Dynamic Resource Allocation for Energy-Constrained Wireless Networks over Time-Varying Channels

ZHANG XIAOLU

B. Eng., Beijing Univ. of Posts & Telecomm.

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPT. OF ELECTRICAL & COMPUTER ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2008

Abstract

Dynamic Resource Allocation for Energy-Constrained Wireless Networks over Time-Varying Channels

by

Zhang Xiaolu

in Department of Electrical and Computer Engineering

National University of Singapore

The focus of this thesis is on the establishment of a theoretical framework on dynamic resource allocation for energy-constrained wireless networks over time-varying channels. This framework chooses the end-user application needs as the optimization objective, establishes the theoretically optimal performance benchmark under system constraints, and designs solution that is easy to be integrated in practice systems using mathematical tools, such as gradient algorithm and dual decomposition. This framework is applied to different network situations including infrastructure-based wireless network, wireless sensor network (WSN) and orthogonal frequency division multiplexing (OFDM)-based multi-hop network. We attempt to address the following three questions: 1) How to jointly optimize average rate and rate oscillation in wireless networks supporting variable rate transmission; 2) How to jointly design quantization and transmission for lifetime maximization in WSNs; 3) how to minimize end-to-end outage and maximize average rate in OFDM-based relay networks. All above problems are investigated using convex optimization-based approaches.

For the first problem, we demonstrate that a proposed utility function can be used

to facilitate the choice of the combinations of average rate and rate oscillation, both of which are important performance metrics. A gradient based scheduling algorithm is developed to maximize the proposed utility function. The dynamics of transmission rate under this algorithm is analyzed using ordinary differential equation. In addition, the condition under which generalized gradient scheduling algorithm (GGSA) is asymptotically optimal is addressed.

Unlike the infrastructure-based network, a WSN cannot centrally allocate resources due to limited computing capacity and energy. We demonstrate how the network lifetime can be maximized by integrated design of quantization and transmission in a partially distributed way, where each node is aware of the local information and little common information. The behavior of the algorithm's convergence is also explored. Numerical examples show significant lifetime gain and the gain is more significant when sensing environment becomes more heterogeneous.

Finally, we study subcarrier, power and time allocation to minimize the end-to-end outage probability and maximize the end-to-end average rate, respectively, in a one-dimensional multi-hop network under an average transmission power constraint. We derive the optimal resource allocation schemes which determine the system performance limits. However, they incur high computational complexity and high signaling overhead. Several suboptimal algorithms with low complexity and reduced overhead are proposed. The tradeoff between performance of these algorithms and their complexity and overhead is also discussed.

to my parents

Contents

List of Figures

Acknowledgements

I am deeply grateful to my supervisors, Meixia Tao and Chun Sum Ng, for their guidance, support and encouragement. Working with them has been an exciting learning experience. In particular, I would like to thank Dr. Tao for her many valuable suggestions in my research, paper writing, presentation skill and her availability to answer my question and listen to my ideas. Her attention to detail and her enthusiasm for research are strongly impressed on me.

I would also like to thank Wenhua Jiao of Bell Labs Research China for his support and encouragement during my internship. In addition, I would like to thank Yan Xin and Vikram Srinivasan, for being the committee member of my Ph.D. Qualifying Examination and for their constructive comments and suggestions.

I would like to thank all friends in communication laboratory, institute for infocomm research (I2R) and elsewhere. I am grateful to Shengwei Hou, Lan Zhang, Qi Zhang, Le Cao, Lokesh Bheema Thiagarajan, Fangming Liu, Bin Da, Yonglan Zhu, Jun He, Jinhua Jiang, Jing Jiang and Zhen Zhang. Their friendship goes through my graduate study.

My most gratitude goes to my parents who provided me endless love, support and pride. This thesis is partly theirs too.

Curriculum Vitæ

Zhang Xiaolu

Education

2001-2005	Beijing Univ. of Posts and Telecomm., Beijing, China B. Eng., Telecomm. Engineering
2005-2008	National Univ. of Singapore, Singapore Ph.D., Electrical and Computer Engineering

Experience

2007	Bell Labs Research China, Beijing, China Intern Researcher
2006-2008	National University of Singapore, Singapore Graduate Assistant

Honors and Awards

2005-2008	NUS Graduate Scholarship
2005	International Mathematical Contest in Modelling (MCM), Meritorious
2004	China Undergraduate Mathematical Contest in Modeling (CUMCM), First Prize
2002	Enterprise Scholarship, BUPT

Activities

Student Member	IEEE, 2006-Present

Reviewer	*IEEE Transactions on Wireless Communications, Elsevier Computer Communications, IEEE ICC 2007, IEEE GLOBECOM 2007, IEEE ICC 2008, IEEE GLOBECOM 2008 and IEEE VTC 2008*

List of Abbreviations

APFT	adaptive power and fixed time allocation
APTA	adaptive power and time allocation
ARRV	average rate and rate variance
BER	bit error rate
BLUE	best linear unbiased estimator
BS	base station
CDMA	code division multiple access
CSI	channel state information
FC	fusion center
FDMA	frequency division multiple access
FPAT	fixed power and adaptive time allocation
GGSA	generalized gradient scheduling algorithm
HDR	high data rate
IAS	iterative algorithm of sub-optimal power and time allocation
ILS	inverse-log scheduling
JTPC	joint time sharing and power control
KKT	Karush-Kuhn-Tucher
LATS	low-complexity algorithm of optimal discrete time-sharing fraction assignment
LTE	long-term-evolution

MAC	media access control
MSE	mean square error
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
pdf	probability density function
PMP	point-to-multi-point
PS	power scheduling
QAM	quadrature amplitude modulation
QoS	quality of service
QTSL	quantized time sharing with limited channel feedback
SNR	signal to noise ratio
TBS	two-nested binary search
TCP	transmission control protocol
TDM	time division multiplexing
TDMA	time division multiple access
TS	time sharing
UPTA	uniform power and time allocation
UTP	uniform TDMA with power control
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	wireless local area networks
WMAN	wireless metropolitan area network
WSN	wireless sensor network

Notations

Standard notations $\mathbb{R}$ and $\mathbb{R}_+$ are used to denote the sets of real and real nonnegative numbers. Correspondingly, $\mathbb{R}^N$ denotes a standard vector space, with elements $\mathbf{a} \in \mathbb{R}^N$ being column vectors $\mathbf{a} := (a_1, \ldots, a_N)^T$.

$Prob(A)$ is the probability of event A.

$pdf(a)$ is the probability density function of the random variable a.

$\mathbb{E}[a]$ is the expectation of the random variable a.

$\mathbf{a}^T$ is the transpose of $\mathbf{a}$.

For vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$,

$$\mathbf{a}^T \cdot \mathbf{b} \doteq \sum_{i=1}^{N} a_i b_i$$

and

$$\mathbf{a} \times \mathbf{b} \doteq (a_1 b_1, \ldots, a_N b_N)^T.$$

Euclidean norm $|\mathbf{a}| \doteq \sqrt{\mathbf{a}^T \cdot \mathbf{a}}$.

Geometric mean $\tilde{\mathbf{a}} = (\prod_{i=1}^{N} a_i)^{1/N}$.

Harmonic mean $\bar{\mathbf{a}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{a_i}}$.

The gradient of a function $U(\mathbf{a}, \mathbf{b})$ with respect to $\mathbf{a}$ is denoted by

$$\nabla_{\mathbf{a}} U(\mathbf{a}, \mathbf{b}) \doteq \left(\frac{\partial U(\mathbf{a}, \mathbf{b})}{\partial a_1}, \ldots, \frac{\partial U(\mathbf{a}, \mathbf{b})}{\partial a_N} \right)^T.$$

Chapter 1

Introduction

Mobile and wireless communications have experienced impressive growth in quality of services and diversity of services. The popularity of wireless cellular networks, wireless sensor networks (WSNs), broadband wireless metropolitan area networks (WMANs) and wireless local area networks (WLANs) demonstrate a high demand for reliable multimedia service, situation awareness application and high-speed data transmission. From the start of this century, various convergence in these networks are taking place for providing an ubiquitous wireless experience.

Three aspects of wireless communication environment [33] present a fundamental technical challenge for wireless system design. Limited radio resources must be shared between many geographically separated users. Due to the broadcast nature of wireless channel, the data transmission to one user may become interference to others. Moreover, wireless channel suffers from time-varying large-scale, small-scale fading and noise, which makes it a problem to keep communication as reliable as that available on wireline networks. In addition, most portable communication devices have limited battery power supply and small size. It is clearly desirable to prolong the recharging interval while supporting the desired quality of service for devices with

rechargeable batteries, e.g., mobile phone. Energy efficient design plays an even more important role in WSNs since if one or more nodes fail due to lack of energy, the WSN may not sustain normal functionality.

One potential approach for addressing these issues is the dynamic wireless resource allocation. The basic idea of resource allocation is to adapt the link transmission scheme to improve the system performance. This is achieved by power control, data rate adaptation and subcarrier allocation, based on the channel state information (CSI), system state information and service characteristics that are available at the transmitter. Goldsmith *et al.* [26] show that adaptation can obtain up to 20 dB power saving over non-adaptive transmission for a single communication link. In multi-user systems, significant system performance gain can also be achieved by optimal transmission scheduling and power control. For instance, in a multi-access channel, Knopp *et al.* [38] proposed a water-filling based power control algorithm to maximize the sum-of-rate capacity subject to the average transmission power constraint of each user.

In the rest of this chapter, a more complete literature review of resource allocation strategies for infrastructure-based wireless networks, wireless sensor networks, and multi-hop wireless networks is given. New resource allocation methods are proposed and compared with others that currently exist or are suggested in thesis chapters, respectively.

1.1 Resource Allocation in Wireless Networks

1.1.1 Infrastructure-based Wireless Networks

All communication in an infrastructure-based wireless network is via a centralized controller (e.g., base station or access point) through single-hop routing. The centralized controller takes charge of channel estimation and resource allocation.

The most prevalent infrastructure-based wireless networks today are cellular systems. A cellular network is a radio network consisting of a number of radio cells. Each cell is served by a base station that directly communicates with mobiles.

Traditional investigations on wireless resource allocation pay much attention to hard real-time services. Therein, the goal is to smooth out channel variation and build "bit pipes" that deliver data at a fixed rate, e.g., [19] and [74].

The rapid growth of the Internet has led to an increasing demand for supporting transmissions of best-effort service in wireless systems. These applications allow variable-rate transmission and are tolerant of high rate oscillations. Therefore, opportunistic communications [67] have been introduced to achieve higher system throughput. The concept of opportunistic communications is essentially to transmit more information in good channel states and less in poor ones. Hard real-time service and best-effort service may be viewed as two extremes of rate-oscillation sensitivity. However, services such as audio and video applications generally expect a balance between average rate and rate oscillation. If constant-rate transmission algorithms are used, the transmission efficiency would be very low. On the other hand, opportunistic scheduling schemes, such as [32] and [62], whose objective is to maximize a utility based on average rates, can improve efficiency in terms of average rate but result in high oscillation in instantaneous transmission rates.

1.1.2 Wireless Sensor Networks (WSNs)

Nowadays, wireless sensor networks have enormous potential for situation awareness applications. A WSN usually consists of a set of sensor nodes deployed in a certain region and performs distributed detection or estimation of application-specific information. In the context of distributed data collection, each node monitors its surrounding area, collects the data, and transmits it to a fusion center (FC). The FC then makes the final estimation using the collected data. A typical low cost sensor node has only limited processing capacity and often powered by small batteries. In many situations, battery recharging is impossible. If one or more nodes fail due to lack of energy, the sensor network may not sustain normal functionality. The lifetime of the network thus often refers to the time it takes for the first node in the network to die. Prolonging the network lifetime while maintaining a reasonably low computational cost has become the major challenge in designing compression strategies and communication protocols for WSNs.

The classical best linear unbiased estimator (BLUE) is designed to enhance the estimation accuracy by linearly combining the real-valued sensor observations [36]. However, it cannot be directly used in bandwidth-limited wireless network where real-valued message transmission is unavailable. Therefore, several local message functions which depend on the underlying sensor observation quality are designed to reduce communication from sensor nodes to FC, e.g, [12; 71; 70]. These compression schemes are proposed from the point of view of signal processing and do not take the channel condition variation into consideration.

Energy efficient transmission strategies for single-hop sensor data collection have recently attracted attention. The concept of most of transmission strategies has been to adapt the transmission parameters, such as transmission power, time and bandwidth, to the underlying channel gain, interference, and system preferences. In

fundamental communication theory, the transmission rate is an increasing and concave function of the transmission power, and the total amount of energy required for transmitting a given amount of data bits can be reduced by lowering the transmission power. This energy saving is achieved at the expense of increased transmission delay. Making use of this delay-energy tradeoff, Yao *et al.* proposed an energy-efficient transmission scheduling scheme in [73]. Given that each sensor node has a fixed number of bits to transmit, the total energy consumption is minimized by varying the transmission times assigned to different sensors.

The integration of signal processing and transmission is shown to lead to a more efficient and fair use of limited energy in [76]. An optimal power control policy for such integrated design is proposed by Xiao *et al.* in [69]. This policy minimizes the energy consumption by varying the transmission power and quantization level. It is suggested that the sensors with better channel condition and/or good observation quality should increase their quantization resolution. It is however important to point out that maximizing energy efficiency does not necessarily lead to network lifetime maximization. In both [73] and [69], minimizing the total energy consumption or some variants of it may result in some nodes running out of energy quickly.

1.1.3 OFDM-Based Multi-hop Relay Networks

Relay networks in the form of point-to-multipoint based tree-type or multipoint-to-multipoint mesh-type architectures are a promising network topology in future wireless systems. The basic concept of relaying is to allow a source node to communicate with a destination node under the help of a single or multiple relay nodes. It has been shown that relaying can bring a wireless network various benefits including coverage extension, throughput and system capacity enhancement. Recently, multi-hop relaying has been widely adopted in wireless networks such as next generation

cellular networks, broadband wireless metropolitan area networks and wireless local area networks. On the other hand, orthogonal frequency division multiplexing (OFDM) is an efficient physical layer modulation technique for broadband wireless transmission. It divides the broadband wireless channel into a set of orthogonal narrowband subcarriers and hence eliminates the inter-symbol interference. OFDM is one of the dominating transmission techniques in many wireless systems, e.g., IEEE 802.16 (WiMax), EV-DO Revision C and the Long-Term-Evolution (LTE) of UMTS. The combination of OFDM and multi-hop relaying has received a lot of attention recently. For example, this OFDM-based relay architecture has been proposed by the current wireless standard IEEE 802.16j [52]. The complexity of relay station is expected to be much less than the one of legacy IEEE 802.16 base stations, thereby reducing infrastructure deployment cost and improving the economic viability of IEEE 802.16 systems [1].

Recently, a large amount of effort has been directed towards ad hoc networks, which can be viewed as generalized relay networks, with each node in the network being able to communicate with any other node. Gupta *et al.* studied the bound of transmission rates in an asymptotic sense with a large number of hops under various network topologies and node capabilities in [30] and [31]. However, the asymptotic results on ad hoc networks do not apply in a network with a small number of relays. Previous works on resource allocation for relay networks are found in [61; 50; 55; 41; 15]. Authors in [61] and [55] studied efficient scheduling and routing schemes in one-dimensional multi-hop wireless networks, where it is assumed that the point-to-point links are frequency-flat fading channels. In [50], Oyman *et al.* introduced two different transmission strategies over multiple hops, and showed merits of multi-hop relaying in cellular mesh networks. Two-dimensional multi-hop networks are investigated in [41] and [15]. In [15], selective orthogonal frequency division multiple access (OFDMA) relaying is proposed in a network where multiple relay nodes are available at each

hop. The optimal source/relay/subcarrier allocation for OFDMA relay networks with fairness constraints is studied in [41], where cooperative transmission among source nodes and relay nodes is assumed. However, both selective OFDMA relaying and cooperative relaying require precise timing and phase synchronization among different nodes, and hence are difficult to be integrated in practical systems.

1.2 Design Approaches: Optimization for Wireless Networks

Optimization methods have been used widely in the design and analysis of wireless networks since last two decades. The most straight-forward understanding of optimization for wireless networks is that the design and analysis of wireless networks can be formulated as a mathematical optimization problem. The optimization problem could be maximizing a utility function, or minimizing a cost over a set of variables under a set of constraints.

1.2.1 Layered Design

Traditional mathematical optimization in wireless networks typically follows layered or modularized approach. In this approach, a networking system is divided into layers. Each layer makes autonomous decisions for achieving its own objective. The layered design approach hides the complexity of one layer from the others and is intuitively considered enabling a scalable and implementable network design. A classic paradigm is OSI seven layer model [77].

1.2.2 Cross-Layer Design

Recently, there has been increased interest in cross-layer design [60] in an effort to improve efficiency and/or fairness in allocation of network resource. The idea of cross-layer design is mainly motivated by the time-varying characteristics of wireless channel, network conditions and the emergence of differential applications. Rate, power and other resources at the physical layer can be dynamically adjusted to meet the quality of service (QoS) of these applications given the current channel and network conditions. To implement it, information must be shared between layers to obtain the highest possible adaptivity [5]. The gains of cross-layer design are particularly shown for Transmission Control Protocol (TCP) traffic over wireless links [38; 65; 64; 4; 62].

Although cross-layer design approach brings great enhancement to the network performance, it may lead to various negative consequences as pointed out in [35]. For instance, cross-layer design can create loops, thus, stability and robustness become paramount issues. In addition, unbridled cross-layer design can also lead to a "spaghetti design".

1.2.3 Layering as Optimization Decomposition

It is illustrated in [14] "layering as optimization decomposition" provides a more unified framework for network design. Chiang *et at.* in Page 255 of [14] point out

"the overall communication network is modelled by a generalized network utility maximization problem, each layer corresponds to a decomposed subproblem, and the interfaces among layers are quantified as functions of the optimization variables coordinating the subproblems."

"Network as an optimizer" and "layering as decomposition" are two key concepts

behind "layering as optimization decomposition". The former emphasizes viewing protocols as a distributed solution to some global optimization problem. The latter indicates the problem itself does not have any predetermined layering architecture, but the optimal solution automatically established the benchmark for all layering schemes through problem decomposition.

Most of the work inspired by "layering as optimization decomposition" focuses on media access control (MAC) layer (e.g.,[10; 44]) , network layer (e.g., [29]), and TCP layer (e.g.,[34; 37; 46]).

Although most of the work in this thesis had been done or started before [14] was published, the idea of "layering as optimization decomposition", putting the end-user application needs as the optimization objective, establishing the globally optimal performance benchmark and design modularized and/or distributed solution through decomposition, runs through the whole thesis.

1.2.4 Convex Optimization

Difference among resource allocation schemes for the networks with different architectures, namely, infrastructure-based wireless networks, wireless sensor networks, and multi-hop wireless networks, arises from different types of traffic that they support and different system preferences. Due to the different limitations of their network architectures, the centralized or partially distributed algorithm is needed. However, these resource allocation schemes are common in the sense that most of them could be cast as or converted into convex optimization problems [7]. Convex optimization solves the problem of minimization of a convex objective function subject to convex constraints. It plays an important role in engineering application because a local optimum is also a global optimum in a convex problem and this optimal solution often

reveals design insights [47]. More detail in convex optimization will be introduced in Chapter 2.

1.3 Objectives and Contributions

The primary objective of this thesis is to provide a generalized optimization framework with different approaches for dynamic resource allocation for energy-constrained wireless networks and provide solutions to some specific networks. The proposed research can be viewed as a combination of multiple disciplines, including signal processing, information theory, optimization, wireless communication theory and networking to address the questions stated in Section 1.3.1, 1.3.2 and 1.3.3, where we have also discussed motivations and contributions.

1.3.1 Problem 1: Joint Optimization of Average Rate and Rate Oscillation in Variable-Rate Wireless Networks

As demonstrated in Section 1.1.1, constant-rate transmission and opportunistic scheduling schemes may be viewed as two extremes in terms of rate-oscillation sensitivity: constant-rate transmission algorithms result in low transmission efficiency, while opportunistic scheduling schemes can improve efficiency in terms of average rate but result in high oscillation in transmission rates (throughout this thesis, we will use the term "transmission rate" to refer to the instantaneous transmission rate in a time frame). As is known, many services such as wireless audio and video transmission generally expect a balance between average rate and rate oscillation, both of which are important performance metrics for these applications. This motivates the search for transmission schemes that can joint optimize average rate and rate oscillation by dynamically adapting transmission rate.

The contributions of the work for this problem can be summarized as follows.

- We find a new criterion for multi-user scheduling by modifying the utility function in a way that penalizes rate oscillation and rewards average rate. Rate oscillation here is measured using a statistical rate variance.

- We demonstrate later that a utility function that increases with average rate but decreases with rate variance can be used to facilitate the choice of the combinations of average rate and rate oscillation.

- A gradient based scheduling algorithm is developed to maximize the proposed utility function. The proposed algorithm reduces to the traditional gradient algorithm in [2] and [62] when we omit the rate variance term in the new utility function so that the utility is a function of the average rate only.

- The dynamics of transmission rate is analyzed using ordinary differential equation. In addition, the condition under which the proposed algorithm is asymptotically optimal is addressed in Chapter 3.

- As a practical example, the proposed gradient based scheduling algorithm applied in time-sharing wireless networks are studied for two cases: (a) perfect CSI and (b) limited channel feedback. Numerical results show how the average rate and rate variance are balanced and the convergence performance of the algorithm.

The above results will be discussed in detail in Chapter 3.

1.3.2 Problem 2: Integrated Designs of Quantization and Transmission for Lifetime Maximization in Wireless Sensor Networks

Signal processing and transmission scheme in wireless sensor network were separately designed in most of existing works. For example, Xiao *et. al.* in [70] and Yao *et. al.* in [73] consider signal processing and transmission scheme in energy efficient wireless sensor network, respectively. How to maximize the network lifetime by varying both transmission parameters and quantization resolution for a single-hop sensor network is still an open issue. Rate-power curve can be viewed as the interface between quantization design and transmission design. Several fundamental features of sensor networks are taken into account in the formulation.

- First, for a sensor network deployed for decentralized estimation, it emphasizes estimation accuracy achieved at the FC more than the total or individual transmission rate and, thus, is taken as one of our optimization constraints.
- Secondly, besides observation quality and channel condition, initial energy is also critical to lifetime maximization design because the information of initial energy helps to balance the energy consumption by giving high priority to the nodes with high initial energy level. Thus our lifetime maximization strategy also considers the available energy left in each sensor.
- Lastly, since centralized algorithms often require large computational cost and significant control signaling overhead, the proposed algorithm is designed to be of partially distributed nature and can be implemented easily in practical systems. Each node only needs to know its local information resulting in very little common information broadcast by the FC. The convergence behavior of the algorithm is also explored.

The above result under two typical scenarios, (a) uncorrelated source observation and (b) common source observation, will be discussed in Chapter 4.

1.3.3 Problem 3: End-to-End Outage Minimization and Average Rate Maximization in Linear OFDM Based Relay Networks

We move to the context of relay networks after addressing the resource allocation in single-hop networks. Relay networks have the potential to expand coverage and enhance throughput. Similarly as in single-hop network, for many real-time services, one has to consider keeping the target transmission rate and avoiding outage in most fading scenarios through dynamic resource allocation. Whereas, non-real-time services expect high average rate transmission. How to minimize end-to-end outage probability and maximize end-to-end average rate in an OFDM-based multi-hop wireless network is yet under-explored. In a linear relay network where no data is allowed to accumulate at any relay nodes, an end-to-end outage is the event that there exists a hop on which transmission rate is lower than the target rate. In this thesis, we study the subcarrier, power and time allocation to minimize the end-to-end outage probability and maximize the end-to-end average rate, respectively, in a one-dimensional multi-hop network under an average transmission power constraint.

The novelty and contributions of the work done for this problem can be summarized as follows:

- The first problem (minimization of end-to-end outage probability) is solved by decomposing into two subproblems; (a) Derive the minimum short-term power required to meet a target transmission rate for any given channel realization. The resulting power and time allocation is obtained through a Two-nested Bi-

nary Search (TBS) which is conducted in a central controller with the knowledge of channel state information (CSI) on all subcarriers and over all hops; (b) Determine the transmission on-off by comparing the required minimum total power with a threshold.

- The second problem (maximization the end-to-end average rate) is first formulated as a max-min problem. Then it is solved by decomposing it into two subproblems: (a) Determine the power and time allocation to maximize the end-to-end instantaneous transmission rate under a given total power constraint for each channel realization; and (b) Determine the instantaneous total power constraint for each channel state so that the end-to-end average transmission rate is maximized under a long-term total power constraint.

- These optimal allocation schemes determine the performance limitation, but also incur high computational complexity and high signaling overhead. Several suboptimal algorithms with low complexity and reduced overhead are proposed. The tradeoff between performance and complexity and overhead is also discussed.

The above results will be discussed in detail in Chapter 4 and 6.

1.4 Organization of Thesis

This thesis consists of 7 chapters including the present Introduction chapter. Chapter 2 presents basic system models, fundamental concepts such as performance metrics and constraints, and design approaches that will be used throughout the thesis.

We first study the dynamic resource allocation in the context of single-hop wire-

less network. In Chapter 3, we consider jointly optimization of average rate and rate oscillation in a multi-user system over a time-varying wireless fading channel. A generalized gradient algorithm is developed to maximize a proposed utility function from a myopic view of the optimization problem. Chapter 4 considers the lifetime maximization for a cluster-based WSN. Integrated designs of quantization and transmission are investigated, where the decision of transmission power, time and quantization resolution may depend on the information of observation quality, channel condition and initial energy. It is also demonstrated that the optimal decision can be implemented in a partially distributed way.

We then turn to end-to-end resource allocation in a multi-hop relaying network. In Chapter 5, our goal is to maximize the end-to-end average transmission rate in an OFDM based multi-hop linear network. We derive the optimal transmission power on each subcarrier over each hop and the transmission time used by each hop in every time frame under a long-term total power constraint. Minimizing end-to-end outage probability is examined in Chapter 6. In the first step, we derive the minimum short-term power required to meet a target transmission rate for any given channel realization. In the second step, the transmission on-off is determined by comparing the required minimum total power with a threshold. We also propose suboptimal algorithms with low complexity and reduced overhead. The tradeoff between performance of these algorithms, and their complexity and overhead is discussed. Conclusions and future work are discussed in Chapter 7.

Chapter 2

Preliminaries

This chapter presents the basic background of wireless systems and mathematical optimization methods to provide a big picture on performance optimization for wireless network. In Section 2.1, wireless channel model, single-hop multiuser wireless system and multi-hop wireless system are introduced, respectively. In Section 2.2, we define performance measures in wireless networks, including bit error rate, transmission rate, outage probability, utility. Finally, introduction to convex optimization theory and optimization of functionals with integral constraints is given in Section 2.4 and 2.5.

2.1 System Models

2.1.1 Wireless Channel Model

The fundamental difference of a wireless channel from wireline lies in its time-varying characteristics. The radio signal transmitted through a wireless channel suffers from attenuation mainly arising from path loss, shadowing and fading.

The instantaneous channel gain between transmitter and receiver is modelled using [56]

$$g = d^{-\alpha} 10^{0.1\varsigma} X^2 \tag{2.1}$$

where d is the distance between transmitter and receiver, α is the path loss exponent, ς is a zero-mean Gaussian distributed random variable (in dB), $10^{0.1\varsigma}$ is the log-normal shadowing, and X represents small-scale fading envelope.

Path Loss

Path loss is the reduction in power density (attenuation) of an electromagnetic wave as it propagates through space. The attenuation depends on the distance between the transmitter and receiver. In wireless systems, path loss can be represented by the path loss exponent, whose value is normally in the range of 2 to 4 (where 2 is for propagation in free space, 4 is for relatively lossy environments and for the case of full specular reflection from the earth surface).

Shadowing

Other than path loss effect, the average received signal may experience random shadowing effects due to different levels of clutter, e.g., tree and building, on the propagation path between transmitter and receiver. The measured signal levels (in dB) at a specific transmitter and receiver pair follows Gaussian distribution.

Path loss and shadowing belong to large-scale propagation models since they predict the mean signal level over large distance. The values of α and the variance of ς can be computed from measured data, using linear regression. For example, in Stanford University Interim (SUI)-3 channel model with a central frequency at around 1.9 GHz to simulate the fixed broadband wireless access channel environments [17], intermediate path loss condition ([16, Category B]) is modelled by

$$PL = A + \alpha \lg \left(\frac{d}{d_n} \right),$$

where $A = 20\lg(4\pi d_0/\lambda)$ (λ being the wavelength in m), α is the path-loss exponent with $\alpha = (a - bh_b + c/h_b)$. Here $h_b = 30m$ is chosen as the height of the base station, $d_0 = 100m$ and a, b, c are 4, 0.0065 and 17.1, respectively, given in [16]. This model will be used in Chapter 5 and 6.

Fading

Fading (or small-scale) model is used to characterize the rapid fluctuations of the received signal level over short travel distances or short time durations. Small-scale fading mainly arises from the combination of multiple replicas of the transmitted signals having different amplitudes, phases and angles of arrival. In the present of a specular (line-of-sight) component, small-scale fading is commonly modelled by the Ricean probability density function (pdf) [57]

$$pdf(x) = \begin{cases} \frac{x}{\sigma^2} \mathrm{e}^{-\frac{x^2+A^2}{2\sigma^2}} I_0\left(\frac{Ax}{\sigma^2}\right) & A \geq 0, x \geq 0 \\ 0 & x < 0 \end{cases}, \tag{2.2}$$

where σ^2 is the time-average power of the received signal, A represent the peak amplitude of the dominant signal and $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind. The Ricean distribution is characterized in parameter $K := A^2/(2\sigma^2)$. When K goes to 0, that is, the specular component diminishes to zero, the Ricean distribution degenerates to a Rayleigh distribution with pdf

$$pdf(x) = \begin{cases} \frac{x}{\sigma^2} \mathrm{e}^{-\frac{x^2}{2\sigma^2}} & x \geq 0 \\ 0 & x < 0 \end{cases}. \tag{2.3}$$

Small-scale fading can be classified based on multipath time delay spread into *flat fading* and *frequency selective fading*. The received signal is said to experience flat fading if the wireless channel has a constant gain and linear phase response over a bandwidth which is greater than the signal bandwidth. On the other hand, the channel is under frequency selective fading if the channel possesses a constant-gain

and linear phase response over a bandwidth that is smaller than the bandwidth of transmitted signal.

Orthogonal frequency division multiplexing (OFDM) is one of the most popular schemes for broadband wireless networks to overcome inter-symbol-interferences caused by multipath propagations in a frequency selective channel. The broadband channel is divided into a number of equally spaced frequency bands, each carrying a portion of the user information.

2.1.2 Single-Hop Multiuser Wireless Systems

Single-hop multiuser wireless system is a simple network model consisting of N users communicating with a common centralized controller through a same channel (Fig. 2.1-a). This model can be used to describe the signal-cell wireless system, satellite system and WLAN. The corresponding centralized controller corresponds to base station, satellite and access point, respectively. The users share the same channel by different multiple access techniques, such as time division multiple access (TDMA), frequency division multiple access (FDMA) and code division multiple access (CDMA). The broadcast and multiple access channels are used to model two-way transmission.

In multiuser systems, the received signals from (or to) different users may experience different attenuation. By adaptively scheduling the users, and/or dynamically assigning resources such as transmission power, subcarriers, we can take advantage of this channel diversity, which is called multiuser diversity. The theory of multiuser diversity has been applied in practice. Take Qualcomm's HDR (High Data Rate) system (1xEV-DO) downlink case for example. HDR downlink transmission operates on a time-division basis, and scheduler decides which user to be served in each time

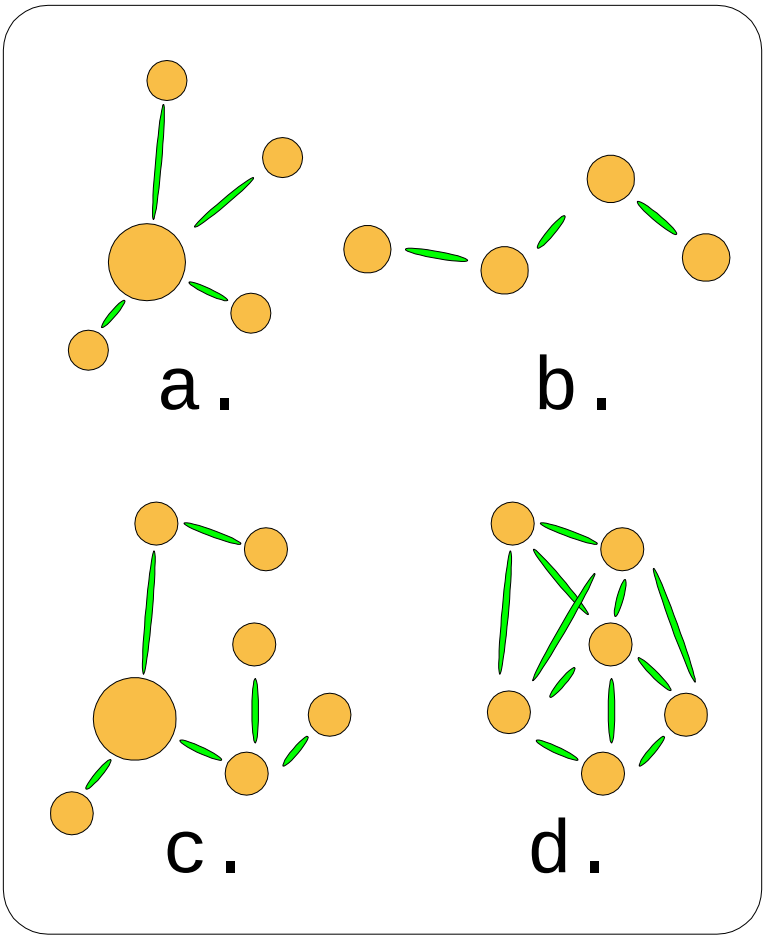

Figure 2.1. Network architecture a. single hop network b. linear multiple hop network c. PMP network d. mesh network

slot. The diversity gain is exploited by scheduling the user with best instantaneous channel condition [32].

2.1.3 Multi-hop Wireless Systems

Deployments of multi-hop relays have the potential to exploit various benefits, such as expanding coverage and enhancing throughput and system capacity since they may shorten the transmission distance and provide the opportunity of frequency reuse.

Point-to-multi-point (PMP) tree and mesh networks are two of the most promising

topologies for future multi-hop wireless networks. For example, these two topologies has been proposed by current wireless standard IEEE 802.16j [52]. Under these two architectures, a source is allowed to communicate with a destination with the help of multiple relaying nodes. PMP network typically has a carrier-owned infrastructure. One end of the path in PMP networks is the base station. Linear multiple hop network (Fig. 2.1-b) is one specific case of PMP networks (Fig. 2.1-c), consisting of a one-dimensional chain of nodes including a source, a destination, and multiple relays. The complexity of relay station is expected to be much less than that of legacy IEEE 802.16 base stations, thereby reducing infrastructure deployment cost and improving the economic viability of IEEE 802.16 systems [1].

In mesh networks, routing is controlled by subscriber equipment and there may be multiple connections between two users. Fig. 2.1-d gives an mesh network layout. A wireless mesh network example is a mini wireless mesh router launched by US-based firm Meraki in early 2007 [49].

2.2 Performance Measures

2.2.1 Bit Error Rate (BER)

Bit error rate is the percentage of bits that have error relative to the total transmitted bits. It measures the reliability of point-to-point communication. The derivation of a closed-form expression of BER is generally difficult except for some specific cases. It has been shown in [18] that, for uncoded M-QAM, the relation between BER and the received signal-to-noise ratio (SNR) and the number of M-QAM is approximately given by

$$\text{BER} \approx 2\text{exp}\left(\frac{-1.5\gamma}{M-1}\right), \tag{2.4}$$

where M is the number of M-QAM constellation points and γ is the average received SNR. Since the number of data bits per symbol is $\log_2 M$, and the bandwidth is equal to the inverse of the duration of each M-QAM symbol, the transmission rate, or say, spectral efficiency is also $\log_2 M$. Equation 2.4 can be used to derive SNR gap expression in the next subsection.

2.2.2 Transmission Rate

Innovations in the physical layer, such as better modulation and coding schemes does not only help to reduce in BERs for a fixed spectral efficiency and SNR, which the users do not directly observe, but also improve transmission rate, for a fixed BER requirement, which the user can more directly observe.

The *instantaneous transmission rate* in the absence of other users' interference depends on the allowable BER, and can be expressed as [54]

$$r = \log_2 \left(1 + \frac{p(g)g}{\Gamma N_0}\right), \tag{2.5}$$

where N_0 is the noise power, $p(g)$ is power allocation scheme according to channel state information g and Γ is the SNR gap [54]. Equation (2.5) gives a generalized rate-power curve. When instantaneous mutual information is used to characterize the achievable transmission rate, we have $\Gamma = 1$ (0dB). In this case, r is the maximum possible information transfer rate per unit bandwidth (in bit/s/Hz) with reliable transmission over a channel, subject to specified constraints. If practical signal constellations are used, Γ is a constant related to a given BER constraint. For example, when uncoded M-QAM constellation is used, we have $\Gamma = -\ln(5 \cdot \text{BER})/1.5$, which can be derived from (2.4).

For best effort traffics, users expect a high average rate and allow a long delay. When channel statistics are assumed to be fixed, and the codeword length can be

chosen arbitrarily long to average over the fading of the channel, the long-term rates, averaged over the fading process

$$\bar{r} = \mathbb{E}(r) \tag{2.6}$$

can be achieved, where $\mathbb{E}(\cdot)$ represents the expectation over the distribution of channel realization. The definition of average rate is relevant for situations when the delay requirement of the users is much longer than the time scale of the channel fading.

2.2.3 Outage Probability

When the delay requirement is shorter than the time scale of channel variations, which occurs in many real-time services, one has to consider maintaining the target transmission rate and avoiding outage in most of fading conditions through dynamic resource allocation. An outage is an event that the actual transmission rate is below a prescribed transmission rate ([11] and [43]). *Outage probability* can be viewed as the fraction of time that a codeword is decoded wrongly. For a given finite average power constraint, it may not be possible to achieve the target rate all the time. Thus, transmission outage is inevitable under severe fading condition. Mathematically, the outage probability is given by

$$P^{out} \triangleq Prob(r < R). \tag{2.7}$$

where R is the target transmission rate.

