Fall 2013

Analysis and Optimization of Cooperative Wireless Networks

Phuong Thanh Tran
Purdue University

Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Phuong Tran

Entitled
Analysis and Optimization of Cooperative Wireless Networks

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

JAMES S. LEHNERT
Chair

DAVID J. LOVE

JAMES V. KROGMEIER

MICHAEL D. ZOLTOWSKI

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): JAMES S. LEHNERT

Approved by: M. R. Melloch 11-20-2013
Head of the Graduate Program Date

ANALYSIS AND OPTIMIZATION OF COOPERATIVE WIRELESS NETWORKS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Phuong T. Tran

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2013

Purdue University

West Lafayette, Indiana

This dissertation is dedicated to my parents.

ACKNOWLEDGMENTS

I would like to express my heartfelt thanks to my advisor, Prof. James S. Lehnert, who taught me how to think of digital communications clearly when I was fortunate enough to be his research assistant, who gave me huge support to my work, from the general ideas to the correction of my manuscript.

This dissertation also owes much gratitude to other members of my Advisory Committee Prof. James Krogmeier, Prof. Michael Zoltowski, and Prof. David Love for taking time out of their busy schedules to review my work as well as giving me the helpful advice to improve my research.

I also would like to send special thanks to Prof. Xiaojun Lin, who taught me the course of Performance Modelling of Computer Communication Network, in which he raised many inspirational questions on convex and nonconvex optimization. One of these questions gave me the motivation to start working on the topic I'm going to introduce in this dissertation.

I would like to acknowledge my labmates Han-Ting Chiang, Mu-Sheng Lin and Hung-Yi Lo for providing me lots of helpful comments and criticism, as well as accepting the non-trivial task of proofreading some portions of this work.

Last, but not least, I would like to thank my parents and other family members for their unconditional love and encouragement, that gave me necessary motivation to overcome the difficulties during my years in graduate school.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

SYMBOLS

$\mathbf{0}$	Zero vector
R	Rate
C	Capacity
$\mathbf{x}$	A vector
$\mathbf{A}$	A matrix
$(\cdot)^T$	Transpose of a vector or matrix
$(\cdot)^H$	Hermitian of a vector or matrix
$[x]^+$	Function $max\{x, 0\}$
$log^+(x)$	Function $max\{ln(x), 0\}$
$\\| \cdot \\|$	Euclidean norm
$E[\cdot]$	Expectation of a random variable
∇f	Gradient of multi-variable function f
$C(x)$	Gaussian capacity function
$p_X(\cdot)$	Probability mass (or density) function of the random variable X
$P_X(\cdot)$	Distribution function of the random variable X
$P_{XY}(\cdot, \cdot)$	Joint distribution function of the random variables X and Y
$P_{Y\\|X}(\cdot)$	Conditional distribution function of Y given X
$I(X;Y)$	Mutual information between X and Y
$Q(\cdot)$	Lattice quantizer

ABBREVIATIONS

AF	Amplify-and-Forward
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BS	Base Station
CDI	Channel Distribution Information
CDMA	Code Division Multiple Access
CF	Compress-and-Forward
CoF	Compute-and-Forward
CSI	Channel State Information
DF	Decode and Forward
DMC	Discrete Memoryless Channel
FDMA	Frequency Division Multiple Access
GP	Geometric Programming
KKT	Karush-Kuhn-Tucker
LDLC	Low Density Lattice Codes
LDPC	Low Density Parity Check
MAC	Medium Access Control
MIMO	Multi-Input Multi-Output
MINLP	Mixed Integer Nonlinear Programming
MIQP	Mixed Integer Quadratic Programming
MS	Mobile Station
MSE	Mean-Squared Error
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access

PHY	Physical
QoS	Quality of Service
SNR	Signal-to-Noise Ratio
SR	Sum Rate
TDMA	Time Division Multiple Access
WSR	Weighted Sum Rate

ABSTRACT

Tran, Phuong T. Ph.D., Purdue University, December 2013. Analysis and Optimization of Cooperative Wireless Networks. Major Professor: James S. Lehnert.

Recently, cooperative communication between users in wireless networks has attracted a considerable amount of attention. A significant amount of research has been conducted to optimize the performance of different cooperative communication schemes, subject to some resource constraints such as power, bandwidth, and time. However, in previous research, each optimization problem has been investigated separately, and the optimal solution for one problem is usually not optimal for the other problems.

This dissertation focuses on joint optimization or cross-layer optimization in wireless cooperative networks. One important obstacle is the non-convexity of the joint optimization problem, which makes the problem difficult to solve efficiently. The first contribution of this dissertation is the proposal of a method to efficiently solve a joint optimization problem of power allocation, time scheduling and relay selection strategy in Decode-and-Forward cooperative networks. To overcome the non-convexity obstacle, the dual optimization method for non-convex problems [1], is applied by exploiting the time-sharing properties of wireless OFDM systems when the number of subcarriers approaches infinity.

The second contribution of this dissertation is the design of practical algorithms to implement the aforementioned method for optimizing the cooperative network. The difficulty of this work is caused by the randomness of the data, specifically, the randomness of the channel condition, and the real-time requirements of computing. The proposed algorithms were analyzed rigorously and the convergence of the algorithms is shown.

Furthermore, a joint optimization problem of power allocation and computational functions for the advanced cooperation scheme, Compute-and-Forward, is also analyzed, and an iterative algorithm to solve this problem is also introduced.

1. INTRODUCTION

1.1 Motivation

One of the most important contributions to the evolution of wireless networks in recent years has been the advent of MIMO technologies, which create the transmission diversity by using multiple receive and transmit antennas. It has been shown that this method can significantly improve the performance of transmission by exploiting the spatial diversity to combat fading [2]. However, today wireless networks require small-size and low-power devices, which cannot be equipped with multiple antennas. In this setting, the cooperation between users in wireless networks becomes an attractive idea.

The idea of cooperative communication has roots in the work of Cover and El Gamal in 1979 [3], and then it is described more rigorously in some papers beginning from 2003 ([4], [5], [6], [7] and [8]). A concise tutorial about cooperative communication can be found in [9]. More theoretical analysis of this technique is introduced in [10]. Briefly speaking, in cooperative communication systems, each wireless user is assumed to transmit data as well as acting as a cooperative agent for another user [9]. The data from each user can reach the base station (BS) by at least two ways: direct transmission to the BS and relayed transmission via another user [1]. It has been shown that this technique can help to enhance the capacity and reliability of transmission systems by exploiting the spatial diversity gain inherent in a multi-user wireless system without the need for multiple antennas at each node.

As mentioned above, the theory of cooperative communication is built from the work of Cover and El Gamal [3] about the capacity of the relay channel, which has been a challenging problem for a couple of decades. Inspired by the early work of Cover and Gamal, many researchers have tried to solve problems involving general relay networks, but there is still no explicit solutions for these problems. Research has also involved finding some communication schemes that can approach the capacity limit while still being implementable with an acceptable complexity.

While solving the capacity problem for general relay networks and finding advanced and implementable schemes that can reach the maximum theoretical capacity of relay networks is complicated and may require a long-term research, researchers are also focusing on how to optimize the performance of current cooperative communication schemes under the constraints of available resources such as transmission power, bandwidth, data rate, etc. Over the last few years, convex optimization theory has provided a powerful tool for the analysis and design of communication systems, and cooperative networks are not the exception. However, not all problems can be solved by the traditional convex optimization tools. One of main challenges is on nonconvexity of the problems in these applications. Specifically, if we consider a joint optimization problem which combines various objectives and constraints, most likely it will be a non-convex problem.

Motivated by these unsolved problems, in this dissertation, algorithms, which are numerically stable and computationally implementable, are proposed to jointly optimize the cooperative communication systems.

1.2 Review of Previous Work

Since 1979, several cooperation strategies have been proposed and studied, including Amplify-and-Forward, Decode-and-Forward, and Compress-and-Forward. Detailed analysis of the capacity of these strategies can be found in [11]. Also, before the paper of Cover and El Gamal, the investigation of the capacity region of some specific relay networks also had been done by R. Ahlswede ([12], [13]) and E.C. van der Meulen [14].

Solving the problems of the capacity region of more complex relay networks has led to the idea of network coding [15]. Ideas from network coding theory have been applied to cooperative communication networks to build good relaying techniques that approach the capacity limit. Also, this has led to lot of research on coded cooperative communications, in which, channel coding or network coding can be used to implement the cooperation between transmitters and relays. Inspired by the invention of LDPC codes [16], which has been proven to approach the Shannon capacity, some coded cooperative communication schemes have been proposed. C. Li and G. Yue proposed a cooperative communication system based on LDPC coding and analyzed its performance [17]. Razaghi and Yu developed a theory that is called parity-forwarding and proposed the Bilayer LDPC code to implement that theory [18], [19]. However, the most recent approach that has attracted most interest from researchers has been the exploiting of interference in multi-user communications by using structured codes. This approach arose naturally from the idea of network coding. In 2009, B. Nazer and M.Gastpar published their work on this problem and proposed a new strategy for cooperative communication networks, namely, the Compute-and-Forward strategy [20], [21]. However, in that paper they only showed the existence of a class of structured codes that can be used to implement that strategy, but they didn't mention how to design those codes in practice.

In addition to the problem of finding new coding schemes for cooperative networks to achieve the capacity limit, recent research in this field is also focusing on the optimization problems in relay networks subject to some constraints on the available wireless resources. Power allocation optimization at the PHY layer has been solved in several papers, for example, [22] and [23]. In the first paper, the authors exploit the convexity of the problems and solve them using dual method for convex optimization problems; while in the latter their approach is based on the geometric programming (GP), a well-studied class of nonlinear and nonconvex optimization that can be readily transformed into an equivalent convex optimization problem. At the MAC layer, the problem of optimizing the scheduling mechanism has been investigated in [24] and [25]. Cross-layer routing optimization algorithm has also been introduced, for instance, in [26]. The optimization problems formulated in these papers are also solved by well-studied convex optimization methods. However, the research on joint optimization or cross-layer optimization is still moderate. One important obstacle is due to the non-convexity of the joint optimization problem.

Several non-convex optimization methods have been proposed for some specific problems. A dual optimization method for non-convex problems that arise in wireless OFDM-based systems has been proposed by W.Yu and R.Lui [1] in 2006. The most important result in this paper is that, if an optimization problem satisfies a special property called the "time-sharing" property, then it can be solved efficiently using the Lagrangian dual method.

The ideas used for optimization problems in Compute-and-Forward have also been mentioned several times since this cooperation scheme was proposed. In [27], the authors argue that the lattice property of the codes introduced by Nazer in his seminal paper about Compute-and-Forward is only applied for integer combinations of codewords, while the combination computed by the channel can be any real number. Nazer solved this problem by scaling the received channel output so that it's close to an integer combination. However, while the larger scaling can make a better approximation, it also results in the amplification of noise. This suggests the problem of optimizing the scaling. A special case of this problem was solved in [28]. In that paper, the authors consider a Compute-and-Forward scheme for a multiple-access relay channel, which includes 2 source nodes communicating with one destination with the support of one relay node.

1.3 Contribution of this Dissertation

Some important contributions have been made in this dissertation. They are outlined below.

First, this research proposes a method to solve the non-convex joint optimization problem of power allocation, scheduling, and the strategy for selecting relays in multi-user relay networks. Several algorithms to implement these results are also proposed.

Three conditions of the channel state information (CSI) are investigated: known CSI, unknown CSI but perfect feedback, and unknown CSI with erroneous feedback.

Secondly, the detailed convergence analysis of the algorithms is also provided. The convergence has been proved both mathematically and numerically. In addition, the condition for the CSI to make the algorithms converge to the optimal solution is established. If the CSI does not satisfy this condition, the error between the solution obtained from the algorithm and the true optimal solution is evaluated.

Finally, the new cooperative scheme, Compute-and-Forward, is also studied. In this dissertation, the optimization of power allocation (or scaling factor) and the selection of the integer coefficients of the linear combination of the codewords computed at the relay nodes is investigated. The network of interest has K source nodes and K relay nodes, which is more general than the case mentioned in [28].

List of early publications from this work

Tran P.T., Lehnert James S., "Joint optimization of relay selection and power allocation in cooperative OFDM networks with imperfect channel estimation," *The Proceedings of the Wireless Communications and Networking Conference, WCNC 2012*, Paris, France, Apr. 2012.

Tran P.T., Lehnert James S., "Joint optimization of power allocation and cooperation in wireless OFDM networks," *The Proceedings of the International Conference on Advanced Technologies in Communications, ATC09'*, Hai Phong, Vietnam, Oct. 2009.

1.4 Organization of this Dissertation

The remainder of this dissertation is organized as follows. The theoretical background on cooperative communication and the capacity theorems about cooperative relay networks are introduced in Chapter 2. The joint optimization problem for power allocation, scheduling and relay selection in cooperative networks, as well as the solutions and algorithms to implement those solutions, are provided in Chapter 3 and

4. After that, the problem of optimization in Compute-and-Forward relay systems, together with the proposed solution and algorithm, are introduced in Chapter 5. Finally, the conclusion of this thesis and the related ideas for future work are mentioned in Chapter 6.

2. THEORETICAL BACKGROUND

This chapter starts with some basic concepts of information theory, focusing on the three-terminal channel, and then generalizing to relay networks. Cooperation strategies and their performance are also briefly investigated. A review of convex optimization theory is introduced at the end of this chapter.

2.1 Relay Channel

The wireless environment is broadcasting in nature. In other words, the signal transmitted from a specific node in a wireless network is heard by all other nearby nodes. Similarly, a destination node can receive the signals from multiple nodes in the network. This process is called multiple access. By exploiting these properties, some immediate nodes in the network (called relays) can help by forwarding the information from source nodes to the destination nodes. This is the basic idea of cooperative communication. This transmission basically happens in two phases: relay nodes receive signals during the source broadcasting (phase 1), and forward them to the destination during multiple access (phase 2). The above ideas are summarized by the three-terminal channel model called the relay channel.

First, let's review some basic concepts and definitions. Fig. 2.1 shows a simple communication link, which consists of one source and one destination [29].

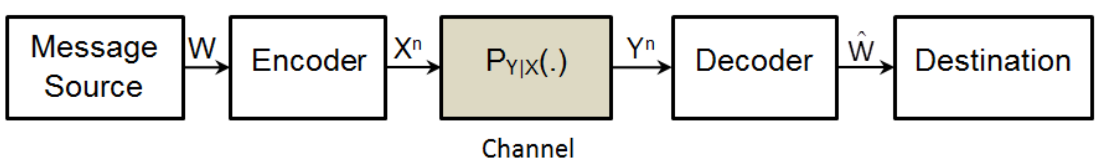

Fig. 2.1. A simple communication model

Suppose that the source wants to transmit a message W that consists of B independent and uniformly distributed bits. This message is first encoded by an encoder, to form an encoded string $X^n = (X_1, X_2, ..., X_n)$, and then this string is transmitted via the channel to the receiver. The signal received at the receiver is denoted by $Y^n = (Y_1, Y_2, ..., Y_n)$. The discrete memoryless channel (DMC) is characterized by a conditional distribution function $P_{Y|X}(\cdot)$, where X and Y are the random variables representing the channel input and output, with alphabets $\mathfrak{X}$ and $\mathfrak{Y}$, respectively. At the receiver, the received signal Y^n is decoded by the decoder to have an estimate $\hat{W}$ that is a function of Y^n.

Definition 2.1.1 *The capacity of the discrete memoryless channel described above is the maximum rate $R = B/n$ bits per channel use for which, for sufficiently large n, there exists a $W - to - X^n$ mapping (an encoder) and a $Y^n - to - \hat{W}$ mapping (a decoder) so that the error probability $Pr\{\hat{W} \neq W\}$ can be made as close to 0 as desired.*

It has been shown that such a capacity exists and is given by the following formula:

$$C = \max_{P_X(\cdot)} I(X;Y) \text{ bits/use}, \tag{2.1}$$

where

$$I(X;Y) = \sum_{a\in\mathfrak{X}, \mathfrak{b}\in\mathfrak{Y}: P_{XY}(a,b)>0} P_{XY}(a,b) \log_2 \frac{P_{XY}(a,b)}{P_X(a)P_Y(b)} \tag{2.2}$$

is the *mutual information* between X and Y. The function P_{XY} represents the joint density of X and Y. The functions P_X and P_Y represent the marginal densities of X and Y, respectively.

A very special case is the AWGN channel described by $Y = X + Z$ with power constraint given by

$$\sum_{i=1}^{n} E\left[|X_i|^2\right]/n \leq P \tag{2.3}$$

where Z is a real Gaussian random variable representing noise with variance N, and $E[\cdot]$ denotes the expectation. The capacity of this channel is given by

$$C = \frac{1}{2} log_2(1 + S) \text{ bits/use} \tag{2.4}$$

where $S = \frac{P}{N}$ is the signal-to-noise ratio (SNR).

The simplest model of a relay channel is the three-terminal relay channel, which includes one relay, in addition to the source node X and the destination node Y. It's assumed that the relay has data of its own to transmit. It's only there to help the receiver.

Even with only one relay, the relay channel capacity is now difficult to determine. The capacity is known only for some special cases, e.g., the physically degraded relay channel ([30], [3]), and the Gaussian relay channel ([31], [32]) (asymptotic capacity).

2.1.1 General Relay Channel

Fig. 2.2 shows a three-terminal relay channel. The channel consists of four finite sets $X, X_1, Y,$ and Y_1, and a collection of probability mass functions $p(y, y_1|x, x_1)$. The symbols x and y represent the input and output of the channel, respectively; y_1 is the relay's observation and x_1 is the input chosen by the relay and depends only on the past observations $(y_{11}, y_{12}, ..., y_{1,i-1})$.

