximizing the secrecy outage capacity on subcarrier i for a fixed $P_k[i]$ in high SNR is given by

$$\begin{aligned} \alpha_k^*[i] &= \frac{\Gamma_k[i] - \sqrt{(\Gamma_k[i])^2\,\Lambda_E[i] - \Gamma_k[i]\,(\Lambda_E[i])^2 + \Gamma_k[i]\,\Lambda_E[i]}}{\Gamma_k[i] - \Gamma_k[i]\,\Lambda_E[i]} \\ &\overset{(a)}{\approx} \frac{\sqrt{\Lambda_E[i]} - 1}{\Lambda_E[i] - 1} \approx \frac{1}{\sqrt{\Lambda_E[i]}}, \end{aligned} \tag{D.4}$$

where (a) is due to the high SNR[25] assumption, i.e., $\Gamma_k[i] \gg \Lambda_E[i] \gg 1$. Note that $\Gamma_k[i] \gg \Lambda_E[i]$ is always valid in the high transmit power regime as $\Gamma_k[i]$ increases with the total transmit power while $\Lambda_E[i]$ remains constant. On the other hand, $\Lambda_E[i] \gg 1$ holds for a reasonably small secrecy outage requirement adopted in practical applications, i.e., $\varepsilon \ll 1$. □

[25]The assumption of high SNR is necessary for arriving at an efficient resource allocation algorithm. Note that the high SNR assumption does not necessarily require a high transmit power. High SNR can be achieved by exploiting multiuser diversity or using MIMO-beamforming for moderate or small transmit powers.

Appendix E

Proof of Proposition 4.1

The outage probability requirement C3 is a complicated non-convex function of data rates and powers, and a closed-form expression for the corresponding distribution function is not available. Therefore, we tackle this issue by the following approximations. We focus on an upper bound on the actual outage probability by bounding $\Pr(\Gamma_{i,k} < c) = \Pr(C_{i,k} < R_{i,k}), 1 \leq k \leq K$, with an outage probability requirement ε, where $\Gamma_{i,k}$ is defined in (4.3) and $c = 2^{\frac{R_{i,k}}{W}} - 1$. For notational simplicity, we define variables $\Phi_j = |\mathbf{h}_{i,k}^T \hat{\mathbf{f}}_{i,j}|^2 P_{i,j} s_{i,j} l_k g_k \geq 0$, $\forall j \neq k$, $\Phi = \sum_{j \neq k} \Phi_j + N_0 W$, and $B = P_{i,k} l_k g_k |\mathbf{h}_{i,k}^T \hat{\mathbf{f}}_{i,k}|^2$. Suppose now we restrict the resource allocator such that $\Pr\Big(\Phi \geq c_2\Big) \leq \frac{\varepsilon}{2}$ and $\Pr\Big(B \leq c_1\Big) = \frac{\varepsilon}{2}$, where $\frac{c_1}{c_2} = c = 2^{\frac{R_{i,k}}{W}} - 1$ is a function of the scheduled data rate, and c_1 and c_2 are positive constants that will be specified in the following. Hence, the actual outage probability can be expressed as

$$\begin{aligned} \Pr\Big(C_{i,k} < R_{i,k}\Big) = \quad & \underbrace{\Pr\Big(\frac{B}{c_1} c_2 < \Phi \Big| B \leq c_1\Big)}_{a'} \Pr\Big(B \leq c_1\Big) \\ & + \underbrace{\Pr\Big(\frac{B}{c_1} c_2 < \Phi \Big| B > c_1\Big)}_{b'} \Pr\Big(B > c_1\Big). \end{aligned} \tag{E.1}$$

For calculating b', it can be observed that $b' \leq \frac{\varepsilon}{2}$ since $\frac{B}{c_1} > 1$ and $\Pr\Big(\Phi \geq c_2\Big) \leq \frac{\varepsilon}{2}$. On the other hand, $a' \leq 1$. Thus, the actual outage probability $\Pr\Big(C_{i,k} < R_{i,k}\Big)$ is

bounded by

$$\Pr\Big(C_{i,k} < R_{i,k}\Big) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}(1 - \frac{\varepsilon}{2}) = \varepsilon - \frac{\varepsilon^2}{4} \approx \varepsilon \quad \text{for} \quad \varepsilon \ll 1. \tag{E.2}$$

In other words, the outage probability requirement $\Pr\Big(C_{i,k} < R_{i,k}\Big) \leq \varepsilon$ is satisfied if we guarantee $\Pr\Big(\Phi \geq c_2\Big) \leq \frac{\varepsilon}{2}$ and $\Pr\Big(B \leq c_1\Big) = \frac{\varepsilon}{2}$.