The minimum outage probability problem can be generally solved in two steps as proposed in [11]. First, for each channel state, the short-term minimum resource, e.g., power, required to guarantee the target end-to-end transmission rate R is to be determined. The second step then determines a threshold to control the transmission on-off subject to certain constraints, e.g., average power constraint.

2.2.4 Utility

Utility is used to indicate the level of satisfaction of the decision maker as a result of its actions and is always used in the area of economics. The formal definition of utility function is given by [21],

Definition 1: A function that assigns a numerical value to the elements of the action set $\mathcal{A}$ ($u : \mathcal{A} \rightarrow \mathbb{R}$) is a utility function, if for all $x, y \in \mathcal{A}$, x is at least as preferred compared to y if and only if $u(x) \geq u(y)$.

The utility function used to guarantee quality of service (QoS) can be understood from two points of view, reverse-engineering and foward-engineering [14]. In reverse-engineering, the given protocols implicitly determine the form of utility function. For example, the improvement in the version of TCP in the FAST Project is enlightened by insights from reverse-engineering TCP. In forward-engineering, utility function typically provide a metric to define optimality of resource allocation. For instance, in the wireless data scheduling context, the proportionally fair (PF)[1] scheduling algorithm is optimal when utility function is defined as $\sum_i \log(\bar{r}_i)$, where $\bar{r}_i$ is the average rate of user i. The objective function in network protocol design could be sum of utility functions of rate, delay, power, etc., of end users. It also could be coupled across the users, e.g., network lifetime defined in Chapter 4, and end-to-end outage probability defined in Chapter 6).

2.3 Constraints

While utility, as defined above, helps to construct objective function, one still needs to find reasonable constraints for problem formulation.

[1]In PF algorithm, at each time slot t, the scheduler serves the user n for which $r_n(t)/\bar{r}_n(t)$ is maximal, where $r_n(t)$ is instantaneous rate of user n at time slot t and $\bar{r}_n(t)$ is exponentially smoothed average service rates (defined in Chapter 3.

2.3.1 Physical Constraints

Time constraint

In most of literatures, time frames are used to denote the shortest time periods within which the channel conditions remain unchanged. This assumption is valid under the condition that the actions in the network take place according to a common timer. We will use this assumption in this thesis. If it is assumed that each time frame can be accessed by all the N users in an adaptive time-sharing fashion, the time constraint is given by

$$\sum_{i=1}^{N} \rho_i = 1. \tag{2.8}$$

where $\boldsymbol{\rho} = [\rho_1, \rho_2, \ldots, \rho_N]$ denotes the time-sharing adaptation policy, and ρ_i represents the fraction of the frame duration allocated to user i.

Power constraint

Two types of power constraint, *peak power constraint* and *average power constraint*, are common in wireless system. Peak power must be contained within fixed bounds in every time slot regardless of previous transmissions. Such a constraint is realistic in cases when the electronics driving wireless transmitters must be operated within a certain power range. In infrastructure-based wireless network, uplink and downlink transmission have different *peak power constraints*. For the uplink scenario, the terminal of each user is power limited, while the total power provided by the base station is limited for the downlink. In a wireless sensor network where nodes are deployed with a fixed amount of energy E_i, the network is guaranteed to last for E_i/p_i units of time, where p_i is the transmission power of node i used in each time unit. Many literatures also consider the *average power constraint*, which allows energy to be stored and used later to either extend network lifetime or enable more powerful future transmissions. The assumption of average power constraint can be found in [11; 42; 43; 20; 66].

2.3.2 Hard QoS Constraints

In contrast to elastic QoS measure which may be treated as objective functions, some hard, inelastic QoS measures must be treated as inflexible constraints. Two specific hard QoS constraints are the following:

Target rate

In transmission of hard real time traffic, a constant transmission rate must be kept with probability one regardless of channel condition. when the real transmission rate is lower than the target rate, an outage occurs. When we face the resource allocation problem in a heterogeneous multiuser system supporting both best effort traffic and hard real time traffic, maintaining target rates for hard real time traffic is necessary. An example can be found in [63].

Estimation accuracy

For a sensor network deployed for decentralized estimation, the estimation accuracy achieved at the FC is usually more important than the total or individual transmission rate from the perspective of WSN design and, thus, is taken as one of the optimization constraints when network lifetime is chosen as the objective function.

2.4 Convex Optimization

In the last two decades, the gradient descent and the least square algorithms were two typical algorithms used in solving optimization problem. Nevertheless, they suffer from slow convergence and sensitivity to initial point and step size. One powerful way to avoid these problems is to derive a convex reformulation or a convex relaxation of the original non-convex formulation. Optimization based approaches have been ubiquitously used in communications and signal processing [47; 45; 53]. Many resource

allocation problems can be investigated using convex optimization approaches. This section introduces the basic convex optimization concepts.

2.4.1 Convex Optimization Problems

A generic optimization problem can be expressed as follows

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m, \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, r \\ & \mathbf{x} \in \mathcal{S} \end{aligned} \tag{2.9}$$

where f_0 is called objective function and $\mathbf{x}$ is called optimization variable. The optimization problem is said to be convex when the functions $f_1, \ldots, f_m$ are convex in $\mathbf{x} \in \mathbb{R}^N$, $h_1, \ldots, h_m$ are affine function and the set $\mathcal{S}$ is convex. If we change "min" to "max" and the inequalities from "$f_i \leq 0$" to "$f_i \geq 0$", then it is still a convex optimization problem if and only if all f_i are concave.

Convexity is important in optimization problems because a local optimum is also a global optimum in a convex problem. Consequently, when a design problem is converted to a convex form, the same optimal solution is achieved. Furthermore, many powerful numerical algorithms exist to solve for the optimal solution of convex problem efficiently.

2.4.2 Lagrangian Duality and Karush-Kuhn-Tucker Condition

Consider problem (2.9), and call it primal optimization problem, and $\mathbf{x}$ the primal vector. We define the Lagrangian function as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) := f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^{r} \nu_j h_j(\mathbf{x}).$$

The vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are called the dual variables associated with problem (2.9). The dual function is defined as

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) := \min_{\mathbf{x} \in \mathcal{S}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}).$$

The corresponding dual optimization problem is

$$\begin{aligned} \max \quad & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \geq 0, \boldsymbol{\nu} \in \mathbb{R}^r. \end{aligned}$$

If the global minimum value of primal problem (2.9) is $\mathbf{f}^*$ and the one of dual problem (2.10) is $\mathbf{g}^*$, then we have $\mathbf{f}^* \geq \mathbf{g}^*$. Usually, we have strong duality in the sense that $\mathbf{f}^* = \mathbf{g}^*$. When the primal problem is convex, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for the primal and dual optimization.

When the primal problem has a coupling constraint such that, when relaxed, the optimization problem decouples into several subproblems, which independently decide the amount of resource to be allocated according to a given resource price. This method is called dual decomposition, which will be used in Chapter 4.

2.5 Optimization of Functionals with Integral Constraints

This section examines a mathematical optimization problem—that of optimizing a function of an unknown function [28]. It is a useful tool to derive the relation between optimal resource allocation and channel condition and/or system preference under time-average type of constraints, e.g., time-average power constraint.

Suppose that we seek some function $u(t)$ which will cause the integral of a known function $\Phi\{u(t), t\}$ to be optimized. This integral, which is a function of an independent variable t and the unknown dependent variable $u(t)$, is known as a functional. The problem of optimizing a functional is often referred to as a variational problem.

Considering the optimization of an integral involving only one unknown function, let

$$I = \max(\min) \int_{t_0}^{t_f} \Phi(u, t) \mathrm{d}t, \tag{2.10}$$

Subject to an integral constraint of the form

$$G = \int_{t_0}^{t_f} \Psi(u, t) \mathrm{d}t = 0. \tag{2.11}$$

where u is an unknown function of t, which we shall assume to be continuous and to have a continuous second derivative with respect to t. Let us assert that the boundary conditions are known; i.e., $u(t_0) = u_0,\ u(t_f) = u_f$. We seek the particular function $u(t)$ that satisfies the boundary conditions and optimizes the above integral. Through the introduction of a Lagrange multiplier, we form the augmented integral

$$\begin{aligned} L &= I + G & (2.12) \\ &= \int_{t_0}^{t_f} \left[\Phi(u, t) + \lambda \Psi(u, t)\right] \mathrm{d}\mathbf{g}, & (2.13) \end{aligned}$$

where λ , the Lagrange multiplier, is an undetermined constant. Application of the Euler-Lagrange equation results in the necessary condition for an optimum to exist,

$$\frac{\partial \Phi}{\partial u} + \lambda \frac{\partial \Psi}{\partial u} = 0, \tag{2.14}$$

where λ can be determined by (2.11).

Chapter 3

Joint Optimization of Average Rate and Rate Oscillation for Variable-Rate Wireless Networks

In this chapter, we consider joint optimization of average rate and rate oscillation in a multi-user system over a time-varying wireless fading channel. Multiple users are served by a common base station. In each time frame a scheduler determines the current transmission parameters of all users, such as power and rate, according to current channel state information and long-term system preferences.

Many existing scheduling and resource allocation schemes are formulated as a utility-based throughput allocation problem, such as [38; 32; 2; 62]. In [38], the scheme is in essence restricted to a linear utility function. It is shown that the sum capacity of a wireless system is maximized when only the user with the maximum achievable rate is chosen for transmission. However, this scheme could result in significant unfairness among users with asymmetric channels. In order to address the fairness concerns, the *proportionally fair* (PF) scheduling algorithm is proposed

in [32] and [68] by using a log-form utility function of average transmission rate. The PF scheduling is then generalized by Agrawal *et al.* in [2] to a *gradient scheduling* algorithm, which applies to any increasing and concave utility function of average transmission rate. The gradient scheduling algorithm suggests that at any time instant, the best scheduler should maximize the sum of weighted transmission rates and the weight depends on system preferences and channel conditions. Although the gradient scheduling algorithm is developed from a myopic view of the optimization problem (maximizing the utility function of average transmission rate), Stolyar in [62] proves its asymptotic optimality for a general model, which allows for simultaneous transmission of multiple users and the set of scheduling decisions to be discrete (as in [32]).

The gradient algorithms in [32; 3; 62] can improve efficiency in terms of average transmission rate but result in high rate oscillation. This phenomenon motivates us to find the answer to the following question: how to joint optimize average transmission rate and rate oscillation? We note that if the utility function increases with average rate but decreases with rate variance, it can be used to facilitate the choice of the combinations of average rate and rate oscillation. A gradient based scheduling algorithm is developed to maximize the proposed utility function from a myopic view (in the sense that the difference between utilities in two consecutive time frames is maximized). The proposed algorithm degenerates to the traditional gradient scheduling algorithm (TGSA) in [2] and [62] when we omit the rate variance term in the new utility function so that the utility is a function of the average rate only. Thus, we refer to our algorithm as the *generalized gradient scheduling algorithm* (GGSA). It can be shown that at each time frame, the best scheduler of GGSA is not to maximize the sum of weighted transmission rates (as in [2] and [62]), but the sum of concave functions of transmission rates. This coincides with our idea in [75] that maximizing the sum of concave functions of transmission rates can obtain a balance between average

transmission rate and rate oscillation. Next, we analyze the dynamics of transmission rate using ordinary differential equation and show that GGSA is asymptotically optimal under the condition that the transmission rate vector, with appropriate scaling, converges to a fixed vector as time goes into infinity.

The rest of this chapter is organized as follows. In Section 3.1, the system and channel model are presented. Section 3.2 presents the TGSA. Section 3.3 derives the GGSA for variable rate transmission. We study the dynamics of average transmission rates and rate variances under GGSA in Section 3.4 and give an example in time-sharing wireless network in Section 3.5 . Numerical results are given in Section 3.6. Finally, Section 3.7 concludes the chapter.

3.1 System Model

Consider a wireless system with a base station (BS) and a finite set of users, denoted by $\mathcal{N} = \{1, 2, \ldots, N\}$. The communication link between each user and the base station is modelled as a time-varying fading channel. And the transmission is centrally coordinated by BS on a time frame basis. In time frame k, the channel state between BS and user i is denoted as $g_i(k) \in \mathcal{G}$, where $\mathcal{G}$ is the channel state space. Here we assume that $\{g_i(k), k = 1, 2, \ldots\}$ is ergodic and stationary with the distribution γ.

We consider the situation where each user always has data for transmission, and are interested in optimizing the pair of the average rates and the rate variances $(\mathbf{m}, \boldsymbol{\sigma}^2) := (m_1, \ldots, m_N, \sigma_1^2, \ldots, \sigma_N^2)$ so as to maximize a utility function $U(\mathbf{m}, \boldsymbol{\sigma}^2)$. Here, rate variance is chosen to quantify the rate oscillation. The utility function U represents the system preferences. Since the system expects a high average transmission rate and a low rate oscillation, the utility is assumed to be increasing in $\mathbf{m}$ but

decreasing in $\boldsymbol{\sigma}^2$. Mathematically, the problem can be expressed as

$$\begin{aligned} \max \quad & U(\mathbf{m}, \boldsymbol{\sigma}^2) \\ \text{s.t.} \quad & (\mathbf{m}, \boldsymbol{\sigma}^2) \in \mathcal{V}, \end{aligned} \tag{3.1}$$

where we further assume that U is continuously differentiable in $\mathbf{m}$ and $\boldsymbol{\sigma}^2$ but not necessarily concave. The steady-state average rate and rate variance (ARRV) region $\mathcal{V}$ is defined as

$$\begin{aligned} \mathcal{V} \;:=\; & \Big\{(\mathbf{m}, \boldsymbol{\sigma}^2) \in \mathbb{R}_+^{2N} : \exists \mathbf{r}(\mathbf{g}) \in \mathcal{R}(\mathbf{g}) \;\forall \mathbf{g} \in \mathcal{G} \\ & \text{such that } \mathbf{m} = \int_{\mathcal{G}} \mathbf{r}(\mathbf{g}) \gamma(\mathrm{d}\mathbf{g}), \text{and} \\ & \boldsymbol{\sigma}^2 = \int_{\mathcal{G}} \left[\mathbf{r}(\mathbf{g}) \times \mathbf{r}(\mathbf{g}) - \mathbf{m} \times \mathbf{m}\right] \; \gamma(\mathrm{d}\mathbf{g}) \Big\}, \end{aligned}$$

where $\mathbf{r}$ is the transmission rate vector, and $\mathcal{R}(\mathbf{g})$ is the achievable rate region when the network channel state is $\mathbf{g}$.

Let $X_i(k)$ and $Y_i(k)$ denote the average rate and rate variance, respectively, for user i up to the k-th time frame. If $r_i(k)$ is the selected transmission rate of user i at time frame k, then $X_i(k)$ and $Y_i(k)$ can be updated as:

$$\begin{aligned} X_i(k+1) &= \frac{1}{k} \sum_{l=1}^{k} r_i(l) \\ &= X_i(k) + \frac{1}{k}[r_i(k) - X_i(k)]; \\ Y_i(k+1) &= \frac{1}{k} \sum_{l=1}^{k} [r_i(l) - X_i(l)]^2 \\ &= Y_i(k) + \frac{1}{k} \left\{[r_i(k) - X_i(k)]^2 - Y_i(k)\right\}. \end{aligned}$$

If we use a small fixed step size μ to replace $1/k$, then the recursive forms can be rewritten as

$$X_i(k+1) = X_i(k) + \mu[r_i(k) - X_i(k)]; \tag{3.2}$$

$$Y_i(k+1) = Y_i(k) + \mu \left\{[r_i(k) - X_i(k)]^2 - Y_i(k)\right\}. \tag{3.3}$$

3.2 Traditional Gradient Scheduling Algorithm

Before moving to the proposed generalized gradient scheduling algorithm, let us review the traditional gradient scheduling algorithm. Considering a myopic view of the optimization problem, we chose $\mathbf{r}(\mathbf{k})$ to maximize $U(\mathbf{X}(k+1))$ given that $\mathbf{r}(0), \cdots, \mathbf{r}(k-1)$ have already been chosen, where $U(\cdot)$ is an increasing, strictly concave, and continuously differentiable utility function on $\mathbb{R}_+$ (this restriction is only valid in this section). Thus, it is key to find the next step action given whatever action was taken in the past. Since we have

$$\begin{aligned} & U\big(\mathbf{X}(k+1)\big) - U\big(\mathbf{X}(k)\big) \\ = \; & U\big(\mathbf{X}(k) + \mu[\mathbf{r}(k) - \mathbf{X}(k)]\big) - U\big(\mathbf{X}(k)\big) \\ \approx \; & \mu \nabla U(\mathbf{X}(k))^T \cdot [\mathbf{r}(k) - \mathbf{X}(k)]. \end{aligned}$$

The last approximation is true since μ is sufficiently small. Thus, the best choice given the past actions is to select a current instantaneous rate that satisfies

$$\mathbf{r}(k) = \arg \max_{\mathbf{v}(\mathbf{k}) \in \mathcal{R}(\mathbf{g})} \nabla_{\mathbf{m}} U(\mathbf{X}(k))^T \cdot \mathbf{v}(k). \tag{3.4}$$

Note that the utility function $U(\cdot)$ defined in this section is only a function of average rate. When all the users have the same derivative of U with respect to average rate, the best scheduler in TGSA reduces to maximum rate rule. Formally, we schedule at time k, the user

$$j = \arg \max_i r_i(k). \tag{3.5}$$

When the utility has a specific form as follows

$$U(\mathbf{m}) = \sum_i^N \log(m_i), \tag{3.6}$$

TGSA degenerates to PF algorithm, given by

$$j = \arg \max_i \frac{r_i(k)}{X_i(k)}. \tag{3.7}$$

3.3 Generalized Gradient Scheduling Algorithm

In this section we develop a generalized gradient scheduling algorithm to solve the problem in (3.1).

Adopting the method in Section 3.2, instead of finding the optimal solution to the problem in (3.1) directly, we consider a simple gradient based scheduling policy which attempts to adapt the current rates such that we have the largest first order change in the utility, namely

$$\begin{aligned} & U\big(\mathbf{X}(k+1),\mathbf{Y}(k+1)\big) - U\big(\mathbf{X}(k),\mathbf{Y}(k)\big) \\ = \;& U\big(\mathbf{X}(k)+\mu[\mathbf{r}(k)-\mathbf{X}(k)],\mathbf{Y}(k)+\mu\left\{[\mathbf{r}(k)-\mathbf{X}(k)]\times\right. \\ & \left.[\mathbf{r}(k)-\mathbf{X}(k)]-\mathbf{Y}(k)\right\}\big) - U\big(\mathbf{X}(k),\mathbf{Y}(k)\big) \\ = \;& \mu\nabla_{\mathbf{m}}U(\mathbf{X}(k),\mathbf{Y}(k))^T\cdot[\mathbf{r}(k)-\mathbf{X}(k)]+ \\ & \mu\nabla_{\boldsymbol{\sigma}^2}U(\mathbf{X}(k),\mathbf{Y}(k))^T\cdot\{[\mathbf{r}(k)-\mathbf{X}(k)]\times \\ & [\mathbf{r}(k)-\mathbf{X}(k)]-\mathbf{Y}(k)\}. \end{aligned} \tag{3.8}$$

In (3.8), the last equality holds when the step size μ is sufficiently small. The best scheduler of the system in state $\mathbf{g}$ is to choose a transmission rate vector $\mathbf{r}(k)$ that satisfies

$$\begin{aligned} \mathbf{r}(k) &= \arg\max_{\mathbf{v(k)}\in\mathcal{R}(\mathbf{g})} \nabla_{\mathbf{m}}U(\mathbf{X}(k),\mathbf{Y}(k))^T\cdot\mathbf{v}(k)+ \\ &\quad \nabla_{\boldsymbol{\sigma}^2}U(\mathbf{X}(k),\mathbf{Y}(k))^T\cdot \\ &\quad [(\mathbf{v}(k)-\mathbf{X}(k))\times(\mathbf{v}(k)-\mathbf{X}(k))] \\ &= \arg\max_{\mathbf{v}\in\mathcal{R}(\mathbf{g})}\sum_{i=1}^{N}\left[a_i(k)v_i^2(k)+b_i(k)v_i(k)\right] \end{aligned} \tag{3.9}$$

where

$$a_i(k)=\frac{\partial U(\mathbf{X}(k),\mathbf{Y}(k))}{\partial\sigma_i^2},$$

and

$$b_i(k)=\frac{\partial U(\mathbf{X}(k),\mathbf{Y}(k))}{\partial m_i}-2X_i(t)\frac{\partial U(\mathbf{X}(k),\mathbf{Y}(k))}{\partial\sigma_i^2}.$$

Remark: Let us observe the best scheduler in (3.9). When U strictly increases in m_i and decreases in σ_i^2 for all i (it is a natural assumption in practice), we have $a_i(k) < 0$ and $b_i(k) > 0$ for all i and k. Hence, unlike the gradient scheduling algorithms in [2] and [62] which maximize the sum of weighted transmission rates at each time frame, the best scheduler of GGSA is to maximize the sum of concave functions of transmission rate

$$u_i(r_i) := a_i r_i^2 + b_i r_i. \tag{3.10}$$

The advantage of GGSA is that it can jointly optimize the average rates and rate oscillations. Consider a simple case of GGSA corresponding to a utility function

$$U(\mathbf{m}, \boldsymbol{\sigma}) = \sum_{i=1}^{N} \left[\log(m_i) - \alpha \log(\sigma_i^2)\right], \tag{3.11}$$

where $\alpha \geq 0$ is a parameter that reflects the sensitivity of the communication service to rate oscillation. It is easy to see that when $\alpha = 0$, it is just the same as the utility in the PF algorithm. Thus, we refer to it as PF-GGSA. The best choice of the PF-GGSA is to select a rate vector that satisfies

$$\mathbf{r} \;=\; \arg \max_{\mathbf{v} \in \mathcal{R}(\mathbf{g})} \sum_{i=1}^{N} u_i(v_i) \tag{3.12}$$

where

$$u_i(x) = -\frac{\alpha}{Y_i}\left[x^2 - 2\left(X_i + \frac{Y_i}{2\alpha X_i}\right)x\right].$$

It is noted that the function u_i increases in the region $[0, X_i + Y_i/(2\alpha X_i)]$ but decreases in the region $[X_i + Y_i/(2\alpha X_i), +\infty)$. This provides the key mechanism that the maximum transmission rate of user i chosen by the best scheduler is always bounded by $X_i + Y_i/(2\alpha X_i)$. Therefore, the shared resource (e.g., time or frequency) can be allocated to other users with lower transmission rate. A larger value of α means more sensitivity to rate oscillation. Correspondingly, the selected transmission rate is restricted within a smaller region.

3.4 Asymptotic Analysis of GGSA

We study the dynamics of average rates and rate variances under the GGSA algorithm when the step size μ approaches zero. For this purpose we define a fluid sample path (FSP) under GGSA. For a fixed step size μ, let $\mathbf{X}^\mu(k)$ and $\mathbf{Y}^\mu(k)$ be the realizations of an average rate and rate variance vector process. For a given channel state realization $\mathbf{g}^\mu = (\mathbf{g}^\mu(k),\ k = 0, 1, 2, \ldots)$ and a fixed initial state $\mathbf{X}^\mu(0)$ and $\mathbf{Y}^\mu(0)$, $\mathbf{X}^\mu(k)$ and $\mathbf{Y}^\mu(k)$ are uniquely determined. Consider the following continuous time process $\big(\mathbf{x}^\mu(t), \mathbf{y}^\mu(t)\big)$,

$$\begin{aligned} \mathbf{x}^\mu(t) &:= \mathbf{X}^\mu(\lfloor t/\mu \rfloor), \quad t \geq 0 \\ \mathbf{y}^\mu(t) &:= \mathbf{Y}^\mu(\lfloor t/\mu \rfloor), \quad t \geq 0, \end{aligned}$$

where $\lfloor x \rfloor := \sup\{i \in \mathbb{Z} : i \leq x\}$.

A pair of vector-functions $\big(\mathbf{x} = (\mathbf{x}(t), t \geq 0), \mathbf{y} = (\mathbf{y}(t), t \geq 0)\big)$ is called a fluid sample path (FSP) [62], if there exist a sequence of positive values of μ such that as $\mu \to 0$, a sequence of sample path $(\mathbf{x}^\mu(t), \mathbf{y}^\mu(t))$ satisfy

$$(\mathbf{x}^\mu(t), \mathbf{y}^\mu(t)) \to (\mathbf{x}(t), \mathbf{y}(t)) \quad u.\ o.\ c.$$

Here, *u. o. c.* means uniform on compact sets convergence of function.

Some properties of FSPs are described in the following lemmas based on which we establish Theorem 1.

Lemma 1: For any FSP, $\mathbf{x}$ and $\mathbf{y}$ are Lipschitz continuous in $[0, \infty)$.

Proof: Recall the update equations (3.2) and (3.3) in Section 3.3. Applying

these equations iteratively, we have

$$\begin{aligned} X_i^\mu(k) &= \sum_{j=1}^{k-1} \mu(1-\mu)^j r_i^\mu(k-j) + (1-\mu)^k X_i^\mu(0) \\ &\leq \max\{\bar{r}, X_i^\mu(0)\}, \quad \forall k \geq 0 \\ Y_i^\mu(k) &= \sum_{j=1}^{k-1} \mu(1-\mu)^j [r_i^\mu(k-j) - X_i^\mu(k-j)]^2 \\ &\quad +(1-\mu)^k Y_i^\mu(0) \\ &\leq \max\{\max\{\bar{r}^2, [x_i^\mu(0)]^2\} + \bar{r}^2 Y_i^\mu(0)\}, \quad \forall k \geq 0, \end{aligned}$$

where $\bar{r}$ is the upper bound on the rate $r_i^\mu(k),\ \forall i, k, \mu > 0$. Then, we have

$$\begin{aligned} |X_i^\mu(k) - X_i^\mu(k-1)| &= |\mu r_i^\mu(k) - \mu X_i^\mu(k-1)| \\ &\leq \mu[r_i^\mu + X_i^\mu(k-1)] \\ &\leq \mu[2\bar{r} + X_i^\mu(0)]. \end{aligned} \tag{3.13}$$

Similarly, we can prove

$$\begin{aligned} & |Y_i^\mu(k) - Y_i^\mu(k-1)| \\ \leq\ & \mu\left[2\max\{\bar{r}^2, [x_i^\mu(0)]^2\} + 2\bar{r}^2 + Y_i^\mu(0)\right]. \end{aligned} \tag{3.14}$$

The inequalities (3.13) and (3.14) imply that $(\mathbf{x}, \mathbf{y})$ is Lipschitz continuous in $[0, \infty)$.

■

Under the assumption that the chosen rate vector $\mathbf{r}(k) = \mathbf{F}\big(\mathbf{X}(k), \mathbf{Y}(k), \mathbf{g}(k)\big)$ is continuous for almost all $\big(\mathbf{X}(k), \mathbf{Y}(k), \mathbf{g}(k)\big)$, we have the Lemma 2.

Lemma 2: The family of FSPs $(\mathbf{x}, \mathbf{y})$ satisfies the ordinary differential equations (ODE),

$$\dot{\mathbf{x}} = \bar{\mathbf{F}}(\mathbf{x}, \mathbf{y}) - \mathbf{x} \tag{3.15}$$

$$\dot{\mathbf{y}} = \overline{\mathbf{F}^2}(\mathbf{x}, \mathbf{y}) - 2\mathbf{x} \times \bar{\mathbf{F}}(\mathbf{x}, \mathbf{y}) - \mathbf{x} \times \mathbf{x} - \mathbf{y} \tag{3.16}$$

where

$$\bar{\mathbf{F}}(\mathbf{x},\mathbf{y}) = \int_{\mathcal{G}} \mathbf{F}\big(\mathbf{x}(s),\mathbf{x}(s),\mathbf{g}\big)\gamma(\mathrm{d}\mathbf{g}) \tag{3.17}$$

$$\begin{aligned}\overline{\mathbf{F}^2}(\mathbf{x},\mathbf{y}) &= \int_{\mathcal{G}} \mathbf{F}\big(\mathbf{x}(s),\mathbf{x}(s),\mathbf{g}\big) \\ &\quad \times \mathbf{F}\big(\mathbf{x}(s),\mathbf{x}(s),\mathbf{g}\big)\gamma(\mathrm{d}\mathbf{g}).\end{aligned} \tag{3.18}$$

Proof: Based on Lemma 1 of [9], we can obtain

$$\begin{aligned}\mathbf{x}(t) &= \mathbf{x}(0) + \int_{\mathcal{G}\times[0,t]} \big[\mathbf{F}\big(\mathbf{x}(s),\mathbf{x}(s),\mathbf{g}\big) - \mathbf{x}(s)\big]\,\gamma(\mathrm{d}\mathbf{g})\mathrm{d}s \\ &= \mathbf{x}(0) + \int_0^t \big[\bar{\mathbf{F}}\big(\mathbf{x}(s),\mathbf{y}(s)\big) - \mathbf{x}(s)\big]\,\mathrm{d}s \\ \mathbf{y}(t) &= \mathbf{y}(0) + \int_{\mathcal{G}\times[0,t]} \big\{\big[\mathbf{F}\big(\mathbf{x}(s),\mathbf{x}(s),\mathbf{g}\big) - \mathbf{x}(s)\big] \times \\ &\quad \big[\mathbf{F}\big(\mathbf{x}(s),\mathbf{x}(s),\mathbf{g}\big) - \mathbf{x}(s)\big] - \mathbf{y}\big\}\,\gamma(\mathrm{d}\mathbf{g})\mathrm{d}s \\ &= \mathbf{y}(0) + \int_0^t \big[\overline{\mathbf{F}^2}\big(\mathbf{x}(s),\mathbf{y}(s)\big) - 2\mathbf{x}(s) \times \\ &\quad \bar{\mathbf{F}}\big(\mathbf{x}(s),\mathbf{y}(s)\big) + \mathbf{x}(s) \times \mathbf{x}(s) - \mathbf{y}(s)\big]\mathrm{d}s\end{aligned}$$

Since $(\mathbf{x},\mathbf{y})$ of an FSP is Lipschitz continuous, its derivatives exist at almost all points and, therefore, for every regular point $t \geq 0$, $(\mathbf{x},\mathbf{y})$ satisfies the ODEs (3.15) and (3.16). ■

Lemma 3: Suppose $(\mathbf{x},\mathbf{y})$ is a stationary FSP, namely

$$\mathbf{x}(t) \equiv \mathbf{x}^*,\ \mathbf{y}(t) \equiv \mathbf{y}^*,\quad \forall t \geq 0. \tag{3.19}$$

Then $(\mathbf{x},\mathbf{y})$ is a solution to the problem in (3.1).

Proof: For simplicity, we define

$$\bar{\mathbf{H}}(\mathbf{x},\mathbf{y}) := \overline{\mathbf{F}^2}(\mathbf{x},\mathbf{y}) - 2\mathbf{x} \times \bar{\mathbf{F}}(\mathbf{x},\mathbf{y}) - \mathbf{x} \times \mathbf{x}.$$

When $\mathbf{x}(t) \equiv \mathbf{x}^*$, we have $\bar{\mathbf{F}}(\mathbf{x},\mathbf{y}) = \mathbf{x}$ from (3.15). It follows that

$$\bar{\mathbf{H}}(\mathbf{x},\mathbf{y}) = \overline{\mathbf{F}^2}(\mathbf{x},\mathbf{y}) - \big[\bar{\mathbf{F}}(\mathbf{x},\mathbf{y})\big] \times \big[\bar{\mathbf{F}}(\mathbf{x},\mathbf{y})\big]\,,$$

By (3.9), (3.17) and (3.18), we have

$$\begin{aligned}&\big(\bar{\mathbf{F}}(\mathbf{x},\mathbf{y}),\bar{\mathbf{H}}(\mathbf{x},\mathbf{y})\big)\\ =\ &\arg\max_{(\mathbf{v},\mathbf{u})\in\mathcal{V}} \nabla_{\mathbf{m}}U(\mathbf{x},\mathbf{y})^T\cdot\mathbf{v}+\nabla_{\boldsymbol{\sigma}^2}U(\mathbf{x},\mathbf{y})^T\cdot\mathbf{u}.\end{aligned}$$

When (3.19) holds, we have

$$\big(\bar{\mathbf{F}}(\mathbf{x},\mathbf{y}),\bar{\mathbf{H}}(\mathbf{x},\mathbf{y})\big)=(\mathbf{x}^*,\mathbf{y}^*),$$

This implies that $(\mathbf{x}^*,\mathbf{y}^*)$ is a solution to the following problem

$$\max_{(\mathbf{v},\mathbf{u})\in\mathcal{V}} \nabla_{\mathbf{m}}U(\mathbf{x}^*,\mathbf{y}^*)^T\cdot\mathbf{v}+\nabla_{\boldsymbol{\sigma}^2}U(\mathbf{x}^*,\mathbf{y}^*)^T\cdot\mathbf{u},$$

which means that $(\mathbf{x}^*,\mathbf{y}^*)$ is a maximal point of the set $\mathcal{V}$. In addition, $\big(\nabla_{\mathbf{m}}U(\mathbf{x}^*,\mathbf{y}^*)$, $\nabla_{\boldsymbol{\sigma}^2}U(\mathbf{x}^*,\mathbf{y}^*)\big)$ is normal to the set $\mathcal{V}$. Therefore, $(\mathbf{x},\mathbf{y})$ is a solution to the problem in (3.1). ■

From Lemma 1, 2 and 3, we have Theorem 1.

Theorem 1: Suppose FSP, $(\mathbf{x},\mathbf{y})$ satisfies $\big(\mathbf{x}(t),\mathbf{y}(t)\big)\to(\mathbf{x}^*,\mathbf{y}^*)$ as $t\to\infty$, then $(\mathbf{x}^*,\mathbf{y}^*)$ is a solution to the problem in (3.1).

Proof: The proof is analogous to [3, Proof of Theorem 1]. ■

Theorem 1 reveals that GGSA is asymptotically optimal under the condition that the transmission rate vector converges to a fixed vector as time goes into infinity. Whether the average rate and rate variance vectors will converge is, however, not known. We shall study the convergence properties using numerical results in the next section.

3.5 GGSA in Time-Sharing Wireless Networks

In this section, We concentrate on GGSA in a time division multiplexing (TDM) system where each time frame can be accessed by all the N users in an adaptive

time-sharing fashion. The channels of different users are statistically independent and follows the exponential distribution, denoted by

$$pdf_i(g) = \begin{cases} \frac{1}{\bar{g}_i}\exp\left(-\frac{g}{\bar{g}_i}\right) & g \geq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3.20}$$

where $\bar{g}_i$ is the average channel gain of user i. Let p_i denote the transmit power allocated to or from user i. The achievable transmission rate of user i in the absence of other users can be expressed as

$$c_i = \log_2\left(1 + \frac{p_i g_i}{\Gamma N_0}\right), \tag{3.21}$$

where N_0 is the noise power, and Γ is the SNR gap as defined in Chapter 2. Let $\boldsymbol{\rho}(\mathbf{g}) = (\rho_1, \rho_2, \ldots, \rho_N)$ denote the time-sharing adaptation policy with respect to the network channel gain $\mathbf{g}$, where ρ_i represents the fraction of the frame duration allocated to user i. The actual transmission rate of user i in each time frame, r_i, can therefore be written as

$$r_i = \rho_i c_i = \rho_i \log_2\left(1 + \frac{p_i g_i}{\Gamma N_0}\right). \tag{3.22}$$

3.5.1 Continuous Time Sharing (TS) with Perfect CSI

We assume that the transmission powers between BS and mobiles are constant and identical for different users, i.e., $p_i(t) = p, \forall i, t$, and that the wireless network is fully loaded. We choose $\sum_i u_i(r_i)$ given in (3.12) as the objective function in each time frame. The goal is to find the optimal time-sharing adaptation policy $\boldsymbol{\rho}^*(\mathbf{g})$ relative to the instantaneous network channel condition $\mathbf{g}$, so as to maximize the objective function. The optimization problem can be expressed mathematically as

$$\max_{\boldsymbol{\rho}} \quad I_{\text{TS}} \triangleq \sum_{i=1}^{N} u_i(r_i) \tag{3.23}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} \rho_i = 1. \tag{3.24}$$

Since the utility $u_i(\cdot)$ is a concave function of r_i by assumption, $u_i(\cdot)$ is also concave in ρ_i. Therefore, taking the derivative of the Lagrangian $\sum_{i=1}^{N} U_i(\rho_i, g_i) + \lambda(1 - \sum_{i=1}^{N} \rho_i)$, and equating it to zero, we obtain ρ_i^* as

$$\rho_i^*(\mathbf{g}) = \left[\left(\frac{\partial U_i(\rho_i, g_i)}{\partial \rho_i} \right)^{-1} (\lambda) \right]^+, \quad i = 1, 2, \ldots, N. \tag{3.25}$$

In (3.25), $(\partial U_i/\partial \rho_i)^{-1}(\cdot)$ is the inverse function of $(\partial U_i/\partial \rho_i)(\cdot)$[1], and $[x]^+ \triangleq \max(0, x)$.

If we allow both the transmission time and power to change with respect to channel conditions in each time frame, the optimization problem of time sharing is extended to finding the joint optimal time-sharing and power control policy (JTPC). Uplink and downlink transmission are considered separately due to different power constraints. For the uplink, the power source is, generally, rechargeable batteries attached to the mobile devices. Thus, the optimization is subject to each user's average power constraint.

The closed-form solution to the joint optimization is generally difficult to obtain due to nonlinearity of the utility function U in ρ and p. The nonlinear Gauss-Seidel algorithm [6] can be used to search for the optimal time-sharing vector $\boldsymbol{\rho}^*$ and vector $\mathbf{s}^* = [s_1^*, s_2^*, \ldots, s_N^*]$ ($s_i := \rho_i p_i$). In Gauss-Seidel algorithm, the optimization of $\boldsymbol{\rho}$ and $\mathbf{s}$ is carried out successively at each iteration, and the iteration stops when the difference of objective function between two last iterations is sufficiently small. The condition that U is continuously differentiable and concave in (ρ, s) guarantees the convergence of the nonlinear Gauss-Seidel algorithm. The proof can be seen in [6, Prop 3.9 in Section 3.3].