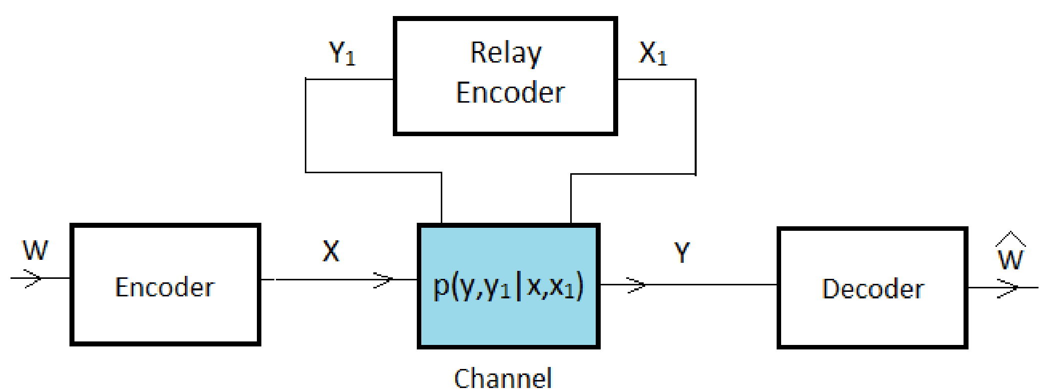

Fig. 2.2. Relay channel

The coding for the relay channel is defined by a set of integers $\mathfrak{M} = \{1, 2, ..., M\}$, a source encoding function $X : \mathfrak{M} \rightarrow \mathfrak{X}^n$, a set of relay encoding functions $\{f_i\}_{i=1}^n$ such that

$$x_{1i} = f_i(Y_{11}, Y_{12}, ..., Y_{1,i-1}), \tag{2.5}$$

and a decoding function

$$g : \mathfrak{Y}^n \rightarrow \mathfrak{M} \tag{2.6}$$

The channel is memoryless in the sense that (Y_i, Y_{1i}) depends on the past only through the current transmitted symbols (X_i, X_{1i}). Thus, for any choice of $p(w), w \in \mathfrak{M}$, code $x : \mathfrak{M} \rightarrow \mathfrak{X}^n$ and relay functions $f_{i\,i=1}^{\,n}$, the joint probability mass function on $\mathfrak{M} \times \mathfrak{X}^n \times \mathfrak{X}_1^n \times \mathfrak{Y}^n \times \mathfrak{Y}_1^n$ is given by

$$p(w, x, x_1, y, y_1) = p(w) \prod_{i=1}^{n} p(x_i|w)p(x_{1i}|y_{11}, y_{12}, ..., y_{1,i-1}) \tag{2.7}$$

The average probability of error is defined by the expression

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{w \in \mathfrak{M}} \Pr\{g(Y) \neq w|w\, sent\} \triangleq \frac{1}{2^{nR}} \sum_{w \in \mathfrak{M}} \lambda(w) \tag{2.8}$$

As mentioned above, this channel consists of a broadcast channel (X to Y and Y_1) and a multiple access channel (X_1 and X to Y). There is no exact capacity formula for the channel capacity of this general relay channel. However, we can apply the max-low-min-cut theorem for general multi-terminal networks to get an upper bound on the capacity.

Theorem 2.1.1 *For any relay channel, the capacity is bounded above by:*

$$C \leq \sup_{p(x,x_1)} \min\{I(X, X_1; Y), I(X; Y, Y_1|X_1)\} \tag{2.9}$$

The first term in (2.9) upper bounds the maximum rate of the multiple access channel from X and X_1 to Y, while the second term upper bounds the rate of the broadcast channel from X to Y and Y_1. However, the destination node Y should decode the relay signal X_1 prior to decoding X, which explains the appearance of the conditioning term X_1 in $I(X; Y, Y1|X_1)$. The detailed proof is introduced in [3].

2.1.2 Degraded Relay Channel

A *degraded relay channel* is a relay channel in which the ultimate receiver y is a degraded version of the relay receiver y_1 (the relay receiver is better than the ultimate receiver), and thus the relay can cooperate to send x. The other case is called the *reversely degraded relay channel*, in which the relay y_1 is worse than y. This case is less interesting, because the relay has no contribution to the destination.

Definition 2.1.2 *The relay channel* $(\mathfrak{X} \times \mathfrak{X}_1, p(y, y_1|x, x_1, \mathfrak{Y} \times \mathfrak{Y}_1)$ *is said to be degraded if:*

$$p(y, y_1|x, x_1) = p(y_1|x, x_1)p(y|y_1, x_1) \tag{2.10}$$

Equivalently, a relay channel is degraded if $p(y|y1, x, x_1) = p(y|y1, x1)$, i.e., $X \rightarrow (X_1, Y_1) \rightarrow Y$ form a Markov chain.

Theorem 2.1.2 *The capacity C of the degraded relay channel is given by*

$$C = \sup_{p(x,x_1)} \min\{I(X, X_1; Y), I(X; Y_1|X_1)\} \tag{2.11}$$

By using the degradedness $I(X; Y, Y_1|X_1) = I(X; Y_1|X_1)$, the proof of the converse directly follows from Theorem 2.1.1. For the achievability part of the proof, see [3].

Gaussian Degraded Relay Channel

Here, we consider an important case of a degraded relay channel, the Gaussian degraded relay channel, which is illustrated in Fig. 2.3. In this figure, Z_1 and Z_2 represent the sequences of i.i.d. normal random variables with zero mean and variances N_1 and N_2, respectively. The ultimate received signal Y is a corrupted version of the relay Y_1, conditioning on X_1.

$$Y_1 = X + Z_1$$

$$Y = Y_1 + X_1 + Z_2$$

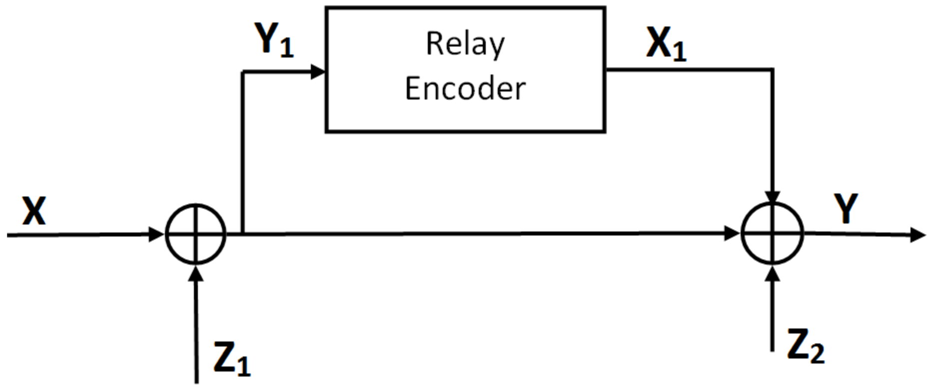

Fig. 2.3. Gaussian degraded relay channel

We also assume that the transmitted power is constrained by

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2(w) \leq P, \ \ w \in \{1, 2, ..., M\} \tag{2.12}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} x_{1i}^2(y_{11}, y_{12}, ..., y_{1,i-1}) \leq P_1, \ (y_{11}, y_{12}, ..., y_{1n}) \in \mathbb{R}^n \tag{2.13}$$

Theorem 2.1.3 *The capacity C^* of the Gaussian degraded relay channel is given by:*

$$C^* = \max_{0\leq\alpha\leq1} \min\left\{C\left(\frac{P + P_1 + 2\sqrt{\overline{\alpha}PP_1}}{N_1 + N_2}\right), C\left(\frac{\alpha P}{N_1}\right)\right\} \tag{2.14}$$

where $\overline{\alpha} = 1 - \alpha$ and $C(x) = \frac{1}{2}log(1 + x)$.

We can find a sketch of the random code that achieves C^* in [3].

Remark 1 *1. If $P_1/N_2 \geq P/N_1$, then $I(X, X1; Y) \geq I(X; Y_1|X_1)$. The relay can forward the cooperative information s to the receiver without error. The capacity $C^* = C(P/N_1)$ is achieved when $\alpha = 1$, which implies that the transmitter does not need to allocate power to send the partition index s. The channel appears to be noise-free after the relay, the rate without the relay $C(P/(N_1 + N_2))$ is increased by the relay to $C(P/N_1)$.*

2. If $P_1/N_2 < P/N_1$, then $I(X, X1; Y) < I(X; Y_1|X_1)$. The relay cannot guarantee perfect transmission of the cooperative information s, then the transmitter must cooperate to send s. Clearly the maximizing $\alpha = \alpha^*$ is strictly less than one, and the capacity is $C^* = C(\alpha^* P_1/N_1)$, where α^* is given by solving

$$\frac{1}{2}ln\left(1 + \frac{P + P_1 + 2\sqrt{\overline{\alpha}PP_1}}{N_1 + N_2}\right) = \frac{1}{2}ln\left(1 + \frac{\alpha P}{N_1}\right) \tag{2.15}$$

The capacity region of the degraded Gaussian relay channel with multiple relays can be obtained by building an inductive argument based on the single-relay capacity Theorem 2.1.3. The details are given in [33].

2.1.3 General Relay Channel with Feedback

In this case, the relay is provided with the information about the y sequence through the feedback link. Therefore, it can decode the x sequence with more reliability than the destination does. In other words, y is a degraded version of the relay signal y_1 with feedback y. Hence, we can consider this as a degraded relay channel with one modification: Y_1 should be replaced by (Y, Y_1).

Theorem 2.1.4 *The capacity C_{FB} of an arbitrary relay channel with feedback is given by*

$$C_{FB} = \sup_{p(x,x_1)} \min\{I(X, X_1; Y), I(X; Y, Y_1|X_1)\} \tag{2.16}$$

In particular, if the channel is degraded or reversely degraded, then feedback does not increase the capacity.

2.1.4 Gaussian Relay Network

In general, the Gaussian relay channel may not be degraded, or may consist of more than one relay (Fig. 2.4(a) shows an example of this case) [31], [34]. The exact capacity region for such general networks is still unknown. However, an asymptotic capacity in the general Gaussian relay network with multiple relays can be derived as in [31]. The "asymptotic" means that the difference between the upper bound and the lower bound of the capacity asymptotically approaches zero as the number of relays goes to infinity. Two additional assumptions are needed for this analysis, however. Firstly, there is a "dead zone" of nonzero radius around the source and the destination node in which there may not be any other node. Secondly, the source node may only send half of the time.

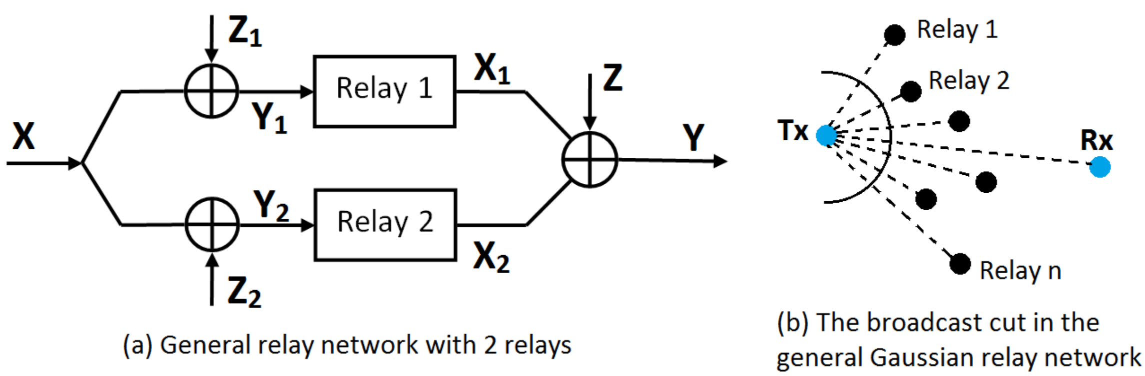

Fig. 2.4. General Gaussian relay network

The upper bound is derived from the max-low-min-cut theorem. The "broadcast cut", which separates the source node from all other nodes, is considered (Fig. 2.4(b)). The lower bound follows from a consideration of almost uncoded transmission of a particular source across the Gaussian relay channel.

Theorem 2.1.5 *(Upper bound) For any particular realization of the random geometry of the network, the capacity of the considered relay network is upper bounded by:*

$$C \leq C_{upper} = \frac{1}{4} log_2 \left(1 + \frac{\|\beta\|^2 P}{N}\right) \tag{2.17}$$

where β is a vector accounting for the fading process, and P is the power constraint for the source signal.

Theorem 2.1.6 *(Lower bound) For any particular realization of the random geometry of the network, the capacity of the considered relay network is at least:*

$$C \geq C_{lower} = \frac{1}{4} log_2 \frac{P}{D_1} \tag{2.18}$$

where D_1 is the mean-square error of the decoded signal.

It is proved that the capacity C lies between C_{upper} and C_{lower}, which meet asymptotically. Hence, the asymptotic capacity of the considered relay network is

$$C_\infty = \frac{1}{4} \log_2 \left(1 + \frac{\|\beta\|^2 P}{N} \right) \tag{2.19}$$

2.2 Network Models and Capacity

In this section, some popular multi-node network models are introduced and their capacity regions are analyzed. Although the problem of finding the capacity region of general relay networks is still unsolved, there have been several significant results for certain networks with AWGN.

2.2.1 Relay Network Models

Depending on the method of relay assistance, different kinds of relay networks can exist. The following are some popular examples of relay networks. For more classification of relay network models, see [35].

a. One-way relay channel: For this model, the information only flows one-way from the source node to the destination node (i.e., the roles of source node and destination node do not change during the communication process).

b. Two-way relay channel: Two nodes can exchange their message with the assistance of a relay node (or maybe several relay nodes). This channel may be half-duplex or full-duplex.

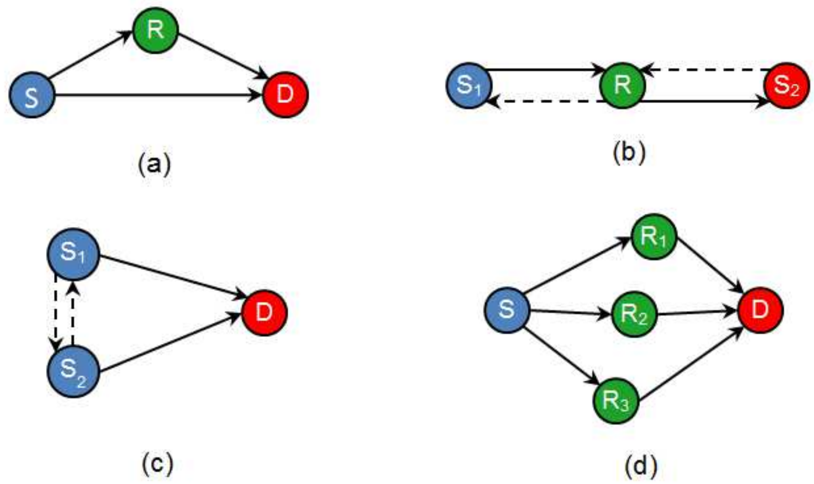

Fig. 2.5. Popular relay networks

c. Multiple-access channel with cooperation: This network has one sink and multiple sources, and the sources can cooperate, that is, some source nodes can act as relays for another source at a specific time. This kind of network includes the relay channel as a special case, by letting node 1 transmit but not receive, and node 2 only forwards the information received from node 1.

d. Parallel relay channel: This network has one source, one sink, and multiple relays. The relays help forwarding the messages from source to sink by one of two mechanisms:

- Simultaneous relaying: At first, the source transmits its signal to all of the relays, then the relays forward this message simultaneously to the sink.

- Successive relaying: In the b^{th} time slot (b = 1, 2,..., B), a non-empty subset of relays is chosen to listen to the source, while the others are sending the new information to the sink. During each time slot, except the first and the last one, both the transmitter and receiver links are active.

2.2.2 Capacity Regions of Some Relay Networks

We extend the concept of channel capacity to general relay networks. Consider a network with M sources, where source m wants to transmit the message W_m including B_m bits, $m = 1, 2, ..., M$, independent of other sources' messages. Let $\mathcal{D}_m$ be the set of nodes that want to decode W_m. We divide the time into n time slots, and denote by $X_{u,i}$ the signal transmitted by node u during time slot i; similarly, $Y_{u,i}$ denotes the signal received by node u during time slot i. Hence, the rate of source message W_m is $R_m = B_m/n$ bits per time slot.

Definition 2.2.1 *The capacity region of the network described above is the closure of the set of rate-tuples $(R_1, R_2, ..., R_M)$ for which, for sufficiently large n, there exist encoders and decoders so that the error probability*

$$\Pr\left[\bigcup_{m=1}^{M}\bigcup_{u\in\mathcal{D}_m}\left\{\hat{W}_m(u) \neq W_m\right\}\right] \tag{2.20}$$

can be made as close to 0 as desired (but not necessarily exactly 0).

Unfortunately, there are still no explicit formulas for the capacity region of general relay networks. However, for some specific cases, especially for AWGN channels, there are several results in capacity analysis of relay networks. This section summarizes those results.

First, we state the common assumptions for our analysis. We consider real-number signals only. The transmitted signals are assumed to have limited power

$$\sum_{i=1}^{n} E\left[|X_{u,i}|^2\right]/n \leq P_u, \quad u = 1, 2 \tag{2.21}$$

where the expectation is over the codewords. The Gaussian noise at the receiver v is denoted as Z_v. It has a variance of N_v. The signal-to-noise ratio (SNR) at receiver u is defined as

$$S_u = \left(\frac{P_u}{N}\right)|h_u|^2 \tag{2.22}$$

For convenience, we introduce the Gaussian capacity function $C(x) = \frac{1}{2}log_2(1+x)$.