Next, we calculate $\Pr\Big(\Phi \geq c_2\Big)$ which represents the probability that the sum power of the $K-1$ inter-user interferers exceeds c_2. Let $c_2 = \sum_{j\neq k} \eta_j + N_0 W$, where η_j are dummy variables. We obtain

$$\begin{aligned} \Pr\Big(\Phi \geq c_2\Big) &= \Pr\Big(\sum_{j\neq k} \Phi_j + N_0 W \geq \sum_{j\neq k} \eta_j + N_0 W\Big) \\ &\overset{(a)}{\leq} \frac{\mathcal{E}\{\sum_{j\neq k} \Phi_j\}}{\sum_{j\neq k} \eta_j} = \frac{\sum_{j\neq k} P_{i,j} s_{i,j} l_k g_k}{\sum_{j\neq k} \eta_j}, \end{aligned} \tag{E.3}$$

where (a) is due to Markov's inequality [141, 142]. Note that although Markov's inequality may not be the tightest upper bound for the corresponding outage probability, it has been widely adopted in the literature [141, 142] for calculating the outage probability in interference channels, since it only requires the first moment of the random variable. As a result, if we set $\eta_j = P_{i,j} s_{i,j} l_k g_k (\frac{2}{\varepsilon})$, then we have[26]

$$\Pr\Big(\Phi \geq c_2\Big) = \Pr\Big(\sum_{j\neq k} |\mathbf{h}_{i,k}^T \hat{\mathbf{f}}_{i,j}|^2 P_{i,j} s_{i,j} l_k g_k \geq \sum_{j\neq k} \frac{2}{\varepsilon} P_{i,j} s_{i,j} l_k g_k\Big) \leq \frac{\varepsilon}{2}. \tag{E.4}$$

[26]Note that in [92, 143], the denominator of the SINR is approximated by only its mean value. However, this approximation cannot guarantee a small channel outage probability requirement ε.

For calculating $\Pr\Big(B \leq c_1\Big)$, we consider

$$\begin{aligned} |\mathbf{h}_{i,k}^T \hat{\mathbf{f}}_{i,k}|^2 &= \Big[\big(\hat{\mathbf{h}}_{i,k}^T + \Delta\mathbf{h}_{i,k}^T\big)\frac{\hat{\mathbf{h}}_{i,k}}{\|\hat{\mathbf{h}}_{i,k}\|}\Big]^2 = \|\hat{\mathbf{h}}_{i,k}\|^2 + 2\Re(\Delta\mathbf{h}_{i,k}^T\hat{\mathbf{h}}_{i,k}) + \frac{\|\Delta\mathbf{h}_{i,k}^T\hat{\mathbf{h}}_{i,k}\|^2}{\|\hat{\mathbf{h}}_{i,k}\|^2} \\ &\overset{(c)}{\approx} \|\hat{\mathbf{h}}_{i,k}\|^2 = \Theta\Big(N_{T_{i,k}}(1-\sigma_e^2)\Big) \qquad \text{for } N_{T_{i,k}} \to \infty, \end{aligned} \tag{E.5}$$

