ENERGY AWARE RESOURCE ALLOCATION IN GREEN WIRELESS NETWORKS

SHIXIN LUO

NATIONAL UNIVERSITY OF SINGAPORE

2015

ENERGY AWARE RESOURCE ALLOCATION IN GREEN WIRELESS NETWORKS

SHIXIN LUO

(*B. Eng., The Hong Kong Polytechnic University*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

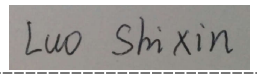
NATIONAL UNIVERSITY OF SINGAPORE

2015

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



Luo Shixin

August 5, 2015

Acknowledgement

I feel very fortunate to have had the opportunity to work with and learn from many remarkable people during my PhD. candidature at the National University of Singapore (NUS).

First of all, I would like to express my sincere gratitude and appreciation to my advisors Prof. Zhang Rui and Prof. Lim Teng Joon for their valuable guidance and helpful support throughout my work. Prof. Zhang has set an example of both a great scholar with the spirit of adventure and a responsible mentor. Prof. Lim, who constantly gives me his support and trust, is not only an erudite scholar, but also a very patient and considerate mentor. Had it not been for their advice, direction and encouragement, this thesis would certainly not have been possible.

I thank many of our group members and visitors with whom I had the best fortune to collaborate. My thanks also go to all the colleagues in the Communication and Network Laboratory at the Department of Electrical and Computer Engineering of NUS for their sincere help. I also thank all my old friends since childhood and my roommate in college for their friendship and encouragement.

Lastly, and most importantly, I would like to express my deepest gratitude to my beloved family members, my parents and my dear girlfriend J. Without their unconditional love, care and support, my studies at NUS would not be successful.

Table of Contents

Summary

Mobile and wireless data traffic is expected to increase many-fold from 2010 to 2020 at an astonishing rate. This many-fold increase in demand can only be met through a judicious combination of improvements in system performance and network infrastructure, which however have triggered fast escalation of overall network energy consumption. This thesis is devoted to investigating various energy efficient/saving communication strategies and their corresponding resource allocation optimizations in wireless networks.

This thesis starts with studying the dynamic adaptation of a base station's (BS's) transmit power levels and coverage area according to channel conditions and traffic load, in an orthogonal frequency-division multiple access (OFDMA) based broadcast channel. It is motivated by the observation that the traffic load in cellular networks exhibits significant fluctuations in both space and time, which can be exploited through cell range adaptation for energy saving. A power scaling law that relates the (short-term) average transmit power at BS with given cell range and mobile terminal (MT) density is first developed. Based on this result, we further derive BS's optimal (long-term) transmit adaptation policy by solving a joint range adaptation and long-term power control (including BS's on-off control) problem.

Then, we consider a similar orthogonal frequency-division multiplexing (OFDM) based multiuser wireless system with one BS and multiple MTs. In particular, the energy consumption at MTs is jointly considered with that of BS, since there is an increasing concern on MTs' limited power supply (e.g. battery). However, the transmitter- and receiver-side energy consumption cannot be minimized simultaneously in general. Thus, we aim to characterize the tradeoffs in minimizing

the BS's versus MTs' energy consumption by investigating a weighted-sum transmitter and receiver energy minimization (WSTREMin) problem. Moreover, we propose a new multiple access scheme, i.e., Time-Slotted OFDMA, to achieve the desired flexible energy consumption tradeoffs between the BS and MTs.

In the third part of this thesis, we study a more general setup with densely deployed access points (APs)[1] cooperatively serving distributed MTs, in the context of emerging cloud radio access network (C-RAN). Different from the previous two parts, both uplink (UL) and downlink (DL) communications are considered jointly. Moreover, APs are allowed to be switched into sleep mode for energy saving. These practical aspects are considered because UL transmission is becoming more crucial with the growing popularity of highly interactive applications, and the amount of energy consumed by a large number of active APs is considerable. To optimize the energy consumption tradeoffs between the active APs and MTs, we investigate the problem of joint DL and UL MT-AP association and beamforming design. By leveraging the celebrated UL-DL duality result and applying sparse optimization techniques, we propose efficient algorithms for joint DL and UL MT-AP association and beamforming design.

Lastly, the design of a wireless transmitter solely powered by means of energy harvesting (EH) from the environment (e.g. solar energy) is studied. We consider the basic point-to-point communication link with one EH transmitter and one constant-power receiver. Assuming a practical model with non-ideal energy storage efficiency and transmit circuit power, a *save-then-transmit* (ST) protocol is proposed to optimize the system outage performance via finding the optimal save-ratio (fraction of time devoted exclusively to EH instead of data transmission). Under this protocol, we characterize how the optimal save-ratio and the minimum outage probability vary with practical system parameters, and further compare the outage performance under random power supply versus constant power supply at the transmitter in fading channel.

[1] AP and BS are used interchangeably in this thesis.

List of Figures

List of Tables

List of Notations

Throughout this thesis, scalars are denoted by lower-case letters, vectors are denoted by bold-face lower-case letters, and matrices are denoted by bold-face upper-case letters. Also, we define the following symbols:

$\boldsymbol{M}^T$	the transpose of matrix $\boldsymbol{M}$
$\boldsymbol{M}^H$	the conjugate transpose of matrix $\boldsymbol{M}$
$\boldsymbol{M}^\dagger$	the pseudo-inverse of matrix $\boldsymbol{M}$
$\boldsymbol{I}_M$	the $M \times M$ identity matrix
$\mathbf{1}_M$	the $M \times 1$ vector with all elements being one
$\mathrm{Tr}(\cdot)$	the matrix trace operation
$\mathrm{rank}(\cdot)$	the matrix rank operation
$\lvert\boldsymbol{S}\rvert$	the determinant of a square matrix $\boldsymbol{S}$
$\boldsymbol{S}^{-1}$	the inverse of a square matrix $\boldsymbol{S}$
$\boldsymbol{S} \succeq (\preceq, \not\succeq)\ \mathbf{0}$	the square matrix $\boldsymbol{S}$ is positive (negative, non-positive) semidefinite
$\mathrm{diag}(\mathbf{x})$	the diagonal matrix with the diagonal elements being vector $\mathbf{x}$
$\lVert\mathbf{x}\rVert$	the Euclidean norm of a complex vector $\mathbf{x}$
$\boldsymbol{x} \preceq \boldsymbol{y}$	element wise inequality, i.e., $x_i \leq y_i,\ \forall i$
$\mathcal{CN}(\boldsymbol{x}, \boldsymbol{\Sigma})$	the distribution of a CSCG random vector with mean $\boldsymbol{x}$ and covariance $\boldsymbol{\Sigma}$
$\mathrm{E}[\cdot]$	the statistical expectation operator
$\max(x, y)$	the maximum element of x and y
$\min(x, y)$	the minimum element of x and y
$\lvert z\rvert$	the magnitude of a complex number z
$\triangleq$	defined as

List of Abbreviations

AP	Access Point
AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BS	Base Station
CDF	Cumulative Distribution Function
CoMP	Coordinated Multi-point
C-RAN	Cloud Radio Access Network
CSCG	Circularly Symmetric Complex Gaussian
CSI	Channel State Information
DL	Downlink
DPC	Dirty Paper Coding
EE	Energy Efficiency
EH	Energy Harvesting
ESD	Energy Storage Device
FDD	Frequency Division Duplex
GSO	Group Sparse Optimization
HPPP	Homogeneous Poisson Point Process
IC	Interference Channel
i.i.d	Independent and Identically Distributed
KKT	Karush-Kuhn-Tucker
LOS	Line-Of-Sight
MAC	Multiple Access Channel

MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MM	Majorization-Minimization
MT	Mobile Terminal
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PDF	Probability Density Function
PHY	Physical Layer
PMF	Probability Mass Function
QoS	Quality-of-Service
RIP	Relaxed-Integer Programming
RB	Resource Block
SE	Spectral Efficiency
SIMO	Single Input Multiple Output
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SOCP	Second Order Cone Programming
SON	Self Organised Network
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
UL	Uplink
WSR	Weighted Sum-Rate

List of Chapter Specific Abbreviations

Chapter 2

ARw/OFC	Adaptive Range with BS On-Off Control
ARw/oOFC	Adaptive Range without BS On-Off Control
DBS	Dynamic Bandwidth Sharing
EBS	Equal Bandwidth Sharing
ESM	Energy Saving Mechanism
FRw/OFC	Fixed Range with BS On-Off Control
FRw/oOFC	Fixed Range without BS On-Off Control
LTPC	Long-term Power Control
STPC	Short-term Power Control

Chapter 3

CCI	Channel Correlation Index
COG	Channel Orthogonality based Grouping
D-TDMA	Dynamic Time Division Multiple Access
TEMin	Transmitter-Side Energy Minimization
TS-OFDMA	Time-Slotted Orthogonal Frequency Division Multiple Access
WSREMin	Weighted-Sum Receiver-Side Energy Minimization
WSTREMin	Weighted-Sum Transmitter and Receiver Joint Energy Minimization

Chapter 4

APIRSS	AP Initiated Reference Signal Strength
HAP	High Power Access Point
LAP	Low Power Access Point
MUIRSS	MU Initiated Reference Signal Strength
PAw/oUL	Proposed Algorithm without Considering UL

Chapter 5

MESD	Main Energy Storage Device
SESD	Secondary Energy Storage Device
ST	Save then Transmit
TDMA-ST	TDMA based ST

List of Publications

Journal Papers

- S. Luo, R. Zhang, and T. J. Lim, "Optimal save-then-transmit protocol for energy harvesting wireless transmitters", *IEEE Transactions on Wireless Communications*, vol. 12, no. 3, pp. 1196-1207, Mar. 2013.
- S. Luo, R. Zhang, and T. J. Lim, "Optimal power and range adaptation for green broadcasting", *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4592-4603, Sep. 2013.
- S. Luo, R. Zhang, and T. J. Lim, "Joint transmitter and receiver energy minimization in multiuser OFDM systems", *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3504-3516, Oct. 2014.
- S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minimization through user association and beamforming in Cloud RAN", *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 494-508, Jan. 2015.
- S. Luo, J. Xu, T. J. Lim, and R. Zhang, "Capacity region of MISO broadcast channel for simultaneous wireless information and power transfer", *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3856-3868, Oct. 2015.

Conference Papers

- S. Luo, R. Zhang, and T. J. Lim, "Optimal save-then-transmit protocol for energy harvesting wireless transmitters", *IEEE International Symposium on*

Information Theory (ISIT), Boston, USA, 1-6 Jul. 2012.

- S. Luo, R. Zhang, and T. J. Lim, "Diversity analysis of an outage minimizing energy harvesting wireless protocol in rayleigh fading", *International Conference on Signal Processing and Communications (SPCOM)*, Bangalore, India, 22-25 Jul. 2012.

- S. Luo, R. Zhang, and T. J. Lim, "Optimal power and range adaptation for green broadcasting", *IEEE International Conference on Communication (ICC)*, Budapest, Hungary, 9-13 Jun. 2013.

- S. Luo, R. Zhang, and T. J. Lim, "Coordinated downlink and uplink user association and beamforming for energy minimization in cloud radio access network", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, 4-9 May 2014.

- S. Luo, J. Xu, T. J. Lim, and R. Zhang, "Capacity region of MISO broadcast channel with SWIPT", *IEEE International Conference on Communication (ICC)*, London, United Kingdom, 8-12 Jun. 2015.

Chapter 1

Introduction

According to CISCO's visual networking index (VNI) statistics, the global mobile data traffic in 2013 was nearly 18 times the size of the entire global Internet in 2000, which will further increase nearly 11-fold by 2018 [1]. This many-fold increase in demand can only be satisfied through a judicious combination of system performance improvement and network infrastructure upgrade, which however have triggered fast escalation of overall network energy consumption.

Consider the classic Shannon capacity formula [2]:

$$C = B \log_2 (1 + \text{SINR}) \tag{1.1}$$

where C denotes the capacity of a channel in bits/second, B denotes the channel bandwidth and SINR $= P/(I + N)$ denotes the signal-to-interference-plus-noise ratio with P, I and N representing the received signal, interference and noise power, respectively. From (1.1), it follows that higher capacity can be achieved by either allocating more bandwidth or improving the overall SINR. Since wireless spectrum is limited and scarce, improving SINR has been the major research drive during the past decades, which in general can be classified into two paradigms: 1) increase received signal power, P; and 2) reduce received interference power, I. These goals can be achieved through e.g., decreasing transmission distance by deploying more dense base stations (BSs); and applying advanced multiple access techniques such as OFDMA, respectively.

However, the technology advancement comes with substantially increased energy

consumption. It is reported that the total energy consumed by the infrastructure of cellular wireless networks, wired communication networks, and the Internet takes up more than 3 percent of the worldwide electric energy consumption nowadays [3]. As a result, energy-efficient operation has become an urgent need for wireless networks today. This thesis is devoted to investigating energy-efficient/saving communication techniques in wireless networks.

1.1 Common Terminology

In this section, we first introduce some important terminologies, which will be used throughout this thesis.

1.1.1 OFDM and OFDMA

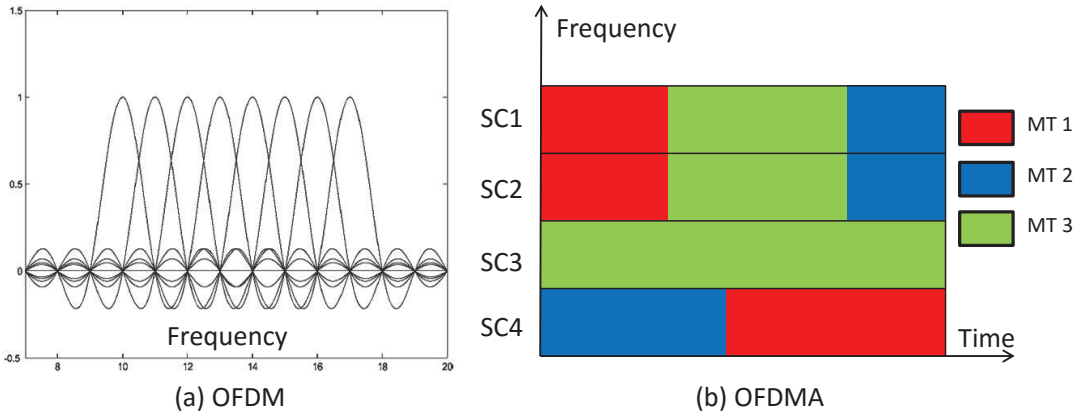


Figure 1.1: OFDM and OFDMA

Orthogonal frequency-division multiplexing (OFDM) is a digital multi-carrier modulation method, which divides each user's data stream into multiple substreams each transmitted over one of (a set of) orthogonal subcarriers (SCs) [4], as shown in Fig. 1.1(a). Each subcarrier is encoded with a conventional modulation scheme (such as quadrature amplitude modulation, QAM) at a low symbol rate, maintaining total data rate similar to that of single-carrier modulation schemes with the same bandwidth.

By employing the cyclic-prefix together with the multi-carrier modulation, OFDM is a robust transmission method that can mitigate the frequency-selective fading caused by multipath propagation inherent in the mobile environment. In practice, OFDM can be efficiently implemented by using fast fourier transform (FFT) algorithm on the receiver-side, and inverse FFT algorithm on the transmitter-side.

Orthogonal frequency-division multiple access (OFDMA) can be treated as a multiuser extension of OFDM. This multiple access scheme assigns orthogonal subcarriers to different users at one time, as shown in Fig. 1.1(b). With OFDMA, adaptive user-to-subcarrier assignment can be achieved, which makes it more flexible to allocate data rates among the users and thus leads to better system spectral efficiency.

1.1.2 Cloud Radio Access Network

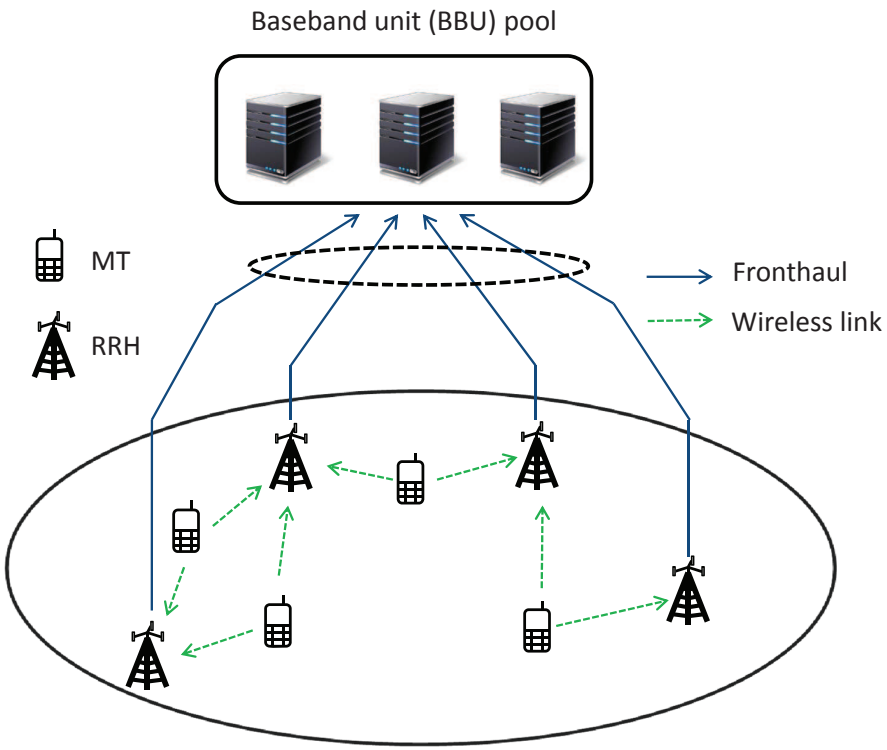


Figure 1.2: Simplified example of cloud radio access network (C-RAN)

Cloud radio access network (C-RAN) has recently been proposed as a promising

wireless network architecture to enable small-cell networks for more efficiently managing the interference and enhancing the network capacity [5]. In a C-RAN, the distributed transmission/reception points, called remote radio heads (RRHs), are connected to the baseband unit (BBU) pool through high-speed fronthaul links (fiber or wireless), as shown in Fig. 1.2. MTs can be cooperatively served by multiple RRHs, and each RRH merely forwards the signals to/from the BBU pool via its fronthaul link while leaving the complex joint encoding/decoding to the BBU pool. This network architecture enables centralized processing, collaborative transmission, and real-time cloud computing. As a result, significant rate improvement can be achieved due to reduced pathloss (resulting from the closer proximity of RRHs to MTs on average) along with joint scheduling and signal processing.

As the baseband processing is migrated to a BBU pool, the data exchanged between the RRHs and the BBU pool includes oversampled real-time digital signals with very high bit rates (in the order of Gbps). Consequently, the capacity requirement for the fronthaul links becomes far more challenging to meet in the C-RAN. Given finite fronthaul capacity, the optimal strategy for fronthaul compression and quantization has been studied recently in e.g., [6–8]. In this thesis, however, we focus on addressing the energy consumption issue by assuming that the fronthaul transport network is provisioned with sufficiently large capacity.

1.1.3 Energy Harvesting Powered Communication System

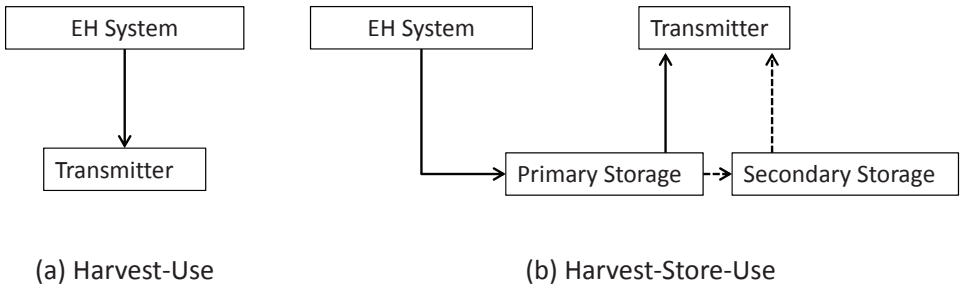


Figure 1.3: Energy harvesting system architecture with or without storage capability

The operation of communication networks powered either largely or exclusively by renewable energy sources has become increasingly attractive, due to the increased need to reduce energy consumption globally [9,10]. In general, energy harvesting (EH) based operation can be realized using one of the following two architectures [11]: (i) Harvest-Use (HU), energy is harvested for immediate use as in Fig. 1.3 (a); and (ii) Harvest-Store-Use (HSU), energy is harvested whenever possible and can be stored for future use as in Fig. 1.3 (b). For the case of HU, the harvested energy directly powers the wireless transmitter, and as a result the output power of the EH system needs to be above the minimum requirement for the operation of the device. For the case of HSU, two energy storage devices (e.g., battery and super-capacitor) are generally needed to store the harvested energy and power the wireless transmitter, respectively, in the same time. Energy storage is useful when the harvested energy does not need to be completely used, and the excess energy can thus be stored for future use. Therefore, the HSU scheme improves the energy efficiency and system performance over the HU counterpart in general.

1.2 Motivations

In this section, we present four important issues in the conventional wireless network design, which have not been properly addressed and can potentially be exploited for energy saving.

1.2.1 Spatial and Temporal Traffic Variation in Cellular Networks

Cell size and capacity are generally static at the phase of network planning, pertaining to the estimated value of peak traffic load. However, traffic load in cellular networks fluctuates substantially over both space and time due to mobility and traffic burstiness. For a cellular network in an urban area, the traffic load is relatively more heavy in workplaces than that in housing areas in the daytime; while the reverse is true during the night. Fig. 1.4 demonstrates the load of three representative cells from U.S. in densely populated urban areas of northern California, over a period of three weeks [12].

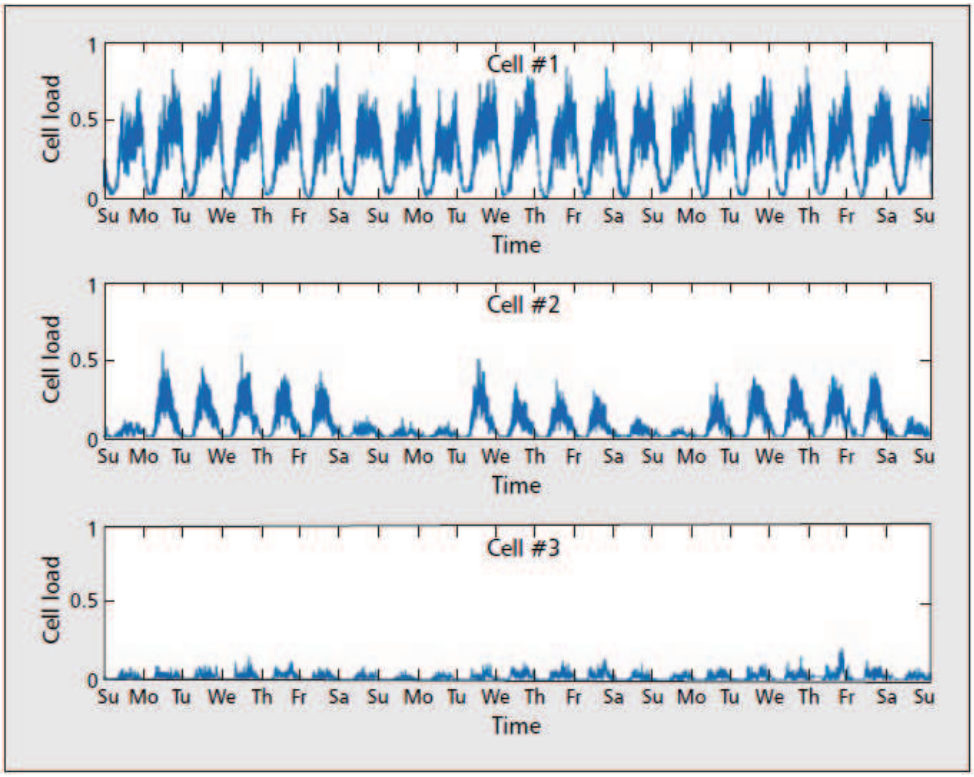


Figure 1.4: Normalized load of three different cell sectors over three weeks

The top cell has low load only at night, whereas the middle cell has low load during the weekends too. The bottom cell always has low load (i.e., during both day and night).

Due to the traffic fluctuation over both space and time, it is likely that some cells are under light load, and others are under heavy load, which suggests that the conventional static cell planning based on peak load is not optimal. Therefore, dynamic cell range adaptation and transmit power control according to channel conditions and traffic load has the potential to achieve energy saving.

1.2.2 Increasing Concern on Receiver-Side Energy Consumption

Another important issue, which has not been properly addressed in the traditional wireless network design, is the receiver-side energy consumption at MT/access point (AP) in downlink (DL)/uplink (UL) transmission. However, under many practical circumstances, it is indeed important to take the energy consumption at both the

transmitter and receiver into account. Considering DL transmission as an example, the limited battery capacity of MTs makes their energy consumption a more serious issue compared to that of AP/BS, which is generally connected to the grid with unlimited energy supply. Optimizing the user experience requires the design of efficient resource allocation schemes that can prolong the operation time of MTs by minimizing their energy usage.

1.2.3 Asymmetries between Downlink and Uplink Transmission

There exist various asymmetries between the DL and UL transmissions in terms of channel condition, traffic load and hardware limitations. As a result, DL or UL oriented design (usually from the DL perspective) may result in inefficient or even infeasible operation at the other end of the link. Consider the problem of MT-AP association as an example, if the decision is made solely based on DL transmission (e.g., the received signal power from AP to MT), the individual cell's coverage could be much larger than that obtained based on UL transmission (from MT to AP). This is because the AP is in general more capable (e.g., infinite energy supply and higher transmit power) than the MT. Therefore, the network operation optimization through joint DL and UL design can result in further performance improvement.

1.2.4 Random Energy Supply in Energy Harvesting Powered Communication System

Consider a simple point-to-point wireless communication system with an EH transmitter and a constant-power receiver, as shown in Fig. 1.5. The transmitter has two queues: the data queue where data packets are stored, and an energy queue where the harvested energy is stored (assuming the HSU architecture in Fig. 1.3). Since EH is intermittent in nature, it results in random energy arrival amount, i.e., E_{in}, in addition to the time-varying channel power, i.e., h. The randomness in E_{in} depends on the EH technology used (e.g., solar or wind energy).

The availability of an inexhaustible but unreliable energy source could change

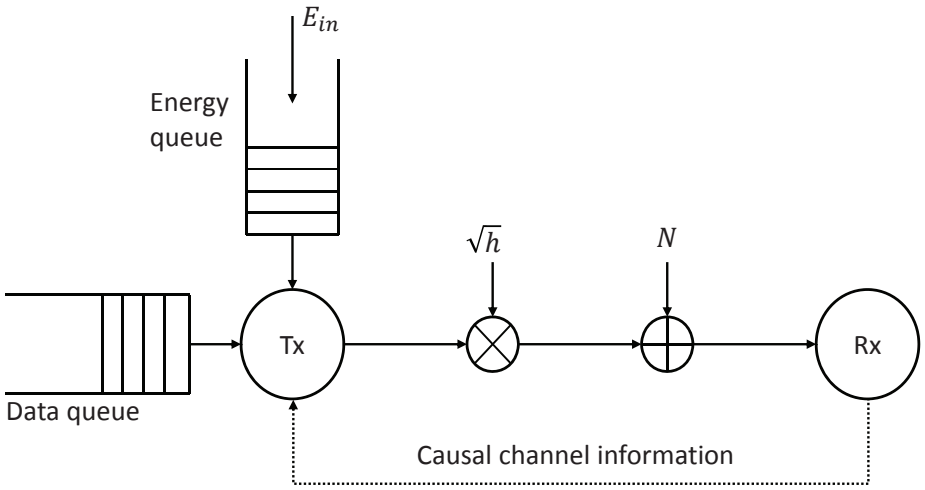


Figure 1.5: Point-to-point fading channel with an energy harvesting transmitter and channel state information (CSI) feedback from the receiver

the communication system design considerably. For example, the randomness in the energy supply makes it challenging to achieve smooth and continuous operation. Furthermore, a new type of transmitter-side energy constraint, namely EH constraint, which refers to that the energy accumulatively consumed up to any time instance cannot exceed that accumulatively harvested, is added to the transmission scheduling. As a result, existing designs and optimization strategies for conventional wireless systems assuming constant power supply are not applicable to an EH powered communication system.

1.3 Overview of the Thesis

Motivated by the above discussions, in this thesis, we investigate four energy-efficient and energy-saving communication problems via resource allocation optimization in wireless networks.

1.3.1 Dynamic Power and Range Adaptation

Chapter 2 of this thesis considers the energy efficient (green) broadcasting in an OFDMA based wireless network with one BS serving multiple randomly distributed MTs. The energy-saving approach studied is the adaptation of a BS's transmit power levels and coverage area according to channel conditions and traffic load. The BS's energy consumption includes both the transmit power and a constant power accounting for all non-transmission related power consumption (e.g., electronic hardware and air conditioning). Under this setup, we investigate short- and long-term BS's power control (termed STPC and LTPC, respectively) policies, where bandwidth is dynamically shared among a random number of MTs. STPC is a function of all MTs' channel gains to achieve the required user-level quality of service (QoS) at all time, while LTPC (including BS's on-off control) is a function of traffic density to minimize the long-term energy consumption at the BS under a minimum throughput constraint.

1.3.2 Joint Transmitter and Receiver Energy Minimization

In Chapter 3 of this thesis, we characterize the tradeoffs in minimizing the BS's versus MTs' energy consumption in multiuser OFDM based DL transmission by investigating a weighted-sum transmitter and receiver joint energy minimization (WSTREMin) problem, subject to the given transmission power constraint at the BS and data requirements of individual MTs. The proposed approach offers the flexibility of assigning different levels of importance to BS and MT power consumption, with the BS being connected to the grid and the MTs relying on batteries. We assume that each subcarrier (SC) can only be allocated to one MT at each time, but can be shared among different MTs over time, a channel allocation scheme that we refer to as SC time sharing. Under this scheme, we obtain the optimal transmission scheduling at the BS, which involves determining the time sharing factors and the transmit power allocations over the SCs for all MTs.

1.3.3 Joint Downlink and Uplink Energy Minimization

Chapter 4 further studies a more general cellular network setup with densely deployed APs cooperatively serving distributed MTs, in the context of emerging cloud radio access network (C-RAN). The total energy consumption in the C-RAN consists of the energy consumed by all APs and all MTs during both DL and UL communications. Under this setup, we study a joint DL and UL MT-AP association and beamforming design problem to minimize the total energy consumption in the network subject to given MTs' DL and UL QoS requirements. The energy saving is achieved by optimally assigning MTs to be served by the minimal subset of active APs, finding the power levels to transmit at all MTs and APs, as well as finding the beamforming vectors to use at the multi-antenna APs.

1.3.4 Save-Then-Transmit Protocol for Energy Harvesting Powered Wireless Transmitter

In Chapter 5, we turn to address the energy saving issue from the EH perspective by studying the design of a wireless communication system relying exclusively on EH. In particular, we consider the basic point-to-point communication link with one EH transmitter and one constant-power receiver. Assuming a practical model with non-ideal energy storage efficiency and transmit circuit power, a *save-then-transmit* (ST) protocol is proposed to optimize the system outage performance via finding the optimal save-ratio (fraction of time devoted exclusively to EH as opposed to data transmission). Important properties of the optimal save-ratio that minimizes outage probability are derived, from which useful design guidelines are drawn. In addition, we compare the outage performance of random power supply to that of constant power supply at the transmitter in the fading channel.

1.4 Major Contributions

The major contributions of this thesis are summarized as follows.

1.4.1 Key Insights from Dynamic Power and Range Adaptation

In Chapter 2, to design the optimal power and range adaptation of BS, we first develop a power scaling law that relates the (short-term) average transmit power at BS with the given cell range and MT density. Based on this result, we further derive the optimal (long-term) transmit adaptation policy by solving a joint range adaptation and LTPC problem. By applying the obtained optimal design, we show that energy saving at BS can be achieved through two major energy saving mechanisms (ESMs), i.e., range adaptation and BS's on-off power control. When the network throughput is low, BS's on-off power control is the most effective ESM, while when the network throughput increases, range adaptation becomes more effective.

1.4.2 Tradeoff between Transmitter and Receiver Energy Minimization

In Chapter 3, we formulate and solve a WSTREMin problem in the DL transmission of an OFDM based multiuser wireless system to characterize the tradeoffs in minimizing the BS's versus MTs' energy consumption. It is shown that Dynamic Time-Division-Multiple-Access (D-TDMA), where MTs are scheduled for single-user OFDM transmissions over orthogonal time slots, is the optimal transmission strategy for weighted-sum receiver-side energy minimization (WSREMin) at MTs; while OFDMA is optimal for transmitter-side energy minimization (TEMin) at the BS. As a hybrid of the two extreme cases, we further propose a new multiple access scheme, i.e., Time-Slotted OFDMA (TS-OFDMA) scheme, in which MTs are grouped into orthogonal time slots with OFDMA applied to users assigned within the same slot, to achieve more flexible energy consumption tradeoffs between the BS and MTs.

1.4.3 Framework for Joint Downlink and Uplink Energy Minimization

The proposed joint DL and UL MT-AP association and beamforming design, in Chapter 4, is unfortunately NP hard. Moreover, due to the new consideration of UL transmission, it is shown that the two state-of-the-art approaches, for finding computationally efficient solutions of joint MT-AP association and beamforming design solely from the DL perspective, i.e., group-sparse optimization (GSO) and relaxed-integer programming (RIP), cannot be modified in a straightforward way to solve our problem. Leveraging the celebrated UL-DL duality result, we show that by establishing a virtual DL transmission for the original UL transmission, the joint DL and UL optimization problem can be converted to an equivalent DL problem in C-RAN with two inter-related subproblems for the original and virtual DL transmissions, respectively. Based on this transformation, two efficient algorithms for joint DL and UL MT-AP association and beamforming design are proposed.

1.4.4 Random Power Supply versus Constant Power Supply

Based on the proposed ST protocol in Chapter 5, we minimize the outage probability when transmitting over a block fading channel with an arbitrary fading distribution. Furthermore, we compare the performance between two system setups: the (new) case with random power supply versus the (conventional) case with constant power supply at the transmitter, over the Rayleigh fading channel. It is shown that EH, which results in time-varying power availability in addition to the random channel fading, may severely degrade the outage performance. To be concrete, we further consider exponentially distributed random power, and show that although the diversity order with exponential power is the same as that with constant power in the Rayleigh fading channel, the outage probability curve may only display the slope predicted by this diversity analysis at substantially higher SNRs in the EH system as compared to in the conventional system.

Chapter 2

Optimal Power and Range Adaptation for Green Broadcasting

2.1 Introduction

In this chapter, we consider the dynamic adaptation of a BS's transmit power levels and coverage area according to channel conditions and traffic load in the DL transmission of an OFDMA based cellular network. Unlike traditional cellular networks using fixed time and/or bandwidth allocation, we consider that the available time-frequency transmission blocks are dynamically allocated to a random number of active MTs. Moreover, the BS is assumed to have two levels of power control: short-term power control (STPC) and long-term power control (LTPC), which correspond to the inherent difference in the time scales of the MTs' average channel gain variations (in e.g. seconds) and traffic density variations (in e.g. hours). STPC sets the transmit power based on each MT's distance from the BS to meet each MT's outage probability requirement over fading at all time, while LTPC (including BS's on-off control) is implemented according to traffic density variations such that the long-term energy consumption at the BS is minimized under a certain system-level throughput constraint. By focusing on two major energy saving mechanisms (ESMs), i.e., range adaptation and BS's on-off power control, we propose suboptimal schemes with various combinations of the two ESMs to investigate their impacts on system energy consumption.

2.2 Literature Review

The drive to make cellular networks more "green" mainly targets for BSs, since they account for a large proportion of the total energy consumed in the cellular network [13, 14] due to their operational units, e.g., processing circuits, air conditioner, besides radio transmission. Cell planning, i.e., placement of BSs and coverage area of each one, is usually based on estimated static (e.g. peak) traffic load. Current research in cellular network planning mainly focus on the BSs deployment design. For example, in [15], the authors used stochastic geometry to analyze the optimal macro/micro BS density for energy-efficient heterogeneous cellular networks with QoS constraints. The energy efficiency of heterogeneous networks and the effects of cell size on cell energy efficiency were investigated in [16] by introducing a new concept called area energy efficiency.

However, traffic load in cellular networks fluctuates substantially over both space and time due to mobility and traffic burstiness. Therefore, there will always be some cells under light load, and others under heavy load, which suggests that static cell planning based on peak load will not be optimal. Load balancing schemes have thus been proposed in both academia and industry [17–19], which react to load variations across time and cells by adaptively re-allocating users to cells. In [17], a network-wide utility maximization problem was considered to jointly optimize partial frequency reuse and load-balancing in a multi-cell network. In [18, 19], the authors proposed the "cell breathing" technique, which shrinks (or expands) the coverage of congested (or under-loaded) cells by reducing (or raising) the transmit power level, so that the load becomes more balanced.

In addition to load balancing, selectively letting some BSs be switched off according to traffic load can yield substantial energy saving. There have been a few BS on-off switching schemes introduced in the literature. For example, energy saving as a function of the daily traffic pattern, i.e the traffic intensity as a function of time, was derived in [20], where it is shown through simulations that energy saving on the order of $25 - 30\%$ is possible. Centralized and distributed BS reconfiguration

algorithms were proposed in [21], with simulations showing that the centralized algorithm outperforms the distributed one at the cost of increased complexity and overhead. In [22], the authors considered a wireless local area network (WLAN) consisting of a high density of APs. The resource on-demand (RoD) strategy was introduced to switch on or off WLAN APs dynamically, based on the volume and location of user demand.

When some BSs are switched off, their coverage areas need to be served by the remaining active BSs in the network. Such a self-organized network (SON) has been introduced in 3GPP LTE [23]. A similar but more flexible method called "Cell Zooming" was proposed in [24], which adaptively adjusts the cell size according to traffic load, user requirements, and channel conditions, in order to balance the traffic load in the network and thereby reduce energy consumption. Energy-efficient cellular network planning with consideration of BSs' ability of cell zooming, which is characterized as cell zooming ratio, was investigated in [25]. However, to the best of our knowledge, a scheme that adapts both coverage range and transmit power (including the possibility of switching off the BS) to minimize the total energy consumed has not been studied in the literature, even under the simple one-cell setup. This motivates our work in this chapter, which studies the extreme case of one single-cell system in order to obtain useful insights that could be applied in a general multi-cell environment.

2.3 System Model

We consider an OFDMA downlink in a given cell with bandwidth W Hz. It is assumed that the BS can adaptively adjust its cell coverage according to MT density and power budget through admission control. In this section, we first introduce a spatial model of cellular traffic based on MTs distributed according to a HPPP. Then, we elaborate on the proposed bandwidth sharing scheme for the OFDMA-based broadcast channel. Finally, we describe the STPC, based on which a power scaling law relating the (short-term) average transmit power at a BS given a pair of coverage range and MT

density is derived.

2.3.1 Traffic Model

The two-dimensional Poisson Point Process (PPP) has been used to model the spatial distribution of the randomly located MTs in a cellular network [26]. In this chapter, we assume that MTs form a HPPP Φ_m of density λ_m in the Euclidean plane[1]. Considering that every MT within the cell coverage requests connection (voice service or data application) randomly and independently with probability q (MTs have no knowledge about their surrounding wireless environment, and thus intend to transmit independently [28]). Then according to the Marking Theorem [29], the active MTs (that need to communicate with a BS) form another HPPP Φ of density λ,[2] where $\lambda = q\lambda_m$. Since we are interested in active MTs, we refer to active MTs simply as MTs in the rest of this chapter. The MT density λ is assumed to be a random variable with finite support, i.e., $0 \leq \lambda \leq \lambda_{\max}$, with $f_\lambda(\cdot)$ and $F_\lambda(\cdot)$ denoting its probability density function (PDF) and cumulative distribution function (CDF), respectively. Let $N \triangleq |\Phi(B)|$ represent the total number of MTs within a cell, denoted by B. Then N is a Poisson random variable with mean $\mu_N \triangleq \lambda\pi R^2$, where R denotes the cell radius, and probability mass function (PMF)

$$\Pr[N = n] = \frac{\mu_N^n}{n!} e^{-\mu_N}, \quad n = 0, 1, \ldots. \tag{2.1}$$

2.3.2 Bandwidth Sharing Model

Practically, dynamic bandwidth sharing (DBS) can be realized by users' time-sharing the available sub-carriers in OFDMA. To be more specific, the available time-frequency resource is divided into Resource Blocks (RBs) over both time and frequency, which are allocated among MTs such that each MT can be ideally assigned an effective bandwidth with arbitrary value from 0 to W Hz. Note that in general,

[1]The MTs density can be practically estimated in two steps [27]: 1) estimate the location and number of active MTs, which can be done by e.g., monitoring the time difference of arrival between signals from multiple BSs at the MT; 2) estimate the MTs density with a properly assumed statistical model.

[2]BS is assumed to support all MTs, within coverage, who request service.

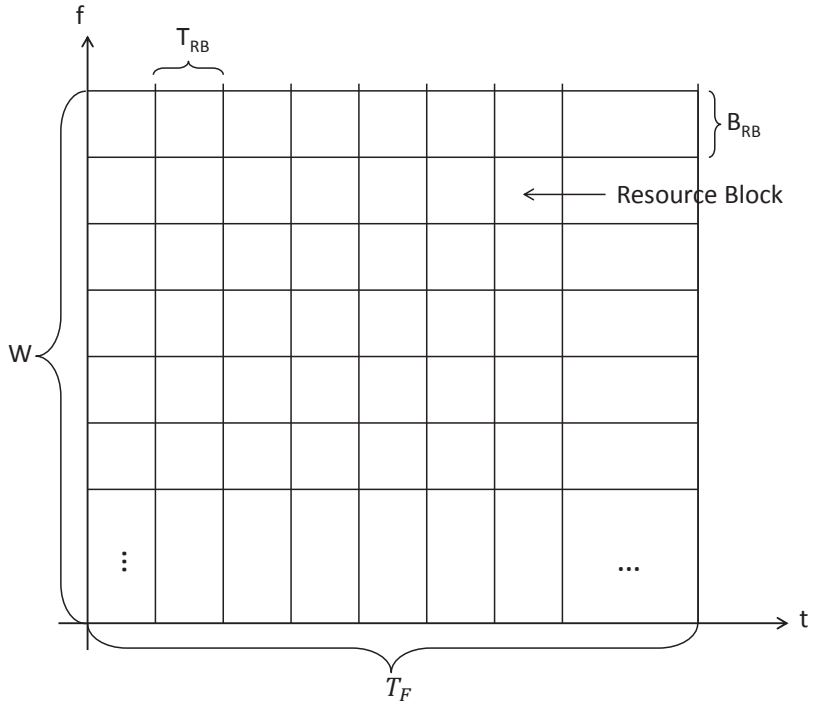


Figure 2.1: Equal bandwidth sharing (EBS)

DBS allocates the available RBs dynamically among MTs in order to optimize certain system-level utility (e.g. throughput) based on the number of MTs, their channels from the BS, and their QoS requirements. For the purpose of exposition, in this chapter we assume a simplified equal bandwidth sharing (EBS) scheme among MTs, i.e., the effective bandwidth allocated to MT i, $i = 1, 2..., N$, is W/N Hz.

An illustration of the EBS within a scheduled transmission frame T_F is shown in Fig. 2.1. The available time-frequency resource is divided into RBs with dimensions T_{RB} and B_{RB} over time and frequency, respectively. T_{RB} and B_{RB} are assumed to be much smaller than the channel coherence time, T_c, and the channel coherence bandwidth, B_c, respectively; thus a flat-fading channel can be assumed in each RB. Let $N_F = \frac{W}{B_{\text{RB}}}$ and $N_T = \frac{T_F}{T_{\text{RB}}}$ be the number of frequency slices and time slices, respectively, within a transmission frame. The total number of available RBs within one frame can be computed as $U = N_F N_T$, which is assumed to be large enough such that each MT can be assigned a continuous effective bandwidth $\frac{U_i}{U}W$, where U_i is the number of RBs allocated to MT i. For example, 4 RB's are allocated to MT i as shown

in Fig. 2.1. The total bandwidth allocated to MT i is therefore $\frac{4}{N_F}W$, over a period of N_T channel uses, where a channel use corresponds to T_{RB} seconds. Therefore, MT i is given $\frac{4W}{N_F N_T} = \frac{4W}{U}$ Hertz of bandwidth per channel use, which also implies that the BS is serving $N = \frac{U}{4}$ active MTs by EBS.

With EBS, the achievable rate for MT i, given received signal power S_i, is

$$V_i = \frac{W}{N} \log_2 \left(1 + \frac{NS_i}{\Gamma N_0 W}\right) \tag{2.2}$$

where Γ accounts for the gap from the channel capacity due to a practical coding and modulation scheme, and N_0 is the power spectral density of the additive white Gaussian noise (AWGN).

Suppose that channel coding is performed over L non-contiguous RBs allocated to a MT (c.f. Fig. 2.1 with $L = 4$). Then from (2.2), the average achievable rate of MT i over $L \geq 1$ RBs is given by [30]

$$\bar{V}_i = \frac{1}{L} \sum_{l=1}^{L} \frac{W}{N} \log_2 \left(1 + \frac{NS_{i,l}}{\Gamma N_0 W}\right) \tag{2.3}$$

where $S_{i,l}$ is the received signal power at the lth allocated RB, $l = 1, ..., L$, and $S_{i,l}$'s are independent over l due to independent channel fading if the L RBs allocated to a MT are sufficiently far apart in time and/or frequency.

2.3.3 Power Scaling Law

We assume a simplified channel model consisting of distance-dependent pathloss with path loss exponent $\alpha > 2$ and an additional random term accounting for short-term fading of the channel from the BS to each MT. With the assumed channel model, the received signal power for the lth RB of MT i is given by

$$S_{i,l} = \begin{cases} P_i h_{i,l} K \left(\frac{r_i}{r_0}\right)^{-\alpha} & \text{if } r_i \geq r_0 \\ P_i h_{i,l} K & \text{otherwise} \end{cases} \tag{2.4}$$

where r_i is a random variable representing the distance between MT i and BS, K is a constant equal to the pathloss at a reference distance r_0, $h_{i,l}$ is an exponential random variable with unit mean accounting for Rayleigh fading with $h_{i,l}$'s being independent

and identically distributed (i.i.d) over both i and l, and P_i is the transmit power for MT i, which is assumed to be identical for all l's since the realizations of $h_{i,l}$'s are not assumed to be known at BS. It is easy to verify that $S_{i,l}$'s are i.i.d over l as previously assumed.

To characterize the required minimum transmit power for MT i, P_i, outage performance is considered as the user-level QoS constraint. An outage event occurs when the link between MT i and BS cannot support a desired target rate $\bar{v}$ bits/sec, which is assumed to be equal for all MTs for simplicity. According to (2.3), the outage probability for MT i is given by

$$\mathrm{P}_{\text{out}}^{i} = \Pr\left\{\sum_{l=1}^{L} \frac{W}{N}\log_2\left(1+\frac{NS_{i,l}}{\Gamma N_0 W}\right) < L\bar{v}\right\}. \tag{2.5}$$

Since outage typically occurs when none of the L parallel channels can support the average rate $\bar{v}$ [30], (2.5) can be properly approximated as

$$\mathrm{P}_{\text{out}}^{i} \approx \prod_{l=1}^{L} \Pr\left\{\frac{W}{N}\log_2\left(1+\frac{NS_{i,l}}{\Gamma N_0 W}\right) < \bar{v}\right\}. \tag{2.6}$$

Since $S_{i,l}$'s are i.i.d over l as discussed at the beginning of this subsection, the outage probability above can be further expressed as

$$\mathrm{P}_{\text{out}}^{i} \approx \left(\Pr\left\{S_{i,1} < \frac{\Gamma N_0 W}{N}(2^{\frac{N\bar{v}}{W}}-1)\right\}\right)^{L}. \tag{2.7}$$

Given r_i, $S_{i,1}$ is an exponential random variable with mean $\bar{S}_{i,1}$, which is given by

$$\bar{S}_{i,1} = \begin{cases} P_i K\left(\dfrac{r_i}{r_0}\right)^{-\alpha} & \text{if } r_i \geq r_0 \\ P_i K & \text{otherwise.} \end{cases} \tag{2.8}$$

Thus, the outage probability for MT i given distance from BS r_i can be simplified as

$$\mathrm{P}_{\text{out}}^{i}(r_i) \approx \left[1-\exp\left(-\frac{\Gamma N_0 W}{N\bar{S}_{i,1}}(2^{\frac{N\bar{v}}{W}}-1)\right)\right]^{L}. \tag{2.9}$$

Let $\bar{\mathrm{P}}_{\text{out}}$ denote the maximum allowable outage probability for all MTs. Then the inequality

$$\mathrm{P}_{\text{out}}^{i} \leq \bar{\mathrm{P}}_{\text{out}} \tag{2.10}$$

needs to be maintained for all i's. From (2.8), (2.9) and (2.10), we can obtain P_i given r_i and N for the BS's STPC as[3]

$$P_i(r_i, N) = \begin{cases} \dfrac{\Gamma N_0 W}{K C_1} \cdot \dfrac{2^{NC_2} - 1}{N} \cdot \dfrac{r_i^\alpha}{r_0^\alpha} & \text{if } r_i \geq r_0 \\ \frac{\Gamma N_0 W}{K C_1} \cdot \frac{2^{NC_2 - 1}}{N} & \text{otherwise} \end{cases} \tag{2.11}$$

where $C_1 = -\ln(1 - \bar{\mathrm{P}}_{\text{out}}^{1/L})$ and $C_2 = \frac{\bar{v}}{W}$. With $P_i(r_i, N)$, the total transmit power P_t at the BS can be expressed as

$$P_t = \sum_{i=1}^{N} P_i(r_i, N). \tag{2.12}$$

Note that P_t is a random variable due to the randomness in the number of MTs, N, and their random distances from the BS, r_i's.