Gaussian Multiple-access Channel

Consider the Gaussian MAC channel with 2 sources:

$$Y = h_1 X_1 + h_2 X_2 + Z \tag{2.23}$$

Theorem 2.2.1 *(Capacity region of Gaussian MAC channel) The capacity region of the Gaussian MAC channel is the set of non-negative pairs* (R_1, R_2) *that satisfy the following bounds [30]*

$$R_1 \leq C(S_1) \tag{2.24a}$$

$$R_2 \leq C(S_2) \tag{2.24b}$$

$$R_1 + R_2 \leq C(S_1 + S_2) \tag{2.24c}$$

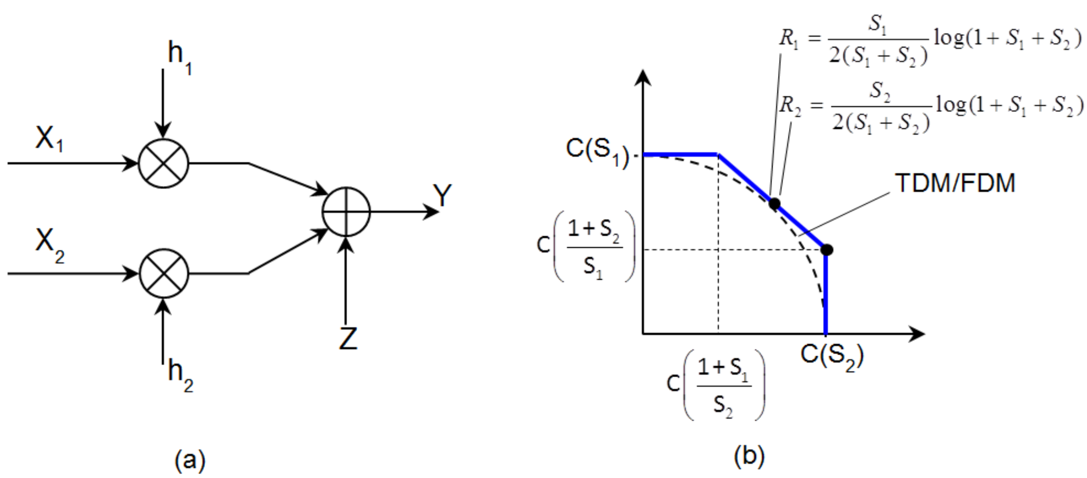

Fig. 2.6. Gaussian MAC channel: (a) model (b) capacity region

Gaussian Broadcast Channel

Although this may not be considered as a relay network, it plays an important role in the capacity analysis of relay networks.

$$Y_1 = h_1 X + Z_1 \tag{2.25a}$$

$$Y_2 = h_2 X + Z_2 \tag{2.25b}$$

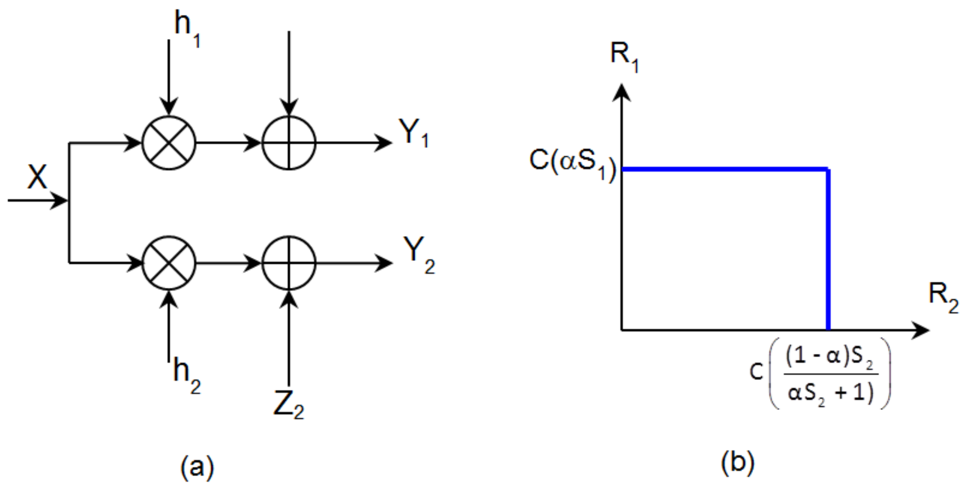

Fig. 2.7. Gaussian broadcast channel: (a) model (b) capacity region

Theorem 2.2.2 *(Capacity region of Gaussian broadcast channel) The capacity region of the Gaussian broadcast channel is the set of non-negative pairs* (R_1, R_2) *such that*

$$R_1 \leq C(\alpha S_1) \tag{2.26a}$$

$$R_2 \leq C(\frac{(1-\alpha)S_2}{\alpha S_2 + 1}) \tag{2.26b}$$

for some $\alpha \in [0, 1]$*, where* $C(x)$ *is the Gaussian capacity function.*

Gaussian Interference Channel

This model consists of 2 sources, which interfere with each other; and 2 sinks, each of which tries to decode the message from the respective source.

$$Y_1 = h_{11}X_1 + h_{12}X_2 + Z_1 \tag{2.27a}$$

$$Y_2 = h_{21}X_1 + h_{22}X_2 + Z_2 \tag{2.27b}$$

Let's denote $I_1 = h_{12}^2 P_2/N_1$ and $I_2 = h_{21}^2 P_1/N_2$. A Gaussian interference channel is said to have strong inteference if $|h_{21}| \geq |h_{11}|$ and $|h_{12}| \geq |h_{22}|$.

Theorem 2.2.3 *(Capacity region of Gaussian inteference channel with strong inteference)*

The capacity region of the Gaussian interference channel with strong interference is the set of non-negative pairs (R_1, R_2) *such that*

$$R_1 \leq C(S_1) \tag{2.28a}$$

$$R_2 \leq C(S_2) \tag{2.28b}$$

$$R_1 + R_2 \leq min\{C(S_1 + I_1), C(S_2 + I_2)\} \tag{2.28c}$$

Cut-Set Bound on Capacity

Because finding the capacity regions of general relay networks is a difficult problem, some researchers have tried to find the capacity bounds. The lower bound can be found by designing protocols and/or codes to achieve some desired rate-tuples. It's more complicated to find the upper bound, because we have to show that this bound holds for all protocols and codes. There is an upper bound that can be applied for most large networks, called the cut-set bound [30].

Consider a set N of network nodes (excluding the sources and the sinks). Let U and V be two disjoint subsets of N. Let (U,V) denote the set of edges connecting from U to V. Consider a set $S \in N$ and denote $\overline{S}$ as the complement of S in N.

Definition 2.2.2 *A cut separating the message* W_m *from one of its estimates* $\hat{W}_m(u)$ *is a pair* $(S, \overline{S})$ *where the* W_m *message node is connected to a node in* S *but not in* $\overline{S}$*, and where the* $\hat{W}_m(u)$ *message-estimate node is connected to a node in* $\overline{S}$*.*

Let $X_S = \{X_u : u \in S\}$, $Y_S = \{Y_u : u \in S\}$, where s and t denote the source node and the destination node, respectively.

Theorem 2.2.4 *(Cut-set bound for general relay channel)*

For a general relay network with a single source and a single destination, the capacity is upper bounded by

$$C \leq C_{cutset} = \sup_{p(x_1, x_2, \ldots, x_n)} \min_{(S,\overline{S}): s \in S, t \in \overline{S}} I(X_S; Y_{\overline{S}} | X_{\overline{S}}) \tag{2.29}$$

Remarks 1. $I(X_S; Y_{\overline{S}}|X_{\overline{S}})$ is concave in $p(x_1, x_2, ..., x_n)$. Hence, finding boundary points of (2.29) is a concave optimization problem.

2. By applying the cut-set bound to the discrete memoryless relay channel in Fig. 2.2, we get the upper bound for the single-relay channel

$$C \leq \max_{p(x,x_1)} \min\{I(X, X_1; Y), I(X; Y_1, Y|X_1)\} \tag{2.30}$$

3. For the Gaussian relay channel, $Y_1 = h_1 X + Z_1$ and $Y = h_2 X + h_{12} X_2 + Z_2$, where h_1, h_2, h_{12} are the channel gain of source-relay, source-destination, and relay-destination, respectively. The random variables $Z_1 \sim N(0, N_1)$ and $Z_2 \sim N(0, N_2)$ represent independent Gaussian noise at relay and destination nodes. By optimizing the bound subject to the power constraints, we can show that it's attained by jointly Gaussian (X, X_1) and the upper bound becomes

$$\begin{aligned} C &\leq \max_{0 \leq \rho \leq 1} \min\{C(S_2 + S_{12} + 2\rho\sqrt{S_2 S_{12}}), C((1-\rho^2)(S_2 + S_{12}))\} \\ &= \begin{cases} C\left(\frac{\left(\sqrt{S_1 S_{12}} + \sqrt{S_2(S_2+S_1-S_{12})}\right)^2}{S_2+S_1}\right) & \text{if } S_1 \geq S_{12}; \\ C(S_2 + S_1) & \text{otherwise} \end{cases} \end{aligned} \tag{2.31}$$

where $S_1 = \frac{h_1^2 P}{N_1}$, $S_2 = \frac{h_2^2 P}{N_2}$ and $S_{12} = \frac{h_{12}^2 P_1}{N_1}$.

2.3 Cooperative Strategies

In wireless networks, the source node can cooperate with other nodes in its vicinity to transmit its message to the destination node. This cooperation creates independent paths between the source and the destination via the introduction of a relay channel as illustrated in Fig. 2.8. This method, which is called cooperative communications ([4], [6]), provides spatial diversity gains to the system, because the users that momentarily experience deep fades in their links to their destinations can utilize the quality channels provided by their partners to achieve the desired Quality of Sevice (QoS).

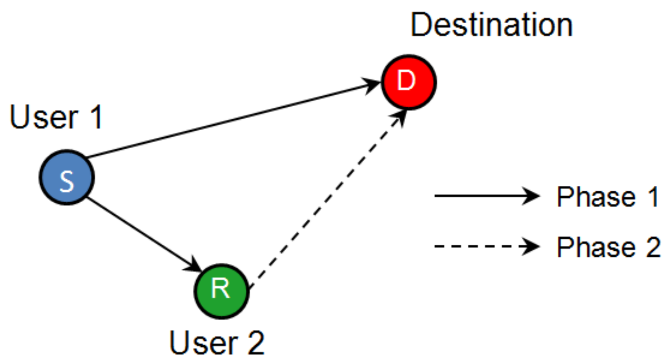

Fig. 2.8. Three-node cooperative network

A cooperative communication scheme is typically modeled with two orthogonal phases so that there is no interference between the two phases. This is illustrated in Fig. 2.8, with two users cooperate in transmitting their messages to the destination. In the first transmission phase, either user 1 or 2 transmits its own message to the destination while the partner, who acts as a relay, receives the message simultaneously due to the broadcast nature of wireless networks. In the second phase, the relay forwards the information that it received in the previous phase to the destination, where optimal combining is then performed for detection. An important aspect of this two-phase process is how the relay processes the signal received from the source node and transfers it to the destination. These different processing schemes result in different cooperative strategies.

Cooperative strategies can be categorized as fixed relaying or adaptive relaying. In fixed relaying, the channel resources are allocated between the sources and the relays in a fixed manner. This is easy to implement, but has the disadvantage of low bandwidth efficiency. For the case in which the source-destination link is not very bad, the relays are not really needed, so the resources allocated to relays is wasted. Some fixed relaying strategies that have been studied in the literature ([4], [36]) are Decode-and-Forward (DF), Amplify-and-Forward (AF) and Compress-and-Forward (CF) [11] methods. Adaptive relaying techniques, such as Selective Relaying, and Incremental Relaying [36], try to overcome the above problem by dynamically allocating the resources to the relays and switching between the fixed strategies depending on the channel conditions of the source-relay, source-destination, and relay-destination links.

In the following section, we consider several types of cooperative strategies and provide the capacity analysis for each of them.

There are some assumptions to make before starting the analysis. First, we consider the full-duplex communication, i.e, devices can transmit and receive at the same time in the same frequency band. Second, the channel state information (CSI) is assumed to be available at the receiver but not at the transmitter. The basic coop-

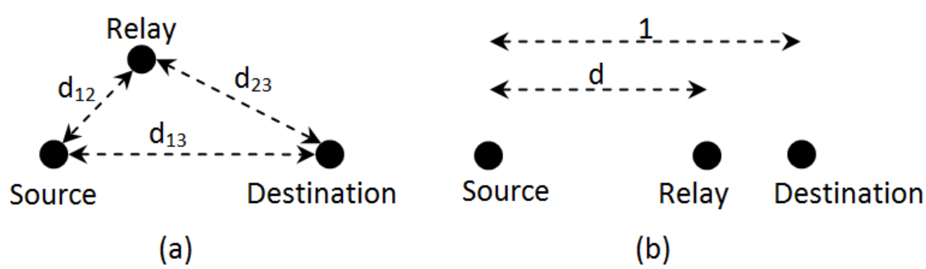

Fig. 2.9. Basic cooperative model: general geometry and linear geometry

erative model consists of one source, one destination, and one relay (see Fig. 2.9). For numerical evaluations, the linear geometry is chosen as in Fig. 2.9(b). The distance between the source and the destination is $d_{13} = 1$, and the relay is a distance $d_{12} = d$ to the right of the source. Two kinds of channels are considered: (1) no fading, and (2) fast Rayleigh fading. The following equations describe the channels:

$$Y_2 = \frac{H_{12}}{|d|^{\alpha}/2} X_1 + Z_2 \tag{2.32a}$$

$$Y_3 = H_{13} X_1 + \frac{H_{23}}{|1-d|^{\alpha/2}} X_2 + Z_3 \tag{2.32b}$$

For no-fading channels, the parameters H_{uv} are constants. For Rayleigh fading channels, H_{uv} are independent Gaussian random variables with zero mean and unit variance (consider at a specific time).

2.3.1 Amplify-and-Forward

For this strategy, the relay just transmits the amplified version of the signal received from the source

$$X_{2,i} = aY_{2,i-1} \tag{2.33}$$

where a is the gain which is chosen to satisfy the power constraint for the relay node. We have the expression

$$\begin{aligned} Y_{3,i} &= \frac{H_{13,i}}{|d_{13}|^{\alpha/2}} X_{1,i} + \frac{H_{23,i}}{|d_{23}|^{\alpha/2}} X_{2,i} + Z_{3,i} \\ &= \frac{H_{13,i}}{|d_{13}|^{\alpha}/2} X_{1,i} + a \frac{H_{12,i-1} H_{23,i}}{(|d_{12}|^{\alpha/2}).(|d_{23}|^{\alpha/2})} X_{1,i-1} + a \frac{H_{23,i}}{|d_{23}|^{\alpha/2}} Z_{2,i-1} + Z_{3,i} \end{aligned} \tag{2.34}$$

The power constraint becomes the inequality given by

$$|a|^2 \leq \frac{P_2}{N + P_1 E\left[|H_{12}|^2\right] / |d_{12}|^{\alpha}} \tag{2.35}$$

The rate of this strategy can be found by performing water-filling optimization (see [11]). The capacity of the relay channel with the Amplify-and-Forward strategy is

$$C_{AF} = log_2(1 + S_{AF}) \text{ bits per time slot} \tag{2.36}$$

where $S_{AF} = \frac{|H_{12}|^2 P_1}{N|d_{12}|^{\alpha}} . \frac{|H_{23}|^2/|d_{23}|^{\alpha}}{1+|H_{23}|^2/|d_{23}|^{\alpha}}$.

Fig. 2.12 shows the results for the channels with no fading and for the linear topology. Here, $P_1/N = P_2/N = 10, H_{uv} = 1$ for all (u,v), and $\alpha = 2$.

The AF strategy may perform better by using the parallel relays. Suppose there are T relays and there is no channel from source node to destination node. Each relay has the power constraint P. We use $0 < d < 1$, and $H_{uv} = 1$ for all (u,v) as before. It has been shown that [35].

$$log_2(T) + log_2\left(\frac{P_1}{N} . \frac{|a|^2}{1 + |a|^2}\right) \leq C \leq 2log_2(T) + log_2\left(\frac{1}{T^2} + \frac{P}{N}\right) \tag{2.37}$$

Thus, C grows as $klog_2(T)$ with T, where $1 \leq k \leq 2$, and AF achieves the scaling law up to a constant factor [35].

2.3.2 Compress-and-Forward

For this strategy, the information is transmitted in blocks with the structure described in Fig. 2.10. In every block b, the source encodes the message w_b by the encoder $x_1(\cdot)$ and transmits $x_1(w_b)$ over the wireless channel. The relay observes $y_{2,b}$ in block b and quantizes it to $\hat{y}_2(s_{b-1}, s_b)$ using its quantization codebook $\hat{y}_2$. Then, the quantized bits s_b are encoded by the relay to $x_2(s_b)$ and transmitted in block b+1 ($\hat{y}_2(s_{b-1}, s_b)$ has 2 indices because of correlation with $x_2(s_{b-1})$). The destination receives the signal $y_{3,b+1}$. It decodes the index s_b first and then uses that to decode w_b, using the codebook $\hat{y}_2(s_{b-1}, s_b)$ and the signal $y_{3,b}$. In the last block, the source transmits the default codeword $x_1(1)$.

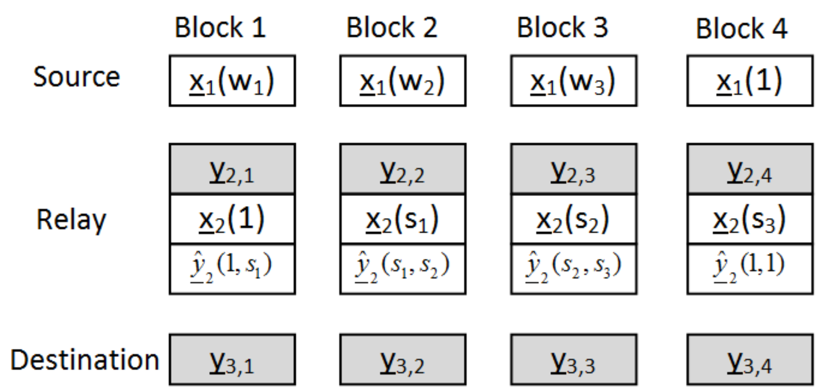

Fig. 2.10. Block structure of Compress-and-Forward strategy

To improve the capacity, a more complicated quantization and destination decoding should be used [3]. Instead of transmitting $x_2(s_b)$, the relay transmits a hash $h(s_b)$ in block b+1 after encoding it to $x_2(h(s_b))$. It also finds a quantization $\hat{y}_2(h(s_{b-1}), s_b)$ in block b. This is called a binning strategy, and $h(s_b)$ is called the bin index.

Using Shannon's rate-distortion theory, we can show that the destination can decode the message w_b with any rate satisfying

$$R = I(X_1; \hat{Y}_2 Y_3 | X_2) \tag{2.38}$$

subject to

$$I(\hat{Y}_2; Y_2|X_2 Y_3) \leq I(X_2; Y_3) \tag{2.39}$$

where the joint probability distribution of the random variables is factored as

$$P(x_1, x_2, y_2, y_3, \hat{y}_2) = P(x_1)P(x_2)P(y_2, y_3|x_1, x_2)P(\hat{y}_2|x_2, y_2) \tag{2.40}$$

For AWGN channels, X_1 and X_2 are Gaussian distributed. Let's choose $\hat{Y}_2 = Y_2 + \hat{Z}_2$ where $\hat{Z}_2 = \hat{Z}_{2R} + j\hat{Z}_{2I}$ and $\hat{Z}_{2R}, \hat{Z}_{2I}$ are independent Gaussian random variables with variance $\hat{N}_2/2$. Then, the resulting rate R is

$$R = log_2\left(1 + \frac{P_1|H_{12}|^2}{d_{12}^{\alpha}(N + \hat{N}_2)} + \frac{P_1|H_{13}|^2}{d_{12}^{\alpha}N}\right) \tag{2.41}$$

where $\hat{N}_2$ is chosen to be

$$\hat{N}_2 = N.\frac{P_1|H_{12}|^2/d_{12}^{\alpha} + P_1|H_{13}|^2/d_{13}^{\alpha} + N}{P_2|H_{23}|^2/d_{23}^{\alpha}} \tag{2.42}$$

to satisfy (2.39) with equality.