While JTPC utilizes two degrees of freedom in resource allocation and has much higher computational complexity, its performance is not expected much higher than

[1]When the utility function is strictly concave, $(\partial U_i/\partial \rho_i)(\cdot)$ is a monotonically decreasing function of ρ and, hence, its inverse exists.

that of TS in the high SNR region. This is attributed to the fact that the transmission rate is linear in time, but concave in transmission power. That is, at high SNR, the gain from power control is smaller than from time sharing adaptation. This is also verified by the simulation results in [75], where we compare the objective function obtained by the proposed TS and JTPC schemes. It is observed that at high SNR the performance gain of JTPC over TS is not noticeable. Hence, in the remaining parts, we assume the absence of power control.

3.5.2 Quantized Time Sharing With Limited Channel Feedback

In scheduling the downlink transmission, the BS needs to know each user's channel state information (CSI). This could be gained by sending the CSI from each user to the base station through a feedback channel upon channel estimation at each user terminal. In practice, perfect channel feedback is not feasible due to limited capacity of the feedback links. We assume in this subsection that the channel estimate of each user is quantized into $K = 2^M$ regions using M bits. Let $\mathbb{G} = \{G_1, G_2, \ldots, G_{K+1}\}$ be the set of channel gain thresholds in increasing order with $G_1 = 0$ and $G_{K+1} = \infty$. If the channel gain of user i falls into range $[G_k, G_{k+1})$, we say user i is in channel state k, and denote it as $S_i = k$. Suppose we apply the equal-probability method to do the channel partitioning and the channel gains follow exponential distribution, the threshold set $\mathbb{G}$ is simply the channel partitioning boundaries.

Furthermore, time sharing fractions in practice cannot be an arbitrary number, but are restricted to a finite set of values due to switching latency and difficulties in rigid synchronization. Therefore, we assume that a time frame is partitioned into L slots with equal length. Correspondingly, the number of users which can transmit in the same frame is limited by L. At the beginning of each frame, the base station

computes the optimal time-sharing vector $\boldsymbol{\rho}^*$ defined in

$$\begin{aligned} \boldsymbol{\rho}^*(\mathbf{S}) &= \arg\max_{\boldsymbol{\rho}(\mathbf{S})} \int_{G_{S_1}}^{G_{S_1+1}} \int_{G_{S_2}}^{G_{S_2+1}} \cdots \int_{G_{S_N}}^{G_{S_N+1}} \left[\sum_{i=1}^{N} u_i(\rho_i(\mathbf{g}), g_i)\right] \mathrm{d}\mathbf{g} \quad (3.26) \\ \text{s.t.} \quad & \rho_i \in \left\{0, \frac{1}{L}, \frac{2}{L}, \ldots, 1\right\}, \ \forall i \in \{1, \ldots, N\} \text{ and } \sum_{i=1}^{N} \rho_i = 1. \end{aligned}$$

upon obtaining the network channel states $\mathbf{S} = [S_1, S_2, \ldots, S_N]$.

The time sharing policy considered here maps the current channel states $\mathbf{S} \in \mathbb{R}_+^N$ to a time-sharing vector. To avoid the exponential complexity in exhaustive search, an online greedy algorithm with complexity of $\mathcal{O}(LN)$ is proposed. Beginning with an initial solution $\boldsymbol{\rho} = [0, 0, \ldots, 0]$, each time slot is assigned at one iteration to the most favorable user that maximizes the increment of the current objective till the total L slots are traversed. The greedy algorithm is outlined below:

1. *Initialization*

   Let $v = 0$ (the index of the time slot), $\rho_i^{(0)} = 0$ and $u_i^{(0)} = 0$ ($\forall i \in \{1, \ldots, N\}$).

2. *Allocate the* $(v+1)^{th}$ *time slot to the user indexed by* i^*

   $$i^* = \arg \max_{i \in \{1,\ldots,N\}} \int_{G_{S_i}}^{G_{S_i+1}} \left[u_i(\rho_i^{(v)} + 1/L) - u_i(\rho_i^{(v)})\right] \mathrm{d}g_i. \quad (3.27)$$

   Let $\rho_{i^*}^{(v+1)} = \rho_{i^*}^{(v)} + 1/L$ and $\rho_i^{(v+1)} = \rho_i^{(v)}$ for $i \neq i^*$.

3. *Let* $v = v + 1$*, and return to Step 2) until* $v = L$

It is shown in the Appendix I that this greedy algorithm leads to the optimal solution to Problem (3.26).

3.6 Numerical Results

In the numerical study, a time-sharing wireless network with symmetric users is considered. The target BER is set to be 10^{-5}, which results in a SNR gap of 8.2dB.

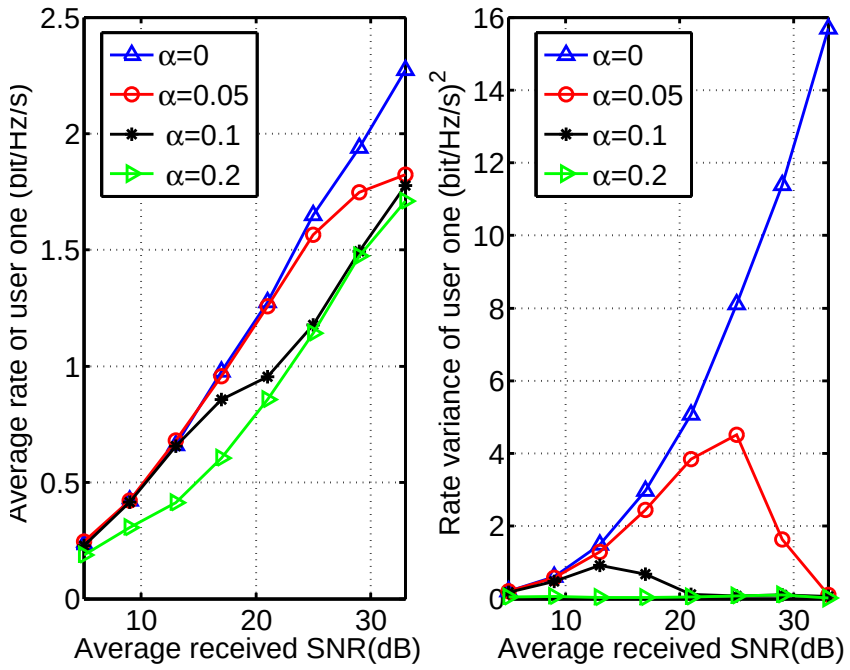

Figure 3.1. Average rate and rate variance of user one versus average received SNR

The average received SNR is varied from 0 dB to 35 dB. We assume that the feedback channel is error free and has no delay. Each run comprises 50000 time frames. The PF-GGSA scheduler defined in (3.12) is used for simplicity.

Fig. 3.1 compares the average rates and the rate variances by varying the parameter α in the set $\{0, 0.05, 0.1, 0.2\}$ using optimal time sharing policy when the network has 4 users. It is observed that the GGSA has a flexible balance between the average rate and the rate oscillation through adjusting the parameter α. When $\alpha = 0$, the GGSA is equivalent to the PF algorithm in [32], and it can obtain the maximum average transmission rate (in the symmetric case) while it suffers a high rate oscillation.

Figs. 3.2 and 3.3 present the trajectories of the GGSA with different starting points and step sizes, μ, using optimal time sharing policy. When the step size is sufficiently small, the trajectories with different starting points and different step

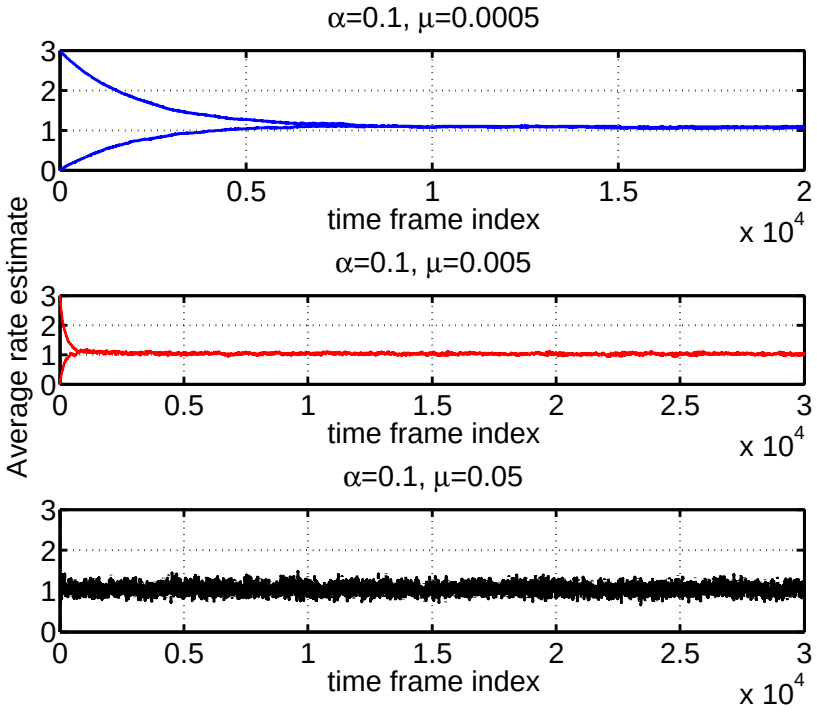

Figure 3.2. Trajectories of the average rate of user one with different starting points and step sizes

sizes always converge to the same optimal point. The speed of convergence depends on the step size. A small step size corresponds to a slow speed of convergence. When the step size is not small enough, the trajectories oscillate around the optimal points.

Next, we compare the performance of the quantized time sharing with limited channel feedback (QTSL) with optimal time sharing.

Fig. 3.4 shows the performance of the quantized time sharing with limited channel feedback (QTSL) policy. The CSI is quantized at 3-bit resolution, and the number of time slots is the same as the number of users in the network. It is seen that the performance of QTSL is very close to that obtained by the optimal TS policy.

In Fig. 3.5, we illustrate the performance of the QTSL policy when the number of time slots in a time frame is half of the number of users in the network. This time we also vary the number of channel feedback bits from 2 to 3. There is a performance

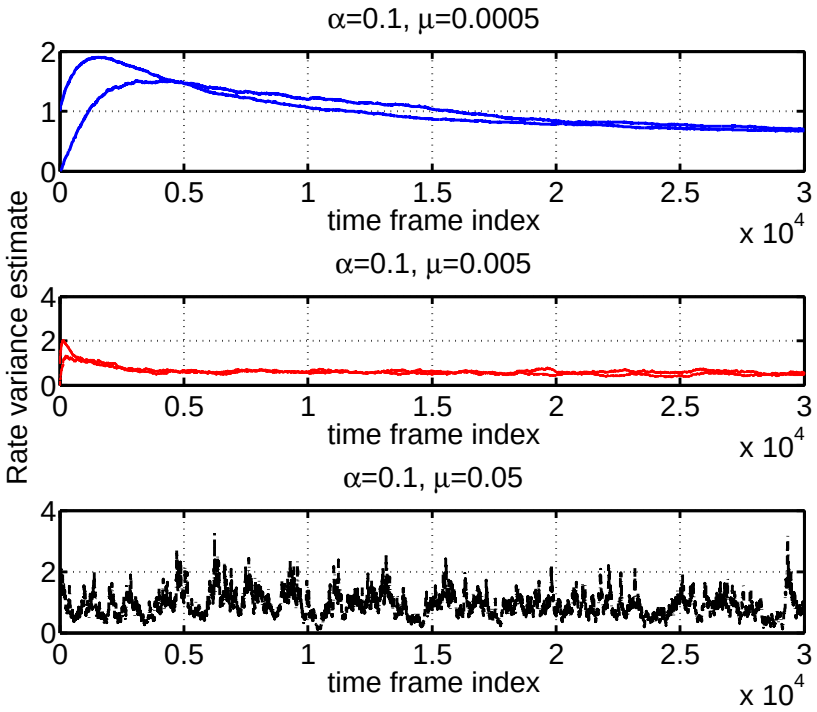

Figure 3.3. Trajectories of the rate variance of user one with different starting points and step sizes

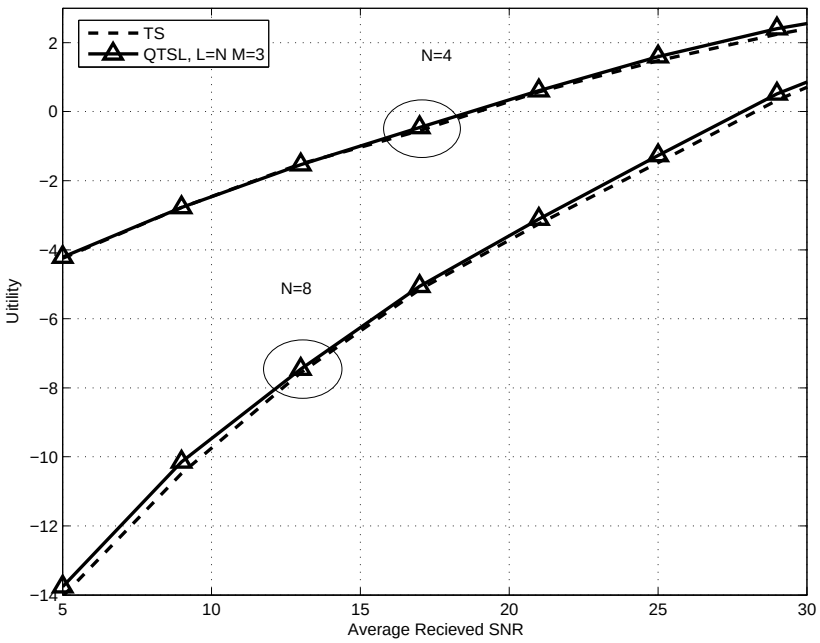

Figure 3.4. Performance comparison of the optimal TS policy and QTSL policy for $N = 4$ and 8 users, with $L = N$ slots and $M = 3$ feedback bits

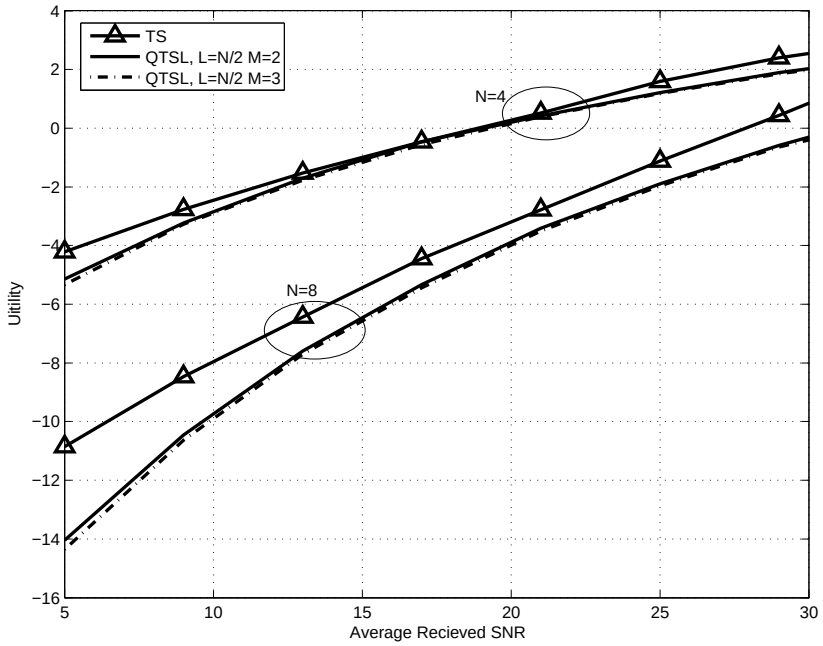

Figure 3.5. Performance comparison of $N = 4$ users with $L = 2$ time slots and $N = 8$ users with $L = 4$ time slots with $M = 2$ and $M = 3$ feedback bits

gap between QTSL and TS. In addition, increasing the CSI quantization resolution to beyond two bits does not lead to noticeable performance gain.

3.7 Conclusions

We proposed a GGSA for wireless networks supporting variable rate transmission. This algorithm can maximize a utility function that increases with average transmission rate but decreases with rate variance. The utility function is capable of facilitating the resource allocation with flexible combinations of average rate and rate oscillation, which are both important performance metrics for variable rate transmission. The GGSA suggests that the best scheduler should maximize the sum of concave functions of instantaneous transmission rate at each time frame. We studied the asymptotic performance of the algorithm, and showed that if the transmission

rates converge, then the convergence point maximizes the desired utility function. Numerical results show a good convergence performance of the algorithm in time division multiplexing systems when the step size is small enough.

Chapter 4

Lifetime Maximization for Wireless Sensor Networks

The problem of dynamic resource allocation studied in Chapter 3 mainly focuses on joint optimization of average rate and rate oscillation by exploring the channel variance over time. That study is expected to be beneficial to the multiuser transmission for soft delay-constrained traffic in the context of infrastructure-based wireless network. In this chapter, we will demonstrate that dynamic resource allocation is still highly beneficial in energy efficient transmission design by exploring the heterogeneity of sensor environment in wireless sensor networks (WSN) , even when there are no channel variation over time.

Energy limitation is one of the greatest distinctions between a wireless sensor network (WSN) and other wireless data networks such as cellular networks and wireless local area networks (WLAN). A natural problem in this area is the maximization of network lifetime, i.e., the time that it takes for the first node to die. The power in WSN is consumed by transmission and reception, data processing, sensing, and switching between active mode and sleeping mode. For simplicity, we consider only

the power required for transmission which is dominant in total power consumption as in [48] and [69].

Adaptive resource allocation design is a potent technique for improving wireless data network performance under application-specific constraints [27]. The concept is to adapt the transmission parameters, such as transmission power, transmission time and channel bandwidth, to the underlying channel, interference, and system preferences. The work [26] shows that adaptation can obtain up to 20 dB power saving over non-adaptive transmission for a single communication link. In multi-user systems, significant system performance gain can also be achieved by optimal transmission scheduling and power control. For instance, in a multi-access channel, the authors in [38] proposed a water-filling based power control algorithm to maximize the sum-of-rate capacity subject to the average transmission power constraint of each user.

The benefit to wireless data networks from adaptive resource allocation motivates us to apply the adaptation in wireless sensor networks. The optimal power control problem for the decentralized estimation of a noise-corrupted deterministic signal in an inhomogeneous sensor network is studied in [69]. The proposed power control scheme suggests that the sensors with bad channel should decrease quantization resolutions for total energy saving. Yao and Giannakis in [73] propose to minimize the total energy consumption by varying the transmission times assigned to different sensor nodes under the individual rate constraint for each node. However, minimizing the total power consumption does not guarantee maximum lifetime for a network since the objective, total power minimization, always results in some nodes with non-rechargeable batteries running out of energy quickly [13]. In addition, existing adaptation schemes relying on centralized algorithms in wireless data networks cannot be applied directly in a WSN. This is because the fusion center only has limited computation capacity and the network cannot afford large control signalling overhead.

We consider the lifetime maximization for a cluster-based WSN. Sensor nodes are organized into clusters. Each cluster consists of multiple sensor nodes and one common fusion center (FC), and is responsible for monitoring a certain geographical area. The focus of this chapter is on single-hop data collection from sensor nodes to the FC in a particular cluster. The responsibility of sensor nodes is to sense the surrounding environment, quantize the observation and report to the FC. We assume that FC is less energy-constrained than sensor nodes.

In this chapter, integrated designs of quantization and transmission are investigated to maximize the network lifetime. The decision of a sensor node to select transmission power, time and quantization resolution may depend on the information of observation quality, channel condition and initial energy. We characterize the optimal decision under two typical scenarios:

1. Uncorrelated source observation: Each sensor nodes are distributed sparsely and observe uncorrelated signals.

2. Common source observation: All sensor nodes are located densely and observe only one common physical phenomenon.

In the former scenario, the optimal solution indicates that the "weak" sensor nodes, the ones with small product of channel gain and initial energy and/or low quality of observation, should take longer time for transmission with a low received power. The aim is to give these "weak" nodes priority to take advantage of "lazy scheduling" for energy saving in order to avoid becoming the bottleneck of the whole WSN. In the latter scenario, the optimal solution not only takes advantage of "lazy scheduling", but also "opportunistic quantization", that is, the sensor nodes with high observation qualities should increase their quantization resolution. We also demonstrate that the optimal decision can be implemented in a partially distributed way. Simulation results

show a significant lifetime gain over existing schemes, especially when the sensing environment becomes more heterogeneous and the number of nodes increases.

The rest of the chapter is organized as follows. Section 4.1 describes the system model. In Section 4.2 and 4.3, we obtain some properties of the optimal solutions and develop partially distributed adaptive optimization algorithms for the two scenarios. Section 4.4 presents and evaluates the simulation results and, finally, Section 4.5 concludes the chapter.

4.1 System Model

Consider a distributed estimation problem in a single-hop WSN depicted in Fig 4.1 with a FC and a set of K sensor nodes, denoted by $\mathcal{K}$.

Source signal and observation noise: Each sensor, say the kth one, observes a scalar x_k, which are described by

$$x_k = \phi_k(\boldsymbol{\theta}) + \omega_k, \quad k \in \mathcal{K}, \tag{4.1}$$

where vector $\boldsymbol{\theta}$ is a $p \times 1$ random signal of interest and ω_k is additive *observation noise* in which ω_k ($k \in \mathcal{K}$) are independent zero-mean random variable with variance σ_k^2. Generally, ϕ_k: $\mathbb{R}^p \rightarrow \mathbb{R}$ is a nonlinear function. However, in this chapter, we particularly focus on two representative scenarios:

1. Uncorrelated source observation: $p = K$ and $\phi_k(\boldsymbol{\theta}) = \theta_k$. It corresponds to the networks where the K sensor nodes observe K uncorrelated signals.

2. Common source observation: $p = 1$ and $\phi_k(\theta) = \theta$. In this case, vector $\boldsymbol{\theta}$ is reduced to a scalar. All sensors observe only one common physical phenomenon θ.

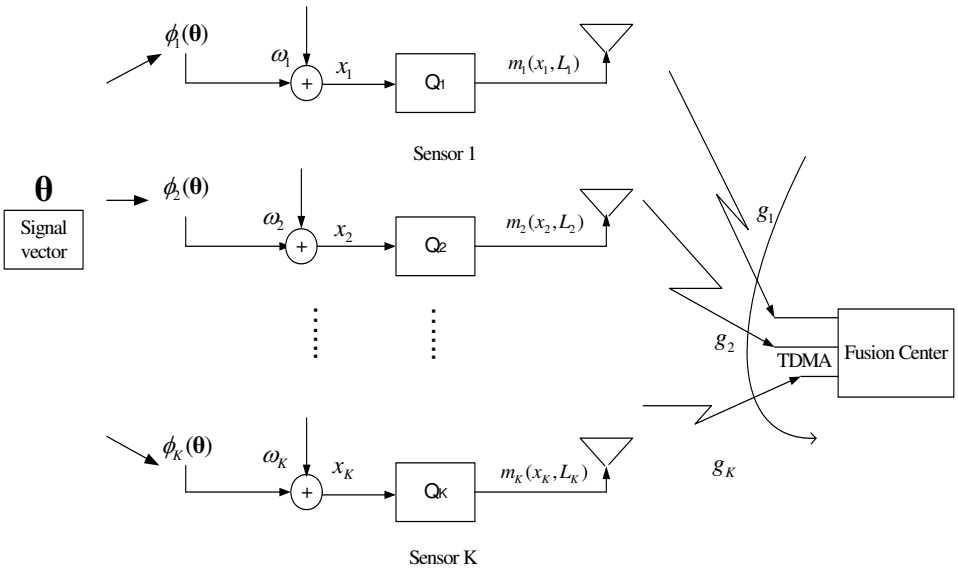

Figure 4.1. Data fusion procedure in a WSN

While it is hard to characterize the probability density function (pdf) of x_k, it is often possible to measure the observation noise variance σ_k^2 at the receiver without the presence of the signal of interest. The value of σ_k^2 can directly be used to measure the observation quality of sensor node k. Difference of observation qualities among sensor nodes arises from different distances from the signals of interest and different sensing resolutions.

Quantization: Due to the bandwidth and power constraints, it may be preferable to let each node transmit a quantized version of x_k instead of the analog-amplitude x_k to the FC to enable the final estimation. Suppose sensor observations, x_k's, have a bounded range in the interval $[-W, W]$ for all k, where W is a known parameter determined by the sensor's observation range. When the pdf of x_k is unavailable, x_k can be quantized into L_k bits by uniformly dividing the region $[-W, W]$ into 2^{L-1} intervals of length Δ, and mapping x_k to a discrete message $m_k(x_k, L_k)$ following the same way as in [69]

$$m_k(x_k, L_k) = \phi_k(\boldsymbol{\theta}) + \omega_k + v_k(x_k, L_k),$$

where $v_k(x_k, L_k)$ denotes the *quantization noise* which is independent of ω_k. It is shown in [70] that $m_k(x_k, L_k)$ is an unbiased estimator of $\boldsymbol{\theta}$, namely, $\mathbb{E}(m_k) = \mathbb{E}(x_k)$. Moreover, the variance of m_k approaches σ_k^2 at an exponential rate as L increases

$$\mathbb{E}[|m_k(x_k, L_k) - \phi_k(\boldsymbol{\theta})|^2] \leq \frac{W^2}{(2^{L_k} - 1)^2} + \sigma_k^2, \ \forall L \geq 1, k \in \mathcal{K}, \tag{4.2}$$

where $\mathbb{E}$ denotes expectation taken over both the sensor observation noise and quantization noise. The inequality in (4.2) comes from the fact that $r(1-r) \leq 1/4$, where r is a random variable and satisfies $0 \leq r \leq 1$.

Communication: Except for quality of the information, timeliness of the information is another important issue. It is in each time frame that the resulting discrete message $m_k(x_k, L_k)$ should be transmitted consecutively from each sensor node to FC in an adaptive time sharing fashion through an additive white Gaussian noise (AWGN) channel. Without loss of generality, the time-frame duration is assumed to be normalized. The channel gain between FC and sensor nodes is modelled as g_k. We assume that g_k remains approximately constant during the network lifetime. This assumption is reasonable in case where the network condition change slowly in a quasi-static manner. Let $\boldsymbol{\rho} = [\rho_1, \rho_2, \ldots, \rho_K]$ denote the time-sharing policy, where ρ_k represents the fraction of the frame duration allocated to node k. It is assumed that the transmission is error-free and the information rate is given by the mutual information. Thus we have:

$$L_k = \left\lfloor \rho_k \log_2 \left(1 + \frac{g_k p_k}{N_0}\right) \right\rfloor, \quad \forall k \in \mathcal{K}, \tag{4.3}$$

where $\lfloor x \rfloor$ represents the largest integer that is not greater than x. When $x \gg 1$, $\lfloor x \rfloor \approx x$. $\mathbf{p} = [p_1, p_2, \ldots, p_K]$ is the transmission power vector. Hence the energy consumed by node k in a time frame is $\rho_k p_k$. Without loss of generality, the noise power, N_0, is assumed to be one throughout this chapter.

Estimation: After collecting the data from all sensor nodes, the FC combines

them to produce a final estimation of $\boldsymbol{\theta}$

$$\bar{\boldsymbol{\theta}} = \Gamma\big(m_1(x_1, L_1), \ldots, m_K(x_K, L_K)\big). \tag{4.4}$$

If we adopt the mean square error (MSE) criterion to measure the estimation accuracy, a linear estimator of $\boldsymbol{\theta}$ based on BLUE [36] can be used in the two scenarios to minimize MSE,

1. Uncorrelated source observation:

$$\bar{\theta}_k = m_k,$$

and the MSE, D_k, satisfies

$$D_k \triangleq \mathbb{E}(|m_k(x_k, L_k) - \theta_k|^2) \leq \frac{W^2}{(2^{L_k} - 1)^2} + \sigma_k^2, \ \forall k \in \mathcal{K}. \tag{4.5}$$

2. Common source observation:

$$\bar{\theta} = \left(\sum_{k\in\mathcal{K}} \frac{1}{\sigma_k^2 + \delta_k^2}\right)^{-1} \sum_{k\in\mathcal{K}} \frac{m_k}{\sigma_k^2 + \delta_k^2}, \ \forall k \in \mathcal{K}$$

where δ_k^2 is the upper bound of quantization noise variance at sensor node k, given by

$$\delta_k^2 = \frac{W^2}{(2^{L_k} - 1)^2}. \tag{4.6}$$

In addition, WSN has a MSE, D, upper-bounded by

$$D \leq \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2 + \delta_k^2}\right)^{-1}. \tag{4.7}$$

Problem formulation: Network lifetime is an important performance metric for WSNs where energy-constrained sensor nodes collaborate on a certain task. If one or more nodes fail due to a lack of energy, the sensor network may not sustain normal functionality. Thus, the lifetime of the network often refers to the time it takes for the first node in the network to die. Apart from the observation quality measured

by $\{\sigma_k^2,\ \forall k \in \mathcal{K}\}$ and CSI by $\{g_k,\ \forall k \in \mathcal{K}\}$, initial energy information also plays a crucial role in the network lifetime by prioritizing nodes with more initial energy for transmission. As is aforementioned, the quantization resolution and the transmission parameters in physical layer are linked by (4.3). Estimation accuracy is dependent on the quantization resolution and observation quality of each nodes following the relationship (4.5) or (4.7). In addition, the time by which the sensor node runs out of energy is determined by initial energy and transmission power and time sharing fraction. Therefore, exploiting dependencies between quantization and transmission is crucial to offer improvement in overall system performances. In this chapter, the goal is to jointly optimize the quantization and transmission schemes by exploiting CSI, observation quality and initial energy to maximize the network lifetime. For both scenarios, the objective and the constraints are given as follows.

The objective: The objective is to maximize the minimum of times by which the sensor nodes die, T, which is also defined as network lifetime,

$$\max\ T \triangleq \min_k T_k.$$

Estimation accuracy constraint: We choose the MSE to quantify the estimation accuracy. For uncorrelated source observation, the estimation of signal θ_k is assumed to have a target MSE constraint, $\bar{D}_k$,

$$D_k \leq \bar{D}_k,\ \forall k \in \mathcal{K}. \tag{4.8}$$

For a common source observation, the overall estimation accuracy constraint is given by

$$D \leq \bar{D}.$$

Energy constraint: T_k and E_k are used to denote the lifetime and initial energy of sensor node k, respectively, with $\rho_k p_k$ being the energy consumed in each time frame. Then,

$$T_k \rho_k p_k = E_k,\ \forall k \in \mathcal{K}. \tag{4.9}$$

Time constraint: Under the assumption that the time sharing fraction can take arbitrary value between 0 and 1, we have

$$\sum_{k=1}^{K} \rho_k = 1$$

$$\rho_k \geq 0, \quad \forall k \in \mathcal{K}.$$

Note that the upper bounds of D and D_k ($\forall k \in \mathcal{K}$) can be directly denoted by $\{\boldsymbol{\rho}, \mathbf{p}\}$ using (4.3) and (4.5)-(4.7). Hence, we can release other variables and focus on joint optimal transmission scheduling and power control (JTPC) scheme, $\{\boldsymbol{\rho}^*, \mathbf{p}^*\}$.

4.2 Uncorrelated Source Observation

Substituting (4.3) into (4.5), we can express the upper bound of D_k as $\{g_k, \rho_k\}$

$$D_k \leq W^2 \left[(1 + g_k p_k)^{\rho_k} - 1\right]^{-2} + \sigma_k^2. \tag{4.10}$$

Substituting (4.10) into estimation accuracy constraint (4.8) results in

$$p_k \geq \frac{1}{g_k} \left[\left(\frac{W}{\sqrt{\bar{D}_k - \sigma_k^2}} \right)^{\frac{1}{\rho_k}} - 1 \right], \tag{4.11}$$

Substituting the minimum required p_k in (4.11) into the energy constraint in (4.9), we obtain the inverse of node k's lifetime as

$$u_k(\rho_k) := \frac{1}{T_k} = \frac{\rho_k}{E_k g_k} \left[\left(\frac{W}{\sqrt{\bar{D}_k - \sigma_k^2}} \right)^{\frac{1}{\rho_k}} - 1 \right]. \tag{4.12}$$

The quantity $u_k(\rho_k)$ can also be interpreted as the energy consumption of sensor node k in one time frame normalized by its initial energy. Thus far, the power optimization variables $\{p_k\}$ and the estimation accuracy constraint (4.8) and energy constraint (4.9) are all absorbed. The maximum network lifetime problem formulated in Section 4.1

is now equivalent to minimizing the maximum of $u_k(\rho_k)$ subject to time constraints, which can be rewritten as

$$\begin{aligned} \textbf{P4-1}: \quad \min_{\boldsymbol{\rho}} \quad & \max_k u_k(\rho_k) \\ \text{s.t.} \quad & \sum_{k\in\mathcal{K}} \rho_k = 1, \\ & 0 \le \rho_k \le 1, \ \forall k \in \mathcal{K}. \end{aligned} \tag{4.13}$$

4.2.1 Some Properties of Optimal Solution

Property 1: The optimal solution to **P4-1** satisfies that

$$u_i(\rho_i) = u_j(\rho_j), \ \forall i, j \in \mathcal{K}. \tag{4.14}$$

In other words, the optimal solution forces all node lifetimes to be equal.

Proof: See Appendix II. ■

Property 2: When $\bar{D}_k = \bar{D}_0, \ \sigma_k^2 = \sigma_0^2 \ (\forall k \in \mathcal{K})$, where $\bar{D}_0, \ \sigma_0^2$ are constants, the optimal time sharing fraction only relates to the product of available energy and channel gain. More time sharing fraction and less received power will be given to the node with smaller product. Mathematically,

$$\begin{cases} \left(\frac{\rho_i}{\rho_j} - 1\right)\left(\frac{E_i g_i}{E_k g_j} - 1\right) \le 0 \\ \left(\frac{g_i p_i}{g_j p_j} - 1\right)\left(\frac{E_i g_i}{E_j g_j} - 1\right) \ge 0 \end{cases} \forall i, j \in \mathcal{K} \tag{4.15}$$

Proof: See Appendix III. ■

This result is in agreement with the idea of "lazy scheduling" [23], which achieves energy saving for the "weak" nodes, the ones with small product of channel gain and initial energy, by lowering the transmission power and hence taking longer transmission time.

Property 3: When $\bar{D}_k = \bar{D}_0,\ E_k g_k = E_0 g_0\ (\forall k \in \mathcal{K})$ where $\bar{D}_0$ and $E_0 g_0$ are two constants, the optimal time sharing fraction relates to the observation quality. More time sharing fraction and less received power will be given to the node with low observation quality. Mathematically,

$$\begin{cases} \left(\frac{\rho_i}{\rho_j} - 1\right)\left(\frac{\sigma_i^2}{\sigma_j^2} - 1\right) \geq 0 \\ \left(\frac{g_i p_i}{g_j p_j} - 1\right)\left(\frac{\sigma_i^2}{\sigma_j^2} - 1\right) \leq 0 \end{cases} \quad \forall i, j \in \mathcal{K} \tag{4.16}$$

Proof: See Appendix IV. ■

Remark: Channel realizations are typically independent of the initial energy. Here, the product of channel gain and initial energy provide us with an integrated measurement of them. Further more, Property 2 and Property 3 indicate that the "weak" sensor nodes, the ones with small product of channel gain and initial energy and/or low quality of observation, should take long time for transmission with a low received power. The aim is to give these "weak" nodes priority to take advantage of "lazy scheduling" for energy saving in order to avoid them becoming the bottleneck of the whole WSN.

4.2.2 Partially Distributed Adaptation

We can now transform **P4-1** into a new problem, **P4-2**, which includes the time-sharing variable vector $\boldsymbol{\rho}$ only,

$$\begin{aligned} \textbf{P4-2:}\ \min_{\boldsymbol{\rho}} \quad & \sum_{k=1}^{K} \left(\bar{u}(\boldsymbol{\rho}) - u_k(\rho_k)\right)^2 \\ \text{s.t.} \quad & \sum_{k=1}^{K} \rho_k = 1 \\ & 0 \leq \rho_k \leq 1, \quad \forall k \in \mathcal{K}, \end{aligned}$$

where $\bar{u}(\boldsymbol{\rho}) := (1/K)\sum_k u_k(\rho_k)$. The optimal solution to **P4-2** obviously achieves the maximum network lifetime.

To implement it in a distributed way, instead of solving the problem **P4-2** at the FC, we add a penalty function to relax the constraint. The equivalent problem can then be expressed as:

$$\textbf{P4-3:} \quad \min_{\boldsymbol{\rho}} \quad f(\boldsymbol{\rho}) = \sum_{k=1}^{K} \left[\bar{u}(\boldsymbol{\rho}) - u_k(\rho_k)\right]^2 + \mu\left(1 - \sum_{k=1}^{K} \rho_k\right)^2$$
$$\text{s.t.} \quad 0 \le \rho_k \le 1, \quad \forall k \in \mathcal{K}$$

where μ is a penalty factor and can take any arbitrary positive value.

Theorem 2: Problems **P4-2** and **P4-3** have the same and unique optimal solution. The unique optimal solution is $\boldsymbol{\rho}^*$ if and only if $\nabla f(\boldsymbol{\rho}^*) = 0$.

Proof: See Appendix V. ■

Using gradient projection, the time sharing fractions are adjusted in the opposite direction to the gradient $\nabla f(\boldsymbol{\rho})$, i.e.,

$$\boldsymbol{\rho}(t+1) = [\boldsymbol{\rho}(t) - \gamma \nabla f(\boldsymbol{\rho}(t))]_0^1, \tag{4.17}$$

where γ is a stepsize, and $[y]_a^b := \min\{\max\{y, a\}, b\}$. As seen in (4.17), the time sharing adaptation can be implemented at each sensor node individually in a distributed manner, and the only common information required is $\bar{u}(\boldsymbol{\rho})$ and $\sum_k \rho_k$. Thus, it has a good scalability.

Theorem 3: There exists a constant $\gamma_0 > 0$ such that if $0 < \gamma < \gamma_0$, then $\lim_{t\to\infty} \nabla f(\boldsymbol{\rho}(t)) = 0$. That is, given an arbitrary initial point, $0 < \boldsymbol{\rho}(0) < 1$, the time sharing policy $\boldsymbol{\rho}$ generated by the iterative function (4.17) converges to the unique optimal solution $\boldsymbol{\rho}^*$.