In this chapter, we assume that the BS can perform a slow LTPC based on the MT density variation, in addition to the more rapid STPC, for the purpose of minimizing the long-term energy consumption (more details will be given in Section 2.4). Considering the fluctuations of P_t given coverage range R and MT density λ, according to (2.12), a power scaling law that averages the random effects of the number of MTs and their locations is desired to facilitate the LTPC design to be studied in Section 2.4. This motivates us to find the (short-term) average transmit power $\bar{P}_t \triangleq \mathbb{E}[P_t]$ at BS for a given pair of R and λ, where the expectation is taken over N and r_i's.

The approach for finding $\bar{P}_t$ is to apply the law of iterated expectations, i.e.,

$$\bar{P}_t = \mathbb{E}_N \left[\mathbb{E}[P_t | N] \right] \tag{2.13}$$

where the inner expectation is taken over the random user locations given $N = n$ number of MTs, and the outer expectation is performed over the Poisson distributed N. This method works because $\mathbb{E}[P_t | N = n]$ in (2.13) can be obtained using the

[3]Note that several other quantities such as $\bar{V}_i$ and $\mathrm{P}_{\text{out}}^i$ are also dependent on N, but to simplify notation, we did not explicitly display this dependency when defining them. However, the manipulations of P_i to follow do involve N and therefore we write P_i as a function of r_i and N below.

following property of conditioned HPPP [29]:

$$\begin{aligned}\mathbb{E}[P_t|N=n] &= \mathbb{E}\left[\sum_{i=1}^{n} P_i(r_i,n)\right] \\ &= n\mathbb{E}[P_i(r_i,n)] \end{aligned} \tag{2.14}$$

where $P_i(r_i,n)$ represents the required transmit power from the BS to any MT i with distance r_i given that $N=n$ number of MTs equally share the total bandwidth W by EBS. It can be further verified that given $N=n$, MT i is uniformly distributed within a circular coverage area with radius R [29]. Thus, $\mathbb{E}[P_i(r_i,n)]$ is identical for all i's, and computed as

$$\mathbb{E}[P_i(r_i,n)] = \int_0^R P_i(r_i,n) f(r_i) dr_i \tag{2.15}$$

where $f(r_i) = \frac{2r_i}{R^2}$, $0 \le r_i \le R$, is the PDF of r_i.

Using (2.15) and averaging $\mathbb{E}[P_t|N=n]$ in (2.14) over the Poisson distribution of N, we obtain a closed-form expression for $\bar{P}_t$, which is given in the following theorem.

Theorem 2.3.1. *Consider an OFDMA-based broadcast channel, where the available bandwidth W Hz is equally shared among all MTs with STPC to support a target rate $\bar{v}$ bits/sec with outage constraint $\bar{P}_{out}$. Suppose that the channels from the BS to all MTs experience independent Rayleigh fading, then the transmit power at the BS averaged over MT population N and BS-MT distance r_i, given a coverage range R and a MT intensity λ, is approximated by*

$$\bar{P}_t(R,\lambda) = D_1 R^{\alpha}\left(2^{D_2 \pi \lambda R^2} - 1\right) \tag{2.16}$$

where $D_1 = \frac{2\Gamma N_0 W}{K(-\ln(1-\bar{P}_{out}^{1/L}))(\alpha+2) r_0^{\alpha}}$ and $D_2 = \frac{\bar{v}}{W}$ is the per-user spectrum efficiency in bps/Hz.

Proof. See Appendix A.1. □

Remark 2.3.1. Theorem 2.3.1 relates the average BS transmit power $\bar{P}_t$ with cell range R and MT density λ. Given R, $\bar{P}_t$ grows exponentially with increasing λ due to the reduced bandwidth equally allocated among (on average) $\mu_N = \lambda\pi R^2$ MTs. On

the other hand, given λ, besides the exponential increment in $\bar{P}_t$ with respect to R^2 due to the similar effect of per-user bandwidth reduction, there exists an extra polynomial term R^α in $\bar{P}_t$, due to the increased power consumption needed to compensate for more significant path loss with growing R. Since $\bar{P}_t$ is a strictly increasing function of both R and λ, to maintain a constant $\bar{P}_t$, R needs to be reduced when λ increases and vice versa. Theorem 2.3.1 therefore quantifies the relationship among BS transmit power, cell size and MT density, which enables the design of the (long-term) cell adaptation strategies introduced in the rest of this chapter.

2.4 Optimal Power and Range Adaptation

Power and range adaptation is the combined task of cell range adaptation and BS LTPC (including on-off control), which are both assumed to be performed on the time scale of MT density variation. Since MT's density variation is much slower as compared with MT's channel variation (which is taken care of by STPC studied in Section 2.3.3), LTPC is implemented over $\bar{P}_t$ given in (2.16) for the purpose of minimizing the BS's long-term energy consumption.

In this section, we first present a practical energy consumption model for BS by considering both transmission and non-transmission related power consumptions. Based on the presented energy consumption model, we study a joint cell range adaptation and LTPC problem to minimize the long-term power consumption at BS under a system-level throughput constraint.

2.4.1 Energy Consumption Model at Base Station

The energy consumption of a BS in general includes two parts: transmit power $\bar{P}_t$ and a constant power P_c accounting for all non-transmission related power consumption of e.g., electronic hardware and air conditioning. When the BS does not need to support any user, it can switch to a "sleep" mode [31], by turning off the power amplifier

to reduce energy consumption[4]. We note that the two cases of $R > 0$ and $R = 0$ correspond to "on" and "off (sleep)" modes of BS, respectively. A power consumption model for the BS is thus given by

$$\bar{P}_{\text{BS}}(R,\lambda) = \begin{cases} a\bar{P}_t(R,\lambda) + P_c, & R > 0 \\ P_{\text{sleep}}, & R = 0 \end{cases} \tag{2.17}$$

where $\bar{P}_{\text{BS}}(R,\lambda)$ represents the (short-term) average power consumption at BS given a pair of R and λ, P_{sleep} denotes the power consumed during the off mode, and $a \geq 1$ corresponds to the scaling of the actual power consumed with the radiated power due to amplifier and feeder losses. In practice, P_{sleep} is generally much smaller than P_c [14] and thus in this chapter, we assume $P_{\text{sleep}} = 0$ for simplicity. Since a is only a scaling constant, we further assume $a = 1$ in our subsequent analysis unless stated otherwise.

2.4.2 Optimal Cell Adaptation

According to (2.16), $\bar{P}_t(R,\lambda)$ is determined by R and λ. LTPC is thus equivalent to range adaptation over λ, i.e., by first finding the range adaptation function $R(\lambda)$ and then obtaining $\bar{P}_t(R,\lambda)$ as $\bar{P}_t(R(\lambda),\lambda)$, the LTPC policy $\bar{P}_{\text{BS}}(R(\lambda),\lambda)$ follows from (2.17). The joint cell range adaptation and LTPC problem can thus be formulated as

$$(\text{P0}): \quad \underset{R(\lambda)\geq 0}{\texttt{Min.}} \ \mathbb{E}_\lambda\left[\bar{P}_{\text{BS}}(R(\lambda),\lambda)\right] \tag{2.18}$$

$$\texttt{s.t.} \ \mathbb{E}_\lambda\left[U(R(\lambda),\lambda)\right] \geq U_{\text{avg}} \tag{2.19}$$

$$\bar{P}_{\text{BS}}(R(\lambda),\lambda) \leq P_{\max}, \ \ \forall\lambda \tag{2.20}$$

where $U(R(\lambda),\lambda) = \pi\lambda R^2(\lambda)$ corresponds to the (short-term) average number of supported MTs, U_{avg} represents the (long-term) system throughput[5] constraint, and

[4]Note that turning off BS is considered in the LTPC of this chapter. Since we focus on the extreme case of a one-cell system in this chapter, we assume that any uncovered spatial holes left by the single cell of our interest are to be filled by the surrounding active cells, which cause no interference to the considered cell by a proper frequency assignment scheme.

[5]Since a constant rate requirement $\bar{v}$ is assumed for all MTs and the effective system throughput equals to $\bar{v}U_{\text{avg}}(1-\bar{\text{P}}_{\text{out}})$, where $\bar{\text{P}}_{\text{out}}$ is a given outage probability target, the average number of supported MTs U_{avg} is an equivalent measure of the effective system throughput.

$P_{\max}$ is the (short-term) power constraint at BS. For convenience, in the rest of this chapter, $\bar{P}_t(R(\lambda),\lambda)$ and $\bar{P}_{\text{BS}}(R(\lambda),\lambda)$ are referred to as (short-term average) transmit power and power consumption at BS for a given λ, respectively, while $\mathbb{E}_\lambda\left[\bar{P}_t(R(\lambda),\lambda)\right]$ and $\mathbb{E}_\lambda\left[\bar{P}_{\text{BS}}(R(\lambda),\lambda)\right]$ are called the (long-term) average transmit power and average power consumption at BS, respectively.

Note that if choosing $R(\lambda)$ such that $\bar{P}_{\text{BS}}(R(\lambda),\lambda) = P_{\max}$ for all $\lambda > 0$ still leads to a violation of constraint (2.19), then Problem (P0) is infeasible. For analytical tractability, we only consider the case where U_{avg} yields a feasible (P0). (P0) is not convex due to the non-convexity of both the objective function (at $R = 0$) and the throughput constraint (2.19) since $U(R(\lambda),\lambda)$ is a non-concave function over $R(\lambda)$.

We start with reformulating (P0) via a change of variable: $x = R^2$, and making the constraint (2.20) implicit, which yields an equivalent problem

$$(\text{P1}): \quad \underset{x(\lambda)\in\mathcal{X}_a}{\texttt{Min.}} \quad \mathbb{E}_\lambda\left[\bar{P}_{\text{BS}}(x(\lambda),\lambda)\right] \tag{2.21}$$

$$\texttt{s.t.} \quad \mathbb{E}_\lambda\left[U(x(\lambda),\lambda)\right] \geq U_{\text{avg}} \tag{2.22}$$

where $\mathcal{X}_a \triangleq \left\{x(\lambda) : x(\lambda) \geq 0, \bar{P}_{\text{BS}}(x(\lambda),\lambda) \leq P_{\max}, \forall\lambda\right\}$. In (P1), the constraint (2.22) becomes convex since $U(x(\lambda),\lambda) = \pi\lambda x(\lambda)$ is affine over $x(\lambda)$. Furthermore, $\mathcal{X}_a$ is a convex set, and $\mathbb{E}_\lambda\left[\bar{P}_{\text{BS}}(x(\lambda),\lambda)\right]$ is the affine mapping of an infinite number of quasi-convex functions $\bar{P}_{\text{BS}}(x(\lambda),\lambda)$ and can be shown to be quasi-convex. Therefore, (P1) is a quasi-convex optimization problem and it can be verified that Lagrangian duality method can be applied to solve (P1) globally optimally [32]. The Lagrangian of Problem (P1) is

$$\mathcal{L}(x(\lambda),\mu) = \mathbb{E}_\lambda\left[\bar{P}_{\text{BS}}(x(\lambda),\lambda)\right] - \mu\left(\mathbb{E}_\lambda\left[U(x(\lambda),\lambda)\right] - U_{\text{avg}}\right) \tag{2.23}$$

where $\mu \geq 0$ is the dual variable associated with the throughput constraint (2.22). Then it can be shown that solving (P1) is equivalent to solving parallel subproblems all having the same structure and each for a different value of λ. For a particular λ, the associated subproblem is expressed as

$$\underset{x(\lambda)\in\mathcal{X}_a}{\texttt{Min.}} \quad L_\lambda(x(\lambda),\mu) \tag{2.24}$$

where $L_\lambda(x(\lambda),\mu) = \bar{P}_{\text{BS}}(x(\lambda),\lambda) - \mu U(x(\lambda),\lambda)$.

To tackle the non-continuity of $\bar{P}_{\text{BS}}(x(\lambda),\lambda)$ at $x(\lambda)=0$ (due to $P_c > P_{\text{sleep}} \triangleq 0$) and the power constraint $\bar{P}_{\text{BS}}(x(\lambda),\lambda) \le P_{\max}$, we first consider the case where BS is always on, i.e., $x(\lambda) > 0$ (thus, $\bar{P}_{\text{BS}}(x(\lambda),\lambda)$ is always differentiable) and there is no power constraint, i.e., $P_{\max} = +\infty$. The power constraint and the non-continuity at $x(\lambda)=0$ will be incorporated into the solution later without loss of optimality.

Denote $x_1^*(\lambda)$ and $x_2^*(\lambda)$ as the roots of the following two equations:

$$\frac{\partial L_\lambda(x(\lambda),\mu)}{\partial x(\lambda)} = 0, \;\; x(\lambda) > 0 \tag{2.25}$$

$$\bar{P}_{\text{BS}}(x(\lambda),\lambda) = P_{\max}, \tag{2.26}$$

respectively, where (2.25) is the optimality condition for $x(\lambda)$ in the case where BS is always on with infinite power budget and (2.26) gives the maximum coverage range due to finite $P_{\max}$ for any given λ. Note that it is difficult to obtain closed-form solutions for $x_1^*(\lambda)$ and $x_2^*(\lambda)$ due to the complex form of $\bar{P}_{\text{BS}}(x(\lambda),\lambda)$ in (2.17). However, since $\bar{P}_{\text{BS}}(x(\lambda),\lambda)$ is a strictly increasing function of $x(\lambda)$, and furthermore is convex in $x(\lambda)$ when $x(\lambda) > 0$, $x_1^*(\lambda)$ and $x_2^*(\lambda)$ can both be obtained numerically by a simple bisection search given μ and/or λ.

Let $x^*(\lambda)$ denote the optimal solution of Problem (2.24) with finite P_c and $P_{\max}$. Then $x^*(\lambda)$ has three possible values: $x_1^*(\lambda)$, $x_2^*(\lambda)$ and 0, where $x_2^*(\lambda)$ is taken when $x_1^*(\lambda)$ violates the power constraint of $P_{\max}$, i.e., $\bar{P}_{\text{BS}}(x_1^*(\lambda),\lambda) > P_{\max}$. In the case of $\bar{P}_{\text{BS}}(x_1^*(\lambda),\lambda) \le P_{\max}$, a comparison between $L_\lambda(x_1^*(\lambda),\mu)$ and $L_\lambda(0,\mu)=0$ is needed to tackle the non-continuity due to $P_c > 0$. If $L_\lambda(x_1^*(\lambda),\mu) < 0$, $x_1^*(\lambda)$ indeed gives the optimal solution; otherwise, we have $x^*(\lambda) = 0$ since it minimizes $L_\lambda(x(\lambda),\mu)$ over $x(\lambda) \ge 0$. On the other hand, if $\bar{P}_{\text{BS}}(x_1^*(\lambda),\lambda) > P_{\max}$, a similar comparison between $L_\lambda(x_2^*(\lambda),\mu)$ and $L_\lambda(0,\mu)=0$ is needed to verify the optimality between $x_2^*(\lambda)$ and 0. Thus, the signs of $L_\lambda(x_1^*(\lambda),\mu)$ and $L_\lambda(x_2^*(\lambda),\mu)$ as well as the value of $\bar{P}_{\text{BS}}(x_1^*(\lambda),\lambda)$

jointly determine $x^*(\lambda)$, as summarized below:

$$x^*(\lambda) = \begin{cases} x_1^*(\lambda) & \text{if } \bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda) \le P_{\max}, \\ & \quad L_\lambda(x_1^*(\lambda), \mu) < 0 \\ x_2^*(\lambda) & \text{if } \bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda) > P_{\max}, \\ & \quad L_\lambda(x_2^*(\lambda), \mu) < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.27}$$

To avoid checking the conditions in (2.27) for all λ's and gain more insights to the optimal power and range adaptation scheme, we proceed to characterize some critical values of λ, based on which the BS can determine $x^*(\lambda)$ with only the knowledge of the current density λ, through the following lemmas.

Lemma 2.4.1. *There exists λ_1, where $L_\lambda(x_1^*(\lambda_1), \mu) = 0$, such that $L_\lambda(x_1^*(\lambda), \mu)$ is positive for all $\lambda < \lambda_1$ and negative for all $\lambda > \lambda_1$.*

Proof. See Appendix A.2. □

Lemma 2.4.2. *$x_1^*(\lambda)$ is a strictly decreasing function of λ; $\bar{P}_{BS}(x_1^*(\lambda), \lambda)$ and $U(x_1^*(\lambda), \lambda)$ are all strictly increasing functions of λ.*

Proof. See Appendix A.3. □

Lemma 2.4.3. *$x_2^*(\lambda)$ is a strictly decreasing function of λ; $U(x_2^*(\lambda), \lambda)$ is a strictly increasing function of λ.*

Proof. The monotonicity of $x_2^*(\lambda)$ can be directly obtained from Remark 2.3.1. The proof for $U(x_2^*(\lambda), \lambda)$ is similar to that of Lemma 2.4.2, and is thus omitted for brevity. □

Since $\bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda)$ is a strictly increasing function of λ, there exists λ_2 with $\bar{P}_{\text{BS}}(x_1^*(\lambda_2), \lambda_2) = P_{\max}$, above which $\bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda) > P_{\max}$. Furthermore, since $U(x_2^*(\lambda), \lambda)$ strictly increases with λ, $L_\lambda(x_2^*(\lambda), \mu) = P_{\max} - \mu U(x_2^*(\lambda), \lambda)$ is thus a strictly decreasing function of λ and there exists λ_3 with $L_\lambda(x_2^*(\lambda_3), \mu) = 0$, such that $L_\lambda(x_2^*(\lambda), \mu) < 0$ for all $\lambda > \lambda_3$. Therefore, the conditions in (2.27) can be simplified as the inequalities among λ_1, λ_2 and λ_3, which is presented in the following theorem.

Theorem 2.4.1. *The optimal solution of Problem (P1) is given by*

- *If* $\lambda_2 \geq \lambda_1$

$$x^*(\lambda) = \begin{cases} 0 & \text{if } \lambda \leq \lambda_1 \\ x_1^*(\lambda) & \text{if } \lambda_1 < \lambda \leq \lambda_2 \\ x_2^*(\lambda) & \text{otherwise}. \end{cases} \tag{2.28}$$

- *If* $\lambda_2 < \lambda_1$

$$x^*(\lambda) = \begin{cases} 0 & \text{if } \lambda \leq \lambda_3 \\ x_2^*(\lambda) & \text{otherwise}. \end{cases} \tag{2.29}$$

Proof. See Appendix A.4. □

Note that Problem (P1) needs to be solved by iteratively solving $x^*(\lambda)$ with a fixed μ based on Theorem 2.4.1, and updating μ via the bisection search until the throughput constraint (2.22) is met with equality. The optimal solution of Problem (P0), $R^*(\lambda)$, can then be obtained as $R^*(\lambda) = \sqrt{x^*(\lambda)}$. From Theorem 2.4.1, Lemma 2.4.2 and Lemma 2.4.3, we obtain the following corollary.

Corollary 1. *$R^*(\lambda)$ and $U(R^*(\lambda), \lambda)$ are strictly decreasing and increasing functions of λ, respectively, if $R^*(\lambda) > 0$; $\bar{P}_{BS}(R^*(\lambda), \lambda)$ is a non-decreasing function of λ if $R^*(\lambda) > 0$.*

Proof. The proof directly follows from Lemmas 2.4.2 and 2.4.3, and thus is omitted for brevity. □

Next, we illustrate the optimal solution $R^*(\lambda)$ to Problem (P0) to gain more insights to the optimal cell adaptation scheme. It is observed that there exists a cut-off value of λ for each of the two cases in Theorem 2.4.1, below which the BS is switched off. This on-off behavior implies that allowing BS be switched off under light load is essentially optimal for energy saving. Since $x_2^*(\lambda)$ is the root of (2.26), which corresponds to the maximum coverage range with finite P_{max} for any given λ, it is worth noticing that when $\lambda_2 < \lambda_1$, constant power transmission with P_{max} is optimal. The reason is that when P_{max} is relatively small for the given throughput constraint

U_{avg}, BS has to transmit at its maximum power at all the "on" time. According to Corollary 1, the average number of supported MTs $U(x^*(\lambda), \lambda)$ strictly increases with λ. This is because that under the optimal scheme, BS should support more MTs when the density is larger to optimize energy-efficiency.

2.4.3 High Spectrum-Efficiency Regime

Although Theorem 2.4.1 reveals the structure of the optimal cell adaptation solution, which can be efficiently obtained numerically, the solution is expressed in terms of critical values of λ, namely λ_1, λ_2 and λ_3, for which closed-form expressions are difficult to be obtained. In this subsection, we obtain closed-form expressions of the solution in Theorem 2.4.1 under a high spectrum-efficiency assumption. It is observed from (2.16) that $D_2\pi\lambda R^2 = \frac{\bar{v}\pi\lambda R^2}{W} = \frac{\bar{v}\mu_N}{W}$, which can be interpreted as the average network throughput in bps divided by the total bandwidth, and is thus the system spectrum-efficiency in bps/Hz. Therefore, the high spectrum-efficiency assumption is equivalent to letting $D_2\pi\lambda R^2 \gg 1$. Under this condition, (2.16) in Theorem 2.3.1 can be simplified as

$$\bar{P}_t(R, \lambda) = D_1 R^\alpha 2^{D_2\pi\lambda R^2}. \tag{2.30}$$

Lemma 2.4.4. *Under the high spectrum-efficiency assumption of $D_2\pi\lambda R^2 \gg 1$, $x_1^*(\lambda)$ and $x_2^*(\lambda)$ in Theorem 2.4.1 are given by*

$$x_1^*(\lambda) = \frac{\alpha}{2D_3\pi\lambda}\mathcal{W}\left(\frac{2D_3\pi\lambda}{\alpha}\left(\frac{\mu}{D_1 D_3}\right)^{\frac{2}{\alpha}}\right) \tag{2.31}$$

$$x_2^*(\lambda) = \frac{\alpha}{2D_3\pi\lambda}\mathcal{W}\left(\frac{2D_3\pi\lambda}{\alpha}\left(\frac{P_{max}^t}{D_1}\right)^{\frac{2}{\alpha}}\right) \tag{2.32}$$

where $D_3 = (\ln 2)D_2$, $P_{max}^t = P_{max} - P_c$, and $\mathcal{W}(\cdot)$ is the Lambert W function defined as $y = \mathcal{W}(y)e^{\mathcal{W}(y)}$ [33].

Proof. See Appendix A.5. □

The accuracy of the above high spectrum-efficiency approximation will be verified by numerical results in Section 2.6. With (2.31) and (2.32),

closed-form expressions of $U(x_1^*(\lambda),\lambda)$, $U(x_2^*(\lambda),\lambda)$ and $\bar{P}_{\text{BS}}(x_1^*(\lambda),\lambda)$ under the high spectrum-efficiency assumption can be easily obtained, which can be verified to preserve the properties given in Lemmas 2.4.1-2.4.3 by using properties of the Lambert W function. For brevity, we omit the details here.

Moreover, we obtain the following corollary from Lemma 2.4.4.

Corollary 2. *Under the high spectrum-efficiency assumption of $D_2\pi\lambda R^2 \gg 1$, λ_1, λ_2 and λ_3 in Theorem 2.4.1 are given by*

$$\lambda_1 = \left(\frac{1}{\pi D_3} + \frac{P_c}{\mu\pi}\right)\left(\frac{D_1 D_3}{\mu}\right)^{\frac{2}{\alpha}} \exp\left(\frac{2}{\alpha} + \frac{2D_3 P_c}{\mu\alpha}\right) \tag{2.33}$$

$$\lambda_2 = \frac{\alpha P_{max}^t}{2\pi(\mu - D_3 P_{max}^t)}\left(\frac{D_1 D_3}{\mu}\right)^{\frac{2}{\alpha}} \exp\left(\frac{D_3 P_{max}^t}{\mu - D_3 P_{max}^t}\right) \tag{2.34}$$

$$\lambda_3 = \frac{P_{max}}{\mu\pi}\left(\frac{D_1}{P_{max}^t}\right)^{\frac{2}{\alpha}} \exp\left(\frac{2D_3 P_{max}}{\mu\alpha}\right). \tag{2.35}$$

Proof. The proof is similar to that of Lemma 2.4.4, and thus omitted for brevity. □

Remark 2.4.1. λ_1, λ_2 and λ_3 in Corollary 2 can be verified to be all strictly decreasing functions of the dual variable μ as follows. Let μ^* be the optimal dual solution of Problem (P1), λ_1^*, λ_2^* and λ_3^* be the corresponding critical values of λ when $\mu = \mu^*$. Since μ^* strictly increases as the throughput constraint U_{avg} increases, it follows from (2.33)-(2.35) that λ_1^*, λ_2^* and λ_3^* are all strictly decreasing functions of U_{avg}. Since in Theorem 2.4.1, λ_1 and λ_3 are the thresholds of the MT density above which BS switches from off to on mode, their decrease with increasing U_{avg} implies that BS needs to be stay on for more time if large system throughput is required.

2.5 Suboptimal Schemes

The optimal power and range adaptation policy presented in Section 2.4 combines cell range adaptation and BS LTPC (including on-off control), suggesting that the energy saving at BS essentially comes from two major energy saving mechanisms (ESMs): range adaptation and BS on-off control. In this section, we propose four suboptimal schemes, which can be considered as suboptimal solutions of (P0) with

various combinations of these two ESMs, to investigate their effects on the system energy consumption.

1. **Fixed range with BS on-off control (FRw/OFC)**: In this scheme, BS is switched off when MT density is lower than a cutoff value λ_c, while the coverage range R is fixed as R_f whenever BS is on. For a given λ_c, since from (2.16) the BS transmission power is a strictly increasing function of R, R_f should be chosen as the minimum value, denoted by $R_f(\lambda_c)$, to satisfy the throughput constraint U_{avg} by applying BS power control with fixed coverage based on λ according to (2.16). Furthermore, λ_c should be optimized to minimize the average BS power (including both transmission and non-transmission related portions) consumption. The optimal cutoff value λ_c^* and its corresponding coverage range $R_f(\lambda_c^*)$ can be found via solving Problem (P0) by assuming the following (suboptimal) range adaptation policy:

$$R(\lambda) = \begin{cases} R_f(\lambda_c) & \text{if } \lambda \geq \lambda_c \\ 0 & \text{otherwise.} \end{cases} \tag{2.36}$$

Specifically, we have

$$\lambda_c^* = \arg \min_{\lambda_c < \lambda_{\max}} \mathbb{E}_{\lambda_c} \left[\bar{P}_{\text{BS}}(R_f(\lambda_c), \lambda) \right] \tag{2.37}$$

where

$$\begin{aligned} R_f(\lambda_c) = \min.\ & R_f \\ \text{s.t. } & \mathbb{E}_{\lambda_c} \left[U(R_f, \lambda) \right] \geq U_{\text{avg}} \\ & \bar{P}_{\text{BS}}(R_f, \lambda) \leq P_{\max}, \forall \lambda \geq \lambda_c \end{aligned} \tag{2.38}$$

where $\mathbb{E}_{\lambda_c} \left[f(\lambda) \right] \triangleq \mathbb{E}_{\lambda} \left[f(\lambda) | \, \lambda \geq \lambda_c \right] \Pr \{ \lambda \geq \lambda_c \}$. For a given λ_c, since $\mathbb{E}_{\lambda_c} \left[U(R_f, \lambda) \right]$ is a strictly increasing function of R_f, Problem (2.38) can be solved efficiently through the bisection search. Then, the optimal cut-off threshold in (2.37) can be found by a line search over $[0, \lambda_{\max}]$.

2. **Fixed range without BS on-off control (FRw/oOFC)**: In this scheme, BS is not allowed to be switched off during operation. The coverage range is fixed as R_f,

which is chosen as the minimum value of R to satisfy the throughput constraint U_{avg} by applying BS power control only based on λ according to (2.16). Note that FRw/oOFC can be treated as a special case of FRw/OFC with λ_c in (2.36) set to be 0. Thus, the fixed coverage R_f can be directly determined by solving Problem (2.38) with $\lambda_c = 0$.

3. **Adaptive range with BS on-off control (ARw/OFC)**: In this scheme, BS is switched off when MT density is lower than a cutoff value λ_c, while BS transmits with constant power $P_f - P_c$ whenever it is powered on by applying range adaptation only based on λ according to (2.16). Given P_f, the corresponding λ_c is chosen as the maximum value of λ, denoted by $\lambda_c(P_f)$, to satisfy the throughput constraint U_{avg}, in order to minimize the BS average power consumption $\mathbb{E}_{\lambda_c(P_f)}[P_f]$; P_f is then optimized to further minimize the average power consumption at BS. The optimal transmit power $P_f^* - P_c$ and its corresponding cutoff value $\lambda_c(P_f^*)$ can be obtained via solving Problem (P0) by assuming the following (suboptimal) range adaptation policy:

$$R(\lambda) = \begin{cases} \bar{P}_{\text{BS}}^{-1}(P_f, \lambda) & \text{if } \lambda \geq \lambda_c(P_f) \\ 0 & \text{otherwise} \end{cases} \tag{2.39}$$

where $\bar{P}_{\text{BS}}^{-1}(P_f, \lambda)$ is the inverse function of (2.17) which computes the coverage range with given BS power consumption P_f and MT density λ. Specifically, we have

$$P_f^* = \arg \min_{P_f \leq P_{\max}} \mathbb{E}_{\lambda_c(P_f)}[P_f] \tag{2.40}$$

where

$$\begin{aligned} \lambda_c(P_f) = \max.\ & \lambda_c \\ \text{s.t. } & \mathbb{E}_{\lambda_c}[U(R(\lambda), \lambda)] \geq U_{\text{avg}} \\ & \bar{P}_{\text{BS}}(R(\lambda), \lambda) = P_f, \forall \lambda \geq \lambda_c. \end{aligned} \tag{2.41}$$

Note that from (2.39) and Remark 2.3.1, $R(\lambda)$ increases strictly with P_f given λ, $U(R(\lambda), \lambda) = \pi\lambda R^2(\lambda)$ is thus a strictly increasing function of P_f. Therefore,

Problem (2.41) can be solved efficiently through the bisection search. Then, the optimal constant BS power consumption in (2.40) can be found by a line search over $[0, P_{\max}]$.

4. **Adaptive range without BS on-off control (ARw/oOFC)**: In this scheme, BS transmits with constant power $P_f - P_c$ and is not allowed to be switched off during operation, i.e., no BS power control is applied. The constant transmit power $P_f - P_c$ is chosen as the minimum value to satisfy the throughput constraint U_{avg} by applying range adaptation only based on λ according to (2.16). Note that ARw/oOFC is a special case of ARw/OFC with λ_c in (2.39) set to be 0. Thus, P_f can be obtained by solving Problem (2.40) with $\lambda_c = 0$.

The suboptimal schemes presented above all yield feasible and in general suboptimal solutions of Problem (P0). In particular, FRw/OFC and ARw/oOFC apply only BS power control (including on-off control) and only range adaptation, respectively; ARw/OFC applies both BS on-off control and range adaptation, while FRw/oOFC does not apply any of them for lowest complexity. By comparing the performance of these suboptimal schemes with the optimal scheme presented in Section 2.4, we can investigate the effect of each individual ESM, namely, BS power control and range adaptation on the BS energy saving, as will be shown in the next section through numerical examples.

2.6 Numerical Results

To obtain numerical results, we assume a time-varying traffic density with PDF: $f(\lambda) = \frac{4\lambda}{\lambda_{\max}^2},\ 0 \le \lambda \le \frac{\lambda_{\max}}{2}$; $f(\lambda) = \frac{4}{\lambda_{\max}} - \frac{4\lambda}{\lambda_{\max}^2},\ \frac{\lambda_{\max}}{2} < \lambda \le \lambda_{\max}$, where $\lambda_{\max} = 1 \times 10^{-4}$ MTs/m^2 is the peak traffic load. We consider pathloss and Rayleigh fading for channels between BS and MTs, where the pathloss exponent α is 3 and the outage probability threshold $\bar{\text{P}}_{\text{out}}$ is 10^{-3}. The bandwidth W and the rate requirement $\bar{v}$ of each MT are set to be 5 MHz and 150 kbits/sec, respectively, if not specified otherwise [24]. We also set a short-term power constraint at BS as $P_{\max} = 160$ W.

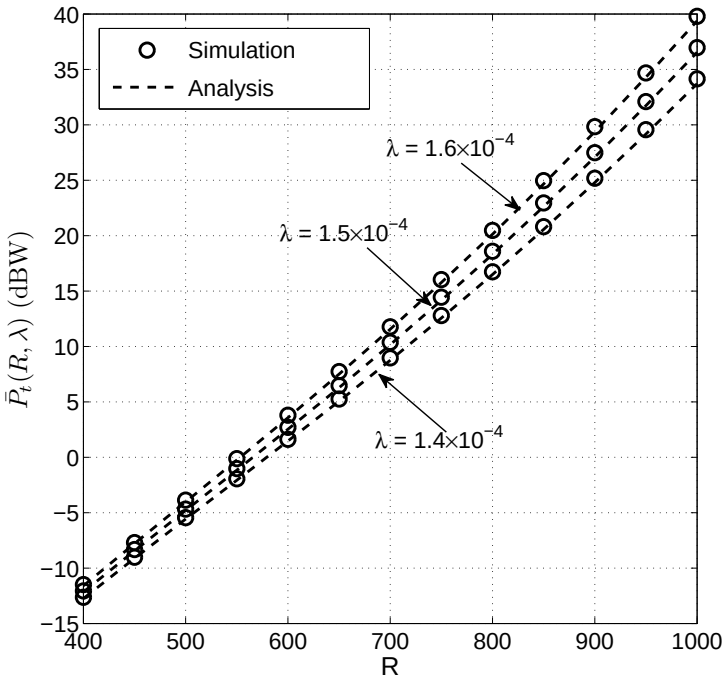


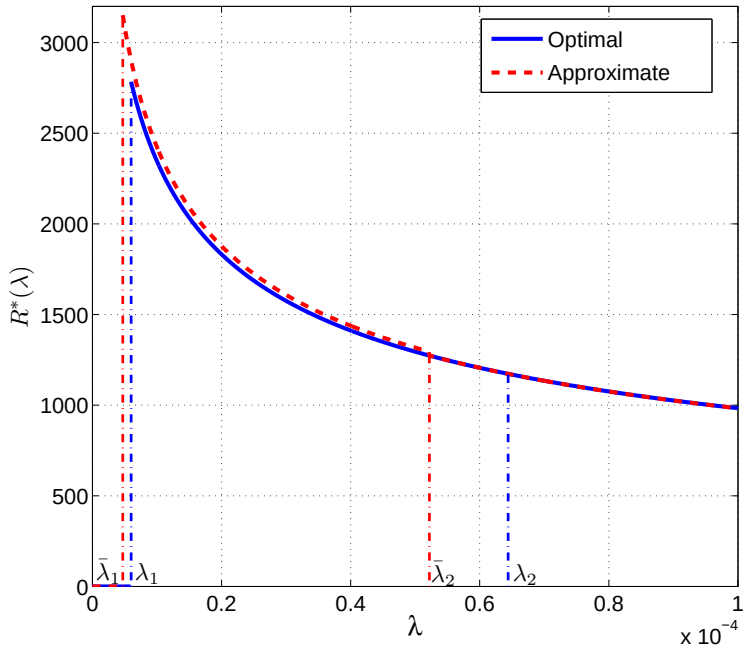
Figure 2.2: Average transmit power $\bar{P}_t(R, \lambda)$ in Theorem 2.3.1

Other parameters are set as $\Gamma = 1$, $N_0 = -174$ dBm/Hz, $r_0 = 10$ m, and $K = -60$ dB. We conduct the simulations by using Matlab on a computer equipped with an Intel Core i5-2500 @3.3GHz processor and 8GB of RAM memory. With the assumed setup, the solution of the optimal range and power adaptation can be obtained within seconds.

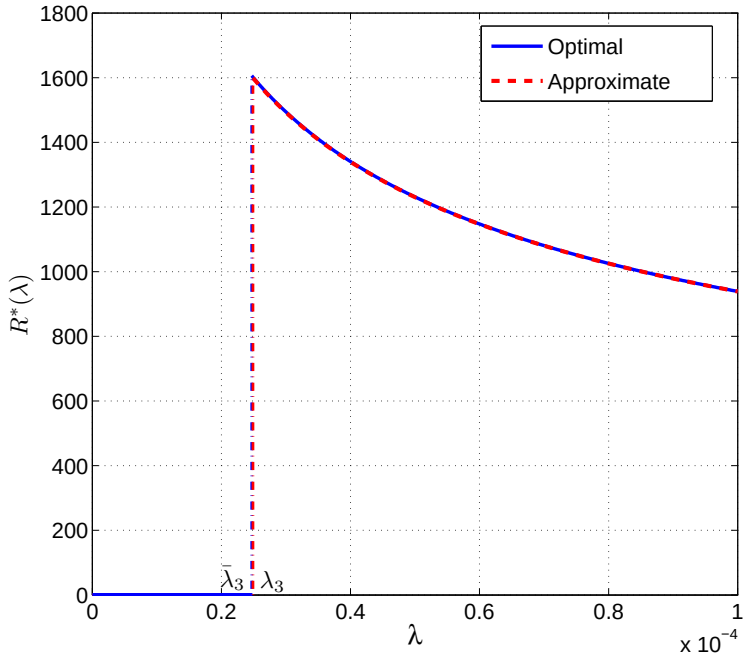
Fig. 2.2 verifies the power scaling law in Theorem 2.3.1. For a given MT density λ, it is observed that the simulation results match well with our analytical result in (2.16).

Fig. 2.3(a) and Fig. 2.3(b) show the optimal range adaptation in Theorem 2.4.1 and the approximate range adaptation in Lemma 2.4.4 under the high spectrum-efficiency assumption as functions of MT density, i.e., $R^*(\lambda) = \sqrt{x^*(\lambda)}$, for the two cases of $\lambda_2 \geq \lambda_1$ and $\lambda_2 < \lambda_1$, respectively. Fig. 2.4 and Fig. 2.5 show the corresponding optimal BS power adaptation and the resulting system throughput (in terms of average number of supported MTs), respectively[6]. For Fig. 2.3(a), Fig.

[6]Since the results by the approximate range adaptation are almost no different from those in Figs. 2.4 and 2.5, we do not show them in these two figures for brevity.

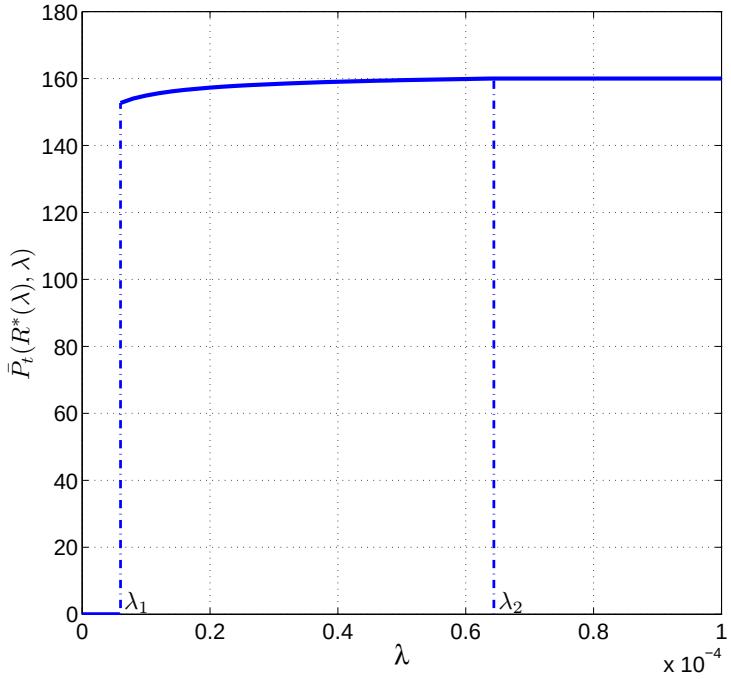


(a)

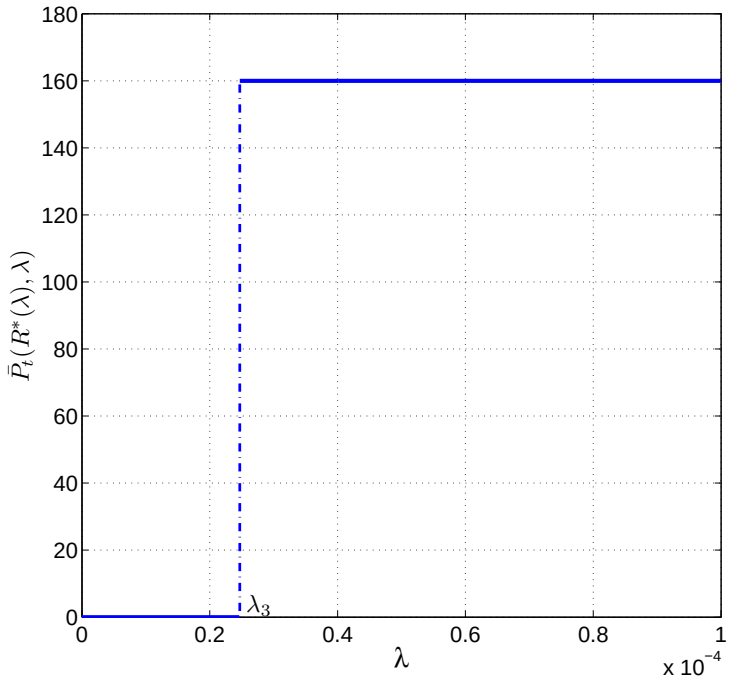


(b)

Figure 2.3: Optimal and approximate cell range adaptation v.s. MT density: (a) $\lambda_2 \geq \lambda_1$; (b) $\lambda_2 < \lambda_1$

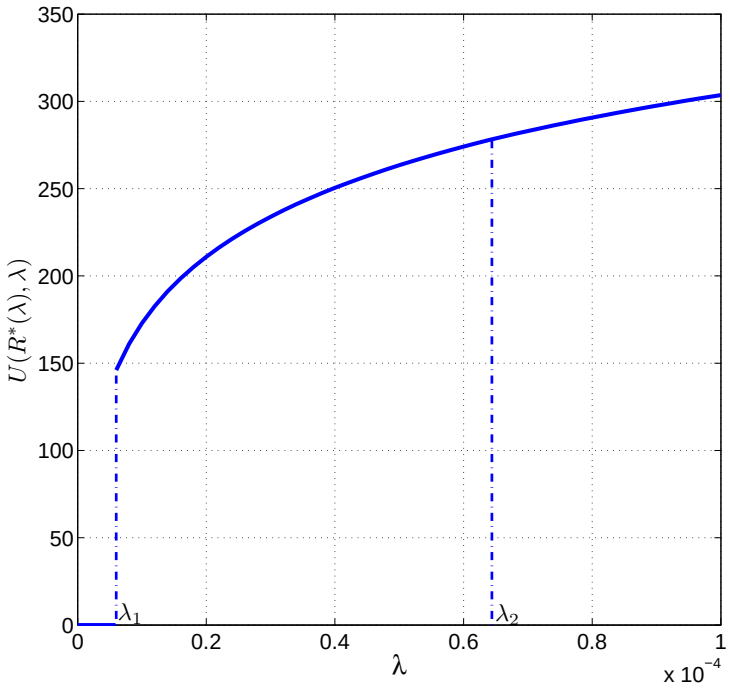


(a)

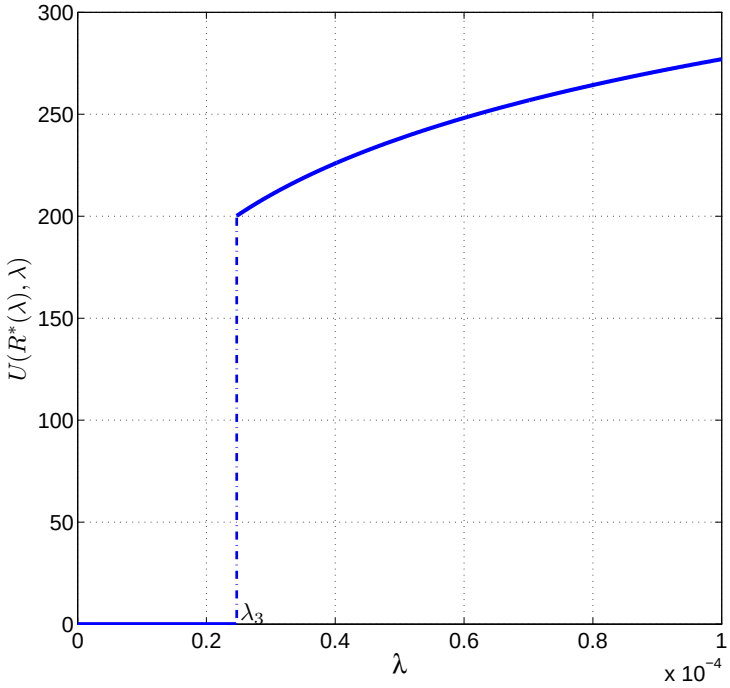


(b)

Figure 2.4: Optimal BS power control v.s. MT density: (a) $\lambda_2 \geq \lambda_1$; (b) $\lambda_2 < \lambda_1$



(a)



(b)

Figure 2.5: Average number of supported users v.s. MT density: (a) $\lambda_2 \geq \lambda_1$; (b) $\lambda_2 < \lambda_1$

2.4(a) and Fig. 2.5(a), it is assumed that $P_c = 120$ W and the corresponding optimal dual solution for Problem (P1) is $\mu^* = 1.05$, with which it can be verified that $\lambda_2 > \lambda_1$, i.e., corresponding to the first case in Theorem 2.4.1. For Fig. 2.3(b), Fig. 2.4(b) and Fig. 2.5(b), it is assumed that $P_c = 140$ W and $\mu^* = 0.8$; thus the critical values of λ satisfy $\lambda_3 > \lambda_1 > \lambda_2$, which is in accordance with the second case of Theorem 2.4.1. It is observed that the numerical examples validate our theoretical results. As shown in Fig. 2.3, a cut-off value of λ exists (note that $\bar{\lambda}_i, i = 1, 2, 3$, represent the approximate critical values of λ obtained by Corollary 2) in either of the two cases of Theorem 2.4.1, which implies that allowing BS to be switched off under light load is optimal for energy saving. Note that from Fig. 2.3, the approximate range adaptation is observed to match well with the optimal range adaptation for both cases. Fig. 2.4 shows the optimal BS power adaptation versus the MT density. It is observed that once the BS is on, it transmits near or at the maximum power budget, which implies that constant power transmission at "on" mode is near or even optimal. This also explains the observation in Fig. 2.3(a) that the deviation of the approximated value of λ_2 or $\bar{\lambda}_2$ from λ_2 does not affect the accuracy of the approximate range adaptation policy, since the accuracy of λ_1 and λ_3 that control BS's on-off behavior is more crucial. The variations of the system throughput $U(R^*(\lambda), \lambda)$ with MT density λ under the optimal scheme is shown in Fig. 2.5. As discussed in Corollary 1, $U(R^*(\lambda), \lambda)$ is observed to increase strictly with λ indicating that the optimal adaptation scheme takes advantage of higher MT density to maximize the system throughput.

Next, we compare the suboptimal schemes in Section 2.5 with the optimal scheme. With $P_c = 60$ W, Fig. 2.6 shows the average power consumption $\bar{P}_{\text{BS}}$ at BS versus the system throughput U_{avg}. From Fig. 2.6, we observe that ARw/OFC performs almost the same as the optimal scheme over the entire range of values of U_{avg}. This is because that constant power transmission at BS "on" mode is near or even optimal (c.f. Fig. 2.4(b)) and ARw/OFC differs from the optimal scheme only in that the (long-term) transmit power control when BS is on (c.f. Fig. 2.4(a) with $\lambda_1 < \lambda < \lambda_2$) is not implemented. It is also observed that when U_{avg} is small, FRw/OFC has similar energy consumption as the optimal scheme and ARw/OFC; however, their

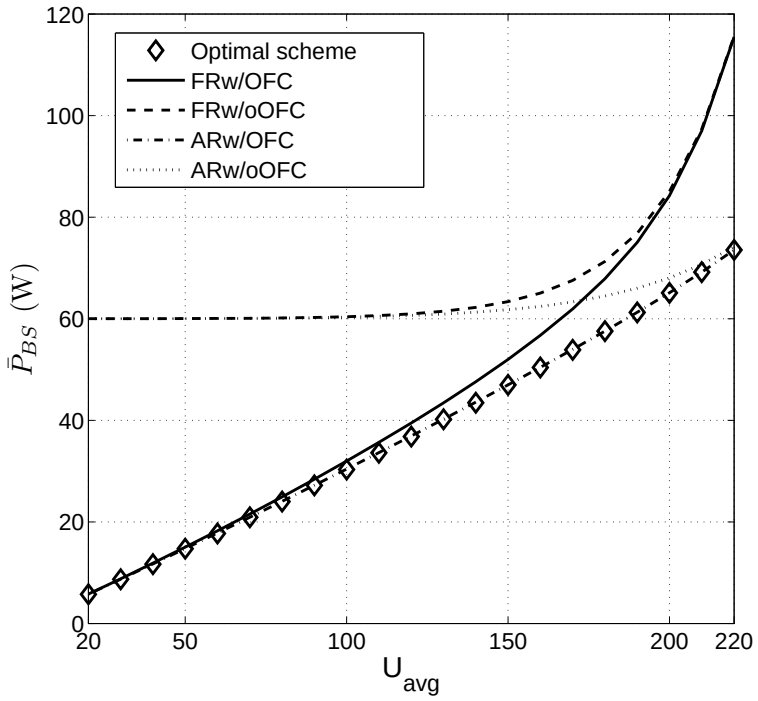


Figure 2.6: Performance comparison with $P_c = 60$ W and $\bar{v} = 150$ Kbps

performance gap is enlarged as U_{avg} increases. A similar observation can be made by comparing ARw/oOFC and FRw/oOFC. From these observations, it follows that BS on-off control is the most effective ESM when the network throughput is low, while range adaptation plays a more important role when the network throughput becomes higher. Finally, we observe that ARw/OFC and FRw/OFC converge to ARw/oOFC and FRw/oOFC, respectively, as U_{avg} increases. This is because that to achieve higher network throughput, BS needs to be "on" for more time to support larger number of MTs; as a result, BS on-off control is less useful for energy saving.