2.3.3 Decode-and-Forward

Decode-and-Forward is a cooperative strategy such that the relay completely decodes the original source message w_b in block b before encoding the message using its own codebook and retransmitting it in block b+1. The block structure of this strategy is shown in 2.11.

The source messages is encoded using *block Markov encoding* in which the source codeword generated in block b, $\underline{x}_1(w_{b-1}, w_b)$, depends on both the messages in block b and in the previous block b-1. In block b, the relay knows the message w_{b-1} and uses this knowledge to decode w_b. Then in block b+1, the relay transmits the encoded message $x_2(w_b)$. The destination receives the signal $\underline{y}_{3,b}$ in block b and the signal $\underline{y}_{3,b+1}$ in block b+1.

For the relay and the destination to be able to decode the message, the rate must be respectively upper bounded by

$$R \leq I(X_1; Y_2|X_2), \tag{2.43}$$

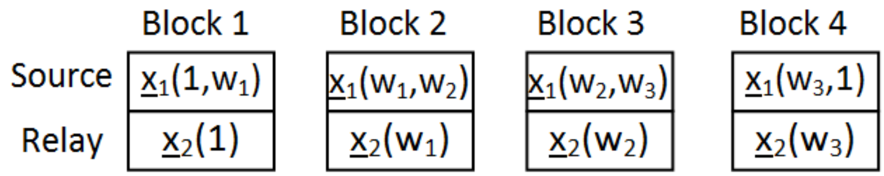

Fig. 2.11. Block structure of Decode-and-Forward strategy

and

$$R \leq I(X_1; Y_3|X_2) + I(X_2; Y_3) = I(X_1X_2; Y_3) \tag{2.44}$$

In summary, the maximum achievable rate is

$$R = \max_{P_{X_1X_2}(.)} \min\{I(X_1; Y_2|X_2), I(X_1X_2; Y_3)\} \tag{2.45}$$

Specifically, consider the DF strategy for a full-duplex Gaussian relay channel. We can construct the codebook $\underline{x}_2(.)$ by superposing the codewords from a Gaussian codebook $\underline{x}_1'(.)$ to codewords from a Gaussian codebook $\underline{x}_2(.)$ scaled by β, i.e, $\underline{x}_1(w_{b-1}, w_b) = \underline{x}_1'(w_b) + \beta\underline{x}_2(w_{b-1})$. The source power is P_1, the relay power is P_2, the codewords in $\underline{x}_1'(.)$ use power $P_1' \leq P_1$, and the scaled codewords use power $P_1 - P_1'$. Hence, $\beta = \sqrt{(P_1 - P_1')/P_2}$. When the destination decodes w_b, the codeword $\underline{x}_1'(w_{b+1})$ is treated as interference. The achievable rate is:

$$\begin{aligned} R = \max_{\rho} \min &\left\{ \log_2\left(1 + \frac{P_1|H_{12}|^2(1-|\rho|^2)}{d_{12}^{\alpha}N}\right), \right. \\ &\left. \log_2\left(1 + \frac{P_1|H_{13}|^2}{d_{13}^{\alpha}N} + \frac{P_2|H_{23}|^2}{d_{23}^{\alpha}N} + \frac{2\sqrt{P_1P_2}\,\mathrm{Re}\{\rho H_{13}H_{23}^*\}}{d_{13}^{\alpha/2}d_{23}^{\alpha/2}N}\right)\right\} \end{aligned} \tag{2.46}$$

where $\rho = E[X_1X_2^*]/\sqrt{P_1P_2} = \sqrt{(1 - P_1'/P_1)}$ is the correlation coefficient of X_1 and X_2.

In [35], the above strategies are compared together for the case in which $P_1/N = P_2/N = 10, H_{uv} = 1$ for all (u,v), and $\alpha = 2$. Assume that the network topology is the linear geometry. The achievable rates from different strategies are shown in Fig. 2.12,

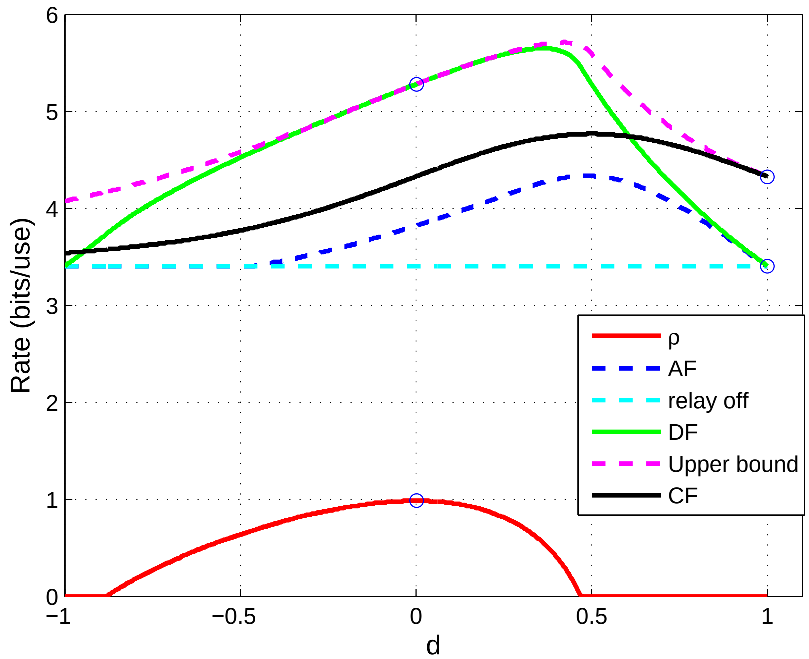

Fig. 2.12. Comparison of cooperative strategies

which is a reproduced version of Fig.4.9 from [35]. It can be observed that CF and DF always outperform AF. The CF strategy works well when the relay is close to the destination ($d \approx 1$), and it reaches the upper bound curve at d = 1. On the other hand, the DF strategy works well when the relay is close to the source ($d \approx 0$) and can reach the upper bound curve at $d = 0$. In general networks, we can have multiple relays, and the relays that are close to the destination should use the CF strategy, while the others which are close to the source should choose the DF strategy.

2.4 Convex Optimization

Optimization is an important tool in the development of wireless communication resource allocation. In particular convex optimization can lead to interesting analytical results and is used in this thesis. A convex optimization problem can be formulated as

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, i = 1, 2, ..., m \end{aligned} \tag{2.47}$$

where the functions $f_i(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}, i = 0, 1, ..., m$ are convex, i.e., satisfy

$$f_i(\alpha\mathbf{x} + \beta\mathbf{y}) \leq \alpha f_i(\mathbf{x}) + \beta f_i(\mathbf{y}), \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \forall \alpha, \beta \in \mathbb{R}^+ : \alpha + \beta = 1.$$

Here, $\mathbf{x}$ is called the optimization variable, $f_0(\mathbf{x})$ is called the objective function, and $f_i(\mathbf{x})$ are constraint functions. A vector $\mathbf{x}^*$ is called the optimal solution if it has the smallest objective values among all vectors that satisfy the constraints; $f^* = f_0(\mathbf{x}^*)$ is called the optimal value of the problem.

The well-known method for solving convex optimization problem is the Lagrangian dual method. The basic idea is to take the constraints into account by adding the weighted sum of the constraint to the objective function. Let a non-negative dual variable λ_i associate with each constraint $f_i(\mathbf{x}) \leq 0$ and let $\boldsymbol{\lambda}$ denote the vector $[\lambda_1, \lambda_2, ..., \lambda_m]^T$. The Lagrangian associated with the problem (2.47) is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) \tag{2.48}$$

The Lagrange dual function is defined as

$$g(\boldsymbol{\lambda}) = \inf_{\mathbf{x} \in \mathbb{R}^n} (L(\mathbf{x}, \boldsymbol{\lambda})) \tag{2.49}$$

It's easy to see that $g(\boldsymbol{\lambda})$ is a lower bound of the optimal value f^*:

$$\begin{aligned} f_0(\mathbf{x}) & \geq f_0(\mathbf{x}) + \textstyle\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) \\ & \geq \inf_{\mathbf{x} \in \mathbb{R}^n} (f_0(\mathbf{x}) + \textstyle\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x})) \\ & = g(\boldsymbol{\lambda}) \end{aligned} \tag{2.50}$$

where the first inequality comes from the definition of the dual variable, $\lambda_i \geq 0$. The result inequality holds for any feasible $\mathbf{x}$, so $f^* \geq max\{g(\boldsymbol{\lambda})\}$.

We define the *dual optimization problem* as

$$\begin{aligned} &\text{maximize} \quad g(\boldsymbol{\lambda}) \\ &\text{subject to} \quad \lambda_i \geq 0, i = 1, 2, ..., m \end{aligned} \tag{2.51}$$

Obviously, this dual optimization problem is convex, whether the functions $f_i(x)$ are convex or not. Let g^* be the optimal value of (2.51), then $f^* - g^*$ is always non-negative and called the *duality gap*. One of the most important results in convex optimization is that when $f_i(\mathbf{x})$ are convex, and the problem satisfies some technical conditions, then the duality gap is equal to zero. There are many results that establish these technical conditions on the problem. One simple condition is called *Slater's condition*: there exists a feasible $\mathbf{x}$ such that $f_i(x) < 0, \forall i = 1, 2, ..., m$. This is stated in the following theorem [37].

Theorem 2.4.1 *If the objective and constraint functions $f_i(\mathbf{x})$ are convex, and the problem* (2.47) *satisfies Slater's condition, then the strong duality holds, i.e., the duality gap is zero.*

Now consider the convex optimization problems that satisfy the "technical conditions". Then,

$$\begin{aligned} f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*) = \inf_{\mathbf{x} \in \mathbb{R}^n} \left(f_0(\mathbf{x}) + \textstyle\sum_{i=1}^m \lambda_i^* f_i(\mathbf{x})\right) \\ &\leq f_0(\mathbf{x}^*) + \textstyle\sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}^*) \\ &\leq f_0(\mathbf{x}^*) \end{aligned}$$

Hence, all the inequalities must become equalities. Hence, we have $\lambda_i^* f_i(\mathbf{x}^*) = 0$, $\forall i = 1, 2, ..., m$. This is called the *complementary slackness condition*. Furthermore, if the functions $f_i(\mathbf{x})$ are differentiable, then the gradient of the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}^*)$ must vanish at $\mathbf{x}^*$ since $\mathbf{x}^*$ minimizes $L(\mathbf{x}, \boldsymbol{\lambda}^*)$ over $\mathbf{x}$. All of these results can be summarized in the following theorem [37].

Theorem 2.4.2 *Let $\mathbf{x}^*$ and $\boldsymbol{\lambda}^*$ be any primal and dual optimal points with zero duality gap. Then the following conditions, called KKT (Karush-Kuhn-Tucker) conditions, must be satisfied. The converse is also true if the primal problem is convex.*

(i) $f_i(\mathbf{x}^*) \leq 0, i = 1, 2, ..., m$.

(ii) $\boldsymbol{\lambda}^* \geq 0, i = 1, 2, ..., m$.

(iii) $\lambda_i^* f_i(\mathbf{x}^*) = 0, \forall i = 1, 2, ..., m$ *(Complementary slackness condition).*

(iv) $\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) = 0$.

Sub-gradient method

The KKT conditions are the key to solve for the optimal solutions of both primal and dual problems. However, in most practical cases, we do not have closed-form solutions for the KKT conditions. In these cases, we need some numerical analysis.

For unconstrained optimization to minimize a convex function $f(x)$, we consider the following gradient algorithm

$$\mathbf{x}(t+1) = \mathbf{x}(t) - \gamma \nabla f(\mathbf{x}(t)) \tag{2.52}$$

Notice that if f is differentiable, then $\nabla f(\mathbf{x}^*) = 0$. Hence, if $\mathbf{x}(t) = \mathbf{x}^*$, then $\mathbf{x}(t+1) = \mathbf{x}^*$.

Theorem 2.4.3 *Assume that the function f is convex, continuously differentiable and* ∇f *is Lipschitz with parameter L, i.e., there exists a constant L such that*

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})||_2 \leq L||\mathbf{x} - \mathbf{y}||_2, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \tag{2.53}$$

Assume further that $0 < \gamma < \frac{2}{L}$. *Then the sequence of* $\mathbf{x}(t)$ *generated by* (2.52) *converges and the limit* $\mathbf{x}_0$ *satisfies* $\nabla f(\mathbf{x}_0) = 0$.

Now consider a constrained optimization problem of minimizing $f(\mathbf{x})$ subject to $\mathbf{x} \in X$, where X is a convex set. There are several algorithms to solve this problem. The basic idea is to add a "penalty function" to the objective function and convert this problem to an unconstrained problem. This idea is applied in the penalty-function method and the interior point method. Here, let's consider the third method, i.e., the projection method.

Definition 2.4.1 *The projection of* $\mathbf{x}$ *onto the convex and closed set X, is the point* $\mathbf{z} \in X$ *that is closest to* $\mathbf{x}$*,i.e.,*

$$[\mathbf{x}]^+ = \underset{\mathbf{z}\in X}{argmin}\ (||\mathbf{z} - \mathbf{x}||_2) \tag{2.54}$$

Theorem 2.4.4 *(Projection Theorem) [38]*
(i) $[\mathbf{x}]^+$ *exists and is unique for each* $\mathbf{x} \in \mathbb{R}^n$*.*
(ii) $\mathbf{z} = [\mathbf{x}]^+$ *if and only if*

$$(\mathbf{y} - \mathbf{z})^T(\mathbf{x} - \mathbf{z}) \leq 0, \forall \mathbf{y} \in X \tag{2.55}$$

(iii) The mapping $f(\mathbf{x}) = [\mathbf{x}]^+$ *is continuous and non-expansive,i.e.,*

$$||[\mathbf{x}]^+ - [\mathbf{y}]^+||_2 \leq ||\mathbf{x} - \mathbf{y}||_2, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{2.56}$$

By using the Projection Theorem and the results from the unconstrained optimization algorithms, we can derive the gradient projection algorithm and prove its convergence.

Theorem 2.4.5 *If f is convex, and* ∇f *is Lipschitz with parameter L, then the sequence of points*

$$\mathbf{x}(t+1) = [\mathbf{x}(t) - \gamma\nabla f(\mathbf{x}(t))]^+ \tag{2.57}$$

converges if $0 < \gamma < \frac{2}{L}$*, and the limit* $\mathbf{x}^*$ *minimizes* $f(\mathbf{X})$ *over X.*

In case that ∇f is not Lipschitz, or if f is even not differentiable, we can replace the gradient by the sub-gradient. However, the sub-gradient projection algorithm does not always converge. It has been shown that it will converge if some conditions on f and on the step size γ are satisfied.

Theorem 2.4.6 *Assume that f is convex and its subgradients are bounded. Consider the subgradient descent algorithm*

$$\mathbf{x}(t+1) = [\mathbf{x}(t) - \gamma_t\nabla f(\mathbf{x}(t))]^+ \tag{2.58}$$

where $\nabla f(\mathbf{x}(t))$ *is a subgradient of* f *at* $\mathbf{x}(t)$*. Assume further that*

$$\sum_{t=1}^{\infty} \gamma_t \to \infty, \sum_{t=1}^{\infty} \gamma_t^2 < \infty \tag{2.59}$$

then as $t \to \infty$*,* $\mathbf{x}(t)$ *converges to* $\mathbf{x}^*$ *and* $\nabla f(\mathbf{x}^*) = 0$*.*

In general, the projection algorithm may be difficult to implement if the constraint set is in complex form. However, for the dual problem, the constraint set is always a "quadrant". Hence, the projection becomes very simple. Furthermore, if $\mathbf{x}^*$ is the minimizer of the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda})$ at λ, then $(f_i(\mathbf{x}^*)$ is a subgradient of the dual function $g(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda}$. As a result, the gradient projection algorithm for the dual problem has the following simple form:

$$\lambda_i(t+1) = \left[\lambda_i(t) + \gamma f_i(\mathbf{x}(t))\right]^+ . \tag{2.60}$$

3. JOINT OPTIMIZATION OF POWER ALLOCATION AND RELAY SELECTION STRATEGY IN WIRELESS OFDM NETWORKS WITH PERFECT CSI

As stated above, there has been some recent research focused on optimizing the power allocation at the PHY layer and scheduling at the MAC layer in cooperative networks, but most of this work includes the consideration of these two problems separately rather than in combination as one joint optimization problem. The most difficult obstacle which we have to overcome is the non-convexity of the problem. In this chapter, a so-called dual optimization method for non-convex problems is applied to jointly optimize the power allocation and the scheduling of wireless cooperative OFDM systems. This problem will be mathematically analyzed to find the optimal solution, and the analysis will be confirmed by some simulation results.

3.1 System Model and Problem Formulation

3.1.1 System Model

We consider the uplink of a wireless network with K mobile stations (MS). The network of interest uses an OFDMA signal format with N tones. The base station (BS) transmits a pilot signal with constant power. The MS measure the channel gain based on this signal and report it to the BS. Perfect power control is assumed. Each user has a limitation of average transmission power, denoted by $p_{i,max}$.