The proof of this theorem uses the following Lemmas.

Lemma 4 (Lipschitz Continuity of $\nabla f(\boldsymbol{\rho})$): The function f is continually differentiable and there exists a constant B such that $\| \nabla f(\boldsymbol{\rho}) - \nabla f(\boldsymbol{\rho}') \|_2 \leq B \| \boldsymbol{\rho} - \boldsymbol{\rho}' \|_2$.

Proof: See Appendix VI. ■

Lemma 5 (Descent Lemma in [6]): If $\nabla f(\boldsymbol{\rho})$ is Liqschitz continuous then $f(\boldsymbol{\rho} + \boldsymbol{\rho}') \leq f(\boldsymbol{\rho}) + (\boldsymbol{\rho}')^T \nabla f(\boldsymbol{\rho}) + 0.5B \| \boldsymbol{\rho}' \|_2^2$.

Define the error term $s_k(t) := [\rho_k(t+1) - \rho_k(t)]/\gamma$. Using Lemma 4 and 5, we can obtain

$$\begin{aligned} f(\boldsymbol{\rho}(t+1)) &\leq f(\boldsymbol{\rho}(t)) + \gamma \mathbf{s}^T(t) \nabla f(\boldsymbol{\rho}(t)) + \gamma^2 \frac{B}{2} \| \mathbf{s}(t) \|_2^2 \\ &\leq f(\boldsymbol{\rho}(t)) - \gamma \left(1 - \frac{B\gamma}{2}\right) \| \mathbf{s}(t) \|_2^2 . \end{aligned}$$

Summing the above inequalities from $t = 0$ yields

$$\begin{aligned} f(\boldsymbol{\rho}(t+1)) &\leq f(\boldsymbol{\rho}(0)) - \gamma \left(1 - \frac{B\gamma}{2}\right) \sum_{t=0}^{t} \| \mathbf{s}(t) \|_2^2 \\ &= f(\boldsymbol{\rho}(0)) - \gamma \left(1 - \frac{\gamma}{\gamma_0}\right) \sum_{t=0}^{t} \| \mathbf{s}(t) \|_2^2, \end{aligned}$$

where $\gamma_0 = 2/B$. Since $f(\boldsymbol{\rho}(t+1)) \geq 0, \forall t$, we obtain

$$\sum_{t=0}^{\infty} \| \mathbf{s}(t) \|_2^2 \leq \frac{1}{\gamma \left(1 - \frac{\gamma}{\gamma_0}\right)} f(\boldsymbol{\rho}(0)) < \infty.$$

This implies that $\lim_{t \to \infty} s(t) = 0$. Combining it with Theorem 2, we can achieve the convergence of $\boldsymbol{\rho}(t)$ generated by (4.17) to the unique optimal point.

4.2.3 Discrete Time Sharing Fraction Assignment

Here, we discuss the implementation issue in a more realistic situation. In the previous analysis, we assume that the time sharing fraction can take arbitrary value between 0 and 1. However, in practical systems, the partition of a time frame is finite. Therefore, the optimization problem turns from continuous to a discrete one.

We assume that a time frame can only be partitioned into M time slots with equal length.

Discrete time sharing fraction assignment falls into the scope of nonlinear combinatorial optimization problems, which insofar do not have a general solution. For this combinatorial optimization problem, there are K^M choices to assign the time slots. An exhaustive search would be a formidable task. Here, we propose a Low-complexity Algorithm of optimal discrete Time-Sharing fraction assignment (LATS). The algorithm only needs M iterations, and each iteration only performs K comparisons.

The basic idea of the algorithm arises from (4.14). That is, each user's normalized energy consumption tends to be identical. Since the node lifetime, $T_k = 1/u_k(\cdot)$, is a monotonically increasing function of ρ_k, each time slot in a time frame can be allocated one by one to the current node with the minimum value of $1/u_k(\rho_k)$. The LATS is described as in detail in Appendix VII.

Fig. 4.2 illustrates the procedure of LATS. Consider a case where a total of 12 time slots in a frame is to be allocated to 4 sensor nodes in a WSN. At the beginning of allocation, the initial value of T_k's are zero, so we randomly allocate the first 4 slots into 4 nodes. As in shown in the first chart in Fig. 4.2, the 4^{th} node has the minimum T_k at this time. Thus, the 5^{th} slot is allocated to node 4. Until all slots are allocated, the network lifetime is the same as the minimum node lifetime, that is, T_{10}. The optimality of this greedy algorithm can be simply proven by the principle of mathematical induction.

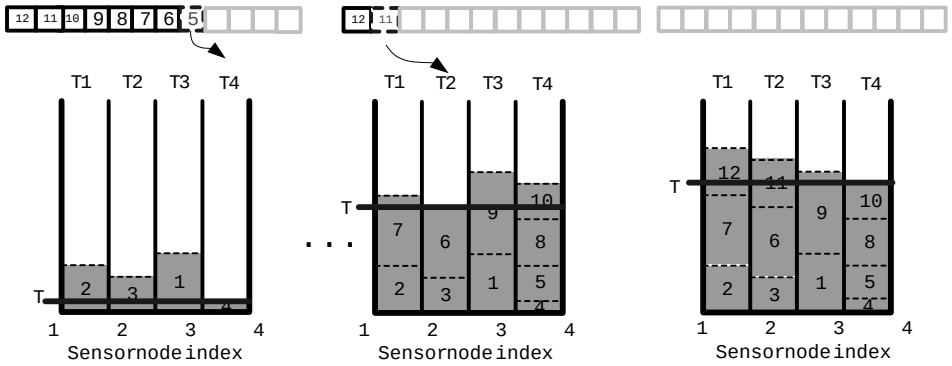

Figure 4.2. Illustrations of LATS

4.3 Common Source Observation

In common source observation, all sensor nodes contribute to a single signal estimation. If we define function $U_k(\cdot)$ as

$$U_k(L_k) := \frac{1}{W^2\left(2^{L_k}-1\right)^{-2}+\sigma_k^2}, \tag{4.18}$$

then, we have $D(\mathbf{L}) \leq 1\big/\left(\sum_{k=1}^{K} U_k(L_k)\right)$ from (4.7). It can be shown that U_k is an increasing and concave function of L_k in the WSN with large observation noise variances. According to the property of convexity preservation[1] [7], the upper bound of D is convex in L_k since the bound is a decreasing and convex function of U_k.

Definition: An average MSE target $\bar{D}$ is said to be a *feasible MSE requirement* if there exists $\{\mathbf{p}, \boldsymbol{\rho}\}$ such that $\bar{D}$ can be achieved. Any feasible MSE requirement $\bar{D}$ must satisfy

$$\bar{D} > D_{\min} = 1 \Big/ \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2}\right), \tag{4.19}$$

where $D_{\min}$ is the smallest MSE in the absence of quantization error. When $\sigma_k^2 = \sigma^2$ $(\forall k \in \mathcal{K})$, we have $D_{\min} = \sigma^2/K$, which describes how the number of sensor nodes enhances the estimation accuracy.

[1]For $f(x) = h(g(x))$, f is convex if h is convex and non-increasing, and g is concave.

4.3.1 Some Properties of Optimal Solution

Property 4: The optimal solution $\{\mathbf{p}^*, \boldsymbol{\rho}^*\}$ to the problem formulated in Section 4.1 for common source observation, denoted as **P4-4**

$$\begin{aligned} \textbf{P4-4}: \quad \max_{\{\mathbf{p},\boldsymbol{\rho}\}} \quad & \min_k \frac{E_k}{\rho_k p_k} \\ \text{s.t.} \quad & D(\mathbf{L}) \leq \bar{D} \\ & \sum_{k\in\mathcal{K}} \rho_k = 1 \\ & \rho_k \geq 0 \end{aligned}$$

satisfies $\rho_k^* p_k^*/E_k = x^*$ $(\forall k \in \mathcal{K})$, where $x^* = \max_k \rho_k^* p_k^*/E_k$. In other words, the normalized energy consumptions in one time frame for all users are identical.

Proof: See Appendix VIII. ■

Since the normalized energy consumption in a time frame is equivalent to the inverse of the lifetime of sensor node, *Property* 4 indicates that all sensor nodes in the network should run out of energy simultaneously. This result is consistent with the one for uncorrelated source observation.

According to *Property* 4, we have $x = 1/T_k$ $(\forall k \in \mathcal{K})$. After substituting $p_k = E_k x/\rho_k$ into (4.9), **P4-4** can be transformed into a new problem, **P4-5**, with optimization variables $\boldsymbol{\rho}$ and x.

$$\begin{aligned} \textbf{P4-5}: \quad \min_{\{\boldsymbol{\rho},\ x\}} \quad & x & (4.20) \\ \text{s.t.} \quad & \sum_{k=1}^{K} U_k(\rho_k, x) \geq \frac{1}{\bar{D}} & (4.21) \\ & \sum_{k=1}^{K} \rho_k = 1 & (4.22) \end{aligned}$$

Here, $U_k(L_k)$ in (4.18) is re-expressed as $U_k(\rho_k, x)$ since L_k is a function of ρ_k and x.

Although it is difficult to obtain a closed-form solution, some properties are pre-

sented to describe the relation between the optimal resource allocation and channel condition, initial power and observation quality.

Property 5: For a network with K users, the following equation

$$\frac{\partial U_i(\rho_i^*, x^*)}{\partial \rho_i} = \frac{\partial U_j(\rho_j^*, x^*)}{\partial \rho_j}, \quad \forall i, j \in \mathcal{K} \tag{4.23}$$

is a necessary condition under which $\boldsymbol{\rho}^*$ and x^* are optimal.

Proof: See Appendix IX. ■

Property 6: When all the nodes have an identical observation quality, i.e., $\sigma_k^2 = \sigma_0^2$ ($\forall k \in \mathcal{K}$), the optimal time sharing fraction only relates to the product of initial energy and channel condition. More time sharing fraction should be given to the user with smaller product of initial energy and channel gain. Mathematically,

$$\begin{cases} \left(\frac{\rho_i}{\rho_j} - 1\right)\left(\frac{E_i g_i}{E_j g_j} - 1\right) \leq 0 \\ \left(\frac{g_i p_i}{g_j p_j} - 1\right)\left(\frac{E_i g_i}{E_j g_j} - 1\right) \geq 0 \end{cases}, \quad \forall i, j \in \mathcal{K}. \tag{4.24}$$

Proof: See Appendix X. ■

Property 7: When all the nodes have an identical product of E_k and g_k, the optimal time sharing fraction only relates to the observation noise variation σ_k^2. More time sharing fraction and low received power should be given to the sensor node with better observation quality. Correspondingly, the sensor node with better observation quality should use higher quantization resolution. Mathematically,

$$\begin{cases} \left(\frac{\rho_i}{\rho_j} - 1\right)\left(\frac{\sigma_i^2}{\sigma_j^2} - 1\right) \leq 0 \\ \left(\frac{g_i p_i}{g_j p_j} - 1\right)\left(\frac{\sigma_i^2}{\sigma_j^2} - 1\right) \leq 0 \\ \left(\frac{L_i}{L_j} - 1\right)\left(\frac{\sigma_i^2}{\sigma_j^2} - 1\right) \leq 0 \end{cases} \quad \forall i, j \in \mathcal{K} \tag{4.25}$$

Proof: See Appendix XI. ■

Property 7 indicates that for prolonging the network lifetime, the optimal solution takes advantage of "opportunistic quantization", that is, the sensor nodes with high observation qualities should increase their quantization resolution.

From *Properties* 4-7, we conclude that in JTPC, more time sharing fraction and lower received power are given to the user with smaller product of initial energy and channel gain and/or to the one with better observation quality. Moreover, the power is adjusted so that all sensor nodes run out of energy at the same time.

4.3.2 Partially Distributed Adaptation

It is observed from the problem **P4-5** that if the optimal x^* is given, the optimal time sharing policy $\boldsymbol{\rho}^*$ obviously satisfies

$$\boldsymbol{\rho}^* = \arg \max_{\substack{\sum_k \rho_k = 1, \\ 0 \le \rho_k \le 1}} \sum_{k=1}^{K} U_k(\rho_k, x^*). \tag{4.26}$$

Therefore, we propose to decompose **P4-5** into two subproblems: (a) binary search for x^*; (b) partially distributed implementation of time-sharing adaptation using gradient projection method. Subproblem (b) is addressed in detail as follows.

For a given x^*, define the Lagrangian

$$\begin{aligned} L(\boldsymbol{\rho}, \lambda) &= \sum_{k=1}^{K} U_k(\rho_k, x^*) - \lambda \left(\sum_{k=1}^{K} \rho_k - 1 \right) \\ &= \sum_{k=1}^{K} \left[U_k(\rho_k, x^*) - \lambda \rho_k \right] + \lambda. \end{aligned}$$

The dual function of (4.26) is

$$F(\lambda) = \max_{\boldsymbol{\rho}} L(\boldsymbol{\rho}, \lambda) = \sum_{k=1}^{K} B_k(\lambda) + \lambda, \tag{4.27}$$

where

$$B_k(\lambda) = \max_{0 \le \rho_k \le 1} U_k(\rho_k, x^*) - \lambda \rho_k. \tag{4.28}$$

If we interpret λ as the price of time sharing fraction, $B_k(\lambda)$ represents the maximum benefit of node k at the given price λ. The K independent subproblems in (4.28) can be solved at each node individually, and the resulting optimal $\rho_k(\lambda)$ is obtained as

$$\rho_k(\lambda) = \left[(U_k')^{-1}(\lambda, x^*)\right]_0^1, \tag{4.29}$$

Here $(U_k')^{-1}(\cdot, x^*)$ is the inverse of $\partial U_k/\partial \rho_k$ with respect to ρ_k which exists since U_k is twice differentiable and strictly concave.

Minimizing dual function in (4.27) can now be readily solved by adjusting the price λ using gradient projection. That is,

$$\lambda(t+1) = \left[\lambda(t) - \gamma\left(\sum_{k=1}^{K}\rho_k - 1\right)\right]^+. \tag{4.30}$$

Here, $\gamma > 0$ is a step size, and $[y]^+ := \max\{y, 0\}$.

The time sharing adaptation and price adjustment have the decentralized nature. It is observed that to implement the update, the FC needs to know the sum of ρ_k's, whereas each node k only needs to know its own function U_k and the price λ. In the partially distributed algorithm, we assume that sensors are numbered and sensor k can reliably monitor the transmission of sensor $k-1$. When sensor $k-1$ finishes transmission, sensor k starts transmission immediately. The FC can obtain the sum of ρ_k's by recording the end time of the transmission of the last node K. The benefit of the partially distributed algorithm is that it dramatically reduces the computational complexity at FC and, moreover, avoids local information report to the FC since FC does not need to know the information $\{E_k, g_k, \sigma_k^2\}$.

Theorem 4: There exists a constant γ_0 such that for any stepsize γ satisfying $0 < \gamma < \gamma_0$, and any initial time sharing fractions $0 \le \boldsymbol{\rho}(0) \le 1$ and price $\lambda(0)$, the sequence $\{\boldsymbol{\rho}(t), \lambda(t)\}$ generated by (4.30) and $\rho_k(t+1) = \rho_k(\lambda(t))$ (see (4.29)) converges to the unique optimal point $(\boldsymbol{\rho}^*, \lambda^*)$.

Theorem 4 holds since the dual objective function F is convex, lower bounded and Lipschitz continuous (see [6, Prop. 3.4]).

Combining the binary search for x^* and partially distributed implementation of time sharing adaptation, the implementation of partially distributed adaptation is described in Appendix XII.

The implementation of binary search can also be realized at the node side only using 1-bit control signal from fusion center since each node only needs to know whether $\sum_k U_k(\rho_k, x) > 1/\bar{D}$ or not.

4.4 Numerical Results

In numerical studies, we compare the network lifetime obtained by our proposed JTPC with the one by several benchmark schemes for uncorrelated source observation and common source observation, respectively.

The simulation is carried out under various network and channel configurations by varying path losses, available energies, and observation noise variances. The channel gain is modelled as $g = \alpha d^{-n}$, where d^{-n} is the effect of path loss, and α is an attenuation constant. Here, the distance between sensor node k and the FC, d_k, is randomly generated following a uniform distribution over the region [1,10]. Similar to [72], observation noise variances σ_k^2 ($\forall k \in \mathcal{K}$) are generated according to the distribution $0.05(1+aY^2)$, where Y^2 follows the exponential distribution. In addition, we set the available energies $E_k = E^{\min}(1+bY^2)$, where $E^{\min}$ is the minimum energy of all nodes. Through adjusting the parameters n, a, and b, we generate g_k, E_k and σ_k^2 for all $k \in \mathcal{K}$ to model different sensing environments. In the simulation, we set the target MSE, $\bar{D}_k = 2\sigma_k^2$ $\forall k \in \mathcal{K}$. We use the normalized deviation $\sqrt{\mathbb{V}(t)}/\mathbb{E}(t)$

defined in [72] to measure the heterogeneity of a variable t, where $\mathbb{V}(t)$ denotes the variance of t and $\mathbb{E}(t)$ the mean of t.

4.4.1 Uncorrelated source observation

First we consider uncorrelated source observation. We compare the network lifetime of our proposed scheme with the following two transmission schemes:

- *Uniform TDMA with power control* (UTP): It can be viewed as a suboptimal solution to **P4-1** by fixing $\rho_k = 1/K,\ \forall k \in \mathcal{K}$. The optimal transmission power p_k under the constraint (4.5) can be obtained from (4.11).

- *Inverse-log scheduling* (ILS) [73]: The total energy in a sensor network is minimized by varying the transmission times assigned to different sensor nodes. The time-sharing fraction in the ILS scheme has an inverse-log form,

$$\rho'_k = \frac{R_k}{\log(\lambda g_k)}, \quad \forall k \in \mathcal{K}, \tag{4.31}$$

  where λ is determined by the constraint $\sum_k \rho'_k = 1$, and it can be obtained by numerical search.

Fig. 4.3 compares the network lifetime (in time frames) of JTPC, UTP and ILS by choosing the simulation parameters $n = 3,\ a = 1$, and $b = 0$, and varying the number of sensor nodes from 10 to 50. It is shown that the network lifetimes decrease exponentially with the number of sensor nodes. This is because the increase of the number of sensor nodes reduces the average transmission times of each node when the duration of a time frame is fixed. Correspondingly, the energy consumption increases dramatically. This observation suggests that partitioning a WSN into several subnetworks could prolong the network lifetime at the expense of increasing the number of FC. The performance of ILS is a little worse than that of UTP. It is because

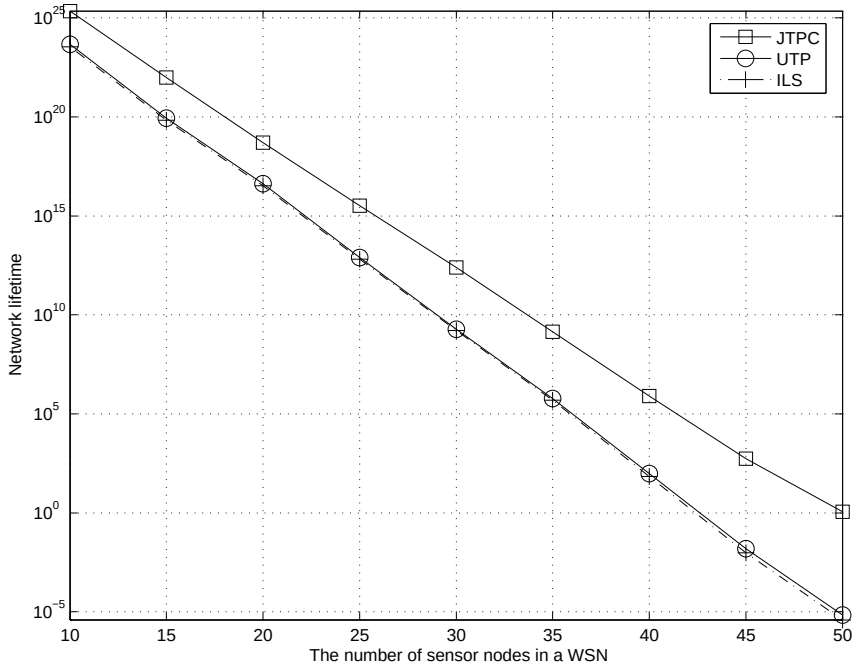

Figure 4.3. The network Lifetimes of JTPC, UTP and ILS vs. the number of sensor nodes

minimizing the total energy consumption in ILS leads to some nodes running out of energy quickly. It is also observed that the lifetime gains of JTPC over UTP and ILS increase exponentially with the number of nodes. Here, we define the lifetime gain of one scheme over another as the ratio of network lifetimes obtained by the two schemes. The following figures illustrate the lifetime gains obtained by the proposed scheme under different sensing environments.

Fig. 4.4 plots the lifetime gain of JTPC over the UTP and ILS schemes versus normalized deviation of channel path losses by varying the path loss exponent n. Both σ_k^2 and E_k are fixed by letting $a = b = 0$. Fig. 4.5 is plotted by fixing $d_k = 5.5$, $\forall k \in \mathcal{K}$, $b = 0$ and varying a. It is observed that the lifetime gains increase when observation noise variances become more heterogeneous. In addition, for fixed normalized deviation of noise variances, the lifetime gain rises as the number of sensors increases. The relationship between the lifetime gains and the normalized

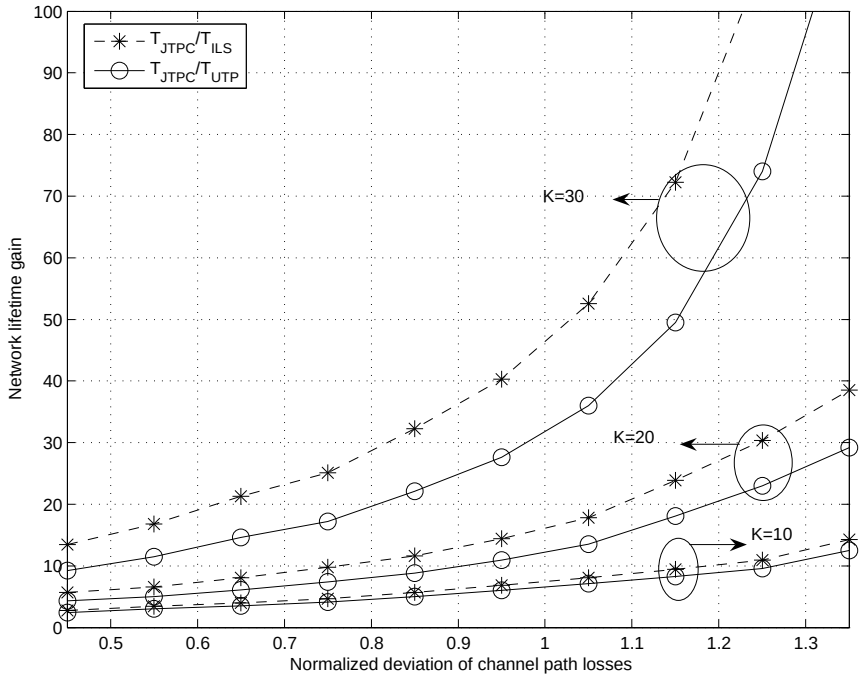

Figure 4.4. Lifetime gains of JTPC over UTP and ILS vs. normalized deviation of channel path losses

deviation of initial energies is shown in Fig. 4.6. A trend similar to that of Figs. 4.4 and 4.5 is observed.

4.4.2 Common source observation

To compare with JTPC under common source observation, We consider two schemes:

1. *Uniform TDMA with power control* (UTP): It is a suboptimal solution of the proposed JTPC by letting $\rho_k = 1/K, \forall k \in \mathcal{K}$. The best power allocation $\mathbf{p}^*$ can be obtained from *Property* 4. By letting $T_k = T,\ \forall k \in \mathcal{K}$, $\mathbf{p}$ can be expressed a function of T. Thus the inequality in (4.21) only has one variable T, the minimum value of which can be obtained by numerical search.

2. *Power scheduling (PS)* [69]: All nodes also have a same time sharing fraction, but the power is adjusted to minimize the L^2-norm of transmission power vector.

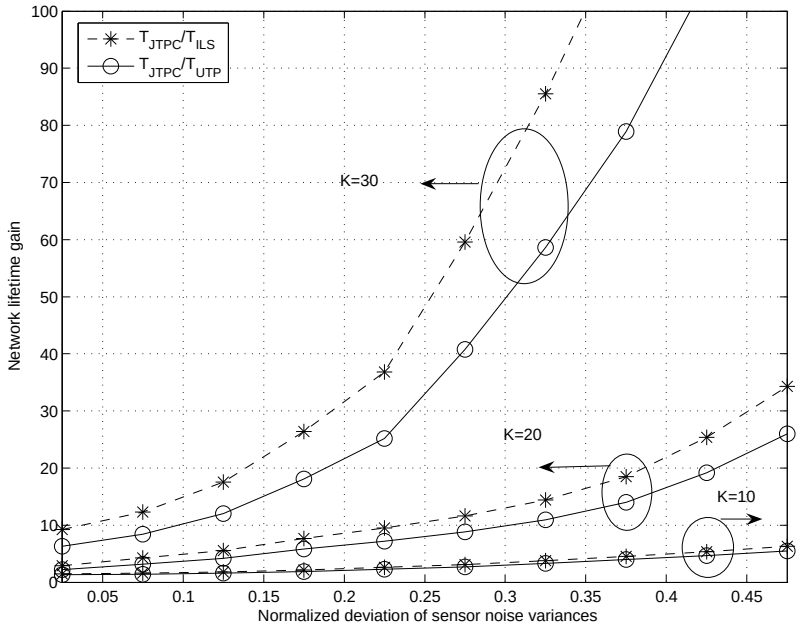

Figure 4.5. Lifetime gains of JTPC over UTP and ILS vs. normalized deviation of observation noise variances

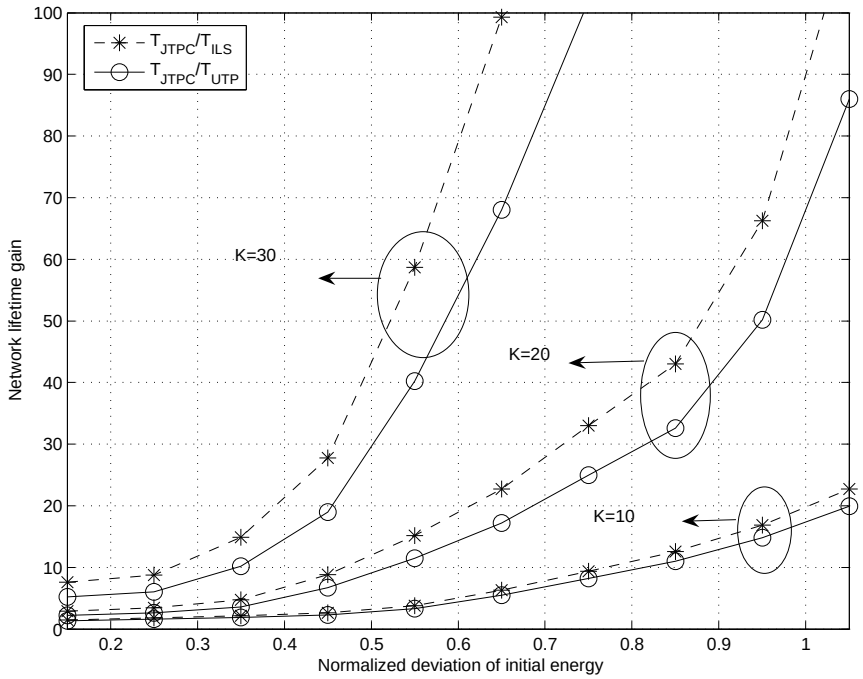

Figure 4.6. Lifetime gains of JTPC over UTP and ILS vs. normalized deviation of initial energies

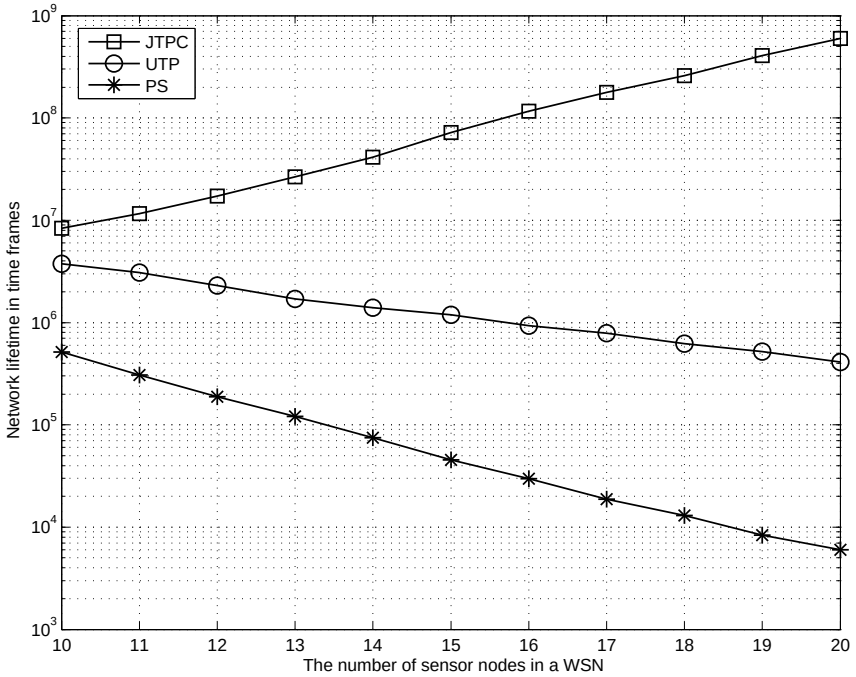

Figure 4.7. The network Lifetimes of JTPC, UTP and PS vs. the number of sensor nodes

Fig. 4.7 compares the network lifetimes (in time frames) of JTPC, UTP and PS schemes by letting $n = 3,\ a = 1$, and $b = 0$, and varying the number of sensor nodes from 10 to 30. The MSE requirement $\bar{D}$ is fixed to be 0.01. It is observed that the network lifetimes of JTPC increases exponentially with the number of sensor nodes. Since transmission time of each user, $\rho = 1/K$, in UTP and PS, deceases linearly with the number of sensor nodes, the increment of number of sensor nodes can not prolong the lifetime of the WSN. The performance of PS is worse than that of UTP since PS suggests that the nodes with good channels and/or observation qualities should increase their quantization resolutions which may lead them to run out of energy quickly.

In the remaining simulations, we study the lifetime gains (defined as the ratio of lifetime obtained by JTPC and that by UTP or PS) of JTPC over UTP and PS under different sensing environments. The MSE requirement $\bar{D}$ is fixed to be $2D_{\min}$, where $D_{\min}$ is the MSE of estimation without quantization noise, and is given in (4.19).

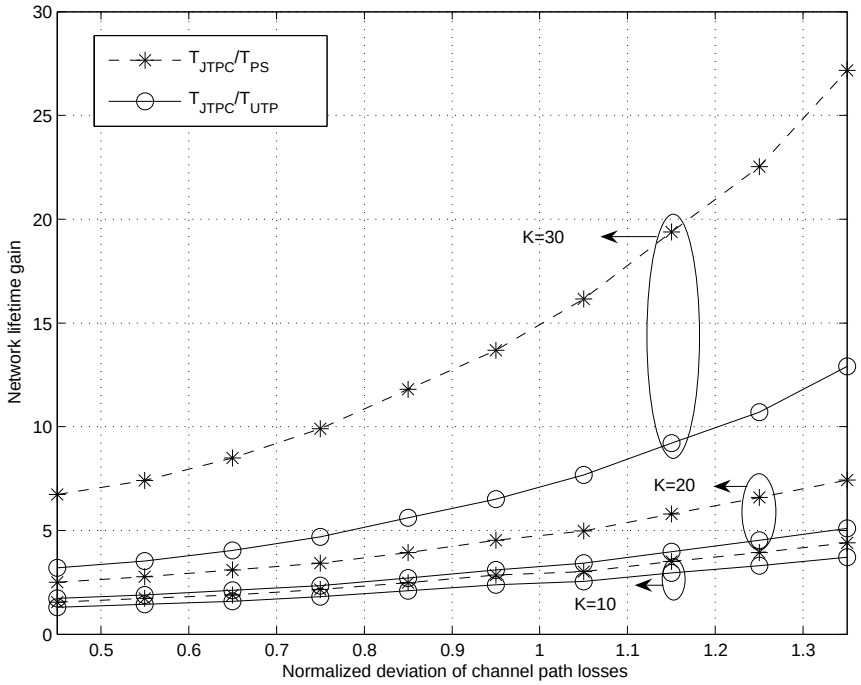

Figure 4.8. Lifetime gains of JTPC over UTP and PS vs. normalized deviation of channel path losses

In Fig. 4.8, we fix σ_k^2 and E_k by choosing $a = b = 0$. The normalized deviation of channel path losses can be changed by adjusting n ($2 \leq n \leq 4$). Fig. 4.8 plots the lifetime gain of JTPC over UTP and PS versus normalized deviation of channel path losses. Fig. 4.9 is plotted by fixing $d_k = 5.5$, $\forall k \in \mathcal{K}$, $b = 0$ and varying a. In Fig. 4.9, the lifetime gain of JTPC over UTP and PS increases when observation noise variances become more heterogeneous. In addition, for fixed normalized deviation noise variances, the lifetime gain increases as the number of users increases. The relation between the lifetime gain and the normalized deviation of initial energies is shown in Fig. 4.10.

From the numerical results, we conclude that JTPC brings significant lifetime gain over UTP and PS scheme in a heterogeneous WSN environment, and the performance gain increases with the number of sensor nodes.

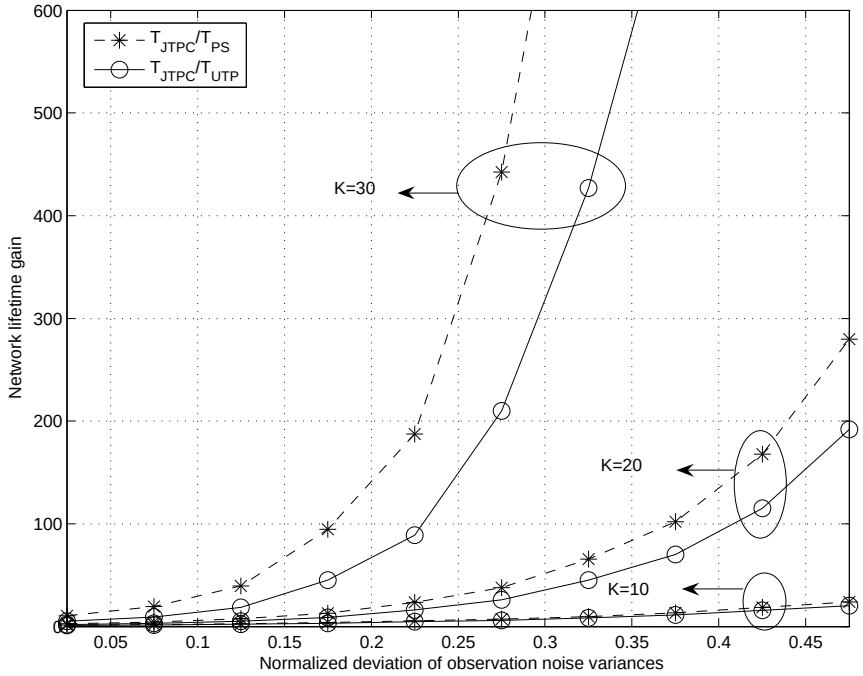

Figure 4.9. Lifetime gains of JTPC over UTP and PS vs. normalized deviation of observation noise variances

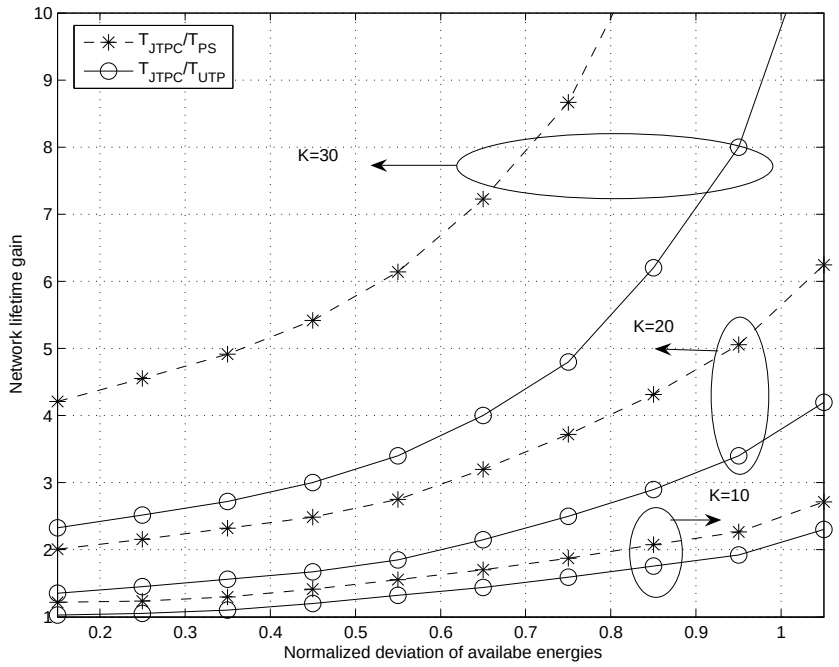

Figure 4.10. Lifetime gains of JTPC over UTP and PS vs. normalized deviation of initial energies

4.5 Conclusions

We have considered a joint quantization and transmission design to prolong the lifetime for cluster-based WSNs under the average estimation accuracy constraints. We focused on single-hop data collection from sensor nodes to a common FC in a particular cluster where sensor nodes observe spatially uncorrelated signals or a common signal. After studying the properties of optimal solutions, it is observed that:

1. For uncorrelated source observation, the product of channel gain and initial energy provide us with an integrated measurement of them. Further more, it is suggested that the "weak" sensor nodes, the ones with small product of channel gain and initial energy and/or low quality of observation, should take long time for transmission with a low received power. The aim is to give these nodes priority to take advantage of "lazy scheduling" for energy saving in order to avoid becoming the bottleneck of the whole WSN.