In Fig. 2.7, we set $P_c = 100$ W to further evaluate the performances of different schemes under a higher non-transmission related power consumption at BS. Similar observations can be made from Fig. 2.7 as in Fig. 2.6. However, it is worth noticing that BS on-off control plays a more dominant role for energy saving when U_{avg} is small, since a higher P_c is required. It is also interesting to observe that the performance gaps among different schemes with and without range adaptation are almost invariant to the change of P_c at high network throughput, which is around 45 W in both Figs. 2.6 and

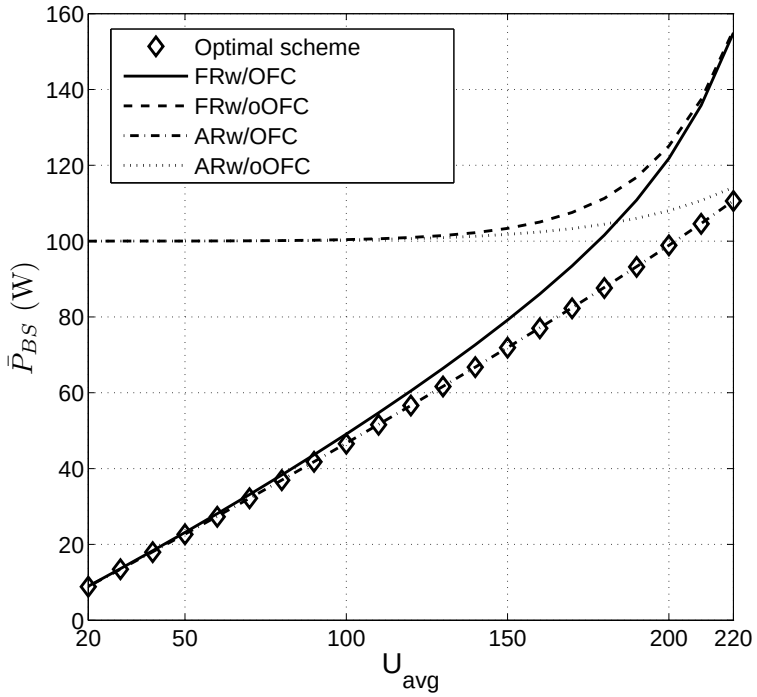


Figure 2.7: Performance comparison with $P_c = 100$ W and $\bar{v} = 150$ Kbps

2.7 with $U_{\text{avg}} = 220$. In Fig. 2.8, P_c is reset as 60 W but the transmission rate for each MT $\bar{v}$ is increased to 500 kbits/sec to model the case with high-rate multimedia traffic. The simulation result shows that the convergence between different schemes with and without BS on-off control is much faster, which implies that range adaptation becomes more effective.

To summarize, we draw the following key conclusions on the effects of different ESMs on the BS energy saving performance:

- BS on-off control is the most effective ESM when the network throughput is not high;
- Cell range adaptation plays a more important role in BS energy saving when the network throughput is higher;
- Finer-grained transmit power control at BS does not introduce significant benefit, i.e., constant power transmission at BS "on" mode is practically optimal.

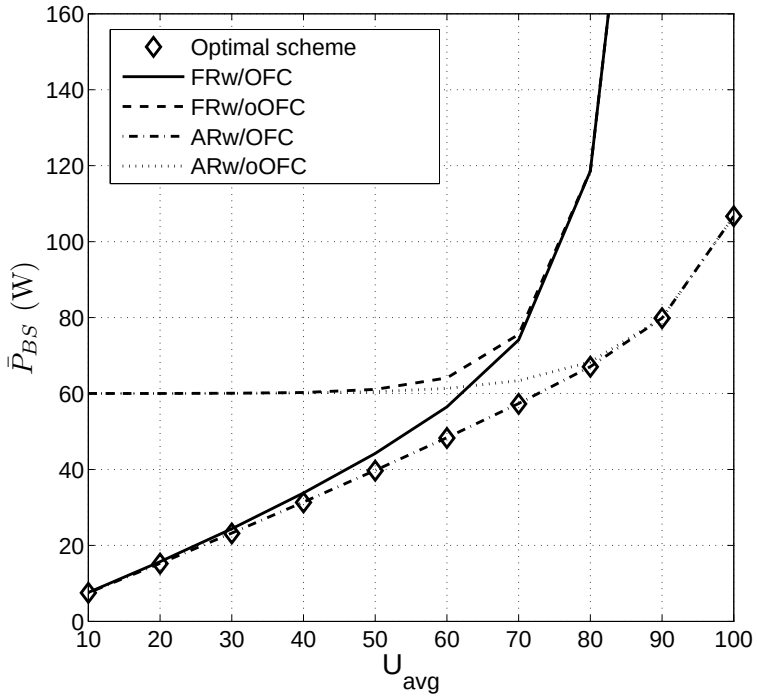


Figure 2.8: Performance comparison with $P_c = 60$ W and $\bar{v} = 500$ Kbps

2.7 Chapter Summary

In this chapter, under an OFDMA-based broadcast channel setup, we investigated optimal power and range adaptation polices for time-varying traffic to minimize the BS's average power consumption subject to the throughput and QoS constraints. A new power scaling law that relates the (short-term) average transmit power at BS with the given cell range and MT density was derived, based on which we obtained the optimal power and range adaptation policy by solving a joint cell range adaptation and (long-term) power control problem. By exploiting the fact that energy saving at BS essentially comes from two major mechanisms, namely BS's on-off power control and range adaptation, suboptimal schemes were proposed to investigate their effects on the system energy saving. It was shown by simulation results that when the network throughput is moderate, BS's on-off power control is the most effective energy saving mechanism, while when the network throughput increases, range adaptation becomes more effective. Note that in this chapter we have only studied the minimization of

BS's energy consumption. In the next chapter, motivated by the discussions in Section 1.3.2, the energy consumption at MTs is jointly minimized with that of the BS.

Chapter 3

Joint Transmitter and Receiver Energy Minimization in Multiuser OFDM System

3.1 Introduction

In Chapter 2, we studied OFDMA based DL broadcasting and minimized the long-term energy consumption at the BS through two levels of power control. In this chapter, to account for both transmitter- and receiver-side energy consumption, we characterize the tradeoffs in minimizing the BS's versus MTs' energy consumption in multiuser OFDM based DL transmission by investigating a weighted-sum transmitter and receiver joint energy minimization (WSTREMin) problem, subject to the given transmission power constraint at the BS and data requirements of individual MTs. It is shown that Dynamic TDMA (D-TDMA) as illustrated in Fig. 3.1(a), where MTs are scheduled in orthogonal time slots for single-user OFDM transmission, is the optimal strategy for weighted-sum receiver-side energy minimization (WSREMin) at MTs. In contrast, OFDMA as shown in Fig. 3.1(b) is proven to be optimal for transmitter-side energy minimization (TEMin) at the BS. To obtain more flexible energy consumption tradeoffs between the BS and MTs for WSTREMin, we further propose a new multiple access scheme, i.e., Time-Slotted OFDMA (TS-OFDMA) scheme as illustrated in Fig. 3.1(c), in which MTs are grouped into orthogonal time slots with OFDMA applied to the set of users that are assigned to the same time slot. TS-OFDMA can be shown to include both the D-TDMA and OFDMA as special cases.

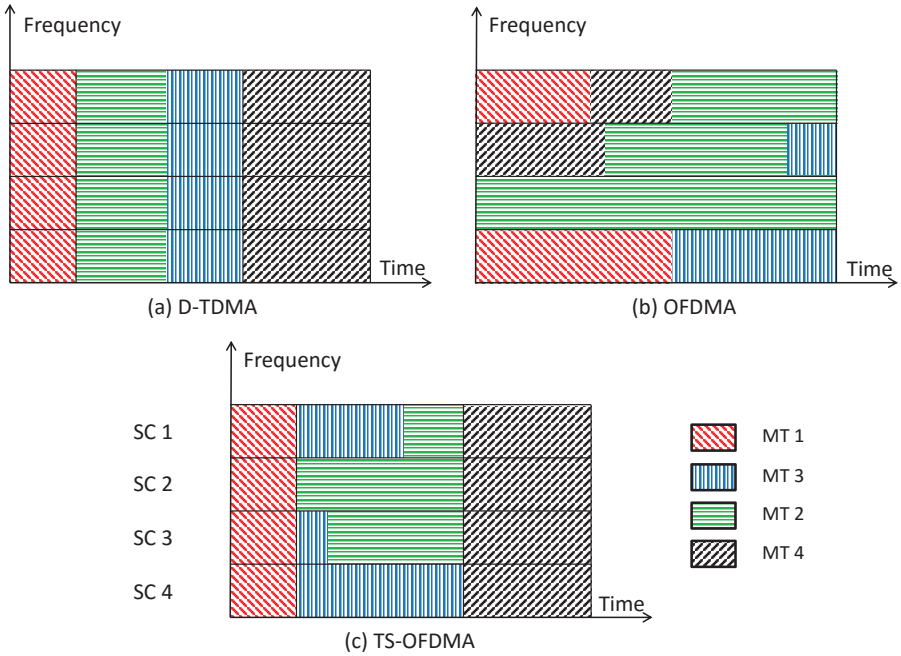


Figure 3.1: Transmission schemes: (a) Dynamic TDMA (D-TDMA); (b) OFDMA; and (c) Time-Slotted OFDMA (TS-OFDMA)

3.2 Literature Review

As introduced in Chapter 1, OFDMA has been adopted in various wireless communication standards, e.g., WiMAX and 3GPP LTE [34], to meet the fast growing mobile traffic volume. However, the complexity of OFDMA leads to increased energy consumption, which thus attracts widespread interest in emphasizing the improvement in EE optimization for OFDMA based networks [35–39].

Prior to the relatively recent emphasis on EE, the research on OFDMA based wireless networks has mainly focused on dynamic resource allocation, which includes dynamic subcarrier (SC) and power allocation, and/or data rate adaptation, for the purposes of either maximizing the throughput [40–43] or minimizing the transmit power [44,45]. The authors in [44] first considered the problem of power minimization in OFDMA, through adaptive SC and power allocation, subject to transmit power and MTs' individual rate constraints. A time sharing factor, taking values within the interval $[0, 1]$, was introduced to relax the original problem to a convex problem, which

can then be efficiently solved. The throughput maximization problem for OFDMA can be more generally formulated as a utility maximization problem [41]. For example, if the utility function is the network sum-throughput itself, then the maximum value is achieved with each SC being assigned to the MT with the largest channel gain together with the water-filling power allocation over SCs [42]. This work has been extended to the case of rate proportional fair scheduling in [43,46]. The Lagrange dual decomposition method [47] was proposed in [45] to provide an efficient algorithm for solving OFDMA based resource allocation problems. Although there has been no proof yet for the optimality of the solution by the dual decomposition method, it was shown in [45] that with a practical number of SCs, the duality gap is virtually zero.

Recently, there has been an upsurge of interest in EE optimization for OFDMA based networks [35–39]. Since energy scarcity is more severe at mobile terminals (MTs), due to the limited capacity of batteries, energy-efficient design for OFDMA networks was first considered under the uplink setup [35]. The sum of MTs' individual EEs, each defined as the ratio of the achievable rate to the corresponding MT's power consumption, was maximized considering both the circuit and transmit power (termed the total power consumption in the sequel). EE maximization for OFDMA downlink transmissions has been studied in [36–39]. A generalized EE, i.e., the weighted-sum rate divided by the total power consumption, was maximized in [37] under prescribed user rate constraints. Instead of modeling circuit power as a constant, the authors in [36, 38] proposed a model of rate-dependent circuit power, in the context of EE maximization, since larger circuit power is generally required to support a higher data rate.

It is worth noting that most of the existing work on EE-based resource allocation for OFDMA has only considered transmitter-side energy consumption. However, in an OFDMA downlink, energy consumption at the receivers of MTs is also an important issue given the limited power supply of MTs. Therefore, it is interesting to design resource allocation schemes that prolong the operation time of MTs by minimizing their energy usage. Since the energy consumption at the receivers is roughly independent of the data rate and merely dependent on the active time of

the MT [48], the dominant circuit power consumption at MTs should be considered. Consequently, fast transmission is more beneficial for reducing the circuit energy consumption at the receivers. A similar idea has also been explored in a recent work [49].

3.3 System Model and Problem Formulation

3.3.1 System Model

Consider a multiuser OFDM-based downlink transmission system consisting of one BS, N orthogonal subcarriers (SCs) each with a bandwidth of W Hz, and K MTs. Let $\mathcal{K}$ and $\mathcal{N}$ denote the sets of MTs and SCs, respectively. We assume that each SC can be assigned to at most one MT at any given time, but the SC assignment is allowed to be shared among MTs over time, i.e., SC time sharing. We also assume that the noise at the receiver of each MT is modeled by an additive white Gaussian noise (AWGN) with one-sided power spectrum density denoted by N_0. Let $p_{k,n}$ be the transmit power allocated to MT k in SC n, $k \in \mathcal{K}$, $n \in \mathcal{N}$, and $r_{k,n}$ be the achievable rate of MT k at SC n in the downlink. Then it follows that

$$r_{k,n} = W \log_2 \left(1 + \frac{h_{k,n} p_{k,n}}{\Gamma N_0 W}\right) \tag{3.1}$$

where $\Gamma \geq 1$ accounts for the gap from the channel capacity due to practical modulation and coding, and $h_{k,n}$ is the channel power gain from the BS to MT k at SC n, which is assumed to be perfectly known at both the BS and MT k.

With time sharing of SCs among MTs, $\rho_{k,n}$, dubbed the time sharing factor, is introduced to represent the fraction of time that SC n is assigned to MT k, where $0 \leq \rho_{k,n} \leq 1, \forall k, n$ and $\sum_{k=1}^{K} \rho_{k,n} \leq 1, \forall n$. Let T denote the total transmission time for our proposed scheduling. The amount of information bits delivered to MT k over time T is thus given by

$$Q_k = T \sum_{n=1}^{N} \rho_{k,n} r_{k,n}. \tag{3.2}$$

The average transmit power is given by

$$\bar{P} = \sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} p_{k,n}. \tag{3.3}$$

We assume that $\bar{Q}_k$ bits of data need to be delivered from the BS to MT k over a slot duration T for the time slot of interest. Then the following constraint must be satisfied:

$$Q_k \geq \bar{Q}_k, \forall k \in \mathcal{K}. \tag{3.4}$$

We further assume that the receiver of each MT is turned on only when the BS starts to send the data it requires, which can be at any time within the time slot, and that it is turned off right after all $\bar{Q}_k$ bits of data are received. Let t_k, $0 \leq t_k \leq T$, denote the "on" period of MT k. It is observed that the following inequalities must hold for all MTs:

$$\max_n \{T\rho_{k,n}\} \leq t_k \leq T, \forall k \in \mathcal{K}. \tag{3.5}$$

The origin of this inequality can be understood from Fig. 3.2, where MT k is turned on and then off within the time interval T.

Energy consumption at the BS in general comprises two major parts: transmit power $\bar{P}$ and a constant power $P_{t,c}$ accounting for all non-transmission related energy consumption due to e.g., processing circuits and cooling. Consequently, the total energy consumed by the BS over duration T, denoted by E_t, can be modeled as

$$E_t = T\bar{P} + TP_{t,c}. \tag{3.6}$$

On the other hand, the power consumption at the receiver of each MT is assumed to be constant [48], denoted by $P_{r,c}$, when it is in the "on" period receiving data from the BS. Otherwise, if the receiver does not receive any data from the BS, its consumed power is in general negligibly small and thus is assumed to be zero. Hence, the receiver energy consumed by each MT k over T, denoted by $E_{r,k}$, can be approximately modeled as

$$E_{r,k} = P_{r,c} t_k, k \in \mathcal{K}. \tag{3.7}$$

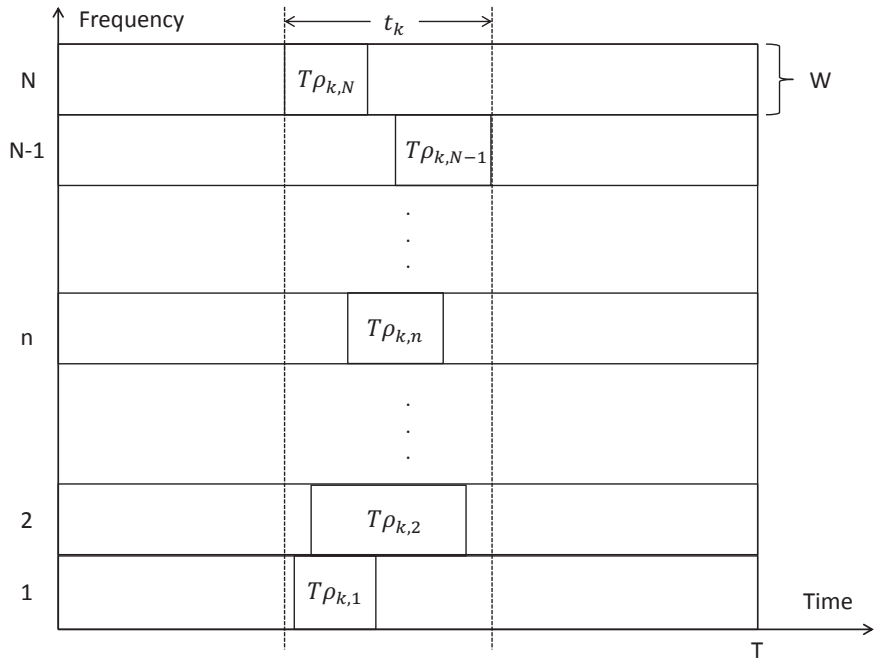


Figure 3.2: Multiuser OFDM transmission with subcarrier time sharing

In general, each MT can be in a different state of energy depletion, and thus it is sensible to define a weighted-sum receiver-side energy (WSRE) consumption of all MTs as

$$E_r^w = \sum_{k=1}^{K} \alpha_k E_{r,k} \tag{3.8}$$

where a larger weight α_k reflects the higher priority of MT k in terms of energy minimization.

It is assumed that all channels $h_{k,n}$'s are constant over the total transmission time of a frame, T. While in theory the optimal T is unbounded, for a practical number of bits to be transmitted per frame, $\bar{Q}_k$'s, and practical transmit power levels $P_{t,c}$ and $P_{r,c}$, the designed optimal T will be finite and in fact usually quite small. If we consider low-mobility and/or short frame lengths, then the assumption of a static channel over an indeterminate T is valid. However, in Section 3.7, we provide detailed discussions on how the obtained results in this chapter can be extended to the case when an explicit maximum transmission time constraint is imposed.

3.3.2 Problem Formulation

We aim to characterize the tradeoffs in minimizing the BS's versus MTs' energy consumption, i.e., E_t versus $E_{r,k}$'s, in multiuser OFDM based downlink transmission by investigating a weighted-sum transmitter and receiver joint energy minimization (WSTREMin) problem, which is formulated as

$$\begin{aligned}
&(\text{WSTREMin}): \\
&\underset{\{p_{k,n}\},\{\rho_{k,n}\},T}{\texttt{Min.}} \quad \sum_{k=1}^{K} \alpha_k t_k P_{r,c} \\
&\quad + \alpha_0 \left(\sum_{k=1}^{K} \sum_{n=1}^{N} T\rho_{k,n} p_{k,n} + TP_{t,c} \right) && (3.9) \\
&\texttt{s.t.} \quad \sum_{k=1}^{K} \rho_{k,n} \leq 1, \forall n && (3.10) \\
&\quad \sum_{n=1}^{N} T\rho_{k,n} r_{k,n} \geq \bar{Q}_k, \forall k && (3.11) \\
&\quad \sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} p_{k,n} \leq P_{\text{avg}} && (3.12) \\
&\quad T > 0, p_{k,n} \geq 0,\ 0 \leq \rho_{k,n} \leq 1, \forall n, k && (3.13)
\end{aligned}$$

where α_0 is an additional weight assigned to the BS, which controls the resulting minimum energy consumption of the BS as compared to those of MTs. Notice that the design variables in the above problem include the power allocation $p_{k,n}$, time sharing factor $\rho_{k,n}$, as well as transmission time T, while the constraints in (3.10) are to limit the total transmission time at each SC to be within T, those in (3.11) are for the data requirements of different MTs, and that in (3.12) specifies the average transmit power at BS, denoted by P_{avg}. The main difficulty in solving problem (WSTREMin) lies in the absence of a functional relationship among t_k, $\rho_{k,n}$'s and T with the inequality in (3.5) being the only known expression that links the three variables. Minimizing over the upper bound of each MT's energy consumption, i.e., $TP_{r,c}$, which could be quite loose as illustrated in Fig. 3.2, may result in conservative or energy-inefficient solution.

In order to obtain useful insights into the optimal energy consumption for the BS and MTs, we first consider two extreme cases separately in the following two sections, i.e., WSRE minimization (WSREMin) corresponding to the case of $\alpha_0 = 0$ in Section 3.4 and transmitter-side energy minimization (TEMin) corresponding to the case of $\alpha_k = 0, \forall k$, respectively, in Section 3.5. Compared with problem (WSTREMin), problems (WSREMin) and (TEMin) have exactly the same set of constraints but different objective functions. We will illustrate how problem (WSTREMin) may be practically solved based on the results from the the two extreme cases in Section 3.6.

Remark 3.3.1. Problem (WSTREMin) could have an alternative interpretation by properly setting the energy consumption weights α_0 and α_k's. Suppose α_0 and α_k represent the unit cost of consumed energy at the BS and MT k, respectively. Since MTs are usually powered by capacity limited batteries in comparison to the electrical grid powered BS, α_0 and α_k's should reflect the energy price in the market for the BS and the risk of running out of energy for each MT k, respectively. With this definition, problem (WSTREMin) can be treated as a network-wide cost minimization problem. How to practically select the values of α_0 and α_k's to achieve this end is beyond the scope of this chapter.

3.4 Receiver-Side Energy Minimization

In this section, we consider minimizing receiver energy consumption at all MTs without regard for BS energy consumption. From (3.7) and (3.8), the WSREMin problem is thus formulated as

$$(\text{WSREMin}):$$

$$\underset{\{p_{k,n}\},\{\rho_{k,n}\},T}{\texttt{Min.}} \quad \sum_{k=1}^{K} \alpha_k P_{r,c} t_k \tag{3.14}$$

$$\texttt{s.t.} \quad (3.10), (3.11), (3.12), \text{ and } (3.13). \tag{3.15}$$

As mentioned in Section 3.2, receiver-side energy minimization has also been considered in [49], in which the available time-frequency resources are divided into

equally spaced RBs over both time and frequency. Flat-fading, i.e., the channels are the same across all the RBs, was assumed for each MT, based on which an integer programme with each MT constrained by the number of required RBs is formulated for RB allocation. Problem (WSREMin), in contrast, assumes a more flexible SC allocation with time sharing factor $\rho_{k,n}$'s to achieve further energy saving. Moreover, the optimal power allocation corresponding to frequency selective channels is obtained.

Similar to problem (WSTREMin), the main difficulty in solving problem (WSREMin) lies in the absence of a functional relationship among t_k, $\rho_{k,n}$'s and T. However, it can be shown that a dynamic TDMA (D-TDMA) based solution, i.e., MTs are scheduled for single-user OFDM transmission over orthogonal slots with respective duration $\rho_k T$, $k = 1, \cdots, K$, with $\sum_{k=1}^{K} \rho_k \leq 1$, is optimal for problem (WSREMin), as given in the following proposition.

Proposition 3.4.1. *Let $\rho_{k,n}^*$, $n = 1, \cdots, N$, and t_k^* denote the optimal set of time sharing factors and the optimal "on" period for MT k, respectively, $k \in \mathcal{K}$, in problem (WSREMin). Then, we have*

$$\rho_{k,n}^* = \rho_k^*, \quad \forall n \in \mathcal{N}, k \in \mathcal{K} \tag{3.16}$$

$$t_k^* = T\rho_k^*, \forall k \in \mathcal{K} \tag{3.17}$$

where ρ_k^* denotes the common value of all $\rho_{k,n}^*$, $\forall n \in \mathcal{N}$, for MT k.

Proof. Proposition 3.4.1 can be proved by first identifying the fact that minimizing the WSRE of all MTs is equivalent to minimizing the weighted-sum "on" time of all MTs. Then, with given allocated transmission power and data requirement, the active period of each MT is minimized by assigning all frequency resource, i.e., all the SCs, exclusively to this particular MT. For a more rigorous proof, please refer to Appendix B.1. □

Remark 3.4.1. Proposition 3.4.1 indicates that the time sharing factors at all SCs should be identical for each MT k to minimize its "on" period, which is achieved by D-TDMA transmission as shown in Fig. 3.1(a). Notice that D-TDMA minimizes

the on time of each MT and therefore their weighted energy consumption, as will be shown next. However, it extends the transmission time of BS, T, and thus may not be optimal from the viewpoint of BS energy saving, as we shall see in Section 3.5.

With Proposition 3.4.1 and t_k's given in (3.17), the WSREMin problem under D-TDMA is formulated as

$$(\text{WSREMin-TDMA}):$$

$$\underset{\{p_{k,n}\geq 0\},\{t_k>0\}}{\texttt{Min.}} \quad \sum_{k=1}^{K} \alpha_k P_{r,c} t_k \tag{3.18}$$

$$\texttt{s.t.} \quad \sum_{n=1}^{N} t_k r_{k,n} \geq \bar{Q}_k, \forall k \tag{3.19}$$

$$\sum_{k=1}^{K}\sum_{n=1}^{N} t_k p_{k,n} \leq P_{\text{avg}} \sum_{k=1}^{K} t_k. \tag{3.20}$$

It is observed that problem (WSREMin-TDMA) is non-convex due to the coupled terms $t_k r_{k,n}$ in (3.19) and $t_k p_{k,n}$ in (3.20). By a change of variables $s_{k,n} = t_k r_{k,n}$, $\forall k, n$, problem (WSREMin-TDMA) can be reformulated as

$$(\text{P1}): \quad \underset{\{s_{k,n}\geq 0\},\{t_k>0\}}{\texttt{Min.}} \quad \sum_{k=1}^{K} \alpha_k P_{r,c} t_k \tag{3.21}$$

$$\texttt{s.t.} \quad \sum_{n=1}^{N} s_{k,n} \geq \bar{Q}_k, \forall k \tag{3.22}$$

$$\sum_{k=1}^{K}\sum_{n=1}^{N} t_k \frac{e^{a\frac{s_{k,n}}{t_k}}-1}{f_{k,n}} \leq P_{\text{avg}} \sum_{k=1}^{K} t_k \tag{3.23}$$

where $f_{k,n} = \frac{h_{k,n}}{\Gamma N_0 W}$ and $a = \frac{\ln 2}{W}$. Note that the objective function in (3.21) and constraints in (3.22) are all affine, while the constraints in (3.23) are convex due to the fact that the function $t_k e^{a\frac{s_{k,n}}{t_k}}$ is the perspective of a strictly convex function $e^{a s_{k,n}}$ with $a > 0$, and thus is a convex function [32]. As a result, problem (P1) is convex. Thus, the Lagrange duality method can be applied to solve this problem exactly [32].

In the rest of this section, instead of solving the dual of problem (P1) directly which involves only numerical calculation and provides no insights, we develop a simple bisection search algorithm by revealing the structure of the optimal solution to problem (WSREMin-TDMA), given in the following theorem.

Theorem 3.4.1. *Let* $\boldsymbol{\lambda}^* = [\lambda_1^*, \cdots, \lambda_K^*] \geq \mathbf{0}$ *and* $\beta^* \geq 0$ *denote the optimal dual solution to problem (P1). The optimal solution of problem (WSREMin-TDMA) is given by*

$$p_{k,n}^* = \left(\frac{\lambda_k^*}{a\beta^*} - \frac{1}{f_{k,n}}\right)^+ \tag{3.24}$$

$$t_k^* = \frac{a\bar{Q}_k}{\sum_{n=1}^N \left(\ln \frac{\lambda_k^* f_{k,n}}{a\beta^*}\right)^+} \tag{3.25}$$

where $\boldsymbol{\lambda}^*$ *and* β^* *need to satisfy*

$$\beta^* - \min_k(\alpha_k) P_{r,c}/P_{avg} < 0 \tag{3.26}$$

$$\alpha_k P_{r,c} - \beta^* P_{avg} + \sum_{n=1}^N u_n(\beta^*, \lambda_k^*) = 0, \forall k \in \mathcal{K} \tag{3.27}$$

where $u_n(\beta, \lambda_k) = \left(\frac{\lambda_k}{a} - \frac{\beta}{f_{k,n}}\right)^+ - \frac{\lambda_k}{a}\left(\ln \frac{\lambda_k f_{k,n}}{a\beta}\right)^+$ *and* $(\cdot)^+ \triangleq \max\{\cdot, 0\}$.

Proof. See Appendix B.2. □

It is observed from (3.24) that the optimal power allocation has a water-filling structure [30], except that the water levels are different over MTs. These are specified by λ_k^* for MT k and need to be found by solving the equations in (3.27). Since it can be shown that $\sum_{n=1}^N u_n(\beta, \lambda_k) \leq 0$ is strictly decreasing in λ_k given $\beta < \min_k\{\alpha_k\} P_{r,c}/P_{\text{avg}}$, with the assumption of identical channels for all the MTs, it is observed that larger α_k results in larger λ_k^* or higher water-level, which means more power should be allocated to the MT that has higher priority in terms of energy minimization.

Based on Theorem 3.4.1, one algorithm to solve problem (WSREMin-TDMA) is given in Table 3.1, in which β^* is obtained through bisectional search until the average power constraint in (3.20) is met with equality. For the algorithm given in Table 3.1, the computation time is dominated by updating the power and time allocation with given β in steps b)-d), which is of order KN. Since the number of iterations required for the bisection search over β is independent of K and N, the overall complexity of the algorithm in Table 3.1 is $\mathcal{O}(KN)$.

Table 3.1: Algorithm for Solving Problem (WSREMin-TDMA)

1. **Given** $\beta_{\text{min}}(\triangleq 0) \leq \beta^* < \beta_{\text{max}}(\triangleq \min_k(\alpha_k)P_{r,c}/P_{\text{avg}})$.
2. **Repeat**
   a) $\beta = \frac{1}{2}(\beta_{\text{min}} + \beta_{\text{max}})$.
   b) Obtain λ_k such that $u(\beta, \lambda_k) = 0$, where $u(\beta, \lambda_k) = \alpha_k P_{r,c} - \beta P_{\text{avg}} + \sum_{n=1}^{N} u_n(\beta, \lambda_k)$, $k = 1, \cdots, K$,.
   c) Obtain $p_{k,n}$ and t_k according to (3.24) and (3.25) for $k = 1, \cdots, K$, $n = 1, \cdots, N$.
   d) If $\sum_{k=1}^{K}\sum_{n=1}^{N} t_k p_{k,n} \geq P_{\text{avg}} \sum_{k=1}^{K} t_k$, set $\beta_{\text{min}} \leftarrow \beta$; otherwise, set $\beta_{\text{max}} \leftarrow \beta$.
3. **Until** $\beta_{\text{max}} - \beta_{\text{min}} < \delta$ where δ is a small positive constant that controls the algorithm accuracy.

3.5 Transmitter-Side Energy Minimization

In this section, we study the case of minimizing the energy consumption at the BS while ignoring the receiver energy consumption at MTs. From (3.3) and (3.6), the transmitter-side energy minimization (TEMin) problem is formulated as

$$(\text{TEMin}):$$

$$\underset{\{p_{k,n}\},\{\rho_{k,n}\},T}{\texttt{Min.}} \quad \sum_{k=1}^{K}\sum_{n=1}^{N} T\rho_{k,n}p_{k,n} + TP_{t,c} \tag{3.28}$$

$$\texttt{s.t.} \quad (3.10), (3.11), (3.12), \text{ and } (3.13). \tag{3.29}$$

A similar formulation has been considered in [36–39], in which the energy efficiency, defined as the ratio of the achievable rate to the total power consumption, is maximized under prescribed user rate constraints. Problem (TEMin), in contrast, considers the data requirements $\bar{Q}_k$'s and includes the transmission time T as a design variable to

explicitly address the tradeoffs between the transmission and non-transmission related energy consumption at BS: longer transmission time results in larger non-transmission related energy consumption $TP_{t,c}$ but smaller transmission related energy consumption $\sum_{k=1}^{K}\sum_{n=1}^{N} T\rho_{k,n}p_{k,n}$ with given data requirements [35].

Problem (TEMin) is also non-convex due to the coupled terms $T\rho_{k,n}r_{k,n}$ in (3.11) and $\rho_{k,n}p_{k,n}$ in (3.12). Compared with [36–39], it is observed that the design variable T further complicates the problem. To solve this problem, we propose to decompose problem (TEMin) into two subproblems as follows.

$$(\text{TEMin-1}): \quad \underset{\{p_{k,n}\},\{\rho_{k,n}\}}{\texttt{Min.}} \quad \sum_{k=1}^{K}\sum_{n=1}^{N} \rho_{k,n}p_{k,n} \tag{3.30}$$

$$\texttt{s.t.} \quad (3.11) \text{ and } (3.12) \tag{3.31}$$

$$p_{k,n} \geq 0,\ 0 \leq \rho_{k,n} \leq 1, \forall n, k. \tag{3.32}$$

$$(\text{TEMin-2}): \quad \underset{T}{\texttt{Min.}} \quad Tv(T) + TP_{t,c} \tag{3.33}$$

$$\texttt{s.t.} \quad v(T) \leq P_{\text{avg}} \tag{3.34}$$

$$T > 0. \tag{3.35}$$

Here, $v(T)$ denotes the optimal value of the objective function in problem (TEMin-1). Note that problem (TEMin-1) minimizes the BS average transmit power with given transmission time T and a set of data constraints $\bar{Q}_k$. Then problem (TEMin-2) searches for the optimal T to minimize the total energy consumption at BS subject to the average transmit power constraint, P_{avg}. In the rest of this section, we first solve problem (TEMin-1) with given $T > 0$. Then, we show that problem (TEMin-2) is convex and can be efficiently solved by a bisection search over T.

3.5.1 Solution to Problem (TEMin-1)

With given $T > 0$, the data requirement $\bar{Q}_k$ for MT k can be equivalently expressed in terms of rate as $c_k = \frac{\bar{Q}_k}{T}$. Similarly as for problem (P1), we make a change of variables as $m_{k,n} = \rho_{k,n}r_{k,n}, \forall k, n$. Moreover, we define $\frac{m_{k,n}}{\rho_{k,n}} = 0$ at $m_{k,n} = \rho_{k,n} = 0$

to maintain continuity at this point. Problem (TEMin-1) is then reformulated as

$$(\mathrm{P2}): \underset{\{m_{k,n}\},\{\rho_{k,n}\}}{\texttt{Min.}} \quad \sum_{k=1}^{K}\sum_{n=1}^{N} \rho_{k,n} \frac{e^{a\frac{m_{k,n}}{\rho_{k,n}}}-1}{f_{k,n}} \tag{3.36}$$

$$\texttt{s.t.} \quad \sum_{k=1}^{K} \rho_{k,n} \leq 1, \forall n \tag{3.37}$$

$$\sum_{n=1}^{N} m_{k,n} \geq c_k, \forall k \tag{3.38}$$

$$m_{k,n} \geq 0, 0 \leq \rho_{k,n} \leq 1, \forall k, n. \tag{3.39}$$

Although problem (P2) can be shown to be convex just as for problem (P1), it does not have the provably optimal structure for SC allocation given in Proposition 3.4.1. In this case, in general the SC's are shared among all MTs at any given time, denoted by the set of time sharing factors $\{\rho_{k,n}\}$, which are different for all k and n in general. Since problem (P2) is convex, the Lagrange duality method can be applied to solve this problem optimally. Another byproduct of solving problem (P2) by this method is the corresponding optimal dual solution of problem (P2), which will be shown in the next subsection to be the desired gradient of the objective function in problem (TEMin-2) required for solving this problem. The details of solving problem (P2) and its dual problem through the Lagrange duality method can be found in Appendix B.3 with one algorithm summarized in Table B.1.

We point out here that the problem of transmit power minimization for OFDMA downlink transmission with SC time sharing has also been studied in [43, 44]. In [43], problem (P2) is solved directly without introducing its dual problem, but in this chapter, the corresponding dual solution is the gradient of the objective function in problem (TEMin-2) and therefore the dual problem is important. In [44], the dual variables are updated one at a time until the data rate constraints for all users are satisfied, and this is extremely slow. In this chapter, the optimal dual solution of problem (P2) is obtained more efficiently by the ellipsoid method [47]. Since with the optimal dual solutions, we may obtain infinite sets of primal solution, and some might not satisfy the constraints in (3.37) and/or (3.38) [50], the optimal solution of problem (P2) is further obtained by solving a linear feasibility problem (more details

are given in Appendix B.3). Finally, in [43, 44], the time sharing factor $\rho_{k,n}$ is treated as a relaxed version of the SC allocation indicator, which needs to be quantized to be 0 or 1 after solving problem (P2). However, since problem (P2) in this chapter is only a subproblem of problem (TEMin), in which the transmission time T is a design variable, SC time sharing can indeed be implemented with proper scheduling at the BS such that each SC is still assigned to at most one MT at any given time.

3.5.2 Solution to Problem (TEMin-2)

With problem (TEMin-1) solved, we proceed to solve problem (TEMin-2) in this subsection. First, we have the following lemma.

Lemma 3.5.1. *Problem (TEMin-2) is convex.*

Proof. See Appendix B.4. □

Since problem (TEMin-2) is convex, and $v(T)$ is continuous and differentiable [51], a gradient based method e.g., Newton method [32] can be applied to solve problem (TEMin-2), where the required gradient is given in the following lemma.

Lemma 3.5.2. *The gradient of $v(T)T + P_{t,c}T$ with respect to T, $T > 0$, is given by*

$$v(T) - \frac{1}{T}\sum_{k=1}^{K} \lambda_k^*(T)\bar{Q}_k + P_{t,c} \tag{3.40}$$

where $\{\lambda_k^*(T)\}$ is the optimal dual solution of problem (P2) with given $T > 0$.

Proof. See Appendix B.5. □

3.5.3 Algorithm for Problem (TEMin)

With both problems (TEMin-1) and (TEMin-2) solved, the solution of problem (TEMin) can be obtained by iteratively solving the above two problems. In summary, an algorithm to solve problem (TEMin) is given in Table 3.2. For the algorithm given in Table 3.2, the computation time is dominated by obtaining $v(T)$ and $\boldsymbol{\lambda}^*(T)$ with given T through the algorithm in Table B.1 of Appendix B.3, which is of order

$K^4 + N^4 + K^3N^3$. Similarly, since the number of iterations required for the bisection search over T is independent of K and N, the overall complexity of the algorithm given in Table 3.2 bears the same order over K and N as that for the algorithm in Table B.1 of Appendix B.3, which is $\mathcal{O}(K^4 + N^4 + K^3N^3)$.

Table 3.2: Algorithm for Solving Problem (TEMin)

1. Define $y(T) \triangleq v(T) - \frac{1}{T}\sum_{k=1}^{K} \lambda_k^*(T)\bar{Q}_k + P_{t,c}$, where $v(T)$ and $\boldsymbol{\lambda}^*(T)$ are obtained by the Algorithm 2 in Table B.1 of Appendix B.3.
2. Obtain $T^{'}$ through bisection search such that $y(T^{'}) = 0$.
3. **If** $v(T^{'}) \leq P_{\text{avg}}$, **then** $T^* = T^{'}$; **otherwise** find T^* through bisection search such that $v(T^*) = P_{\text{avg}}$.
4. Obtain the optimal solution of problem (P2), i.e., $\{\{m_{k,n}^*\}, \{\rho_{k,n}^*\}\}$, with T^* by the Algorithm 2 in Table B.1 of Appendix B.3.
5. Obtain the optimal solution of problem (TEMin), i.e., $\{\{p_{k,n}^*\}, \{\rho_{k,n}^*\}\}$, as $p_{k,n}^* = \left(2^{m_{k,n}^*/\rho_{k,n}^*} - 1\right)/f_{k,n}, \forall k, n$.

Remark 3.5.1. Compared with the D-TDMA based solution in Section 3.4 for the case of receiver-side energy minimization, the optimal solution of problem (TEMin) for transmitter-side energy minimization implies that OFDMA (c.f. Fig. 3.1(b)), in which the N SCs are shared among all MTs at any given time, needs to be employed. However, OFDMA may prolong the active time of individual MTs, i.e., t_k's, and is thus not energy efficient in general from the perspective of MT energy saving.

3.6 Joint Transmitter and Receiver Energy Minimization

From the two extreme cases studied in Sections 3.4 and 3.5, we know that D-TDMA as shown in Fig. 3.1(a) is the optimal transmission strategy to minimize the weighted-sum receive energy consumption at the MT receivers; however, OFDMA as shown in Fig. 3.1(b) is optimal to minimize the energy consumption at the BS transmitter. There is evidently no single strategy that can minimize the BS's and MTs' energy consumptions in OFDM-based multiuser downlink transmission. In this section, motivated by the solutions derived from the previous two special cases, we propose a new multiple access scheme termed Time-Slotted OFDMA (TS-OFDMA) transmission scheme, which includes D-TDMA and OFDMA as special cases, and propose an efficient algorithm to approximately solve problem (WSTREMin) using the proposed TS-OFDMA.

3.6.1 Time-Slotted OFDMA

The TS-OFDMA scheme is described as follows. The total transmission time T is divided into J orthogonal time slots with $1 \leq J \leq K$. The K MTs are then assigned to each of the J slots for downlink transmission. Let Φ_j represent the set of MTs assigned to slot j, $j = 1, \cdots, J$. We thus have

$$\Phi_j \cap \Phi_k = \varnothing, \forall j \neq k \tag{3.41}$$

$$\bigcup_j \Phi_j = \mathcal{K}. \tag{3.42}$$

The period that each MT k is switched on (versus off) then equals the duration of its assigned slot, denoted by T_j, i.e., $t_k = T_j$ if $k \in \Phi_j$, with $\sum_{j=1}^{J} T_j = T$. Notice that TS-OFDMA includes D-TDMA (if $J = K$) and OFDMA (if $J = 1$) as two special cases[1]. An illustration of TS-OFDMA for a multiuser OFDM system with $K = 4$, $N = 4$, and $J = 3$ is given in Fig. 3.1(c).

[1]Note that OFDMA is considered as a flexible transmission scheme, in which each MT can use any subcarrier at any time during the transmission, and TS-OFDMA may be seen as a special form of

3.6.2 Solution to Problem (WSTREMin)

In this subsection, we solve problem (WSTREMin) based on TS-OFDMA with given J and MT grouping. We first study two special cases, i.e., $J = K$ and $J = 1$, which can be regarded as the extensions of the results in Section 3.4 and Section 3.5, respectively, by considering the weighted-sum transmitter and receiver energy consumption as the objective function. We thus have the following results.

1. $J = K$ and $|\Phi_j| = 1, j = 1, \cdots, J$: problem (WSTREMin) can be reformulated as

$$\begin{aligned} \underset{\{p_{k,n}\geq 0\},\{t_k>0\}}{\texttt{Min.}} \quad & \sum_{k=1}^{K} t_k \left(\alpha_k P_{r,c} + \alpha_0 P_{t,c}\right) \\ & + \alpha_0 \sum_{k=1}^{K} t_k \sum_{n} p_{k,n} \\ \texttt{s.t.} \quad & (3.19) \text{ and } (3.20). \end{aligned} \tag{3.43}$$

   Note that for $J = K$, $T_k = t_k, \forall k$. Although problem (3.43) and problem (WSREMin-TDMA) differ in their objective functions, problem (3.43) can be recast as a convex problem similarly as problem (WSREMin-TDMA), and it can be shown that their optimal solutions possess the same structure. Therefore, problem (3.43) can be solved by the algorithm similar to that in Table 3.1.

2. $J = 1$ and $|\Phi_J| = K$: problem (WSTREMin) can be simplified to

$$\begin{aligned} \underset{\{p_{k,n}\},\{\rho_{k,n}\},T}{\texttt{Min.}} \quad & \alpha_0 T \sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} p_{k,n} \\ & + T \left(\alpha_0 P_{t,c} + \sum_{k=1}^{K} \alpha_k P_{r,c}\right) \\ \texttt{s.t.} \quad & (3.10), (3.11), (3.12), \text{ and } (3.13). \end{aligned} \tag{3.44}$$

OFDMA. However, as mentioned in the previous sections, it is difficult to quantify the "on" period of each MT with the inequality in (3.5) being the only known expression. The proposed TS-OFDMA is thus more "general" than OFDMA and D-TDMA in the sense that it explicitly allows each MT to be off for a fraction of a frame (outside its assigned time slot) to save energy, and yet allows subcarriers sharing among users within the same time slot.

Since problem (3.44) has exactly the same structure as problem (TEMin), it can be solved by the algorithm similar to that in Table 3.2.

Next, consider the general case of $1 < J < K$. In this case, we divide J slots into two sets as

$$\mathcal{B}_1 = \{j \ : \ |\Phi_j| = 1, j = 1, \cdots, J\} \tag{3.45}$$

$$\mathcal{B}_2 = \{j \ : \ |\Phi_j| \geq 2, j = 1, \cdots, J\} \tag{3.46}$$

where $\mathcal{B}_1$ and $\mathcal{B}_2$ include slots that correspond to transmissions to single MT and multiple MTs, respectively. For slots in $\mathcal{B}_1$, we can further group them together and thereby formulate one single WSTREMin problem similarly as for the case of $J = K$. On the other hand, for slots in $\mathcal{B}_2$, we can perform WSTREMin in each slot separately similarly as for the case of $J = 1$. Furthermore, we assume that the average power assigned to all the slots in $\mathcal{B}_1$ and each slot in $\mathcal{B}_2$ are P_{avg} to avoid coupled power allocation over these slots, so that each problem can be solved independently. Note that it is possible to jointly optimize the power allocation across all the slots. However, it requires extra complexity and thus this approach was not pursued.

The final tasks remaining in solving problem (WSTREMin) is to find the the optimal number of slots and to optimally assign MTs to each of these slots. Since finding the optimal grouping is a combinatorial problem, an exhaustive search can incur a large complexity if K is large. To avoid the high complexity of exhaustive search, we propose a suboptimal MT grouping algorithm for $1 < J < K$ in Section 3.6.3 next. The optimal J can then be found by a one-dimension search.

3.6.3 Suboptimal Mobile Terminal Grouping

In this subsection, we propose a suboptimal grouping algorithm for given $1 < J < K$, termed as *channel orthogonality based grouping* (COG), with low complexity. The proposed algorithm is motivated by the observation that grouping MTs, whose strongest channels are orthogonal to each other (i.e., in different SCs), into one slot will not affect the power allocation and transmission time of each MT but will shorten the total transmission time, and thus reduce the total energy consumption.

For the purpose of illustration, we first define the following terms. Let $\mathbf{h}_k = [h_{k,1}, \cdots, h_{k,N}]^T$ and $\hat{\mathbf{h}}_k$ denote the original and normalized (nonnegative) channel vector from the BS to MT k across all SCs, respectively, where $\hat{\mathbf{h}}_k = \frac{\mathbf{h}_k}{\|\mathbf{h}_k\|}$. Furthermore, let $\pi_{k,l}$ denote the channel correlation index (CCI) between MTs k and l, which is defined as the inner product between their normalized channel vectors, i.e.,

$$\pi_{k,l} = \hat{\mathbf{h}}_k^T \hat{\mathbf{h}}_l, \forall k, l \neq k. \tag{3.47}$$

Note that $\pi_{k,l} = \pi_{l,k}$, and smaller (larger) $\pi_{k,l}$ indicates that MT k is more (less) orthogonal to MT j in terms of channel power realization across different SCs, which can be utilized as a cost associated with grouping MTs k and l into one slot. Finally, define the sum-CCI Π_j of slot j as

$$\Pi_j = \sum_{l,k \in \Phi_j, l \neq k} \pi_{k,l}, j = 1, \cdots, J. \tag{3.48}$$

We are now ready to present the proposed COG algorithm for given J:

1. Compute the sum-CCI of MT k to all other MTs, i.e., $\sum_{l \neq k}^{K} \pi_{k,l}$, $k = 1, \cdots, K$.

2. Assign the J MTs corresponding to the first J largest sum-CCI each to an individual time slot.

3. Each of the remaining $K - J$ MTs is successively assigned to one of the J slots, which has the minimum increase of Π_j, $j = 1, ..., K$.

3.6.4 Algorithm for Problem (WSTREMin)

Combining the results in Section 3.6.2 and Section 3.6.3, our complete algorithm for problem (WSTREMin) based on TS-OFDMA is summarized in Table 3.3.