We make some assumptions before conducting our analysis. First, we assume that the channel information can be estimated completely by the receivers and then can be sent back to the transmitters without error. Indeed, this requires only a small quantity of additional feedback, that is, the amplitude information on the forward

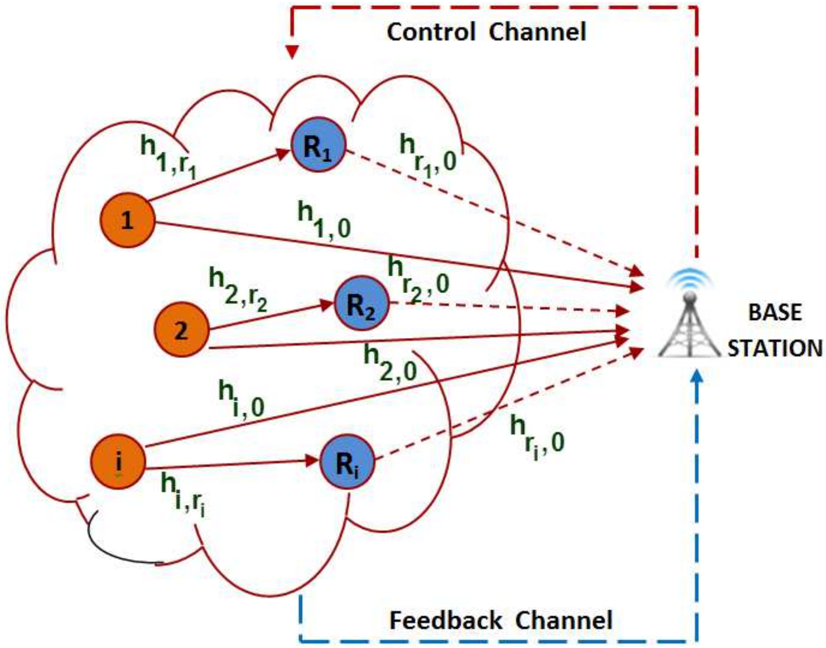

Fig. 3.1. System model

links over the systems requiring coherent combining [4], [5]. Secondly, we assume that the channel conditions are different between users and between each user and the base station, that is, for each transmission from user i to user j (base station is considered as user 0), we have a specific channel gain h_{ij}, and assume that the channel distribution of each channel gain is Rayleigh distributed. We further assume that the number of subcarriers is large enough so that the bandwidth of each subcarrier is sufficiently small. Hence, we can consider the fading on each sub-channel as flat fading.

Scheduling

To avoid the interference between users, the data transmission of each user is scheduled by a mechanism similar to time-division multiplexing, in which each user transmits data in a specific interval that does not overlap with any interval of other users. We denote the time interval for user i over the n^{th} subcarrier by $\tau_i^{(n)}$ and assume a normalization such that

$$\sum_{i=1}^{K} \tau_i^{(n)} \leq 1 \text{ for } n = 1, 2, ..., N \tag{3.1}$$

For a specific subcarrier n, each user is assigned one user from (K-1) remaining users as the relay node. We denote the relay of user i over n^{th} tone by $r_i^{(n)}$.

Each user's interval then contains two equal-length time slots, so there are 2K time slots in total. In the first time slot of interval $\tau_i^{(n)}$, user i transmits its own data to the BS and to its relay $r_i^{(n)}$. In the second time slot, user $r_i^{(n)}$, after receiving data from user i, forwards the data to the BS. The relay-selection process is controlled by the BS by using the CSI from all users via the feedback channel. Note that if $r_i^{(n)} =$ i, user i transmits its data directly to the BS, without using any relay.

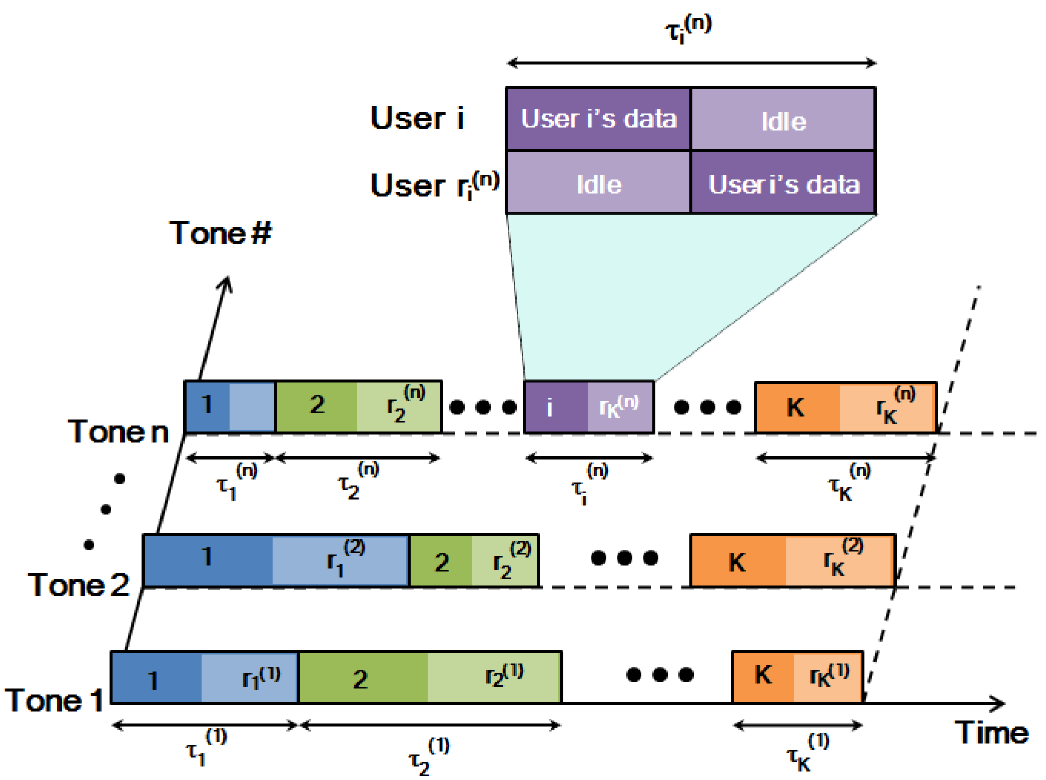

Fig. 3.2. Cooperation protocol (during interval $\tau_i^{(n)}$)

Power Allocation

The system is modeled by the following equations

$$y_0 = \sqrt{\gamma_{i0}} h_{i0} x_i + \sqrt{\gamma_{r_i^{(n)}0}} h_{r_i^{(n)}0} x_{r_i^{(n)}} + n_0 \tag{3.2a}$$

$$y_{r_i^{(n)}} = \sqrt{\gamma_{ir_i^{(n)}}} h_{ir_i^{(n)}} x_i + n_{r_i^{(n)}} \tag{3.2b}$$

where x_i is the signal transmitted from node i, y_i is the signal received at node i, n_i is the zero-mean AWGN at node i, and γ_{ij} is the signal-to-noise ratio at node j corresponding to the signal transmitted from node i, which can be determined by the formula

$$\gamma_{ij} = \frac{P_t G_i}{d_{ij}^{\alpha} N_0 W} \tag{3.3}$$

where P_t is the transmission power from node i, d_{ij} is the distance between node i and node j, α is the path-loss index, N_0 is the noise power density, W is the total bandwidth of the network, and G_i is the antenna gain of user i.

As mentioned above, we assume that each user is allocated a specific time interval on each subchannel to transmit data and denote the transmission power of user i in the time interval $\tau_j^{(n)}$ by $P_{ij}^{(n)}$ (j may, or may not be, equal to i because user i may act as a relay node during the time interval of user j). Here, the superscript (n) indicates the n^{th} subchannel of the system. The average transmission power of the i^{th} user over the unit interval is limited by $p_{i,max}$. The average transmission power of user i is

$$\bar{P}_i = E\left[\sum_{n=1}^{N}\left(\frac{1}{2}\tau_i^{(n)} P_{ii}^{(n)} + \frac{1}{2}\sum_{j:r_j=i}\tau_j^{(n)} P_{ij}^{(n)}\right)\right] \tag{3.4}$$

where the averaging is over the overall channel distributions.

We also assume that the distance between a user and its relay node is much smaller than the distance between that user and its BS. If we use the Decode-and-Forward protocol with code combining for cooperation, then the maximum capacity that user i can get is determined by the following equation [6]

$$C_i = E\left[\frac{1}{2}\sum_{n=1}^{N}\tau_i^{(n)}\log_2\left|1+\rho_{i0}\left|h_{i0}^{(n)}\right|^2 P_{ii}^{(n)}\right| + \frac{1}{2}\sum_{n=1}^{N}\tau_i^{(n)}\log_2\left|1+\rho_{r_i,0}\left|h_{r_i,0}^{(n)}\right|^2 P_{r_i,i}^{(n)}\right|\right] \tag{3.5}$$

where the averaging is over the distribution of channel gains and ρ_{ij} is a constant proportional to γ_{ij}.

3.1.2 Problem Formulation

Our goal is selecting a cooperation strategy (that is, to assign a vector $\mathbf{r} = \left[r_1^{(n)}, r_2^{(n)}, ..., r_K^{(n)}\right]^T$, where $r_i^{(n)} \in \{1, 2, ..., K\}$), and the resource (time and power) allocation, i.e., to find

$$\mathbf{P} = \left[P_{ij}^{(n)} \,|i, j = 1, 2, ..., K; n = 1, ..., N\right]$$
$$\boldsymbol{\tau} = \left[\tau_i^{(n)} \,|i = 1, 2, ..., K; n = 1, ..., N\right]^T$$

so that the weighted sum rate (WSR) (similar to one defined in [39]) is maximized, where

$$WSR = \sum_{i=1}^{K} w_i C_i =$$
$$= \sum_{i=1}^{K} w_i E\left[\frac{1}{2}\sum_{n=1}^{N}\tau_i^{(n)}\log_2\left|1+\rho_{i0}\left|h_{i0}^{(n)}\right|^2 P_{ii}^{(n)}\right| + \frac{1}{2}\sum_{n=1}^{N}\tau_i^{(n)}\log_2\left|1+\rho_{r_i,0}\left|h_{r_i,0}^{(n)}\right|^2 P_{r_i,i}^{(n)}\right|\right] \tag{3.7}$$

(w_i is some weight indicating the importance of user i), subject to the following constraints:

(i) Power constraint: the average consumed power of user i must be less than $p_{i,max}$.

$$E\left[\sum_{n=1}^{N}\left(\frac{1}{2}\tau_i^{(n)}P_{ii}^{(n)} + \frac{1}{2}\sum_{j:r_j=i}\tau_j^{(n)}P_{ij}^{(n)}\right)\right] \le p_{i,\max} \text{ i = 1,2,...,K} \tag{3.8}$$

(ii) Time constraint: $\{\tau_i^{(n)}\}, i = 1, 2, ..., K; n = 1, 2...N$ must satisfy (3.1).

(iii) Rate constraint: the average data rate for user i must be greater than a certain rate R_i to guarantee the quality of service for user i, and hence, guarantee the fairness of the network.

$$E\left[\frac{1}{2}\sum_{n=1}^{N}\tau_i^{(n)}\log_2\left|1+\rho_{i0}\left|h_{i0}^{(n)}\right|^2 P_{ii}^{(n)}\right| + \frac{1}{2}\sum_{n=1}^{N}\tau_i^{(n)}\log_2\left|1+\rho_{r_i,0}\left|h_{r_i,0}^{(n)}\right|^2 P_{r_i,i}^{(n)}\right|\right] \ge R_i \tag{3.9}$$

Now the problem can be stated as follows.

Problem 1:

$$\begin{aligned} \underset{P_{ij},\tau_i,j\in\{0,r_i\},r_i\in\{1,\ldots,K\}}{\text{maximize}} \quad & WSR = \sum_{i=1}^{K} w_i C_i \\ &= \sum_{i=1}^{K} w_i E\left[\frac{1}{2}\sum_{n=1}^{N}\tau_i^{(n)}\log_2\left|1+\rho_{i0}\left|h_{i0}^{(n)}\right|^2 P_{ii}^{(n)}\right|\right] \\ &+ \sum_{i=1}^{K} w_i E\left[\frac{1}{2}\sum_{n=1}^{N}\tau_i\log_2\left|1+\rho_{r_i,0}\left|h_{r_i,0}^{(n)}\right|^2 P_{r_i,i}^{(n)}\right|\right] \end{aligned} \tag{3.10}$$

subject to (3.1), (3.8), (3.9).

3.2 Solution and Algorithm

3.2.1 Scenario 1: two users cooperate with each other

To make our analysis easier to understand, we consider the simplest case with 2 users first, and then generalize the results to the case of K users.

Problem 1a:

$$\underset{\substack{P_{ij}^{(n)},\tau_i^{(n)} \\ i,j=1,2 \\ n=1,2\ldots N}}{\text{maximize}} \quad WSR = \sum_{i=1}^{2} w_i C_i = \sum_{i=1}^{2} w_i E\left[\frac{1}{2}\sum_{n=1}^{N}\tau_i^{(n)}\sum_{j=1}^{2}\log_2\left|1+\rho_{j0}\left|h_{j0}^{(n)}\right|^2 P_{ji}^{(n)}\right|\right] \tag{3.11}$$

subject to (3.1), (3.8), (3.9) with $K = 2$.

Case 1: Known CDI (Channel Distribution Information)

Here, we consider the case in which both the BS and the MS know the channel distribution information (CDI). If the goal is only to optimize the power allocation, the problem is similar to the traditional water-filling problem, and it's convex. However, when combining the power allocation with scheduling, the achievable rate region $\Re(\mathbf{P}, \boldsymbol{\tau})$ is not convex.

Fortunately, it has been shown in [1] that if a non-convex problem satisfies the "time-sharing property," then the duality gap between the primal problem and the

Lagrangian dual problem is zero. Furthermore, it's also emphasized in [1] that in OFDM systems, the time-sharing property is satisfied regardless of the convexity as long as N is sufficiently large and the per-tone objective functions $f_n, ..., f_{n+k}$ and the per-tone constraint functions $h_n, ..., h_{n+k}$ are sufficiently similar for small values of k. This is the case in almost all OFDM systems because sub-channel widths in OFDM systems are chosen so that the channel response is approximately flat within each sub-channel.

Hence, we can use the Lagrangian dual method [37] to solve the original problem. This problem has a special structure so that we can also use the Lagrangian dual decomposition as is shown below.

We let λ_i and ν_i $(i = 1, 2)$ be the Lagrange multipliers associated with the power constraint and the rate constraint, respectively.

Denote:

$$\boldsymbol{\tau} = \left[\tau_i^{(n)} \,|i = 1, 2, n = 1, ..., N\right]^T$$
$$\mathbf{P} = \left[P_{ij}^{(n)} \,|i, j = 1, 2, n = 1, ..., N\right]$$
$$\boldsymbol{\nu} = [\nu_1 \ \nu_2]^T$$
$$\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2]^T$$

Then, the Lagrangian dual function can be written as

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\boldsymbol{\tau}, \mathbf{P}} \left\{ L(\boldsymbol{\tau}, \mathbf{P}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \,|(\boldsymbol{\tau}, \mathbf{P}) \in S \right\} \tag{3.12}$$

where

$$S = \left\{ (\boldsymbol{\tau}, \mathbf{P}) \,\middle|\, (\tau_1^{(n)} + \tau_2^{(n)}) \leq 1, \tau_i^{(n)} \geq 0, P_{ij}^{(n)} \geq 0, \forall i, j = 1, 2, n = 1, ..., N \right\} \tag{3.13}$$

where the set $r_i^{(n)} \in \{1, 2\}$ for i = 1,2. The function $L(\boldsymbol{\tau}, \mathbf{P}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is given by

$$\begin{aligned} L(\boldsymbol{\tau}, \mathbf{P}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = &- \sum_{i=1}^{2} w_i E \left[\frac{1}{2} \sum_{n=1}^{N} \left(\tau_i^{(n)} \sum_{j=1}^{2} \log_2 \left| 1 + \rho_{j0} \left| h_{j0}^{(n)} \right|^2 P_{ji}^{(n)} \right| \right) \right] + \\ &+ \sum_{i=1}^{2} \lambda_i \left[E \left[\frac{1}{2} \sum_{n=1}^{N} \sum_{j=1}^{2} \tau_j^{(n)} P_{ij}^{(n)} \right] - p_{i,\max} \right] + \\ &+ \sum_{i=1}^{2} \nu_i \left\{ R_i - E \left[\frac{1}{2} \sum_{n=1}^{N} \left(\tau_i^{(n)} \sum_{j=1}^{2} \log_2 \left| 1 + \rho_{j0} \left| h_{j0}^{(n)} \right|^2 P_{ji}^{(n)} \right| \right) \right] \right\}, \end{aligned}$$

or alternatively,

$$L(\boldsymbol{\tau},\mathbf{P},\boldsymbol{\lambda},\boldsymbol{\nu}) = -\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{1}{2}\sum_{n=1}^{N} E\left[\tau_i^{(n)} f\left(P_{ji}^{(n)}\right)\right] + \sum_{i=1}^{2}(\nu_i R_i - \lambda_i p_{i,\max}) \tag{3.14}$$

where

$$f(P_{ji}^{(n)}) = (w_i + v_i)\log_2\left|1 + \rho_{j0}\left|h_{j0}^{(n)}\right|^2 P_{ji}^{(n)}\right| - \lambda_j P_{ji}^{(n)} \tag{3.15}$$

The Lagrangian dual problem **(D)** is

$$\begin{aligned} &\text{maximize } g(\boldsymbol{\lambda},\boldsymbol{\nu}) \\ &\text{subject to } \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\nu} \geq \mathbf{0} \end{aligned}$$

where $\mathbf{x} \geq \mathbf{0}$ means that each element of $\mathbf{x}$ is greater than or equal to zero.

If the "time-sharing" property is satisfied, then there is no duality gap between the primal and the dual problems. Hence, the solution to the Lagrangian dual problem **(D)** in (3.16) provides the solution to Problem 1a. Now, we can decompose this problem into three sub-problems.

$$g(\boldsymbol{\lambda},\boldsymbol{\nu}) = \underbrace{\min_{\tau_1^{(n)}+\tau_2^{(n)}\leq 1, \tau_i^{(n)}\geq 0, i=1,2} \left\{ -\sum_{i=1}^{2}\sum_{n=1}^{N}\sum_{j=1}^{2}\frac{1}{2}\underbrace{\max_{P_{ij}^{(n)}\geq 0} E\left[\tau_i^{(n)} f\left(P_{ji}^{(n)}\right)\right]}_{Power\,Allocation\,Subproblem} \right\}}_{Scheduling\,Subproblem} + \sum_{i=1}^{2}(\nu_i R_i - \lambda_i p_{i,\max}) \tag{3.16}$$

We first fix $\boldsymbol{\tau}$ and solve the inner sub-problem for $\mathbf{P}$, and then solve the outer sub-problem for $\boldsymbol{\tau}$.