2. For common source observation, more time sharing fraction and lower received power are given to the user with smaller product of initial energy and channel gain and/or to the one with better observation quality. Moreover, the power is adjusted so that all sensor nodes run out of energy at the same time. The optimal solution not only takes advantage of "lazy scheduling" but also "opportunistic quantization", that is, the sensor nodes with high observation qualities should increase their quantization resolution.

3. For both scenarios, numerical examples show that significant lifetime gain can be achieved when compared with three existing schemes, and the gain becomes more significant when sensing environment becomes more heterogeneous and the number of nodes increases.

Optimal resource allocation in WSNs, while being efficient in prolonging network lifetime, can come with high complexity in design and analysis. The statement can also be generalized to most of wireless systems. Therefore, beside studying the optimal resource allocation solution, the main objective of this chapter is to develop a partially distributed algorithm where the JTPC policy can be calculated at each node with the local information of the path loss, available energy and observation quality, considering the limited initial energy and computational capacity of FC in practice.

A challenge of the future work would be to extend JTPC policies to a more generic distributed estimation framework discussed in [72] where the observed signals could be partially correlated. Also the definition of network lifetime considered in this chapter is very simple. In a real WSN, the death of some nodes may not affect the functioning of the whole network. More generalized definitions of network lifetime, for example, the time when the L^{th} $(L < K)$ node dies, are needed.

Chapter 5

End-to-End Average Rate Maximization in Linear OFDM Based Relay Networks

In Chapter 3 and 4, dynamic resource allocation is studied in a single-hop wireless system over a common channel. Our focus in this chapter is in the context of an OFDM based multi-hop network consisting of a one-dimensional chain of nodes including a source, a destination, and multiple relays. The problem is to maximize the end-to-end average transmission rate under a long-term total power constraint by adapting the transmission power on each subcarrier over each hop and the transmission time used by each hop in every time frame. The solution to the problem is derived by decomposing it into two subproblems: short-term time and power allocation given an arbitrary total power constraint for each channel realization, and total power distribution over all channel realizations. We show that the optimal solution has the following features: the power allocation on subcarriers over each hop has the water-filling structure and a higher water level is given to the hop with relatively

poor channel condition; meanwhile, the fraction of transmission time allocated to each hop is adjusted to keep the instantaneous rates over all hops equal. To reduce computational complexity and signalling overhead, three suboptimal resource allocation algorithms are also proposed. Numerical results are presented for different network settings and channel environments.

Future wireless systems expect a coverage enhancement and a throughput growth at a low cost. This leads to increasing attentions to the concept of relaying in wireless networks such as next generation cellular networks, broadband wireless metropolitan area networks (WMANs) and wireless local area networks (WLANs) [8; 52]. For example, the mesh network architecture is proposed in IEEE 802.11 networks [8], and the multihop relay scheme is to be developed for IEEE 802.16 standard [52]. In recent years, much academic work focuses on ad hoc networks, which can be viewed as generalized relay networks, where each network node may help in relaying each other's data packets. Gupta *et al.* study the asymptotic bounds of transmission rates with a large number of hops under various network topologies and node capabilities in [30] and [31]. These results, however, do not reveal the actual capacity of a network with a given number of nodes, especially when the number is small. More recently, cooperative communications among the source node and relay nodes in a two-hop scenario are proposed to increase network capacity [40; 59]. The basic idea of cooperative relaying is to achieve diversity through independent channels by controlling how the transmitted signals add up at the receiver. Note that cooperative relaying requires precise timing (and possibly phase) synchronization among different nodes which makes it difficult to be integrated in many practical systems. As such, a relay network with a finite number of hops and without synchronous cooperation is of most practical interest.

Relay networks have two possible architectures: a tree of point to multiple points (PMP) architecture and a mesh one. When only a single route is active at any

particular time duration, a two-dimensional network can be viewed as multiple one-dimensional chains of nodes. Such a one-dimensional chain of nodes, a so-called linear network as in [61], consists of one pair of source and destination nodes and several intermediate relay nodes that are aligned on the straight line between the source and the destination. Transmissions can only occur between two neighboring nodes. In [50], Oyman *et al.* introduce two different transmission strategies over multiple hops: fixed-rate relaying and rate-adaptive relaying, and showed merits of multi-hop relaying over frequency-flat fading channels. In this work, we consider orthogonal frequency division multiplexing (OFDM) based linear multi-hop networks over broadband wireless channels. OFDM is an underlying transmission technology to overcome the inter-symbol interference imposed by multipath. In fact, the OFDM-based relaying architecture has been accepted by current wireless standards, e.g. IEEE 802.16j. Moreover we adopt time division (TD) for transmissions over different hops so that there is no interference between nodes. Further, time is divided into frames of multiple time slots and the transmission parameters are determined at the beginning of each frame. The transmission parameters include the transmission power on each subcarrier over each hop and the transmission time over each hop, as shown in Fig. 5.1 (For example, 8, 2 and 6 time slots are allocated to hop 1, 2 and 3, respectively, in time frame 1. $p_{k,n}$ is used to indicate transmission power on subcarrier k over hop n). When the scheduler knows the global channel state information (CSI), it can perform adaptive relaying. This global information gives a potential for enhancing performance through dynamic resource allocation.

In this chapter, we aim to derive the end-to-end maximum average transmission rate in the OFDM-based linear multi-hop network with adaptive relaying. The problem is first formulated as a max-min problem. We then decompose it into two sub-problems: (a) Determine the power and time allocation to maximize the end-to-end instantaneous transmission rate under a given short-term total power constraint for

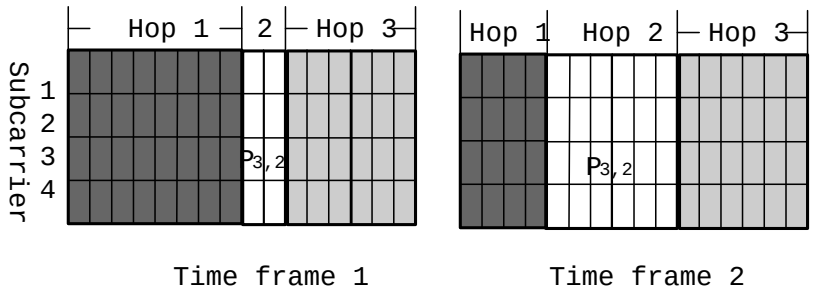

Figure 5.1. Illustration of the transmission scheme for an OFDM-based relaying system with 4 subcarriers and 3 hops

each channel realization; (b) Determine the short-term total power constraint for each channel realization so that the end-to-end average transmission rate is maximized under a long-term total power constraint. Through theoretical analysis, we show that the optimal resource allocation strategy has the following features: the power allocation on the subcarriers at each hop follows the water-filling structure, and the water level varies over time and among different hops. Meanwhile, the fraction of transmission time allocated to each hop is adaptive so that the actual transmission volumes over all hops are equal. The analytical expression for the optimal resource allocation also suggests that the hops with bad channel condition should be given a high water level, and if the channel of any one hop is in deep fade during a certain time frame, the system may be turned off in order to save power. The proposed optimal allocation scheme determines the upper bound of the end-to-end average transmission rate, but also incurs high computational complexity and high signaling overhead. We also propose three suboptimal algorithms with low complexity and reduced overhead. The tradeoff between performance, complexity and overhead is discussed.

The remainder of this chapter is organized as follows. The system model is presented in Section 5.1, followed by the problem formulation in Section 5.2. The optimal time sharing fractions and power allocation are derived in Section 5.3. In Section 5.4, three suboptimal resource allocation algorithms are introduced. Numerical results are given in Section 5.5. Finally, we conclude this chapter in Section 5.6.

5.1 System Model

We consider a linear multi-hop wireless network consisting of a source node R_0, and a destination node R_N and $N-1$ relay nodes R_n with $n = 1, \ldots, N-1$. As shown in Fig. 5.2, the relays are uniformly spaced along a straight line from the source to destination. The hop between node R_{n-1} and R_n is indexed by n, and the set of hops is denoted by $\mathcal{N}$. The transmission time is divided into frames of multiple time slots as shown in Fig. 5.1. In general, multiple nodes can transmit in a same time slot. However, it will result in interference to each other and increase decoding complexity. In addition, it is not easy for a relay to transmit and receive simultaneously. Thus, we forbid multiple nodes transmitting simultaneously over a same time slot. In each time frame, the message from the source is sequentially relayed at each hop using decode-and-forward scheme [40]. That is, each relay decodes the message transmitted by the previous node, re-encodes it, and then forwards it to the next relay. Over each hop, OFDM with K subcarriers is used as the physical layer transmission technology so that the channel on each subcarrier can be treated as a flat fading channel. The set of subcarriers is denoted by $\mathcal{K}$. We assume that the channel between two nodes is a block fading Gaussian channel, and the channel gain on subcarrier k over hop n, denoted by $g_{k,n}$, remains constant during the entire time frame but changes randomly from one frame to another.

Let the transmission power on subcarrier k over hop n be denoted as $p_{k,n}$ and the average total transmission power be constrained by P. Let T_n represent the number of OFDM symbols that hop n is scheduled to transmit a total of T OFDM symbols in a time frame and define time-sharing fraction $\rho_n \triangleq \frac{T_n}{T}$[1]. The instantaneous transmission rate (Nat/OFDM symbol) in a time frame achieved by subcarrier k over

[1]We assume the number of time slots in a time frame is large enough so that ρ_n can take arbitrary value between 0 and 1

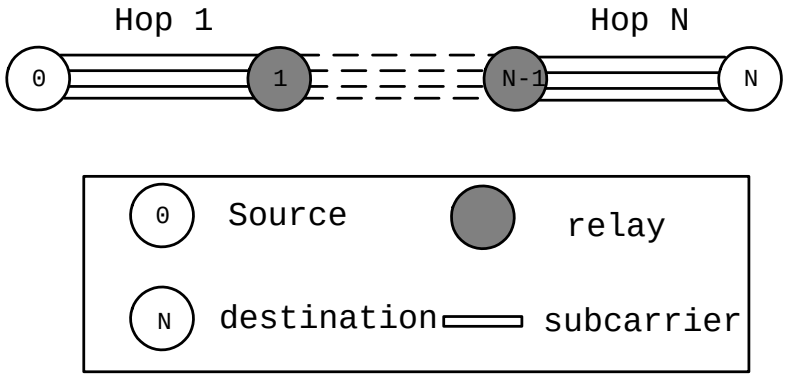

Figure 5.2. Illustration of linear multi-hop networks

hop n is expressed as

$$r_{k,n} = \rho_n \ln\left(1 + \frac{g_{k,n}p_{k,n}}{\Gamma N_0}\right), \tag{5.1}$$

where N_0 is the noise power, and Γ is the signal-to-noise ratio (SNR) gap for characterizing the difference between the SNR needed to achieve a certain data rate for a practical system and the theoretical limit. For notation brevity, in the remaining part of this chapter, we redefine $g_{k,n}$ as the normalized channel gain $g_{k,n} := g_{k,n}/\Gamma N_0$, and let $\mathbf{g} = \{g_{k,n},\ k \in \mathcal{K}, k \in \mathcal{N}\} \in \mathcal{G}$, where $\mathcal{G}$ is the set of all possible normalized channel states. The total instantaneous transmission rate over hop n is denoted by $r_n = \sum_{k\in\mathcal{K}} r_{k,n}$. Under the assumption that no data is allowed to accumulate at any of relay nodes, the end-to-end instantaneous transmission rate is equivalent to the minimum of instantaneous transmission rates over N hops [50; 51]. That is, $r = \min_{n\in\mathcal{N}} r_n$.

5.2 Problem Formulation

In this chapter, we consider the scenario where the delay requirement of the application is much longer than the coherence time of the fading channel and, hence, the channel codeword can span multiple time frames. The end-to-end average trans-

mission rate is used to capture the system performance for such delay-tolerant applications. Under the assumption that the channel processes are ergodic, time-average rate is equal to rate averaged over all channel states. In the remainder of this chapter, we will use the term *average rate* for short to distinguish it from the instantaneous rate defined in a particular time frame. We focus on the situations where the current state of the channels over every hop is available at all the transmitters and receivers. Dynamic time and power allocation in response to the channel variations is considered in the maximization of the end-to-end average rate. More specifically, for a given average power constraint P, the problem is to determine the optimal transmission power set, $\mathbf{p}(\mathbf{g}) = \{p_{k,n}, k \in \mathcal{K}, n \in \mathcal{N}\}$, and time-sharing fraction set, $\boldsymbol{\rho}(\mathbf{g}) = \{\rho_n, n \in \mathcal{N}\}$, as functions of global CSI, $\mathbf{g} = \{g_{k,n}, k \in \mathcal{K}, n \in \mathcal{N}\}$. This functional optimization problem is formulated as

P5-1 :

$$\begin{aligned}
\max_{\mathbf{p}(\mathbf{g}),\ \boldsymbol{\rho}(\mathbf{g})} \quad & R\big(\boldsymbol{\rho}(\mathbf{g}), \mathbf{p}(\mathbf{g})\big) \\
\triangleq\ & \mathbb{E}\left[\min_{n\in\mathcal{N}} r_n\right] \\
=\ & \int_{\mathbf{g}} \left[\min_{n\in\mathcal{N}} \rho_n(\mathbf{g}) \sum_{k\in\mathcal{K}} c_{k,n}\big(p_{k,n}(\mathbf{g})\big)\right] pdf(\mathbf{g})\mathrm{d}\mathbf{g}, \\
\text{s.t.} \quad & \mathbb{E}\left(\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}} \rho_n(\mathbf{g}) p_{k,n}(\mathbf{g})\right) \leq P \qquad (5.2) \\
& \sum_{n\in\mathcal{N}} \rho_n(\mathbf{g}) = 1
\end{aligned}$$

where

$$c_{k,n}(x) \triangleq \ln\left(1 + g_{k,n}x\right) \qquad (5.3)$$

and $pdf(\mathbf{g})$ is the joint probability density function (pdf) of the global CSI $\mathbf{g}$.

When $N = 1$ and $K = 1$, the above problem reduces to characterizing transmission schemes consisting of the direct transmission from source to destination with a single

subcarrier. The optimal R^* would be the maximum average rate for the single hop transmission in a flat fading channel, and the solution is given by the classic water-filling scheme [25].

5.3 Optimal Resource Allocation

In this section, we derive an optimum time and power allocation for **P5-1**. The difficulty in solving the problem arises from the max-min nature of the objective and the co-existence of both long-term and short-term constraints. The long-term power constraint makes it possible to adapt the power not only over hops within one frame but also over frames in the time domain. To make this problem more tractable, we decompose the original problem **P5-1** into two subproblems.

1. *Short-term time and power allocation given an arbitrary total power constraint for each channel realization*

   Consider any given channel state realization $\mathbf{g} \in \mathcal{G}$. Assume that the total power over all N hops assigned to this channel state realization is p. The objective is to maximize the end-to-end instantaneous transmission rate. Mathematically, it can be expressed as

$$\begin{aligned} \textbf{P5-2}: \quad & r\big(\mathbf{g}, p\big) \\ \triangleq \quad & \max_{\boldsymbol{\rho},\ \mathbf{p}} \min_{n\in\mathcal{N}} \left[\rho_n \sum_{k\in\mathcal{K}} c_{k,n}\big(p_{k,n}\big)\right] \\ \text{s.t.} \quad & \sum_{n\in\mathcal{N}} \rho_n \sum_{k\in\mathcal{K}} p_{k,n} = p \qquad (5.4) \\ & \sum_{n\in\mathcal{N}} \rho_n = 1. \qquad (5.5) \end{aligned}$$

2. *Total power distribution $p(\mathbf{g})$ over all channel realizations*

   The second problem is to determine the instantaneous total power constraint

function $p(\mathbf{g})$ over all N hops for each channel realization $\mathbf{g}$ such that the end-to-end transmission rate averaged over all channel states is maximized. That is,

$$\textbf{P5-3}: \max_{p(\mathbf{g})} \quad \mathbb{E}\Big[r\big(\mathbf{g}, p(\mathbf{g})\big)\Big] \tag{5.6}$$

$$\text{s.t.} \quad \mathbb{E}[p(\mathbf{g})] \leq P. \tag{5.7}$$

where $r\big(\mathbf{g}, p(\mathbf{g})\big)$ is defined in **P5-2**.

Proposition 1: The problem **P5-1** is equivalent to **P5-2** when p in (5.4) is given by the solution $p^*(\mathbf{g})$ to **P5-3**.

Proof: See Appendix XIII. ■

We solve subproblem **P5-2** and **P5-3**, separately, in the following two subsection.

5.3.1 Short-Term Time and Power Allocation

Problem **P5-2** is a max-min problem with inseparable variables, but it can be transformed into a equivalent convex optimization problem given by

$$\textbf{P5-4} \max_{\{r,\ \boldsymbol{\rho},\ \mathbf{s}\}} \quad r$$

$$\text{s.t.} \quad \rho_n \sum_{k\in\mathcal{K}} \ln\left(1 + \frac{g_{k,n}s_{k,n}}{\rho_n}\right) \geq r,\ \forall n \tag{5.8}$$

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} s_{k,n} = p \tag{5.9}$$

$$\sum_{n\in\mathcal{N}} \rho_n = 1 \tag{5.10}$$

where r is a new variable and $s_{k,n} \triangleq \rho_n p_{k,n}$. It can be shown that $\rho_n \ln(1+g_{k,n}s_{k,n}/\rho_n)$ is concave in both ρ_n and $s_{k,n}$. Thus, **P5-4** is a convex optimization problem. The

Lagrangian of this problem is given by

$$\begin{aligned} & J(r, \boldsymbol{\rho}, \mathbf{s}, \boldsymbol{\mu}, \lambda_0, \nu) \\ & = r + \sum_{n \in \mathcal{N}} \mu_n \left[\rho_n \sum_{k \in \mathcal{K}} \ln \left(1 + \frac{g_{k,n} s_{k,n}}{\rho_n} \right) - r \right] + \\ & \lambda_0 \left(p - \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} s_{k,n} \right) + \nu \left(1 - \sum_{n \in \mathcal{N}} \rho_n \right), \end{aligned} \tag{5.11}$$

where $\boldsymbol{\mu}$, λ_0 and ν are all positive and represent the Lagrange multipliers associated with the constraints (5.8), (5.9) and (5.10), respectively.

Since this problem is convex, the following KKT conditions are necessary and also sufficient for the optimal solution,

$$\begin{aligned} \frac{\partial J(\ldots)}{\partial r} &= 1 - \sum_{n \in \mathcal{N}} \mu_n \\ &= 0, \ \forall k \in \mathcal{K}, n \in \mathcal{N} \end{aligned} \tag{5.12}$$

$$\begin{aligned} \frac{\partial J(\ldots)}{\partial s_{k,n}} &= \frac{\mu_n g_{k,n}}{1 + g_{k,n} s_{k,n} / \rho_n} - \lambda_0 \\ &= 0, \ \forall k \in \mathcal{K}, n \in \mathcal{N} \end{aligned} \tag{5.13}$$

$$\begin{aligned} \frac{\partial J(\ldots)}{\partial \rho_n} &= \mu_n \left\{ \sum_{k \in \mathcal{K}} \left[\ln \left(1 + \frac{g_{k,n} s_{k,n}}{\rho_n} \right) - \right. \right. \\ & \left. \left. \frac{g_{k,n} s_{k,n}}{\rho_n + g_{k,n} s_{k,n}} \right] \right\} - \nu = 0, \ \forall n. \end{aligned} \tag{5.14}$$

From (5.13), we obtain the optimal power allocation as

$$p^*_{k,n} = \frac{s_{k,n}}{\rho_n} = \left(\frac{\mu_n}{\lambda_0} - \frac{1}{g_{k,n}} \right)^+, \tag{5.15}$$

where $(\cdot)^+ = \max(0, \cdot)$.

Substituting (5.15) into (5.14), we obtain the condition for the Lagrange multipliers,

$$\begin{aligned} & \nu \triangleq h_n(\mu_n, \lambda_0, \mathbf{g}) \\ & = \mu_n \cdot \sum_{k \in \mathcal{K}} \left\{ \left[\ln \left(\frac{\mu_n g_{k,n}}{\lambda_0} \right) \right]^+ - \left(1 - \frac{\lambda_0}{\mu_n g_{k,n}} \right)^+ \right\} \end{aligned} \tag{5.16}$$

for all $n \in \mathcal{N}$. From (5.16), for a fixed λ_0 and ν, finding the optimal $\boldsymbol{\mu}$ are N independent problems. It can be shown easily that the derivative of the function $h_n(\mu_n, \lambda_0, \mathbf{g})$ with respect to μ_n is positive when $\mu_n \geq \min_k\{\lambda_0/g_{k,n}\}$ and, thus, $h_n(\mu_n, \lambda_0, \mathbf{g})$ is monotonically increasing in μ_n. The inverse function $h_n^{-1}(\nu, \lambda_0, \mathbf{g})$ thus exists and is an increasing function of ν. As a result, the exact value of μ_n for a given ν can be obtained numerically using binary search. Substituting $\mu_n = h_n^{-1}(\nu, \lambda_0, \mathbf{g})$ into (5.12), we have

$$\sum_{n\in\mathcal{N}} h_n^{-1}(\nu, \lambda_0, \mathbf{g}) = 1. \tag{5.17}$$

The left side of (5.17) is monotonically increasing in ν, thus the optimal ν can also be obtained via binary search from (5.17). In other words, for a given λ_0, the optimal power allocation can be obtained through two-nested binary search. The outer loop varies the Lagrange multiplier ν to meet (5.17). The inner loop searches μ_n at a given value of ν to satisfy (5.16). There exists a unique λ_0 such that a positive total power constraint is satisfied.

Proposition 2: The optimal solution to **P5-2** forces the instantaneous rates over all hops, r_n's, at each time frame to be equal, i.e.,

$$\rho_n \sum_{k\in\mathcal{K}} \ln\left(1 + \frac{g_{k,n}s_{k,n}}{\rho_n}\right) = \rho_i \sum_{k\in\mathcal{K}} \ln\left(1 + \frac{g_{k,i}s_{k,i}}{\rho_i}\right),$$

for all $n \neq i$.

Proof: This proposition can be easily obtained by using the KKT conditions in the convex problem **P5-4**. An alternative proof is given in Appendix XIV. ■

Based on *Proposition* 2 and the time-sharing constraint (5.10), each ρ_n can be expressed as

$$\rho_n^* = \frac{\prod_{i\neq n}\sum_{k\in\mathcal{K}}\left[\ln\left(\frac{\mu_i g_{k,i}}{\lambda_0}\right)\right]^+}{\sum_{i\in\mathcal{N}}\prod_{j\neq i}\sum_{k\in\mathcal{K}}\left[\ln\left(\frac{\mu_j g_{k,j}}{\lambda_0}\right)\right]^+}, \quad \forall n \in \mathcal{N}. \tag{5.18}$$

From (5.18), a more meaningful expression of ρ_n^* is obtained as

$$\rho_n^* = \frac{1}{N} \cdot \frac{\bar{c}}{c_n}, \quad \forall n \in \mathcal{N}. \tag{5.19}$$

Here, $c_n = \sum_{k\in\mathcal{K}} [\ln(\mu_n g_{k,n}/\lambda_0)]^+$ represents the achievable rate of hop n given the power allocation in (5.15) without time sharing, and $\bar{c} = N/\sum_{n\in\mathcal{N}} \frac{1}{c_n}$ is the harmonic mean of $\{c_n\}$ over all the N hops. Now it is clear that the optimal time fraction allocated to each hop is equal to the fraction $1/N$ scaled by the inverse of the normalized achievable rate on this hop with respect to the harmonic mean of the achievable rates on all hops.

Next, we find out the relation between the maximum end-to-end instantaneous transmission rate r and the short-term total transmission power consumption p. We assume that $\{r^*, \boldsymbol{\rho}^*, \mathbf{s}^*\}$ is the solution to **P5-4**. Define $\mathcal{K}_n$ as the set of active subcarriers ($p_{k,n} > 0$) over hop n, and k_n as the size of the set $\mathcal{K}_n$. According (5.15) and (5.9), the short-term total transmission power can be expressed as

$$p = \sum_{n\in\mathcal{N}} \rho_n \sum_{n\in\mathcal{K}_n} \left(\frac{\mu_n}{\lambda_0} - \frac{1}{g_{k,n}} \right). \tag{5.20}$$

From the above equation, we have

$$\frac{1}{\lambda_0} = \left(p + \sum_{n\in\mathcal{N}} \rho_n \sum_{k\in\mathcal{K}_n} \frac{1}{g_{k,n}} \right) \frac{1}{\sum_{n\in\mathcal{N}} k_n \mu_n \rho_n}. \tag{5.21}$$

We now write the rate r^* as a function of p given by

$$\begin{aligned} r^*(p) &= \sum_{n\in\mathcal{N}} \mu_n \left[\sum_{k\in\mathcal{K}} r_{k,n}(\rho_n^*, s_{k,n}^*) \right] \\ &= \sum_{n\in\mathcal{N}} \mu_n \rho_n^* \left[\sum_{k\in\mathcal{K}} \ln \left[\frac{\mu_n g_{k,n}}{\lambda_0} \right) \right] \\ &= \sum_{n\in\mathcal{N}} \mu_n \rho_n^* \left\{ \sum_{k\in\mathcal{K}} \ln \left[\left(p + \sum_{i\in\mathcal{N}} \rho_i \sum_{k\in\mathcal{K}_i} \frac{1}{g_{k,i}} \right) \cdot \right. \right. \\ &\quad \left. \left. \frac{\mu_n g_{k,n}}{\sum_{i\in\mathcal{N}} k_i \mu_i \rho_i} \right] \right\}. \end{aligned} \tag{5.22}$$

The first equation in (5.22) arises from *Proposition* 2. By taking the derivative of (5.22), and comparing it with (5.21), it is easy to find that

$$\frac{dr^*(p)}{dp} = \lambda_0. \tag{5.23}$$

5.3.2 Total Power Distribution

With the expression of $r^*(p)$ in (5.22), we now solve **P5-3** in this subsection. The Lagrangian of Problem (5.6) is given by

$$L\big(p(\mathbf{g}), \boldsymbol{\mu}(\mathbf{g}), \lambda_0(\mathbf{g}), \lambda\big) = \int_{\mathbf{g}} l\big(p(\mathbf{g}), \boldsymbol{\mu}(\mathbf{g}), \lambda_0(\mathbf{g}), \lambda\big) pdf(\mathbf{g}) \mathrm{d}\mathbf{g},$$

where

$$l\big(p(\mathbf{g}), \boldsymbol{\mu}(\mathbf{g}), \lambda_0(\mathbf{g}), \lambda\big) = r^*(p) - \lambda p(\mathbf{g}). \tag{5.24}$$

According to the generalized KKT necessary condition theorem for variational optimization [28] and using (5.23), the optimal solution to **P5-3** satisfies

$$\frac{\partial l(\ldots)}{\partial p(\mathbf{g})} = \lambda_0(\mathbf{g}) - \lambda = 0. \tag{5.25}$$

Remarks: One can interpret the Lagrange multiplier λ_0 in (5.11) as short-term power price for a given channel realization $\mathbf{g}$, and interpret the Lagrange multiplier λ in (5.24) as long-term power price. Equation (5.25) indicates that the short-term power price is independent of channel condition and equal to the long-term power price.

Combining the results in Sections 5.3.1 and 5.3.2 yields the optimal solution to the original problem **P5-1** that satisfies

$$p^*_{k,n}(\mathbf{g}) = \left[\frac{\mu_n(\mathbf{g})}{\lambda} - \frac{1}{g_{k,n}}\right]^+, \quad \forall k \in \mathcal{K}, n \in \mathcal{N}, \tag{5.26}$$

$$\rho^*_n(\mathbf{g}) = \frac{\prod_{i \neq n} \sum_{k \in \mathcal{K}} \left[\ln\left(\frac{\mu_n(\mathbf{g}) g_{k,n}}{\lambda}\right)\right]^+}{\sum_{i \in \mathcal{N}} \prod_{j \neq i} \sum_{k \in \mathcal{K}} \left[\ln\left(\frac{\mu_j(\mathbf{g}) g_{k,j}}{\lambda}\right)\right]^+}, \quad \forall i \in \mathcal{N}. \tag{5.27}$$

Thus, in each frame, under the current power price λ, one can first obtain the optimal transmission power using (5.26) after two-nested binary search for $\boldsymbol{\mu}$ as described in Section 5.3.1, and then determine the time-sharing fractions using (5.27) which ensures that the transmission rates r_n over all hops are equal. At a larger time scale, we adjust the power price to meet the average power constraint. The adjustment can be implemented using binary search since it can be shown that $\mathbb{E}\left[\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}}\rho_n(\mathbf{g})p_{k,n}(\mathbf{g})\right]$ is monotonically decreasing in the power price.

5.3.3 Properties of Optimal Power and Time Allocation

It is observed from (5.26) that the best power allocation has the water-filling structure, where $\mu_n(\mathbf{g})/\lambda$ can be viewed as the water level.

Proposition 3: If $\sum_{k\in\mathcal{K}_n}\frac{1}{g_{k,n}} > \sum_{k\in\mathcal{K}_i}\frac{1}{g_{k,i}}$, then we have $\mu_n k_n > \mu_i k_i$.

Proof: See Appendix XV. ■

We treat the total transmission power over hop n as a whole,

$$p_n = \frac{\mu_n k_n}{\lambda_0} - \sum_{k\in\mathcal{K}_n}\frac{1}{g_{k,n}}.$$

where $\mu_n k_n/\lambda$ is viewed as the water level and $\sum_{k\in\mathcal{K}_n}\frac{1}{g_{k,n}}$ is inversely proportional to the harmonic mean of channel gains on the active subcarriers over hop n. *Proposition* 3 suggests that a high water level is given to the hop with small harmonic mean of channel gains to avoid the bottleneck of the whole network it brings. Note that when one of the hops n suffers from a severe fading (the water level is less than $\min_k 1/g_{k,n}$) at a particular time frame, equation (5.27) results in that $\rho_n = 1$ and $\rho_i = 0, i(\neq n) \in \mathcal{N}$, which means the transmission over the whole time frame should be turned off in order to save power for further transmission.

The water-filling power allocation scheme for linear multi-hop networks is different from the one for multi-user fading broadcast channels in [42] and multi-user uplink

channels in [38], where a user uses the same water level over time while different users may have different water levels. This is because the optimal solution for linear multi-hop networks requires that each instantaneous transmission rate over all hops to be equal.

5.4 Suboptimal Solutions

In the previous section, we derived the optimal power and time allocation for OFDM based linear multi-hop networks, and refer to it as *alg-opt* for short. This scheme determines the maximum achievable end-to-end average transmission rate. However, it incurs high computational complexity and large signaling overhead, hence may be infeasible in practice. It is desirable to have suboptimal solutions by imposing some constraints on the network operation. In this section, we develop three suboptimal algorithms, $alg1$, $alg2$ and $alg3$, and discuss their respective computational complexity and signaling overhead.

5.4.1 A Solution With A Constant Water Level

Unlike the optimal scheme, the water level used for power allocation in this suboptimal solution, $alg1$, is set to a constant. We express this solution as

$$p_{k,n}(\mathbf{g}) = \left[\frac{1}{\lambda N} - \frac{1}{g_{k,n}}\right]^+, \quad \forall k \in \mathcal{K}, n \in \mathcal{N}, \tag{5.28}$$

$$\rho_n(\mathbf{g}) = \frac{\prod_{i\neq n} \sum_{k\in\mathcal{K}} \left[\ln\left(\frac{g_{k,i}}{\lambda N}\right)\right]^+}{\sum_{i\in\mathcal{N}} \prod_{j\neq i} \sum_{k\in\mathcal{K}} \left[\ln\left(\frac{g_{k,j}}{\lambda N}\right)\right]^+}, \quad \forall n \in \mathcal{N}, \tag{5.29}$$

where λ satisfies average power constraint.

We can comprehend $alg1$ as follows: first we relax the requirement that transmission rates over all hops equal and maximize the weighted sum of the transmission

rates over all hops with weights given by $\{\mu_n\}$. By letting $\mu_n = 1/N, \forall n \in \mathcal{N}$ in (5.26), we can have the expression of $p_{k,n}$ in (5.28). Then we adjust the time-sharing fractions so that the transmission rates over all hops remain equal. By *Proposition* 3, $alg1$ approaches to *alg-opt* when the harmonic mean of the channel gains over active subcarriers, i.e., $k_n/(\sum_{k=\in\mathcal{K}_n} \frac{1}{g_{k,n}})$ of each hop is nearly the same at every time frame. Since the same water level is shared by all subcarriers over all hops and is invariant in the time domain, the computational complexity and signal exchanging overhead can be significantly reduced. However, a central controller is still needed. The difference from the optimal allocation scheme is that there is no need to compute the weight vector $\boldsymbol{\mu}$ at each time frame.

5.4.2 Partially Distributed Power and Time Allocation

The suboptimal time and power allocation, $alg2$, is denoted as

$$\rho_n = \frac{\prod_{i\neq n} \sum_{k\in\mathcal{K}} \mathbb{E}\big[c_{k,i}(P/K)\big]}{\sum_{i\in\mathcal{N}} \prod_{j\neq i} \sum_{k\in\mathcal{K}} \mathbb{E}\big[c_{k,j}(P/K)\big]}, \quad \forall n \in \mathcal{N}. \tag{5.30}$$

$$p_{k,n}(\mathbf{g}_n) = \left[\frac{1}{\lambda_n} - \frac{1}{g_{k,n}}\right]^+, \quad \forall k \in \mathcal{K}, n \in \mathcal{N}, \tag{5.31}$$

where $c_{k,n}(\cdot)$ is given by (5.3), $\mathbf{g}_n \triangleq \{g_{k,n}, k \in \mathcal{K}\}$ and λ_n satisfies

$$\rho_n \mathbb{E}_{\mathbf{g}_n}\left[\sum_{k\in\mathcal{K}} p_{k,n}(\mathbf{g}_n)\right] = \frac{P}{N}.$$

$alg2$ is a partially distributed algorithm. To avoid centralized management, we require that the average total transmission power is to be allocated equally to each hop. Since only statistical information is needed to compute the time-sharing fractions and water levels, the signaling overhead can be significantly reduced. Upon receiving the time-sharing fractions, this algorithm performs power allocation across subcarriers in its locality by binary search for λ_n, separately.

5.4.3 Equal Resource Allocation

The equal time and power allocation scheme, $alg3$, is denoted as

$$\begin{aligned} p_{k,n} &= \frac{P}{K}, \quad \forall k \in \mathcal{K}, n \in \mathcal{N} \\ \rho_n &= \frac{1}{N}, \quad \forall n \in \mathcal{N}. \end{aligned}$$

Transmission power and time are equally distributed among all subcarriers and hops. It corresponds to the scenario that the transmitters have no channel information. The disadvantage is that any one of hops which suffers from bad channel condition may become the bottleneck of the whole link. This scheme needs no signal exchange and computation.

We have listed three suboptimal algorithms above in the decreasing order of computational complexity and signaling overhead. Both *alg-opt* and $alg1$ need a central controller while $alg2$ can perform resource allocation at each node side using local information. All of the three suboptimal algorithms are not required to do binary search in each time frame.

According to practical requirements and restrictions, we can choose one of these schemes by finding the optimal tradeoff between performance of the algorithms, and their complexity and signal exchanging overhead. In the next section, we compare the performance of these algorithms.

5.5 Numerical Results

In this section, we present numerical results of end-to-end average rates under different algorithms. Consider a linear network with N hops. We fix the bandwidth to be 1MHz and the distance between the source and destination to be 1km. The relays are equally spaced. In all simulations, the channel over each hop is modelled

by Stanford University Interim (SUI)-3 channel model with a central frequency at around 1.9 GHz to simulate the fixed broadband wireless access channel environments [17]. The SUI-3 channel is a 3-tap channel. The received signal fading on the first tap is characterized by a Ricean distribution with K-factor equal to 1. The fading on the other two taps follows a Rayleigh distribution. The root-mean-square (rms) delay spread is 0.305μs. Then the coherence bandwidth is approximately 65KHz. Hence, the number of subcarriers K should be greater than 15.2 so that the subcarrier bandwidth is small enough to experience the flat fading. Throughout the numerical analysis in this section, we choose $K = 16$. The maximal Doppler frequency is set to 0.4 Hz. The path loss model is given by the intermediate path loss condition [16, Category B]:

$$PL = A + \alpha \lg \left(\frac{d}{d_0} \right),$$

where d_0 is a reference distance and set to 100m, $A = 20 \lg(4\pi d_0/\lambda)$ with λ being the wavelength, and α is the path-loss exponent. The SNR gap in (5.1) is set to 8.2dB (which corresponds to a BER requirement of 10^{-5} if adaptive QAM modulation is used).