Next, we analyze the complexity of the proposed algorithm in Table 3.3. For step 1), the time complexity of the two extreme cases have been analyzed in Section 3.4 and Section 3.5, which are of order KN and $K^4 + N^4 + K^3N^3$, respectively. Therefore, the time complexity of step 1) is $\mathcal{O}(K^4 + N^4 + K^3N^3)$. For step 2), in each iteration with given $1 < J < K$, the computation time is dominated by solving

Table 3.3: Algorithm for Solving Problem (WSTREMin)

1. Solve the two extreme cases, i.e., $J = K$ in (3.43) and $J = 1$ in (3.44), as described in Section 3.6.2.
2. For $1 < J < K$
   a) Obtain the MT grouping by the COG algorithm.
   b) Obtain $\mathcal{B}_1$ and $\mathcal{B}_2$ according to (3.45) and (3.46), respectively.
   c) For slots in $\mathcal{B}_1$, solve one single WSTREMin problem similarly as for the case of $J = K$; for slots in $\mathcal{B}_2$, perform WSTREMin in each slot separately similarly as for the case of $J = 1$.
3. Identify the optimal J and MT grouping as the one resulting in the smallest WSTRE, and obtain its corresponding time and power allocations from the previous two steps.

separate WSTREMin problems for slots in $\mathcal{B}_1$ and $\mathcal{B}_2$ in step c), which depends on the MT grouping obtained by the COG algorithm. However, from the complexity analysis of the two extreme cases, it is observed that the worst case in terms of computation complexity is to assign as many as MTs into one slot, i.e., there are $J-1$ slots in $\mathcal{B}_1$ but one slot in $\mathcal{B}_2$, which is of order $(K-J+1)^4+N^4+(K-J+1)^3N^3$. Therefore, the overall worst case complexity of the algorithm in Table 3.3 is $\mathcal{O}(KN^4 + \sum_{J=1}^{K-1}(K - J + 1)^4 + (K - J + 1)^3N^3)$, which is upper bounded by $\mathcal{O}(K^5 + K^4N^3 + KN^4)$.

3.7 Time-Constrained Optimization

We note that the total transmission time T is practically bounded by $T \leq T_{\max}$, where $T_{\max}$ may be set as the channel coherence time or the maximum transmission delay constraint, whichever is smaller. In this section, we highlight the consequences of introducing the maximum transmission time constraint, and discuss in details how the

obtained results in the previous sections can be extended to the case of time-constrained optimization.

Note that the maximum transmission time constraint, i.e., $T \leq T_{\max}$, does not affect the solvability of problem (TEMin) in Section 3.5 and problem (WSTREMin) under TS-OFDMA in Section 3.6. However, in the case of maximum time constraint, the optimality of the TDMA structure for WSREMin may not hold in general (for example, when $\sum_{k=1}^{K} t_k^* > T_{\max}$). However, Proposition 3.4.1 reveals that orthogonalizing MTs' transmission in time is beneficial for WSREMin, which is useful even for the case of time-constrained optimization, since we may still assume D-TDMA structure to approximately solve problem (WSREMin). In the rest of this section, we discuss how to solve problems (WSREMin-TDMA), (TEMin) and (WSTREMin) under TS-OFDMA in the case with maximum time constraint $T_{\max}$, which are termed as (WSREMin-TDMA-T), (TEMin-T) and (WSTREMin-T), respectively.

First, for problem (WSREMin-TDMA-T), it is observed that adding $\sum_{k=1}^{K} t_k \leq T_{\max}$ does not affect its convexity after the same change of variables as problem (WSREMin-TDMA). Furthermore, the water-filling structure presented in Theorem 3.4.1 still holds for problem (WSREMin-TDMA-T). Therefore, the algorithm in Table 3.1 can still be applied to solve problem (WSREMin-TDMA-D) with one additional step of bisection search over the maximum transmission time to ensure that it is no larger than $T_{\max}$. On the other hand, the feasibility of problem (WSREMin-TDMA-T) can be verified by setting $\alpha_k = 1/P_{r,c}, \forall k \in \mathcal{K}$ with the algorithm in Table 3.1. If the obtained optimal value is smaller than $T_{\max}$, it is feasible; otherwise, it is infeasible. It should be noted that, for the case with $T_{\max}$, problem (WSTREMin-T) being feasible does not guarantee the feasibility of problem (WSREMin-TDMA-T) due to the prior assumed D-TDMA scheme.

For problem (TEMin-T), the maximum transmission time constraint does not affect its solvability compared with problem (TEMin), where the same decomposition method can be applied since problem (TEMin-2) with $T \leq T_{\max}$ added, termed as (TEMin-2-T), is still convex. As a result, Lagrange duality method can again be

applied to solve this problem optimally. Besides, the feasibility of (TEMin-T) can be checked by solving problem (TEMin-1) in Section 3.5 with $T = T_{\max}$ using the algorithm in Table B.1. If the obtained optimal value is smaller than the average power limit P_{avg}, problem (TEMin-T) is feasible; otherwise, it is infeasible.

Finally, for problem (WSTREMin-T) under TS-OFDMA with given J and MT grouping (by the same suboptimal MT grouping algorithm as proposed in Section 3.6.3), time allocation needs to be optimized among different time slots to ensure the new maximum transmission time constraint. Since it can be shown that the optimal value of problem (WSREMin-TDMA-T) or (TEMin-T) is convex with respect to $T_{\max}$ similarly as that in Lemma 3.5.1, gradient based method, e.g., Newton method [32], can be applied. Last, the feasibility of problem (WSTREMin-T) for given J and MT grouping under TS-OFDMA can be checked by setting $\alpha_0 = 0$ and $\alpha_k = 1/P_{r,c}, \forall k \in \mathcal{K}$. If the obtained total transmission time is smaller than $T_{\max}$, it is feasible; otherwise, it is infeasible.

3.8 Numerical Results

In this section, we present simulation results to verify our theoretical analysis and demonstrate the tradeoffs in energy consumption at the BS and MTs. It is assumed that there are $K = 4$ active MTs with distances to the BS as 400, 600, 800 and 700 meters, and data requirements $\bar{Q}_k$ as 8.5, 11.5, 14.5 and 17.5 Kbits, respectively. The total number of SCs N is set to be 16, and the bandwidth of each SC W is 20kHz. Independent multipath fading channels, each with six equal-energy independent consecutive time-domain taps, are assumed for each transmission link between each pair of the BS and MTs. Each tap coefficient consists of both small-scale fading and distance dependent attenuation components. The small-scale fading is assumed to be Rayleigh distributed with zero mean and unit variance, and the distance-dependent attenuation has a path-loss exponent equal to four [52]. The power consumption of each MT, when turned on, is set to be 0.5W. For the BS, we assume a constant non-transmission related power of $P_{t,c} = 20$W and an average transmit power of

$P_{\text{avg}} = 30$W. We also set $\alpha_k = 1$ for all MTs, i.e., we consider the sum-energy consumption of all MTs. Finally, we set the receiver noise spectral density as $N_0 = -174$ dBm/Hz, which corresponds to a typical thermal noise at room temperature.

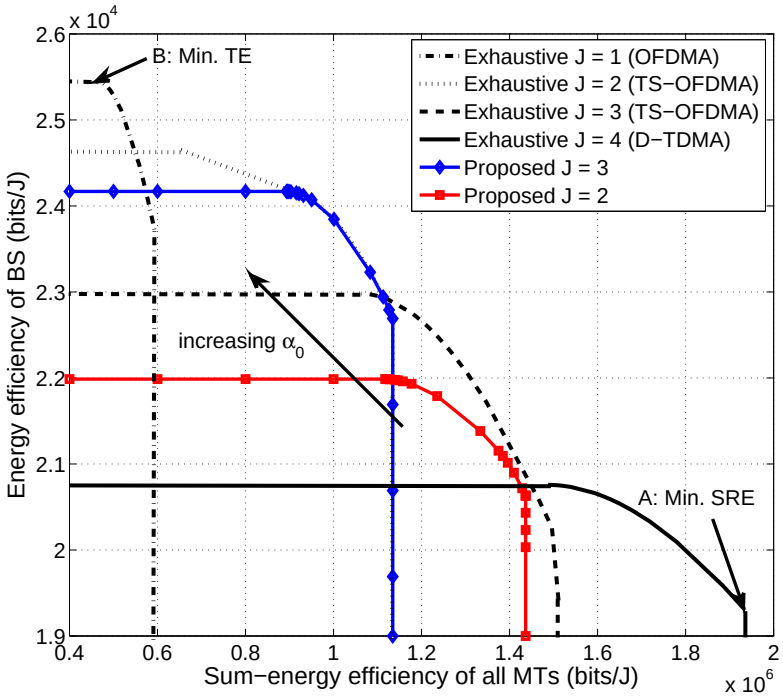


Figure 3.3: Energy efficiency tradeoffs with different transmission schemes. The points "Min. SRE" and "Min. TE" represent the results obtained by methods in Section 3.4 and Section 3.5, respectively

Fig. 3.3 shows the energy efficiency tradeoffs (in bits/joule) between BS and MTs[2] with various values of J, which is the number of orthogonal time slots in our proposed TS-OFDMA scheme in Section 3.6, and by varying the value of the BS energy consumption weight α_0 for each given J. In particular, the curves Exhaustive $J = 2$ and $J = 3$ are obtained by exhaustively searching all possible MT groupings, which serve as performance benchmark. The curves Proposed $J = 2$ and $J = 3$ are obtained by the COG algorithm presented in Section 3.6.3. The performance gap between the proposed algorithm and the benchmark is the price paid for lower

[2]For the ease of illustration, we treat the K MTs as an ensemble, whose energy efficiency is defined as the ratio of sum-data received and sum-energy consumed at all MTs, i.e., $\sum_{k=1}^{K} \bar{Q}_k / \sum_{k=1}^{K} E_{r,k}$.

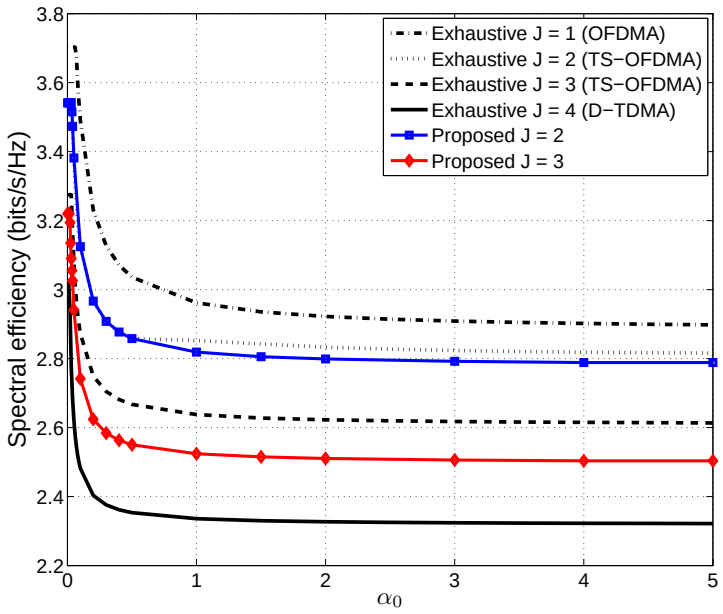


Figure 3.4: Spectral efficiency comparison with different transmission schemes

computation complexity. It is observed that as α_0 increases, the energy efficiency of BS increases and that at MTs decreases, respectively, for each J. It is easy to identify two boundary points of these tradeoff curves, namely, point A (on the curve of $J = 4$ with $\alpha_0 = 0$) and B (on the curve of $J = 1$ with $\alpha_0 = \infty$) correspond to the two special cases of TS-OFDMA, i.e., D-TDMA in Section 3.4 and OFDMA in Section 3.5, respectively. By comparing the two boundary points, we observe that if BS's energy efficiency is reduced by 25%, then the sum-energy efficiency of MTs can be increased by around three times. Furthermore, it is observed that more flexible energy efficiency tradeoffs between BS and MTs than those in the cases of $J = 1$ and $J = 4$ can be achieved by applying the proposed TS-OFDMA transmission scheme with $J = 2$ or 3.

Next, in Fig. 3.4, we show the spectral efficiency (in bits/s/Hz) of the considered multiuser downlink system over α_0 with different values of J, which is defined as the total amount of transmitted data per unit time and bandwidth, i.e., $\sum_{k=1}^{K} \bar{Q}_k/TNW$. First, it is observed that the spectral efficiency decreases and finally converges as α_0 increases for each value of J. The decreasing of spectral efficiency is the price to be

paid for less energy consumption of BS (c.f. Fig. 3.3), which is due to the increase of the required transmission time T and hence results in more energy consumption of MTs. It is also observed that for a given α_0, the spectral efficiency decreases as J increases, which is intuitively expected as $J = 1$, i.e., OFDMA, is known to be most spectrally efficient for multiuser downlink transmission.

Remark 3.8.1. For the proposed TS-OFDMA scheme with given user grouping, the power consumption at MTs can be mathematically interpreted as extra non-transmission related power at the BS. As a result, problem (WSTREMin) can be treated as an equivalent merely transmitter-side energy minimization problem. As α_0 increases, with proper normalization, it can be verified that the effective non-transmission related power decreases. Therefore, the optimal (most energy-efficient) transmission time T will increase [35], which results in the decreasing of the spectral efficiency as shown in Fig. 3.4. Furthermore, as J increases (less MTs in each slot), MTs have more opportunity to be in the "off" mode to save energy, while on the contrary, BS has less opportunity to gain from so-called multiuser diversity [30] to improve spectral efficiency. Consequently, the results in Fig. 3.3 and Fig. 3.4 are expected.

3.9 Chapter Summary

In this chapter, for cellular systems under an OFDM-based downlink communication setup, we have characterized the tradeoffs in minimizing the BS's versus MTs' energy consumptions by investigating a weighted-sum transmitter and receiver joint energy minimization (WSTREMin) problem, subject to an average transmit power constraint at the BS and data requirements of individual MTs. Two extreme cases, i.e., weighted-sum receiver-side energy minimization (WSREMin) for the MTs and transmitter-side energy minimization (TEMin) for the BS, were solved separately. It was shown that Dynamic TDMA (D-TDMA) is the optimal transmission strategy for WSREMin, while OFDMA is optimal for TEMin. Based on the obtained resource allocation solutions in these two cases, we proposed a new multiple access scheme

termed Time-Slotted OFDMA (TS-OFDMA) transmission scheme, which includes D-TDMA and OFDMA as special cases, to achieve more flexible energy consumption tradeoffs between the BS and MTs. In the next chapter, besides continuing to investigate both transmitter- and receiver-side energy consumption, the joint DL and UL transmission design with multiple BSs cooperatively serving multiple MTs is studied to further improve the overall performance.

Chapter 4

Downlink and Uplink Energy Minimization through User Association and Beamforming in Cloud Radio Access Network

4.1 Introduction

In this chapter, we extend our study in Chapter 3 to the case of joint DL and UL communications with multiple transmitters and receivers. In particular, we consider the emerging cloud radio access network (C-RAN), in which densely deployed APs are empowered by cloud computing to cooperatively support distributed MTs, to improve mobile data rates. We propose a joint DL and UL MT-AP association and beamforming design to coordinate interference in the C-RAN for energy minimization, a problem which is shown to be NP hard. Leveraging the celebrated UL-DL duality result, we show that by establishing a virtual DL transmission for the original UL transmission, the joint DL and UL optimization problem can be converted to an equivalent DL problem in C-RAN with two inter-related subproblems for the original and virtual DL transmissions, respectively. Based on this transformation, two efficient algorithms for joint DL and UL MT-AP association and beamforming design are proposed.

4.2 Literature Review

As introduced in Chapter 1, C-RAN has recently been proposed and drawn a great deal of attention [5]. However, with densely deployed APs, several new challenges arise in C-RAN. First, close proximity of many active APs results in increased interference, and hence the transmit power of APs and/or MTs needs to be increased to meet any given quality of service (QoS). Second, the amount of energy consumed by a large number of active APs [14] as well as by the transport network to support high-capacity connections with the BBU pool [53] will also become considerable. Such facts motivate us to optimize the energy consumption in C-RAN, which is the primary concern of this chapter. In particular, both DL and UL transmissions are considered jointly. The studied C-RAN model consists of densely deployed APs jointly serving a set of distributed MTs, where CoMP based joint transmit/receive processing (beamforming) over all active APs is employed for DL/UL transmissions. Under this setup, we study a joint DL and UL MT-AP association and beamforming design problem to minimize the total energy consumption in the network subject to MTs' given DL and UL QoS requirements. The energy saving is achieved by optimally assigning MTs to be served by the minimal subset of active APs, finding the power levels to transmit at all MTs and APs, and finding the beamforming vectors to use at the multi-antenna APs.

This problem has not been investigated to date, and the closest prior studies are [54–58]. However, these studies have all considered MT association and/or active AP selection problems for various objectives from the DL perspective. Note that the MT association and/or active AP selection based on DL only may result in inefficient transmit power of MTs or even their infeasible transmit power in the UL considering various possible asymmetries between the DL and UL in terms of channel, traffic and hardware limitation. Furthermore, with users increasingly using applications with high-bandwidth UL requirements, UL transmission is becoming more important. For example, the upload speed required for full high definition (HD) 1080p Skype video calling is about 20 Mbps [59]. Therefore, we need to account for both DL and UL

transmissions in designing the MTs association and active AP selection scheme. The UL-only MT association problem has also been considered extensively in the literature [18, 60, 61]; however, their solutions are not applicable in the context of this work due to their assumption of one-to-one MT-AP association.

In terms of other related work, there have been many attempts to optimize the energy consumption in cellular networks, but only over a single dimension each time, e.g., power control [62], AP "on/off" control [24, 63, 64], and CoMP transmission [65, 66]. A judicious combination of these techniques should provide the best solution, and this is the direction of our work. Unfortunately, the considered joint DL and UL MT-AP association and beamforming design problem in this chapter involves integer programming and is NP hard as shown for a similar problem in [56, Theorem 1]. To tackle this difficulty, two different approaches, i.e., group-sparse optimization (GSO) and relaxed-integer programming (RIP), have been adopted in [55, 56] and [58], respectively, to solve a similar DL-only problem, where two polynomial-time algorithms were proposed and shown to achieve good performance through simulations. In particular, the GSO approach is motivated by the fact that in the C-RAN with densely deployed APs, only a small fraction of the total number of APs needs to be active for meeting all MTs' QoS. However, due to the new consideration of UL transmission in this chapter, we will show that the algorithms proposed in [55, 56, 58] cannot be applied directly to solve our problem, and therefore the methods derived in this chapter are important advances in this field.

4.3 System Model

We consider a densely deployed C-RAN [5, 67] consisting of N APs, denoted by the set $\mathcal{N} = \{1, \cdots, N\}$. The set of distributed APs jointly support K randomly located MTs, denoted by the set $\mathcal{K} = \{1, \cdots, K\}$, for both DL and UL communications. In this chapter, for the purpose of exposition, we consider linear precoding and decoding in the DL and UL, respectively, which is jointly designed at the BBU pool assuming the perfect channel knowledge for all MTs. The results in this chapter can be readily

extended to the case of more complex successive precoding/decoding, e.g., dirty-paper coding (DPC) [68] and multiuser detection with successive interference cancelation (SIC) [30], with fixed coding orders among the users. We also assume that each AP n, $n \in \mathcal{N}$, is equipped with $M_n \geq 1$ antennas, and all MTs are each equipped with one antenna. It is further assumed that there exist ideal low-latency backhaul transport links with sufficiently large capacity (e.g. optical fiber) connecting the set of APs to the BBU pool, which performs all the baseband signal processing and transmission scheduling for all APs. The centralized architecture results in efficient coordination of the transmission/reception among all the APs, which can also be opportunistically utilized depending on the traffic demand.

We consider a quasi-static fading environment, and denote the channel vector in the DL from AP n to MT i and that in the UL from MT i to AP n as $\mathbf{h}_{i,n}^H \in \mathbb{C}^{1\times M_n}$ and $\mathbf{g}_{i,n} \in \mathbb{C}^{M_n\times 1}$, respectively. Let the vector consisting of the channels from all the APs to MT i and that consisting of the channels from MT i to all the APs be $\mathbf{h}_i^H = \left[\mathbf{h}_{i,1}^H, \cdots, \mathbf{h}_{i,N}^H\right]$ and $\mathbf{g}_i = \left[\mathbf{g}_{i,1}^T, \cdots, \mathbf{g}_{i,N}^T\right]^T$, respectively. There are two main techniques for separating DL and UL transmissions on the same physical transmission medium, i.e., time-division duplex (TDD) and frequency-division duplex (FDD). If TDD is assumed, channel reciprocity is generally assumed to hold between DL and UL transmissions, which means that the channel vector $\mathbf{g}_i$ in the UL is merely the transpose of that $\mathbf{h}_i^H$ in the DL, i.e., $\mathbf{g}_i = \mathbf{h}_i^*, \forall i \in \mathcal{K}$. However, if FDD is assumed, $\mathbf{h}_i$'s and $\mathbf{g}_i$'s are different in general.

4.3.1 Downlink Transmission

In DL transmission, the transmitted signal from all APs can be generally expressed as

$$\mathbf{x}^{\text{DL}} = \sum_{i=1}^{K} \mathbf{w}_i^{\text{DL}} s_i^{\text{DL}} \tag{4.1}$$

where $\mathbf{w}_i^{\text{DL}} \in \mathbb{C}^{M\times 1}$ is the beamforming vector for all APs to cooperatively send one single stream of data signal s_i^{DL} to MT i, which is assumed to be a complex random variable with zero mean and unit variance. Note that $\sum_{n=1}^{N} M_n = M$. Then, the

transmitted signal from AP n can be expressed as

$$\mathbf{x}_n^{\mathrm{DL}} = \sum_{i=1}^{K} \mathbf{w}_{i,n}^{\mathrm{DL}} s_i^{\mathrm{DL}}, \;\; n = 1, \cdots, N \tag{4.2}$$

where $\mathbf{w}_{i,n}^{\mathrm{DL}} \in \mathbb{C}^{M_n \times 1}$ is the nth block component of $\mathbf{w}_i^{\mathrm{DL}}$, corresponding to the transmit beamforming vector at AP n for MT i. Note that $\mathbf{x}^{\mathrm{DL}} = [\left(\mathbf{x}_1^{\mathrm{DL}}\right)^T, \cdots, \left(\mathbf{x}_N^{\mathrm{DL}}\right)^T]^T$ and $\mathbf{w}_i^{\mathrm{DL}} = [\left(\mathbf{w}_{i,1}^{\mathrm{DL}}\right)^T, \cdots, \left(\mathbf{w}_{i,N}^{\mathrm{DL}}\right)^T]^T$, $i = 1, \cdots, K$. From (4.2), the transmit power of AP n in DL is obtained as

$$p_n^{\mathrm{DL}} = \sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\mathrm{DL}}\|^2, \;\; n = 1, \cdots, N. \tag{4.3}$$

We assume that there exists a maximum transmit power constraint for each AP n, i.e.,

$$p_n^{\mathrm{DL}} \leq P_{n,\max}^{\mathrm{DL}}, \;\; n = 1, \cdots, N. \tag{4.4}$$

The received signal at the ith MT is then expressed as

$$y_i^{\mathrm{DL}} = \mathbf{h}_i^H \mathbf{w}_i^{\mathrm{DL}} s_i^{\mathrm{DL}} + \sum_{j \neq i}^{K} \mathbf{h}_i^H \mathbf{w}_j^{\mathrm{DL}} s_j^{\mathrm{DL}} + z_i^{\mathrm{DL}}, \;\; i = 1, \cdots, K \tag{4.5}$$

where z_i^{DL} is the receiver noise at MT i, which is assumed to be a circularly symmetric complex Gaussian (CSCG) random variable with zero mean and variance σ^2, denoted by $z_i^{\mathrm{DL}} \sim \mathcal{CN}(0, \sigma^2)$. Treating the interference as noise, the signal-to-interference-plus-noise ratio (SINR) in DL for MT i is given by

$$\mathrm{SINR}_i^{\mathrm{DL}} = \frac{|\mathbf{h}_i^H \mathbf{w}_i^{\mathrm{DL}}|^2}{\sum_{j \neq i} |\mathbf{h}_i^H \mathbf{w}_j^{\mathrm{DL}}|^2 + \sigma^2}, \;\; i = 1, \cdots, K. \tag{4.6}$$

4.3.2 Uplink Transmission

In UL transmission, the transmitted signal from MT i is given by

$$x_i^{\mathrm{UL}} = \sqrt{p_i^{\mathrm{UL}}} s_i^{\mathrm{UL}}, \;\; i = 1, \cdots, K \tag{4.7}$$

where p_i^{UL} denotes the transmit power of MT i, and s_i^{UL} is the information bearing signal which is assumed to be a complex random variable with zero mean and unit variance. With $P_{i,\max}^{\mathrm{UL}}$ denoting the transmit power limit for MT i, it follows that

$$p_i^{\mathrm{UL}} \leq P_{i,\max}^{\mathrm{UL}}, \;\; i = 1, \cdots, K. \tag{4.8}$$

The received signal at all APs is then expressed as

$$\mathbf{y}^{\mathrm{UL}} = \sum_{i=1}^{K} \mathbf{g}_i \sqrt{p_i^{\mathrm{UL}}} s_i^{\mathrm{UL}} + \mathbf{z}^{\mathrm{UL}} \tag{4.9}$$

where $\mathbf{z}^{\mathrm{UL}} \in \mathbb{C}^{M\times 1}$ denotes the receiver noise vector at all APs consisting of independent CSCG random variables each distributed as $\mathcal{CN}(0, \sigma^2)$. Let $\mathbf{v}_i^{\mathrm{UL}} \in \mathbb{C}^{M\times 1}$ denote the receiver beamforming vector used to decode s_i^{UL} from MT i. Then the SINR in UL for MT i after applying $\mathbf{v}_i^{\mathrm{UL}}$ is given by

$$\mathrm{SINR}_i^{\mathrm{UL}} = \frac{p_i^{\mathrm{UL}} |(\mathbf{v}_i^{\mathrm{UL}})^T \mathbf{g}_i|^2}{\sum_{j\neq i} p_j^{\mathrm{UL}} |(\mathbf{v}_i^{\mathrm{UL}})^T \mathbf{g}_j|^2 + \sigma^2 \|\mathbf{v}_i^{\mathrm{UL}}\|^2}, \quad i = 1, \cdots, K. \tag{4.10}$$

Let $\mathbf{v}_{i,n}^{\mathrm{UL}} \in \mathbb{C}^{M_n\times 1}$ denote the nth block component in $\mathbf{v}_i^{\mathrm{UL}}$, corresponding to the receive beamforming vector at AP n for MT i. We thus have $\mathbf{v}_i^{\mathrm{UL}} = [\left(\mathbf{v}_{i,1}^{\mathrm{UL}}\right)^T, \cdots, \left(\mathbf{v}_{i,N}^{\mathrm{UL}}\right)^T]^T$, $i = 1, \cdots, K$.

4.3.3 Energy Consumption Model

The total energy consumption in the C-RAN comprises of the energy consumed by all APs and all MTs. From (4.1) and (4.7), the total transmit power of all APs in DL and that of all MTs in UL can be expressed as $P_t^{\mathrm{DL}} = \sum_{i=1}^{K} \|\mathbf{w}_i^{\mathrm{DL}}\|^2$ and $P_t^{\mathrm{UL}} = \sum_{i=1}^{K} p_i^{\mathrm{UL}}$, respectively.

Besides the static power consumption at each AP n due to e.g., real-time A/D and D/A processing, denoted as $P_{s,n}, \forall n \in \mathcal{N}$, in C-RAN with centralized processing, the extensive use of high-capacity backhaul links to connect all APs with the BBU pool makes the power consumption of the transport network no more negligible [53]. For example, consider the passive optical network (PON) to implement the backhaul transport network [69]. The PON assigns an optical line terminal (OLT) to connect to a set of associated optical network units (ONUs), which coordinate the set of transport links connecting all the APs to the BBU pool, each through a single fiber. For simplicity, the resulting power consumption in the PON can be modeled as [69]

$$P_{\mathrm{PON}} = P_{\mathrm{OLT}} + \sum_{n=1}^{N} P_{\mathrm{ONU},n} \tag{4.11}$$

where P_{OLT} and $P_{\text{ONU},n}$ are both constant and denote the power consumed by the OLT and the transport link associated with AP n, respectively.

Moreover, we consider that for energy saving, some APs and their associated transport links can be switched into sleep mode [31, 69] (compared with active mode) with negligible power consumption[1]; thus, the total static power consumption of AP n, denoted by $P_{c,n} = P_{s,n} + P_{\text{ONU},n}$, $n \in \mathcal{N}$, can be saved if AP n and its associated transport link are switched into sleep mode for both transmission in DL and UL. For convenience, we express the total static power consumption of all active APs as

$$P_c = \sum_{n=1}^{N} \mathbf{1}_n \left(\{\mathbf{w}_{i,n}^{\text{DL}}\}, \{\mathbf{v}_{i,n}^{\text{UL}}\} \right) P_{c,n} \tag{4.12}$$

where $\mathbf{1}_n(\cdot)$, $n \in \mathcal{N}$, is an indicator function for AP n, which is defined as

$$\mathbf{1}_n \left(\{\mathbf{w}_{i,n}^{\text{DL}}\}, \{\mathbf{v}_{i,n}^{\text{UL}}\} \right) = \begin{cases} 0 & \text{if } \mathbf{w}_{i,n}^{\text{DL}} = \mathbf{v}_{i,n}^{\text{UL}} = \mathbf{0}, \forall i \in \mathcal{K} \\ 1 & \text{otherwise.} \end{cases} \tag{4.13}$$

Note that in practical PON systems, the OLT in general cannot be switched into sleep mode as it plays the role of distributor, arbitrator, and aggregator of the transport network, which has a typical fixed power consumption of $P_{\text{OLT}} = 20$W [69]. We thus ignore P_{OLT} since it is only a constant. From (4.13), MT i is associated with an active AP n if its corresponding transmit and/or receive beamforming vector at AP n is nonzero, i.e., $\mathbf{w}_{i,n}^{\text{DL}} \neq \mathbf{0}$ and/or $\mathbf{v}_{i,n}^{\text{UL}} \neq \mathbf{0}$. Under this setup, it is worth pointing out that each MT i is allowed to connect with two different sets of APs for DL and UL transmissions, respectively, e.g., $\mathbf{w}_{i,n}^{\text{DL}} \neq \mathbf{0}$ but $\mathbf{v}_{i,n}^{\text{UL}} = \mathbf{0}$ for some $n \in \mathcal{N}$, which is promising to be implemented in next generation cellular networks [70]. Furthermore, from (4.13), AP n could be switched into sleep mode only if it does not serve any MT.

We aim to minimize the total energy consumption in the C-RAN, including that due to transmit power of all MTs (but ignoring any static power consumption of MT terminals) as well as that due to transmit power and static power of all active APs.

[1]It is assumed that when the AP is in the sleep mode, it acts as a passive node and listens to the pilot signals transmitted from the MTs for channel estimation, which consumes negligible power compared with being in the active mode for data transmission. It is further assumed that each AP can switch between the active and sleep modes frequently.

Therefore, we consider the following weighted sum-power as our design metric:

$$P_{\text{total}}\left(\{\mathbf{w}_i^{\text{DL}}\},\{\mathbf{v}_i^{\text{UL}}\}\right)=\left(\sum_{n=1}^{N}\mathbf{1}_n\left(\{\mathbf{w}_{i,n}^{\text{DL}}\},\{\mathbf{v}_{i,n}^{\text{UL}}\}\right)P_{c,n}+\sum_{i=1}^{K}\|\mathbf{w}_i^{\text{DL}}\|^2\right)+\lambda\left(\sum_{i=1}^{K}p_i^{\text{UL}}\right) \tag{4.14}$$

where $\lambda \geq 0$ is a weight to trade off between the total energy consumptions between all the active APs and all MTs.

4.4 Problem Formulation and Two Solution Approaches

To minimize the weighted power consumption in (4.14), we jointly optimize the DL and UL MT-AP association and transmit/receive beamforming by considering the following problem.

$$(\text{P1}): \underset{\{\mathbf{w}_i^{\text{DL}}\},\{\mathbf{v}_i^{\text{UL}}\},\{p_i^{\text{UL}}\}}{\texttt{Min.}} \quad P_{\text{total}}\left(\{\mathbf{w}_i^{\text{DL}}\},\{\mathbf{v}_i^{\text{UL}}\}\right) \tag{4.15}$$

$$\texttt{s.t.} \quad \text{SINR}_i^{\text{DL}} \geq \gamma_i^{\text{DL}}, \forall i \in \mathcal{K} \tag{4.16}$$

$$\text{SINR}_i^{\text{UL}} \geq \gamma_i^{\text{UL}}, \forall i \in \mathcal{K} \tag{4.17}$$

$$p_n^{\text{DL}} \leq P_{n,\max}^{\text{DL}}, \ \ \forall n \in \mathcal{N} \tag{4.18}$$

$$0 \leq p_i^{\text{UL}} \leq P_{i,\max}^{\text{UL}}, \forall i \in \mathcal{K} \tag{4.19}$$

where γ_i^{DL} and γ_i^{UL} are the given SINR requirements of MT i for the DL and UL transmissions, respectively. In the rest of this chapter, the constraints in (4.18) and (4.19) are termed per-AP and per-MT power constraints, respectively. Problem (P1) can be shown to be non-convex due to the implicit integer programming involved due to indicator function $\mathbf{1}_n(\cdot)$'s in the objective. Prior to solving problem (P1), we first need to check its feasibility. Since the DL and UL transmissions are coupled only by the objective function in (4.15), the feasibility of problem (P1) can be checked by considering two separate feasibility problems: one for the DL and the other for the UL, which have both been well studied in the literature [71] and thus the details are omitted

here for brevity. For the rest of this chapter, we assume that problem (P1) is always feasible if all APs are active.

As mentioned in Section 4.2, the problem of joint MT-AP association and transmit beamforming subject to MTs' QoS and per-AP power constraints for power minimization in the DL-only transmission has been recently studied in [55, 56, 58] using the approaches of GSO and RIP, respectively, where two different polynomial-time algorithms were proposed and shown to both achieve good performance by simulations. In contrast, problem (P1) in this chapter considers both DL and UL transmissions to address possible asymmetries between the DL and UL in terms of channel realization, traffic load and hardware limitation. Furthermore, considering that MTs are usually powered by finite-capacity batteries as compared to APs that are in general powered by the electricity grid, we study the power consumption tradeoffs between APs and MTs by minimizing the weighted sum-power $P_{\text{total}}\left(\{\mathbf{w}_i^{\text{DL}}\}, \{\mathbf{v}_i^{\text{UL}}\}\right)$ in (P1). Therefore, the problems considered in [55, 56, 58] can be treated as special cases of (P1).

In the following, we show that due to the new consideration of UL transmission, the algorithms proposed in [55, 56, 58] based on GSO and RIP for solving the DL optimization cannot be applied directly to solve (P1), which thus motivates us to find a new method to resolve this issue in Section 4.5.

4.4.1 Group-Sparse Optimization based Solution

Given the fact that the static power, i.e., $P_{c,n}$, is in practice significantly larger than the transmit power at each AP n, to minimize the total network energy consumption [69, 72], it is conceivable that for the optimal solution of (P1) only a subset of N APs should be active. As a result, a "group-sparse" property can be inferred from the following concatenated beamforming vector:

$$\left[[\hat{\mathbf{w}}_1^{\text{DL}}, \hat{\mathbf{v}}_1^{\text{UL}}], \cdots, [\hat{\mathbf{w}}_N^{\text{DL}}, \hat{\mathbf{v}}_N^{\text{UL}}]\right] \tag{4.20}$$

in which $\hat{\mathbf{w}}_n^{\text{DL}} = \left[(\mathbf{w}_{1,n}^{\text{DL}})^T, \cdots, (\mathbf{w}_{K,n}^{\text{DL}})^T\right]$ and $\hat{\mathbf{v}}_n^{\text{UL}} = \left[(\mathbf{v}_{1,n}^{\text{UL}})^T, \cdots, (\mathbf{v}_{K,n}^{\text{UL}})^T\right]$, $n = 1, \cdots, N$, i.e., the beamforming vectors are grouped according to their associated APs

. If AP n is in the sleep mode, its corresponding block $[\hat{\mathbf{w}}_n^{\mathrm{DL}}, \hat{\mathbf{v}}_n^{\mathrm{UL}}]$ in (4.20) needs to be zero. Consequently, the fact that a small subset of deployed APs is selected to be active implies that the concatenated beamforming vector in (4.20) should contain only a very few non-zero block components.

One well-known approach to enforce desired group sparsity in the obtained solutions for optimization problems is by adding to the objective function an appropriate penalty term. The widely used group sparsity enforcing penalty function, which was first introduced in the context of the group least-absolute selection and shrinkage operator (LASSO) problem [73], is the mixed $\ell_{1,2}$ norm. In our case, such a penalty is expressed as

$$\sum_{n=1}^{N} \left\| [\hat{\mathbf{w}}_n^{\mathrm{DL}}, \hat{\mathbf{v}}_n^{\mathrm{UL}}] \right\|. \tag{4.21}$$

The $\ell_{1,2}$ norm in (4.21), similar to ℓ_1 norm, offers the closest convex approximation to the ℓ_0 norm over the vector consisting of ℓ_2 norms $\left\{ \left\| [\hat{\mathbf{w}}_n^{\mathrm{DL}}, \hat{\mathbf{v}}_n^{\mathrm{UL}}] \right\| \right\}_{n=1}^{N}$, implying that each $\left\| [\hat{\mathbf{w}}_n^{\mathrm{DL}}, \hat{\mathbf{v}}_n^{\mathrm{UL}}] \right\|$ is desired to be set to zero to obtain group sparsity.

More generally, the mixed $\ell_{1,p}$ norm has also been shown to be able to recover group sparsity with $p > 1$ [74], among which the $\ell_{1,\infty}$ norm, defined as

$$\sum_{n=1}^{N} \max \left(\max_{i,j} \left| w_{i,n}^{\mathrm{DL}}(j) \right|, \max_{i,j} \left| v_{i,n}^{\mathrm{UL}}(j) \right| \right) \tag{4.22}$$

has been widely used [75]. Compared with $\ell_{1,2}$ norm, $\ell_{1,\infty}$ norm has the potential to obtain more sparse solution but may lead to undesired solution with components of equal magnitude. In this chapter, we focus on the $\ell_{1,2}$ norm in (4.21) for our study. We will compare the performance of $\ell_{1,2}$ and $\ell_{1,\infty}$ norms by simulations in Section 4.6.

According to [54–56], at first glance it seems that using the $\ell_{1,2}$ norm, problem (P1) can be approximately solved by replacing the objective function with

$$\sum_{n=1}^{N} \beta_n \sqrt{\sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\mathrm{DL}}\|^2 + \|\mathbf{v}_{i,n}^{\mathrm{UL}}\|^2} + \sum_{i=1}^{K} \|\mathbf{w}_i^{\mathrm{DL}}\|^2 + \lambda \sum_{i=1}^{K} p_i^{\mathrm{UL}} \tag{4.23}$$

where $\sum_{n=1}^{N} \beta_n \sqrt{\sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\mathrm{DL}}\|^2 + \|\mathbf{v}_{i,n}^{\mathrm{UL}}\|^2}$ can be treated as a convex relaxation of the indicator functions in (4.14), and $\beta_n \geq 0$ indicates the relative importance of the

penalty term associated with AP n. However, problem (P1) with (4.23) as the objective function is still non-convex due to the constraints in (4.16) and (4.17). Furthermore, since the UL receive beamforming vector $\mathbf{v}_i^{\mathrm{UL}}$'s can be scaled down to be arbitrarily small without affecting the UL SINR defined in (4.10), minimizing (4.23) directly will result in all $\mathbf{v}_i^{\mathrm{UL}}$'s going to zero. To be more specific, let $\hat{\mathbf{w}}_i^{\mathrm{DL}}$ and $\hat{\mathbf{v}}_i^{\mathrm{UL}}$ denote the optimal solution of problem (P1) with (4.23) as the objective function. Then, it follows that

$$\hat{\mathbf{v}}_i^{\mathrm{UL}} \approx \mathbf{0}, \ \ \forall i \in \mathcal{K} \tag{4.24}$$

and $\hat{\mathbf{w}}_i^{\mathrm{DL}}, \forall i \in \mathcal{K}$, preserves the "group-sparse" property where the non-zero block components correspond to the active APs. Two issues thus arise: first, the UL does not contribute to the selection of active APs; second, the set of selected active APs based on the DL only cannot guarantee the QoS requirements for the UL. As a result, the $\ell_{1,2}$ norm penalty term in (4.23) or more generally the $\ell_{1,p}$ norm penalty does not work for the joint DL and UL AP selection in our problem, and hence the algorithm proposed in [54–56], which involves only the DL transmit beamforming vector $\mathbf{w}_i^{\mathrm{DL}}$'s, cannot be modified in a straightforward way to solve our problem.

4.4.2 Relaxed-Integer Programming based Solution

Next, we reformulate problem (P1) by introducing a set of binary variable ρ_n's indicating the "active/sleep" state of each AP as follows.

$$(\mathrm{P2}): \underset{\{\mathbf{w}_i^{\mathrm{DL}}\},\{\mathbf{v}_i^{\mathrm{UL}}\},\{p_i^{\mathrm{UL}}\},\{\rho_n\}}{\texttt{Min.}} \ \left(\sum_{n=1}^{N} \rho_n P_{c,n} + \sum_{i=1}^{K} \|\mathbf{w}_i^{\mathrm{DL}}\|^2\right) + \lambda \left(\sum_{i=1}^{K} p_i^{\mathrm{UL}}\right) \tag{4.25}$$

$$\texttt{s.t.} \ \ (4.16), (4.17), (4.18), \text{and } (4.19) \tag{4.26}$$

$$\sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\mathrm{DL}}\|^2 + \|\mathbf{v}_{i,n}^{\mathrm{UL}}\|^2 \leq \rho_n (P_{n,\max}^{\mathrm{DL}} + \eta), \forall n \in \mathcal{N} \tag{4.27}$$

$$\rho_n \in \{0, 1\}, \forall n \in \mathcal{N} \tag{4.28}$$

where $\eta > 0$ is a constant with arbitrary value. Note that the active-sleep constraints in (4.27) are inspired by the well-known big-M method [76]: if $\rho_n = 0$, the constraint

(4.27) ensures that $\mathbf{w}_{i,n}^{\text{DL}} = \mathbf{v}_{i,n}^{\text{UL}} = \mathbf{0}, \forall i \in \mathcal{K}$; if $\rho_n = 1$, the constraint has no effect on $\mathbf{w}_{i,n}^{\text{DL}}$ and $\mathbf{v}_{i,n}^{\text{UL}}, \forall i \in \mathcal{K}$, as $P_{n,\max}^{\text{DL}} + \eta$ represents an upper bound on the term $\sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\text{DL}}\|^2 + \|\mathbf{v}_{i,n}^{\text{UL}}\|^2$. Notice that η can be chosen arbitrarily due to the scaling invariant property of UL receive beamforming vector $\mathbf{v}_i^{\text{UL}}$'s. With the active-sleep constraints in (4.27), the equivalence between problems (P1) and (P2) can be easily verified.

In [58], a similar problem to (P2) was studied corresponding to the case with only DL transmission. For problem (P2) without $\mathbf{v}_i^{\text{UL}}$ and $p_i^{\text{UL}}, \forall i \in \mathcal{K}$ and their corresponding constraints, the problem can be transformed to a convex second-order cone programming (SOCP) by relaxing the binary variable ρ_n as $\rho_n \in [0, 1], \forall n \in \mathcal{N}$. Under this convex relaxation, a BnC algorithm, which is a combination of the branch-and-bound (BnB) and the cutting plane (CP) methods [76], was proposed in [58] to solve the DL problem optimally. However, the computational complexity of BnB is prohibitive for large networks in practice, which grows exponentially with the number of APs. To obtain polynomial-time algorithm with near-optimal performance, in [58], the authors further proposed an incentive measure based heuristic algorithm to determine the set of active APs. The incentive measure reflects the importance of each AP to the whole network and is defined as the ratio of the total power received at all MTs to the total power expended for each AP.

However, with both DL and UL transmissions, it is observed that problem (P2) can no longer be transformed to a convex form by relaxing ρ_n's as continuous variables due to the constraints in (4.17). Furthermore, because of the scaling invariant property of UL receive beamforming vectors, solving the relaxed problem of (P2) will result in all $\mathbf{v}_i^{\text{UL}}$'s going to zero, similar to the case of GSO based solution. Particularly, the value of the relaxed indicator $\rho_n, \forall n \in \mathcal{N}$, will not be related to $\mathbf{v}_i^{\text{UL}}$'s, which in fact contributes to the penalty incurred in the objective due to the static power of AP n, i.e., $\rho_n P_{c,n}$. Finally, it is nontrivial to find an incentive measure that reflects the importance of each AP to both DL and UL transmissions.

4.5 Proposed Solution

In this section, we provide two efficient algorithms to approximately solve problem (P1) based on the GSO and RIP approaches, respectively.

4.5.1 Proposed Algorithm for Problem (P1) via GSO

First, we consider the approach of GSO and present a new method to address the joint DL and UL optimization. To obtain an efficient solution for problem (P1), we first assume that all the APs and MTs have infinite power budget, i.e., $P_{n,\max}^{\mathrm{DL}} = +\infty$, $\forall n \in \mathcal{N}$ and $P_{i,\max}^{\mathrm{UL}} = +\infty, \forall i \in \mathcal{K}$. The resulting problem is termed (P1-1). An equivalent reformulation of problem (P1-1) is then provided to overcome the receive beamforming scaling issue mentioned in Section 4.4. Then, we discuss the challenges of dealing with finite per-AP and per-MT power constraints and provide efficient methods to handle them.

4.5.1.1 Solution for problem (P1-1)

First, we consider the following transmit sum-power minimization problem in the UL:

$$\begin{aligned} \underset{\{\mathbf{v}_i^{\mathrm{UL}}\},\{p_i^{\mathrm{UL}}\}}{\mathtt{Min.}} \quad & \sum_{i=1}^{K} p_i^{\mathrm{UL}} \\ \mathtt{s.t.} \quad & \mathrm{SINR}_i^{\mathrm{UL}} \geq \gamma_i^{\mathrm{UL}}, \forall i \in \mathcal{K} \\ & p_i^{\mathrm{UL}} \geq 0, \forall i \in \mathcal{K}. \end{aligned} \tag{4.29}$$

From [77], it follows that Problem (4.29) can be solved in a virtual DL channel as

$$\begin{aligned} \underset{\{\mathbf{w}_i^{\mathrm{VDL}}\}}{\mathtt{Min.}} \quad & \sum_{i=1}^{K} \|\mathbf{w}_i^{\mathrm{VDL}}\|^2 \\ \mathtt{s.t.} \quad & \mathrm{SINR}_i^{\mathrm{VDL}} \triangleq \frac{|\mathbf{g}_i^H \mathbf{w}_i^{\mathrm{VDL}}|^2}{\sum_{j \neq i} |\mathbf{g}_i^H \mathbf{w}_j^{\mathrm{VDL}}|^2 + \sigma^2} \geq \gamma_i^{\mathrm{UL}}, \forall i \in \mathcal{K} \end{aligned} \tag{4.30}$$

where $\mathbf{w}_i^{\mathrm{VDL}} \in \mathbb{C}^{M \times 1}$ is the virtual DL transmit beamforming vector over N APs for MT i. Denote $(\mathbf{v}_i^{\mathrm{UL}})'$, $(p_i^{\mathrm{UL}})'$ and $(\mathbf{w}_i^{\mathrm{VDL}})'$, $i = 1, \cdots, K$ as the optimal solutions to problems (4.29) and (4.30), respectively. Then from [77] it follows that $(\mathbf{v}_i^{\mathrm{UL}})'$ and

$(\mathbf{w}_i^{\mathrm{VDL}})'$ can be set to be identical, $i = 1, \cdots, N$, and furthermore $\sum_{i=1}^{K}(p_i^{\mathrm{UL}})' = \sum_{i=1}^{K} \|(\mathbf{w}_i^{\mathrm{VDL}})'\|^2$.

By establishing a virtual DL transmission for the UL transmission based on the above UL-DL duality, we have the following lemma.

Lemma 4.5.1. *Problem (P1-1) is equivalent to the following problem.*

$$(P3): \quad \underset{\{\mathbf{w}_i^{DL}\},\{\mathbf{w}_i^{VDL}\}}{\texttt{Min.}} \quad \sum_{n=1}^{N} \mathbf{1}_n\left(\{\mathbf{w}_{i,n}^{DL}\},\{\mathbf{w}_{i,n}^{VDL}\}\right) P_{c,n} + \sum_{i=1}^{K}\|\mathbf{w}_i^{DL}\|^2 + \lambda\sum_{i=1}^{K}\|\mathbf{w}_i^{VDL}\|^2 \tag{4.31}$$

$$\texttt{s.t.} \quad SINR_i^{DL} \geq \gamma_i^{DL}, \forall i \in \mathcal{K} \tag{4.32}$$

$$SINR_i^{VDL} \geq \gamma_i^{UL}, \forall i \in \mathcal{K}. \tag{4.33}$$

Proof. For any given feasible solution to problem (P3), we can always find a corresponding feasible solution to problem (P1-1) achieving the same objective value as that of problem (P3), and vice versa, similar as [78, Proposition 1]; thus, problems (P1-1) and (P3) achieve the same optimal value with the same set of optimal DL/UL beamforming vectors. Lemma 4.5.1 is thus proved. □

Since problem (P3) is a DL-only problem that has the same "group-sparse" property as (P1-1), it can be approximately solved by replacing the objective function with

$$\sum_{n=1}^{N}\beta_n\sqrt{\sum_{i=1}^{K}\|\mathbf{w}_{i,n}^{\mathrm{DL}}\|^2 + \|\mathbf{w}_{i,n}^{\mathrm{VDL}}\|^2} + \sum_{i=1}^{K}\|\mathbf{w}_i^{\mathrm{DL}}\|^2 + \lambda\sum_{i=1}^{K}\|\mathbf{w}_i^{\mathrm{VDL}}\|^2. \tag{4.34}$$

Comparing (4.34) and (4.23), we have successfully solved the scaling issue of UL receive beamforming vector, $\mathbf{v}_i^{\mathrm{UL}}$'s, by replacing them with the equivalent DL transmit beamforming vector, $\mathbf{w}_i^{\mathrm{VDL}}$'s, since from (4.30) it follows that the virtual DL SINR of each MT i is no more scaling invariant to $\mathbf{w}_i^{\mathrm{VDL}}$'s.