where (c) is due to the fact that $\|\hat{\mathbf{h}}_{i,k}\|^2$ scales with $N_{T_{i,k}}$ in the order of $\Theta(N_{T_{i,k}}(1-\sigma_e^2))$ for $N_{T_{i,k}} \to \infty$, thanks to the law of large numbers. Note that $\|\hat{\mathbf{h}}_{i,k}\|^2$ is a random variable if $N_{T_{i,k}}$ is an unknown before solving the optimization problem. On the other hand, the term $2\Re(\Delta\mathbf{h}_{i,k}^T\hat{\mathbf{h}}_{i,k}) + \frac{\|\Delta\mathbf{h}_{i,k}^T\hat{\mathbf{h}}_{i,k}\|^2}{\|\hat{\mathbf{h}}_{i,k}\|^2}$ scales only in the order of $\mathcal{O}(1)$ which can be neglected for large $N_{T_{i,k}}$. By choosing $c_1 = P_{i,k}l_k g_k N_{T_{i,k}}(1-\sigma_e^2)(1-\delta)$, $\Pr\Big(B \leq c_1\Big)$ can be upper bounded by its Chernoff bound as

$$\begin{aligned} \Pr\Big(B \leq c_1\Big) &\approx \Pr\Big(\|\hat{\mathbf{h}}_{i,k}\|^2 \leq N_{T_{i,k}}(1-\sigma_e^2)(1-\delta)\Big) \\ &\leq \phi^{N_{T_{i,k}}} \exp\Big((1-\phi)N_{T_{i,k}}\Big) = \frac{\varepsilon}{2}, \end{aligned} \tag{E.6}$$

where $\phi = (1-\sigma_e^2)(1-\delta)$ and $0 < \delta < 1$ is a constant backoff factor. Mathematically, δ represents the deviation of $(1-\sigma_e^2)$ from $\frac{\|\hat{\mathbf{h}}_{i,k}\|^2}{N_{T_{i,k}}}$, for a finite value of $N_{T_{i,k}}$. For a given outage probability requirement ε and backoff factor δ, solving (E.6) for $N_{T_{i,k}}$ results in the minimum required N_{th} for satisfying the outage requirement. Note that for a target outage probability requirement ε, the actual outage probability for the case of $N_{T_{i,k}} \geq N_{th}$ will be less than ε since $\phi^{N_{T_{i,k}}} \exp\Big((1-\phi)N_{T_{i,k}}\Big)$ is a decreasing function of $N_{T_{i,k}}$. Therefore, by combining (E.4) and (E.6), a scheduled data rate of $R_{i,k} = (1-\varepsilon)W\log_2\Big(1+\frac{c_1}{c_2}\Big) = (1-\varepsilon)W\log_2\Big(1+\frac{P_{i,k}l_k g_k N_{T_{i,k}}(1-\sigma_e^2)(1-\delta)}{WN_0+\sum_{j\neq k}(\frac{2}{\varepsilon})P_{i,j}s_{i,j}l_k g_k}\Big)$ can satisfy the outage probability requirement $\Pr\Big(C_{i,k} < R_{i,k}\Big) \leq \varepsilon$ which proves in Proposition 1. We note that the use of the strong law of large numbers in (E.5) makes the

optimization of $N_{T_{i,k}}$ possible since $N_{T_{i,k}}$ becomes a part of the equivalent channel gain.

Appendix F

Proof of the Concavity of the Transformed Problem in (4.14)

For notational simplicity, we drop the *subindices* and *scaling constants* of all optimization variables in this section such that the transformed objective function in (4.14) can be expressed as the summation of two functions with variables P, s, and N_T, i.e., $y = f + t$, where $f = s\log_2(P/s) + s\log_2(N_T/s)$ and $t = -q\tilde{U}(\mathcal{P}, \mathcal{A}, \mathcal{R}, \mathcal{S})$. Let $\mathbf{H}(f)$ and $\lambda_1, \lambda_2, \lambda_3$ be the Hessian matrix of function f and the eigenvalues of $\mathbf{H}(f)$, respectively. The Hessian matrix of function f and the trace of the Hessian matrix are given by

$$\mathbf{H}(f) = \begin{bmatrix} \frac{-2}{s\ln(2)} & \frac{1}{P\ln(2)} & \frac{1}{N_T\ln(2)} \\ \frac{1}{P\ln(2)} & \frac{-s}{P^2\ln(2)} & 0 \\ \frac{1}{N_T\ln(2)} & 0 & \frac{-s}{N_T^2\ln(2)} \end{bmatrix} \quad \text{and} \tag{F.1}$$

$$\mathrm{tr}\Big(\mathbf{H}(f)\Big) = \sum_{n=1}^{3}\lambda_n = -\frac{s^2\,P^2 + s^2\,N_T^2 + 2\,P^2\,N_T^2}{s\,P^2\,N_T^2\,\ln(2)}, \tag{F.2}$$