Inner sub-problem

Theorem 3.2.1 *The optimal solution of the power allocation sub-problem defined in* (3.16) *is*

$$P_{ji}^{(n)*} = \frac{(w_i+\nu_i)}{\lambda_j \ln 2} - \left(\rho_{j0}\left|h_{j0}^{(n)}\right|^2\right)^{-1} \quad (i,j = 1,2) \tag{3.17}$$

Proof. The inner sub-problem can be solved by using the KKT optimality conditions [37]. To see this, first construct a new Lagrangian function by associating the constraint $P_{ij}^{(n)} \geq 0$ with a Lagrange multiplier υ_{ij}. Write down the KKT conditions for the primal-dual optimal solution P^*, υ^* :

$$\text{i/ } P_{ij}^{(n)*} \geq 0 \text{ (i = 1,2, j = 0,1,2,} j \neq i) \tag{3.18}$$

$$\text{ii/ } \boldsymbol{\upsilon}^* \geq \mathbf{0} \tag{3.19}$$

$$\text{iii/ } \upsilon_{ij}^* P_{ij}^{(n)*} = 0, i, j = 1, 2 \tag{3.20}$$

$$\text{iv/ } \mathbf{P}^* \text{ minimizes the Lagrangian function } L(\mathbf{P}, \boldsymbol{\upsilon}^*) \tag{3.21}$$

The condition (3.21) means that the derivatives of $L(\mathbf{P}, \boldsymbol{\upsilon}^*)$ with respect to $P_{ij}^{(n)}$ must be equal to zero. From (3.20) and by the complementary slackness condition [37], we have $\upsilon_{ij}^* = 0$ because $P_{ij}^{(n)} = 0$ is not optimal. Now solving equations with derivatives of the above Lagrangian function with respect to $P_{ij}^{(n)}$ set to zero, we have

$$\frac{\partial}{\partial P_{ji}^{(n)}} \left[(w_i + v_i) \log_2 \left| 1 + \rho_{j0} \left| h_{j0}^{(n)} \right|^2 P_{ji}^{(n)} \right| - \lambda_j P_{ji}^{(n)}) \right] = 0$$

$$\Rightarrow (w_i + v_i) \left(\left| 1 + \rho_{j0} \left| h_{j0}^{(n)} \right|^2 P_{j0}^{(n)} \right| \ln 2 \right)^{-1} \rho_{j0} \left| h_{j0}^{(n)} \right|^2 - \lambda_j = 0$$

$$\Rightarrow P_{ji}^{(n)*} = \frac{(w_i + v_i)}{\lambda_j \ln 2} - \left(\rho_{j0} \left| h_{j0}^{(n)} \right|^2 \right)^{-1} \tag{3.22}$$

■

Outer sub-problem

After solving the inner sub-problem, the scheduling becomes simple.

Theorem 3.2.2 *The optimal solution to the Scheduling Sub-problem which is defined in* (3.16) *is*

$$\tau_i^{(n)*} = \begin{cases} 1, & \text{if } i = \arg\max\limits_k \left(\sum\limits_{j=1}^{2} E\left[f\left(P_{jk}^{(n)*} \right) \right] \right) \\ 0, & \text{otherwise} \end{cases} \quad \text{for } i = 1, 2 \tag{3.23}$$

Proof. We define

$$A^{(n)} = E\left[f(P_{11}^{(n)*}) + f(P_{21}^{(n)*})\right]$$
$$B^{(n)} = E\left[f(P_{12}^{(n)*}) + f(P_{22}^{(n)*})\right]$$

The outer sub-problem becomes a linear optimization problem.

$$\begin{aligned} \text{maximize} \quad & \sum_{n=1}^{N} (A^{(n)}\tau_1^{(n)} + B^{(n)}\tau_2^{(n)}) \\ \text{subject to} \quad & \sum_{n=1}^{N} \tau_1^{(n)} + \tau_2^{(n)} \leq 1, \tau_1^{(n)}, \tau_2^{(n)} \geq 0 \end{aligned}$$

We can write

$$\begin{aligned} \sum_{n=1}^{N} (A^{(n)}\tau_1^{(n)} + B^{(n)}\tau_2^{(n)}) &= \sum_{n=1}^{N} B^{(n)}\left(\tau_1^{(n)} + \tau_2^{(n)}\right) + \left(A^{(n)} - B^{(n)}\right)\tau_1^{(n)} \leq \\ &\leq \sum_{n=1}^{N} A^{(n)} + \sum_{n=1}^{N} \left(A^{(n)} - B^{(n)}\right)\tau_1^{(n)} \end{aligned} \tag{3.24}$$

If $A^{(n)} \geq B^{(n)}$, then the n^{th} term of (3.24) is maximized when $\tau_1^{(n)} = 1$. Otherwise, it's maximized when $\tau_1^{(n)} = 0$. Hence, $\tau_1^{(n)*} = 1$ if $A^{(n)} \geq B^{(n)}$, and $\tau_1^{(n)*} = 0$, otherwise. ■

Now, it remains to determine the optimal solutions $\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*$. When this is accomplished, $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ can be replaced by $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ in (3.17) and (3.23). This then yields the optimal solution for the original problem described by (3.11).

From (3.23), we see that the dual objective function is not differentiable, so we have to use a subgradient projection method to compute the optimal dual solution. The subgradient projection method is an algorithm to compute the sequence of points $\{x(t)\}_{t=0}^{\infty}$, which converges to the optimal solution, according to the following formula

$$x(t+1) = [x(t) - \gamma(t)g_f(x(t))]^+ \tag{3.25}$$

where $g_f(x)$ is the subgradient of f evaluated at x, $\{\gamma(t)\}_{t=0}^{\infty}$ is a step-size sequence, and $[\cdot]^+$ denotes the projection onto the feasible set. In [38], it's shown that if the sequence $\{\gamma(t)\}_{t=0}^{\infty}$ satisfies the following conditions

$$\gamma(t) \to \infty, \sum_{t=0}^{\infty} \gamma(t) \to \infty, \text{ and } \sum_{t=0}^{\infty} [\gamma(t)]^2 < \infty \tag{3.26}$$

then the subgradient algorithm (3.25) converges with probability 1 to the optimal solution of the convex optimization problem with objective function f.

Back to our dual problem, the subgradients of the dual function $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ with respect to λ_i and ν_i are, respectively, determined by

$$g_f(\lambda_i(t)) = E\left[\frac{1}{2}\sum_{n=1}^{N}\sum_{j=1}^{2}\tau_j^{(n)*}(t)P_{ij}^{(n)*}(t)\right] - p_{i,\max} \tag{3.27}$$

and

$$g_f\left(\nu_i(t)\right) = R_i - E\left[\frac{1}{2}\sum_{n=1}^{N}\left(\tau_i^{(n)*}(t)\sum_{j=1}^{2}\log_2\left|1+\rho_{j0}\left|h_{j0}^{(n)}\right|^2 P_{ji}^{(n)*}(t)\right|\right)\right] \tag{3.28}$$

In the equations (3.27) and (3.28), $\tau_i^{(n)*}(t)$ and $P_{ji}^{(n)*}(t)$ are computed by (3.23) and (3.17), respectively, with $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is replaced by $(\boldsymbol{\lambda}(t), \boldsymbol{\nu}(t))$.

Now, an algorithm can be designed for joint optimization of power allocation and scheduling when the CDI is known.

Algorithm 1

1. Initial condition: Choose $\boldsymbol{\lambda}(0)$ *and* $\boldsymbol{\nu}(0)$*, then compute* $P_{ij}^{(n)*}(0)$ *by* (3.17) *and* $\tau_i^{(n)*}(0)$ *by* (3.23) *for i,j =1,2, n=1,2,...,N.*

2. Iteration: At step t + 1, choose step size $\gamma(t) = \gamma/t$*, where* γ *is some positive constant. Compute* $g_f(\lambda_i(t))$ *and* $g_f(\nu_i(t))$ *for i = 1,2 using* (3.27) *and* (3.28)*. Update* $\lambda_i(t+1)$ *and* $\nu_i(t+1)$*, i = 1, 2 using the following formulas:*

$$\lambda_i(t+1) = [\lambda_i(t) + \gamma(t)g_f(\lambda_i(t))]^+ = \max\{\lambda_i(t) + \gamma(t)g_f(\lambda_i(t)), 0\} \tag{3.29}$$

$$\nu_i(t+1) = [v_i(t) + \gamma(t)g_f(\nu_i(t))]^+ = \max\{v_i(t) + \gamma(t)g_f(\nu_i(t)), 0\} \tag{3.30}$$

3. Termination: Let ε *be the maximum error allowed, then the algorithm stops whenever* $\|\boldsymbol{\lambda}(t+1) - \boldsymbol{\lambda}(t)\| < \varepsilon$ *and* $\|\boldsymbol{\nu}(t+1) - \boldsymbol{\nu}(t)\| < \varepsilon$ *for i = 1, 2. Otherwise, increase t by 1 and repeat step 2.*

Case 2: Unknown CDI but perfect CSI (Channel State Information) feedback

In this section, we assume that the channel distribution is unknown and the mobile users must estimate the channel state information based on the sample symbols sent from the BS. This estimated information is then fed back to the BS. We assume that the feedback channel is error free.

Without knowledge of the full distribution of the channel gain matrices, we can no longer use (3.27) and (3.28), since we cannot average over the total distribution of channel gains. Instead, we use an adaptive stochastic approximation algorithm [40], that is, we use the current values of $P_{ij}^{(n)*}$ and $\tau_i^{(n)*}$ to compute the subgradients, rather than using the expectation of all these values:

$$\hat{g}_f(\hat{\lambda}_i(t)) = \left[\sum_{n=1}^{N}\sum_{j=1}^{2}\frac{1}{2}\tau_j^{(n)*}(t)P_{ij}^{(n)*}(t)\right] - p_{i,\max} \tag{3.31}$$

and

$$\hat{g}_f\left(\hat{\nu}_i(t)\right) = R_i - \frac{1}{2}\sum_{n=1}^{N}\left(\tau_i^{(n)*}(t)\sum_{j=1}^{2}\log_2\left|1+\rho_{j0}\left|h_{j0}^{(n)}\right|^2 P_{ji}^{(n)*}(t)\right|\right) \tag{3.32}$$

Now, we can state the Algorithm 2:

Algorithm 2

1. Initial condition: Choose $\hat{\boldsymbol{\lambda}}(0)$ *and* $\hat{\boldsymbol{\nu}}(0)$*, then compute* $P_{ij}^{(n)*}(0)$ *by* (3.17) *and* $\tau_i^{(n)*}(0)$ *for* $i,j=1,2,n=1,2,...,N$ *by* (3.23).

2. Iteration: At step $t+1$*, choose step size* $\gamma(t)=\gamma/t$*, where* γ *is some positive constant. Compute* $g_f\left(\hat{\lambda}_i(t)\right)$ *and* $g_f\left(\hat{\nu}_i(t)\right)$ *using* (3.31), (3.32)*. Update* $\hat{\lambda}_i(t+1)$ *and* $\hat{\nu}_i(t), i=1,2$ *using*

$$\hat{\lambda}_i(t+1) = \left[\hat{\lambda}_i(t)-\gamma(t)g_f(\hat{\lambda}_i(t))\right]^+ = \min\left\{\hat{\lambda}_i(t)-\gamma(t)g_f(\hat{\lambda}_i(t)),0\right\} \tag{3.33}$$

$$\hat{\nu}_i(t+1) = \left[\hat{\nu}_i(t)-\gamma(t)g_f(\hat{\nu}_i(t))\right]^+ = \min\left\{\hat{\nu}_i(t)-\gamma(t)g_f(\hat{\nu}_i(t)),0\right\} \tag{3.34}$$

3. Termination: Let ε *be the maximum error allowed, then the algorithm stops whenever* $\left\|\hat{\boldsymbol{\lambda}}(t+1)-\hat{\boldsymbol{\lambda}}(t)\right\| < \varepsilon$ *and* $\left\|\hat{\boldsymbol{\nu}}(t+1)-\hat{\boldsymbol{\nu}}(t)\right\| < \varepsilon$ *for i = 1, 2. Otherwise, increase* j *by 1 and repeat step 2.*

3.2.2 Scenario 2: K users, each user has exactly one relay

The preceding development can be easily generalized for the case of K mobile users in the system, as long as each user has one and only one specific relay. In other words, if we have a fixed relay assignment, then we can apply the algorithms 1 and 2 to optimize the WSR of the system. The new question in this case is, "Which relay assignment strategies give the largest WSR?". This general problem involves a search over all possible relay assignment strategies to find the best strategy, which results in a mixed integer programming problem [38].

For simplicity, we consider the case in which the channel distribution information is known. Roughly speaking, if the CDI is unknown, but the CSI can be estimated and feedback perfectly, the algorithm also converges in probability to the optimal solution, as shown in the convergence analysis later.

Utilizing the work of W. Yu and R. Liu [1] again, we can equivalently solve the dual problem when the number of OFDM tones goes to infinity. The Lagrangian function in this case is

$$\begin{aligned} L(\boldsymbol{\tau},\mathbf{P},\boldsymbol{\lambda},\boldsymbol{\nu}) = &-\sum_{i=1}^{K} w_i E\left[\frac{1}{2}\sum_{n=1}^{N}\left(\tau_i^{(n)}\sum_{j\in\{1,r_i\}}\log_2\left|1+\rho_{j0}\left|h_{j0}^{(n)}\right|^2 P_{ji}^{(n)}\right|\right)\right]+ \\ &+\sum_{i=1}^{K}\lambda_i\left\{E\left[\sum_{n=1}^{N}\left(\frac{1}{2}\tau_i^{(n)}P_{ii}^{(n)}+\frac{1}{2}\sum_{j:r_j=i}\tau_j^{(n)}P_{ij}^{(n)}\right)\right]-p_{i,\max}\right\}+ \\ &+\sum_{i=1}^{K}\nu_i\left\{R_i - E\left[\frac{1}{2}\sum_{n=1}^{N}\left(\tau_i^{(n)}\sum_{j\in\{1,r_i\}}\log_2\left|1+\rho_{j0}\left|h_{j0}^{(n)}\right|^2 P_{ji}^{(n)}\right|\right)\right]\right\} \end{aligned}$$

$$g(\boldsymbol{\lambda},\boldsymbol{\nu}) = \min_{\boldsymbol{\tau},\mathbf{P},\mathbf{r}}\left\{L(\boldsymbol{\tau},\mathbf{P},\boldsymbol{\lambda},\boldsymbol{\nu})\,|(\boldsymbol{\tau},\mathbf{P},\mathbf{r})\in S\right\}$$

where

$$S = \left\{(\boldsymbol{\tau},\mathbf{P},\mathbf{r})\left|\begin{array}{l}\sum_{i=1}^{M}\tau_i^{(n)}\le 1\\ \tau_i^{(n)}\ge 0\\ P_{ij}^{(n)}\ge 0\\ r_i\in\{1,2...,K\}\end{array}\right.,\forall i,j=1,2,...,K;n=1,...,N\right\}$$

The dual problem is to maximize $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ subject to $\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\nu} \geq \mathbf{0}$. We're going to find the minimum point of $L(\boldsymbol{\tau}, \mathbf{P}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ over the feasible set S. However, it's equivalent to finding N optimal solutions of N per-tone problems.

For the relay selection strategy, we notice that each user can be a relay for more than one user, but these operations must be in different tones. This means, in the per-tone maximization problem, the set $S_i = \{j \in \{1, 2, ..., K\} \, | r_j = i\}$ has at most one element. It implies that the set $\left\{r_1^{(n)}, r_2^{(n)}, ..., r_K^{(n)}\right\}$ is a permutation of the set $\{1, 2, ..., K\}$. Hence, the Lagrangian function can be rewritten in the following form as

$$L(\boldsymbol{\tau}, \mathbf{P}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = -\sum_{n=1}^{N} \sum_{i=1}^{K} \frac{1}{2} E\left[\tau_i^{(n)} f_i(P_{ii}^{(n)}, P_{r_i i}^{(n)})\right] + \sum_{i=1}^{K} (\nu_i R_i - \lambda_i p_{i,\max}) \tag{3.35}$$

where

$$\begin{aligned} f_i(P_{ii}^{(n)}, P_{r_i i}^{(n)}) = (w_i + \nu_i)\log_2 &\left(\left|1 + \rho_{i0}\left|h_{i0}^{(n)}\right|^2 P_{ii}^{(n)}\right| . \left|1 + \rho_{r_i 0}\left|h_{r_i 0}^{(n)}\right|^2 P_{r_i i}^{(n)}\right|\right) \\ &- \lambda_i P_{ii}^{(n)} - \lambda_{r_i} P_{r_i i}^{(n)} \end{aligned} \tag{3.36}$$

Example: To illustrate the above idea, consider the case in which K = 4 and the relay selection strategy is $(r_1, r_2, r_3, r_4) = (4, 2, 1, 3)$. Consider a particular tone n (we omit the superscript n for simplicity). The Lagrangian function term corresponding to the n^{th} tone (excluding the constant term $(\sum_{i=1}^{K} (\nu_i R_i - \lambda_i p_{i,\max}))$ is given by

$$\begin{aligned} &\frac{1}{2}\left[\tau_1(w_1 + \nu_1)\log_2\left(\left|1 + \rho_{10}|h_{10}|^2 P_{11}\right| . \left|1 + \rho_{40}|h_{40}|^2 P_{41}\right|\right) - \lambda_1(\tau_1 P_{11} + \tau_3 P_{13})\right] \\ &+ \frac{1}{2}\left[\tau_2(w_2 + \nu_2)\log_2\left(\left|1 + \rho_{20}|h_{20}|^2 P_{22}\right| . \left|1 + \rho_{20}|h_{22}|^2 P_{22}\right|\right) - \lambda_2(\tau_2 P_{22} + \tau_2 P_{22})\right] \\ &+ \frac{1}{2}\left[\tau_3(w_3 + \nu_3)\log_2\left(\left|1 + \rho_{30}|h_{30}|^2 P_{33}\right| . \left|1 + \rho_{10}|h_{10}|^2 P_{13}\right|\right) - \lambda_3(\tau_3 P_{33} + \tau_4 P_{34})\right] \\ &+ \frac{1}{2}\left[\tau_4(w_4 + \nu_4)\log_2\left(\left|1 + \rho_{40}|h_{40}|^2 P_{44}\right| . \left|1 + \rho_{30}|h_{30}|^2 P_{34}\right|\right) - \lambda_4(\tau_4 P_{44} + \tau_1 P_{41})\right] \\ &= \frac{1}{2}\sum_{i=1}^{4} \tau_i f_i(P_{ii}, P_{r_i i}) \end{aligned}$$

Hence, we can decompose the dual function $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ into two sub-problems in a similar way to Problem 1a to obtain

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = -\left[\underbrace{\sum_{n=1}^{N} \max_{\mathbf{r}^{(n)}} \underbrace{\left\{ \max_{\substack{\tau_i^{(n)} \geq 0 \\ \sum_{i=1}^{K} \tau_i^{(n)} \leq 1}} \left\{ \sum_{i=1}^{K} \frac{1}{2} \tau_i^{(n)} \underbrace{\max_{P_{ii}^{(n)}, P_{r_i i}^{(n)} \geq 0} E\left[f_i(P_{ii}^{(n)}, P_{r_i i}^{(n)})\right]}_{Power\ Allocation\ Subproblem} \right\} \right\}}_{Scheduling\ Subproblem}}_{Relay\ Selection\ Strategy\ Subproblem}\right] + \sum_{i=1}^{K} (\nu_i R_i - \lambda_i p_{i,\max}) \tag{3.37}$$

The optimal solutions for the power allocation sub-problem and the scheduling sub-problem can be derived by the same method as Problem 1a. The results are stated in the following theorem.