Furthermore, since any arbitrary phase rotation of the beamforming vectors does not affect both (4.34) and the SINR constrains in (4.32) and (4.33), (P3) with (4.34) as

the objective function can be reformulated as a convex SOCP [32], which is given by

$$(\text{P4}):$$

$$\underset{\{\mathbf{w}_i^{\text{DL}}\},\{\mathbf{w}_i^{\text{VDL}}\},\{t_n\}}{\texttt{Min.}} \quad \sum_{n=1}^{N}\beta_n t_n + \sum_{i=1}^{K}\|\mathbf{w}_i^{\text{DL}}\|^2 + \lambda\sum_{i=1}^{K}\|\mathbf{w}_i^{\text{VDL}}\|^2 \tag{4.35}$$

$$\texttt{s.t.} \quad \left\| \begin{array}{c} \mathbf{h}_i^H\mathbf{W}^{\text{DL}} \\ \sigma \end{array} \right\| \leq \sqrt{1+\frac{1}{\gamma_i^{\text{DL}}}}\mathbf{h}_i^H\mathbf{w}_i^{\text{DL}}, \forall i \in \mathcal{K} \tag{4.36}$$

$$\left\| \begin{array}{c} \mathbf{g}_i^H\mathbf{W}^{\text{VDL}} \\ \sigma \end{array} \right\| \leq \sqrt{1+\frac{1}{\gamma_i^{\text{UL}}}}\mathbf{g}_i^H\mathbf{w}_i^{\text{VDL}}, \forall i \in \mathcal{K} \tag{4.37}$$

$$\sqrt{\sum_{i=1}^{K}\|\mathbf{w}_{i,n}^{\text{DL}}\|^2 + \|\mathbf{w}_{i,n}^{\text{VDL}}\|^2} \leq t_n, \forall n \in \mathcal{N} \tag{4.38}$$

where $\mathbf{W}^{\text{DL}} = [\mathbf{w}_1^{\text{DL}}, \cdots, \mathbf{w}_K^{\text{DL}}]$, $\mathbf{W}^{\text{VDL}} = [\mathbf{w}_1^{\text{VDL}}, \cdots, \mathbf{w}_K^{\text{VDL}}]$, and t_n's are auxiliary variables with $t_n = 0$ and $t_n > 0$ indicating that AP n is in active and sleep mode, respectively. Notice that without $\ell_{1,2}$ norm penalty or $\beta_n = 0, \forall n \in \mathcal{N}$, problem (P4) can be decomposed into two separate minimum-power beamforming design problems: one for the original DL transmission, and the other for the virtual DL transmission.

Remark 4.5.1. Conventionally, the UL transmit sum-power minimization problem, as in (4.29), has a convenient analytical structure and thus is computationally easier to handle, as compared to the DL minimum-power beamforming design problem, as in (4.30). Consequently, most existing studies in the literature have transformed the DL problem to its virtual UL formulation for convenience. The motivation of exploiting the reverse direction in this work, however, is to overcome the scaling issue of UL receive beamforming in GSO, so that we can solve the AP selection problem jointly for both DL and UL transmissions.

Next, we present the complete algorithm for problem (P1-1) based on GSO, in which three steps need to be performed sequentially.

1. *Identify the subset of active APs denoted as $\mathcal{N}_{on}$.* This can be done by iteratively solving problem (P4) with different β_n's. Notice that how to set the parameter β_n's in (P4) plays a key role in the resulting APs selection. To optimally set

the values of β_n's, we adopt an iterative method similar as in [79], shown as follows. In the lth iteration, $l \geq 1$, $t_n^{(l)}$'s are obtained by solving Problem (P4) with $\beta_n = \beta_n^{(l)}, \forall n \in \mathcal{N}$. The $\beta_n^{(l)}$'s are derived from the solution $t_n^{(l-1)}$'s of the $(l-1)$th iteration as

$$\beta_n^{(l)} = \frac{P_{c,n}}{t_n^{(l-1)} + \varepsilon}, n = 1, \cdots, N \tag{4.39}$$

where ε is a small positive number to ensure stability. Notice that the initial values of $t_n^{(0)}$'s are chosen as

$$t_n^{(0)} = \sqrt{\sum_{i=1}^{K} \|\tilde{\mathbf{w}}_{i,n}^{\text{DL}}\|^2 + \|\tilde{\mathbf{w}}_{i,n}^{\text{VDL}}\|^2}, n = 1, \cdots, N \tag{4.40}$$

where $\tilde{\mathbf{w}}_{i,n}^{\text{DL}}$ and $\tilde{\mathbf{w}}_{i,n}^{\text{VDL}}$ are the beamforming vector solution of Problem (P4) with $\beta_n = 0, \forall n \in \mathcal{N}$. The above update is repeated until $|\beta_n^{(l)} - \beta_n^{(l-1)}| < \eta, \forall n \in \mathcal{N}$, where η is a small positive constant that controls the algorithm accuracy.

Let $\mathbf{t}^\star = [t_1^\star, \cdots, t_N^\star]$ denote the sparse solution after the convergence of the above iterative algorithm. Then the nonzero entries in $\mathbf{t}^\star$ correspond to the APs that need to be active, i.e., $\mathcal{N}_{\text{on}} = \{n | t_n^\star > 0, n \in \mathcal{N}\}$.

2. *Obtain the optimal transmit/receive beamforming vectors* $(\mathbf{w}_i^{DL})^\star$ *and* $(\mathbf{w}_i^{VDL})^\star$, $i = 1, \cdots, K$, *given the selected active APs*. This can be done by solving (P4) with $\beta_n = 0, \forall n \in \mathcal{N}_{\text{on}}$ and $\mathbf{w}_{i,n}^{\text{DL}} = \mathbf{w}_{i,n}^{\text{VDL}} = \mathbf{0}, i = 1, ..., K, \forall n \notin \mathcal{N}_{\text{on}}$.

3. *Obtain the optimal transmit power values of MTs* $(p_i^{UL})^\star$, $i = 1, \cdots, K$. This can be done by solving problem (4.29) with $\mathbf{v}_i^{\text{UL}} = \left(\mathbf{w}_i^{\text{VDL}}\right)^\star, \forall i \in \mathcal{K}$, which is a simple linear programming (LP) problem.

The iterative update given in (4.39) is designed to make small entries in $\{t_n\}_{n=1}^N$ converge to zero. Furthermore, as the updating evolves, the penalty associated with AP n in the objective function, i.e., $\beta_n t_n$, will converge to two possible values:

$$\beta_n t_n \to \begin{cases} P_{c,n} & \text{if } t_n^\star > 0, \text{i.e., AP } n \text{ is active} \\ 0 & \text{otherwise.} \end{cases} \tag{4.41}$$

In other words, only the active APs will incur penalties being the exact same values as their static power consumption, which has the same effect as the indicator function in problem (P1-1) or (P1). Convergence of this algorithm can be shown by identifying the iterative update as a Majorization-Minimization (MM) algorithm [80] for a concave minimization problem, i.e., using $\log(\cdot)$ function, which is concave, to approximate the indicator function given in (4.13). The details are thus omitted due to space limitations.

4.5.1.2 Per-AP and Per-MT Power Constraints

It is first observed that the per-AP power constraints in (4.18), i.e.,

$$\sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\text{DL}}\|^2 \leq P_{n,\text{max}}^{\text{DL}}, \;\; n = 1, \cdots, N \tag{4.42}$$

are convex. Therefore, adding per-AP power constraints to problem (P1-1) does not need to alter the above algorithm. Thus, we focus on the per-MT power constraints in the UL transmission in this subsection.

Again, we consider the following transmit sum-power minimization problem in the UL with per-MT power constraints:

$$\begin{aligned} \underset{\{\mathbf{v}_i^{\text{UL}}\},\{p_i^{\text{UL}}\}}{\texttt{Min.}} \;\; & \sum_{i=1}^{K} p_i^{\text{UL}} \\ \texttt{s.t.} \;\; & \text{SINR}_i^{\text{UL}} \geq \gamma_i^{\text{UL}}, \forall i \in \mathcal{K} \\ & 0 \leq p_i^{\text{UL}} \leq P_{i,\text{max}}^{\text{UL}}, \forall i \in \mathcal{K}. \end{aligned} \tag{4.43}$$

Although it has been shown in [81] that the sum-power minimization problem in the DL with per-AP power constraints can be transformed into an equivalent min-max optimization problem in the UL, we are not able to find an equivalent DL problem for problem (4.43) as in Section 4.5.1.1 which is able to handle the per-MT power constraints. The fundamental reason is that the power allocation obtained by solving problem (4.29) is already component-wise minimum, which can be shown by the uniqueness of the fixed-point solution for a set of minimum SINR requirements in the UL given randomly generated channels [82]. The component-wise minimum power allocation indicates that it is not possible to further reduce one particular MT's power

consumption by increasing others', i.e., there is no tradeoff among different MTs in terms of power minimization. Consequently, solving problem (4.43) requires only one additional step compared with solving problem (4.29), i.e., checking whether the optimal power solution to problem (4.29) satisfies the per-MT power constraints. If this is the case, the solution is also optimal for problem (4.43); otherwise, problem (4.43) is infeasible.

Next, we present our complete algorithm for problem (P1) with the per-AP and per-MT power constraints. Compared to the algorithm proposed for problem (P1-1) in Section 4.5.1.1 without the per-AP and per-MT power constraints, the new algorithm differs in the first step, i.e., to identify the subset of active APs. The main idea is that a set of candidate active APs is first obtained by ignoring the per-MT power constraints but with a new sum-power constraint in the UL (or equivalently its virtual DL), i.e., we iteratively solve the following problem similarly as in the first step of solving problem (P1-1) in Section 4.5.1.1.

$$(\text{P5}): \underset{\{\mathbf{w}_i^{\text{DL}}\},\{\mathbf{w}_i^{\text{VDL}}\},\{t_n\}}{\texttt{Min.}} \quad \sum_{n=1}^{N} \beta_n t_n + \sum_{i=1}^{K} \|\mathbf{w}_i^{\text{DL}}\|^2 + \lambda \sum_{i=1}^{K} \|\mathbf{w}_i^{\text{VDL}}\|^2 \tag{4.44}$$

$$\texttt{s.t.} \ \ (4.36), (4.37) \text{ and } (4.38) \tag{4.45}$$

$$\sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\text{DL}}\|^2 \leq P_{n,\max}^{\text{DL}}, \forall n \in \mathcal{N} \tag{4.46}$$

$$\sum_{i=1}^{K} \|\mathbf{w}_i^{\text{VDL}}\|^2 \leq \sum_{i=1}^{K} P_{i,\max}^{\text{UL}}. \tag{4.47}$$

The sum-power constraint in (4.47) is added to impose a mild control on the transmit powers of all MTs in the UL. After obtaining the candidate set, the feasibility of the UL transmission is then verified. If the candidate set can support the UL transmission with the given per-MT power constraints, then the optimal solution of (P1) is obtained; otherwise, one or more APs need to be active for the UL transmission.

To be more specific, denote the set of candidate active APs obtained by iteratively solving problem (P5) as $\tilde{\mathcal{N}}_{\text{on}}$. Problem (4.29) is then solved with $\tilde{\mathcal{N}}_{\text{on}}$, for which the feasibility is guaranteed due to the virtual DL SINR constraints in (4.37). We denote the obtained power allocation as $\tilde{p}_i^{\text{UL}}$, $i = 1, \cdots, K$.

- If $\tilde{\mathcal{N}}_{\text{on}}$ can support the UL transmission without violating any MT's power constraint, i.e.,

$$\tilde{p}_i^{\text{UL}} \leq P_{i,\max}^{\text{UL}}, \forall i \in \mathcal{K} \tag{4.48}$$

the candidate set can be finalized as the set of active APs and the algorithm proceeds to find the optimal transmit/receive beamforming vectors similarly as that in Section 4.5.1.1.

- If $\tilde{\mathcal{N}}_{\text{on}}$ cannot support the UL transmission with the given MT's power constraints, we propose the following price based iterative method to determine the additional active APs. Specifically, in each iteration, for those APs that are not in the candidate set each will be assigned a price $\theta_m, m \notin \tilde{\mathcal{N}}_{\text{on}}$, which is defined as

$$\theta_m = \frac{1}{P_{c,m}} \sum_{i \in \mathcal{B}} \frac{\tilde{p}_i^{\text{UL}} - P_{i,\max}^{\text{UL}}}{P_{i,\max}^{\text{UL}}} \|\mathbf{g}_{i,m}\|^2, \ \forall m \notin \tilde{\mathcal{N}}_{\text{on}} \tag{4.49}$$

where $\mathcal{B} \triangleq \left\{ i | \tilde{p}_i^{\text{UL}} > P_{i,\max}^{\text{UL}}, i \in \mathcal{K} \right\}$. The price θ_m is set to be the normalized (by its corresponding static power consumption) weighted-sum power gains of the channels from AP m to all the MTs that have their power constraints being violated. The weights are chosen as the ratios of MTs' required additional powers to their individual power limits. According to the definition of θ_m in (4.49), the AP having smaller static power consumption and better channels to MTs whose power constraints are more severely violated will be associated with a larger price. The candidate set is then updated by including the AP that corresponds the largest θ_m as

$$\tilde{\mathcal{N}}_{\text{on}} \leftarrow \tilde{\mathcal{N}}_{\text{on}} \cup \left(\arg \max_{m \notin \tilde{\mathcal{N}}_{\text{on}}} \theta_m \right). \tag{4.50}$$

With updated $\tilde{\mathcal{N}}_{\text{on}}$, the feasibility of the UL transmission needs to be re-checked by obtaining a new set of power allocation, which will be used to compute the new θ_m's in next iteration if further updating is required. The above process is repeated until all the MTs' power constraints are satisfied. Its convergence is guaranteed since problem (P1) has been assumed to be feasible if all APs are active.

Combining with the algorithm in Section 4.5.1.1, our complete algorithm for problem (P1) based on GSO is summarized in Table 4.1. For the algorithm given in Table 4.1, there are two problems that need to be iteratively solved, i.e., problems (4.29) and (P5). Since problem (4.29) can be efficiently solved by the fixed-point algorithm [82], the computation time is dominated by solving the SOCP problem (P5). If the primal-dual interior point algorithm [32] is used by the numerical solver for solving (P5), the computational complexity is of order $M^{3.5}K^{3.5}$. Furthermore, since the convergence of the iterative update in steps 4)-5), governed by the MM algorithm, is very fast (approximately 10-15 iterations) as observed in the simulations, the overall complexity of the algorithm in Table 4.1 is approximately $\mathcal{O}(M^{3.5}K^{3.5})$.

Table 4.1: Algorithm for Solving Problem (P1) via GSO

1. Set $l = 0$, initialize the set of candidate active APs as $\tilde{\mathcal{N}}_{\text{on}} = \mathcal{N}$.
2. Obtain $\tilde{\mathbf{w}}_{i,n}^{\text{DL}}$'s and $\tilde{\mathbf{w}}_{i,n}^{\text{VDL}}$'s by solving problem (P5) with $\beta_n = 0, \forall n \in \mathcal{N}$.
3. Set $t_n^{(0)} = \sqrt{\sum_{i=1}^{K} \|\tilde{\mathbf{w}}_{i,n}^{\text{DL}}\|^2 + \|\tilde{\mathbf{w}}_{i,n}^{\text{VDL}}\|^2}, n = 1, \cdots, N$.
4. **Repeat:**
   a) $l \leftarrow l + 1$.
   b) Set $\beta_n^{(l)} = \frac{P_{c,n}}{t_n^{(l-1)} + \varepsilon}, \forall n \in \mathcal{N}$.
   c) Obtain $\mathbf{t}^{(l)} = [t_1^{(l)}, \cdots, t_N^{(l)}]$ by solving problem (P5) with $\beta_n = \beta_n^{(l)}, \forall n \in \mathcal{N}$.
5. **Until** $|\beta_n^{(l)} - \beta_n^{(l-1)}| \leq \eta, \forall n \in \mathcal{N}$ or $l = l_{\max}$.
6. Set $\tilde{\mathcal{N}}_{\text{on}}$ as $\tilde{\mathcal{N}}_{\text{on}} = \{n | t_n^\star > 0, n \in \mathcal{N}\}$.
7. **Repeat:**
   a) Obtain $\tilde{p}_i^{\text{UL}}, i = 1, \cdots, K$, by solving problem (4.29) with $\tilde{\mathcal{N}}_{\text{on}}$.
   b) Set $\mathcal{B} = \left\{i | \tilde{p}_i^{\text{UL}} > P_{i,\max}^{\text{UL}}, i \in \mathcal{K}\right\}$.
   c) Set $\theta_m = \frac{1}{P_{c,m}} \sum_{i \in \mathcal{B}} \frac{\tilde{p}_i^{\text{UL}} - P_{i,\max}^{\text{UL}}}{P_{i,\max}^{\text{UL}}} \|\mathbf{g}_{i,m}\|^2, \forall m \notin \tilde{\mathcal{N}}_{\text{on}}$.
   d) Set $\tilde{\mathcal{N}}_{\text{on}} \leftarrow \tilde{\mathcal{N}}_{\text{on}} \cup \left(\arg \max_{m \notin \tilde{\mathcal{N}}_{\text{on}}} \theta_m\right)$.
8. **Until** $\mathcal{B} = \emptyset$.
9. Obtain $\left(\mathbf{w}_i^{\text{DL}}\right)^\star$ and $\left(\mathbf{w}_i^{\text{V-DL}}\right)^\star$, $i = 1, \cdots, K$, by solving (P5) with $\beta_n = 0$, $\forall n \in \mathcal{N}$ and $\mathbf{w}_{i,n}^{\text{DL}} = \mathbf{w}_{i,n}^{\text{VDL}} = \mathbf{0}, i = 1, ..., K, \forall n \notin \tilde{\mathcal{N}}_{\text{on}}$.
10. Set $\left(\mathbf{v}_i^{\text{UL}}\right)^\star = \left(\mathbf{w}_i^{\text{VDL}}\right)^\star, \forall i \in \mathcal{K}$, and compute the $\left(p_i^{\text{UL}}\right)^\star$, $i = 1, \cdots, K$, by solving problem (4.29).

4.5.2 Proposed Algorithm for Problem (P1) via RIP

In this subsection, an alternative algorithm for problem (P1) is developed based on RIP by applying the same idea of establishing a virtual DL transmission for the original UL. Similar to the case with GSO, the per-MT power constraints are first replaced with a sum-power constraint in the UL. The resulting problem is further reformulated as a convex SOCP by relaxing the binary variables $\{\rho_n\}$, which is given as follows.

$$
\begin{aligned}
(\text{P6}): \underset{\{\mathbf{w}_i^{\text{DL}}\},\{\mathbf{w}_i^{\text{VDL}}\},\{\rho_n\}}{\texttt{Min.}} \quad & \sum_{n=1}^{N} \rho_n P_{c,n} + \sum_{i=1}^{K} \|\mathbf{w}_i^{\text{DL}}\|^2 + \lambda \sum_{i=1}^{K} \|\mathbf{w}_i^{\text{VDL}}\|^2 \\
\texttt{s.t.} \quad & (4.36), (4.37) \text{ and } (4.47) \\
& \sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\text{DL}}\|^2 \le \rho_n P_{n,\max}^{\text{DL}}, \forall n \in \mathcal{N} && (4.51) \\
& \sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\text{VDL}}\|^2 \le \rho_n \sum_{i=1}^{K} P_{i,\max}^{\text{UL}}, \forall n \in \mathcal{N} && (4.52) \\
& \sum_{n=1}^{N} \rho_n \ge 1 && (4.53) \\
& 0 \le \rho_n \le 1, \forall n \in \mathcal{N}. && (4.54)
\end{aligned}
$$

Note that instead of implementing the active-sleep constraints jointly for the actual DL and virtual DL as (4.27) in problem (P2), i.e., $\sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\text{DL}}\|^2 + \|\mathbf{w}_{i,n}^{\text{VDL}}\|^2 \le \rho_n (P_{n,\max}^{\text{DL}} + \sum_{i=1}^{K} P_{i,\max}^{\text{UL}})$, we divide them into two sets of coupled active-sleep constraints as in (4.51) and (4.52) via ρ_n's. For the non-relaxed problem of (P6) with binary ρ_n's, i.e., $\rho_n \in \{0,1\}, \forall n \in \mathcal{N}$, it can be shown that these two formulations are equivalent. However, for the case of the relaxed problem (P6) with continuous valued ρ_n's, the separated active-sleep constraints are designed to avoid the situation that the difference between $P_{n,\max}^{\text{DL}}$ and $\sum_{i=1}^{K} P_{i,\max}^{\text{UL}}$ is too large such that the optimal value of ρ_n is dominated by either DL or UL transmission. To implement the big-M method with active-sleep constraints [76], an appropriate upper bound for the term $\sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\text{VDL}}\|^2$ needs to be found. According to the UL-DL duality, the minimum sum-power achieved is the same for the UL and its virtual DL transmissions. Therefore, $\sum_{i=1}^{K} P_{i,\max}^{\text{UL}}$ can be chosen as the upper bound of $\sum_{i=1}^{K} \|\mathbf{w}_{i,n}^{\text{VDL}}\|^2, \forall n \in \mathcal{N}$. Finally, it is evident that the optimal value of problem (P6) serves as a lower bound of its non-relaxed problem

with binary ρ_n's. In order to further tighten this lower bound, one way is to reduce the feasible set of design variables. Constraint in (4.53) is introduced specifically to achieve this end, which can be shown to be redundant for the non-relaxed problem of (P6).

We adopt the same idea of incentive measure based AP selection as in [58] to design a polynomial-time algorithm for problem (P2). However, it remains to find an incentive measure that reflects the importance of each AP to both DL and UL transmissions based on problem (P6). It is interesting to observe that after transforming the UL related terms to their virtual DL counterparts, the optimal relaxed binary variable solution of problem (P6) becomes a good choice to serve this purpose. Let $\breve{\rho}_n$, $\breve{\mathbf{w}}_i^{\mathrm{DL}}$ and $\breve{\mathbf{w}}_i^{\mathrm{VDL}}$ denote the optimal solution to problem (P6). Intuitively, the AP that has larger static power consumption and worse channels to all MTs is more desired to be switched into sleep mode from the perspective of energy saving. In (P6), for AP n having larger $P_{c,n}$, $\breve{\rho}_n$ is desired to be smaller in order to achieve the minimum value of the objective function. Furthermore, it is practically valid that for DL power minimization problem, the optimal transmit power of APs that have worse channels to MTs is in general smaller. As a result, solving problem (P6) yields smaller $\sum_{i=1}^{K} \|\breve{\mathbf{w}}_{i,n}^{\mathrm{DL}}\|^2$ and $\sum_{i=1}^{K} \|\breve{\mathbf{w}}_{i,n}^{\mathrm{VDL}}\|^2$ and thus smaller $\breve{\rho}_n$ for AP n that has worse channels to MTs for both the DL and UL transmissions. To summarize, the AP that corresponds to smaller $\breve{\rho}_n$ is more desired to be switched into sleep mode.

An iterative process is then designed to determine the set of active APs based on $\breve{\rho}_n$'s that are taken as incentive measures. The process starts with assuming all APs are active. In each iteration, problem (P6) is solved with a candidate set of active APs, and the AP corresponding to the smallest $\breve{\rho}_n$ will be removed from the candidate set. This process is repeated until one of the following conditions occurs:

- The weighted sum-power cannot be further reduced;
- Problem (P6) becomes infeasible;
- Problem (4.43) becomes infeasible.

Note that the feasibility checking for problem (4.43) is the same as that in Algorithm

I, which ensures the per-MT power constraints. An overall algorithm for problem (P1) based on RIP is summarized in Table 4.2. For the algorithm given in Table 4.2, the computation time is dominated by solving the SOCP problem (P6). If the primal-dual interior point algorithm [32] is used by the numerical solver for solving (P6), the computational complexity is of order $M^{3.5}K^{3.5}$. Furthermore, since the worst case complexity for the iteration in steps 2)-3) is $\mathcal{O}(N)$, the overall complexity of the algorithm in Table 4.2 is $\mathcal{O}(NM^{3.5}K^{3.5})$.

Table 4.2: Algorithm for Solving Problem (P1) via RIP

1. Set $l = 0$, $\Phi^{(0)}$ a sufficiently large value, and initialize the set of candidate active APs as $\tilde{\mathcal{N}}_{\text{on}} = \mathcal{N}$.
2. **Repeat:**
   a) $l \leftarrow l + 1$.
   b) Solve problem (P6) with $\rho_n = 0, \forall n \notin \tilde{\mathcal{N}}_{\text{on}}$.
   c) Set $\tilde{\mathcal{N}}_{\text{on}} \leftarrow \tilde{\mathcal{N}}_{\text{on}} \setminus \left(\arg \min_{n \in \tilde{\mathcal{N}}_{\text{on}}} \breve{\rho}_n^{(l)} \right)$.
   d) Set $\Phi^{(l)}$ as the optimal value of problem (P6) with $\rho_n = 1, \forall n \in \tilde{\mathcal{N}}_{\text{on}}$ and $\rho_n = 0, \forall n \notin \tilde{\mathcal{N}}_{\text{on}}$.
3. **Until** $\Phi^{(l)} > \Phi^{(l-1)}$ or problem (P6) is infeasible or problem (4.43) is infeasible.
4. Obtain $\left(\mathbf{w}_i^{\text{DL}}\right)^\star$ and $\left(\mathbf{w}_i^{\text{V-DL}}\right)^\star$, $i = 1, \cdots, K$, by solving problem (P6) with $\rho_n = 1, \forall n \in \tilde{\mathcal{N}}_{\text{on}}$ and $\rho_n = 0, \forall n \notin \tilde{\mathcal{N}}_{\text{on}}$.
5. Set $\left(\mathbf{v}_i^{\text{UL}}\right)^\star = \left(\mathbf{w}_i^{\text{VDL}}\right)^\star, \forall i \in \mathcal{K}$, and compute the $\left(p_i^{\text{UL}}\right)^\star$, $i = 1, \cdots, K$, by solving problem (4.29).

4.6 Numerical Results

In this section, we present numerical results to verify our proposed algorithms from three perspectives: ensuring feasibility for both DL and UL transmissions; achieving network power saving with optimal MT-AP association; and adjusting minimum power consumption tradeoffs between active APs and MTs. We consider two possible C-RAN configurations [55]:

1. Homogeneous setup: all APs are assumed to have the same power consumption model with $P_{c,n} = 2$W and $P_{n,\max}^{\mathrm{DL}} = 1$W, $\forall n \in \mathcal{K}$, if not specified otherwise.

2. Heterogeneous setup: two types of APs are assumed, namely, high-power AP (HAP) and low-power AP (LAP), where the static power consumption for HAP and LAP are set as 50W and 2W, respectively, and the transmit power budgets for HAP and LAP are set as 20W and 1W, respectively.

We assume that each AP n, $n \in \mathcal{N}$, is equipped with $M_n = 2$ antennas. For the single-antenna MT, we set the transmit power limit as $P_{i,\max}^{\mathrm{UL}} = 0.5$W, $\forall i \in \mathcal{K}$. For simplicity, we assume that the SINR requirements of all MTs are the same in the UL or DL. All the APs (except HAPs under the heterogeneous setup) and MTs are assumed to be uniformly and independently distributed in a square area with the size of 3Km$\times 3$Km. For all the simulations under heterogeneous setup, it is assumed that there are 2 HAPs with fixed location at $[-750\mathrm{m}, 0\mathrm{m}]$ and $[750\mathrm{m}, 0\mathrm{m}]$, respectively. We assume a simplified channel model consisting of the distance-dependent attenuation with pathloss exponent $\alpha = 3$ and a multiplicative random factor (exponentially distributed with unit mean) accounting for short-term Rayleigh fading. We also set $\lambda = 1$ if not specified otherwise, i.e., we consider the sum-power consumption of all active APs and MTs. Finally, we set the receiver noise power for all the APs and MTs as $\sigma^2 = -50$dBm.

4.6.1 Feasibility Performance

First, we demonstrate the importance of active AP selection by jointly considering both DL and UL transmission in terms of the SINR feasibility in C-RAN. Since feasibility is our focus here instead of power consumption, it is assumed that the selected active APs will support all MTs for both the DL and UL transmissions. The simulation results compare our proposed algorithms (i.e., Algorithms I and II) with the following three AP selection schemes:

- **AP initiated reference signal strength (APIRSS) based selection**: In this scheme, APs first broadcast orthogonal reference signals. Then, for each MT, the AP corresponding to the largest received reference signal strength will be included in the set of active APs. Note that this scheme has been implemented in practical cellular systems [83].

- **MT initiated reference signal strength (MUIRSS) based selection**: In this scheme, MTs first broadcast orthogonal reference signals. Then, for each MT, the AP corresponding to the largest received reference signal strength will be included in the set of active APs. Note that since all MTs are assumed to transmit reference signals with equal power and pathloss in general dominates short-term fading, the AP that is closest to each MT will receive strongest reference signal in general. Also note that in the previous APIRSS based scheme, if all APs are assumed to transmit with equal reference signal power (e.g., for the homogenous setup), the selected active APs will be very likely to be the same as those by the MUIRSS based scheme.

- **Proposed algorithm without considering UL (PAw/oUL)**: In this algorithm, the set of active APs are chosen from the conventional DL perspective by modifying our proposed algorithms. Specifically, Algorithm I is used here and similar results can be obtained with Algorithm II. Note that Algorithm I without considering UL transmission is similar to that proposed in [55].

With the obtained set of active APs, the feasibility check of problem (P1) can be decoupled into two independent feasibility problems: one for the DL and the other

for the UL, while the network feasibility is achieved only when both the UL and DL SINR feasibility of all MTs are guaranteed.

In Fig. 4.1, we illustrate the set of active APs generated by different schemes under the heterogeneous setup, and also compare them with that by the optimal exhaustive search. It is assumed that there are 2 HAPs and 8 LAPs jointly supporting 8 MTs. The SINR targets for both DL and UL transmissions of all MTs are set as 8dB. First, it is observed that Algorithm I and Algorithm II obtain the same set of active APs as shown in Fig. 4.1(a), which is also identical to that found by exhaustive search. Second, it is observed that the 2 HAPs are both chosen to be active in Fig. 4.1(b) for the APIRSS based scheme. This is due to the significant difference between HAP and LAP in terms of transmit power, which makes most MTs receive the strongest DL reference signal from the HAP. The above phenomenon is commonly found in heterogenous network (HetNet) [67] with different types of BSs (e.g. macro/micro/pico BSs). Third, from Fig. 4.1(c), the active APs by the MUIRSS based scheme are simply those closer to the MTs, which is as expected. Finally, in Fig. 4.1(d), only two LAPs are chosen to support all MTs with the PAw/oUL algorithm. This is because the algorithm does not consider UL transmission, and as a result Fig. 4.1(d) only shows the most energy-efficient AP selection for DL transmission.

To compare the feasibility performance, we run the above algorithms with different DL and UL SINR targets. It is assumed that $N = 6$ and $K = 4$. The results are summarized in Table 4.3 and Table 4.4, where the number of infeasible cases for each scheme is shown over 200 randomly generated network and channel realizations, for homogeneous setup and heterogeneous setup, respectively. Note that in these examples, Algorithm I and Algorithm II have identical feasibility performance, since the system is infeasible only when the DL/UL SINR requirements cannot be supported for given channels and power budgets even with all APs being active.

From both Table 4.3 and Table 4.4, it is first observed that the three comparison schemes, i.e., APIRSS based scheme, MUIRSS based scheme and PAw/oUL, all incur much larger number of infeasible cases as compared to our proposed algorithms. It is also observed that among the three comparison schemes, PAw/oUL has the best

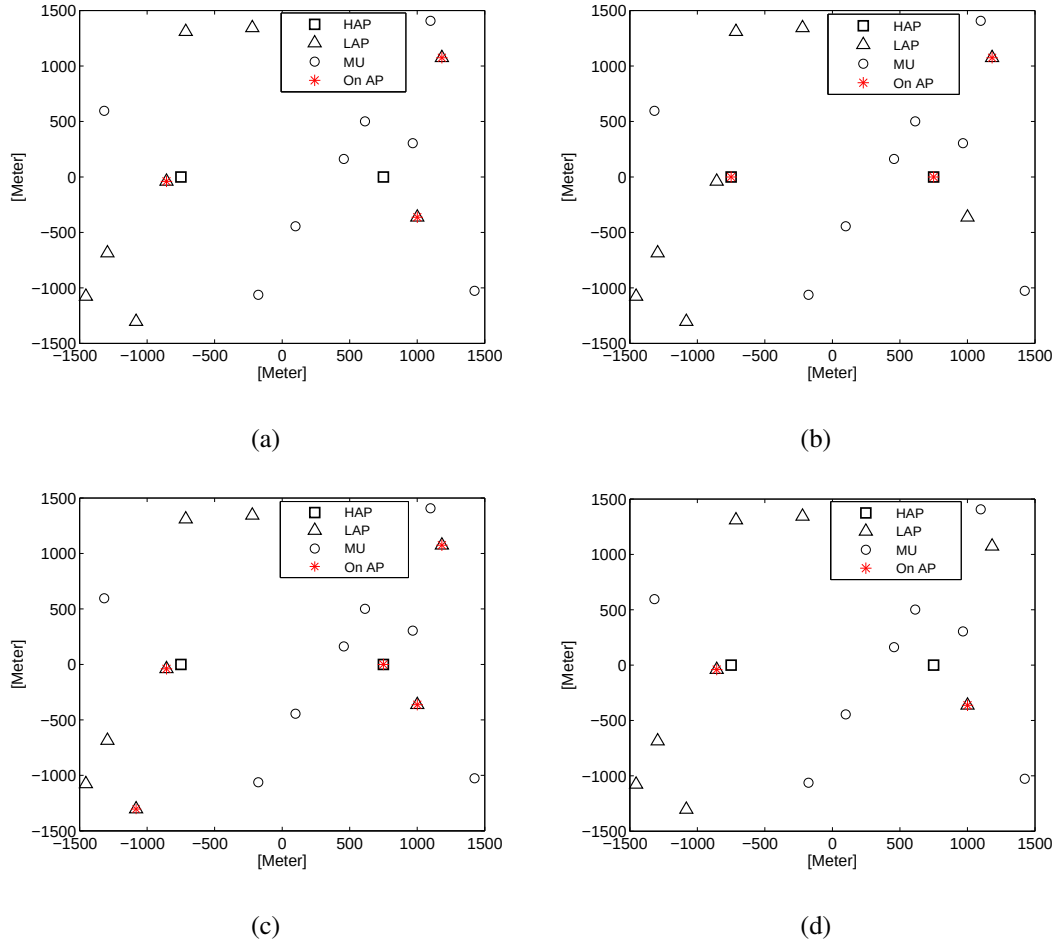


Figure 4.1: The set of active APs generated by: (a) proposed algorithms; (b) APIRSS; (c) MUIRSS; and (d) PAw/oUL

performance (or the minimum number of infeasible cases) when the DL transmission is dominant (i.e., $\gamma_i^{\mathrm{DL}} > \gamma_i^{\mathrm{UL}}$); however, it performs the worst in the opposite situation (i.e., $\gamma_i^{\mathrm{DL}} < \gamma_i^{\mathrm{UL}}$). This observation indicates that DL oriented scheme could result in infeasible transmit power of MTs in the UL for the cases with stringent UL requirements. From the last two rows of Table 4.4, it is observed that the APIRSS based scheme performs worse than the MUIRSS based scheme when the UL SINR target is high. This is because that under heterogeneous setup, as shown in Fig. 4.1(b), MTs are attached to the HAPs under APIRSS based scheme although the HAPs may be actually more distant away from MTs compared with the distributed LAPs. This imbalanced association causes much higher transmit powers of MTs or even their

infeasible transmit power in the UL. There has been effort in the literature to address this traffic imbalance problem in HetNet. For example in [84], the reference signal from picocell BS is multiplied by a factor with magnitude being larger than one, which makes it appear more appealing for MT association than the heavily-loaded macrocell BS.

Table 4.3: Feasibility Performance Comparison under Homogeneous Setup

Parameters			Number of Infeasible Cases			
K	γ_i^{DL}(dB)	γ_i^{UL}(dB)	APIRSS	MUIRSS	PAw/oUL	Proposed Algorithms
2	6	6	0	0	0	0
2	12	6	14	18	2	0
2	6	12	86	90	150	54
2	12	12	88	92	118	50
4	6	6	0	0	4	0
4	12	6	52	56	2	0
4	6	12	140	144	194	80
4	12	12	152	150	164	86

4.6.2 Sum-Power Minimization

Next, we compare the performance of the proposed algorithms in terms of sum-power minimization in C-RAN with the following benchmark schemes:

- **Exhaustive search**: In this scheme, the optimal set of active APs are found by exhaustive search, which serves as the performance upper bound (or lower bound on the sum-power consumption) for other considered schemes. With any set of active APs, the minimum-power DL and UL beamforming problems can be separately solved. Since the complexity of exhaustive search grows exponentially with N, it can only be implemented for C-RAN with small number

Table 4.4: Feasibility Performance Comparison under Heterogeneous setup

Parameters			Infeasibility			
K	γ_i^{DL}(dB)	γ_i^{UL}(dB)	APIRSS	MUIRSS	PAw/oUL	Proposed Algorithms
2	6	6	0	0	0	0
2	12	6	12	18	2	0
2	6	12	124	82	154	46
2	12	12	124	86	118	52
4	6	6	0	0	2	0
4	12	6	42	36	6	0
4	6	12	178	132	196	84
4	12	12	178	134	170	82

of APs.

- **Joint processing among all APs** [65]: In this scheme, all the APs are assumed to be active and only the total transmit power consumption is minimized by solving two separate (DL and UL) minimum-power beamforming design problems.

- **Algorithm I with $\ell_{1,\infty}$ norm penalty**: This algorithm is the same as that given in Table 4.1 except that the sparsity enforcing penalty is replaced with $\ell_{1,\infty}$ norm as given in (4.22).

In our simulations, we consider the homogeneous C-RAN setup with $N = 6$ and plot the performance by averaging over 500 randomly generated network and channel realizations[2]. The SINR requirements are set as $\gamma_i^{\text{UL}} = \gamma_i^{\text{DL}} = 8$dB for all MTs $i \in \mathcal{K}$. Fig. 4.2 and Fig. 4.3 show the sum-power consumption achieved by different

[2]Note that since the channel attenuation is dominated by pathloss, the user association scheme is in general more sensitive to pathloss than to small-scale fading. Therefore, the rate of adjusting user association is much slower than the time scale of small-scale fading. Consequently, hundreds of random realizations are enough to obtain accurate performance estimation.

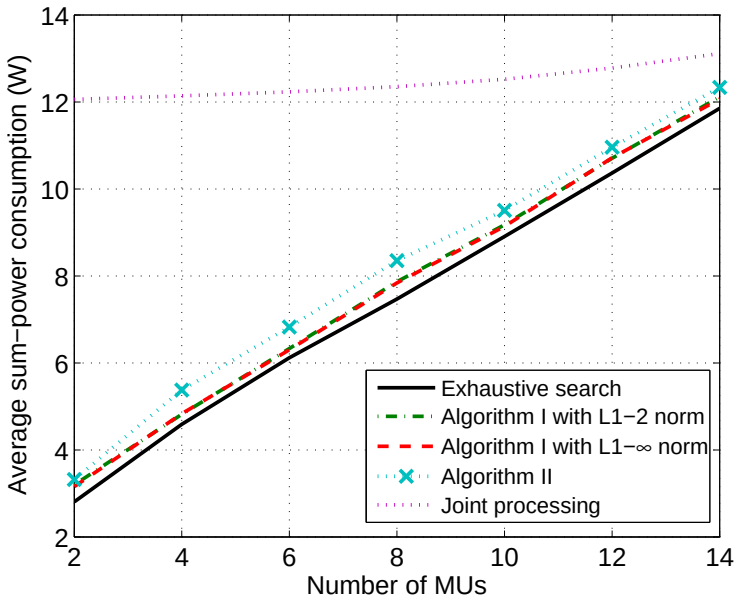


Figure 4.2: Sum-power consumption versus number of MTs under homogeneous setup with $P_{c,n} = 2\text{W}, \forall n \in \mathcal{N}$

algorithms versus the number of MTs K and AP static power consumption $P_{c,n}$ (assumed to be identical for all APs), respectively. From both figures, it is observed that the proposed algorithms have similar performance as the optimal exhaustive search and achieve significant power saving compared with joint processing algorithm. It is also observed that the penalty term based on either $\ell_{1,2}$ or $\ell_{1,\infty}$ norm has small impact on the performance of Algorithm I. Finally, Algorithm I always outperforms Algorithm II although the performance gap is not significant.

4.6.3 Power Consumption Tradeoff

Finally, we compare the sum-power consumption tradeoffs between active APs and all MTs for the proposed algorithms as well as the optimal exhaustive search, by varying the weight parameter λ in our formulated problems. We consider a homogenous C-RAN setup with $N = 6$ and $K = 4$, where $\gamma_i^{\text{UL}} = \gamma_i^{\text{DL}} = 8\text{dB}$ for all MTs $i \in \mathcal{K}$. Since it has been shown in the pervious subsection that Algorithm I with $\ell_{1,2}$ norm and $\ell_{1,\infty}$ norm achieves similar performance, we choose $\ell_{1,2}$ norm in this simulation.

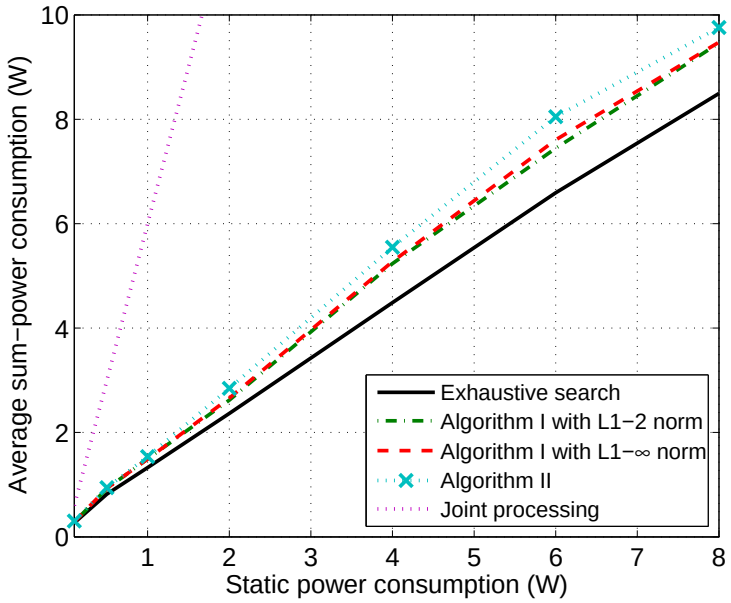


Figure 4.3: Sum-power consumption versus AP static power consumption under homogeneous setup with $K = 4$

Furthermore, since JS assumes that all the APs are active, which decouples DL and UL transmissions and thus has no sum-power consumption tradeoffs between APs and MTs, it is also not included. From Fig. 4.4, it is first observed that for all considered algorithms, as λ increases, the sum-power consumption of active APs increases and that of all MTs decreases, which is as expected. It is also observed that Algorithm I achieves trade-off performance closer to exhaustive search and outperforms Algorithm II, which is in accordance with the results in Figs. 4.2 and 4.3.

4.7 Chapter Summary

In this chapter, we considered C-RAN with densely deployed APs cooperatively serving distributed MTs for both the DL and UL transmissions. We studied the problem of joint DL and UL MT-AP association and beamforming design to optimize the energy consumption tradeoffs between the active APs and MTs. Leveraging the celebrated UL-DL duality result, we showed that by establishing a virtual

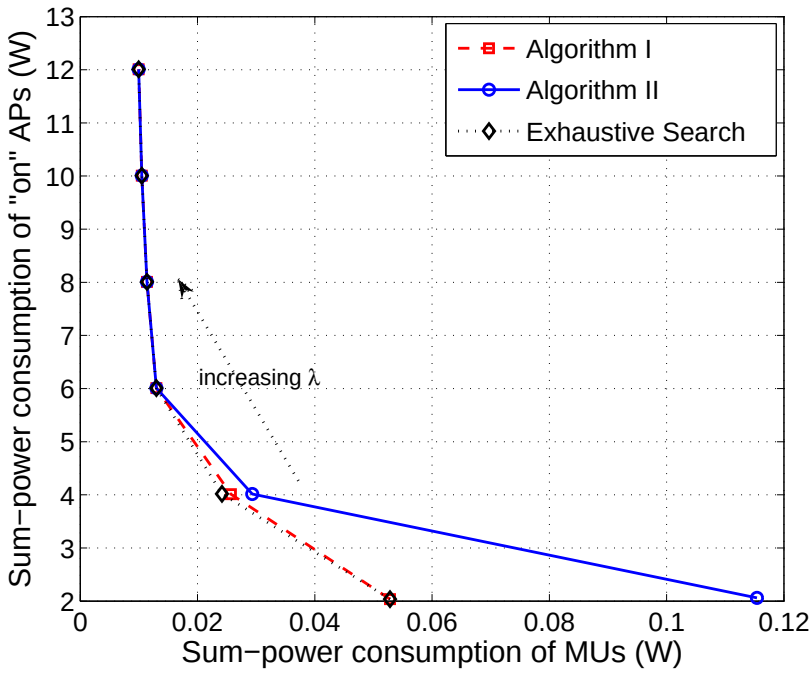


Figure 4.4: Sum-power consumption tradeoffs between active APs and MTs under homogeneous setup

DL transmission for the original UL transmission, the joint DL and UL problem can be converted to an equivalent DL problem in C-RAN with two inter-related subproblems for the original and virtual DL transmissions, respectively. Based on this transformation, two efficient algorithms for joint DL and UL MT-AP association and beamforming design were proposed based on GSO and RIP techniques, respectively. By extensive simulations, it was shown that our proposed algorithms improve the network reliability/feasibility, energy efficiency, as well as power consumption tradeoffs between APs and MTs, as compared to other existing methods in the literature.

Chapter 5

Optimization with Save-Then-Transmit Protocol for Energy Harvesting Powered Wireless Transmission

5.1 Introduction

In this chapter, different from the previous three chapters, we turn to address the energy saving issue from the energy harvesting (EH) perspective. In particular, the design of a wireless communication transmitter relying exclusively on EH is considered in a point-to-point communication link. Two rechargeable energy storage devices (ESDs) are assumed so that at any given time, there is always one ESD being recharged. The EH rate is assumed to be a random variable that is constant over each time interval of interest, but can vary over different intervals. A *save-then-transmit* (ST) protocol is introduced, in which a fraction of time ρ (dubbed the save-ratio) is devoted exclusively to EH, with the remaining fraction $1 - \rho$ used for data transmission. Important properties of the optimal save-ratio that minimizes transmission outage probability are derived, from which useful design guidelines are drawn. In addition, we compare the outage performance of random power supply to that of constant power supply in the Rayleigh fading channel. Finally, we extend the proposed ST protocol to wireless networks with multiple EH transmitters.

5.2 Literature Review

The availability of an inexhaustible but unreliable energy source changes a communication system designer's options considerably, compared to the conventional cases of an inexhaustible reliable energy source (powered by the grid), and an exhaustible reliable energy source (powered by batteries).

There has been recent research on understanding data packet scheduling with an EH transmitter, most of which employed a deterministic energy harvesting (EH) model. In [85], the transmission time for a given amount of data was minimized through power control based on known energy arrivals over all time. Structural properties of the optimum solution were then used to establish a fast search algorithm. This work has been extended to battery limited cases in [86], battery imperfections in [87, 88]. EH with channel fading has been investigated in [89] and [90], wherein a water-filling energy allocation solution where the so-called water levels follow a staircase function was proved to be optimal. An information theoretic analysis of EH communication systems has been provided in [91, 92]. In [91], the authors proved that the capacity of the AWGN channel with stochastic energy arrivals is equal to the capacity with an average power constraint equal to the average recharge rate. This work has been extended in [92] to the fading Gaussian channels with perfect/no channel state information at the transmitter.

In scenarios where multiple EH wireless devices interact with each other, the design needs to adopt a system-level approach [93–95]. In [93], the medium access control (MAC) protocols for single-hop wireless sensor networks, operated by EH capable devices, were designed and analyzed. In [94], N EH nodes with independent data and energy queues were considered, and the queue stability was analyzed under different MAC protocols. In [95], the authors proposed efficient algorithm for simultaneously achieving proportional fairness and perpetual operation in wireless sensor networks. For the cases that the channel condition is such that the source node cannot transmit data directly to the destination node, two-hop or multi-hop transmission with intermediate relay is necessary [96, 97]. The optimal transmission

policy with a non-EH source and an EH relay was developed in [96]. This work has been extended to the case where both the source and relay are EH nodes with different delay constraints in [97].