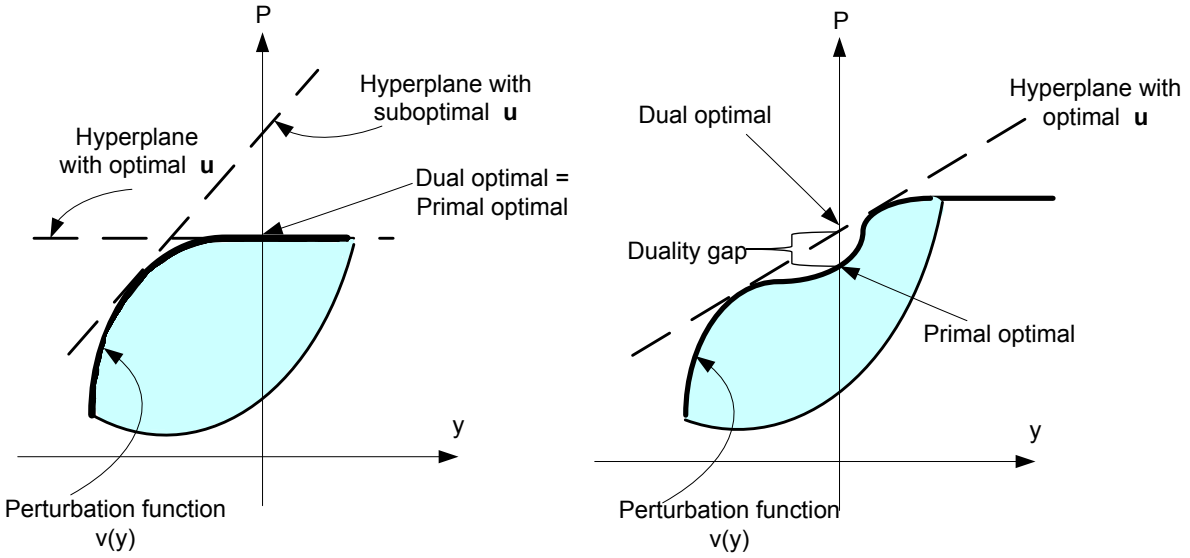
respectively. Besides, it can be shown that the eigenvalues of the Hessian matrix are given by

$$\lambda_1 \times \lambda_2 = \frac{s^2 + P^2 + N_T^2}{P^2\,N_T^2\ln^2(2)} \geq 0, \quad \lambda_3 = 0. \tag{F.3}$$

From (F.3), λ_1 and λ_2 must be either both positive or both negative. Therefore, by combining the above with $\mathrm{tr}\Big(\mathbf{H}(f)\Big) \leq 0$, it implies that $\lambda_1, \lambda_2 \leq 0$. Since $\lambda_n \leq 0,\ \forall n$, so $\mathbf{H}(f)$ is a negative semi-definite matrix and f is jointly concave with respect to P, s, and N_T. On the other hand, function t is a jointly concave function of P, s, and N_T so the concavity of function f is not destroyed by adding function f and function t. Therefore, the transformed objective function is jointly concave with respect to all the optimization variables.

Appendix G

Proof of Theorem 5.1



(a) Zero duality gap in maximizing a concave problem.

(b) Non-zero duality gap in maximizing a non-convex problem.

Figure G.1: Geometric interpretation of duality and perturbation function for concave and non-convex optimization problem with a 1-dimensional perturbation vector $y \in \mathbb{R}^1$ for illustration. The shaded areas represent the set of values of the primal problem under different perturbations $y \in \mathbb{R}^1$.