Theorem 3.2.3 *The optimal solutions to the Power Allocation Sub-problem and the Scheduling Sub-problem which are defined in* (3.37) *are*

$$P_{ji}^{(n)*} = \frac{(w_i + \nu_i)}{\lambda_j \ln 2} - \left(\rho_{j0} \left| h_{j0}^{(n)} \right|^2\right)^{-1} (i = 1, 2, j = 1, r_i) \tag{3.38}$$

and

$$\tau_i^{(n)*} = \begin{cases} 1, & if\ i = \arg\max_k \left(E\left[f_k \left(P_{kk}^{(n)*}, P_{r_k k}^{(n)*} \right) \right] \right) \\ 0, & otherwise \end{cases} \quad for\ i = 1, 2, ..., K \tag{3.39}$$

respectively.

Proof. The power allocation sub-problem can be solved by using the KKT conditions. Here, the proof is similar to the one that is used for Theorem 3.2.1.

For the scheduling sub-problem, we define

$$A_k^{(n)} = E\left[f_k(P_{kk}^{(n)*}, P_{r_k k}^{(n)*})\right]$$

The scheduling subproblem becomes a linear optimization problem.

$$\begin{aligned} \text{maximize} \quad & \sum_{n=1}^{N}\sum_{k=1}^{K} A_k^{(n)}\tau_k^{(n)} \\ \text{subject to} \quad & \sum_{k=1}^{K}\tau_k^{(n)} \leq 1, \forall n \quad \textit{and} \quad \tau_k^{(n)} \geq 0, \forall n, \forall k \end{aligned}$$

Let $k^* = \arg\max_k \left(E\left[f_k\left(P_{kk}^{(n)*}, P_{r_k k}^{(n)*} \right)\right]\right)$ then we can write

$$\sum_{n=1}^{N}\sum_{k=1}^{K} A_k^{(n)}\tau_k^{(n)} \leq \sum_{n=1}^{N} A_{k^*}^{(n)} \sum_{k=1}^{K}\tau_k^{(n)} \leq \sum_{n=1}^{N} A_{k*}^{(n)} \tag{3.40}$$

The first inequality in (3.40) becomes an equality if and only if $A_k^{(n)} = A_{k^*}^{(n)}$ or $\tau_k^{(n)} = 0$. That means $\tau_k^{(n)} = 0, \quad \forall k \neq k^*$. In addition, the second inequality in (3.40) becomes an equality if and only if $\sum_{k=1}^{K}\tau_k^{(n)} = 1$. Hence $\tau_i^{(n)*} = 1$ if $i = k^*$ and $\tau_i^{(n)*} = 0$ otherwise. ■

Finally, the best relay selection strategy is chosen by an exhaustive search over all possible values of $\left\{r_1^{(n)}, r_2^{(n)}, ..., r_K^{(n)}\right\}$(K! possibilities).

The original problem can be solved after the optimal dual solution $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is determined. The solution can be obtained by the subgradient projection method. The subgradient of the dual function $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ with respect to λ_i and ν_i at the t^{th} iteration are respectively determined by the formulas given by

$$g_f(\lambda_i(t)) = E\left[\sum_{n=1}^{N}\left(\frac{1}{2}\tau_i^{(n)}P_{ii}^{(n)} + \frac{1}{2}\tau_j^{(n)}P_{ij}^{(n)}\right)\right] - p_{i,\max}, \ \textit{where}\ r_j = i \tag{3.41}$$

and

$$g_f\left(\nu_i(t)\right) = R_i - E\left[\frac{1}{2}\sum_{n=1}^{N}\left(\tau_i^{(n)}\sum_{j\in\{i,r_i\}}\log_2\left|1 + \rho_{j0}\left|h_{j0}^{(n)}\right|^2 P_{ji}^{(n)}\right|\right)\right] \tag{3.42}$$

In the equations (3.41) and (3.42), $\tau_i^{(n)*}(t)$ and $P_{ji}^{(n)*}(t)$ are computed by (3.39) and (3.38), respectively, with $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is replaced by $(\boldsymbol{\lambda}(t), \boldsymbol{\nu}(t))$.

Now, we summarize the algorithm for joint optimization of power allocation and cooperation strategy in the condition of known CDI.

Algorithm 3

1. Choose an initial relay selection strategy for each tone of the OFDMA system.

2. Solve each per-tone optimization problem:

a) Initial condition: Choose $\boldsymbol{\lambda}(0)$ *and* $\boldsymbol{\nu}(0)$*, then compute* $P_{ij}^{(n)*}(0)$ *by* (3.38) *and* $\tau_i^{(n)*}(0)$ *by* (3.39) *for i,j =1,2,...,K; n=1,2,...,N.*

b) Iteration: At step t + 1, choose step size $\gamma(t) = \gamma/t$ *, where* γ *is some positive constant. Compute* $g_f\left(\lambda_i(t)\right)$ *and* $g_f\left(\nu_i(t)\right)$ *for* $i = 1, 2, ..., K$ *using* (3.41) *and* (3.42)*. Update* $\lambda_i(t+1)$ *and* $\nu_i(t+1)$ *,* $i = 1, 2, ..., K$ *using the following formulas*

$$\lambda_i(t+1) = [\lambda_i(t) + \gamma(t)g_f(\lambda_i(t))]^+ = \max\left\{\lambda_i(t) + \gamma(t)g_f(\lambda_i(t)), 0\right\} \tag{3.43}$$

$$v_i(t+1) = [v_i(t) + \gamma(t)g_f(v_i(t))]^+ = \max\left\{v_i(t) + \gamma(t)g_f(v_i(t)), 0\right\} \tag{3.44}$$

c) Termination: Let ε *be the maximum error allowed. Stop the algorithm whenever* $\|\boldsymbol{\lambda}(t+1) - \boldsymbol{\lambda}(t)\| < \varepsilon$ *and* $\|\boldsymbol{\nu}(t+1) - \boldsymbol{\nu}(t)\| < \varepsilon, i = 1, 2, ..., K$*. Otherwise, increase t by 1 and repeat Step 2b.*

d) Store the optimal solution, and then change to another relay selection strategy. Repeat steps 2a through 2c. If the optimal value is better than previous value, store the new optimal solution. Otherwise, discard it. Repeat Step 2d until there is no remaining strategy to consider.

3. Repeat Step 2 for other tones until all tones have been processed.

The flow chart of Algorithm 3 is illustrated in Fig. 3.3.

If the CDI is unknown, but the CSI is fed back without error, we can modify the Algorithm 3 to a more practical Algorithm 4.

Algorithm 4

Same as Algorithm 3, except at Step 2b:

2. b) Iteration: At step t + 1, choose step size $\gamma(t) = \gamma/t$ *, where* γ *is some positive constant. Compute* $\hat{g}_f\left(\hat{\lambda}_i(t)\right)$ *and* $\hat{g}_f\left(\hat{\nu}_i(t)\right)$ *for* $i = 1, 2, ..., K$ *using* (3.41) *and* (3.42)*. Update* $\hat{\lambda}_i(t+1)$ *and* $\hat{\nu}_i(t+1)$ *,* $i = 1, 2, ..., K$ *using the following formulas*

$$\begin{aligned} \hat{\lambda}_i(t+1) &= \left[\hat{\lambda}_i(t) + \gamma(t)g_f(\hat{\lambda}_i(t))\right]^+ = \min\left\{\hat{\lambda}_i(t) + \gamma(t)g_f(\hat{\lambda}_i(t)), 0\right\} \\ \hat{\nu}_i(t+1) &= [\hat{\nu}_i(t) + \gamma(t)g_f(\hat{\nu}_i(t))]^+ = \min\left\{\hat{\nu}_i(t) + \gamma(t)g_f(\hat{\nu}_i(t)), 0\right\} \end{aligned} \tag{3.45}$$

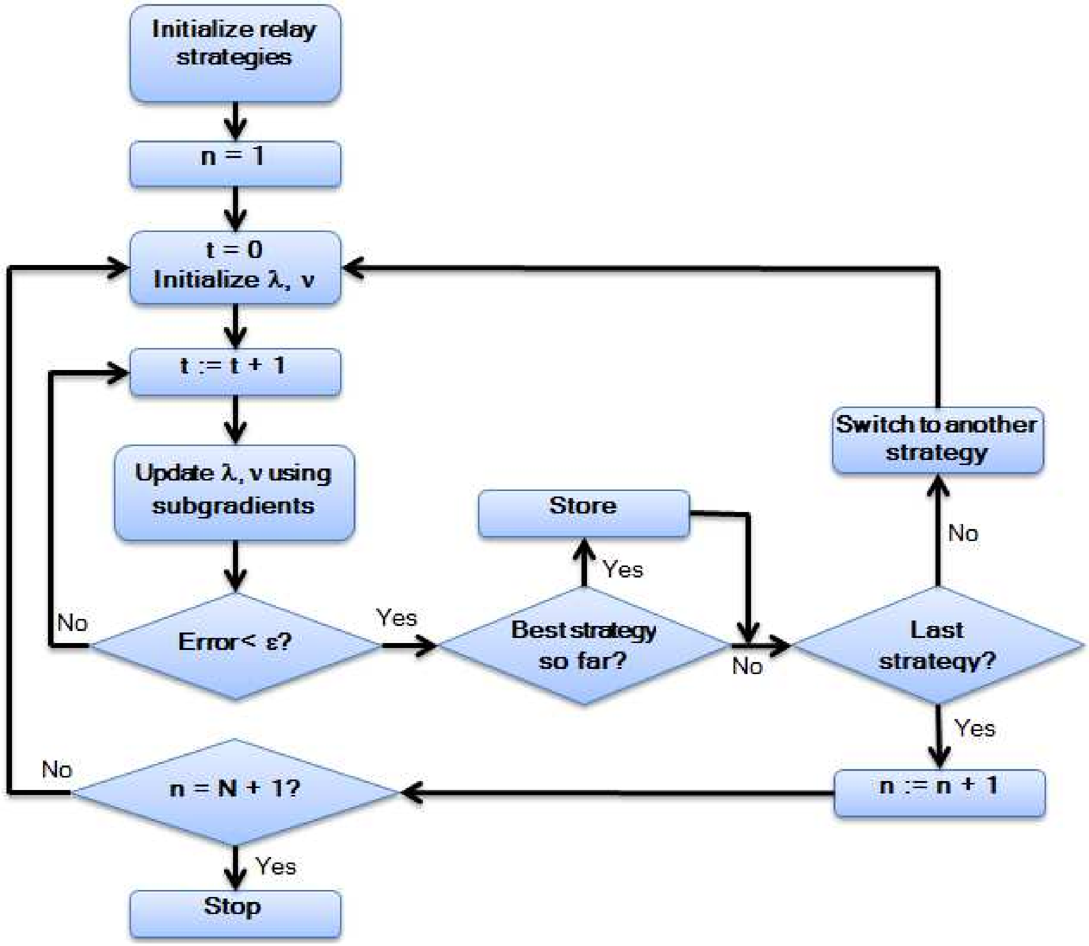

Fig. 3.3. Flow chart of Algorithm 3

3.3 Convergence Analysis

In this section, we prove the convergence of Algorithm 3 and 4 by mathematical analysis. The proof follows similar ones in [40]. Later on, we also illustrate the convergence of these 2 algorithms by simulation results. It's sufficient to show the convergence of Algorithm 3, because when we achieve that, the convergence of Algorithm 4 will follow immediately.

Theorem 3.3.1 *If the step size* $\gamma(t)$ *satisfies the conditions* (3.26)

$$\gamma(t) \to \infty, \sum_{t=0}^{\infty} \gamma(t) \to \infty, \textit{ and } \sum_{t=0}^{\infty} [\gamma(t)]^2 < \infty$$

then Algorithm 3 converges to the optimal solution $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ *with probability 1.*

Proof. Define $B_\varepsilon = \left\{ \left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\nu}}\right) : \left\|\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*\right\|^2 + \|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*\|^2 < \varepsilon \right\}$. We need to show that for any $\varepsilon > 0$, $\left(\hat{\boldsymbol{\lambda}}(t), \hat{\boldsymbol{\nu}}(t)\right) \in B_\varepsilon$ with probability 1 for t large enough. The proof is divided into 2 steps. The first step shows that for any $\varepsilon > 0$, $\left(\hat{\boldsymbol{\lambda}}(t), \hat{\boldsymbol{\nu}}(t)\right) \in B_\varepsilon$ for infinitely large j with probability 1. Then, using this result, we show that for any $\varepsilon > 0$, there exists t_0 such that $\left(\hat{\boldsymbol{\lambda}}(t), \hat{\boldsymbol{\nu}}(t)\right) \in B_\varepsilon$ for any $t > t_0$ with probability 1 in Step 2.

Step 1: We can rewrite the recursive equations for $\hat{\lambda}_i$ and $\hat{\nu}_i$ as

$$\hat{\lambda}_i(t+1) = \hat{\lambda}_i(t) - \gamma(t)[\hat{g}_f(\hat{\lambda}_i(t)) + \delta_{\lambda_i}(t)] + \zeta_{\lambda_i}(t) \tag{3.46}$$

$$\hat{\nu}_i(t+1) = \hat{\nu}_i(t) - \gamma(t)[\hat{g}_f(\hat{\nu}_i(t)) + \delta_{\nu_i}(t)] + \zeta_{\nu_i}(t) \tag{3.47}$$

where

$$\delta_{\lambda_i}(t) = \hat{g}_f(\hat{\lambda}_i(t)) - E[\hat{g}_f(\hat{\lambda}_i(t))] \tag{3.48}$$

$$\delta_{\nu_i}(j) = \hat{g}_f(\hat{\nu}_i(t)) - E[\hat{g}_f(\hat{\nu}_i(t))] \tag{3.49}$$

and $\zeta_{\lambda_i}(t)$ and $\zeta_{\nu_i}(t)$ are some correction terms. Then, we have

$$\begin{aligned} \left\|\hat{\boldsymbol{\lambda}}(t+1) - \boldsymbol{\lambda}^*\right\|^2 &\le \left\|\hat{\boldsymbol{\lambda}}(t) - \boldsymbol{\lambda}^*\right\|^2 \\ &- 2\gamma(t)\left(\hat{\boldsymbol{\lambda}}(t) - \boldsymbol{\lambda}^*\right)^T \left[\hat{g}_f(\hat{\boldsymbol{\lambda}}(t)) + \boldsymbol{\delta}_\lambda(t)\right] + [\gamma(t)]^2 \left\|\hat{g}_f(\hat{\boldsymbol{\lambda}}(t)) + \boldsymbol{\delta}_\lambda(t)\right\|^2 \end{aligned} \tag{3.50}$$

where $\boldsymbol{\delta}_\lambda(t) = [\delta_{\lambda_1}, \delta_{\lambda_2}, ..., \delta_{\lambda_K}]^T$.

It can be seen that $\left\|\hat{\boldsymbol{\lambda}}(t) - \boldsymbol{\lambda}^*\right\|$, $\left\|\hat{g}_f(\hat{\boldsymbol{\lambda}}(t))\right\|$, and $\|\boldsymbol{\delta}_\lambda(t)\|$ are upper bounded, and $E[\boldsymbol{\delta}_\lambda(t)] = \mathbf{0}$. Hence, (3.50) implies that

$$E\left[\left\|\hat{\boldsymbol{\lambda}}(t+1) - \boldsymbol{\lambda}^*\right\|^2\right] \le \left\|\hat{\boldsymbol{\lambda}}(t) - \boldsymbol{\lambda}^*\right\|^2 - 2\gamma(t)\left(\hat{\boldsymbol{\lambda}}(t) - \boldsymbol{\lambda}^*\right)^T \hat{g}_f(\hat{\boldsymbol{\lambda}}(t)) + o\left([\gamma(t)]^2\right) \tag{3.51}$$

Similarly,

$$E\left[\|\hat{\boldsymbol{\nu}}(t+1) - \boldsymbol{\nu}^*\|^2\right] \le \|\hat{\boldsymbol{\nu}}(t) - \boldsymbol{\nu}^*\|^2 - 2\gamma(t)(\hat{\boldsymbol{\nu}}(t) - \boldsymbol{\nu}^*)^T \hat{g}_f(\hat{\boldsymbol{\nu}}(t)) + o\left([\gamma(t)]^2\right) \tag{3.52}$$

Adding (3.51) and (3.52) together, we obtain

$$E\left[\left\|\hat{\boldsymbol{\lambda}}(t+1)-\boldsymbol{\lambda}^*\right\|^2\right]+E\left[\|\hat{\boldsymbol{\nu}}(t+1)-\boldsymbol{\nu}^*\|^2\right]\le\left\|\hat{\boldsymbol{\lambda}}(t)-\boldsymbol{\lambda}^*\right\|^2+\|\hat{\boldsymbol{\nu}}(t)-\boldsymbol{\nu}^*\|^2$$
$$-2\gamma(t)\left[\left(\hat{\boldsymbol{\lambda}}(t)-\boldsymbol{\lambda}^*\right)^T\hat{g}_f(\hat{\boldsymbol{\lambda}}(t))+(\hat{\boldsymbol{\nu}}(t)-\boldsymbol{\nu}^*)^T\hat{g}_f(\hat{\boldsymbol{\nu}}(t))\right]+o\left([\gamma(t)]^2\right) \tag{3.53}$$

Now, we finish Step 1 by using the following theorem, which was shown in [41].