Due to the theoretical intractability of online power scheduling under the energy causality constraint (the cumulative energy consumed is not allowed to exceed the cumulative energy harvested at every point in time), most current research is focused on an offline strategy with deterministic channel and energy state information, which is not practical and can only provide an upper bound on system performance. An earlier line of research considers the problem of energy management, with only causal energy state information, in communications satellites [98], which formulated the problem of maximizing a reward that is linear in the energy as a dynamic programming problem. In [99], energy management policies which stabilize the data queue have been proposed for single-user communication under linear energy-rate approximations.

5.3 System Model

5.3.1 Definitions and Assumptions

The block diagram of the system is given in Fig. 5.1. Because rechargeable energy storage devices (ESDs) cannot both charge and discharge simultaneously (the *energy half-duplex constraint*), an EH transmitter needs two ESDs, which we call the main ESD (MESD) and secondary ESD (SESD). The energy harvested from the environment[1] is first stored in either the MESD or the SESD at any given time, as indicated by switch a, before it is used in data transmission. The MESD powers the transmitter directly and usually has high power density, good recycle ability and high efficiency, e.g., a super-capacitor [100], which is suitable for applications undergoing frequent charge and discharge cycles at high current and short duration. Since the MESD cannot charge and discharge simultaneously, a SESD (e.g. rechargeable battery) stores up harvested energy while the transmitter is on, and transfers all its

[1] Wind, solar, geothermal, etc.

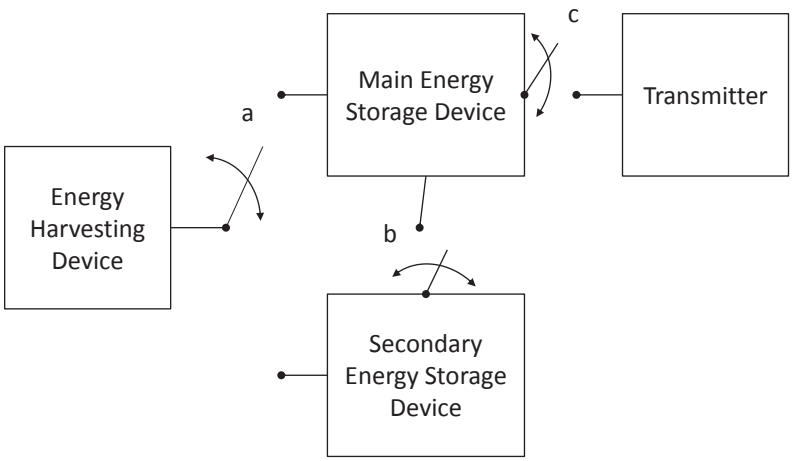


Figure 5.1: Energy harvesting circuit model

stored energy to the MESD once the transmitter is off. We assume in the rest of this chapter that the SESD is a battery with an efficiency η,[2] where $\eta \in [0, 1]$. This means that a fraction η of the energy transferred into the SESD during charging can be subsequently recovered during discharging. The other $1 - \eta$ fraction of the energy is thus lost, due to e.g., battery leakage and/or circuit on/off overhead. The reason of choosing a single-throw switch (switch a in Fig. 5.1) between the EH device (EHD) and ESDs is that splitting the harvested energy with a portion going to the SESD, when the transmitter does not draw energy from the MESD, is not energy efficient due to the SESD's lower efficiency. Note that at the current stage of research, the optimal detailed structure of an EH transmitter is not completely known and there exist various models in the literature (see e.g., [88–90]). The proposed circuit model, given in Fig. 5.1, provides one possible practical design.

We assume that Q bits of data are generated and must be transmitted within a time slot of duration T seconds (i.e., delay constrained) using a *save-then-transmit* (ST) protocol (see Fig. 5.2). In the proposed ST protocol, the save-ratio ρ is the reserved fraction of time for EH by the MESD within one transmission slot. In other

[2]In practice, the battery efficiency can vary from 60% to 99%, depending on different recharging technologies [11].

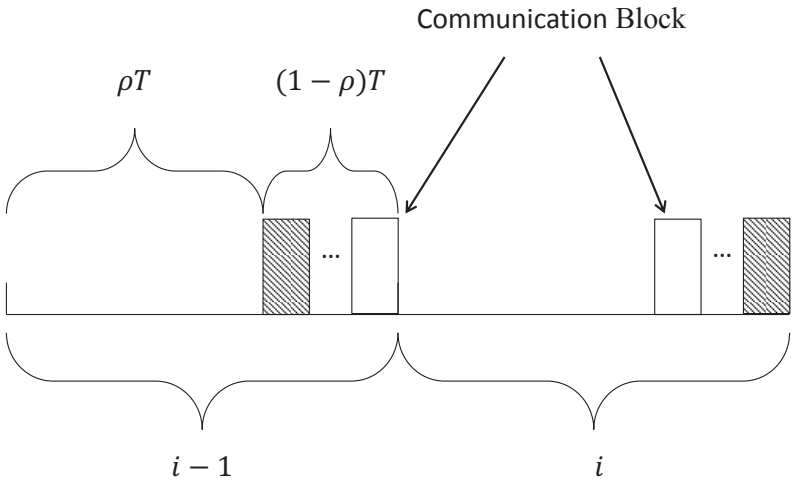


Figure 5.2: Save-Then-Transmit (ST) protocol

words, data delivery only takes place in the last $(1-\rho)T$ seconds of each time slot, which results in an effective rate of $R_{\text{eff}} = \frac{Q}{(1-\rho)T}$ bits/sec. We also allow for a constant power consumption of P_c Watts by the transmitter hardware whenever it is powered on. The combined influence of ρ, η and P_c on outage probability is quantified in this work.

Assume a block-fading frequency-nonselective channel, where the channel is constant over the time slot T. Over any time slot, the baseband-equivalent channel output is given by

$$y = h \cdot x + n \tag{5.1}$$

where x is the transmitted signal, y is the received signal, and n is i.i.d. circularly symmetric complex Gaussian (CSCG) noise with zero mean and variance σ_n^2.

For any frame, the ST protocol (cf. Fig. 5.2) is described as follows:

- During time interval $(0, \rho T]$, harvested energy accumulates in the MESD, which corresponds to the situation that switches b, c are open and a connects to the MESD in Fig. 5.1;

- From time ρT to T, the transmitter is powered on for transmission with energy from the MESD. Since the transmitter has no knowledge of the channel state, we assume that all the buffered energy in the MESD is used up (best-effort

transmission) in each frame. Since the MESD cannot charge and discharge at the same time, the SESD starts to store up harvested energy while the transmitter is on. Referring to Fig. 5.1, c is closed, b is open and a switches to the SESD;

- At time T, the transmitter completes the transmission and powers off. The SESD transfers all its buffered energy to the MESD within a negligible charging time, at efficiency η. In other words, b is closed and switches a and c are open in Fig. 5.1.

5.3.2 Outage Probability

The frame interval T is assumed to be small relative to the time constant of changes in the ESD charging rate (or energy arrival rate). The energy arrival rate is therefore modeled as a random variable X in Joules/second, which is assumed to be constant over a frame. It is clear that X is a non-negative random variable with finite support, i.e., $0 \leq X \leq P_H < \infty$, as the maximum amount of power one can extract from any source is finite. Suppose $f_X(x)$ and $F_X(x)$ represent its probability density function (PDF) and cumulative distribution function (CDF), respectively. According to the proposed ST protocol, the total buffered energy in the MESD at $t = \rho T$ (the start of data transmission within a transmission slot) is given by

$$E_T = X\left[\rho + \eta(1-\rho)\right]T. \tag{5.2}$$

Denote $P = \frac{E_T}{(1-\rho)T} = X\left[\frac{\rho}{1-\rho} + \eta\right]$ as the average total power, which is constant over the entire transmission period, and P_c as the circuit power (i.e., the power consumed by the hardware during data transmission), again assumed constant. The mutual information of the channel (5.1) conditioned on X and the channel gain h is (assuming $P > P_c$)

$$R_T = \log_2\left(1 + \frac{(P-P_c)|h|^2}{\sigma_n^2}\right) = \log_2\left(1 + (P-P_c)\Gamma\right) \tag{5.3}$$

where $\Gamma = \frac{|h|^2}{\sigma_n^2}$ with PDF $f_\Gamma(\cdot)$ and CDF $F_\Gamma(\cdot)$.

For a transmitter with EH capability and working under the ST protocol, the outage event is the union of two mutually exclusive events: Circuit Outage and

Channel Outage. Circuit outage occurs when the MESD has insufficient energy stored up at $t = \rho T$ to even power on the hardware for the duration of transmission i.e., $E_T < P_c(1-\rho)T$ or equivalent $P < P_c$. Channel outage is defined as the MESD having sufficient stored energy but the channel realization does not support the required target rate $R_{\text{eff}} = \frac{Q}{(1-\rho)T}$ bits/s.

Recalling that $X \in [0, P_H]$, the probabilities of Circuit Outage and Channel Outage are therefore:

$$\begin{aligned} P_{out}^{circuit} &= \Pr\{P < P_c\} \\ &= \begin{cases} F_X[\phi(\cdot)] & \text{if } P_H > \phi(\cdot) \\ 1 & \text{otherwise.} \end{cases} \end{aligned} \tag{5.4}$$

$$\begin{aligned} P_{out}^{channel} &= \Pr\{\log_2(1+(P-P_c)\Gamma) < R_{\text{eff}}, P > P_c\} \\ &= \Pr\left\{\Gamma < \frac{2^{R_{\text{eff}}}-1}{P-P_c}, P > P_c\right\} \\ &= \begin{cases} \int_{\phi(\cdot)}^{P_H} f_X(x) F_\Gamma[g(\cdot)]\, dx & \text{if } P_H > \phi(\cdot) \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{5.5}$$

where $g(\rho,\eta,P_c) = \frac{2^{\frac{Q}{(1-\rho)T}}-1}{x[\frac{\rho}{1-\rho}+\eta]-P_c}$ and $\phi(\rho,\eta,P_c) = \frac{P_c}{\frac{\rho}{1-\rho}+\eta}$. Since Circuit Outage and Channel Outage are mutually exclusive, it follows that

$$\begin{aligned} P_{out} &= P_{out}^{circuit} + P_{out}^{channel} \\ &= \begin{cases} F_X[\phi(\cdot)] + \\ \int_{\phi(\cdot)}^{P_H} f_X(x) F_\Gamma[g(\cdot)]\, dx & \text{if } P_H > \phi(\cdot) \\ 1 & \text{otherwise.} \end{cases} \end{aligned} \tag{5.6}$$

For convenience, we define

$$\hat{P}_{out}(\rho,\eta,P_c) = F_X[\phi(\cdot)] + \int_{\phi(\cdot)}^{P_H} f_X(x) F_\Gamma[g(\cdot)]\, dx \tag{5.7}$$

where $\hat{P}_{out}(\rho,\eta,P_c) < 1$ and $P_H > \phi(\cdot)$.

Unlike the conventional definition of outage probability in a block fading channel, which is dependent only on the fading distribution and a *fixed* average transmit power constraint, in an energy harvesting system with block fading and the ST protocol, *both*

transmit power and channel are random, and the resulting outage is thus a function of the save-ratio ρ, the battery efficiency η and the circuit power P_c.

5.4 Outage Minimization

In this section, we design the save-ratio ρ for the ST protocol by solving the optimization problem

$$(\mathrm{P1}): \ \min_{0\le\rho\le 1} \ P_{out}$$

i.e., minimize average outage performance P_{out} in (5.6) over ρ, for any given $\eta \in [0,1]$ and $P_c \in [0,\infty)$. Denote the optimal (minimum) outage probability as $P^*_{out}(\eta, P_c)$ and the optimal save-ratio as $\rho^*(\eta, P_c)$. Note that $\rho \nearrow 1$ represents transmission of a very short burst at the end of each frame, and the rest of each frame is reserved for MESD EH. $\rho = 0$ is another special case, in which the energy consumed in frame i was collected (by the SESD) entirely in frame $i-1$. (P1) can always be solved through numerical search, but it is challenging to give a closed-form solution for $\rho^*(\eta, P_c)$ in terms of P_c and η in general. We will instead analyze how $\rho^*(\eta, P_c)$ varies with P_c and η and thereby get some insights in the rest of this section.

Proposition 5.4.1. *$P_{out}(\rho, \eta, P_c)$ in (5.6) is a non-increasing function of battery efficiency η and a non-decreasing function of circuit power P_c for $\rho \in [0,1)$. The optimal value of (P1) $P^*_{out}(\eta, P_c)$ is strictly decreasing with η and strictly increasing with P_c.*

Proof. Please refer to Appendix C.1. □

The intuition of Proposition 5.4.1 is clear: If η grows, the energy available to the transmitter can only grow or remain the same, whatever the values of ρ and P_c, hence P_{out} must be non-increasing with η; if P_c grows, the energy available for transmission decreases, leading to higher P_{out}.

5.4.1 Ideal System: $\eta = 1$ and $P_c = 0$

Suppose that circuit power is negligible, i.e., all the energy is spent on transmission, and the SESD has perfect energy-transfer efficiency. The condition $P_H > P_c/(\frac{\rho}{1-\rho}+\eta)$ is always satisfied, and problem (P1) is simplified to

$$(\mathrm{P2}): \min_{0\le\rho\le1} \int_0^{P_H} f_X(x) F_\Gamma \left[\frac{(2^{\frac{Q}{(1-\rho)T}}-1)(1-\rho)}{x}\right] dx$$

where the optimal value of (P2) is denoted as $P^*_{out}(1,0)$, and the optimal save-ratio is denoted as $\rho^*(1,0)$.

Lemma 5.4.1. *The minimum outage probability when $\eta = 1$ and $P_c = 0$ is given by*

$$P^*_{out}(1,0) = \int_0^{P_H} f_X(x) F_\Gamma \left[\frac{2^{Q/T}-1}{x}\right] dx \tag{5.8}$$

and is achieved with the save-ratio $\rho^*(1,0) = 0$.

Proof. Please refer to Appendix C.2. □

Lemma 5.4.1 indicates that the optimal strategy for a transmitter that uses no power to operate its circuitry powered by two ESDs with 100 percent efficiency, is to transmit continuously.[3] This is not surprising because the SESD collects energy from the environment just as efficiently as the MESD does, and so idling the transmitter while the MESD harvests energy wastes transmission resources (time) while not reaping any gains (energy harvested). However, we will see that this is only true when there is no circuit power and perfect battery efficiency.

5.4.2 Inefficient Battery: $\eta < 1$ and $P_c = 0$

When the SESD energy transfer efficiency $\eta < 1$ and $P_c = 0$, (P1) becomes

$$(\mathrm{P3}): \min_{0\le\rho\le1} \int_0^{P_H} f_X(x) F_\Gamma \left[\frac{(2^{\frac{Q}{(1-\rho)T}}-1)}{x(\frac{\rho}{1-\rho}+\eta)}\right] dx$$

where the optimal value of (P3) is denoted as $P^*_{out}(\eta,0)$, and the optimal save-ratio is denoted as $\rho^*(\eta,0)$.

[3] Except for the time needed in each slot to transfer energy from the SESD to the MESD, which we assume to be negligible.

Lemma 5.4.2. *When SESD energy transfer efficiency $\eta < 1$ and circuit power $P_c = 0$, the optimal save-ratio ρ has the following properties.*

1. *A "phase transition" behavior:*

$$\begin{cases} \rho^*(\eta, 0) = 0, & \eta \in \left[\frac{2^{\frac{Q}{T}} - 1}{2^{\frac{Q}{T}} (\ln 2) \frac{Q}{T}}, 1 \right) \\ \rho^*(\eta, 0) > 0, & \eta \in \left[0, \frac{2^{\frac{Q}{T}} - 1}{2^{\frac{Q}{T}} (\ln 2) \frac{Q}{T}} \right) \end{cases} \tag{5.9}$$

2. *$\rho^*(\eta, 0)$ is a non-increasing function of η, for $0 \leq \eta \leq 1$.*

Proof. Please refer to Appendix C.3. □

According to (5.9), if the SESD efficiency is above a threshold, the increased energy available to the transmitter if the MESD rather than the SESD collects energy over $[0, \rho T]$ is not sufficient to overcome the extra energy required to transmit at the higher rate R_{eff} over $(\rho T, T]$. The result is that the optimal ρ is 0. On the other hand, if η is below that threshold, then some amount of time should be spent harvesting energy using the higher-efficiency MESD even at the expense of losing transmission time. Lemma 5.4.2 quantifies precisely the interplay among η, Q, T and ρ.

We should note here that even though we consider the case of having two ESD's, by setting $\eta = 0$, we effectively remove the SESD and hence our analysis applies also to the single-ESD case. According to (5.9), if we only have one ESD, the optimal save-ratio is $\rho^*(0, 0)$, which is always larger than 0. This is intuitively sensible, because with only one ESD obeying the energy half-duplex constraint, it would be impossible to transmit all the time ($\rho = 0$) because that would leave no time at all for EH.

5.4.3 Non-Zero Circuit Power: $\eta \leq 1$ and $P_c > 0$

Non-zero circuit power P_c leads to two mutually exclusive effects: (i) inability to power on the transmitter for the $(1 - \rho)T$ duration of transmission – this is when $P_H < \phi(\cdot)$ in (5.6); and (ii) higher outage probability if $P_H > \phi(\cdot)$ because some power is devoted to running the hardware.

Since $\frac{P_c}{\frac{\rho}{1-\rho} + \eta}$ decreases as ρ increases, its maximum value is $\frac{P_c}{\eta}$. Therefore, if $P_H > \frac{P_c}{\eta}$, the transmitter would be able to recover enough energy (with non-zero

probability) to power on the transmitter, i.e., $\rho \in [0,1)$. If $P_H \leq \frac{P_c}{\eta}$, by condition $P_H \leq \frac{P_c}{\frac{\rho}{1-\rho}+\eta}$, save-ratio ρ is required to be larger than $\frac{\frac{P_c}{P_H}-\eta}{1-\eta+\frac{P_c}{P_H}}$. In summary,

- If $P_c < P_H\eta$

$$P_{out} = \hat{P}_{out}(\rho, \eta, P_c), \quad \forall \rho \in [0,1)$$

- If $P_c \geq P_H\eta$

$$P_{out} = \begin{cases} 1, & \rho \leq \frac{\frac{P_c}{P_H}-\eta}{1-\eta+\frac{P_c}{P_H}} \\ \hat{P}_{out}(\rho, \eta, P_c), & \rho > \frac{\frac{P_c}{P_H}-\eta}{1-\eta+\frac{P_c}{P_H}}. \end{cases} \tag{5.10}$$

If $P_c \geq \eta P_H$, referring to (5.10), we may conclude that $\rho^*(\eta, P_c) > \frac{\frac{P_c}{P_H}-\eta}{1-\eta+\frac{P_c}{P_H}}$ due to the need to offset circuit power consumption. If $P_c < \eta P_H$, theoretically, the transmitter is able to recover enough energy (with non-zero probability for all $\rho \in [0,1)$) to transmit.

Lemma 5.4.3. *For an EH transmitter with battery efficiency η and non-zero circuit power P_c,*

$$\eta - \frac{P_c}{P_H} < \frac{2^{\frac{Q}{T}}-1}{2^{\frac{Q}{T}}(\ln 2)\frac{Q}{T}} \implies \rho^*(\eta, P_c) > 0. \tag{5.11}$$

Proof. Please refer to Appendix C.4. □

Intuitively, the smaller the circuit power, the more energy we can spend on transmission; the larger the battery efficiency is, the more energy we can recover from EH. Small circuit power and high battery efficiency suggests continuous transmission ($\rho^*(\eta, P_c) = 0$), which is consistent with our intuition. According to Lemma 5.4.3, larger circuit power may be compensated by larger ESD efficiency (when the threshold for η is smaller than 1). A non-zero save-ratio is only desired if there exists significant circuit power to be offset or substantial ESD inefficiency to be compensated. The threshold depends on required transmission rate.

Remark 5.4.1. It is worth noticing that if we ignore the battery inefficiency or set $\eta = 1$, Lemma 5.4.3 could be simplified as

$$P_c > \frac{2^{\frac{Q}{T}}(\ln 2)\frac{Q}{T} - 2^{\frac{Q}{T}} + 1}{2^{\frac{Q}{T}}(\ln 2)\frac{Q}{T}} P_H \implies \rho^*(1, P_c) > 0 \tag{5.12}$$

where only circuit power P_c impacts the save-ratio. Since the MESD and the SESD are equivalent ($\eta = 1$), harvesting energy using the MESD is not the reason for delaying transmission. Instead, $\rho^* > 0$ when P_c is so large that we should transmit over a shorter interval at a higher power, so that the actual transmission power minimizes P_{out}. Circuit power similarly determined the fundamental tradeoff between energy efficiency and spectral efficiency (data rate) in [101], in which it was shown that with additional circuit power making use of all available time for transmission is not the best strategy in terms of both energy and spectral efficiency. In this chapter, outage is minimized through utilizing available (random) energy efficiently, wherein circuit power causes a similar effect.

5.5 Diversity Analysis

The outage performance of wireless transmission over fading channels at high SNR can be conveniently characterized by the so-called *diversity order* [30], which is the high-SNR slope of the outage probability determined from a SNR-outage plot in the log-log scale. Mathematically, the diversity order is defined as

$$d = -\lim_{\bar{\gamma}\to\infty} \frac{\log_{10}(P_{out})}{\log_{10}(\bar{\gamma})} \tag{5.13}$$

where P_{out} is the outage probability and $\bar{\gamma}$ is the average SNR.

Diversity order under various fading channel conditions has been comprehensively analyzed in the literature (see e.g., [30] and references therein). Generally speaking, if the fading channel power distribution has an accumulated density near zero that can be approximated by a polynomial term, i.e., $\Pr\left(|h|^2 \leq \epsilon\right) \approx \epsilon^k$, where ϵ is an arbitrary small positive constant, then the constant k indicates the diversity order of the fading channel. For example, in the case of Rayleigh fading channel with $\Pr\left(|h|^2 \leq \epsilon\right) \approx \epsilon$, the diversity order is thus 1 according to (5.13).

However, the above diversity analysis is only applicable to conventional wireless systems in which the transmitter has a constant power supply. Since EH results in

random power availability in addition to fading channels, the PDF of the receiver SNR due to both random transmit power and random channel power may not necessarily be polynomially smooth at the origin (as we will show later). As a result, the conventional diversity analysis with constant transmit power cannot be directly applied. In this section, we will investigate the effect of random power on diversity analysis, as compared with the conventional constant-power case. For clarity, in the rest of this section, we consider the ideal system with $\eta = 1$ and $P_c = 0$, and the Rayleigh fading channel with $\mathbb{E}[\Gamma] = \mathbb{E}[\frac{|h|^2}{\sigma_n^2}] = \frac{\sigma_h^2}{\sigma_n^2} = \lambda_\gamma$.

From (5.5) and (5.6), the outage probability when $\eta = 1$ and $P_c = 0$ is given by

$$P_{out} = \Pr\left\{\log_2(1 + P\Gamma) < \frac{Q}{(1-\rho)T}\right\}. \tag{5.14}$$

Based on Lemma 5.4.1, the minimum outage probability is achieved with the save-ratio $\rho = 0$. Therefore, the outage probability is simplified as[4]

$$P_{out}^* = \Pr\{P\Gamma < C\} = \int_0^\infty \int_0^{\frac{C}{P}} f_P(p) f_\Gamma(\gamma) d\gamma dp \tag{5.15}$$

where $C = 2^{\frac{Q}{T}} - 1$ and the last equality comes from the assumption of Rayleigh fading channel so the Γ is exponential distributed. It is worth noting that in this case with $\eta = 1$ and $\rho = 0$, according to (5.2), the energy arrival rate X and the average total power P are identical and thus can be used interchangeably.

Clearly, the near-zero behavior of P_{out}^* critically depends on the PDF of random power $f_P(p)$, while intuitively we should expect that random power can only degrade the outage performance. We choose to use the Gamma distribution to model the random power P, because the Gamma distribution models many positive random variables (RVs) [102, 103]. The Gamma distribution is very general, including exponential, Rayleigh, and Chi-Square as special cases; furthermore, the PDF of any positive continuous RV can be properly approximated by the sum of Gamma PDFs. Supposing that P follows a Gamma distribution denoted by $P \sim \mathcal{G}(\beta, \lambda_p)$, then its PDF is given by

$$f_P(p) = \frac{p^{\beta-1}\exp\left(-\frac{p}{\lambda_p}\right)}{\lambda_p^\beta \Gamma(\beta)} U(p) \tag{5.16}$$

[4]For convenience, P_{out}^* is used to represent $P_{out}^*(1, 0)$ in the rest of this section.

where $U(\cdot)$ is the unit step function, $\Gamma(\cdot)$ is the Gamma function, and $\beta > 0$, $\lambda_p > 0$ are given parameters. Referring to [104, Lemma 2], which gives the distribution of the product of m Gamma RVs, the outage probability in (5.15) can be computed as

$$P_{out}^{*} = \frac{1}{\Gamma(\beta)} G_{13}^{21}\left(\frac{C\lambda_\gamma}{\lambda_p}\middle|\begin{matrix}1\\ 1,\beta,0\end{matrix}\right) \tag{5.17}$$

where $G(\cdot)$ is the Meijer G-function [102].

Meijer G-function can in general only be numerically evaluated and does not give much insights about how random power affects the outage performance. Next, we further assume that the random power P is exponentially distributed (as a special case of Gamma distribution with $\beta = 1$) to demonstrate the effect of random power.

Lemma 5.5.1. *Suppose that P is exponentially distributed with mean λ_p, the channel is Rayleigh fading with $\mathbb{E}[\Gamma] = \frac{\sigma_h^2}{\sigma_n^2} = \lambda_\gamma$, and thus the average received SNR $\bar{\gamma} = \lambda_p\lambda_\gamma = \frac{\lambda_p\sigma_h^2}{\sigma_n^2}$. The minimum outage probability P_{out}^{*}, under an ideal system with $\eta = 1$ and $P_c = 0$, is given by*

$$P_{out}^{*} = \sum_{k=0}^{\infty} \frac{C^{k+1}}{(k!)^2(k+1)\bar{\gamma}^{k+1}}\left[\frac{1}{k+1} - \ln\frac{C}{\bar{\gamma}} + 2\psi(k+1)\right] \tag{5.18}$$

where $\psi(x)$ is the digamma function [105] and $\ln(\cdot)$ represents the natural logarithm.

Proof. Please refer to Appendix C.5. □

In the asymptotically high-SNR[5] regime, we can approximate P_{out}^{*} by taking only the first term of (5.18) as

$$P_{out}^{*} \approx \frac{C}{\bar{\gamma}}\left(1 - \ln\frac{C}{\bar{\gamma}} + 2\psi(1)\right) \approx \frac{\ln\bar{\gamma}}{\bar{\gamma}}. \tag{5.19}$$

As observed, P_{out}^{*} decays as $\bar{\gamma}^{-1}\ln(\bar{\gamma})$ rather than $\bar{\gamma}^{-1}$ as in the conventional case with constant power, which indicates that the PDF of the receiver SNR is no longer polynomially smooth near the origin. Hence, the slope of P_{out}^{*} in the SNR-outage plot, or the diversity order, will converge much more slowly to $\bar{\gamma}^{-1}$ with SNR than in the

[5]We assume that high SNR is achieved via decreasing noise power σ_n^2, while fixing the average harvested energy.

constant-power case, suggesting that random energy arrival has a significant impact on the diversity performance. More specifically, we obtain the diversity order in the case of exponentially distributed random power as

$$d = -\lim_{\bar{\gamma}\to\infty} \frac{-\log_{10}\bar{\gamma} + \log_{10}(\ln\bar{\gamma})}{\log_{10}\bar{\gamma}} = 1 \tag{5.20}$$

which is, in principle, the same as that over the Rayleigh fading channel with constant power. We thus conclude that diversity order may not be as meaningful a metric of evaluating outage performance in the presence of random power, as in the conventional case of constant power.

5.6 Optimization with Multiple Transmitters

In this section, we extend the ST protocol for the single-channel case to the more practical case of multiple transmitters in a wireless network, and quantify the system-level outage performance as a function of the number of transmitters in the network.

5.6.1 TDMA based Save-Then-Transmit Protocol

We consider a wireless network with N transmitters, each of which needs to transmit Q bits of data within a time frame of duration T seconds to a common fusion center (FC). It is assumed that each transmitter is powered by the same EH circuit model as shown in Fig. 5.1, and transmits over the baseband-equivalent channel model given in (5.1). We also assume a homogeneous system setup, in which the channel gains, EH rates or additive noises for all transmitter-FC links are independent and identically distributed (i.i.d).

In order to allow multiple transmitters to communicate with the FC, we propose a TDMA based ST (TDMA-ST) protocol as follows (cf. Fig. 5.3):

- Every frame is equally divided into N orthogonal time slots with each slot equal to $\frac{T}{N}$ seconds.

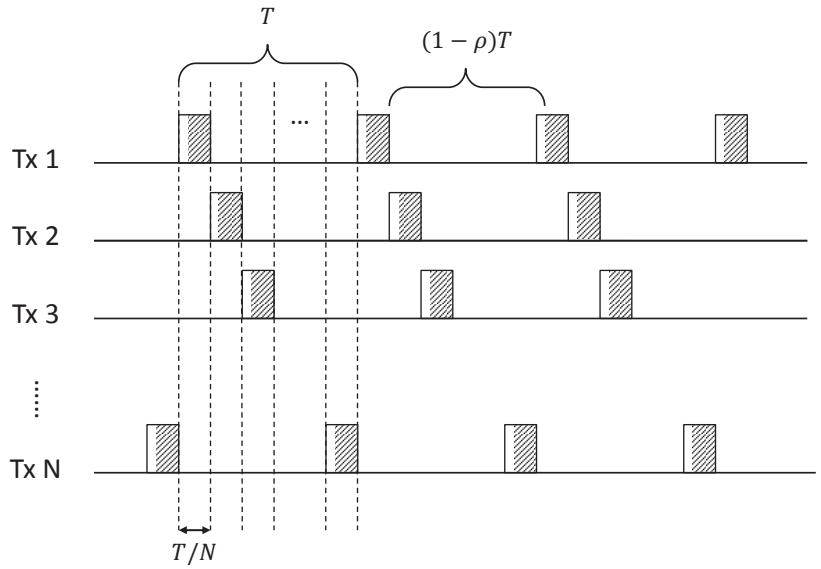


Figure 5.3: TDMA based ST (TDMA-ST)

- Assuming perfect time synchronization, each transmitter is assigned a different (periodically repeating) time slot for transmission, i.e., in each frame, transmitter i is allocated the time slot $\left[\frac{(i-1)}{N}T,\ \frac{i}{N}T\right)$, $1 \leq i \leq N$.
- Assuming $\rho_i = \rho$ for all i's, each transmitter implements the ST protocol with the transmission time in each frame aligned to be within its assigned time slot; as a result, the maximum *transmit-ratio*, denoted by $1-\rho$, for each transmitter cannot exceed $1/N$, which means that $\rho \geq 1 - \frac{1}{N}$.

The protocol described above is illustrated in Fig. 5.3. Unlike the single-channel case where the transmitter can select any save-ratio ρ in the interval $0 \leq \rho \leq 1$, in the case of TDMA-ST, ρ is further constrained by $\rho \geq 1 - \frac{1}{N}$ to ensure orthogonal transmissions by all transmitters. Due to this limitation, each transmitter may not be able to work at its individual minimum outage probability unless the corresponding optimal save-ratio ρ^* satisfies $\rho^* \geq 1 - \frac{1}{N}$ or $N \leq \frac{1}{1-\rho^*}$. In this case, ST protocol naturally extends to TDMA-ST with every transmitter operating at the optimal save-ratio ρ^*. However, if N exceeds the threshold $\frac{1}{1-\rho^*}$, transmitters have to deviate from ρ^* to maintain orthogonal transmissions. Next, we evaluate the performance

of TDMA-ST for two types of source data at transmitters: Independent Data and Common Data.

5.6.2 Independent Data Transmission

First, consider the case where all transmitters send independent data packets to the FC in each frame, which are decoded separately at the FC. Under the symmetric setup, for a given ρ, all transmitters should have the same average outage performance. Consequently, the system-level outage performance in the case of independent data can be equivalently measured by that of the individual transmitter, i.e.,

$$P_{out}^{s} = P_{out}(\rho, \eta, P_c). \tag{5.21}$$

We can further investigate the following two cases:

- $N \leq \frac{1}{1-\rho^*}$

  In this case, the additional constraint due to TDMA, $\rho^* \geq 1 - \frac{1}{N}$, is satisfied. Since P_{out}^{s} is the same as that of the single-transmitter case, the system is optimized when all transmitters work at their individual minimum outage with save-ratio ρ^*. Thus, the minimum system outage probability is $P_{out}^{s*} = P_{out}^{*}(\eta, P_c)$.

- $N > \frac{1}{1-\rho^*}$

  In this case, the TDMA constraint on ρ^* is violated and thus we are not able to allocate all transmitters the save-ratio ρ^*, which means that each transmitter has to deviate from its minimum outage point. Since in this case $\rho^* < 1 - \frac{1}{N} \leq \rho$, the best strategy for each transmitter is to choose $\rho = 1 - \frac{1}{N}$. Thus, $P_{out}^{s*} = P_{out}(1 - \frac{1}{N}, \eta, P_c)$.

To summarize, the optimal strategy for each transmitter in the case of independent data is given by

$$\rho = \begin{cases} \rho^*, & N \leq \frac{1}{1-\rho^*} \\ 1 - \frac{1}{N}, & N > \frac{1}{1-\rho^*} \end{cases} \tag{5.22}$$

which implies that the number of transmitters should be kept below the reciprocal of the single-channel optimal transmit-ratio; otherwise, the system outage performance will degrade.

5.6.3 Common Data Transmission

Next, consider the case where all transmitters send identical data packets in each frame to the FC, which applies diversity combining techniques to decode the common data. For simplicity, we consider selection combining (SC) at the receiver, but similar results can be obtained for other diversity combining techniques [30]. With SC, the system outage probability is given by [30]

$$P_{out}^{s} = \left(P_{out}(\rho, \eta, P_c)\right)^N . \tag{5.23}$$

Similarly to the case of independent data, we can get exactly the same result for the optimal transmit strategy given in (5.22) for the common-data case, with which the minimum system outage probability is obtained as

$$P_{out}^{s*} = \begin{cases} \left(P_{out}^{*}(\eta, P_c)\right)^N , & N \leq \frac{1}{1-\rho^*} \\ \left(P_{out}(1-\frac{1}{N}, \eta, P_c)\right)^N , & N > \frac{1}{1-\rho^*}. \end{cases} \tag{5.24}$$

From the above, it is evident that the system outage probability initially drops as N increases, provided that $N \leq \frac{1}{1-\rho^*}$. However, when $N > \frac{1}{1-\rho^*}$, it is not immediately clear whether the system outage increases or decreases with N, since increasing N improves the SC diversity, but also makes each transmitter deviate even further from its minimum outage save-ratio according to (5.22).

5.7 Numerical Results

In this section, we provide numerical examples to validate our claims. We assume that the EH rate X follows a uniform distribution (unless specified otherwise) within $[0, 100]$ (i.e., $P_H = 100$ J/s), and the channel is Rayleigh fading with exponentially

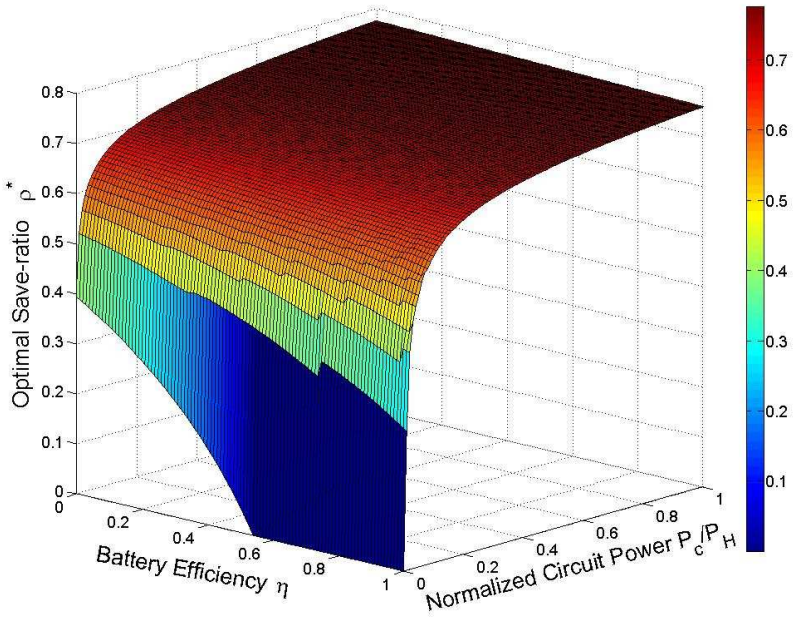


Figure 5.4: Optimal save-ratio ρ^*

distributed Γ with parameter $\lambda = 0.02$. We also assume the target transmission rate $R_{\text{req}} = \frac{Q}{T} = 2$ bits/s.[6]

Fig. 5.4 demonstrates how battery efficiency η and circuit power P_c affect the optimal save-ratio ρ^* for the single-channel case. As observed, larger P_c and smaller η result in larger ρ^*, i.e., shorter transmission time. Since the increment is more substantial along P_c axis, circuit power has a larger influence on the optimal save-ratio compared with battery efficiency. $\rho^*(1,0) = 0$ verifies the result of Lemma 5.4.1 for an ideal system, while $\rho^*(\eta, 0)$ along the line $P_c = 0$ demonstrates the "phase transition" behavior stated in Lemma 5.4.2. The transition point is observed to be $\eta = 0.541$, which can also be computed from (5.9).

Fig. 5.5 shows the optimal (minimum) outage probability $P^*_{out}(\eta, P_c)$ corresponding to ρ^* in Fig. 5.4. Consistent with Proposition 5.4.1, $P^*_{out}(\eta, P_c)$ is observed to be monotonically decreasing with battery efficiency η and monotonically increasing with circuit power P_c. Again, P_c affects outage performance more

[6]This is normalized to a bandwidth of 1 Hz, i.e., R_{req} is the spectral efficiency in bits/s/Hz.

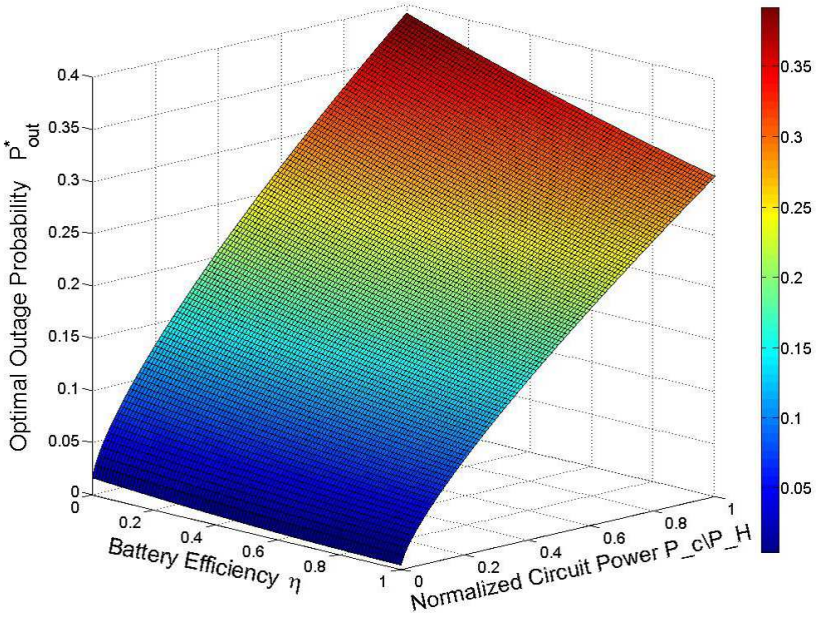


Figure 5.5: Optimal outage probability P_{out}^*

significantly than η. From Fig. 5.5, we see that for a reasonable outage probability e.g. below 0.05, P_c has to be small and η has to be close to 1. Our results can thus be used to find the feasible region in the $\eta - P_c$ plane for a given allowable P_{out}.

Figs. 5.6 and 5.7 compare the outage performance with versus without save-ratio optimization. In Fig. 5.6 we fix the normalized circuit power $\frac{P_c}{P_H} = 0.5$, while in Fig. 5.7 we fix the battery efficiency $\eta = 0.8$. We observe that optimizing the save-ratio can significantly improve the outage performance. It is worth noting that P_{out} has an approximately linear relationship with the normalized circuit power $\frac{P_c}{P_H}$ as observed in Fig. 5.7, which indicates that P_c considerably affects the outage performance as stated previously.

In Fig. 5.8, the outage probability for an ideal system ($\eta = 1$, $P_c = 0$) is shown by numerically evaluating (5.17). By fixing the mean value of P as $\mathbb{E}[P] = 50$ J/s and varying β for the Gamma distributed power from 1 to 5, the resulting outage performance is compared with the case of constant power. As observed, the outage probability increases due to the existence of power randomness. As β increases, the

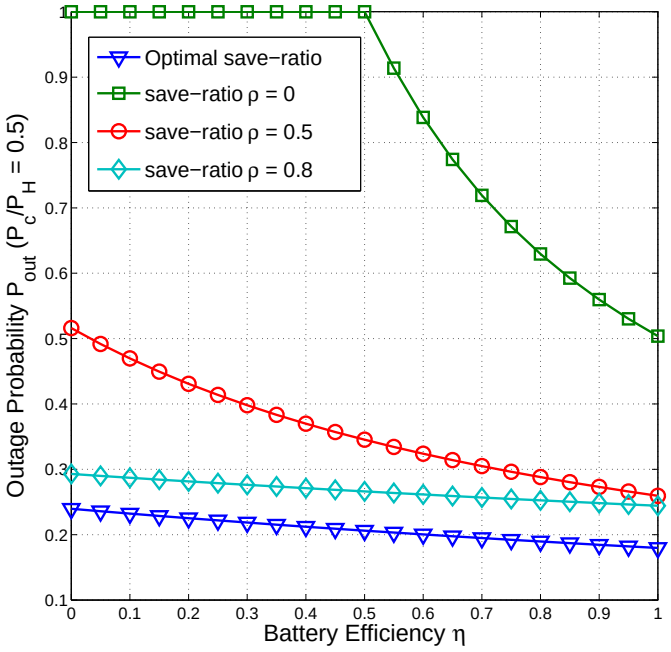


Figure 5.6: Outage performance comparison: $\frac{P_c}{P_H} = 0.5$

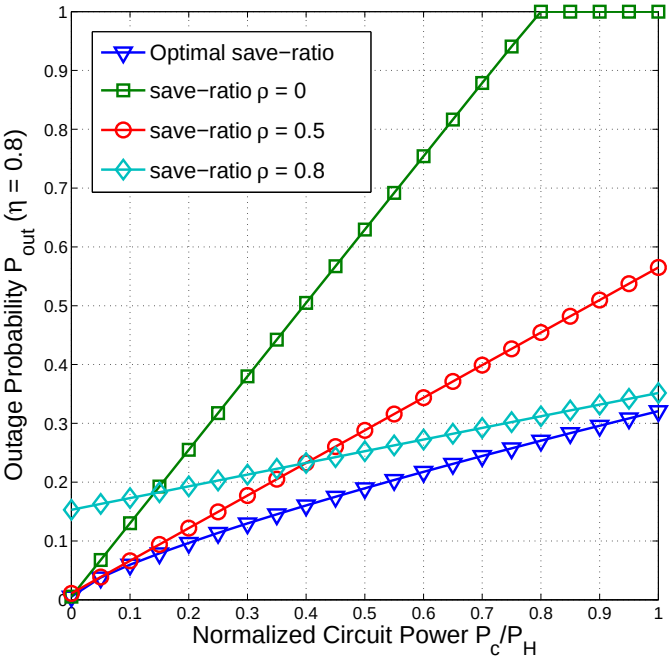


Figure 5.7: Outage performance comparison: $\eta = 0.8$

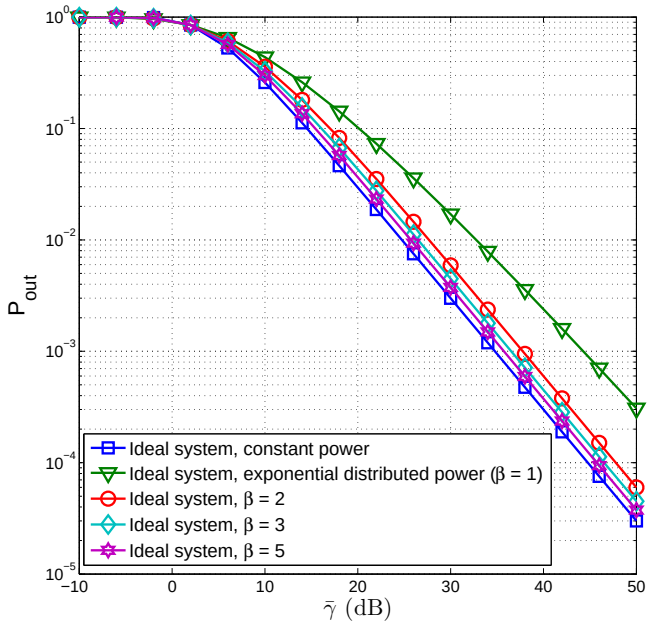


Figure 5.8: Outage probability for an ideal ($\eta = 1$, $P_c = 0$) system with constant power versus random power

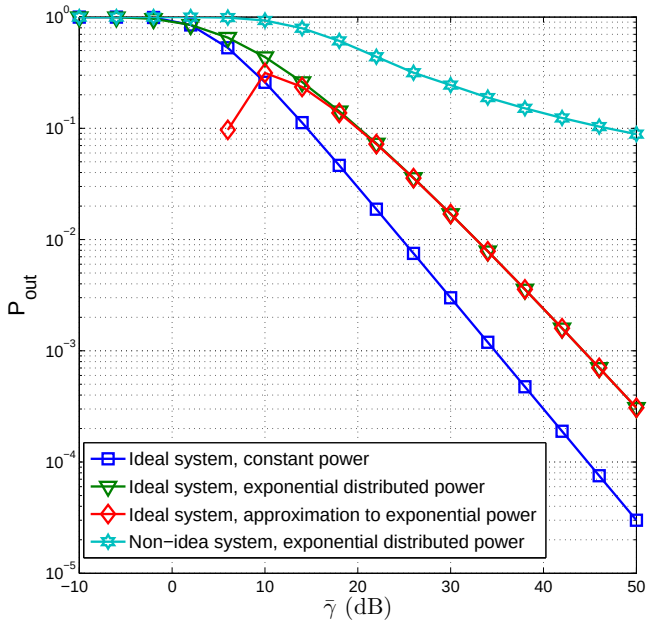


Figure 5.9: Outage probability comparison for ideal ($\eta = 1$, $P_c = 0$) versus non-ideal ($\eta = 0.8$, $P_c = 0.1 * \mathbb{E}[P]$) systems

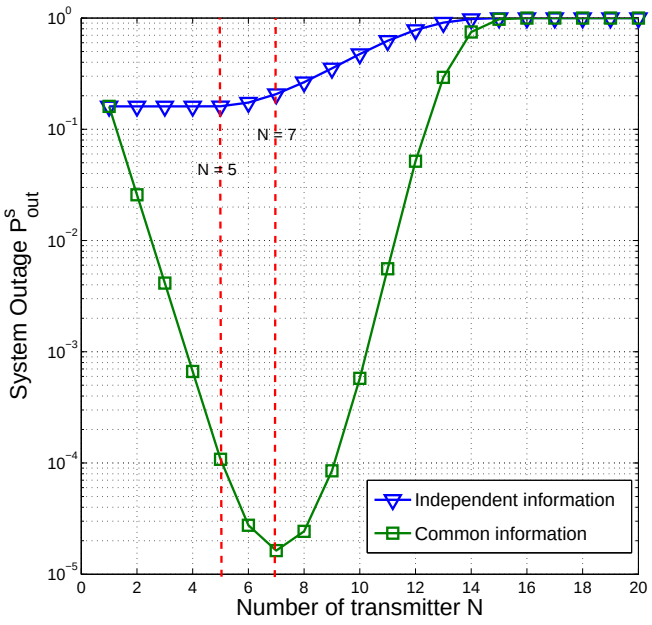


Figure 5.10: Outage performance of multiple transmitters under TDMA-ST protocol, with $\frac{1}{1-\rho^*} = 4.83$

outage curve approaches the case of constant power. In Fig. 5.9, we also plot the outage probability for the ideal system with exponentially distributed power based on the approximation given in (5.19), as well as for a non-ideal system with the normalized circuit power $\frac{P_c}{\mathbb{E}[P]} = 0.1$ and battery efficiency $\eta = 0.8$. In comparison with the constant-power case, for the case of ideal system we observe that the high-SNR slope or diversity order with random power clearly converges much slower with SNR, which is in accordance with our analysis in Section 5.5. Furthermore, at $P_{out} = 10^{-3}$, there is about 10 dB power penalty observed due to exponential random power, even with the same diversity order as the constant-power case. It is also observed that there is a small rising part for the outage approximation given in (5.19), since this approximation is only valid for sufficiently high SNR values ($\bar{\gamma} > 10$ dB). Finally, it is worth noting that the outage probability for the non-ideal system eventually saturates with SNR (regardless of how small the noise power is or how large the SNR is), which indicates that the diversity order is zero for any non-ideal system.