As mentioned in the main text of Chapter 5, the transformed optimization problem for a given parameter q, selected user set, and ZFBF transmission is a non-convex optimization problem due to constraint C3 in (5.9). In general, a non-zero duality gap exists if we solve the transformed problem by solving its dual, cf. Figure G.1(b). However, we demonstrate in the following that when a non-convex optimization problem satisfies certain conditions, the duality gap is always zero. Before introducing

another important theorem for the proof of Theorem 5.1, we first introduce the concept of *perturbation function*. For the sake of notational simplicity and to avoid ambiguity, we use the following notations. We use $v\langle\cdot\rangle$ to denote a function v and $(\cdot)$ to denote brackets. Without loss of generality, the primal optimization problem in (5.12) can be written in general form as

$$\begin{aligned} \mathrm{P} = \max_{\mathbf{p}_i^k \succeq \mathbf{0}} \quad & \sum_{i=1}^{n_F} f_i\langle \mathbf{p}_i^k \rangle \\ \text{s.t.} \quad & \sum_{i=1}^{n_F} \mathbf{g}_i\langle \mathbf{p}_i^k \rangle \preceq \mathbf{0}, \end{aligned} \tag{G.1}$$

where $f_i\langle\cdot\rangle : \mathbb{C}^K \to \mathbb{R}$ and $\mathbf{g}_i\langle\cdot\rangle : \mathbb{C}^K \to \mathbb{R}^L$ are arbitrary continuous functions. L and $\mathbf{0}$ are the total number of inequality constraints and a column vector with all zero elements, respectively. $\mathbf{p}_i^k \in \mathbb{R}^K$ represents a feasible solution vector of the primal problem in general form. Indeed, (G.1) is a general representation of an optimization problem and the physical meaning of vector $\mathbf{p}_i^k$ is not limited to transmit power. Note that we do not make any assumption on the concavity of functions $f_i\langle\cdot\rangle$ and $\mathbf{g}_i\langle\cdot\rangle$. Then, the perturbation function is defined as [144, 138]

$$\begin{aligned} v\langle \mathbf{y} \rangle = \max_{\mathbf{p}_i^k \succeq \mathbf{0}} \quad & \sum_{i=1}^{n_F} f_i\langle \mathbf{p}_i^k \rangle \\ \text{s.t.} \quad & \sum_{i=1}^{n_F} \mathbf{g}_i\langle \mathbf{p}_i^k \rangle \preceq \mathbf{y}, \end{aligned} \tag{G.2}$$

where $\mathbf{y} \in \mathbb{R}^L$ is a *perturbation vector*. A geometrical interpretation of the perturbation function is given in Figure G.1. The perturbation function $v\langle\mathbf{y}\rangle$ corresponds to the upper envelope of the shaded areas in Figure G.1. Note that $v\langle\mathbf{y}\rangle$ is a *non-decreasing* function of $\mathbf{y}$ since a larger value of each element in $\mathbf{y}$ results in a larger

feasible set. The Lagrangian function of (G.1) can be expressed as

$$\mathcal{L}\langle \mathbf{p}_i^k, \mathbf{u} \rangle = \sum_{i=1}^{n_F} f_i\langle \mathbf{p}_i^k \rangle - \mathbf{u}^T \Big(\mathbf{g}_i\langle \mathbf{p}_i^k \rangle \Big) \tag{G.3}$$

where $\mathbf{u} \in \mathbb{R}^L, \mathbf{u} \geq \mathbf{0}$ is a vector of Lagrange multipliers. Thus, the corresponding dual problem is given by

$$\mathrm{D} = \min_{\mathbf{u} \geq \mathbf{0}} \max_{\mathbf{p}_i^k} \mathcal{L}\langle \mathbf{p}_i^k, \mathbf{u} \rangle. \tag{G.4}$$

Indeed, from a geometrically point of view, the dual problem is equivalent to finding the slope $\mathbf{u}$ of the supporting hyperplane of the perturbation function at $v\langle \mathbf{0} \rangle$, such that its intercept on P-axis is minimal, cf. Figure G.1(a).

We are now ready to introduce the following theorem.