Fig. 5.3 and Fig. 5.4 show the end-to-end average transmission rates achieved by using different algorithms with and without shadowing (4dB standard deviation), respectively. From Fig. 5.3 we observe that the performance gaps between $alg2$ and $alg3$ are very small, especially in high power region. Since the relays are equally spaced and there is no shadowing, for $alg2$, the average rate over each hop is the same. Thus, same as $alg3$, $alg2$ has uniform time allocation $\rho_n = 1/N,\ \forall n \in \mathcal{N}$. Although $alg2$ can vary the transmission power over subcarriers, the impact of power control is limited since the rate is a concave function of the transmission power, especially when the water level is far higher than the inverse of the normalized channel gains. This result in the single-hop case was also demonstrated in [25]. The improvement of *Alg-opt* and $alg1$ over $alg2$ is obvious because they adapt the time-sharing fractions according to

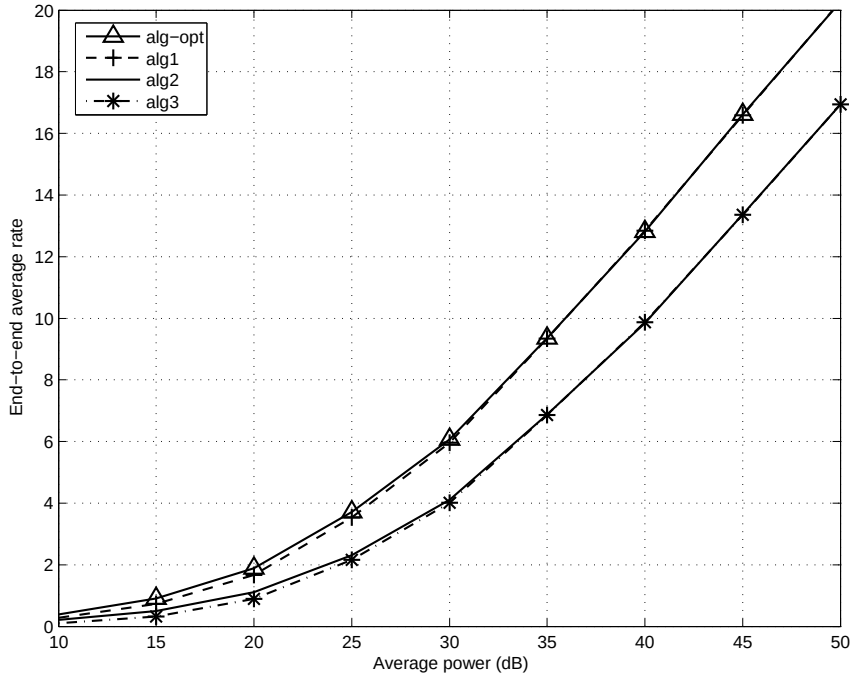

Figure 5.3. End-to-end average rate vs. average total transmission power with path loss exponent $\alpha = 3.5$, no shadowing and $N = 5$

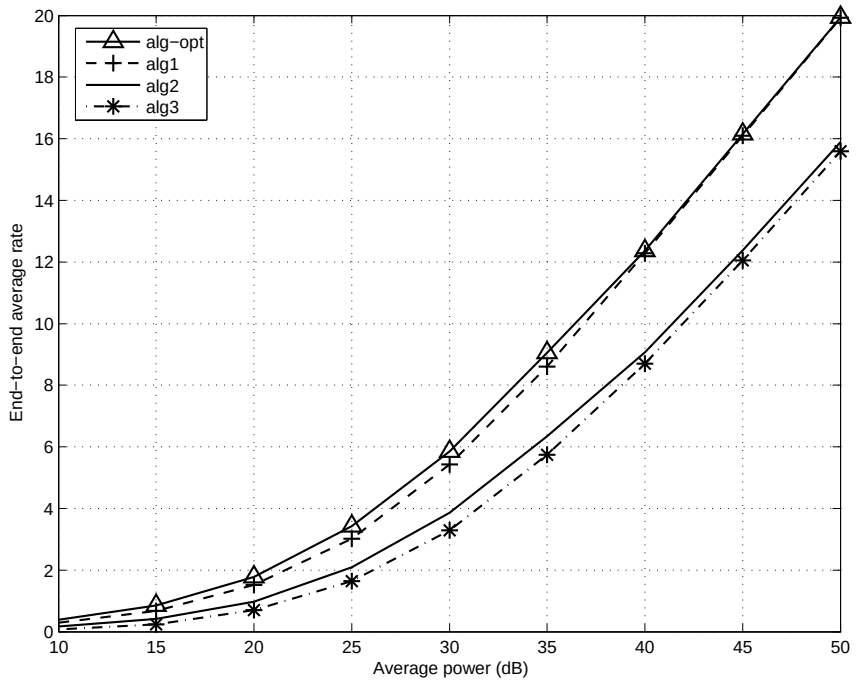

Figure 5.4. End-to-end average rate vs. average total transmission power with path loss exponent $\alpha = 3.5$, shadowing and $N = 5$

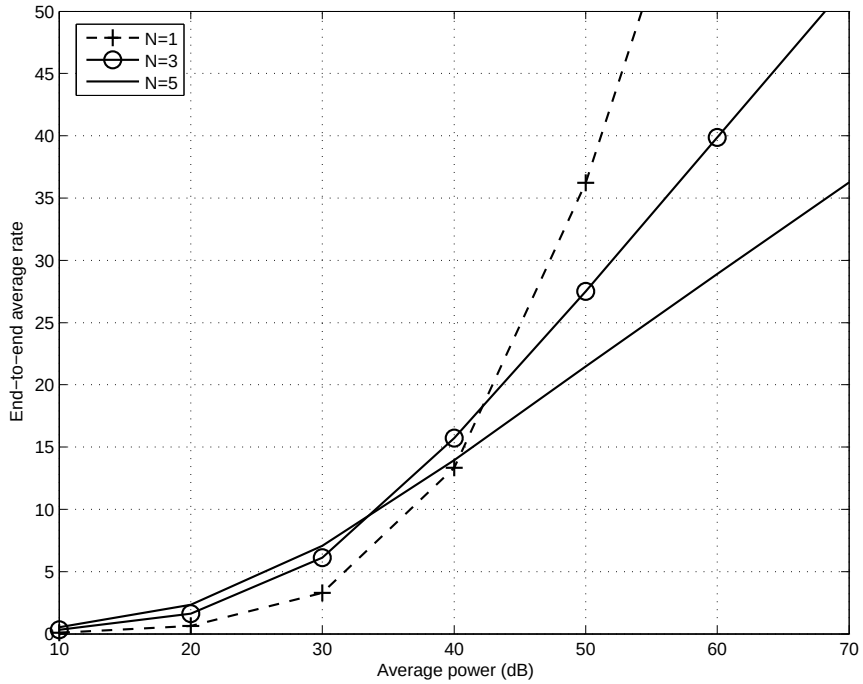

Figure 5.5. End-to-end average rate vs. average total transmission power with path loss exponent $\alpha = 3$ and no shadowing for *alg-opt*

instantaneous CSI at each time frame and the rate is linear in time-sharing fraction. The similarity in performance of $alg1$ and *alg-opt* at the high SNR region indicates that varying water level over time yields a negligible performance gain. Therefore, $alg1$ and $alg3$ are recommended due to their relatively low computational complexity and small signaling overhead. Since different hops suffer independent shadowing, it is more likely for one of the hops to experience bad channel condition and becomes the bottleneck of the whole link. Thus, the shadowing deteriorates the performance of the algorithm with multiple hops. For instance, at 35 dB average power, end-to-end average rate for no shadowing case are 0.5-1 higher than that for shadowing case. It also highlights the performance gaps between $alg2$ and $alg3$ since $alg2$ can reduce the impact of the bottleneck by adapting time-sharing fractions according to the channel statistics. On the other hand, *alg-opt* also does not exhibit a significant rate increase over $alg1$.

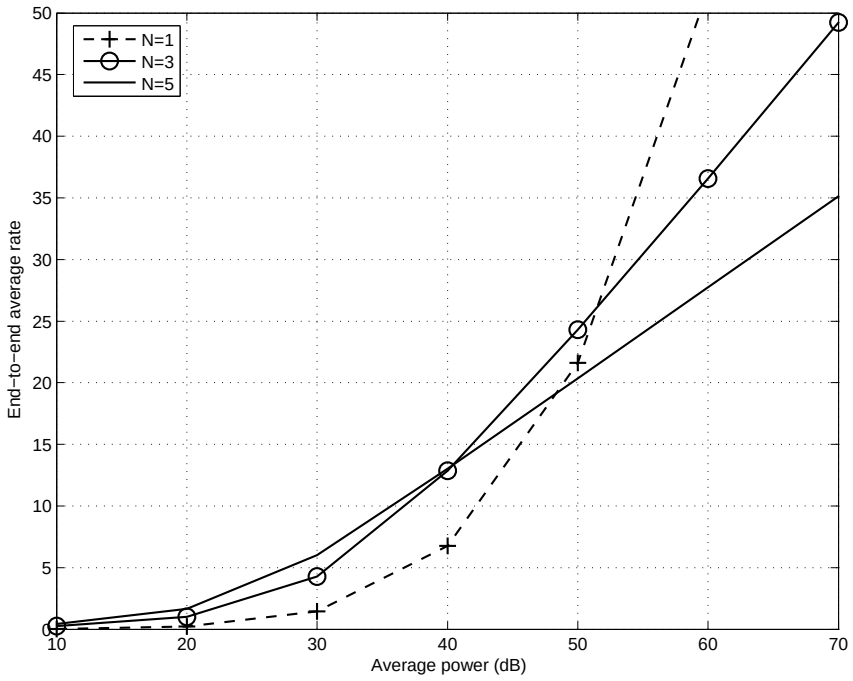

Figure 5.6. End-to-end average rate vs. average total transmission power with path loss exponent $\alpha = 3.5$ and no shadowing for *alg-opt*

Fig. 5.5 and Fig. 5.6 show the impact of the number of hops at $\alpha = 3$ and $\alpha = 3.5$, respectively. The optimal algorithm *alg-opt* is used. It is seen that the optimal number of hops deceases as the average total transmission power increases for a fixed path loss exponent. The optimal number of hops are 5, 3, and 1, respectively for the region [10, 34.2), [34.2, 42.4) and $[42.4, +\infty)$ in Fig. 5.5. In addition, the intersections of performance curves with different N shift to the left side as the path loss exponent increases.

5.6 Conclusions

We have obtained the maximum end-to-end average transmission rate for OFDM based linear multi-hop networks and its corresponding optimal power and time allocation. The optimization problem to maximize end-to-end average transmission rate

is solved by decomposing it into two subproblems which are equivalent to the original one. The optimal strategy indicates that the transmission power has the water-filling structure and the hops with relatively bad channel condition are given higher water levels. Further, time-sharing fractions are adaptive so that the hops have equal transmission rates. We also compare the optimal strategy with three suboptimal ones that require lower computational complexity and lower signaling overhead. The numerical findings are that the optimal strategy has a better performance than the equal resource allocation but it does not exhibit a significant rate increase over the suboptimal algorithm with a constant water level. However, the study in this chapter cannot be applied in networks supporting delay-constrained services where a constant target rate is desired. For this reasion, in the next chapter, the end-to-end resource allocation for minimizing the service outage will be addressed with the same network architecture as used in this chapter.

Chapter 6

End-to-End Outage Minimization in Linear OFDM Based Relay Networks

In this chapter, we continue studying the end-to-end resource allocation in an OFDM based wireless relay network. Unlike the previous chapter, our goal here is to find an optimal power and time adaptation to minimize end-to-end outage probability under a long-term total power constraint.

For many real-time services, one has to consider keeping the target transmission rate and avoiding outage in most fading conditions through dynamic resource allocation. An outage is an event that the actual transmission rate is below a prescribed transmission rate ([11] and [43]). Outage probability can be viewed as the fraction of time that a codeword is decoded wrongly. For a given finite average power constraint, the network may not be able to support the target rate all the time. Thus, transmission outage is inevitable under severe fading conditions. In [11], the information outage probability is minimized with an optimal power controller in a single-user,

single-hop and single channel setting. How to minimize outage probability in an OFDM-based multi-hop wireless network is currently under-explored. In a linear relay network where no data is allowed to accumulate at any of relay nodes, an end-to-end outage is the event that there exists a hop on which transmission rate is lower than the target rate.

In this chapter, we study the transmission power and time allocation to minimize the end-to-end outage probability in a one-dimensional multi-hop network under an average transmission power constraint. In the first step, we derive the minimum short-term power required to meet a target transmission rate for any given channel realization. The resulting power and time allocation can be obtained through a Two-nested Binary Search (TBS) which is conducted in a central controller with the knowledge of channel state information (CSI) of all subcarriers and over all hops. Such scheme gives a theoretical performance limit over multi-hop networks, but typically rely on highly complex computation and involve significant signalling between nodes and central controller, making it difficult to implement. For this reason, a low computation cost Iterative Algorithm of Sub-optimal power and time allocation (IAS) is proposed to approximate the aforementioned scheme based on TBS. The required information for signalling exchange only involve the geometric mean and harmonic mean of channel gains averaged over active subcarriers and the number of active subcarriers on each hop. This sub-optimal allocation scheme suggests prolonging the transmission time for the hop with low geometric mean of channel gains while lowering the transmission power for the hop with low harmonic mean. In the second step, the transmission on-off is determined by comparing the required minimum total power with a threshold. Specifically, the transmission will be shut off if the required minimum total power for a given channel realization exceeds the threshold. The value of the threshold is chosen so that the long-term total power constraint is satisfied.

Numerical results show that a significant power saving can be achieved by the

proposed optimal power and time allocation compared with the uniform power and time allocation under the same end-to-end outage probability. In addition, the proposed sub-optimal power and time allocation serves as a good approximation to the optimal solution when the target rate is sufficiently high. The relation between the optimal number of relays in the sense of the minimum required power for a given target rate is also shown numerically.

The remainder of this chapter is organized as follows. In Section 6.1, end-to-end rate and outage probability are introduced. The optimal and sub-optimal resource allocation schemes to minimize the end-to-end outage probability under an average total power constraint are proposed in Section 6.2. Numerical results are given in Section 6.3. Finally we conclude this chapter in Section 6.4.

6.1 End-to-End Rate and Outage Probability

Consider a same linear relay wireless network as in Chapter 5. To keep this chapter integrated, we repeat the system setting. We denote the transmission power on subcarrier k over hop n by $p_{k,n}$ and let the long-term total transmission power be limited by P. As given by Chapter 5, the transmission rate (in Nat/OFDM symbol) in a time frame achieved over hop n is given by

$$r_n = \sum_{k\in\mathcal{K}} \rho_n \ln\left(1 + \frac{g_{k,n}p_{k,n}}{\Gamma N_0}\right), \quad \forall n \in \mathcal{N}, \tag{6.1}$$

where N_0 is the noise power, and Γ is the signal-to-noise ratio (SNR) gap [54]. For notation brevity, we also redefine $g_{k,n}$ as $g_{k,n} := g_{k,n}/(\Gamma N_0)$. Under the assumption that no data is allowed to accumulate at any relay nodes (also called "information-continuous relaying in [50]), the total number of bits received at the destination node at the end of time frame, B, is the minimum of the number of bits transmitted over each hop, B_n, where $B_n = r_n T_n$. If the end-to-end transmission rate r is defined as

the number of bits per OFDM symbol received at the destination node, then it can be expressed as

$$r = \min_{n \in \mathcal{N}} r_n. \tag{6.2}$$

Uniform power and time allocation (UPTA): When the receivers can perfectly track the CSI but the transmitters have no such information, or transmitters do not exploit CSI due to high signalling overhead, the end-to-end transmission rate has the form

$$r_{UPTA} = \frac{1}{N} \min_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \ln \left(1 + \frac{g_{k,n} P}{K} \right).$$

If the actual end-to-end transmission rate in a certain time frame is less than a target transmission rate R, then an outage is said to have occurred. The corresponding end-to-end outage probability at target rate R is given by

$$P_{UPTA}^{out} \triangleq Prob(r_{UPTA} < R) = Prob \left(\sum_{k \in \mathcal{K}} \ln(1 + g_{k,n} P/K) < NR, \forall n \in \mathcal{N} \right).$$

In this case, the transmission scheme is independent of CSI and follows uniform power and time allocation.

Fixed power and adaptive time allocation (FPAT): When the transmitters have CSI to some extend (not necessarily global CSI), each node can perform rate adaptation to avoid the situation where the bad conditioned hop becomes the bottleneck of the whole link. We assume that the transmission power on each subcarrier over each hop to be unchanged in the time domain, and rate-adaptation can be performed by adjusting time-sharing fraction such that $r_n = r_i, \forall i, n \in \mathcal{N}$. In this scenario, the maximum end-to-end transmission rate under reliable communication guarantee is given by

$$r_{FPAT} = \left(\sum_{n \in \mathcal{N}} \frac{1}{\sum_{k \in \mathcal{K}} \ln(1 + g_{k,n} P/K)} \right)^{-1}.$$

This rate is achieved by assigning the time sharing fraction to be

$$\rho_i(\mathbf{g}) = \frac{\prod_{n \neq i} \sum_{k \in \mathcal{K}} \left[\ln \left(1 + g_{k,n} P/K \right) \right]^+}{\sum_{n \in \mathcal{N}} \prod_{n \neq i} \sum_{k \in \mathcal{K}} \left[\ln \left(1 + g_{k,n} P/K \right) \right]^+}, \tag{6.3}$$

The corresponding end-to-end outage probability at target rate R is given by

$$P_{FPAT}^{out} \triangleq Prob(r_{FPAT} < R).$$

To implement FPTA, the only information needed at each transmit node is the value of $\sum_{k\in\mathcal{K}} \ln(1+g_{k,n}P/K)$, and r_{FPAT}, while the knowledge of global CSI is not required at each transmit node. This reduces signalling exchange load greatly. This case can be viewed as a simple generalization of rate-adaptive relaying in flat fading channels discussed in [50].

Adaptive power and fixed time allocation (APFT): In the case that time-sharing fractions are required to be allocated uniformly while transmission power on each subcarrier over each hop is adjusted to minimize the end-to-end outage probability, the minimum end-to-end outage probability can be expressed as follows,

$$P_{APFT}^{out} = \min_{\mathbf{p}} Prob\left(\frac{1}{N}\sum_{k\in\mathcal{K}} \ln(1+g_{k,n}p_{k,n}) < R \mid \frac{1}{N}\mathbb{E}\left[\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} p_{k,n}\right] \leq P\right).$$

APFT can be viewed as a special case of APTA that will be presented next by forcing by fixing $\rho_n = 1/N$, and as will be demonstrated, it can also be solved using the following two steps: short-term power minimization and long-term power threshold determination. Whereas, unlike APTA, the first step can be performed locally, i.e., each transmitter only need to know the local CSI over the associated hop to solve the problem

$$\begin{aligned} \min_{\mathbf{p}} \quad & \sum_{k\in\mathcal{K}} p_{k,n} \\ \text{s.t.} \quad & \frac{1}{N}\sum_{k\in\mathcal{K}} \ln(1+g_{k,n}p_{k,n}) \geq R, \end{aligned}$$

for all $n \in \mathcal{N}$.

Adaptive power and time allocation (APTA): we shall now focus on the scenario of interest, where both transmission power over each subcarrier over each hop and time over each hop are allowed to be dynamically allocated. We assume that at the start

of each time frame, the CSI on all subcarriers and over all hops is fully available at a central controller, which could be embedded in the source node. Given the global CSI, our goal is to find an optimal power and time allocation to minimize the end-to-end outage probability for a given target transmission rate subject to a long-term total power constraint. Namely,

$$\textbf{P6-1:} \quad \min_{\{\boldsymbol{\rho},\ \mathbf{p}\}} \quad \mathbb{E}\{\mathcal{I}_F[r(\mathbf{g},\boldsymbol{\rho},\mathbf{p}) < R]\} \tag{6.4}$$

$$\text{s.t.} \quad \mathbb{E}\left[\sum_{n\in\mathcal{N}} \rho_n(\mathbf{g}) \left(\sum_{k\in\mathcal{K}} p_{k,n}(\mathbf{g})\right)\right] \leq P$$

$$\sum_{n\in\mathcal{N}} \rho_n(\mathbf{g}) = 1, \tag{6.5}$$

where $\mathbf{g} = \{g_{k,n}\ , k \in \mathcal{K}, n \in \mathcal{N}\}$, $\boldsymbol{\rho} = \{\rho_n\ , n \in \mathcal{N}\}$, $\mathbf{p} = \{p_{k,n}\ , k \in \mathcal{K}, n \in \mathcal{N}\}$, $\mathcal{I}_F$ is an indicator function and $\mathbb{E}(x)$ denotes the expectation of x over the joint distribution of $\mathbf{g}$.

6.2 Adaptive Power and Time Allocation

The minimum outage probability problem **P6-1** defined in the previous section can in general be solved in two steps as proposed in [11]. First, for each global channel state $\mathbf{g}$, the short-term minimum total power $P_{\min}(\mathbf{g}, R)$ required to guarantee the target end-to-end transmission rate R is to be determined. The second step then determines a threshold to control the transmission on-off subject to a long-term power constraint.

6.2.1 Short-Term Power Minimization

In this subsection, we shall find the optimal time sharing fraction ρ_n^* $(\forall n \in \mathcal{N})$ and the optimal power allocation $p_{k,n}^*$ $(\forall n \in \mathcal{N}, k \in \mathcal{K})$ to minimize the short-term

total power needed to achieve a target end-to-end transmission rate R. Then a sub-optimal solution with reduced complexity is developed. The sub-optimal solution has a closed-form expression from which a few attractive properties regarding time and power allocation can be observed. Comparisons of average powers and computational complexity between the optimal and the sub-optimal solution are also given.

Optimal power and time allocation

The optimal power and time allocation problem to minimize short-term total power can be formulated as

$$\textbf{P6-2: } p_{\min}(\mathbf{g}) = \min_{\{\boldsymbol{\rho},\ \mathbf{p}\}} \sum_{n\in\mathcal{N}} \rho_n(\mathbf{g}) \left(\sum_{k\in\mathcal{K}} p_{k,n}(\mathbf{g}) \right) \tag{6.6}$$

$$\text{s.t. } r(\mathbf{g}, \boldsymbol{\rho}, \mathbf{p}) \geq R \tag{6.7}$$

$$\sum_{n\in\mathcal{N}} \rho_n = 1.$$

Unfortunately, the term $r(\mathbf{g}, \boldsymbol{\rho}, \mathbf{p})$ in constraint (6.7) is not concave in $\boldsymbol{\rho}$ and $\mathbf{p}$. To make the problem **P6-2** more tractable, we introduce a new variable $s_{k,n}$ defined as $s_{k,n} := \rho_n p_{k,n}$. This new variable can be viewed as the actual amount of energy consumed by hop n on subcarrier k in a time frame interval. In addition, it follows from (6.2) that constraint (6.7) can be rewritten as N sub-constraints. By doing these, problem **P6-2** can be transformed into a new problem with optimization variables ρ_n $(\forall n \in \mathcal{N})$ and $s_{k,n}$ $(\forall n \in \mathcal{N},\ k \in \mathcal{K})$:

$$\textbf{P6-3: } \min_{\{\boldsymbol{\rho},\ \mathbf{s}\}} \quad \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} s_{k,n} \tag{6.8}$$

$$\text{s.t.} \quad \rho_n \sum_{k\in\mathcal{K}} \ln\left(1 + \frac{g_{k,n} s_{k,n}}{\rho_n}\right) \geq R, \quad \forall n \in \mathcal{N} \tag{6.9}$$

$$\sum_{n\in\mathcal{N}} \rho_n = 1. \tag{6.10}$$

Since its Hessian matrix is negative semidefinite, the function $\rho_n \ln(1 + g_{k,n} s_{k,n}/\rho_n)$ is concave in ρ_n and $s_{k,n}$. Therefore, problem **P6-3** is a convex optimization problem

and there exists a unique optimal solution. To observe the structure of the optimal solution, we write the Lagrangian of Problem **P6-3** as follows:

$$J(\{\rho_n\},\{s_{k,n}\},\{\lambda_n\},\beta) = \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} s_{k,n} + \beta\left(\sum_{n\in\mathcal{N}}\rho_n - 1\right) + \sum_{n\in\mathcal{N}}\lambda_n\left[R - \rho_n\sum_{k\in\mathcal{K}}\ln\left(1+\frac{g_{k,n}s_{k,n}}{\rho_n}\right)\right] \tag{6.11}$$

where $\lambda_n \geq 0$ ($n \in \mathcal{N}$) and $\beta \geq 0$ are the Lagrange multipliers for the constraints (6.9) and (6.10), respectively. If $\{\rho_n^*\}$ and $\{s_{k,n}^*\}$ are the optimal solution of **P6-3**, they should satisfy the Karush-Kuhn-Tucker (KKT) conditions [7], which are necessary and sufficient for the optimality. The KKT conditions are listed as follows:

$$\frac{\partial J(\ldots)}{\partial s_{k,n}}\begin{cases} = 0 & \text{if } s_{k,n}^* > 0 \\ > 0 & \text{if } s_{k,n}^* = 0 \end{cases}, \ \forall n \in \mathcal{N}, k \in \mathcal{K} \tag{6.12}$$

$$\frac{\partial J(\ldots)}{\partial \rho_n}\begin{cases} > 0 & \text{if } \rho_n^* = 0 \\ = 0 & \text{if } 0 < \rho_n^* < 1 \\ < 0 & \text{if } \rho_n^* = 1 \end{cases}, \forall n \in \mathcal{N} \tag{6.13}$$

$$\lambda_n\left[\rho_n^*\sum_{k\in\mathcal{K}}\ln\left(1+\frac{g_{k,n}s_{k,n}^*}{\rho_n^*}\right) - R\right] = 0, \ \forall n \in \mathcal{N}. \tag{6.14}$$

It can be obtained from the KKT condition (6.12) that the optimal power distribution $\{p_{k,n}^*\}$ has a water-filling structure, and is given by

$$p_{k,n}^* = \frac{s_{k,n}^*}{\rho_n^*} = \left(\lambda_n - \frac{1}{g_{k,n}}\right)^+, \ \forall k \in \mathcal{K}, \ n \in \mathcal{N}, \tag{6.15}$$

where $(x)^+ \triangleq \max(0, x)$, and λ_n can be regarded as the water level on hop n. Different hops may have different water levels, and for each hop, more power is allocated to the subcarrier with higher channel gain and vice versa.

Let $\mathcal{K}_n$ denote the set of subcarriers over hop n that satisfies $g_{k,n} \geq 1/\lambda_n$ ($\forall k \in \mathcal{K}_n$), and let k_n be the size of the set. The subcarriers in the set are said to be active.

The value of λ_n in (6.14) is non-zero, otherwise $p_{k,n}^* = 0$ $(\forall k, n)$. Hence, we can obtain the closed-form expression for ρ_n^* by substituting (6.15) into (6.14):

$$\rho_n^* \triangleq h_n(\mathbf{g}, \lambda_n) = \frac{R}{\sum_{k\in\mathcal{K}_n} \ln g_{k,n} + k_n \ln \lambda_n}, \ \forall n \in \mathcal{N}. \tag{6.16}$$

From (6.16), it can be shown that ρ_n^* is monotonically decreasing in λ_n (note that k_n in (6.16) is treated as a function of λ_n).

In the following, we derive the relation between λ_n and β. Taking the derivative of Lagrangian of **P6-3** in (6.11) with respect to ρ_n, we have

$$\frac{\partial J(\ldots)}{\partial \rho_n} = \lambda_n \left[\sum_{k\in\mathcal{K}} \ln\left(1 + \frac{g_{k,n}s_{k,n}}{\rho_n}\right) - \sum_{k\in\mathcal{K}} \frac{g_{k,n}s_{k,n}}{\rho_n + g_{k,n}s_{k,n}} \right] - \beta. \tag{6.17}$$

Suppose that there existed a $n \in \mathcal{N}$ such that $\rho_n^* = 0$ or 1, then the constraint (6.14) would be violated. Thus, we have $0 < \rho_n^* < 1$ $(\forall n \in \mathcal{N})$. Substituting (6.15) into (6.17) and using (6.13), we can express β as a function of λ_n given by

$$\beta \triangleq f_n(\mathbf{g}, \lambda_n) = \lambda_n \left(\sum_{k\in\mathcal{K}_n} \ln g_{k,n} + k_n \ln \lambda_n - k_n \right) + \sum_{k\in\mathcal{K}_n} \frac{1}{g_{k,n}}, \ \forall n \in \mathcal{N}. \tag{6.18}$$

It is seen from (6.18) that finding the optimal water levels $\{\lambda_n\}$ for a given β are N independent problems. It can be proven that $f_n(\mathbf{g}, \lambda_n)$ is a monotonically increasing function of λ_n in the region $\left[\min_k \left(\frac{1}{g_{k,n}}\right), +\infty\right]$ by evaluating its deviation with respect to λ_n. Hence, the inverse function, $f_n^{-1}(\mathbf{g}, \beta)$, exists and is an increasing function of β. Therefore, the exact value of λ_n for a given β can be obtained numerically using binary search when the upper bound is known.

Substituting $\lambda_n = f_n^{-1}(\mathbf{g}, \beta)$ into (6.16), we can express ρ_n^* as $\rho_n^* = h_n(f_n^{-1}(\mathbf{g}, \beta))$. Since ρ_n^* is decreasing in λ_n and λ_n is increasing in β, it follows that ρ_n^* is decreasing in β. Therefore, the optimal β can also be obtained via binary search from the constraint (6.10). Hence, the optimal solution $\{\rho_n^*, s_{k,n}^*\}$ of **P6-3** and the resulting $p_{\min}$ can be obtained through two-nested binary search. The outer loop varies the

Lagrange multiplier β to meet the total transmission time constraint. The inner loop searches the water level for each hop at a given value of β to satisfy (6.18). The algorithm is outlined as follows.

Two-nested binary search for minimum short-term power (TBS)

Binary search for β

1. Find the upper bound and lower bound of β
   (a) For all $n \in \mathcal{N}$, let $\bar{\lambda}_n = \underline{\lambda}_n = \min_k\{1/g_{k,n}\}$
   (b) Compute $\bar{\rho}_n = h_n(\mathbf{g}, \bar{\lambda}_n)$ using (6.16)
   (c) If $\bar{\rho}_n > 1/N$, update $\bar{\lambda}_n = 2\bar{\lambda}_n$ and repeat Step 1)-b) and c)
   else, go to Step 1)-d)
   (d) Set $\beta^{\min} = \max_{n\in\mathcal{N}} f_n(\mathbf{g}, \underline{\lambda}_n)$, and $\beta^{\max} = \max_{n\in\mathcal{N}} f_n(\mathbf{g}, \bar{\lambda}_n)$
2. Set $high = \beta^{\max}$, $low = \beta^{\min}$
3. Let $center = (low+high)/2$ and **binary search for** λ_n $(\forall n \in \mathcal{N})$ at $\beta = center$
   (a) Find the upper bound and lower bound of λ_n, $\lambda_n^{\max}$ and $\lambda_n^{\min}$, respectively
      i. Let $\lambda_n^{\min} = \lambda_n^{\max} = \min_k\{1/g_{k,n}\}$
      ii. Compute $\beta' = f_n(\mathbf{g}, \lambda_n^{\max})$ using (6.18)
      iii. If $\beta' < \beta$, update $\lambda_n^{\min} = \lambda_n^{\max}$ and $\lambda_n^{\max} = 2\lambda_n^{\max}$ and repeat Step 3)-a)-ii) and iii)
      else, let $high_n = \lambda_n^{\max}$, $low_n = \lambda_n^{\min}$, and go to Step 3)-b)
   (b) Set $center_n = (low_n+high_n)/2$. If $f_n(\mathbf{g}, center_n) > \beta$, let $high_n = center_n$; otherwise, let $low_n = center_n$
   (c) Repeat Step 3)-b) until $high_n - low_n < \varepsilon'$ and let $\lambda_n = center_n$

4. If $\sum_{n\in\mathcal{N}} h_n(\mathbf{g},\lambda_n) > 1$, let $low = center$; otherwise, let $high = center$

5. Repeat Step 3) and Step 4) until $high - low < \varepsilon$

6. Using the found $\{\lambda_n\}$ and β, obtain ρ_n^* and $p_{k,n}^*$ based on (6.16) and (6.15), respectively.

7. Compute $p_{\min} = \sum_{n\in\mathcal{N}} \rho_n^* (\sum_{k\in\mathcal{K}} p_{k,n}^*)$

In Step 1), the boundaries of β are determined in order to proceed with the binary search in the outer loop. From (6.18), a common Lagrange multiplier β is shared by all hops and it is a monotonically increasing function of λ_n in the region of $\left[\min_{k\in\mathcal{K}}\left(\frac{1}{g_{k,n}}\right), +\infty\right]$ for all n. We use $\underline{\lambda}_n = \min_{k\in\mathcal{K}}\left(\frac{1}{g_{k,n}}\right)$ to represent the lower bound of λ_n[1], then the lower bound of β is the maximum of $f_n(\mathbf{g},\underline{\lambda}_n)$ among N hops. The same lower bound of λ_n will also be used in Step 3) for the inner loop. The upper bound of β is obtained from the fact that there exists at least an n^* such that $\rho_{n^*} \geq \frac{1}{N}$. Correspondingly, we find a water level $\bar{\lambda}_n = h_n^{-1}(\mathbf{g},\frac{1}{N})$ for all n, where $h_n^{-1}(\mathbf{g},\cdot)$ is the inverse function of $h_n(\mathbf{g},\cdot)$. Then for hop n^*, we have $\lambda_{n^*} \leq \bar{\lambda}_{n^*}$ since $h_n^{-1}(\mathbf{g},\cdot)$ is a decreasing function. Therefore, due to the monotonicity of $f_n(\mathbf{g},\cdot)$, the upper bound of β can be obtained from

$$\beta = f_{n^*}(\mathbf{g},\lambda_{n^*}) \leq f_{n^*}(\mathbf{g},\bar{\lambda}_{n^*}) \leq \max_{n\in\mathcal{N}} f_n(\mathbf{g},\bar{\lambda}_n). \tag{6.19}$$

The algorithm then updates β using binary chop until the sum of the corresponding time-sharing fraction converges to 1. The convergence of the outer loop is guaranteed by the fact that the actual sum of time-sharing fractions is also monotonically decreasing in β.

The aim of the inner loop in Step 3) is to find λ_n ($\forall n \in \mathcal{N}$) for a given β. We first initialize the upper bound of λ_n, $\lambda_n^{\max} = \min_k\left\{\frac{1}{g_{k,n}}\right\}$ and then keep increasing it until

[1] $\lambda_n \geq \min_{k\in\mathcal{K}}\left(\frac{1}{g_{k,n}}\right)$ so that there exists at least one subcarrier on each hop is allocated non-zero power.

the corresponding β' goes beyond the given β. In each iteration, the binary search guess an halfway λ_n between the new *high* and *low* values and repeats it until the actual β' approach the given β. The iteration converges because of the monotonicity of β in λ_n.

The outer loop involves $\log_2\left(\frac{\beta^{\max}-\beta^{\min}}{\varepsilon}\right)$ iterations where ε represents outer loop accuracy. The inner loop has N binary searches, and each involves $\log_2\left(\frac{\lambda_n^{\max}-\lambda_n^{\min}}{\varepsilon'}\right)$ iterations, where ε' is the inner loop accuracy. It is observed from (6.16) and (6.18) that $\beta^{\max} = \mathcal{O}\left(NR\mathrm{e}^{NR/K}\right)$ and $\lambda_n^{\max} = \mathcal{O}\left(\mathrm{e}^{NR/K}\right)$ when the target rate is so high that all subcarriers are active. Therefore, the average computational complexity of this algorithm is upper bounded by the magnitude of $\frac{N^3R^2}{K^2}\ln(\frac{NR}{\varepsilon})\ln(\frac{1}{\varepsilon'})$ in the asymptotical sense with a high target rate.

Sub-optimal time and power allocation

In the optimal time and power allocation, it is infeasible to obtain an closed-form expression for the solution. In this part, we will observe that when the target rate is sufficiently large, the optimal transmission time can be approximated by an explicit function of the geometric mean of channel gains averaged over the active subcarriers and the number of active subcarriers

$$\rho_n' = \frac{R}{k_n R\mu + k_n a_n}, \tag{6.20}$$

where k_n is the size of the set $\mathcal{K}_n = \{k \mid g_{k,n} \geq 1/\lambda_n\}$, and a_n is a function of geometric mean of $g_{k,n}$ of the active subcarrers at hop n and the number of active subcarrers, which will be expressed in (6.23). In addition, the product of water level and the number of active subcarriers for each hop tends to be the same. In the following, we shall investigate this sub-optimal solution and show that it has a low computational complexity and little signalling exchange.

Let $\{\rho_n,\ n \in \mathcal{N}\}$ be any given time allocation that satisfies $\sum_{n\in\mathcal{N}} \rho_n = 1$ and $0 \leq$

$\rho_n \leq 1$. The optimal power distribution for the given time allocation, $\{\rho_n,\ n \in \mathcal{N}\}$, is expressed by (6.15). Substituting (6.15) into (6.1) and letting $p_n = \sum_{k\in\mathcal{K}} p_{k,n}$, the close-form expression of λ_n can be obtained as [63]

$$\lambda_n = \left(\frac{\mathrm{e}^{\frac{R}{\rho n}}}{\prod_{k\in\mathcal{K}_n} g_{k,n}} \right)^{1/k_n}, \tag{6.21}$$

where k_n is the size of the set $\mathcal{K}_n = \{k \mid g_{k,n} \geq 1/\lambda_n\}$ Moreover, substituting (6.21) back into (6.15), we have

$$p_n^*(\rho_n) = \mathrm{e}^{\frac{R}{kn}\frac{1}{\rho n} - a_n} - b_n, \tag{6.22}$$

where, for notation brevity, we define

$$a_n \triangleq \frac{1}{k_n}\left(\sum_{k\in\mathcal{K}_n} \ln g_{k,n}\right) - \ln k_n = \ln \tilde{g}_n - \ln k_n, \tag{6.23}$$

and

$$b_n \triangleq \sum_{k\in\mathcal{K}_n} \frac{1}{g_{k,n}} = \frac{k_n}{\bar{g}_n}. \tag{6.24}$$

In (6.23) and (6.24), $\tilde{g}_n$ and $\bar{g}_n$ represent the geometric mean and harmonic mean of $g_{k,n}$ on active subcarriers at hop n, respectively.

For the moment, let us assume that k_n's are fixed and then both a_n and b_n are constants. Then, the problem of minimizing total power for supporting the target end-to-end transmission rate can be reformulated as **P6-4** only with optimization variables $\{\rho_n,\ n \in \mathcal{N}\}$

$$\begin{aligned} \textbf{P6-4}: \quad p_{\min}(\mathbf{g}) &= \min_{\{\rho_n\}} \sum_{n\in\mathcal{N}} \rho_n p_n^* && (6.25)\\ &= \min_{\{\rho_n\}} \sum_{n\in\mathcal{N}} \rho_n \left(\mathrm{e}^{\frac{R}{kn}\frac{1}{\rho n} - a_n} - b_n\right) \\ &\quad \text{s.t.} \sum_{n\in\mathcal{N}} \rho_n = 1. && (6.26) \end{aligned}$$

Problem **P6-4** can be also solved using Lagrange multiplier method since it is convex. The Lagrangian of this problem is given by

$$L(\boldsymbol{\rho}, \nu) = \sum_{n\in\mathcal{N}} \left(\rho_n \mathrm{e}^{\frac{R}{kn}\frac{1}{\rho n} - a_n} - b_n\right) + \nu\left(1 - \sum_{n\in\mathcal{N}} \rho_n\right),$$

where the Lagrange multiplier ν satisfies constraint (6.26). Applying KKT condition, the optimal time-sharing fraction ρ_n should satisfy

$$\frac{\partial L(\boldsymbol{\rho},\nu)}{\partial \rho_n} = \mathrm{e}^{\frac{R}{kn}\frac{1}{\rho n}-a_n} - \frac{R}{k_n}\frac{1}{\rho_n}\mathrm{e}^{\frac{R}{kn}\frac{1}{\rho n}-a_n} - \nu = 0. \tag{6.27}$$

The closed-form solution to (6.27) is difficult to obtain.