Fig. 5.10 shows the outage performance for the case of multiple transmitters

operating under the TDMA-ST protocol. We set the normalized circuit power $\frac{P_c}{P_H} = 0.5$ and the battery efficiency $\eta = 0.9$. Then, the optimal save-ratio ρ^* for single-transmitter outage minimization can be obtained as 0.7930 by numerical search. Therefore, the threshold value for N in the optimal rule of assigning save-ratio values in (5.22) is $\frac{1}{1-\rho^*} = 4.83$. For the case of independent data, it is observed that when $N \leq 4$, the system outage probability is constantly equal to the optimal single-transmitter outage probability $P^*_{out}(0.9, 0.5P_H)$; however, as $N > 4$, the outage probability increases dramatically. In contrast, for the case of common data, it is observed that the system outage probability decreases initially as N increases, even after the threshold value and until $N = 7$, beyond which it starts increasing. This implies that there is an optimal decision on the number of transmitters to achieve the optimal outage performance.

5.8 Chapter Summary

In this chapter, we studied a wireless system under practical energy harvesting conditions. Assuming a general model with non-ideal energy storage efficiency and transmit circuit power, we proposed a Save-then-Transmit (ST) protocol to optimize the system outage performance via finding the optimal save-ratio. We characterized how the optimal save-ratio and the minimum outage probability vary with practical system parameters. We compared the outage performance between random power and constant power under the assumption of Rayleigh fading channel. It is shown that random power considerably degrades the outage performance. Furthermore, we presented a TDMA-ST protocol for wireless networks with multiple EH transmitters. In particular, two types of source data were examined: independent data and common data. It was shown that if the number of transmitters is smaller than the reciprocal of the optimal transmit-ratio for single-transmitter outage minimization, each transmitter should work with its minimum outage save-ratio; however, when the number of transmitters exceeds this threshold, each transmitter has to deviate from its individual optimal operating point.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Minimizing the overall energy consumption and yet meeting the increasing traffic demand has become an urgent need for wireless networks today. This thesis has made an innovative and comprehensive investigation of energy-efficient and energy-saving communication techniques via resource allocation optimizations in wireless networks. We summarize the main contributions of this thesis as follows.

- In Chapter 2, under an OFDMA-based broadcast channel setup, we investigated optimal power and range adaptation polices with time-varying traffic to minimize the BS's average power consumption subject to the network throughput and user-level QoS constraints. A new power scaling law that relates the (short-term) average transmit power at BS with the given cell range and MT density was derived, based on which we obtained the optimal power and range adaptation policy by solving a joint cell range adaptation and (long-term) power control problem. The obtained results provide a preliminary unified framework for evaluating the performance of existing cell adaptation schemes such as BS's on-off switching and cell zooming.

- In Chapter 3, under an OFDM-based DL communication setup, we characterized the tradeoffs in minimizing the BS's versus MTs' energy consumptions by investigating a weighted-sum transmitter and receiver joint energy minimization

(WSTREMin) problem, subject to an average transmit power constraint at the BS and data requirements of individual MTs. Moreover, we proposed a new multiple access scheme termed Time-Slotted OFDMA (TS-OFDMA) transmission scheme, to achieve more flexible energy consumption tradeoffs between the BS and MTs. Our results provide important new insights to the optimal design of next generation cellular networks with their challenging requirements on both the spectral and energy efficiency of the network.

- In Chapter 4, we studied the problem of joint DL and UL MT-AP association and beamforming design to optimize the energy consumption tradeoffs between the active APs and MTs in the context of the emerging C-RAN. Leveraging the celebrated UL-DL duality result, the joint DL and UL problem was converted to an equivalent DL problem. Based on this transformation, two efficient algorithms were proposed based on GSO and RIP techniques, respectively. Our study is the first attempt to unify the joint DL and UL MU-AP association and beamforming design into one general framework.

- In Chapter 5, we studied a point-to-point communication link with one EH transmitter and one constant-power receiver. Assuming a practical model with non-ideal energy storage efficiency and transmit circuit power, we proposed a save-then-transmit (ST) protocol to optimize the system outage performance via finding the optimal save-ratio. Moreover, we compared the outage performance between random power and constant power under the assumption of Rayleigh fading channel, which shows that random power considerably degrades the outage performance. Our studies provide important insights for designing EH enabled wireless systems.

6.2 Future Work

We highlight several future work directions in the following which we deem important and worthy of further investigations by extending the results presented in this thesis.

- For the dynamic power and range adaptation in Chapter 2, we focused on the simplified case of one single cell for the purpose of obtaining useful insights, which needs to be extended to the more practical multi-cell scenario. It is thus interesting as well as important to investigate the optimal cell adaptation policy in a cooperative multi-cell setup via balancing between the cellular network energy consumption and its coverage performance by extending the mathematical framework developed in Chapter 2.

- In Chapter 4, we focused on addressing the energy consumption issue in the C-RAN by assuming that the fronthaul transport network is provisioned with sufficiently large capacity. However, as mentioned in Chapter 1, the data exchanged between the APs and the BBU pool includes oversampled real-time digital signals with very high bit rates (in the order of Gbps). As a result, the capacity requirement for the backhaul/fronthaul links becomes far more stringent in the C-RAN as compared to in traditional networks. Thus, practical strategies for fronthaul capacity allocation as well as fronthaul compression and quantization are worth pursuing.

- For the proposed ST protocol, only a simple point-to-point communication link with one EH transmitter and one constant-power receiver was studied in Chapter 5. It is thus interesting to extend the ST protocol to the general case of multiple EH powered transmitters and receivers, or multiple-hop transmission. Another possible direction is to consider the effect of different configurations of battery/supercapacitor and MESD/SESD on the system performance, for the purpose of finding the most efficient circuit model for EH powered transmitter and/or receiver.

- Last but not least, in this thesis, the circuit energy consumption has been considered for all communication devices, which is assumed to be relatively independent of data rate at both the transmitter and receiver and thus generally be constant for simplicity. However, studying a more general energy consumption model, which includes the transmission power, signal processing power (may

be related to the transmission rate and channel bandwidth) and fixed operation power from both the transmitter and receiver, is expected to further improve the practical system performance in terms of energy efficiency.

Appendix A

Appendices to Chapter 2

A.1 Proof of Theorem 2.3.1

First, $\mathbb{E}[P_i(r_i,n)]$ is computed based on (2.15) as follows, where $P_i(r_i,n)$ is given by (2.11) with N replaced by n.

$$\mathbb{E}[P_i(r_i,n)] = \frac{2\Gamma N_0 W(2^{nC_2}-1)}{KC_1(\alpha+2)r_0^{\alpha}n}\left(R^{\alpha}+\frac{\alpha r_0^{\alpha+2}}{2R^2}\right). \tag{A.1}$$

Since $\mathbb{E}[P_i(r_i,n)]$ is identical for all i's, according to (2.14), $\mathbb{E}[P_t|N]$ can be simply obtained through multiplying $\mathbb{E}[P_i(r_i,n)]$ by the number of MTs n, i.e.

$$\begin{aligned}\mathbb{E}[P_t|N] &= n\mathbb{E}[P_i(r_i,n)] \\ &= \frac{2\Gamma N_0 W(2^{nC_2}-1)}{KC_1(\alpha+2)r_0^{\alpha}}\left(R^{\alpha}+\frac{\alpha r_0^{\alpha+2}}{2R^2}\right).\end{aligned} \tag{A.2}$$

Averaging (A.2) over the Poisson distribution of N, we finally obtain $\bar{P}_t$ as

$$\bar{P}_t = \sum_{n=0}^{\infty}\frac{2\Gamma N_0 W(2^{nC_2}-1)}{KC_1(\alpha+2)r_0^{\alpha}}\left(R^{\alpha}+\frac{\alpha r_0^{\alpha+2}}{2R^2}\right)\frac{\mu_N^n}{n!}e^{-\mu_N} \tag{A.3}$$

$$= D_1\left(R^{\alpha}+\frac{\alpha r_0^{\alpha+2}}{2R^2}\right)\left(\sum_{n=0}^{\infty}\frac{(\mu_N 2^{C_2})^n}{n!}e^{-\mu_N}-1\right) \tag{A.4}$$

$$= D_1\left(R^{\alpha}+\frac{\alpha r_0^{\alpha+2}}{2R^2}\right)\left(e^{D_1'\pi\lambda R^2}-1\right) \tag{A.5}$$

$$\approx D_1 R^{\alpha}\left(e^{D_1'\pi\lambda R^2}-1\right) \tag{A.6}$$

where $D_1 = \frac{2\Gamma N_0 W}{KC_1(\alpha+2)r_0^{\alpha}}$ and $D_1' = 2^{\frac{\bar{v}}{W}}-1$. Note that since cell radius R is practically much larger than the reference distance r_0, we have ignored the term $\frac{\alpha r_0^{\alpha+2}}{2R^2}$ in (A.5).

It is worth noting that

$$D_1^{'} = (2^{\frac{\bar{v}}{W}} - 1) = (2^{\frac{r_{se}}{\bar{N}}} - 1) \tag{A.7}$$

where r_{se} is the system spectrum efficiency in bps/Hz and $\bar{N}$ is the nominal number of supported users, both of which are pre-designed system parameters. In practice, $r_{se} = 2 \sim 6$ bps/Hz and $\bar{N}$ is a couple of hundreds and even thousands. Therefore, $\frac{r_{se}}{\bar{N}}$ is generally a very small number such that

$$D_1^{'} \approx \frac{\bar{v}}{W} \ln 2. \tag{A.8}$$

Thus, (A.6) can be further simplified as

$$\bar{P}_t \approx D_1 R^{\alpha} \left(2^{D_2 \pi \lambda R^2} - 1\right) \tag{A.9}$$

where $D_2 = \frac{\bar{v}}{W}$. Theorem 2.3.1 is thus proved.

A.2 Proof of Lemma 2.4.1

To prove Lemma 2.4.1, the following two facts are first verified:

1. For any P_c, which yields feasible (P0), there always exist some λ such that $L_\lambda(x_1^*(\lambda), \mu) < 0$;
2. If $L_\lambda(x_1^*(\lambda_a), \mu) \leq 0$, then $L_\lambda(x_1^*(\lambda_b), \mu) < 0$ for all $\lambda_b > \lambda_a$.

The first fact can be shown by contradiction as follows. Suppose that $L_\lambda(x_1^*(\lambda), \mu)$ is always non-negative, i.e.

$$L_\lambda(x_1^*(\lambda), \mu) \geq 0, \quad \forall x > 0, \lambda \geq 0. \tag{A.10}$$

Then, according to (2.27) we have

$$x^*(\lambda) = 0, \quad \forall \lambda \geq 0 \tag{A.11}$$

which violates the throughput constraint $\mathbb{E}_\lambda \left[U(x(\lambda), \lambda)\right] \geq U_{\text{avg}}$. The first fact is thus proved.

Next, we verify the second fact. According to the first fact, there always exists a λ such that $L_\lambda(x_1^*(\lambda),\mu) < 0$. Therefore, without loss of generality, we can assume $L_\lambda(x_1^*(\lambda_a),\mu) \leq 0$, i.e.

$$\min_{x(\lambda_a)>0} \bar{P}_{\text{BS}}(x(\lambda_a),\lambda_a) - \mu U(x(\lambda_a),\lambda_a) \leq 0. \tag{A.12}$$

Then there exists at least one $x_a(\lambda_a) > 0$ such that

$$\bar{P}_{\text{BS}}(x_a(\lambda_a),\lambda_a) - \mu U(x_a(\lambda_a),\lambda_a) \leq 0 \tag{A.13}$$

or equivalently,

$$D_1 x_a(\lambda_a)^{\frac{\alpha}{2}} \left(2^{D_2\pi\lambda_a x_a(\lambda_a)} - 1\right) + P_c \leq \mu\pi\lambda_a x_a(\lambda_a). \tag{A.14}$$

For any given $\lambda_b > \lambda_a$, by letting $x_b(\lambda_b) = x_a(\lambda_a)\frac{\lambda_a}{\lambda_b}$, then

$$\begin{aligned}
& D_1 x_b(\lambda_b)^{\frac{\alpha}{2}} \left(2^{D_2\pi\lambda_b x_b(\lambda_b)} - 1\right) + P_c && \text{(A.15)}\\
=\ & D_1 x_b(\lambda_b)^{\frac{\alpha}{2}} \left(2^{D_2\pi\lambda_a x_a(\lambda_a)} - 1\right) + P_c && \text{(A.16)}\\
<\ & D_1 x_a(\lambda_a)^{\frac{\alpha}{2}} \left(2^{D_2\pi\lambda_a x_a(\lambda_a)} - 1\right) + P_c && \text{(A.17)}\\
\leq\ & \mu\pi\lambda_a x_a(\lambda_a) = \mu\pi\lambda_b x_b(\lambda_b). && \text{(A.18)}
\end{aligned}$$

Thus for any $\lambda_b > \lambda_a$, we can always find an $x_b(\lambda_b)$ such that $\bar{P}_{\text{BS}}(x_b(\lambda_b),\lambda_b) - \mu U(x_b(\lambda_b),\lambda_b) < 0$, which implies $L_\lambda(x_1^*(\lambda_b),\mu) < 0$. The second fact is thus proved.

We are now ready to prove Lemma 2.4.1. The proof is by first showing the fact that $L_\lambda(x_1^*(\lambda),\mu)$ is positive for sufficiently small λ's, and then combining this result with the two facts previously shown.

According to the first-order Taylor expansion, we have

$$\begin{aligned}
& D_1 x(\lambda)^{\frac{\alpha}{2}} \left(2^{D_2\pi\lambda x(\lambda)} - 1\right) + P_c && \text{(A.19)}\\
>\ & (\ln 2) D_1 D_2 \pi\lambda x(\lambda)^{\frac{\alpha+2}{2}} + P_c,\ \ \forall x > 0. && \text{(A.20)}
\end{aligned}$$

Let $h(x(\lambda)) = (\ln 2) D_1 D_2 \pi\lambda x(\lambda)^{\frac{\alpha+2}{2}} + P_c - \mu\pi\lambda x(\lambda)$; then the minimum value of $h(x(\lambda))$ could be easily found by its first-order differentiation, given by

$$h(x(\lambda))_{\min} = P_c - x_{\min}\lambda\mu\pi\frac{\alpha}{\alpha+2} \tag{A.21}$$

where $x_{\min} = \left(\frac{2\mu}{(\alpha+2)(\ln 2)D_1 D_2}\right)^{\frac{2}{\alpha}}$. It is easy to verify that if $\lambda < \frac{(\alpha+2)P_c}{\alpha\mu\pi x_{\min}}$, $h(x(\lambda))_{\min} > 0$. Since $L_\lambda(x(\lambda),\mu)$ is an upper bound of $h(x(\lambda))$, we have

$$L_\lambda(x(\lambda),\mu) > 0, \ \ \forall x(\lambda) > 0 \text{ and } \lambda < \frac{(\alpha+2)P_c}{\alpha\mu\pi x_{\min}} \tag{A.22}$$

which implies that

$$L_\lambda(x_1^*(\lambda),\mu) > 0, \ \ \forall \lambda < \frac{(\alpha+2)P_c}{\alpha\mu\pi x_{\min}}. \tag{A.23}$$

We thus show that $L_\lambda(x_1^*(\lambda_b),\mu)$ is positive for λ's satisfying (A.23). With the two facts given earlier, it follows that $L_\lambda(x_1^*(\lambda),\mu)$ cannot be positive for all λ's and $L_\lambda(x_1^*(\lambda),\mu)$ will remain negative once it turns to be negative for the first time as λ increases; thus, we conclude that there must exist a critical value for λ, i.e., $\lambda_1 > 0$ as given in Lemma 2.4.1. Lemma 2.4.1 is thus proved.

A.3 Proof of Lemma 2.4.2

Using the series expansion $2^x = \sum_{k=0}^{\infty} \frac{(x(\ln 2))^k}{k!}$, (2.25) is expanded as

$$x_1^*(\lambda)^{\frac{\alpha}{2}} \sum_{k=1}^{\infty} \frac{(k+\frac{\alpha}{2})((\ln 2)D_2\pi)^k(\lambda x_1^*(\lambda))^{k-1}}{k!} = \frac{\mu\pi}{D_1}. \tag{A.24}$$

It can be verified that the left-hand-side (LHS) of (A.24) is a strictly increasing function of both λ and $x_1^*(\lambda)$. Thus, to maintain the equality in (A.24), $x_1^*(\lambda)$ needs to be decreased when λ increases and vice versa.

Since $U(x_1^*(\lambda),\lambda) = \pi\lambda x_1^*(\lambda)$, checking the monotonicity of $U(x_1^*(\lambda),\lambda)$ is equivalent to checking that of $\lambda x_1^*(\lambda)$. It is observed that if λ increases, decreasing $x_1^*(\lambda)$ with $\lambda x_1^*(\lambda)$ being a constant will decrease the LHS of (A.24) due to the term $x_1^*(\lambda)^{\frac{\alpha}{2}}$. Therefore, $\lambda x_1^*(\lambda)$ needs to be an increasing function of λ and so does $U(x_1^*(\lambda),\lambda)$.

To prove the monotonicity of $\bar{P}_{\text{BS}}(x_1^*(\lambda),\lambda)$, we expand (2.25) as

$$\begin{aligned} &\frac{\alpha}{2} D_1 x_1^*(\lambda)^{\frac{\alpha-2}{2}} \left(2^{D_2\pi\lambda x_1^*(\lambda)} - 1\right) \\ &+ (\ln 2) D_1 D_2 \pi\lambda x_1^*(\lambda)^{\frac{\alpha}{2}} 2^{D_2\pi\lambda x_1^*(\lambda)} = \mu\pi\lambda \end{aligned} \tag{A.25}$$

which can be rearranged as

$$\begin{aligned} & D_1 x_1^*(\lambda)^{\frac{\alpha}{2}} \left(2^{D_2 \pi \lambda x_1^*(\lambda)} - 1\right) \frac{\alpha}{2\lambda x_1^*(\lambda)} \\ + & D_1 x_1^*(\lambda)^{\frac{\alpha}{2}} \left(2^{D_2 \pi \lambda x_1^*(\lambda)} - 1\right) (\ln 2) \pi D_2 \\ + & (\ln 2) D_1 D_2 \pi x_1^*(\lambda)^{\frac{\alpha}{2}} = \mu\pi \end{aligned} \tag{A.26}$$

or equivalently,

$$\begin{aligned} & \left(\bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda) - P_c\right) \frac{\alpha}{2\lambda x_1^*(\lambda)} \\ + & \left(\bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda) - P_c\right) (\ln 2) \pi D_2 \\ + & (\ln 2) D_1 D_2 \pi x_1^*(\lambda)^{\frac{\alpha}{2}} = \mu\pi. \end{aligned} \tag{A.27}$$

Suppose that $x_1^*(\lambda_1)$ and $x_1^*(\lambda_2)$ are the two roots of (2.25) when $\lambda = \lambda_1$ and $\lambda = \lambda_2$, respectively, where $\lambda_2 > \lambda_1$. Based on the monotonicity of $x_1^*(\lambda)$ and $U(x_1^*(\lambda), \lambda)$ proved above, we have

$$x_1^*(\lambda_1)\lambda_1 < x_1^*(\lambda_2)\lambda_2, \tag{A.28}$$

$$x_1^*(\lambda_1)^{\frac{\alpha}{2}} > x_1^*(\lambda_2)^{\frac{\alpha}{2}}. \tag{A.29}$$

Due to the equality in (A.27) for all $\lambda > 0$, we have

$$\bar{P}_{\text{BS}}(x_1^*(\lambda_1), \lambda_1) < \bar{P}_{\text{BS}}(x_1^*(\lambda_2), \lambda_2), \forall \lambda_2 > \lambda_1. \tag{A.30}$$

Lemma 2.4.2 is thus proved.

A.4 Proof of Theorem 2.4.1

First, we consider the case of $\lambda_2 \geq \lambda_1$, in which three subcases are addressed as follows:

1. If $\lambda \leq \lambda_1$, according to the definition of λ_1 given in Lemma 2.4.1, $L_\lambda(x_1^*(\lambda), \mu) \geq 0$ for $\lambda \leq \lambda_1$, which corresponds to the third condition in (2.27). Therefore, we have

$$x^*(\lambda) = 0.$$

2. If $\lambda_1 < \lambda \leq \lambda_2$, we have $L_\lambda(x_1^*(\lambda), \mu) < 0$. Since $\bar{P}_{\text{BS}}(x_1^*(\lambda_2), \lambda_2) = P_{\max}$ and $\bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda_2)$ increases with λ from Lemma 2.4.2, it can be easily verified that $\bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda_1) < P_{\max}$ for the assumed range of λ, which is in accordance with the first condition in (2.27). Therefore, we have

$$x^*(\lambda) = x_1^*(\lambda).$$

3. Otherwise, if $\lambda > \lambda_2 \geq \lambda_1$, similar to the previous subcase, we know that $\bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda_1) > P_{\max}$. Next, we need to check the sign of $L_\lambda(x_2^*(\lambda), \mu) = P_{\max} - \mu\pi\lambda x_2^*(\lambda)$. Note that $L_\lambda(x_2^*(\lambda_2), \mu) = L_\lambda(x_1^*(\lambda_2), \mu)$, which is non-positive due to $\lambda_2 \geq \lambda_1$. Since $U(x_2^*(\lambda), \lambda)$ strictly increases with λ, $L_\lambda(x_2^*(\lambda), \mu)$ is thus a strictly decreasing function of λ. Therefore $L_\lambda(x_2^*(\lambda), \mu) < 0$ for $\lambda > \lambda_2$, which implies

$$x^*(\lambda) = x_2^*(\lambda).$$

Second, consider the case of $\lambda_2 < \lambda_1$. It is first verified that $\lambda_3 > \lambda_1 > \lambda_2$ in this case as follows: since $x_1^*(\lambda_1)$ minimizes $L_\lambda(x(\lambda), \mu)$ when $\lambda = \lambda_1$ to attain a zero value, and $L_\lambda(x(\lambda), \mu)$ is strictly convex in $x(\lambda)$, it follows that $L_\lambda(x_2^*(\lambda_1), \mu) > 0$. Since $L_\lambda(x_2^*(\lambda_3), \mu) = 0$ and $L_\lambda(x_2^*(\lambda), \mu)$ is a strictly decreasing function of λ, we conclude that $\lambda_3 > \lambda_1$. Next, we consider the following three subcases:

1. If $\lambda \leq \lambda_1$, according to Lemma 2.4.1, it is easy to verify that $L_\lambda(x_2^*(\lambda), \mu) > L_\lambda(x_1^*(\lambda), \mu) \geq 0$. Therefore, we have

$$x^*(\lambda) = 0.$$

2. If $\lambda_1 < \lambda \leq \lambda_3$, we have $\bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda) > P_{\max}$ and $L_\lambda(x_2^*(\lambda), \mu) \geq 0$, which implies

$$x^*(\lambda) = 0.$$

3. Otherwise, if $\lambda > \lambda_3$, we have $\bar{P}_{\text{BS}}(x_1^*(\lambda), \lambda) > P_{\max}$ and $L_\lambda(x_2^*(\lambda), \mu) < 0$, which is in accordance with the second condition in (2.27). Therefore, we have

$$x^*(\lambda) = x_2^*(\lambda).$$

Combining the above two cases, Theorem 2.4.1 is thus proved.

A.5 Proof of Lemma 2.4.4

From (2.16) and (2.25), we obtain the following equation

$$D_1 x_2^*(\lambda)^{\frac{\alpha}{2}} \left(2^{D_2\pi\lambda x_2^*(\lambda)} - 1\right) = P_{\max} - P_c. \tag{A.31}$$

With the high spectrum-efficiency assumption of $D_2\pi\lambda x_2^*(\lambda) \gg 1$, (A.31) is simplified as

$$D_1 x_2^*(\lambda)^{\frac{\alpha}{2}} 2^{D_2\pi\lambda x_2^*(\lambda)} = P_{\max} - P_c \tag{A.32}$$

which can be rearranged as

$$2^{-\frac{2D_2\pi\lambda}{\alpha} x_2^*(\lambda)} = \left(\frac{D_1}{P_{\max} - P_c}\right)^{\frac{2}{\alpha}} x_2^*(\lambda). \tag{A.33}$$

By utilizing

$$p^{ax+b} = cx + d \Rightarrow x = -\frac{\mathcal{W}\left(-\frac{a\ln p}{c} p^{b-\frac{ad}{c}}\right)}{a\ln p} - \frac{d}{c} \tag{A.34}$$

with $p > 0$, $a, c \neq 0$, it is easy to verify that $a = -\frac{2D_2\pi\lambda}{\alpha}$, $b = 0$, $c = \left(\frac{D_1}{P_{\max}-P_c}\right)^{\frac{2}{\alpha}}$, $d = 0$ and $p = 2$ in (A.33). Thus, $x_2^*(\lambda)$ is given by

$$x_2^*(\lambda) = \frac{\alpha}{2D_3\pi\lambda} \mathcal{W}\left(\frac{2D_3\pi\lambda}{\alpha}\left(\frac{P_{\max} - P_c}{D_1}\right)^{\frac{2}{\alpha}}\right). \tag{A.35}$$

We then proceed to derive the expression of $x_1^*(\lambda)$. Note that $x_1^*(\lambda)$ is the root of equation (A.25), which can be expressed as

$$x_1^*(\lambda)^{\frac{\alpha-2}{2}} 2^{D_2\pi\lambda x_1^*(\lambda)} \left[\frac{\alpha}{2} + (\ln 2) D_2\pi\lambda x_1^*(\lambda)\right] = \frac{\mu\pi\lambda}{D_1} \tag{A.36}$$

by applying the high spectrum-efficiency assumption of $D_2\pi\lambda x_1^*(\lambda) \gg 1$. Furthermore, it is observed that (A.36) can be simplified as

$$(\ln 2) D_1 D_2 x_1^*(\lambda)^{\frac{\alpha}{2}} 2^{D_2\pi\lambda x_1^*(\lambda)} = \mu \tag{A.37}$$

due to the fact that $(\ln 2) D_2\pi\lambda x_1^*(\lambda) \gg \frac{\alpha}{2}$, where $\alpha = 2 \sim 6$ in practice. Similar to the case for obtaining $x_2^*(\lambda)$, $x_1^*(\lambda)$ can be solved from (A.37) and given by

$$x_1^*(\lambda) = \frac{\alpha}{2D_3\pi\lambda} \mathcal{W}\left(\frac{2D_3\pi\lambda}{\alpha}\left(\frac{\mu}{D_1 D_3}\right)^{\frac{2}{\alpha}}\right). \tag{A.38}$$

Lemma 2.4.4 is thus proved.

Appendix B

Appendices to Chapter 3

B.1 Proof of Proposition 3.4.1

We prove this proposition by contradiction. Suppose that $\left\{\{\rho_{k,n}^a\}, \{p_{k,n}^a\}, T^a\right\}$ (termed Solution A) is the optimal solution of problem (WSREMin), which does not satisfy the condition given in Proposition 3.4.1, i.e., there exists at least one MT k, such that its associated time sharing factors are not all identical. Next, we construct a new solution $\left\{\{\rho_{k,n}^b\}, \{p_{k,n}^b\}, T^b\right\}$ (termed Solution B) for problem (WSREMin), which satisfies the condition given in Proposition 3.4.1 and also achieves a weighted-sum receiver energy consumption no larger than that of Solution A. The details of constructing Solution B are given as follows:

$$T^b = \sum_{k=1}^{K} \max_n \{\rho_{k,n}^a T^a\} \tag{B.1}$$

$$\rho_{k,n}^b = \max_j \{\rho_{k,j}^a\} / \sum_{i=1}^{K} \max_j \{\rho_{i,j}^a\}, \forall k, n \tag{B.2}$$

$$r_{k,n}^b = \begin{cases} r_{k,n}^a \rho_{k,n}^a T^a / \rho_{k,n}^b T^b & \text{if } \rho_{k,n}^a > 0 \\ 0 & \text{otherwise.} \end{cases} \forall k, n. \tag{B.3}$$

Note that $\rho_{k,n}^b = \rho_k^b, \forall n \in \mathcal{N}, k \in \mathcal{K}$.

Next, we check that Solution B is also feasible for problem (WSREMin). Since

$$\sum_{k=1}^{K} \rho_{k,n}^b = \sum_{k=1}^{K} \max_j \{\rho_{k,j}^a\} / \sum_{i=1}^{K} \max_j \{\rho_{i,j}^a\} = 1, \forall n \tag{B.4}$$

$$\sum_{n=1}^{N} T^b \rho_{k,n}^b r_{k,n}^b = \sum_{n=1}^{N} T^a \rho_{k,n}^a r_{k,n}^a \geq \bar{Q}_k, \forall k \tag{B.5}$$

we verify that both the constraints in (3.10) and (3.11) are satisfied. Moreover, from (B.1) and (B.2), it is observed that

$$T^b \rho_{k,n}^b = T^a \max_j \{\rho_{k,j}^a\} \geq T^a \rho_{k,n}^a, \forall k, n \tag{B.6}$$

i.e., the time allocated to MT k on SC n in Solution B is no smaller than that in Solution A. From (B.3), we also have that

$$\rho_{k,n}^b T^b r_{k,n}^b = \rho_{k,n}^a T^a r_{k,n}^a, \forall k, n \tag{B.7}$$

i.e., the amount of data delivered to MT k on SC n is the same for both Solution A and Solution B. Since $r_{k,n}$ is a strictly concave and increasing function of $p_{k,n}$, it is easy to verify that the amount of energy consumed for delivering the same amount of data decreases as the transmission time increases. Therefore, we have

$$\sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n}^b p_{k,n}^b \leq \sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n}^a p_{k,n}^a \leq P_{\text{avg}} \tag{B.8}$$

i.e., the constraint in (3.12) is satisfied by Solution B.

Finally, we show that Solution B achieves a weighted-sum receiver-side energy that is no larger than that by Solution A as follows. According to (3.5), we infer that

$$t_k^a \geq \max_n \{T^a \rho_{k,n}^a\}, \forall k \tag{B.9}$$

where t_k^a is the on time of MT k in Solution A. Let t_k^b denote the on time of MT k in Solution B. Since $\rho_{k,n}^b$'s are identical for given MT k, we can find t_k^b's such that

$$t_k^b = T^b \rho_{k,n}^b = \max_n \{T^a \rho_{k,n}^a\} \leq t_k^a, \forall k \tag{B.10}$$

which indicates that Solution B achieves a weighted-sum receiver-side energy no larger than that by Solution A. Thus, Proposition 3.4.1 is proved.

B.2 Proof of Theorem 3.4.1

Denote $\{s_{k,n}^*\}$ and $\{t_k^*\}$ as the optimal solution of problem (P1). Let β and $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \cdots, \lambda_K]$ be the dual variables of problem (P1) associated with the average

transmit power constraint in (3.23) and the data requirements in (3.22), respectively. Then the Lagrangian of problem (P1) can be expressed as

$$\begin{aligned}
&\mathcal{L}^{\text{P1}}(\{s_{k,n}\},\{t_k\},\boldsymbol{\lambda},\beta)\\
&=\sum_{k=1}^{K}\alpha_k P_{r,c}t_k-\sum_{k=1}^{K}\lambda_k\left(\sum_{n=1}^{N}s_{k,n}-\bar{Q}_k\right)\\
&+\beta\left(\sum_{k=1}^{K}\sum_{n=1}^{N}t_k\frac{e^{a\frac{s_{k,n}}{t_k}}-1}{f_{k,n}}-P_{\text{avg}}\sum_{k=1}^{K}t_k\right) \qquad \text{(B.11)}\\
&=\sum_{k=1}^{K}\left(\alpha_k P_{r,c}t_k+\beta\sum_{n=1}^{N}t_k\frac{e^{a\frac{s_{k,n}}{t_k}}-1}{f_{k,n}}-\beta P_{\text{avg}}t_k\right)\\
&-\sum_{k=1}^{K}\lambda_k\sum_{n=1}^{N}s_{k,n}+\sum_{k=1}^{K}\lambda_k\bar{Q}_k. \qquad \text{(B.12)}
\end{aligned}$$

The Lagrange dual function of $\mathcal{L}^{\text{P1}}(\cdot)$ in (B.12) is defined as

$$g^{\text{P1}}(\boldsymbol{\lambda},\beta)=\underset{\{s_{k,n}\geq 0\},\{t_k>0\}}{\texttt{Min.}}\mathcal{L}^{\text{P1}}(\{s_{k,n}\},\{t_k\},\boldsymbol{\lambda},\beta). \qquad \text{(B.13)}$$

The dual problem of problem (P1) is expressed as

$$(\text{P1}-\text{D}):\ \underset{\boldsymbol{\lambda}\geq 0,\beta\geq 0}{\texttt{Max.}}\ g^{\text{P1}}(\boldsymbol{\lambda},\boldsymbol{\beta}). \qquad \text{(B.14)}$$

Since (P1) is convex and satisfies the Salter's condition [32], strong duality holds between problem (P1) and its dual problem (P1-D). Let $\boldsymbol{\lambda}^*\geq 0$ and $\beta^*\geq 0$ denote the optimal dual solutions to problem (P1); then we have the following lemma.

Lemma B.2.1. *The optimal solution to problem (P1-D) satisfies that*

$$\lambda_k^*>0,\forall k \qquad \text{(B.15)}$$

$$\beta^*>0 \qquad \text{(B.16)}$$

$$\beta^*-\min_k(\alpha_k)P_{r,c}/P_{avg}<0 \qquad \text{(B.17)}$$

$$\alpha_k P_{r,c}-\beta^*P_{avg}+\sum_{n=1}^{N}u_n(\beta^*,\lambda_k^*)=0,\forall k \qquad \text{(B.18)}$$

where $u_n(\beta,\lambda_k)=\left(\frac{\lambda_k}{a}-\frac{\beta}{f_{k,n}}\right)^+-\frac{\lambda_k}{a}\left(\ln\frac{\lambda_k f_{k,n}}{a\beta}\right)^+$ *and* $(\cdot)^+\triangleq\max\{\cdot,0\}$.

Proof. From (B.12), it follows that the minimization of $\mathcal{L}^{\text{P1}}(\{s_{k,n}\},\{t_k\},\boldsymbol{\lambda},\beta)$ can be decomposed into K independent optimization problems, each for one MT and given by

$$\underset{t_k>0,\{s_{k,n}\geq 0\}}{\texttt{Min.}} \ \mathcal{L}_k^{\text{P1}}(\{s_{k,n}\},t_k,\lambda,\beta),\ k=1,\cdots,K \tag{B.19}$$

where $\mathcal{L}_k^{\text{P1}}(\{s_{k,n}\},t_k,\lambda,\beta) \triangleq \alpha_k t_k - \lambda_k \sum_{n=1}^{N} s_{k,n} + \beta \sum_{n=1}^{N} t_k \frac{e^{a\frac{s_{k,n}}{t_k}}-1}{f_{k,n}} - \beta P_{\text{avg}} t_k$. Note that $\mathcal{L}^{\text{P1}}(\{s_{k,n}\},\{t_k\},\boldsymbol{\lambda},\beta) = \sum_{k=1}^{K} \mathcal{L}_k^{\text{P1}}(\cdot) + \sum_{k=1}^{K} \lambda_k \bar{Q}_k$. By taking the derivative of $\mathcal{L}_k^{\text{P1}}(\cdot)$ with respect to $s_{k,n}$, we have

$$\frac{\partial \mathcal{L}_k^{\text{P1}}}{\partial s_{k,n}} = \frac{a\beta}{f_{k,n}} e^{a\frac{s_{k,n}}{t_k}} - \lambda_k. \tag{B.20}$$

Let $\{s_{k,n}^{\star}(\lambda_k,\beta)\}$ and $t_k^{\star}(\lambda_k,\beta)$ denote the optimal solution of problem (B.19) given λ_k and β.

Next, we show that $\beta^* > 0$ and $\lambda_k^* > 0, \forall k$ by contradiction. If $\beta^* = 0$ and $\lambda_k^* = 0, \forall k$, from (B.19), it follows that $g^{\text{P1}}(\boldsymbol{\lambda}^*,\beta^*) = 0$, which is approached as $t_k \to 0, \forall k$, and the optimal value of problem (P1-D) is thus 0, which contradicts with the fact that strong duality holds between problems (P1) and (P1-D). If $\beta^* = 0$ and $\exists i \in \mathcal{K}$ such that $\lambda_i^* > 0$, it follows that $\frac{\partial \mathcal{L}_k^{\text{P1}}}{\partial s_{i,n}} < 0, \forall n$ at the optimal dual solution, which implies that $s_{i,n}^* = \infty, \forall n$. Since $s_{i,n} = t_i r_{i,n}$, which is the amount of data delivered to MT i on SC n over the transmission, $s_{i,n}^* = \infty, \forall n$ indicates that $Q_i = \infty$, which is evidently suboptimal for problem (P1). If $\exists j \in \mathcal{K}$ such that $\lambda_j^* = 0$ and $\beta^* > 0$, it follows that $\frac{\partial \mathcal{L}_j^{\text{P1}}}{\partial s_{j,n}} > 0, \forall n$ at the optimal dual solution, which implies that $s_{j,n}^* = 0, \forall n$ or $Q_j = 0$. Then it contradicts with the fact that $\bar{Q}_j > 0$. Combining all the three cases above, it concludes that $\beta^* > 0$ and $\lambda_k^* > 0, \forall k$.

With $\beta > 0$ and $\lambda_k > 0, \forall k$ as proved above and from (B.20), the ratio $\frac{s_{k,n}^{\star}(\lambda_k,\beta)}{t_k^{\star}(\lambda_k,\beta)}$ thus needs to satisfy

$$\frac{s_{k,n}^{\star}(\lambda_k,\beta)}{t_k^{\star}(\lambda_k,\beta)} = \frac{1}{a}\left(\ln\frac{\lambda_k f_{k,n}}{a\beta}\right)^{+}, \forall n. \tag{B.21}$$

Substituting (B.21) back to $\mathcal{L}_k^{\text{P1}}(\cdot)$ yields

$$\mathcal{L}_k^{\text{P1}}(\{s_{k,n}\},t_k,\lambda,\beta) = \left(\alpha_k P_{r,c} - \beta P_{\text{avg}} + \sum_{n=1}^{N} u_n(\beta,\lambda_k)\right) t_k \tag{B.22}$$

which is a linear function of t_k, and thus t_k^* would be finite only if $\alpha_k P_{r,c} - \beta^* P_{\text{avg}} + \sum_{n=1}^{N} u_n(\beta^*, \lambda_k^*) = 0$. Condition (B.18) is thus verified.

Finally, we show that $\beta^* < \alpha_k P_{r,c}/P_{\text{avg}}$. Since it can be shown that given β, $\sum_{n=1}^{N} u_n(\beta, \lambda_k)$ equals zero when $\lambda_k \leq \frac{a\beta}{\max_n\{f_{k,n}\}}$ and is a strictly decreasing function of λ_k when $\lambda_k > \frac{a\beta}{\max_n\{f_{k,n}\}}$, we have $\beta^* \leq \alpha_k P_{r,c}/P_{\text{avg}}$ from (B.15). If $\beta^* = \alpha_k P_{r,c}/P_{\text{avg}}$, it follows that $\lambda_k^* \leq \frac{a\beta^*}{\max_n\{f_{k,n}\}}$, which implies that $s_{k,n}^* = 0, \forall n$ from (B.21). This again contradicts with the fact that $\bar{Q}_k > 0$. Lemma B.2.1 is thus proved. □

Next, we proceed to show the structural property of the optimal solution to problem (WSREMin-TDMA). Let the optimal solution of this problem be given by $\{p_{k,n}^*\}$ and $\{t_k^*\}$ with $s_{k,n}^* = r_{k,n}^* t_k^*, \forall n, k$, as in problem (P1). From the change of variables and (3.1), it follows that

$$\frac{s_{k,n}^*}{t_k^*} = W \log_2 \left(1 + f_{k,n} p_{k,n}^*\right), \forall n, k. \tag{B.23}$$

Furthermore, from (B.21) we have

$$\frac{s_{k,n}^*}{t_k^*} = \frac{1}{a}\left(\ln \frac{\lambda_k^* f_{k,n}}{a\beta}\right)^+, \forall n, k. \tag{B.24}$$

Combining (B.23) and (B.24), (3.24) can be easily verified.

From Lemma B.2.1 and the complementary slackness conditions [32] satisfied by the optimal solution of problem (P1), it follows that

$$\sum_{n=1}^{N} s_{k,n}^* = \bar{Q}_k, \forall k \tag{B.25}$$

$$\sum_{k=1}^{K}\sum_{n=1}^{N} t_k^* \frac{e^{a\frac{s_{k,n}^*}{t_k^*}} - 1}{f_{k,n}} = P_{\text{avg}} \sum_{k=1}^{K} t_k^*. \tag{B.26}$$

In other words, the optimal solutions of problem (P1) or problem (WSREMin-TDMA) are always attained with all the data constraints in (3.22) or (3.19) and average power constraint in (3.23) or (3.20) being met with equality. Substituting (B.24) into (B.25), (3.25) then easily follows. Theorem 3.4.1 is thus proved.

B.3 Solution to Problem (P2)

The Lagrangian of problem (P2) can be expressed as

$$\begin{aligned}
&\mathcal{L}^{\text{P2}}(\{m_{k,n}\},\{\rho_{k,n}\},\boldsymbol{\lambda},\boldsymbol{\beta}) \\
&= \sum_{k=1}^{K}\sum_{n=1}^{N}\rho_{k,n}\frac{e^{a\frac{m_{k,n}}{\rho_{k,n}}}-1}{f_{k,n}} - \sum_{k=1}^{K}\lambda_k\left(\sum_{n=1}^{N}m_{k,n}-c_k\right) \\
&+ \sum_{n=1}^{N}\beta_n\left(\sum_{k=1}^{K}\rho_{k,n}-1\right) \qquad \text{(B.27)} \\
&= \sum_{k=1}^{K}\sum_{n=1}^{N}\left(\rho_{k,n}\frac{e^{a\frac{m_{k,n}}{\rho_{k,n}}}-1}{f_{k,n}} - \lambda_k m_{k,n} + \beta_n\rho_{k,n}\right) \\
&+ \sum_{k=1}^{K}\lambda_k c_k - \sum_{n=1}^{N}\beta_n \qquad \text{(B.28)}
\end{aligned}$$

where $\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots,\lambda_K]$ and $\boldsymbol{\beta}=[\beta_1,\beta_2,\cdots,\beta_N]$ are the vectors of dual variables associated with the constraints in (3.38) and (3.37), respectively.

Then, the corresponding dual function is defined as

$$g^{\text{P2}}(\boldsymbol{\lambda},\boldsymbol{\beta}) = \underset{\{m_{k,n}\geq 0\},\{0\leq\rho_{k,n}\leq 1\}}{\texttt{Min.}} \mathcal{L}^{\text{P2}}(\{m_{k,n}\},\{\rho_{k,n}\},\boldsymbol{\lambda},\boldsymbol{\beta}). \qquad \text{(B.29)}$$

The dual problem of problem (P2) is thus expressed as

$$(\text{P2}-\text{D}): \quad \underset{\boldsymbol{\lambda}\geq 0,\boldsymbol{\beta}\geq 0}{\texttt{Max.}} \; g^{\text{P2}}(\boldsymbol{\lambda},\boldsymbol{\beta}). \qquad \text{(B.30)}$$

Since (P2) is convex and satisfies the Salter's condition [**?**], strong duality holds between problem (P2) and its dual problem (P2-D). To solve (P2-D), in the following we first solve problem (B.29) to obtain $g(\boldsymbol{\lambda},\boldsymbol{\beta})$ with given $\boldsymbol{\lambda}\geq 0$ and $\boldsymbol{\beta}\geq 0$.

The expression of (B.28) suggests that the minimization of $\mathcal{L}^{\text{P2}}(\{m_{k,n}\},\{\rho_{k,n}\},\boldsymbol{\lambda},\boldsymbol{\beta})$ can be decomposed into NK parallel subproblems, each of which is for one given pair of n and k and expressed as

$$\underset{m_{k,n}\geq 0,0\leq\rho_{k,n}\leq 1}{\texttt{Min.}} \; \mathcal{L}^{\text{P2}}_{k,n}(m_{k,n},\rho_{k,n},\lambda_k,\beta_n) \qquad \text{(B.31)}$$

where $\mathcal{L}^{\text{P2}}_{k,n}(m_{k,n},\rho_{k,n},\lambda_k,\beta_n) \triangleq \rho_{k,n}\frac{e^{a\frac{m_{k,n}}{\rho_{k,n}}}-1}{f_{k,n}} - \lambda_k m_{k,n} + \beta_n\rho_{k,n}$. Note that $\mathcal{L}^{\text{P2}}(\cdot) = \sum_{k=1}^{K}\sum_{n=1}^{N}\mathcal{L}^{\text{P2}}_{k,n}(\cdot) + \sum_{k=1}^{K}\lambda_k c_k - \sum_{n=1}^{N}\beta_n$.

Lemma B.3.1. *The optimal solution of problem (P2-D) satisfies that* $\boldsymbol{\lambda}^* > 0$ *and* $\boldsymbol{\beta}^* > 0$.

Proof. The proof is similar to that of Lemma B.2.1, and thus is omitted for brevity. □

With Lemma B.3.1, in the following, we only consider the case that $\boldsymbol{\lambda} > 0$ and $\boldsymbol{\beta} > 0$.

Lemma B.3.2. *For a given pair of* n *and* k *with* $\lambda_k > 0$ *and* $\beta_n > 0$*, the optimal solution of problem (B.31) is given by*

$$m^{\star}_{k,n}(\lambda_k, \beta_n) = \frac{\rho^{\star}_{k,n}(\lambda_k, \beta_n)}{a} \left(\ln \frac{\lambda_k f_{k,n}}{a} \right)^+ \tag{B.32}$$

$$\rho^{\star}_{k,n}(\lambda_k, \beta_n) = \begin{cases} 1 & o(\lambda_k, \beta_n) < 0 \\ 0 & \textit{otherwise} \end{cases} \tag{B.33}$$

where $o(\lambda_k, \beta_n) = \left(\frac{\lambda_k}{a} - \frac{1}{f_{k,n}} \right)^+ - \frac{\lambda_k}{a} \left(\ln \frac{\lambda_k f_{k,n}}{a} \right)^+ + \beta_n$.

Proof. First, consider the case of $\rho_{k,n} = 0$, in which $m_{k,n} = 0$ and $\frac{m_{k,n}}{\rho_{k,n}} = 0$. It follows that $\mathcal{L}^{\text{P2}}_{k,n}(\cdot) = 0$.

Second, consider the case of $\rho_{k,n} > 0$. Taking the derivative of $\mathcal{L}^{\text{P2}}_{k,n}(\cdot)$ over $m_{k,n}$ and $\rho_{k,n}$, respectively, we have

$$\frac{\partial \mathcal{L}^{\text{P2}}_{k,n}}{\partial m_{k,n}} = \frac{a}{f_{k,n}} e^{a \frac{m_{k,n}}{\rho_{k,n}}} - \lambda_k \tag{B.34}$$

$$\frac{\partial \mathcal{L}^{\text{P2}}_{k,n}}{\partial \rho_{k,n}} = \frac{1}{f_{k,n}} e^{a \frac{m_{k,n}}{\rho_{k,n}}} \left(1 - a \frac{m_{k,n}}{\rho_{k,n}} \right) - \frac{1}{f_{k,n}} + \beta_n. \tag{B.35}$$

Then it is easy to see that given $\lambda_k > 0$ and $\beta_n > 0$, from (B.34), the optimal solution of problem (B.31) needs to satisfy the following equation:

$$m^{\star}_{k,n}(\lambda_k, \beta_n) = \frac{\rho^{\star}_{k,n}(\lambda_k, \beta_n)}{a} \left(\ln \frac{\lambda_k f_{k,n}}{a} \right)^+. \tag{B.36}$$

Substituting (B.36) into (B.35), it then follows that $\frac{\partial \mathcal{L}^{\text{P2}}_{k,n}}{\partial \rho_{k,n}} = o(\lambda_k, \beta_n)$, which is a constant implying

$$\rho^{\star}_{k,n}(\lambda_k, \beta_n) = \begin{cases} 1 & \text{if } o(\lambda_k, \beta_n) < 0 \\ (0, 1] & \text{if } o(\lambda_k, \beta_n) = 0 \\ \to 0 & \text{otherwise} \end{cases} \tag{B.37}$$

where $\to 0$ means here that the optimal value cannot be attained but can be approached as $\rho^{\star}_{k,n}(\lambda_k, \beta_n) \to 0$. Then, substituting (B.36) into $\mathcal{L}^{\text{P2}}_{k,n}(\cdot)$, it follows that $\mathcal{L}^{\text{P2}}_{k,n}(\cdot) = \rho^{\star}_{k,n}(\lambda_k, \beta_n) o(\lambda_k, \beta_n)$. Thus, (B.37) achieves the optimal value of $\mathcal{L}^{\text{P2}}_{k,n}(\cdot)$ as

$$\mathcal{L}^{\text{P2}}_{k,n}(\cdot) = \begin{cases} o(\lambda_k, \beta_n) & \text{if } o(\lambda_k, \beta_n) < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{B.38}$$

Combining the two cases above, Lemma B.3.2 is thus proved. □

With Lemma B.3.2, we can solve the NK subproblems in (B.31) and thus obtain $g(\boldsymbol{\lambda}, \boldsymbol{\beta})$ with given $\boldsymbol{\lambda} > 0$ and $\boldsymbol{\beta} > 0$. Then, we solve problem (P2-D) by finding the optimal $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ to maximize $g(\boldsymbol{\lambda}, \boldsymbol{\beta})$. Although problem (P2-D) is convex, the dual function $g(\boldsymbol{\lambda}, \boldsymbol{\beta})$ is not differentiable and as a result analytical expressions for its differentials do not exist. Hence, conventional methods with gradient based search, such as Newton method, cannot be applied for solving problem (P2-D). An alternative method is thus the ellipsoid method [47], which is capable of minimizing non-differentiable convex functions based on the so-called subgradient.[1] Hence, the optimal solution of (P2-D) can be obtained as $\boldsymbol{\lambda}^*$ and $\boldsymbol{\beta}^*$ by applying the ellipsoid method.