Theorem G.1 *If the perturbation function $v\langle \mathbf{y} \rangle$ is a concave function of $\mathbf{y}$, then the duality gap is zero despite the convexity of the primal problem*[27]*, i.e., $D = P$.*

Proof of Theorem G.1:

The main idea of the proof is based on [138, Theorem 6.2.7] which states that a zero duality gap is equivalent to the existence of a saddle point of the Lagrangian function. Before proceeding to the proof of Theorem G.1, let us first show how the concavity of the perturbation function $v\langle \mathbf{y} \rangle$ can be used to prove the existence of a saddle point of the Lagrangian function.

Suppose $v\langle \mathbf{y} \rangle$ is a concave function with respect to $\mathbf{y}$, then there exists a hyperplane that supports the hypograph of $v\langle \mathbf{y} \rangle$ for any $\mathbf{y} \in \mathbb{R}^L$, cf. Figure G.1. Thus, by

[27] It is obvious that if the primal problem is concave, then the perturbation function $v\langle \mathbf{y} \rangle$ will be a concave function of $\mathbf{y}$. However, the reverse is not necessarily true.

the definition of concavity [75, Chapter 3.1.3], there exists some vector $\tilde{\mathbf{u}}$ such that

$$v\langle \mathbf{y} \rangle \leq v\langle \mathbf{0} \rangle + \tilde{\mathbf{u}}^T(\mathbf{y} - \mathbf{0}), \tag{G.5}$$

where $\tilde{\mathbf{u}} \in \mathbb{R}^L$ is known as the sub-gradient of $v\langle \cdot \rangle$.

Next, let $\mathbf{p}_i^{k*}$ be the optimal solution of optimization problem (G.1). Then, $(\mathbf{p}_i^{k*}, \tilde{\mathbf{u}})$ is a saddle point of the Lagrangian function if $\tilde{\mathbf{u}} \geq \mathbf{0}$ and $(\mathbf{p}_i^{k*}, \tilde{\mathbf{u}})$ satisfies

$$\mathcal{L}\langle \mathbf{p}_i^{k*}, \mathbf{u} \rangle \geq \mathcal{L}\langle \mathbf{p}_i^{k*}, \tilde{\mathbf{u}} \rangle \geq \mathcal{L}\langle \mathbf{p}_i^{k}, \tilde{\mathbf{u}} \rangle, \forall\, \mathbf{p}_i^k \geq \mathbf{0}, \mathbf{u} \geq \mathbf{0}. \tag{G.6}$$

First, we prove $\tilde{\mathbf{u}} \geq \mathbf{0}$ by contradiction. Consider a vector $\boldsymbol{\Delta} \in \mathbb{R}^L, \boldsymbol{\Delta} \geq \mathbf{0}$. By the non-decreasing property of the permutation function, we have $v\langle \mathbf{y} + \boldsymbol{\Delta} \rangle \geq v\langle \mathbf{y} \rangle$. As a result, we obtain the following inequality

$$v\langle \mathbf{y} \rangle \leq v\langle \mathbf{y} + \boldsymbol{\Delta} \rangle \leq v\langle \mathbf{0} \rangle + \tilde{\mathbf{u}}^T(\mathbf{y} + \boldsymbol{\Delta} - \mathbf{0}) \tag{G.7}$$

which holds for arbitrary vectors $\boldsymbol{\Delta} \geq \mathbf{0}$ and $\mathbf{y}$. Now, we put $\mathbf{y} = \mathbf{0}$ into (G.7) which yields

$$v\langle \mathbf{0} \rangle \leq v\langle \mathbf{0} \rangle + \tilde{\mathbf{u}}^T \boldsymbol{\Delta}. \tag{G.8}$$