It is known that when the target transmission rate is sufficiently small, the power saving through time adaptation is insignificant [73]. This result motivates us to focus on the high target transmission rate with $R \gg K$. We consider two particular hops, n_1 and n_2. Under the assumption of a high target transmission rate, the equation (6.27) can be approximated by

$$\mathrm{e}^{-a_{n_1}}\frac{R}{k_{n_1}\rho_{n_1}}\mathrm{e}^{\frac{R}{kn_1\rho n_1}} \approx \mathrm{e}^{-a_{n_2}}\frac{R}{k_{n_2}\rho_{n_2}}\mathrm{e}^{\frac{R}{kn_2\rho n_2}}.$$

From the above approximation, we can obtain a ratio

$$\frac{k_{n_1}\rho_{n_1}}{k_{n_2}\rho_{n_2}} \approx 1 + \frac{k_{n_1}\rho_{n_1}}{R}(a_{n_2}-a_{n_1}) - \frac{k_{n_1}\rho_{n_1}}{R}\ln\left(\frac{k_{n_1}\rho_{n_1}}{k_{n_2}\rho_{n_2}}\right). \tag{6.28}$$

Without loss of generality, we assume $a_{n_2} \geq a_{n_1}$, then we have $(k_{n_1}\rho_{n_1})/(k_{n_2}\rho_{n_2}) \geq 1$ from (6.28). Thus, (6.28) leads to

$$1 \leq \frac{k_{n_1}\rho_{n_1}}{k_{n_2}\rho_{n_2}} \leq 1 + \frac{k_{n_1}\rho_{n_1}}{R}(a_{n_2}-a_{n_1}). \tag{6.29}$$

Using the inequality $\ln(x) \leq x-1$ and (6.28), we obtain a lower bound of $(k_{n_1}\rho_{n_1})/(k_{n_2}\rho_{n_2})$ after some manipulations,

$$\frac{k_{n_1}\rho_{n_1}}{k_{n_2}\rho_{n_2}} \geq 1 + \frac{k_{n_1}\rho_{n_1}(a_{n_2}-a_{n_1})}{R\left(1+\frac{k_{n_1}\rho_{n_1}}{R}\right)}. \tag{6.30}$$

Since $R \gg K$, inequalities (6.29) and (6.30) lead to

$$\frac{k_{n_1}\rho_{n_1}}{k_{n_2}\rho_{n_2}} \approx 1 + \frac{k_{n_1}\rho_{n_1}}{R}(a_{n_2}-a_{n_1}).$$

After manipulation, we have

$$\frac{1}{k_{n_2}\rho_{n_2}} - \frac{a_{n_2}}{R} \approx \frac{1}{k_{n_1}\rho_{n_1}} - \frac{a_{n_1}}{R},\ \forall n_1, n_2 \in \mathcal{N}.$$

Let

$$\frac{1}{k_n \rho_n} - \frac{a_n}{R} = \mu, \ \forall n \in \mathcal{N}.$$

We can obtain the approximated but close-form solution to (6.27) as follows

$$\rho'_n = \frac{R}{k_n R\mu + k_n a_n}, \forall n \in \mathcal{N}, \tag{6.31}$$

where μ is determined by constraint (6.26), and can be obtained through binary search. Substituting (6.31) into (6.22), the corresponding transmission power allocated to hop n is given by

$$p'_n = \mathrm{e}^{R\mu} - b_n, \tag{6.32}$$

Furthermore, substituting (6.31) into (6.21) yields the sub-optimal water level for hop n as

$$\lambda_n = \frac{\mathrm{e}^{R\mu}}{k_n}. \tag{6.33}$$

From (6.15), (6.32) and (6.33), we can regard the sub-optimal power allocation as a two-level water filling scheme. First, the power is poured among all the hop according to (6.32) using the water level $\mathrm{e}^{R\mu}$ and the hop with small b_n will be given more power. In each hop, the power obtained from the previous level is then poured among different subcarriers following (6.15), and the water level is equal to $\mathrm{e}^{R\mu}/k_n$.

Consider a special case where $\mathrm{e}^{R\mu}$ is sufficiently large so that all subcarriers are active, i.e., $k_n = K$. It follows immediately from (6.31) that the hops with low geometric mean of channel gains over the subcarriers should be assigned with longer transmission time. Also, it follows from (6.32) one should lower the transmission power for the hops with low harmonic mean of channel gains. An intuitive understanding of this result is that a high priority is given to the hop with poor channel condition to take advantage of "Lazy Scheduling" [23] to prevent this hop in becoming the bottleneck of the whole link. The idea behind "Lazy Scheduling" is that energy required to transmit a certain amount of information decease when prolonging transmission time.

We now relax the assumption made earlier that k_n's fixed and propose an iterative procedure to find the best k_n's for this sub-optimal short-term total power minimization.

Iterative Algorithm of Sub-Optimal Power and Time Allocation (IAS)

1. Initialization of k_n

   Set $k_n = K,\ \forall n \in \mathcal{N}$

2. Binary search for μ for a given $\{k_n,\ \forall n\}$

   (a) Set $high = \mu_{\max}$, $low = \mu_{\min}$

   (b) Let $center = (low + high)/2$, and calculate $\{\rho'_n,\ \forall n\}$ when $\mu = center$ according to (6.31)

   (c) If $\sum_{n\in\mathcal{N}} \rho'_n > 1$, let $low = center$; otherwise, let $high = center$

   (d) Repeat Step 2)-b) and c) until $high - low < \varepsilon''$

3. Find k_n $(\forall n)$ in the set $\{1, \ldots, K\}$ for a given ρ'_n to meet the target rate R based on (6.34)

4. Repeat Step 2) and 3) until k_n's are unchanged

5. Compute $p'_{k,n}$ through substituting (6.21) into (6.15)

6. Obtain the required total power $p'_{\min} = \sum_n \rho'_n (\sum_k p'_{k,n})$

In Step 2)-a) $\mu_{\max}$ and $\mu_{\min}$ represent the upper bound and lower bound of μ, respectively. The exact value $\mu_{\min} = \max_{n\in\mathcal{N}} \left(\frac{1}{k_n} - \frac{a_n}{R}\right)$ can be obtained from the time constraint $0 \le \rho'_n \le 1$. Its upper bound $\mu_{\max}$ could be $\min_{n\in\mathcal{N}} \left(\frac{NR - k_n a_n}{k_n R}\right)$, since if $\mu > \min_{n\in\mathcal{N}} \left(\frac{NR - k_n a_n}{k_n R}\right)$, $\rho'_n < 1/N$ $(\forall n)$ from (6.31), which violates the constraint $\sum_n \rho'_n = 1$.

The implementation of the above algorithm can be done as follows. At the beginning of each time frame, we first assume that the transmission is on for all subcarriers. The central controller searches for μ and broadcast it to all relays. The relays and source node then compute their own transmission time $\{\rho'_n,\ \forall n \in \mathcal{N}\}$ using (6.31) locally. The required power allocation for hop n to meet the target rate should satisfy

$$\sum_{k\in\mathcal{K}} \ln(1 + g_{k,n}p_{k,n}) = \sum_{k\in\mathcal{K}_n} \ln(g_{k,n}\lambda_n) \overset{(a)}{=} \frac{R}{\rho'_n},\ \forall n \in \mathcal{N}. \tag{6.34}$$

The left side of the above equation (a) can be shown to be a monotonically increasing function of λ_n, and is denoted as $z_n(\lambda_n)$. Without loss of the generality, we assume $g_{1,n} \geq g_{2,n} \geq \ldots \geq g_{K,n}$ $(\forall n \in \mathcal{K})$. A λ_n maps to a unique k_n which satisfies that $\frac{1}{g_{kn,n}} \leq \lambda_n \leq \frac{1}{g_{kn+1,n}}$. Thus, we have

$$z_n\left(\frac{1}{g_{k_n,n}}\right) \leq \frac{R}{\rho'_n} \leq z_n\left(\frac{1}{g_{k_n+1,n}}\right).$$

Therefore, the desired k_n in Step 3) can be obtain through binary search in the set of $\{1,\ldots,K\}$ by comparing $z_n(1/g_{k_n,n})$ with R/ρ'_n. The found k_n and the geometric and harmonic mean of channel gains on these k_n subcarriers are returned to the input of (6.31) in the central controller. This procedure repeats until the k_n's are unchanged. Although the convergence of this algorithm cannot be guaranteed theoretically, divergent behaviors were never observed in the simulation. In the following, we shall use simulation to examine the average number of iterations for the algorithm to converge and the average required short-term total transmission power.

In the simulation, SUI channel model for the fixed broadband wireless access channel environments [17] is used and the channel parameters will be detailed in Section 6.3. The simulation is run for 10^3 time frames to evaluate the average performance. The number of subcarriers is set to 16.

Fig. 6.1 shows the average iterations in the outer loop over 10^3 independent channel realizations required for the search of $\{k_n\}$ to converge. It is shown that the

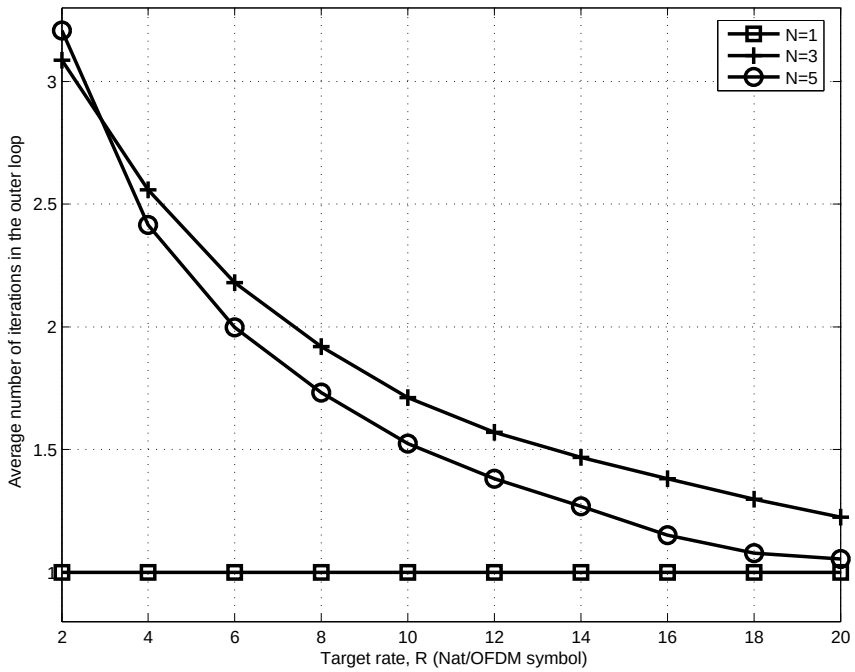

Figure 6.1. Average number of iterations in the outer loop required for the search of $\{k_n\}$

average iteration numbers, denoted as M, is decreasing in R and approaches 1 when the target rate is sufficiently large. It can be explained by the fact that $k_n = K$ when R goes infinity.

Since the binary search for μ in the inner loop involves $\log_2\left(\frac{\mu_{\max}-\mu_{\min}}{\varepsilon''}\right)$ iterations and finding k_n for a given R/ρ_n involves $\log_2(K)$ ones, the total number of iterations required for the IAS can be express as

$$C_{IAS} = M\left\{\log_2\left[\frac{\min_{n\in\mathcal{N}}\left(\frac{NR-k_n a_n}{k_n R}\right) - \max_{n\in\mathcal{N}}\left(\frac{1}{k_n}-\frac{a_n}{R}\right)}{\varepsilon''}\right] + N\log_2(K)\right\} \tag{6.35}$$

Since M is decreasing in R, C_{IAS} is also decreasing in R and upper bounded by a linear function of N. Fig. 6.2 compares average total complexities between TBS and IAS for different R and different N.

Fig. 6.3 compares the average power required to meet the target rate between

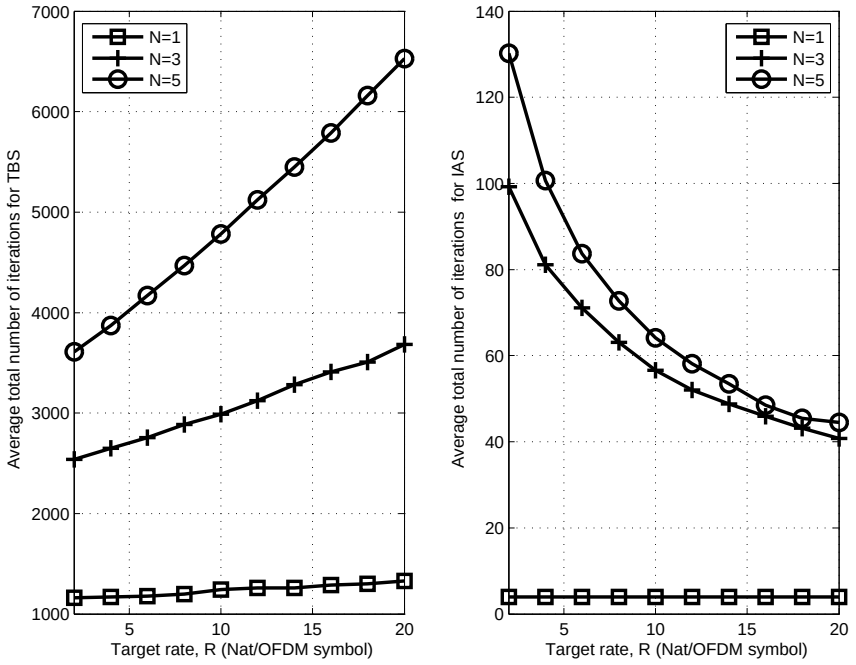

Figure 6.2. Average total number of iterations using TBS and IAS

TBS developed in Section 6.2.1 and its sub-optimal algorithm, IAS. It is shown that IAS serves as a good approximation of TBS, especially for a high target rate.

As we discuss previously, the required controlling signals from the feedback channel are only geometric mean, harmonic mean of $g_{k,n}$ and the number of active subcarriers over each hop instead of $\{g_{k,n}\ , \forall k \in \mathcal{K}, n \in \mathcal{N}\}$ as in TBS. Thus the signalling exchange is greatly reduced when the number of subcarriers is large and/or the target rate is high. Furthermore, since IAS has low complexity and near-optimal power consumption performance, it is a good candidate for a sufficiently high target rate in a real system.

6.2.2 Long-Term Power Threshold Determination

We have discussed the short-term total transmission power minimization. If the transmission is on for every possible channel realization, the long term power con-

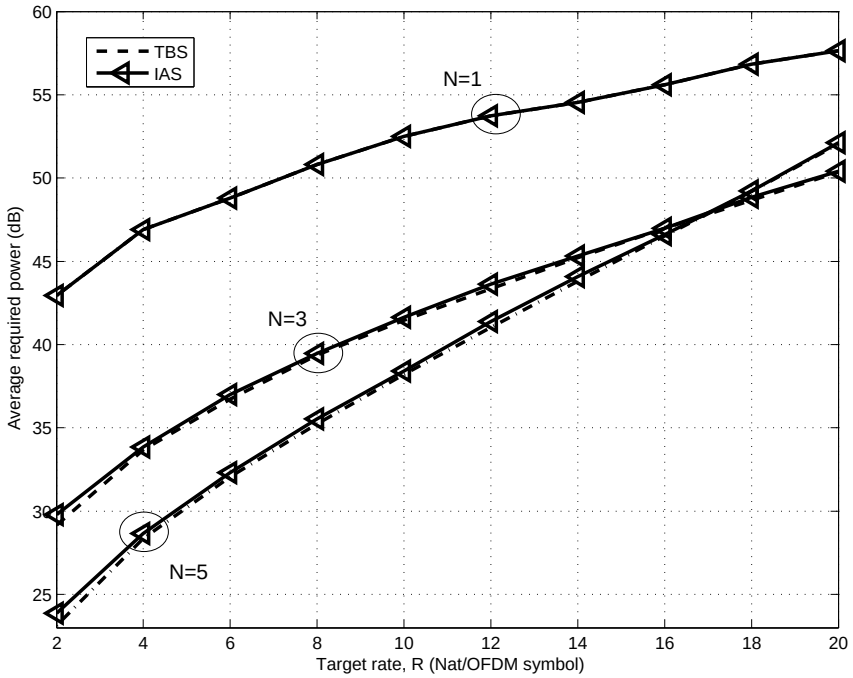

Figure 6.3. Average short-term power required to meet the target rate, R

straint may be violated. Similar to the single user case [11], the optimal power allocated to all hops for **P6-1** with a long term power constraint must have the following structure,

$$p(\mathbf{g}) = \begin{cases} p_{\min}(\mathbf{g}) & \text{with probability } w(\mathbf{g}) \\ 0 & \text{with probability } 1 - w(\mathbf{g}) \end{cases}. \tag{6.36}$$

Thus, the outage probability is $\mathbb{E}\{\mathcal{I}_F[r(\mathbf{g}, \boldsymbol{\rho}, \mathbf{p}) < R]\} = \mathbb{E}[1 - w(\mathbf{g})]$. Then solving **P6-1** is equivalent to finding the optimal weighting function $w(\mathbf{g})$ to the following problem,

$$\begin{aligned} \min_{w(\mathbf{g})} \quad & \mathbb{E}[1 - w(\mathbf{g})] \\ \text{s.t.} \quad & 0 \leq w(\mathbf{g}) \leq 1 \\ & \mathbb{E}[p_{\min}(\mathbf{g}) w(\mathbf{g})] = P. \end{aligned}$$

According to the result of [11, Lemma 3], the optimal weighting function has the form

$$w^*(\mathbf{g}) = \begin{cases} 1 & \text{for } p_{\min}(\mathbf{g}) < s^* \\ w_0 & \text{for } p_{\min}(\mathbf{g}) = s^* \\ 0 & \text{for } p_{\min}(\mathbf{g}) > s^* \end{cases} . \tag{6.37}$$

The power threshold s^* is given by

$$s^* = \sup\{s : \mathcal{P}(s) < P\},$$

and w_0 is given by

$$w_0 = \frac{P - \mathcal{P}(s^*)}{\bar{\mathcal{P}}(s^*) - \mathcal{P}(s^*)},$$

where the region $\mathcal{R}(s)$ and $\bar{\mathcal{R}}(s)$ are defined as

$$\mathcal{R}(s) = \{\mathbf{g} : p_{\min}(\mathbf{g}) < s\},\ \bar{\mathcal{R}}(s) = \{\mathbf{g} : p_{\min}(\mathbf{g}) \leq s\},$$

and the corresponding average power over the two sets are:

$$\mathcal{P}(s) = \mathbb{E}_{\mathbf{g}\in\mathcal{R}(s)}[p_{\min}(\mathbf{g})], \bar{\mathcal{P}}(s) = \mathbb{E}_{\mathbf{g}\in\bar{\mathcal{R}}(s)}[p_{\min}(\mathbf{g})]$$

The resulting minimum outage probability is denoted as

$$P_{\text{out}} = 1 - \text{Prob}\{\mathbf{g} \in \mathcal{R}(s^*)\} - w_0\text{Prob}\{p_{\min}(\mathbf{g}) = s\}.$$

From (6.36) and (6.37), we see that when the minimum total power for all hops required to support the target transmission rate is beyond the threshold s^*, transmission is turned off. When the required power is less than the threshold, the transmission follows the minimum transmission power strategy derived from Section 6.2.1.

The value of s^* can be computed a priori if the fading statistics are known. Otherwise, the threshold can be estimated using fading samples. During the estimation of the threshold, since the channel is assumed to be ergodic, the ensemble average transmission power is equal to the time average

$$\mathbb{E}_{\mathbf{g}\in\mathcal{R}(s)}[p_{\min}(\mathbf{g})] = \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \hat{p}(i),$$

where $\hat{p}_{(i)}$ represents the actual transmission power at time frame i. Thus, the threshold is always adjusted in the opposite direction of $P - \frac{1}{t}\sum_{i=1}^{t}\hat{p}_{(i)}$ as

$$s^*(t+1) = s^*(t)\left[1 + \epsilon\left(P - \frac{1}{t}\sum_{i=1}^{t}\hat{p}_{(i)}\right)\right]. \tag{6.38}$$

where t is the time frame index.

Combining the short-term power minimization and long-term power threshold determination, the full algorithm for APTA is outlined as follows.

APTA

1. Set $t = 1$ and $s^*_{(t)} = P$
2. Search for minimum short-term power (developed in Section 6.2.1)
3. *On-off decision*

   If $p_{\min} > s^*_{(t)}$, turn off the transmission and let $\hat{p}_{(t)} = 0$; otherwise, turn on the transmission and let $\hat{p}_{(t)} = p_{\min}$.
4. *Update the threshold* s^*

$$s^*(t+1) = s^*(t)\left[1 + \epsilon\left(P - \frac{1}{t}\sum_{i=1}^{t}\hat{p}_{(i)}\right)\right]. \tag{6.39}$$

5. Let $t = t + 1$ and return to Step 1).

If TBS in Section 6.2.1 is used in Step 2), we name the optimal APTA as APTA-opt for short. If IAS in Section 6.2.1 is used, we denote it as APTA-sub.

6.3 Numerical results

In this section, we present numerical results to illustrate the performance of the proposed adaptive power and time allocation for OFDM based linear relay networks.

The proposed algorithms, APTA-opt and APTA-sub, are compared with UPTA, FPAT and APFT mentioned in Section 6.1.

We consider an N-hop linear wireless network. The acceptable BER is chosen to be 10^{-5}, which corresponds to 8.2dB SNR gap. We fix the bandwidth to be 1MHz and the end-to-end distance to be 1km. The relays are equally spaced. In all simulations, the channel over each hop is modelled by Stanford University Interim (SUI)-3 channel model with a central frequency at around 1.9 GHz to simulate the fixed broadband wireless access channel environments [17]. The SUI-3 channel is a 3-tap channel. The received signal fading on the first tap is characterized by a Ricean distribution with K-factor equal to 1. The fading on the other two taps follows a Rayleigh distribution. The root-mean-square (rms) delay spread is 0.305μs. Then the coherence bandwidth is approximately 65KHz. Hence, the number of subcarrier K should be greater than 15.2 so that the subcarrier bandwidth is small enough to experience the flat fading. Here we choose $K = 16$. Doppler maximal frequency is set to 0.4 Hz. Intermediate path loss condition ([16, Category B]) is chosen as the path loss model, which is given by

$$PL = A + \alpha \lg \left(\frac{d}{d_n} \right),$$

where $A = 20 \lg(4\pi d_0/\lambda)$ (λ being the wavelength in m), α is the path-loss exponent with $\alpha = (a - bh_b + c/h_b)$. Here $h_b = 30m$ is chosen as the height of the base station , $d_0 = 100m$ and a, b, c are 4, 0.0065 and 17.1, respectively, as given in [16]. The corresponding α will be used in all simulations except the one in Fig Fig. 6.5. In each simulation, 10^4 time frames are used to estimate the outage probability.

Fig. 6.4 shows the end-to-end outage probabilities versus average total transmission power for $R = 1, 20$ and 40 Nat/OFDM symbol using APTA-opt when N varies in the set of $\{1, 3, 5\}$. From the figure, it is shown that multi-hop transmission can help to save total power consumption when the target transmission rate is low (e.g.,

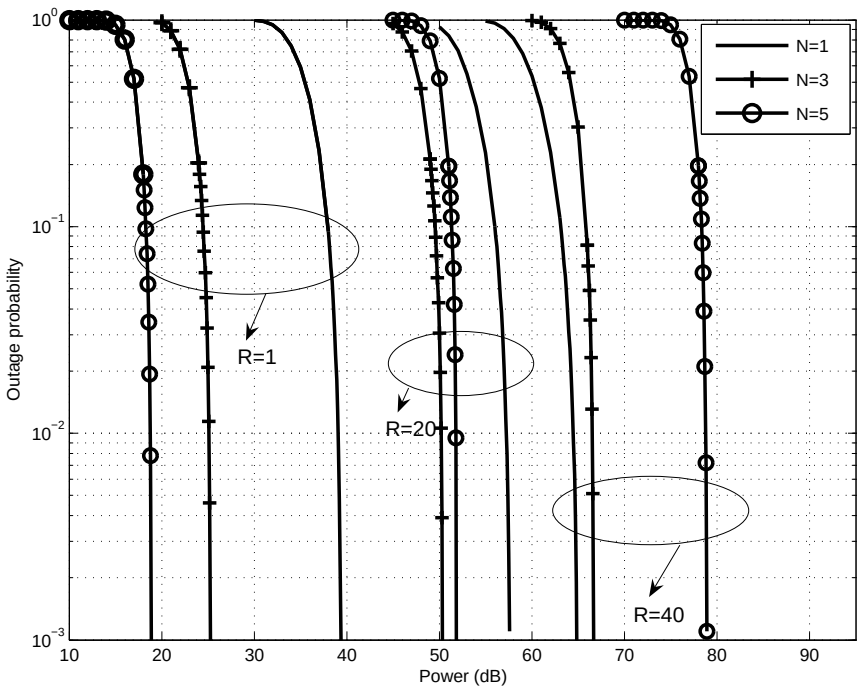

Figure 6.4. End-to-end outage probability vs. average total transmission power under APTA when $K = 16$

$R = 1$) whereas it is better to send data directly to a destination if the target transmission rate is high (e.g., $R = 40$). That can be explained by the following fact. As the number of hops increases, the path loss attenuation on each hop reduces. But the transmission time spent at each hop also reduces since the total frame length is fixed. It is observed from (6.1) that the transmission rate is linear in transmission time and concave in channel gain. Hence, when target transmission rate increases, the loss due to the reduction of transmission time goes beyond the benefit derived from the reduction of path loss attenuation.

Fig. 6.5 shows the relationship between the optimal number of hops in the sense of the minimum required power consumption and the target transmission rate, when the outage probability is fixed to 1%, and the path loss exponent α =2.5 and 4, respectively. It is observed that the optimal number of hops is roughly proportional

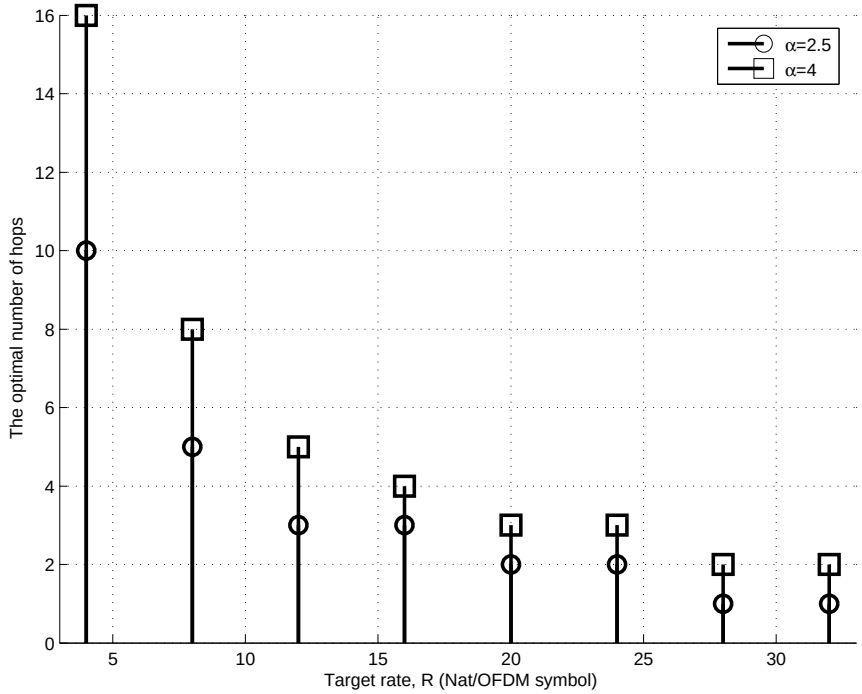

Figure 6.5. The optimal number of hops vs. target rate under APTA when $\alpha = 2.5$ and 4

to the inverse of R. In our simulation, we also find that the optimal number of hops is independent of the end-to-end distance.

Fig. 6.6 and Fig. 6.7 compares the end-to-end outage probabilities of APTA-opt, APTA-sub, FPAT, APFT and UPTA when R =1 and 20 Nat/OFDM symbol. There is a tremendous loss in performance by performing UPTA and FPAT compared with the other three algorithms for both two cases, especially when the the required outage probability is low. UPTA and FPAT have a fixed short-term total power consumption while APTA-opt, APTA-sub and APFT use a long-term power constraint, which yield a significant gain by turning off the transmission when the channels suffer from deep fade. The performance of APTA-sub is even worse than that of APFT when the target rate is low (e.g. R=1). But when the target rate is medium (e.g. R=20) and at the outage probability of 10^{-2}, APTA-sub performs almost the same as APTA-opt and outperforms APFT by 1 to 2dB for N=3 and N=5 .

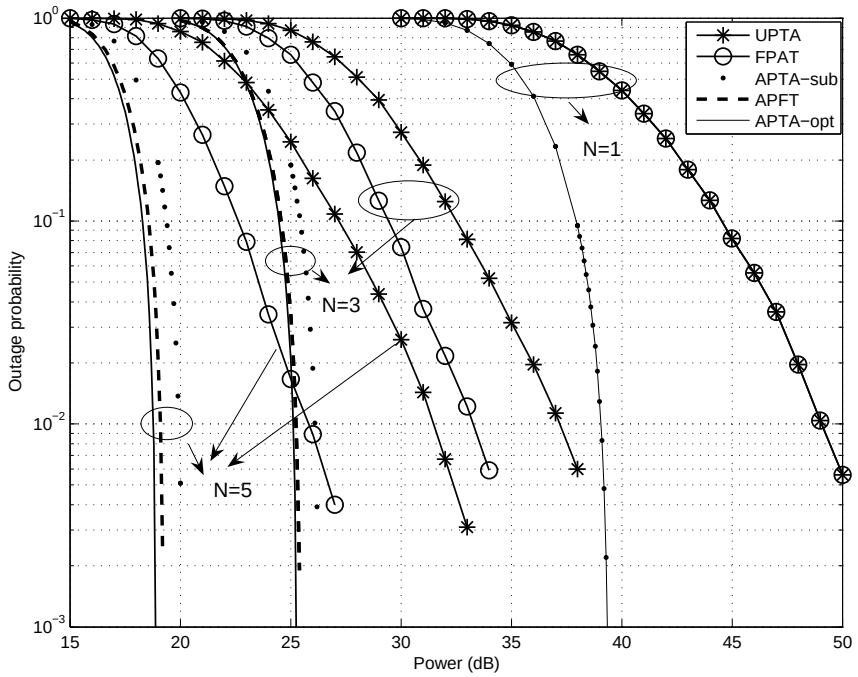

Figure 6.6. End-to-end outage probability vs. average total transmission power under APTA-opt, APTA-sub, APFT, FPAT and UPTA when $K = 16$ and $R = 1$ Nat/OFDM symbol

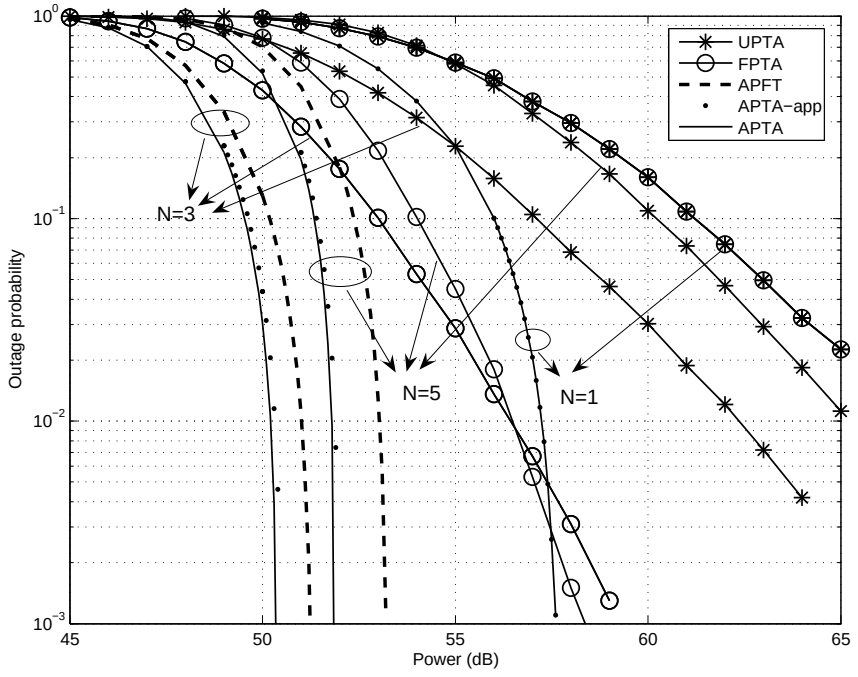

Figure 6.7. End-to-end outage probability vs. average total transmission power under APTA-opt, APTA-sub, APFT, FPAT and UPTA when $K = 16$ and $R = 20$ Nat/OFDM symbol

The numerical results suggest that multi-hop transmission is favorable at a low or medium target rate, whereas a direct transmission from source to destination is preferred if the target rate is high. APFT is a good choice in practice for a low target rate since it has similar performance with APTA-opt and yet is much less complex. For the similar reason, APTA-sub is recommended at a medium target rate.

6.4 Conclusions

In this chapter, we consider adaptive power and time allocations for OFDM based linear relay networks for end-to-end outage probability minimization. First, we derive the minimum short-term total power to meet the target transmission rate. Second, the transmission on-off is determined by comparing the required minimum total power with a threshold, which is selected to satisfy the long-term total power constraint. To avoid high computational complexity and signalling exchange, a sub-optimal solution, APTA-sub, is proposed to approximate APTA-opt when the target rate is sufficiently high. The information needed by the controller are only geometric mean and harmonic mean of channel gains averaged over active subcarriers and the number of active subcarriers. It suggests prolonging the transmission time for the hop with low geometric mean of channel gains averaged over subcarriers while lowering the transmission power for the hop with low harmonic mean. Numerical results demonstrate the relation between the optimal number of hops and the target rate and path loss exponent, and compare the outage probability of APTA-opt and APTA-sub with APFT, FPAT and UPTA. It is shown that APTA-opt, APTA-sub and APFT has a significant performance gain over the other two. It is noted that the resource allocation studied in this chapter is particularly developed for linear relay networks.

Chapter 7

Conclusions and Future Work

This thesis investigates dynamic resource allocation for energy-constrained wireless networks over time-varying channels. For different network architectures and different user's traffic types, resource allocation problems have been formulated in a generalized optimization framework. The possible solutions were developed using different mathematical tools. The results have shown the advantage of adaptive resource allocation over non-adaptive one by exploiting multi-user diversity, channel time diversity, frequency diversity and heterogeneous QoS requirements.

The thesis started with joint optimization of average rate and rate oscillation in a multi-user system over a time-varying wireless fading channel. It has been shown that a utility function that increases with average rate but decreases with rate variance, can be used to facilitate the choice of the combinations of average rate and rate oscillation. A generalized gradient based scheduling algorithm was developed to maximize the proposed utility function from a myopic view (in the sense that the difference of utilities between two consecutive time frames is maximized). This proposed algorithm degenerates to the traditional gradient algorithm when the utility is a function of average rate only. It has also been shown that GGSA is asymptotically optimal when

the transmission rate vector, under an appropriate scaling, converges to a fixed vector as time goes into infinity. However, the conditions of convergence cannot be elucidated theoretically from current research. Further work would be needed to identify the convergence conditions and study the effect of GGSA on other performance metrics, such as, average delay and delay jitter. The numerical results in this study also suggest that GGSA plays a significant role in balancing average rate and rate variance and has a good convergence performance in a time-sharing wireless network. This finding shows that an additional benefit of GGSA might be also obtained in a more generalized network architecture, i.e., resource-shared time-varying wireless network, though more theoretical analysis and verification are still needed at the current stage.

In chapter 4, we looked at the integrated design of quantization and transmission in wireless sensor networks. The optimal solution suggests that the network lifetime can be prolonged by taking advantage of "lazy scheduling" and/or "opportunistic quantization". It can be attributed to the fact that "lazy scheduling" helps to avoid the weak nodes becoming the bottleneck of the whole network, and that "opportunistic quantization" allows the WSN to exploit the limited energy most efficiently. The developed algorithms were shown to contribute the partially distributed property. Finally, this integrated design of quantization and transmission was shown to provide a significant lifetime increase over conventional techniques, and this increase is pronounced in heterogenous sensing environment. A challenge in future work would be to extend the proposed policies to a more general distributed estimation framework as discussed in [72] where the observed signals may be partially correlated. Also the definition of network lifetime considered in this thesis is simple. In a real WSN, the death of some nodes may not affect the functioning of the whole network. More generalized definitions of network lifetime, for example, the time when the L^{th} $(L < K)$ node dies, would be an interesting subject of investigation.

Chapter 5 and Chapter 6 studied the end-to-end resource allocation in an OFDM-

based multi-hop relaying network. The research investigated how to optimize the transmission power on each subcarrier over each hop and the transmission time used by each hop in every time frame so that the end-to-end average transmission rate (outage) is maximized (minimized) under a long-term total power constraint. We have shown that for average rate maximization, the optimal resource allocation strategy has the following features: the power allocation on the subcarriers at each hop follows the water-filling structure, and the water level varies over time and among different hops. Meanwhile, the fraction of transmission time allocated to each hop is adaptive so that the actual transmission volumes over all hops are equal. The analytical expression for optimal resource allocation also suggests that the hops with bad channel condition should be given a high water level, and if the channel of any one hop is in deep fade during a certain time frame, the system may be turned off in order to save power. For outage minimization, the proposed scheme suggests prolonging the transmission time for the hop with low geometric mean of channel gains while lowering the transmission power for the hop with low harmonic mean. Such schemes give a theoretical performance limit over multi-hop networks, but typically involve highly complex computation and involve significant signalling between nodes and central controller, making it difficult to implement. Interestingly, the proposed suboptimal schemes with less computation and signalling can be used to approximate the aforementioned optimal scheme. Thus, a suboptimal scheme may becomes a good candidate in practical systems. However, the study in these two chapters are restricted to linear network, a one-dimensional chain of nodes, under the assumption that no data is allowed to accumulate in each relay. It will be interesting to study the more complex situation where the network has PMP or mesh architecture. More performance analysis, e.g., delay and stability, may be needed when data accumulation is allowed.