After obtaining the dual solution $\boldsymbol{\lambda}^*$ and $\boldsymbol{\beta}^*$, we can substitute them into (B.32) and (B.33), and obtain the corresponding $\{m^{\star}_{k,n}\}$ and $\{\rho^{\star}_{k,n}\}$. However, notice that the obtained $\{m^{\star}_{k,n}\}$ and $\{\rho^{\star}_{k,n}\}$ may not necessarily be the optimal solution of problem (P2), denoted by $\{m^{*}_{k,n}\}$ and $\{\rho^{*}_{k,n}\}$, since they may not satisfy the constraints in (3.37) and (3.38). The reason is that when $o(\lambda^*_k, \beta^*_n) = 0$ for certain pairs of n and k, the corresponding $\rho^{\star}_{k,n}$ can actually take any value within $[0, 1]$ according to (B.37), each of which would result in a different $m^{\star}_{k,n}$ accordingly. Therefore, with $\boldsymbol{\lambda}^*$ and $\boldsymbol{\beta}^*$, we may obtain infinite sets of $\{m^{\star}_{k,n}\}$ and $\{\rho^{\star}_{k,n}\}$, some of which might not satisfy the constraints in (3.37) and/or (3.38) [50]. In such cases, a linear programming (LP) needs to be further solved to obtain a feasible optimal solution for problem (P2).

[1]The subgradient of $g(\boldsymbol{\lambda}, \boldsymbol{\beta})$ at given $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ for the ellipsoid method can be shown to be $\sum_n m^{\star}_{k,n}(\lambda_k, \beta_n) - c_k$ for λ_k, $k = 1, \cdots, K$ and $1 - \sum_k \rho^{\star}_{k,n}(\lambda_k, \beta_n)$ for β_n, $n = 1, \cdots, N$.

To be more specific, we first define the following two sets with given $\boldsymbol{\lambda}^*$ and $\boldsymbol{\beta}^*$:

$$\mathcal{A}_1 = \{(k,n) \mid o(\lambda_k^*, \beta_n^*) \neq 0, \forall k, n\} \tag{B.39}$$

$$\mathcal{A}_2 = \{(k,n) \mid o(\lambda_k^*, \beta_n^*) = 0, \forall k, n\}. \tag{B.40}$$

From (B.32) and (B.33), we know that for any pair of n and k with $(k,n) \in \mathcal{A}_1$, the corresponding $m_{k,n}^\star$ and $\rho_{k,n}^\star$ can be uniquely determined, which implies

$$m_{k,n}^* = m_{k,n}^\star, \rho_{k,n}^* = \rho_{k,n}^\star, \forall (k,n) \in \mathcal{A}_1. \tag{B.41}$$

The problem remains to find $m_{k,n}^*$ and $\rho_{k,n}^*$ with $(k,n) \in \mathcal{A}_2$. It is then observed that the optimal solution of problem (P2) needs to satisfy the following linear equations:

$$m_{k,n}^* = \frac{\rho_{k,n}^*}{a}\left(\ln \frac{\lambda_k^* f_{k,n}}{a}\right)^+, \forall k, n \tag{B.42}$$

$$\sum_k \rho_{k,n}^* = 1, \forall n, \quad \sum_n m_{k,n}^* = c_k, \forall k \tag{B.43}$$

where (B.42) is due to (B.32), and (B.43) is due to Lemma B.3.1 and the complementary slackness conditions [32] satisfied by the optimal solution of problem (P2). Therefore, $m_{k,n}^*$ and $\rho_{k,n}^*$ with $(k,n) \in \mathcal{A}_2$ can be found through solving the above linear equations by treating $m_{k,n}^*$ and $\rho_{k,n}^*$ with $(k,n) \in \mathcal{A}_1$ as given constants, which is a linear programming (LP) and can be efficiently solved. In summary, one algorithm for solving problem (P2) and its dual problem (P2-D) is given in Table B.1 as follows.

For the algorithm given in Table B.1, the computation time is dominated by the ellipsoid method in steps 1)-3) and the LP in step 4). In particular, the time complexity of steps 1)-3) is of order $(K+N)^4$ [47] , step 4) is of order K^3N^3 [32]. Therefore, the time complexity of the algorithm in Table B.1 is $\mathcal{O}(K^4 + N^4 + K^3N^3)$.

B.4 Proof of Lemma 3.5.1

To show problem (TEMin-2) is convex, we need to prove that both $v(T)$ and $v(T)T + P_{t,c}T$ are convex functions of T. Since $P_{t,c}T$ is linear in T, we only need to show the convexity of $v(T)$ and $v(T)T$.

Table B.1: Algorithm for Solving Problem (P2) and (P2-D)

1. Initialize $\boldsymbol{\lambda} > \mathbf{0}$ and $\boldsymbol{\beta} > 0$.
2. **Repeat:**
   a) Obtain $\{m_{k,n}^*(\lambda_k, \beta_n)\}$ and $\{\rho_{k,n}^*(\lambda_k, \beta_n)\}$ using (B.32) and (B.33), respectively, with given $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$.
   b) Compute the subgradient of $g(\boldsymbol{\lambda}, \boldsymbol{\beta})$ and update $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ accordingly using the ellipsoid method [47].
3. **Until** both $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ converge to $\boldsymbol{\lambda}^*$ and $\boldsymbol{\beta}^*$, respectively, within a prescribed accuracy.
4. Determine $\{\{m_{k,n}^\star\}, \{\rho_{k,n}^\star\}\}$ with $\boldsymbol{\lambda}^*$ and $\boldsymbol{\beta}^*$. If it is feasible for problem (P2), set $\{\{m_{k,n}^*\}, \{\rho_{k,n}^*\}\} = \{\{m_{k,n}^\star\}, \{\rho_{k,n}^\star\}\}$; otherwise solve a LP to find $\{\{m_{k,n}^*\}, \{\rho_{k,n}^*\}\}$.

First, we check the convexity of function $v(T)$, which is sufficient to prove that for any convex combination $T = \theta T_1 + (1-\theta)T_2$ with $T_1, T_2 > 0$ and $\theta \in (0,1)$, we have $v(T) \leq \theta v(T_1) + (1-\theta)v(T_2)$. Denote the optimal solution to problem (TEMin-1) with T_1 and T_2 as $\{\dot{p}_{k,n}^*\}$, $\{\dot{\rho}_{k,n}^*\}$ (termed Solution 1) and $\{\ddot{p}_{k,n}^*\}$, $\{\ddot{\rho}_{k,n}^*\}$ (termed Solution 2), respectively. Then we have

$$\begin{aligned}\theta v(T_1) + (1-\theta)v(T_2) = \theta \sum_{k=1}^{K}\sum_{n=1}^{N} \dot{\rho}_{k,n}^* \dot{p}_{k,n}^* \\ + (1-\theta)\sum_{k=1}^{K}\sum_{n=1}^{N} \ddot{\rho}_{k,n}^* \ddot{p}_{k,n}^*. \end{aligned} \tag{B.44}$$

Next we construct another solution $\{\bar{p}_{k,n}^*\}$, $\{\bar{\rho}_{k,n}^*\}$ (termed Solution 3) of problem (TEMin-1) with given T, which is achieved by properly allocating power for each MT on each SC such that the average power consumption is the same as that with time sharing between Solution 1 and Solution 2. The details of constructing Solution 3 are

given as follows:

$$\bar{\rho}^*_{k,n} = \frac{\dot{\rho}^*_{k,n}\theta T_1 + \ddot{\rho}^*_{k,n}(1-\theta)T_2}{\theta T_1 + (1-\theta)T_2} \tag{B.45}$$

$$\bar{p}^*_{k,n} = \frac{\dot{p}^*_{k,n}\dot{\rho}^*_{k,n}\theta T_1 + \ddot{p}^*_{k,n}\ddot{\rho}^*_{k,n}(1-\theta)T_2}{\bar{\rho}^*_{k,n}[\theta T_1 + (1-\theta)T_2]}. \tag{B.46}$$

It can then be shown that

$$\begin{aligned}
\sum_{k=1}^{K} \bar{\rho}^*_{k,n} &= \frac{\theta T_1 \sum_{k=1}^{K} \dot{\rho}^*_{k,n} + (1-\theta)T_2 \sum_{k=1}^{K} \ddot{\rho}^*_{k,n}}{\theta T_1 + (1-\theta)T_2} \\
&\leq \frac{\theta T_1 + (1-\theta)T_2}{\theta T_1 + (1-\theta)T_2} = 1 \qquad \text{(B.47)} \\
\sum_{n=1}^{N} T\bar{\rho}^*_{k,n}\bar{r}^*_{k,n} &= \left(\theta T_1 \sum_{n=1}^{N} \dot{\rho}^*_{k,n} + (1-\theta)T_2 \sum_{n=1}^{N} \ddot{\rho}^*_{k,n}\right) \bar{r}^*_{k,n} \\
&\geq \theta T_1 \sum_{n=1}^{N} \dot{\rho}^*_{k,n}\dot{r}^*_{k,n} + (1-\theta)T_2 \sum_{n=1}^{N} \ddot{\rho}^*_{k,n}\ddot{r}^*_{k,n} \\
&\geq \theta\bar{Q}_k + (1-\theta)\bar{Q}_k = \bar{Q}_k \qquad \text{(B.48)} \\
\sum_{k=1}^{K}\sum_{n=1}^{N} \bar{\rho}^*_{k,n}\bar{p}^*_{k,n} &= \frac{\theta T_1 \sum_{k=1}^{K}\sum_{n=1}^{N} \dot{\rho}^*_{k,n}\dot{p}^*_{k,n}}{\theta T_1 + (1-\theta)T_2} \\
&\quad + \frac{(1-\theta)T_2 \sum_{k=1}^{K}\sum_{n=1}^{N} \ddot{\rho}^*_{k,n}\ddot{p}^*_{k,n}}{\theta T_1 + (1-\theta)T_2} \\
&\leq \sum_{k=1}^{K}\sum_{n=1}^{N} \dot{\rho}^*_{k,n}\dot{p}^*_{k,n} + \sum_{k=1}^{K}\sum_{n=1}^{N} \ddot{\rho}^*_{k,n}\ddot{p}^*_{k,n} \qquad \text{(B.49)}
\end{aligned}$$

i.e., Solution 3 is feasible for problem (TEMin-1) with the given T, and also achieves the same objective value as that in (B.44). Since Solution 3 is only a feasible solution for problem (TEMin-1) with given T, which is not necessary to be optimal, we have

$$v(T) \leq \theta v(T_1) + (1-\theta)v(T_2). \tag{B.50}$$

The convexity of $v(T)$ is thus proved.

Similar arguments can be applied to verify the convexity of $v(T)T$; Lemma 3.5.1 is thus proved.

B.5 Proof of Lemma 3.5.2

First, we find the gradient of $v(T)$. Since $v(T)$ is differentiable, its gradient and subgradient are equivalent. We provide the definition of subgradient [47] as follows. A vector $y \in \mathbf{R}^n$ is said to be the subgradient of function $q : \mathbf{R}^n \to \mathbf{R}$ at $x \in \mathbf{dom}\ q$ if for all $z \in \mathbf{dom}\ q$,

$$q(z) \geq q(x) + y^T(z - x). \tag{B.51}$$

The dual function (B.29) can be expressed as

$$\begin{aligned} & g(\boldsymbol{\lambda}, \boldsymbol{\beta}) \\ & = \inf_{\{m_{k,n}\},\{\rho_{k,n}\}} \sum_{k=1}^{K}\sum_{n=1}^{N} \left(\rho_{k,n} \frac{e^{a\frac{m_{k,n}}{\rho_{k,n}}} - 1}{f_{k,n}} - \lambda_k m_{k,n} + \beta_n \rho_{k,n} \right) \\ & + \frac{1}{T}\sum_{k=1}^{K} \lambda_k \bar{Q}_k - \sum_{n=1}^{N} \beta_n. \end{aligned} \tag{B.52}$$

Then, we have

$$v(T) = \underset{\boldsymbol{\lambda}\geq 0, \boldsymbol{\beta}\geq 0}{\texttt{Max.}}\ g(\boldsymbol{\lambda}, \boldsymbol{\beta}) \tag{B.53}$$

$$\begin{aligned} & = \inf_{\{m_{k,n}\},\{\rho_{k,n}\}} \sum_{k=1}^{K}\sum_{n=1}^{N} \left(\rho_{k,n} \frac{e^{a\frac{m_{k,n}}{\rho_{k,n}}} - 1}{f_{k,n}} - \lambda_k^*(T) m_{k,n} \right. \\ & \left. + \beta_n^*(T) \rho_{k,n} \right) + \frac{1}{T}\sum_{k=1}^{K} \lambda_k^*(T) \bar{Q}_k - \sum_{n=1}^{N} \beta_n^*(T) \end{aligned} \tag{B.54}$$

where $\{\lambda_k^*(T)\}$ and $\{\beta_n^*(T)\}$ is the optimal solution of problem (P2-D) with given $T > 0$. For any $T^{'} > 0$ and $T^{'} \neq T$, we have

$$\begin{aligned} v(T^{'}) = & \underset{\boldsymbol{\lambda}\geq 0, \boldsymbol{\beta}\geq 0}{\texttt{Max.}} \inf_{\{m_{k,n}\},\{\rho_{k,n}\}} \sum_{k=1}^{K}\sum_{n=1}^{N} \left(\rho_{k,n} \frac{e^{a\frac{m_{k,n}}{\rho_{k,n}}} - 1}{f_{k,n}} \right. \\ & \left. - \lambda_k m_{k,n} + \beta_n \rho_{k,n} \right) + \frac{1}{T^{'}}\sum_{k=1}^{K} \lambda_k \bar{Q}_k - \sum_{n=1}^{N} \beta_n \\ \geq & \inf_{\{m_{k,n}\},\{\rho_{k,n}\}} \sum_{k=1}^{K}\sum_{n=1}^{N} \left(\rho_{k,n} \frac{e^{a\frac{m_{k,n}}{\rho_{k,n}}} - 1}{f_{k,n}} - \lambda_k^*(T) m_{k,n} \right. \end{aligned} \tag{B.55}$$

$$+ \beta_n^*(T)\rho_{k,n}\Bigg) + \frac{1}{T'}\sum_{k=1}^{K}\lambda_k^*(T)\bar{Q}_k - \sum_{n=1}^{N}\beta_n^*(T) \tag{B.56}$$

$$= v(T) + \left(\frac{1}{T'} - \frac{1}{T}\right)\sum_{k=1}^{K}\lambda_k^*(T)\bar{Q}_k \tag{B.57}$$

$$= v(T) + \left(-\frac{1}{T^2}\sum_{k=1}^{K}\lambda_k^*(T)\bar{Q}_k\right)(T - \frac{T^2}{T'}) \tag{B.58}$$

$$\geq v(T) + \left(-\frac{1}{T^2}\sum_{k=1}^{K}\lambda_k^*(T)\bar{Q}_k\right)(T' - T) \tag{B.59}$$

where the last inequality is due to $\left(T - \frac{T^2}{T'}\right) - (T' - T) = (T' - T)\left(\frac{T}{T'} - 1\right) < 0$. Thus, the subgradient (gradient) of $v(T)$ is given by

$$v^{'}(T) = -\frac{1}{T^2}\sum_{k=1}^{K}\lambda_k^*(T)\bar{Q}_k. \tag{B.60}$$

With the gradient of $v(T)$, Lemma 3.5.2 can be easily verified.

Appendix C

Appendices to Chapter 5

C.1 Proof of Proposition 5.4.1

According to the Fundamental Theorem of Calculus [106], we can derive the first derivative of $\hat{P}_{out}(\rho, \eta, P_c)$ in (5.7) with respect to η, P_c and ρ as

$$\begin{aligned}\frac{\partial \hat{P}_{out}}{\partial \beta} &= \left(\frac{P_c}{\frac{\rho}{1-\rho}+\eta}\right)' f_x\left(\frac{P_c}{\frac{\rho}{1-\rho}+\eta}\right) \\ &\quad - \left(\frac{P_c}{\frac{\rho}{1-\rho}+\eta}\right)' f_x\left(\frac{P_c}{\frac{\rho}{1-\rho}+\eta}\right) F_\Gamma(\infty) \\ &\quad + \int_{\frac{P_c}{\frac{\rho}{1-\rho}+\eta}}^{P_H} f_X(x) f_\Gamma\left[g(\cdot)\right] \frac{\partial g(\cdot)}{\partial \beta} dx \\ &= \int_{\frac{P_c}{\frac{\rho}{1-\rho}+\eta}}^{P_H} f_X(x) f_\Gamma\left[g(\cdot)\right] \frac{\partial g(\cdot)}{\partial \beta} dx \end{aligned} \tag{C.1}$$

where β could be η, P_c or ρ and $g(\rho, \eta, P_c) = \frac{2^{\frac{Q}{(1-\rho)T}}-1}{x[\frac{\rho}{1-\rho}+\eta]-P_c}$.

It is easy to verify that $\frac{\partial g(\rho,\eta,P_c)}{\partial \eta} < 0, \forall \eta \in [0,1]$ and $\frac{\partial g(\rho,\eta,P_c)}{\partial P_c} > 0, \forall P_c \in [0,\infty]$. Therefore $\hat{P}_{out}(\rho, \eta, P_c)$ is strictly decreasing with battery efficiency η and strictly increasing with circuit power P_c. Next, we are going to prove the monotonicity of P_{out} and P^*_{out} with battery efficiency η, where circuit power P_c is treated as constant.

The condition $P_H > \frac{P_c}{\frac{\rho}{1-\rho}+\eta}$ in (5.6) could be expressed in terms of battery

efficiency: $\eta > \frac{P_c}{P_H} - \frac{\rho}{1-\rho}$, then

$$P_{out} = \begin{cases} 1, & \eta \leq \frac{P_c}{P_H} - \frac{\rho}{1-\rho} \\ \hat{P}_{out}, & \eta > \frac{P_c}{P_H} - \frac{\rho}{1-\rho} \end{cases}. \tag{C.2}$$

Consider the following two cases:

- Suppose $\eta_1 < \eta_2$ and $\frac{P_c}{P_H} - \frac{\rho}{1-\rho} > 0$
  - If $\frac{P_c}{P_H} - \frac{\rho}{1-\rho} < \eta_1 < \eta_2$, then $P_{out}(\rho, \eta_1, P_c) = \hat{P}_{out}(\rho, \eta_1, P_c)$ and similarly $P_{out}(\rho, \eta_2, P_c) = \hat{P}_{out}(\rho, \eta_2, P_c)$. Since $\hat{P}_{out}(\rho, \eta, P_c)$ is strictly decreasing with battery efficiency η, we have
  $$P_{out}(\rho, \eta_1, P_c) > P_{out}(\rho, \eta_2, P_c).$$
  - If $\eta_1 \leq \frac{P_c}{P_H} - \frac{\rho}{1-\rho} < \eta_2$, then $P_{out}(\rho, \eta_1, P_c) = 1$ and $P_{out}(\rho, \eta_2, P_c) = \hat{P}_{out}(\rho, \eta_2, P_c)$. Therefore
  $$P_{out}(\rho, \eta_1, P_c) = 1 > P_{out}(\rho, \eta_2, P_c).$$
  - If $\eta_1 < \eta_2 \leq \frac{P_c}{P_H} - \frac{\rho}{1-\rho}$, then $P_{out}(\rho, \eta_1, P_c) = P_{out}(\rho, \eta_2, P_c) = 1$, which means
  $$P_{out}(\rho, \eta_1, P_c) = P_{out}(\rho, \eta_2, P_c).$$
- Suppose $\eta_1 < \eta_2$ and $\frac{P_c}{P_H} - \frac{\rho}{1-\rho} \leq 0$, we have $\frac{P_c}{P_H} - \frac{\rho}{1-\rho} \leq \eta_1 < \eta_2$. Then it could be easily verified that
$$P_{out}(\rho, \eta_1, P_c) > P_{out}(\rho, \eta_2, P_c).$$

Combining all the above cases, we can conclude that $P_{out}(\rho, \eta, P_c)$ is a non-increasing function of battery efficiency η given any non-zero circuit power P_c for $\rho \in [0, 1)$. Next, we proceed to prove the monotonicity of $P^*_{out}(\eta, P_c)$.

Assuming $\eta_1 < \eta_2$ again, then we could argue that $\frac{P_c}{P_H} - \frac{\rho_1^*}{1-\rho_1^*} < \eta_1$ and $\frac{P_c}{P_H} - \frac{\rho_2^*}{1-\rho_2^*} < \eta_2$, where ρ_1^* and ρ_2^* are the optimal save-ratio for $\eta = \eta_1$ and $\eta = \eta_2$, respectively. Therefore we only need to consider two cases: $\max\left\{\frac{P_c}{P_H} - \frac{\rho_1^*}{1-\rho_1^*}, \frac{P_c}{P_H} - \frac{\rho_2^*}{1-\rho_2^*}\right\} < \eta_1 < \eta_2$ and $\eta_1 \leq \max\left\{\frac{P_c}{P_H} - \frac{\rho_1^*}{1-\rho_1^*}, \frac{P_c}{P_H} - \frac{\rho_2^*}{1-\rho_2^*}\right\} <$

η_2. From the arguments we have given for the proof of the monotonicity of P_{out} we know that, under these two conditions we have

$$P_{out}(\rho, \eta_1, P_c) > P_{out}(\rho, \eta_2, P_c).$$

Therefore,

$$P^*_{out}(\eta_1, P_c) > P_{out}(\rho^*_1, \eta_2, P_c) \geq P^*_{out}(\eta_2, P_c)$$

which completes the proof of the monotonicity for $P^*_{out}(\eta, P_c)$. With similar arguments, we could get the results regarding circuit power P_c. Proposition 5.4.1 is thus proved.

C.2 Proof of Lemma 5.4.1

Since $F_\Gamma(\cdot)$ is non-negative and non-decreasing, we have

$$a < b \;\; \Rightarrow \;\; F_\Gamma\left(\frac{a}{x}\right) \leq F_\Gamma\left(\frac{b}{x}\right) \tag{C.3}$$

for any $x \in [0, P_H]$. Since $f_X(\cdot)$ is non-negative, this leads to

$$\begin{aligned} a < b \;\; \Rightarrow \;\; & \int_0^{P_H} f_X(x) F_\Gamma\left(\frac{a}{x}\right) dx \\ & \leq \int_0^{P_H} f_X(x) F_\Gamma\left(\frac{b}{x}\right) dx. \end{aligned}$$

Given the form of P_{out} in Problem (P2), with ρ appearing only in the numerator of the argument of $F_\Gamma(\cdot)$, we conclude that P_{out} is a non-decreasing function of $g(\rho) = \left(2^{\frac{Q}{(1-\rho)T}} - 1\right)(1-\rho)$. Hence minimizing $g(\rho)$ is equivalent to minimizing P_{out}. The first and second derivatives of $g(\rho)$ are

$$\begin{aligned} g'(\rho) &= 2^{\frac{Q}{(1-\rho)T}} (\ln 2) \frac{Q}{(1-\rho)T} - 2^{\frac{Q}{(1-\rho)T}} + 1 \\ g''(\rho) &= 2^{\frac{Q}{(1-\rho)T}} (\ln 2)^2 \frac{Q^2}{T^2(1-\rho)^3} > 0 \quad \text{since } Q > 0. \end{aligned}$$

Let $h(\rho) = g'(\rho)$. From the second equation above, $h(\rho)$ is an increasing function. In the range $0 \leq \rho \leq 1$, $h(\rho)$ is thus minimized at $\rho = 0$, i.e. the minimum of $g'(\rho)$ is

$h(0)$, given by

$$g'_{\min} = 2^{\frac{Q}{T}}(\ln 2)\frac{Q}{T} - 2^{\frac{Q}{T}} + 1 \quad \text{(C.4)}$$

$$= 2^{\frac{Q}{T}}\left(\ln 2^{Q/T} - 1\right) + 1 > 0 \quad \text{(C.5)}$$

for $Q > 0$. In other words, the smallest value that the gradient of $g(\rho)$ can take in the range $0 \leq \rho \leq 1$ for any feasible Q is positive, which implies that $g(\rho)$ is increasing and therefore minimized at $\rho = 0$, as claimed. The proof of Lemma 5.4.1 is thus completed.

C.3 Proof of Lemma 5.4.2

To prove Property 1, we observe that as noted in the proof of Lemma 5.4.1, P_{out} is a monotonic function of $g(\rho) = \frac{(2^{\frac{Q}{(1-\rho)T}}-1)}{(\frac{\rho}{1-\rho}+\eta)}$ in Problem (P3), hence minimizing $g(\rho)$ leads to the same solution as minimizing P_{out}. The first derivative of $g(\rho)$ is

$$g'(\rho) = \frac{2^{\frac{Q}{(1-\rho)T}}(\ln 2)\frac{Q}{(1-\rho)T}[\rho + \eta(1-\rho)] - 2^{\frac{Q}{(1-\rho)T}} + 1}{[\rho + \eta(1-\rho)]^2}$$

$$= \frac{u(\rho)}{[\rho + \eta(1-\rho)]^2}.$$

It is clear in the above that the sign of $g'(\rho)$ is the same as that of $u(\rho)$. Since $u(1) = +\infty$ and $u(\rho)$ is a differentiable function, if $u(0)$ is negative then there exists a value $\rho_c \in (0,1)$ such that $u(\rho_c) = 0 = g'(\rho_c)$. It is easily verified that $u'(\rho) > 0$; hence ρ_c is the unique optimal value of ρ in this case. Conversely, if there exists an ρ_c such that $u(\rho_c) = 0$, then $u(0)$ must be negative. Hence $u(0) < 0$ is a necessary and sufficient condition for the optimal ρ to lie in $(0,1)$.

The condition $u(0) < 0$ translates into the following condition on η, which proves the first part of the lemma:

$$u(0) < 0 \;\Rightarrow\; 2^{\frac{Q}{T}}(\ln 2)\frac{Q}{T}\eta - 2^{\frac{Q}{T}} + 1 < 0$$

$$\Rightarrow\; \eta < \frac{2^{\frac{Q}{T}} - 1}{2^{\frac{Q}{T}}(\ln 2)\frac{Q}{T}}. \quad \text{(C.6)}$$

To prove the second point, suppose $\rho_1^*(\eta_1, 0)$ and $\rho_2^*(\eta_2, 0)$ are optimal save-ratios of (P3) for SESD efficiencies η_1 and η_2, where $\eta_1 < \eta_2$. Then, $u(\rho_1^*, \eta_1) = 0$ and $u(\rho_2^*, \eta_2) = 0$. Since $\eta_1 < \eta_2$ and $u(\rho, \eta)$ is an increasing function of η, we have $u(\rho_1^*, \eta_2) > 0$. Combining what we have that $u(\rho, \eta)$ is an increasing function of ρ, $u(\rho_2^*, \eta_2) = 0$ and $u(\rho_1^*, \eta_2) > 0$, we may conclude $\rho_2^*(\eta_2, 0) < \rho_1^*(\eta_1, 0)$. Lemma 5.4.2 is thus proved.

C.4 Proof of Lemma 5.4.3

According to the proof of Proposition 5.4.1, the first derivative of $\hat{P}_{out}(\rho, \eta, P_c)$ with respect to η, P_c and ρ is,

$$\frac{\partial \hat{P}_{out}}{\partial \beta} = \int_{\frac{P_c}{\frac{\rho}{1-\rho}+\eta}}^{P_H} f_X(x) f_\Gamma\left[g(\cdot)\right] \frac{\partial g(\cdot)}{\partial \beta} dx$$

where β could be η, P_c or ρ and $g(\rho, \eta, P_c) = \frac{2^{\frac{Q}{(1-\rho)T}}-1}{x[\frac{\rho}{1-\rho}+\eta]-P_c}$. Furthermore, we have

$$\begin{aligned}
\frac{\partial g(\rho)}{\partial \rho} &= \frac{2^{\frac{Q}{(1-\rho)T}} (\ln 2) \frac{Q}{(1-\rho)T} \left[\rho + \eta(1-\rho) - (1-\rho)\frac{P_c}{x}\right]}{x\left[\rho + \eta(1-\rho) - (1-\rho)\frac{P_c}{x}\right]^2} \\
&\quad - \frac{2^{\frac{Q}{(1-\rho)T}} - 1}{x\left[\rho + \eta(1-\rho) - (1-\rho)\frac{P_c}{x}\right]^2} \\
&= \frac{v(\rho)}{x\left[\rho + \eta(1-\rho) - (1-\rho)\frac{P_c}{x}\right]^2}.
\end{aligned}$$

With similar arguments about $u(\rho)$ in the proof of Lemma 5.4.2, we claim that $v(0) < 0, \forall x \in (\frac{P_c}{\frac{\rho}{1-\rho}+\eta}, P_H]$ is a sufficient condition of having $\rho^*(\eta, P_c) > 0$ while $P_c < \eta P_H$.

Since $v(0)$ is an increasing function of x, the condition $v(0) < 0, \forall x \in (\frac{P_c}{\frac{\rho}{1-\rho}+\eta}, P_H]$ translates into the following condition on η and P_c

$$\begin{aligned}
v(0) &= 2^{\frac{Q}{T}} (\ln 2) \frac{Q}{T} \left(\eta - \frac{P_c}{x}\right) - 2^{\frac{Q}{T}} + 1 \\
&< 2^{\frac{Q}{T}} (\ln 2) \frac{Q}{T} \left(\eta - \frac{P_c}{P_H}\right) - 2^{\frac{Q}{T}} + 1 < 0 \\
&\Longrightarrow 0 < \eta - \frac{P_c}{P_H} < \frac{2^{\frac{Q}{T}} - 1}{2^{\frac{Q}{T}} (\ln 2) \frac{Q}{T}}.
\end{aligned}$$

Combined with the fact that $\rho^*(\eta, P_c) > \frac{\frac{P_c}{P_H} - \eta}{1 - \eta + \frac{P_c}{P_H}}$ when $P_c \geq \eta P_H$, we may conclude

$$\rho^*(\eta, P_c) > 0, \quad \eta - \frac{P_c}{P_H} < \frac{2^{\frac{Q}{T}} - 1}{2^{\frac{Q}{T}} (\ln 2) \frac{Q}{T}}. \tag{C.7}$$

Lemma 5.4.3 is thus proved.

C.5 Proof of Lemma 5.5.1

Let $Z = P\Gamma$, where P and Γ are exponential random variables with mean λ_p and λ_γ respectively. Then the PDF of Z could be derived as follows,

$$\begin{aligned} F_Z(z) &= \Pr\{P\Gamma \leq z\} \\ &= 1 - \frac{1}{\lambda_p} \int_0^\infty e^{-\frac{z}{p\lambda_\gamma}} e^{-\frac{p}{\lambda_p}} dp \\ &= 1 - 2\sqrt{\frac{z}{\lambda_p \lambda_\gamma}} K_1\left(2\sqrt{\frac{z}{\lambda_p \lambda_\gamma}}\right) \end{aligned} \tag{C.8}$$

where $K_1(x)$ is the first-order modified Bessel function of the second kind and the last equality is given by [107, §3.324.1]:

$$\int_0^\infty \exp\left(-\frac{\beta}{4x} - \gamma x\right) dx = \sqrt{\frac{\beta}{\gamma}} K_1\left(\sqrt{\beta\gamma}\right)$$

where $\Re(\beta) \geq 0, \Re(\gamma) \geq 0$. Let $M = \frac{1}{\sqrt{\lambda_p \lambda_\gamma}}$. Taking the derivative of $F(z)$ yields

$$\begin{aligned} f(z) &= M\left\{-\frac{1}{\sqrt{z}} K_1\left(2M\sqrt{z}\right) - 2\sqrt{z}\left(K_1\left(2M\sqrt{z}\right)\right)'\right\} \\ &= M\left\{-\frac{1}{\sqrt{z}} K_1\left(2M\sqrt{z}\right)\right. \\ &\quad \left. - 2\sqrt{z}\left(-K_0(2M\sqrt{z}) - \frac{1}{2M\sqrt{z}} K_1(2M\sqrt{z})\right)\frac{M}{\sqrt{z}}\right\} \\ &= 2M^2 K_0\left(2M\sqrt{z}\right) \\ &= \frac{2}{\lambda_p \lambda_\gamma} K_0\left(2\sqrt{\frac{z}{\lambda_p \lambda_\gamma}}\right) \end{aligned} \tag{C.9}$$

where $\frac{\partial K_v(z)}{\partial z} = -K_{v-1}(z) - \frac{v}{z} K_v(z)$.

Next, we characterize the outage probability using (C.9). According to (5.15), we have

$$\begin{aligned} P_{out}^{*} &= \Pr\left[P\Gamma < C\right] \\ &= \int_{0}^{C} \frac{2}{\lambda_p \lambda_\gamma} K_0\left(2\sqrt{\frac{z}{\lambda_p \lambda_\gamma}}\right) dz. \end{aligned} \tag{C.10}$$

Let $X = \frac{z}{\lambda_p \lambda_\gamma}$ and $D = \frac{C}{\lambda_p \lambda_\gamma}$. We then have

$$P_{out}^{*} = 2\int_{0}^{D} K_0\left(2\sqrt{x}\right) dx. \tag{C.11}$$

Using the series presentation [107, §8.447.3], we have

$$K_0(x) = -\ln\left(\frac{x}{2}\right) I_0(x) + \sum_{k=0}^{\infty} \frac{x^{2k}}{2^{2k}(k!)^2} \psi(k+1) \tag{C.12}$$

with the series expansion for the modified Bessel function given by

$$I_0(x) = \sum_{k=0}^{\infty} \frac{x^{2k}}{2^{2k}(k!)^2}. \tag{C.13}$$

(C.11) could be expanded as

$$P_{out}^{*} = \sum_{k=0}^{\infty} \frac{2}{(k!)^2} \left[-\frac{1}{2}\int_{0}^{D} x^k \ln x dx + \psi(k+1)\int_{0}^{D} x^k dx\right] \tag{C.14}$$

where

$$\psi(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma(x)'}{\Gamma(x)} \tag{C.15}$$

is the digamma function [105]. Since the two integrals in (C.14) could be evaluated as

$$\begin{aligned} \int_{0}^{D} x^k dx &= \frac{D^{k+1}}{k+1} \\ \int_{0}^{D} x^k \ln x dx &= x^{k+1}\left(\frac{\ln x}{k+1} - \frac{1}{(k+1)^2}\right)\Bigg|_{x=0}^{x=D} \\ &= D^{k+1}\left(\frac{\ln D}{k+1} - \frac{1}{(k+1)^2}\right) \end{aligned}$$

where $\lim_{x\to 0}(x\ln x) = 0$. Then we have

$$P_{out}^{*} = \sum_{k=0}^{\infty} \frac{2}{(k!)^2}\frac{D^{k+1}}{k+1}\left[-\frac{1}{2}\left(\ln D - \frac{1}{k+1}\right) + \psi(k+1)\right]. \tag{C.16}$$

Since $D = \frac{C}{\lambda_p\lambda_\gamma} = \frac{C\sigma_n^2}{\lambda_p\sigma_h^2} = \frac{C}{\bar{\gamma}}$, (5.18) follows. Lemma 5.5.1 is thus proved.

Bibliography

[1] CISCO, *CISCO VNI Mobile Forecast*, Aug. 2013. [Online]. Available: http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html

[2] Shannon, "A mathematical theory of communication," *Bell Labs Tec. J.*, vol. 27, pp. 379–423, Jul. and Oct. 1948.

[3] G. P. Fettweis and E. Zimmermann, "ICT energy consumption-trends and challenges," in *Proc. Int. Symp. Wireless Personal Multimedia Commun. (WPMC)*, Lapland, Finland, Sep. 2008.

[4] T. Hwang, C. Yang, G. Wu, S. Li, and G. Y. Li, "OFDM and its wireless applications: a survey," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1673–1694, May 2009.

[5] *"C-RAN: the road towards green RAN,"*, White paper, ver. 2.5 ed., China Mobile Res. Inst., Oct. 2011.

[6] Y. Zhou and W. Yu, "Approximate bounds for limited backhaul uplink multicell processing with single-user compression," in *Canadian Workshop on Information Theory (CWIT)*, Toronto, ON, Jun. 2013.

[7] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5646–5658, Nov. 2013.

[8] J. Zhao, T. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2762–2775, Jun. 2013.

[9] I. F. Akyidiz, W. Su, Y. S. Subramaniam, and E. Cayirei, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[10] A. Sinha and A. Chandrakasan, "Dynamic power management in wireless sensor networks," *IEEE Design Test Comp.*, vol. 18, no. 2, pp. 62–74, Mar./Apr. 2001.

[11] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: survey and implications," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443–461, Third Quarter 2011.

[12] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary user behavior in cellular networks and implications for dynamic spectrum access," *IEEE Commun. Mag.*, vol. 47, no. 3, pp. 88–95, Mar. 2009.

[13] F. Richter, A. J. Fehske, and G. P. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks," in *Proc. IEEE Vehic. Tech. Conf. (VTC)*, Anchorage, AK, Sep. 2009.

[14] 3GPP TR 32.826, *Telecommunication management: Study on energy savings management (ESM)*, release 10 ed., Mar. 2010.

[15] S. V. Hanly and R. Mathar, "On the optimal base-station density for cdma cellular networks," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1274–1281, Aug. 2002.

[16] W. Wang and G. Shen, "Energy efficiency of heterogeneous cellular network," in *Proc. IEEE Vehic. Tech. Conf. (VTC)*, Ottawa, ON, Sep. 2010.

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

[18] S. V. Hanly, "An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity," *IEEE J. Select. Areas Commun.*, vol. 13, no. 7, pp. 1332–1340, Sep. 1995.

[19] S. Das, H. Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," in *Proc. IEEE INFOCOM*, San Francisco, CA, Mar. 2003.

[20] M. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *Proc. IEEE ICC GreenCom Wkshps.*, Dresden, Jun. 2009.

[21] K. Samdanis, D. Kutscher, and M. Brunner, "Dynamic energy-aware network re-configuration for cellular urban infrastructures," in *Proc. IEEE GLOBECOM GreenCom Wkshps.*, Miami, FL, Dec. 2010.

[22] A. P. Jardosh, K. Papagiannaki, E. M. Belding, K. C. Almeroth, G. Iannaccone, and B. Vinnakota, "Green wlans: on-demand wlan infrastructures," *Mobile Networks and Applications*, vol. 14, Dec. 2009.

[23] 3G Americas, *The benefits of SON in LTE: Self-optimizing and self organizing networks*, white paper ed., Dec. 2009.

[24] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green celular networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 74–79, Nov. 2010.

[25] X. Weng, D. Cao, and Z. Niu, "Energy-efficient cellular network planning under insufficient cell zooming," in *Proc. IEEE Vehic. Tech. Conf. (VTC)*, Yokohama, May 2011.

[26] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 996–1019, Third Quarter 2013.

[27] R. R. Collman, "Evaluation of methods for determining the mobile traffic distribution in cellular radio networks," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1628–1635, Nov. 2001.

[28] Y. Che, R. Zhang, and Y. Gong, "On spatial capacity of wireless ad hoc networks with threshold based scheduling," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6915–6927, Dec. 2014.

[29] J. Kingman, *Poisson Process*. Oxford, England: Oxford University Press, 1993.

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

[31] O. Blume, H. Eckhardtm, S. Klein, E. Kuehn, and W. M. Wajda, "Energy savings in mobile networks based on adaptation to traffic statistic," *Bell Labs Tec. J.*, vol. 15, no. 2, pp. 77–94, Sep. 2010.

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

[33] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the lambert w function," *Advances in Computational Mathematics*, 1996.

[34] *3GPP; Technical specification group radio access network; Physical layer aspects for evolved UTRA (R7)*, Tr 25.814 v7.0.0 ed., Jun. 2006. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/25814.htm

[35] M. Guowang, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.

[36] C. Xiong, Y. G. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink ofdma networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3874–3886, Nov. 2011.

[37] C. Xiong, Y. G. Li, Y. Chen, and S. Xu, "Energy-efficient resource allocation in ofdma networks," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3767–3778, Dec. 2012.

[38] C. Isheden and G. P. Fettweis, "Energy-efficient multi-carrier link adaptation with sum rate-depend circuit power," in *Proc. IEEE Globecom*, Miami, FL, Dec. 2012.

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

[40] H. Yin and H. Liu, "An efficient multiser loading algorithm for ofdm-based broadband wireless systems," in *Proc. IEEE Globecom*, San Francisco, CA, Nov. 2000.

[41] G. Song and Y. G. Li, "Cross-layer optimization for ofdm wireless networks - part ii: algorithm development," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 625–634, Mar. 2005.

[42] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser ofdm systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.

[43] M. Ergen, S. Coleri, and P. Varaiya, "Qos aware adaptive resource allocation techniques for fair scheduling in ofdma based broadband wireless access systems," *IEEE Trans. Broadcasting*, vol. 49, no. 4, pp. 362–370, Dec. 2003.

[44] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser ofdm with adaptive sub-carrier, bit and power allocation," *IEEE J. Select. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[45] D. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for ofdma downlink systems," in *Proc. IEEE ISIT*, Seattle, WA, Jul. 2006.

[46] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser ofdm systems with proportional rate constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726–2734, Nov. 2005.

[47] S. Boyd, *Convex optimization II*. Stanford University. [Online]. Available: http://www.stanford.edu/class/ee364b/lectures.html

[48] H. Kim and G. Veciana, "Leveraging dynamic spare capacity in wireless systems to conserve mobile terminals' energy," *IEEE/ACM Trans. Networking*, vol. 18, no. 3, pp. 802–815, Jun. 2010.

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

[50] M. Mohseni, R. Zhang, and J. M. Cioffi, "Optimized transmission for fading multiple-access and broadcast channels with multiple antennas," *IEEE J. Select. Areas Commun.*, vol. 24, no. 8, pp. 1627–1639, Aug. 2006.

[51] P. Milgrom and I. Segal, "Envelope theorems for arbitrary choice sets," *Econometrica*, vol. 70, no. 2, pp. 583–601, Mar. 2002.

[52] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer in multiuser ofdm systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2282–2294, Apr. 2014.

[53] S. Tombaz, P. Monti, K. Wang, A. Vastberg, M. Forzati, and J. Zander, "Impact of backhauling power consumption on the deployment of heterogeneous mobile networks," in *Proc. IEEE Globecom*, Houston, TX, Dec. 2011.

[54] M. Hong, R. Sun, H. Baligh, and Z. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 226–240, Feb. 2013.

[55] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud radio access networks," in *Proc. IEEE Globecom*, Atlanta, GA, Sep. 2013.

[56] W. Liao, M. Hong, Y. Liu, and Z. Luo, "Base station activation and linear tansceiver design for optimal resource management in heterogenous networks," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3939–3952, Aug. 2014.

[57] B. Dai and W. Yu, "Sparse beamforming for limited-backhaul network mimo system via reweighted power minimization," in *Proc. IEEE Globecom*, Atlanta, GA, Sep. 2013.

[58] Y. Cheng, M. Pesavento, and A. Philipp, "Joint network optimization and downlink beamforming for comp transmissions using mixed integer conic programming," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3972–3987, Aug. 2013.

[59] Skype, *How much bandwidth does Skype need?* [Online]. Available: https://support.skype.com/EN/faq/FA1417/how-much-bandwidth-does-skype-need

[60] F. R. Farrokhi, L. Tassiulas, and K. J. R. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1313–1324, Oct. 1998.

[61] M. Hong and Z. Luo, "Distributed linear precoder optimization and base station selection for an uplink heterogeneous network," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3214–3228, Jun. 2013.

[62] M. Rasti, A. R. Sharafat, S. Member, and J. Zander, "Pareto and energy-efficient distributed power control with feasibility check in wireless networks," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 245–255, Jan. 2011.

[63] S. Luo, R. Zhang, and T. J. Lim, "Optimal power and range adaptation for green broadcasting," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4592–4603, Sep. 2013.

[64] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1525–1536, Sep. 2011.

[65] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell mimo cooperative networks: a new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[66] R. Zhang and S. Cui, "Cooperative interference management with miso beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.

[67] I. Hwang, B. Song, and S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, Jun. 2013.

[68] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[69] A. Dhaini, P. H. Ho, G. Shen, and B. Shihada, "Energy efficiency in tdma-based next-generation passive optical access networks," *IEEE/ACM Trans. Netw.*, vol. 22, no. 3, pp. 850–863, Jun. 2013.

[70] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs, "Lte release 12 and beyond," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 154–160, Jul. 2013.

[71] L. Liu, R. Zhang, and K. C. Chua, "Achieving global optimality for weighted sum-rate maximization in the k-user gaussian interference channel with multiple antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1933–1945, May 2012.

[72] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[73] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, Feb. 2006.

[74] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Foundations and Trends in Machine Learning*, vol. 4, no. 1, pp. 1–106, Jan. 2012.

[75] O. Mehanna, N. Sidiropoulos, and G. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2660–2674, May 2013.

[76] S. Leyffer, *Mixed integer nonlinear programming*. Springer, 2012.

[77] M. Schubert and H. Boche, "Iterative multiuser uplink and downlink beamforming under sinr constraints," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2324–2334, Jul. 2005.

[78] L. Zhang, R. Zhang, Y. C. Liang, Y. Xin, and H. V. Poor, "On Gaussion mimo bc-mac duality with multiple transmit covariance constraints," *IEEE Trans. Inform. Theory*, vol. 58, no. 4, pp. 2064–2078, Apr. 2012.

[79] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, Dec. 2008.

[80] D. R. Hunter and K. Lange, "Quantile regression via an mm algorithm," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, Mar. 2000.

[81] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.

[82] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed mimo receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[83] J. Sangiamwong, Y. Saito, N. Miki, T. Abe, S. Nagata, and Y. Okumura, "Investigation on cell selection methods associated with inter-cell interference coordination in heterogeneous networks for lte-advanced downlink," in *European Wireless Conference Sustainable Wireless Technologies*, Vienna, Austria, Apr. 2011.

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

[85] J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Trans. Commun.*, vol. 60, no. 1, pp. 220–230, Jan. 2012.

[86] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1180–1189, Mar. 2012.

[87] B. Devillers and D. Gunduz, "A general framework for the optimization of energy harvesting communication systems with battery imperfections," *J. Commun. and Netw., Special Issue on Energy Harvesting in Wireless Networks*, vol. 14, no. 2, pp. 130–139, Apr. 2012.

[88] K. Tutuncuoglu and A. Yener, "Optimal power policy for energy harvesting transmitters with inefficient energy storage," in *Proc. Annual Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 2012.

[89] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, Sep. 2011.

[90] C. K. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4808–4818, Sep. 2012.

[91] O. Ozel and S. Ulukus, "Information-theoretic analysis of an energy harvesting communication system," in *Proc. IEEE PIMRC*, Instanbul, Sep. 2010.

[92] R. Rajesh, V. Sharma, and P. Viswanath, "Capacity of fading gaussian channel with an energy harvesting sensor node," in *Proc. IEEE Globecom*, Houston, TX, Dec. 2011.

[93] F. Iannello, O. Simeone, and U. Spagnolini, "Medium access control protocols for wireless sensor networks with energy harvesting," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1381–1389, May. 2012.

[94] V. Sharma, U. Mukherji, and V. Joseph, "Efficient energy management policies for networks with energy harvesting sensor nodes," in *Proc. Annual Allerton Conference on Communication, Control and Computing*, Urbana-Champaign, IL, Sep. 2008.

[95] L. Ren-Shiou, P. Sinha., and C. E. Koksal, "Joint energy management and resource allocation in rechargeable sensor networks," in *Proc. IEEE INFOCOM*, San Diego, CA, Mar. 2010.

[96] B. D. D. Gunduz, "Two-hop communication with energy harvesting," in *Proc. Int. Workshop Computational Advances Multi-Sensor Adaptive Process*, San Juan, Dec. 2011.

[97] C. Huang, R. Zhang, and S. Cui, "Throughput maximization for the gaussian relay channel with energy harvesting constraints," *IEEE J. Sel. Areas Commun., Special Issue on Theories and Methods for Advanced Wireless Relays*, vol. 31, no. 8, pp. 1469–1479, Dec. 2012.

[98] A. C. Fu, E. Modiano, and J. N. Tsitsiklis, "Optimal energy allocation and admission control for communications satellites," *IEEE/ACM Trans. Netw.*, vol. 11, no. 3, pp. 488–500, Jun. 2003.

[99] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy harvesting sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1326–1336, Apr. 2010.

[100] M. Jayalakshmi and K. Balasubramanian, "Simple capacitors to supercapacitors - an overview," *Int. J. Electrochem. Sci.*, vol. 3, pp. 1196–1217, Oct. 2008.

[101] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 20, pp. 545–554, Feb. 2010.

[102] M. Springer, *The Algebra of Random Variables*. Wiley New York, 1979.

[103] S. A. Halim and H. Yanikomeroglu, "On the approximation of the generalized-k distribution by a gamma distribution for modeling composite fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 706–713, Jul. 2008.

[104] S. Nadarajah, "Exact distribution of the product of m gamma and n pareto random variables," *Journal of Computational and Applied Mathematics*, vol. 235, no. 15, pp. 4496–4512, Jun. 2011.

[105] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Courier Dover Publications, 1965.

[106] R. Walter, *Real and Complex Analysis*. New York: McGraw-Hill Book, 1987.

[107] I. Gradshteyn and I. Ryzhik, *Table of integrals, series, and products*. Academic Press, Inc., 1980.