Suppose now, there exists one element in $\tilde{\mathbf{u}}$ which takes a negative value. Then, we can always choose a vector $\boldsymbol{\Delta}$ such that $\tilde{\mathbf{u}}^T\boldsymbol{\Delta} < 0$ which violates the inequality in (G.8). Thus, $\tilde{\mathbf{u}} \geq \mathbf{0}$ has to be true. Second, we prove $\tilde{\mathbf{u}}^T \sum_{i=1}^{n_F} \mathbf{g}_i \langle \mathbf{p}_i^{k*} \rangle = 0$. Again, we consider the hyperplane in (G.5) with input vector $\mathbf{y} = \sum_{i=1}^{n_F} \mathbf{g}_i \langle \mathbf{p}_i^{k*} \rangle$. Since $\mathbf{p}_i^{k*}$ is the optimal solution of the primal problem in (G.1), $\sum_{i=1}^{n_F} \mathbf{g}_i \langle \mathbf{p}_i^{k*} \rangle \leq \mathbf{0}$ must

hold. Therefore, $v\Big\langle \sum_{i=1}^{n_F} \mathbf{g}_i\langle \mathbf{p}_i^{k*}\rangle \Big\rangle = v\langle \mathbf{0}\rangle$ and $\tilde{\mathbf{u}}^T \sum_{i=1}^{n_F} \mathbf{g}_i\langle \mathbf{p}_i^{k*}\rangle = 0$ must be true for satisfying (G.5). Now, we are ready to prove the right hand side of (G.6). Let us first consider the following:

$$\begin{aligned} \mathcal{L}\langle \mathbf{p}_i^{k*}, \tilde{\mathbf{u}}\rangle &= \sum_{i=1}^{n_F} f_i\langle \mathbf{p}_i^{k*}\rangle - \tilde{\mathbf{u}}^T\Big(\mathbf{g}_i\langle \mathbf{p}_i^{k*}\rangle\Big) = \sum_{i=1}^{n_F} f_i\langle \mathbf{p}_i^{k*}\rangle \\ &= v\langle \mathbf{0}\rangle \geq v\langle \mathbf{y}\rangle - \tilde{\mathbf{u}}^T\mathbf{y},\ \forall \mathbf{y}. \end{aligned} \tag{G.9}$$

Suppose $\mathbf{p}_i^k$ is a feasible solution of the primal problem. Then, $\mathbf{p}_i^k$ is also a feasible solution of the perturbation function $\mathbf{v}\langle\cdot\rangle$ if we set the perturbation vector $\mathbf{y}$ such that $\mathbf{y} = \sum_{i=1}^{n_F} f\langle \mathbf{p}_i^k\rangle$. Then, we substitute $\mathbf{y} = \sum_{i=1}^{n_F} f\langle \mathbf{p}_i^k\rangle$ into (G.9) which yields

$$\begin{aligned} \mathcal{L}\langle \mathbf{p}_i^{k*}, \tilde{\mathbf{u}}\rangle = v\langle \mathbf{0}\rangle &\geq v\Big\langle \sum_{i=1}^{n_F} \mathbf{g}_i\langle \mathbf{p}_i^k\rangle \Big\rangle - \tilde{\mathbf{u}}^T \sum_{i=1}^{n_F} \mathbf{g}_i\langle \mathbf{p}_i^k\rangle \\ &\overset{(a)}{\geq} v\Big\langle \sum_{i=1}^{n_F} \mathbf{g}_i\langle \mathbf{p}_i^k\rangle \Big\rangle + \tilde{\mathbf{u}}^T \sum_{i=1}^{n_F} \mathbf{g}_i\langle \mathbf{p}_i^k\rangle = \mathcal{L}\langle \mathbf{p}_i^k, \tilde{\mathbf{u}}\rangle, \end{aligned} \tag{G.10}$$

where (a) is due to $\sum_{i=1}^{n_F} \mathbf{g}_i\langle \mathbf{p}_i^k\rangle \leq \mathbf{0}$.

On the other hand, the left hand side inequality in (G.6) can be proved as follows:

$$\begin{aligned} \mathcal{L}\langle \mathbf{p}_i^{k*}, \tilde{\mathbf{u}}\rangle &= \sum_{i=1}^{n_F} f_i\langle \mathbf{p}_i^{k*}\rangle - \tilde{\mathbf{u}}^T\Big(\mathbf{g}_i\langle \mathbf{p}_i^{k*}\rangle\Big) = \sum_{i=1}^{n_F} f_i\langle \mathbf{p}_i^{k*}\rangle \qquad (G.11) \\ &\leq \sum_{i=1}^{n_F} f_i\langle \mathbf{p}_i^{k*}\rangle - \mathbf{u}^T\Big(\mathbf{g}_i\langle \mathbf{p}_i^{k*}\rangle\Big) = \mathcal{L}\langle \mathbf{p}_i^{k*}, \mathbf{u}\rangle \quad \because \mathbf{g}_i\langle \mathbf{p}_i^{k*}\rangle \leq \mathbf{0}, \mathbf{u} \geq \mathbf{0}. \end{aligned}$$