Appendix I

Optimality Proof of The Greedy Algorithm

Consider one realization of $\boldsymbol{S}$ and $\rho_i \in \{0, 1/L, 2/L, \ldots, 1\}$. For simplicity, we define $\tilde{u}_i(\rho_i) := \int_{G_{S_i}}^{G_{S_i+1}} u_i(\rho_i(\mathbf{g}), g_i)\mathrm{d}g_i$ and $d_i(\rho) = \tilde{u}_i(\rho) - \tilde{u}_i(\rho - 1/L)$. Let D_L denote the set of L largest elements in $D = \{d_i(\rho) | i = 1, 2, \ldots, N, \rho = 0, 1/L, 2/L, \ldots 1\}$. Then, the Lagrangian can be written as:

$$\sum_{i=1}^{N}(\tilde{u}_i(\rho_i) - \lambda\rho_i) = \sum_{i=1}^{N}[\tilde{U}_i(0) + (d_i(1/L) - \lambda) + \ldots + (d_i(\rho_i) - \lambda)]. \tag{7.1}$$

Let λ be the smallest $d_i(\rho)$ in D_L, then

$$d_i(\rho) - \lambda \begin{cases} \geq 0 & \text{if } d_i(\rho) \in D_L \\ < 0 & \text{otherwise} \end{cases}. \tag{7.2}$$

Therefore the Lagrangian (7.1) is maximized by

$$\rho_i^* = \begin{cases} 0 & \text{if } d_i(1/L) \notin D_L \\ 1 & \text{if } d_i(1) \in D_L \\ \rho & \text{if } d_i(\rho) \in D_L \text{ and } d_i(\rho + 1/L) \notin D_L. \end{cases}. \tag{7.3}$$

Due to the concavity of the utility function, we have $d_i(1/L) > d_i(2/L) > \ldots > d_i(N_i/L) > d_i((N_i + 1)/L) > \ldots > d_i(1)$. Here suppose that $d_i(N_i/L) \in D_L$ and $d_i((N_i + 1)/L) \notin D_L$, then $\sum_i N_i = |D_L|$. Since $\rho_i^* = N_i/L$ from (7.3), $\sum_i \rho_i^* = |D_L|/L = 1$ holds for the chosen λ. ρ_i^* is an optimal solution of Problem (3.26).

Since Step 2) computes the L largest $d_i(\rho)$ in the decreasing order of their values, the greedy algorithm obtains the optimal solution.

Appendix II

Proof of Property 1

Proof: The equivalent problem of **P4-1** is expressed as

$$\begin{aligned} \min_{\boldsymbol{\rho}} \quad & a \\ \text{s.t.} \quad & u_k(\rho_k) \leq a \qquad (7.4) \\ & \sum_{k\in\mathcal{K}} \rho_k = 1, \qquad (7.5) \end{aligned}$$

The optimal solution satisfies the following Karush-Kuhn-Tucker (KKT) conditions,

$$\begin{aligned} & \lambda_k u_k'(\rho_k) + \mu = 0, \ \forall k \in \mathcal{K} \\ & \textstyle\sum_{k\in\mathcal{K}} \lambda_k = 1 \\ & \textstyle\sum_{k\in\mathcal{K}} \rho_k = 1 \\ & u_i(\rho_i) = u_j(\rho_j), \ \forall i,j \in \mathcal{K} \end{aligned} \qquad (7.6)$$

where $\lambda_k, \ \forall k \in \mathcal{K}$ and μ are Lagrange multipliers corresponding to the constraints (7.4) and (7.4). The last condition shows that the optimal solution forces all node lifetime equal. ■

Appendix III

Proof of Property 2

Proof: We define a new generalized function for all k, $\hat{u}(\cdot)$, satisfying $\hat{u}(\rho_k, E_k g_k, \sigma_k^2) := u_k(\rho_k)$, where the product of E_k and g_k is treated as a variable. It can be shown that $\hat{u}$ is a monotonically decreasing function of ρ_k and $E_k g_k$. Therefore, to maintain the equality in (4.14), the first inequality in (4.16) must hold. Consequentially, the second inequality in (4.16) can be obtained through (4.3). ■

Appendix IV

Proof of Property 3

Proof: Suppose that $\sigma_i^2 > \sigma_j^2$ $(i, j \in \mathcal{K})$. It results in $L_i > L_j$ under the assumption $\bar{D}_k = \bar{D}_0$ $(\forall k \in \mathcal{K})$. When $E_k g_k$ $(\forall k \in \mathcal{K})$ are the same, $\hat{u}$ is monotonically increasing in L_k but decreasing in ρ_k. Therefore, we have $\rho_i > \rho_j$ to maintain this equality in (4.14). From the assumption $E_k g_k = E_j g_j$ and the energy constraint in (4.9), we have $g_i p_i \rho_i T_k = g_j p_j \rho_j T_j$. Since it was said that $T_i = T_j$ in (4.14), it is true that $g_i p_i < g_j p_j$. ■

Appendix V

Proof of Theorem 2

Proof: According to the discussion above, if $\boldsymbol{\rho}^*$ is an optimal solution of Problem **P4-2**, it must satisfy the following condition:

$$\begin{cases} u_i(\rho_i^*) = u_j(\rho_j^*), & \forall i, j \in \mathcal{K} \\ \sum_{k=1}^{K} \rho_k^* = 1 \end{cases} . \tag{7.7}$$

We now prove the existence and uniqueness of the above solution $\boldsymbol{\rho}^*$.

Existence and uniqueness: when $x \to 0$, $\sum_k u_k^{-1}(x) \to +\infty$ (here, $u_k^{-1}(\cdot)$ is the inverse function of $u_k(\cdot)$), and when $x \to +\infty$, $\sum_k u_k^{-1}(x) \to 0$. Meanwhile, $\sum_k u_k^{-1}(x)$ is a continuous and strictly decreasing function of x. Thus there exists the unique x^* satisfying $\sum_k u_k^{-1}(x^*) = 1$. It is easy to see that a unique solution $\rho_k^* = u_k^{-1}(x^*)$ satisfies (7.7), and condition (7.7) results in the minimum of the objective function of Problem **P4-2** and **P4-3**.

The following shows that condition (7.7) holds if and only if $\nabla f(\boldsymbol{\rho}^*) = 0$. When K

is sufficiently large and/or the step size (it will be mentioned in latter half of Section 4.2.2) is sufficiently small, $\bar{u}(\boldsymbol{\rho})$ can be viewed as a constant at each updating ρ_k. Hence, we have

$$\nabla_k f(\boldsymbol{\rho}^*) = -2\left\{ u_k'(\rho_k)\left[\bar{u}(\boldsymbol{\rho}) - u_k(\rho_k)\right] + \mu\left(1 - \sum_{k=1}^{K} \rho_k\right)\right\},$$

for all $k \in \mathcal{K}$. When $\nabla f(\boldsymbol{\rho}^*) = 0$, if we assume that $1 - \sum_k \rho_k > 0$, then it holds that $\bar{u}(\boldsymbol{\rho}) - u_k(\rho_k) < 0,\ \forall k \in \mathcal{K}$. However, it contradicts with the definition of $\bar{u}(\boldsymbol{\rho}) := \frac{1}{K}\sum_k u_k(\rho_k)$. Similarly, it can be proven that $1 - \sum_k \rho_k < 0$ does not hold when $\nabla f(\boldsymbol{\rho}^*) = 0$. When condition (7.7) is satisfied, obviously $\nabla f(\boldsymbol{\rho}^*) = 0$. Therefore, $\boldsymbol{\rho}^*$ is the unique optimal solution if and only if $\nabla f(\boldsymbol{\rho}^*) = 0$. ■

Appendix VI

Proof of Lemma 4

First, we will show that $\| \nabla^2 f(\boldsymbol{\rho}) \|_2 \leq B$, where B is a constant. Then Lemma 4 follows from [58, Theorem 9.19].

Because $\| \nabla^2 f(\boldsymbol{\rho}) \|_2^2 \leq \| \nabla^2 f(\boldsymbol{\rho}) \|_\infty \cdot \| \nabla^2 f(\boldsymbol{\rho}) \|_1$ (see [6, page 635]) and $\| \nabla^2 f(\boldsymbol{\rho}) \|_\infty = \| \nabla^2 f(\boldsymbol{\rho}) \|_1$ ($\nabla^2 f(\boldsymbol{\rho})$ is symmetric), we have

$$\begin{aligned}
\| \nabla^2 f(\boldsymbol{\rho}) \|_2 &\leq \| \nabla^2 f(\boldsymbol{\rho}) \|_1 \\
&= \max_i \sum_{j=1}^{K} \left[\nabla^2 f(\boldsymbol{\rho})\right]_{ij} \\
&\leq 2 \max_i \left\{ |u_i''(\rho_i)| \cdot |\bar{u}(\boldsymbol{\rho}) - u_i(\rho_i)| + \right. \\
&\quad \left. |u_i'(\rho_i)|^2 + \frac{1}{K}|u_i'(\rho_i)| \sum_{j=1}^{K} |u_j(\rho_i)| + \mu K \right\} \\
&\leq B. \qquad (7.8)
\end{aligned}$$

The inequality (7.8) results from the fact that $|u_i(\rho_i)|, |u_i'(\rho_i)|$ and $|u_i''(\rho_i)|$ are bounded over the feasible region of $\boldsymbol{\rho}$. Hence, $\nabla f(\boldsymbol{\rho})$ is Liqschitz continuous as desired.

Appendix VII

Algorithm Description

Low-Complexity Algorithm of Optimal Discrete Time-Sharing Fraction Assignment

1. *Initialization*

   Let $v = 0$ (v is the index of the time slot to be allocated in a time frame), and $\rho_k^{(0)} = 0,\ \forall k \in \mathcal{K}$.

2. *Allocate the $(v+1)^{th}$ time slot to the user indexed by i^**

   If $\arg\min_k 1/u(\rho_k)$ is unique, let

   $$i^* = \arg\min_k 1/u(\rho_k).$$

   Otherwise, we randomly choose one of them as i^*.

   Let $\rho_{i^*}^{(v+1)} = \rho_{i^*}^{(v)} + 1/M$ and $\rho_i^{(v+1)} = \rho_i^{(v)}$ for $i \neq i^*$.

3. *Let $v = v + 1$, and return to Step 2) until $v = M$*

4. *The optimal time sharing policy $\boldsymbol{\rho}^*$ is obtained as*

   $$\boldsymbol{\rho}^* = \boldsymbol{\rho}^{(M)}.$$

Appendix VIII

Proof of Property 4

Proof: We assume that $\{\boldsymbol{\rho}^*, \mathbf{p}^*\}$ is the optimal solution to **P4-4**. We let $\Delta_k = x^* - \rho_k^* p_k^*/E_k$ ($\forall k \in \mathcal{K}$). All the nodes are arranged in the non-decreasing order of Δ_k.

Suppose that $0 = \Delta_1 = \ldots = \Delta_i < \Delta_{i+1} \leq \ldots \leq \Delta_K$. We define a new transmission scheme $\{\boldsymbol{\rho'}, \mathbf{p}'\}$. Let $\boldsymbol{\rho'} = \boldsymbol{\rho}^*$ and

$$p'_k = \begin{cases} p^*_k - \epsilon & \text{for } k = 1, 2, \ldots, i, \\ p^*_k + \delta p_k & \text{for } k = i+1, \ldots, K. \end{cases} \tag{7.9}$$

where δp_k satisfies

$$0 < \delta p_k < E_k / \rho^*_k \Delta_k, \quad \forall k \in \mathcal{K}. \tag{7.10}$$

It can be shown that for any δp_k, there always exists a positive value of ϵ such that constraint (4.9) is guaranteed. Hence, if we use the transmission scheme $\{\boldsymbol{\rho'}, \mathbf{p}'\}$, it will result in $x' < x^*$. In other words, the new policy $\{\boldsymbol{\rho'}, \mathbf{p}'\}$ can obtain a longer network lifetime than $\{\boldsymbol{\rho}^*, \mathbf{p}^*\}$, which contradicts the assumption that $\{\boldsymbol{\rho}^*, \mathbf{p}^*\}$ is optimal. ■

Appendix IX

Proof of Property 5

Proof: Since the inequality constraint function in (4.21) can be shown to be concave in ρ_k and the equality constraint in (4.22) is affine, one of Karush-Kuhn-Tucker conditions in (4.23) becomes the necessary condition for the optimality [39]. ■

Appendix X

Proof of Property 6

Proof: When $\sigma_k^2 = \sigma^2$ $(\forall k \in \mathcal{K})$, substituting $p_k = E_k x/\rho_k$ into L_k in (4.3), and substituting L_k into $U_k(L_k)$ in (4.18), we have

$$U_k(L_k) = \frac{1}{W^2\left[\left(1+\frac{g_k E_k x}{\rho_k N_0}\right)^{\rho_k} - 1\right]^2 + \sigma^2}.$$

Thus, $U_k(L_k)$ can be treated as a function of ρ_k, $E_k g_k$ and x. Here, the product of E_k and g_k is viewed as a variable. Then, we can define a function $U(\rho, Eg, x)$ such that $U(\rho_k, E_k g_k, x) := U_k(L_k)$. Let the derivative of $U(\rho, Eg, x)$ with respect to ρ be denoted by $f(\rho, Eg, x) = \partial U(\rho, Eg, x)/\partial\rho$. It can be shown easily that the function $f(\rho, Eg, x)$ is monotonically decreasing in both ρ and Eg. According to *Property* 5, we have $f(\rho_i^*, E_i g_i, x^*) = f(\rho_k^*, E_k g_k, x^*)$, for all i, k. Therefore, to maintain this equality, one must have $\rho_i > \rho_j$ $(\rho_i < \rho_j)$ if $E_i g_i < E_j g_j$ $(E_i g_i > E_j g_j)$. The second inequality in (4.24) can be derived from Property 4. ■

Appendix XI

Proof of Property 7

Proof: When $E_k g_k = E_i g_i$ $(\forall i, k \in \mathcal{K})$, $U_k(L_k)$ can be viewed as a function of ρ_k and σ_k^2. Then, a similar approach used in the proof of *Property* 6 also applies to the proof of the first inequality of (4.25). Since $g_k p_k = \frac{E_k x g_k}{\rho_k}$, the second inequality in (4.25) holds. According to "lazy scheduling" theory, for a fixed amount of energy for transmission, more data can be delivered using longer transmission time and lower transmission power. Thus we have third inequality in (4.25). ■

Appendix XII

Algorithm Description

Partially Distributed Adaptation in Common Source Observation

1. *Set* $low = 0$, $high = x^{\max}$

2. *Update* x *at each node*

   Let $center \leftarrow \lfloor (low + high)/2 \rfloor$ and $x \leftarrow center$.

3. *Assign time sharing fraction*

   (a) Set $\lambda(t) = 0$, $\rho_k(t) = 1/K, \forall k \in \mathcal{K}$

   (b) Compute a new price according to (4.30) at the FC

   (c) Compute a new time sharing fraction according to (4.29)

   (d) Compute the increment of the update at the FC

   Let $t = t + 1$ and $I(t) = \sum_k [U_k(\rho_k(t), x) - \lambda(t)\rho_k] + \lambda(t)$,

   if $I(t) - I(t-1) < \Delta I$, go to 4), otherwise return to (b).

4. *Compare with the target value at FC*

   If $\sum_k U_k(\rho_k, x) > 1/D_0$, set $high \leftarrow center$ at each node.

   Otherwise, $low \leftarrow center$.

5. *Return to Step 2), until* $high - low < \triangle$.

In practice, the transmission power usually has a peak value constraint $p^{\max}$. In that case the maximum possible value of x, $x^{\max} = \max_k p_k^{\max}/E_k$ and the constraint of time sharing fraction, $0 \leq \rho_k \leq 1$, should be changed to $xE_k/p^{\max} \leq \rho_k \leq 1$ if x is given.

Appendix XIII

Proof of Proposition 1

Proof: The equivalence can be proven by contradiction. We assume that $\big(\boldsymbol{\rho}'(\mathbf{g}), \mathbf{p}'(\mathbf{g})\big)$ is an optimal solution to **P5-1**. Let $p'(\mathbf{g}) = \sum_{n\in\mathcal{N}} \rho'_n(\mathbf{g}) \sum_{k\in\mathcal{K}} p'_{k,n}(\mathbf{g})$ denote the corresponding total power consumption function. Suppose that $\big(\boldsymbol{\rho}'(\mathbf{g}), \mathbf{p}'(\mathbf{g})\big)$ is not the optimal solution to **P5-2** for some channel realization $\mathbf{g} \in \mathcal{G}'$ when the short-term power constraint in (5.4) is given by $p = p'(\mathbf{g})$, and $\mathcal{G}'$ is the subset of $\mathcal{G}$. We also assume that $\big(\boldsymbol{\rho}''(\mathbf{g}), \mathbf{p}''(\mathbf{g})\big)$ is the solution to **P5-2** when $\sum_{n\in\mathcal{N}} \rho''_n(\mathbf{g}) \sum_{k\in\mathcal{K}} p''_{k,n}(\mathbf{g}) = p'(\mathbf{g})$ for all $\mathbf{g} \in \mathcal{G}$. From the above assumptions, we have

$$\min_{n\in\mathcal{N}} \left[\rho''_n(\mathbf{g}) \sum_{k\in\mathcal{K}} c_{k,n}\big(p''_{k,n}(\mathbf{g})\big) \right] > \min_{n\in\mathcal{N}} \left[\rho'_n(\mathbf{g}) \sum_{k\in\mathcal{K}} c_{k,n}\big(p'_{k,n}(\mathbf{g})\big) \right], \tag{7.11}$$

for $\mathbf{g} \in \mathcal{G}'$. Taking expectation over all $\mathbf{g} \in \mathcal{G}$ for both sides of (7.11), we have $R\big(\boldsymbol{\rho}''(\mathbf{g}), \mathbf{p}''(\mathbf{g})\big) > R\big(\boldsymbol{\rho}'(\mathbf{g}), \mathbf{p}'(\mathbf{g})\big)$ while $\big(\boldsymbol{\rho}''(\mathbf{g}), \mathbf{p}''(\mathbf{g})\big)$ satisfies all constraints in **P5-1**. This result contradicts with the assumption that $\big(\boldsymbol{\rho}'(\mathbf{g}), \mathbf{p}'(\mathbf{g})\big)$ is an optimal solution to **P5-1**. Therefore, $\big(\boldsymbol{\rho}'(\mathbf{g}), \mathbf{p}'(\mathbf{g})\big)$ must be the optimal solution to **P5-2** when the short-term total power constraint is given by $p'(\mathbf{g})$. This result also indicates that the end-to-end instantaneous transmission rate can be expressed in terms of the short-term total power as $r\big(\mathbf{g}, p'(\mathbf{g})\big)$. Further, if $p'(\mathbf{g})$ is not the optimal solution to **P5-3**, there always exists a $p^*(\mathbf{g})\big(\neq p'(\mathbf{g})\big)$ such that $\mathbb{E}\Big(r\big(p^*(\mathbf{g})\big)\Big) > \mathbb{E}\Big(r\big(p'(\mathbf{g})\big)\Big)$. To sum up, the optimal solution to **P5-1** is the same as the one to **P5-2** where the total power constraint in **P5-2** is the optimal solution to **P5-3**. ■

Appendix XIV

Proof of Proposition 2

Proof: Let ($\mathbf{p}^*$, $\boldsymbol{\rho}^*$) be the optimal power allocation and time-sharing fraction at time frame t. The corresponding instantaneous transmission rate over hop n is denoted as r_n^* ($\forall n \in \mathcal{N}$). Suppose that for a certain hop i, $r_i^* > r_n^*$ ($\forall n \neq i$). Since r_i is a continuous and increasing function of $p_{k,i}$ ($\forall k \in \mathcal{K}$), we can always find a power allocation $p'_{k,i} < p^*_{k,i}$ ($\forall k \in \mathcal{K}$) such that the corresponding r'_i satisfies $r_i^* > r'_i > r_n^*$ ($\forall n \neq i$), while keeping all time-sharing fraction the same ($\mathbf{p}' = \mathbf{p}^*$). That is, we can use a less transmission power to obtain the same instantaneous end-to-end transmission rate at time frame t. If we increase the transmission power in time slot $t+1$ by equally allocating the extra power $\sum_{k\in\mathcal{K}}(p^*_{k,i} - p'_{k,i})$ at time slot t, the corresponding end-to-end transmission rates at time slot $t+1$ satisfy $r'(t+1) > r^*(t+1)$. Thus, $\mathbb{E}(r') > \mathbb{E}(r^*)$. This contradicts the assumption that ($\mathbf{p}^*$, $\boldsymbol{\rho}^*$) is optimal. Therefore, we have Proposition 2, i.e., $r_i^* = r_n^*,\ \forall i, n \in \mathcal{N}$. ■

Appendix XV

Proof of Proposition 3

Proof: We consider two-hop case, the results of which can be generalized to N-hop case. Define

$$\begin{aligned} f_n(p_n) \quad &\triangleq \quad \max_{\{p_{k,n}, k\in\mathcal{K}\}} \mu_n \sum_{k\in\mathcal{K}} \ln(1 + g_{k,n}p_{k,n}) \\ \text{s.t.} \quad &\quad \sum_{k\in\mathcal{K}} p_{k,n} = p_n. \end{aligned} \tag{7.12}$$

Here, $f_n(\cdot)$ denotes the weighted total achievable transmission rate on all subcarriers over hop n, which is a function of total power on all subcarriers over hop n, p_n.

We let $p_n = x$. According to the basic water-filling theorem [22], the optimal $p_{k,n}$ is denoted as

$$p_{k,n} = \left(\frac{\mu_n}{\lambda_n} - \frac{1}{g_{k,n}}\right)^+ \quad \forall k \in \mathcal{K}, \tag{7.13}$$

where λ_n is selected to meet $\sum_{k\in\mathcal{K}} p_{k,n} = x$. Substituting (7.13) into the above equation, we have

$$\frac{\mu_n}{\lambda_n} = \frac{1}{k_n}\left(x + \sum_{k\in\mathcal{K}_n} \frac{1}{g_{k,n}}\right) \tag{7.14}$$

By combining (7.12), (7.13) and (7.14), $f_n(x)$ can be denoted as

$$f_n(x) = \mu_n \sum_{k\in\mathcal{K}_n} \ln\left[\frac{g_{k,n}}{k_n}\left(x + \sum_{k\in\mathcal{K}_n} \frac{1}{g_{k,n}}\right)\right],$$

Therefore,

$$f_n'(x) = \frac{\mu_n k_n}{x + \sum_{k\in\mathcal{K}_n} \frac{1}{g_{k,n}}}$$

Define

$$f_0(p) = f_2(p) - f_1(p),$$

then we have

$$\begin{aligned} f_0'(p) &= \frac{\mu_2 k_2}{p + \sum_{k\in\mathcal{K}_2} \frac{1}{g_{k,2}}} - \frac{\mu_1 k_1}{p + \sum_{k\in\mathcal{K}_1} \frac{1}{g_{k,1}}} \\ &= \frac{(\mu_2 k_2 - \mu_1 k_1)p + \left(\mu_2 k_2 \sum_{k\in\mathcal{K}_1} \frac{1}{g_{k,1}} - \mu_1 k_1 \sum_{k\in\mathcal{K}_2} \frac{1}{g_{k,2}}\right)}{\left(p + \sum_{k\in\mathcal{K}_1} \frac{1}{g_{k,1}}\right)\left(p + \sum_{k\in\mathcal{K}_2} \frac{1}{g_{k,2}}\right)}. \end{aligned} \tag{7.15}$$

If $\sum_{k\in\mathcal{K}_1} \frac{1}{g_{k,1}} > \sum_{k\in\mathcal{K}_2} \frac{1}{g_{k,2}}$, then we have $\mu_1 k_1 > \mu_2 k_2$. Suppose that $\mu_1 k_1 \le \mu_2 k_2$, the numerator of (7.15) is non-negative for any $p > 0$. Thus, $\forall p > 0$, $f_0'(p) > 0$. Therefore, $f_2(p) > f_1(p)$. For $p_1 \ge 0,\ p_2 \ge 0$ and $\rho_1 p_1 + (1-\rho_1)p_2 = p$, we have

$$\begin{aligned} \rho_1 f_1(p_1) + (1-\rho_1) f_2(p_2) &\le \rho_1 f_2(p_1) + (1-\rho_1) f_2(p_2) \\ &\le f_2(p). \end{aligned}$$

The last inequality is due to the concavity of $f_2(p_2)$. This assumption means that all the resource is allocated to hop 2 only. This contradicts the condition that each has the same transmission rate. Therefore, $\mu_1 k_1 > \mu_2 k_2$. ■

Bibliography

[1] "Amendment to IEEE standard for local and metropolitan area networks - part 16: Air interface for fixed and mobile broadband wireless access systems - multihop relay specification," March 2006. [Online]. Available: ieee802.org/16/relay/

[2] R. Agrawal and V. Subramanian, "Optimality of certain channel aware scheduling policies," in *Proc. 40th Annual Allerton Conf. Comm., Control, Comput.*, Monticello, IL, USA, Oct. 2002, pp. 1532–1541.

[3] M. Andrews, L. Qian, and A. Stolyar, "Optimal utility based multi-user throughput allocation subject to throughput constraints," in *Proc. IEEE INFOCOM*, Miami, FL, USA, March 2005.

[4] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "Cdma data qos scheduling on the forward link with variable channel conditions," *Bell Labs Tech. Rep.*, April 2000.

[5] F. Aune, *Cross-layer design tutorial.* Publsihed under Creative Commons License, 2004.

[6] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods.* Athena Scientific, 1997.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, United Kingdom: Cambridge Univ. Press, 2004.

[8] R. Bruno, M. Conti, and E. Gregori, "Mesh networks: commodity multihop ad hoc networks," *IEEE Commun. Mag.*, vol. 43, no. 9, pp. 123–131, 2005.

[9] J. A. Bucklwew, T. Kurtz, and W. A. Sethares, "Weak convergence and local stability properties of fixed step size recursive algorithms," *IEEE Trans. Info. Theory*, vol. 39, no. 3, pp. 966–978, May 1993.

[10] L. Bui, A. Eryilmaz, R. Srikant, and X. Wu, "Joint congestion control and distributed scheduling in multihop wireless networks with a node exclusive interference model," in *Proc. IEEE INFOCOM*, Barcelona, Catalunya, Spain, April 2006.

[11] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channelsl," *IEEE Trans. Info. Theory*, vol. 45, no. 5, pp. 1468–1489, 1999.

[12] Z. Chair and P. K. Varshney, "Distributed bayesian hypothesis testing with distributed data fusion," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 5, pp. 695–699, Sep. 1988.

[13] J. H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc netowrks," in *Proc. IEEE INFOCOM*, Tel Aviv, Israel, March 2000.

[14] M. Chiang, S. H. Low, and A. R. Calderbank, "Layering as optimization decomposition: a mathematical theory of network architectures," *Proc. IEEE*, vol. 95, no. 1, pp. 256–312, Jan. 2007.

[15] L. Dai, B. Gui, and L. J. Cimini Jr., "Selective relaying in OFDM multihop cooperative networks," in *Proc. IEEE WCNC*, Hong Kong, March 2007, pp. 963–968.

[16] V. Erceg, L. Greenstein, S. Tjandra, S. Parkoff, A. Gupta, B. Kulic, A. Julius, and R. Jastrzab, "An empirically based path loss model for wireless channels insuburban environments," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 11, pp. 1205–1211, 8-12 Nov. 1999.

[17] V. Erceg, K. Hari, M. Smith, and D. Baum *et al*, "Channel models for fixed wireless applications," IEEE 802.16.3c-01/29r1, 23 Feb 2001.

[18] G. J. Foschini and J. Salz, "Digital communications over fading radio channels," *Bell System Tech. J.*, vol. 62, no. 2, pp. 429–256, 1983.

[19] G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Tech.*, vol. 42, no. 4, pp. 641–646, Nov. 1993.

[20] A. Fu, E. Modiano, and J. Tsitsiklis, "Optimal energy allocation for delay-constrained data transmission over a time-varying channel," in *Proc. IEEE INFOCOM*, 2003.

[21] D. Fudenberg and J. Tirole, *Game Theory.* The MIT Press, 1991.

[22] R. G. Gallager, *Information theroy and reliable communication.* New York, Wiley, 1986.

[23] A. E. Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *IEEE INFOCOM*, New York, USA, March 2002.

[24] Goldsmith, "Design and performance of high-speed communication systems over time-varying radio channels," Ph.D. dissertation, U.C. Berkeley, 1994.

[25] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Info. Theory*, vol. 43, no. 11, pp. 1986–1992, Nov. 1997.

[26] A. J. Goldsmith and S. G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, no. 10, pp. 1218–1230, Oct. 1997.

[27] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wirel. Commun. Mag.*, vol. 9, no. 4, pp. 8–27, Aug. 2002.

[28] B. S. Gottfried and J. Weisman, *Introduction to optimization theory.* Prentice-Hall, 1973.

[29] T. G. Griffin, F. B. Shepherd, and G. Wilfong, "The stable path problem and interdomain routing," *IEEE/ACM Trans. Netw.*, vol. 10, no. 2, pp. 232–243, Apr. 2002.

[30] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Info. Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[31] ——, "Toward an information theory of large network: An achievable rate region," *IEEE Trans. Info. Theory*, vol. 49, no. 8, pp. 1877–1897, 2003.

[32] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of cdma-hdr a high efficiency-high data rate personal communication wireless system," in *Proc. IEEE VTC*, Tokyo, Japan, May 2000, pp. 1854 – 1858.

[33] H. Jiang, W. Zhuang, X. Shen, and Q. Bi, "Quality-of-service provisioning and efficient resource utilization in cdma cellular communications," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 1, pp. 4–15, Jan. 2006.

[34] C. Jin, D. Wei, and S. Low, "Fast tcp: motivation, architecture, algorithms, performance," in *Proc. IEEE INFOCOM*, vol. 4, 7-11 March 2004, pp. 2490–2501vol.4.

[35] V. Kawadia and P. R. Kumar, "A cautionary perspective of cross-layer design," *IEEE Wirel. Commun. Mag.*, vol. 12, no. 1, pp. 3–11, Feb. 2005.

[36] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[37] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Operations Res.*, vol. 49, no. 3, pp. 237–252, Mar. 1998.

[38] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE ICC*, Seattle, USA, June 1995.

[39] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Proc. 2nd Berkeley Symposium*, 1951.

[40] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Info. Theory*, vol. 50, pp. 3062–3080, 2004.

[41] G. Li and H. Liu, "Resource allocation for OFDMA relay networks with fairness constraints," *IEEE J. Sel. Areas Commun.*, vol. 11, pp. 2061–2069, Nov. 2006.

[42] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels: Part I: ergodic capacity," *IEEE Trans. Info. Theory*, vol. 47, pp. 100–119, 3 2000.

[43] ——, "Capacity and optimal resource allocation for fading broadcast channels: Part II: outage capacity," *IEEE Trans. Info. Theory*, vol. 47, no. 3, pp. 120–145, 2000.

[44] X. Lin and S. Rasool, "Constant time distributed scheduling policies for ad hoc wireless networks," in *Proc. IEEE CDC*, San Diego, CA, USA, Dec. 2006.

[45] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer design in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 1452–1463, 2006.

[46] S. H. Low, "A duality model of tcp and queue management algorithms," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 525–536, Aug 2003.

[47] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug. 2006.

[48] R. J. Marks, A. K. Das, M. El-Sharkawi, and P. A. andA. Gray, "Maximizing lifetime in an energy constrained wireless sensor array using team optimization of cooperating systems," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Honolulu, Hawaii, May 2002.

[49] meraki, "http://meraki.com/," online, 2007.

[50] O. Oyman, J. N. Laneman, and S. Sandhu, "Multihop relaying for broadband wireless mesh networks: From theory to practice," *IEEE Commun. Mag.*, vol. 45, no. 11, pp. 116–122, Nov. 2007.

[51] O. Oyman and S. Sandhu, "Non-ergodic power-bandwidth tradeoff in linear multi-hop networks," in *Proc. ISIT*, Seattle, Washington, USA, July 2006.

[52] R. Pabst, B. Walke, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. Falconer, and G. Fettweis, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 80–89, 2004.

[53] D. Palomar and M. Chiang, "A tutorial to decomposition method for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 1439–1450, 2006.

[54] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 884–895, June 1999.

[55] B. Radunovic and J.-Y. L. Boudec, "Joint scheduling, power control and routing in symmetric, one-dimensional, multi-hop wireless networks," in *WiOpt*, France, March 2003.

[56] T. S. Rappaport, *Wireless Communications: Principles and Practice.* Prentice Hall, 1999.

[57] S. O. Rice, "Statistical properties of a ine wave plus random noise," *Bell Sys. Tech. J.*, vol. 27, pp. 109–157, Jan. 1948.

[58] W. Rudin, *Principles of Mathemetical Analysis.* McGraw-hill, 1976.

[59] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part I: System description," *IEEE. Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, 2003.

[60] S. Shakkottai, T. Rappaport, and P. Karlsson, "Cross-layer design for wireless networks," *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 74– 80, Oct 2003.

[61] M. Sikora, J. N. Laneman, M. Haenggi, J. Daniel J. Costello, and T. E. Fuja, "Bandwidth-and power-efficient routing in linear wireless networks," *IEEE Trans. Info. Theory*, vol. 52, pp. 2624–1633, 2006.

[62] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation," *J. Operations Res.*, vol. 53, no. 1, pp. 12–25, Jan. 2005.

[63] M. Tao, Y. Liang, and F. Zhang, "Adaptive resource allocation for delay differentiated traffics in multiuser ofdm systems," in *Proc. IEEE ICC*, Istanbul, Turkey, June 2006.

[64] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Trans. Info. Theory*, vol. 39, no. 2, pp. 466–478, 1993.

[65] D. Tse, "Forward link multiuser diversity through proportional fair scheduling," *Bell Lab J.*, 1999.

[66] E. Uysal-Biyikoglu, B. Prabhakar, and A. E. Gamal., "Energy-efficient packet transmission over a wireless link," *IEEE/ACM Trans. Netw.*, vol. 10, pp. 487–499, Aug 2002.

[67] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Info. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[68] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 802–811, June 2003.

[69] J. Xiao, S. Cui, Z. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 413–422, Feb. 2006.

[70] J. Xiao and Z. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Info. Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.

[71] J. Xiao, Z. Luo, and J. N. Tsitsiklis, "Data fusion with minimal communication," *IEEE Trans. Info. Theory*, vol. 40, no. 5, pp. 1551–1563, Sep. 1994.

[72] J. Xiao, A. Ribeiro, G. B. Giannakis, and Z. Luo, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Proc. Mag.*, vol. 23, no. 4, pp. 27–41, July 2006.

[73] Y. Yao and G. B. Giannakis, "Energy-efficient scheduling for wireless sensor networks," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1333–1342, Aug. 2005.

[74] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Trans. Veh. Tech.*, vol. 41, no. 1, Feb. 1992.

[75] X. Zhang, M. Tao, and C. S. Ng, "Utility-based wireless resource allocation for wariable rate transmission," *to appear in IEEE Trans. Wirel. Commun.*, 2007.

[76] Q. Zhao, A. Swami, and L. Tong, "The interplay between signal processing and networking in sensor networks," *IEEE Signal Proc. Mag.*, vol. 06, no. 6, pp. 84–93, July 2006.

[77] H. Zimmermann, "OSI reference model-the ISO model of architecutre for open systems interconnection," *IEEE Trans. Commun.*, vol. COM-28, no. 4, pp. 425–432, Apr. 1980.

List of Publications

Journal Papers

[1] Xiaolu Zhang, Meixia Tao and Chun Sum Ng, "Utility-Based Resource Allocation in Time-Shared Wireless Networks for Variable Rate Transmission," *IEEE Trans. Wirel. Commun.*, vol. 7, no. 9, pp. 3292-3296, Sep. 2008.

[2] Xiaolu Zhang, Meixia Tao and Chun Sum Ng, " End-to-end Outage Minimization in OFDM Based Relay Networks", to appear in *IEEE Trans. Commun.*

[3] Xiaolu Zhang, Meixia Tao and Chun Sum Ng, " Transmission schemes for network lifetime maximization in wireless sensor networks", submitted to *IEEE Trans. Commun.*

Conference Papers

[4] Xiaolu Zhang, Wenhua Jiao and Meixia Tao, "End-to-End Resource Allocation in OFDM Based Linear Multi-Hop Networks", to appear in *INFOCOM'08*.

[5] Xiaolu Zhang, Meixia Tao, Wenhua Jiao and Chun Sum Ng, "End-to-end outage probability minimization in OFDM based linear multi-hop networks", in Proc. of *IEEE ICC'08*, Beijing, China, May 2008.

[6] Xiaolu Zhang, Meixia Tao and Chun Sum Ng, "Maximum lifetime transmission in wireless sensor networks for a common source observation", in Proc. of *IEEE SPAWC'08*, Pernambuco, Brazil, July 2008 .

[7] Xiaolu Zhang, Meixia Tao and Chun Sum Ng, "Transmission schemes for lifetime maximization in wireless sensor networks: uncorrelated source observations", in Proc. of *IEEE GLOBECOM'07*, Washington, DC, USA, Nov. 2007.

[8] Xiaolu Zhang, Meixia Tao and Chun Sum Ng, "A generalized gradient scheduling algorithm in wireless networks for variable rate transmission", in proc. of *IEEE GLOBECOM'07*, Washington, DC, USA, Nov. 2007.

[9] Xiaolu Zhang, Meixia Tao and Chun Sum Ng, "Non-cooperative power control for faded wireless ad hoc networks", in proc. of *IEEE GLOBECOM'07*, Washington, DC, USA, Nov. 2007.

[10] Xiaolu Zhang, Meixia Tao, and Chun Sum Ng, "Time sharing policy in wireless networks for variable rate transmission ", in Proc. of *IEEE ICC'07*, Glasgow, UK, June 2007.