Therefore, $\langle \mathbf{p}_i^{k*}, \tilde{\mathbf{u}}\rangle$ is a saddle point of the Lagrangian function and by [138, Theorem 6.2.5], the duality gap is zero. □

In other words, the concavity of the perturbation function $v\langle \mathbf{y}\rangle$ with respect to $\mathbf{y}$ is the key to prove that the duality gap is zero. The final step in proving Theorem

5.1 is to prove that $v\langle \mathbf{y} \rangle$ is a concave function of $\mathbf{y}$, i.e.,

$$v\langle \rho\mathbf{y} + (1-\rho)\mathbf{x} \rangle \geq \rho v\langle \mathbf{y} \rangle + (1-\rho) v\langle \mathbf{x} \rangle \tag{G.12}$$

for $0 \leq \rho \leq 1$, where $\mathbf{x} \in \mathbb{R}^L$ is another perturbation vector such that $\mathbf{x} - \mathbf{y} \neq \mathbf{0}$. Indeed, the concavity condition of the perturbation function is always satisfied in multi-carrier systems if frequency sharing is allowed. Suppose the spectrum of each subcarrier is flat due to a large number of subcarriers[28]. Then, we can increase the number of subcarriers for a fixed amount of total bandwidth $\mathcal{B}$ such that the original bandwidth of each subcarrier is divided into two portions, i.e., $(1-\rho)$ and ρ. Let $\mathbf{p}_{\mathbf{x}i}^{k*}$ and $\mathbf{p}_{\mathbf{y}i}^{k*}$ be the two optimal resource allocation policies with respect to the perturbation functions $v\langle \mathbf{x} \rangle$ and $v\langle \mathbf{y} \rangle$, respectively. Then, for the perturbation function $v\langle \rho\mathbf{y} + (1-\rho)\mathbf{x} \rangle$, by construction, we implement the resource allocation policies $\mathbf{p}_{\mathbf{x}i}^{k*}$ and $\mathbf{p}_{\mathbf{y}i}^{k*}$ in portion one and portion two, respectively. Then, the constraints associated with $v\langle \rho\mathbf{y} + (1-\rho)\mathbf{x} \rangle$ become a linear combination of the constraints in $v\langle \mathbf{x} \rangle$ and $v\langle \mathbf{y} \rangle$. On the other hand, the value of $v\langle \rho\mathbf{y} + (1-\rho)\mathbf{x} \rangle$ becomes a linear combination of $v\langle \mathbf{x} \rangle$ and $v\langle \mathbf{y} \rangle$ due to the flatness of the channel over neighbouring subcarriers. As a result, $v\langle \rho\mathbf{y} + (1-\rho)\mathbf{x} \rangle \geq \rho v\langle \mathbf{y} \rangle + (1-\rho) v\langle \mathbf{x} \rangle$ holds immediately[29] due to linearity. Thus, the frequency sharing condition always holds when $n_F \to \infty$ for a fixed amount of total bandwidth. In other words, the perturbation function is a concave function with respect to $\mathbf{y}$ under the frequency sharing condition.

So, by combining Theorem G.1 and the condition of $n_F \to \infty$, Theorem 5.1 is

[28] In practical systems such as LTE, the coherence bandwidth is in the order of 100 kHz [123] and the subcarrier spacing is in the order of 10 kHz. Thus, the spectrum of each subcarrier is virtually flat due to the highly correlated fading within each subcarrier.

[29] In [87], the frequency sharing condition was used to show the zero duality gap visually for multi-carrier systems. The proof provided in this chapter is more rigorous as it is based on showing the existence of a saddle point of the Lagrangian function by using the concavity of the perturbation function.

proved. □