Theorem 3.3.2 *(Supermartingale Convergence Theorem)*
Let $\{X_n\}$ be an $\mathbb{R}^r$-valued stochastic process, and $V(\cdot)$ be a real-valued non-negative function in $\mathbb{R}^r$. Suppose that $\{Y_n\}$ is a sequence of random variables satisfying that $E_n|Y_n|<\infty$ with probability 1. Let $\{\mathfrak{F}_n\}$ be a sequence of σ-algebra generated by $\{X_i, Y_i, i\le n\}$. Suppose that there exists a compact set $B\in\mathbb{R}^r$ such that for all n, $E_n[V(X_n)]-V(X_n)\le -s_n\delta+Y_n$ for $X_n\notin B$, where s_n satisfies (3.26) and δ is a positive constant. Then $X_n\in B$ for infinitely large n with probability 1.

Applying this theorem with $X_t=\left(\hat{\boldsymbol{\lambda}}(t),\hat{\boldsymbol{\nu}}(t)\right), Y_t=o\left([\gamma(t)]^2\right)$, and $V\left(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\nu}}\right)=\left\|\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^*\right\|^2+\|\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}^*\|^2$, we obtain the expected result of Step 1.

Step 2: Pick any $\varepsilon>0$. We let $\varepsilon_0=\varepsilon/3$ and use the result of Step 1 for ε_0. Assume that t_1 is large enough and

$$\left(\hat{\boldsymbol{\lambda}}(t+1),\hat{\boldsymbol{\nu}}(t+1)\right)\notin B_{\varepsilon_0}.$$

We first show that

$$\left(\hat{\boldsymbol{\lambda}}(t+1),\hat{\boldsymbol{\nu}}(t+1)\right)\in B_{2\varepsilon_0}$$

with probability 1. This holds by using Chebyshev's inequality

$$\Pr\left\{\gamma(t_1)\|\boldsymbol{\delta}_\lambda(t_1)\|>\frac{\varepsilon_0}{2}\right\}\le\frac{4[\gamma(t_1)]^2E\left[\|\boldsymbol{\delta}_\lambda(t_1)\|^2\right]}{\varepsilon_0^2} \tag{3.54}$$

$$\Pr\left\{\gamma(t_1)\|\boldsymbol{\delta}_\nu(t_1)\|>\frac{\varepsilon_0}{2}\right\}\le\frac{4[\gamma(t_1)]^2E\left[\|\boldsymbol{\delta}_\nu(t_1)\|^2\right]}{\varepsilon_0^2} \tag{3.55}$$

Adding the results together, we obtain

$$\Pr\left\{\gamma(t_1)\left[\|\boldsymbol{\delta}_\lambda(t_1)\|+\|\boldsymbol{\delta}_\nu(t_1)\|\right]>\varepsilon_0\right\}\le\frac{4[\gamma(t_1)]^2E\left[\|\boldsymbol{\delta}_\lambda(t_1)\|^2+\|\boldsymbol{\delta}_\nu(t_1)\|^2\right]}{\varepsilon_0^2} \tag{3.56}$$

Because $\left(\hat{\boldsymbol{\lambda}}(t), \hat{\boldsymbol{\nu}}(t)\right) \in B_{\varepsilon_0}$ with probability 1, we conclude $\left(\hat{\boldsymbol{\lambda}}(t+1), \hat{\boldsymbol{\nu}}(t+1)\right) \in B_{2\varepsilon_0}$ with probability 1. Now, using the martingale inequality ([38], Eq. (1.4)), we have

$$\Pr\left\{\sup\left|\sum_{t=t_1+1}^{r} \gamma(t)\left\|\boldsymbol{\delta}_{\boldsymbol{\lambda}}(t)\right\|\right| \geq \frac{\varepsilon_0}{2}\right\} \leq \frac{4\limsup_t E\left[\left\|\boldsymbol{\delta}_{\lambda}(t)\right\|^2\right]}{\varepsilon_0^2} \sum_{t=t_1+1}^{\infty} [\gamma(t)]^2$$

for $r \geq t_1 + 1$. Then,

$$\lim_{r\to\infty} \Pr\left\{\sup\left|\sum_{t=t_1+1}^{r} \gamma(t)\left\|\boldsymbol{\delta}_{\lambda}(t)\right\|\right| \geq \frac{\varepsilon_0}{2}\right\} = 0, \forall \varepsilon_0 > 0 \tag{3.57}$$

and similarly,

$$\lim_{r\to\infty} \Pr\left\{\sup\left|\sum_{t=t_1+1}^{r} \gamma(t)\left\|\boldsymbol{\delta}_{\nu}(t)\right\|\right| \geq \frac{\varepsilon_0}{2}\right\} = 0, \forall \varepsilon_0 > 0 \tag{3.58}$$

From (3.57), (3.58), we get

$$\lim_{r\to\infty} \Pr\left\{\sup\left|\sum_{t=t_1+1}^{r} \gamma(t)\left[\left\|\boldsymbol{\delta}_{\lambda}(t)\right\| + \left\|\boldsymbol{\delta}_{\nu}(t)\right\|\right]\right| \geq \varepsilon_0\right\} = 0, \forall \varepsilon_0 > 0 \tag{3.59}$$

Therefore, all $\left(\hat{\boldsymbol{\lambda}}(t), \hat{\boldsymbol{\nu}}(t)\right)$, where $t > t_1$, are in $B_{3\varepsilon_0} = B_{\varepsilon}$. This completes the proof. ■

3.4 Numerical Analysis

In this section, we verify our analysis by simulation results. The simulation program has been developed using the MATLAB environment. Two simulation scenarios illustrate our analysis, a network of two users and a network of three users, respectively. Scenario 1 illustrates the convergence of the proposed algorithms as well as the effect of step size selection and the number of OFDM tones on these algorithms. Scenario 2 illustrates the optimal relay selection strategy when Algorithm 2 is used. Table 1 shows the common simulation parameters. The users are assumed to have the same priority. Hence, their weights are all set to 1. The minimum rate and the maximum transmitted power of each user are 10Mbps and 20dB (normalized by the average noise power N_0W), respectively. The network topologies for these two scenarios are depicted in Fig. 3.4.

Table 3.1
Common parameters for numerical analysis of the algorithms

Symbol	Parameter name	Value
N	Number of subcarriers	16, 32
K	Number of users	2, 3
w_k	Weight of user k, k = 1, 2, 3	[1 1 1]
R_k	Minimum rate of user k	[10 10 10] Mbps
$P_{i,max}/(N_0W)$	Power constraint for all users	20dB
ε	Error goal	10^{-4}

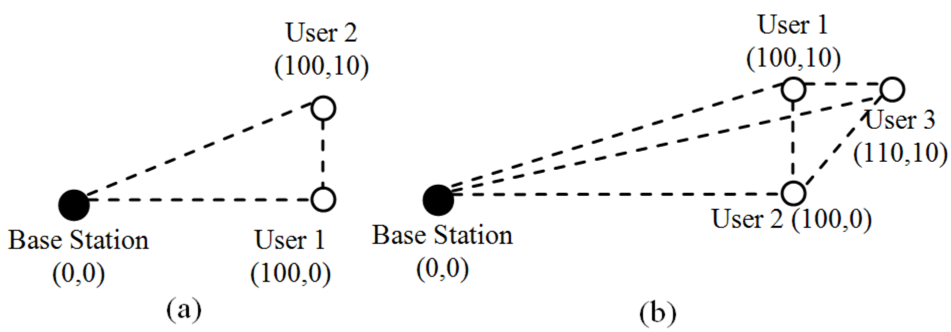

Fig. 3.4. Simulated network topology

3.4.1 Scenario 1: Network with 2 users

Assume that the base station is located at (0,0) while user 1 and user 2 are located at (100,0) and (100,10), respectively, on a Cartesian coordinate plane. Some numerical results follow together with the conclusions drawn from these results.

Convergence of the Algorithm:

Figure 3.5 and 3.6 shows the numerical results when using Algorithm 1 in the condition of known CDI. We observe that this algorithm converges after 500 iterations. Figure 3.7 shows the similar results for the unknown CDI, but with perfect CSI. The number of iterations is about 180, and the running time has been improved significantly (89.2 seconds, while the algorithm 1 takes 1292 seconds). This fact confirms the theoretical analysis that was developed. Although Algorithm 2 is just an approximation approach, it improves the computational efficiency because it relaxes the computation from the complexity of evaluating the expectation over the distribution of all channels and users. However, Algorithm 2 may not converge uniformly. It may suffer from oscillations even at a large number of iterations due to the randomness of the channel gain.

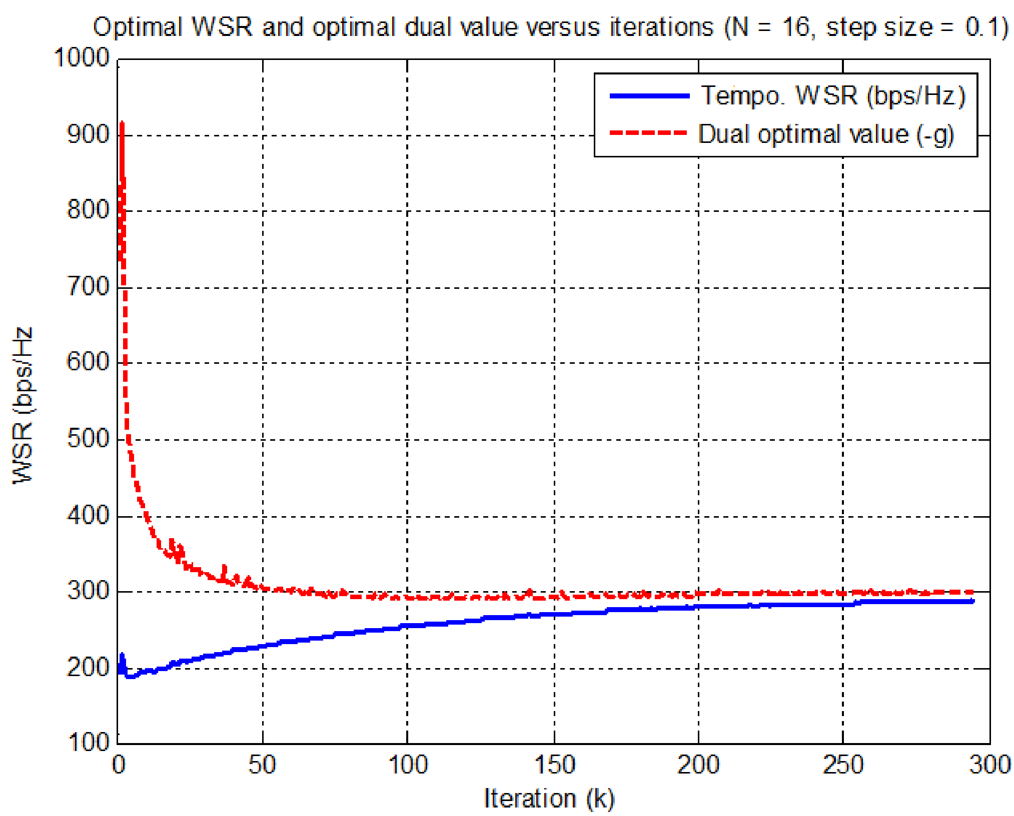

Fig. 3.5. The convergence of Algorithm 1 with the condition of known CDI

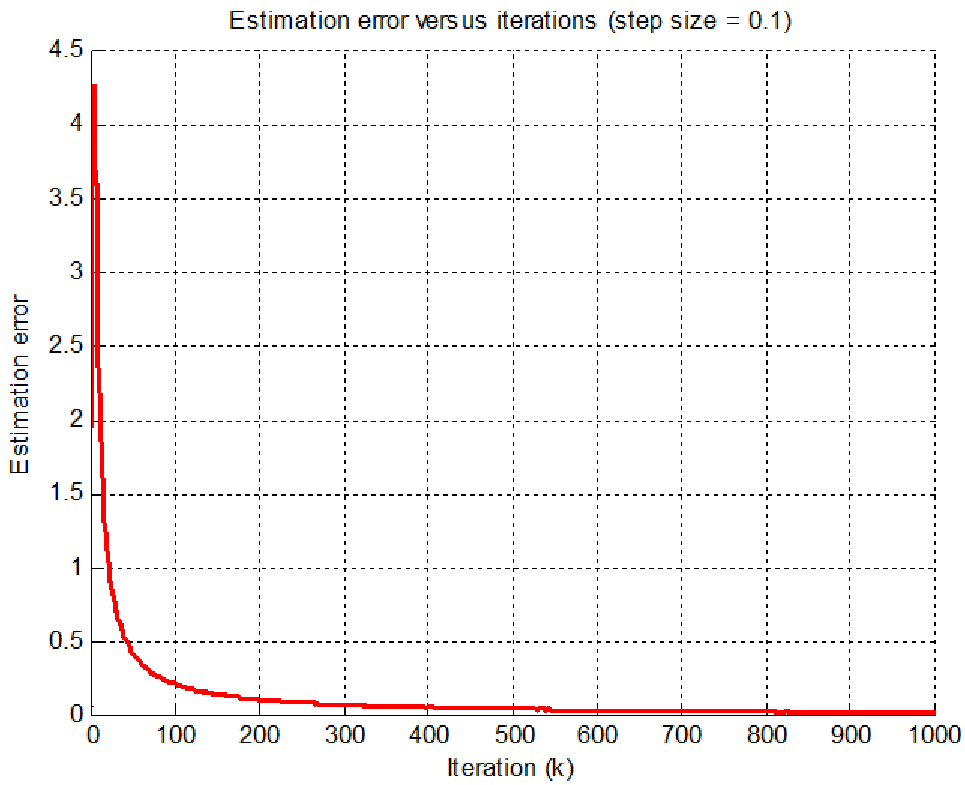

Fig. 3.6. Estimation error of Algorithm 1 in condition of known CDI

The optimal dual solutions are $\boldsymbol{\lambda}^* = [9.6410\ 9.6220]$ and $\boldsymbol{\nu}^* = [0\ 0]$. This can be interpreted in the following way. The optimal power allocation and scheduling should satisfy the rate constraint strictly to have the maximum sum rate. Hence, by the complementary slackness condition, we must have $\boldsymbol{\nu}^* = [0\ 0]$. On the other hand, to get the maximum rate, most likely the users take their full power capability, so the equality must occur in the power constraints. Hence, also by complementary slackness condition, $\boldsymbol{\lambda}^* > 0$. If the constraints are so strict, for example, if $R_{i,min}$ is too large, the Slater condition does not hold, and the algorithms above will not converge. The duality gap in this case is $(-d*) - (p*) = 295.89 - 280.32 = 15.57(bps/Hz)$. This is because the number of tones is not large enough (N = 16 in this case).

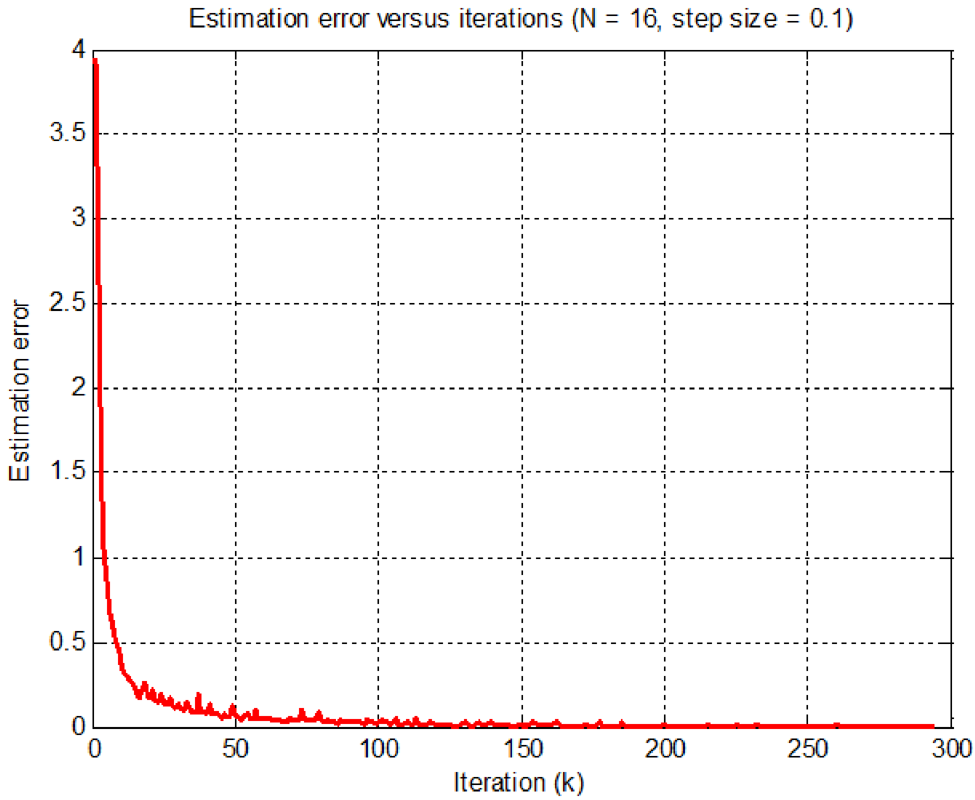

Fig. 3.7. Estimation error of Algorithm 2 with the condition of perfect CSI

Effectiveness of step size on the convergence rate

Figure 3.8 shown that the number of required iterations of Algorithm 1 is nearly proportional to the initial step size for small values of step size. Algorithm 2 converges faster and requires less running than Algorithm 1. In a manner similar to Algorithm 1, in general, Algorithm 2 converges faster when the initial step size is small. However, if the step size is too small, the number of iterations may increase again.

Effectiveness of the number of tones

As mentioned before, the duality gap will eventually go to zero when the number of tones goes to infinity. For the case in which N = 16 as shown in Fig. 3.5, we observed that the duality gap is 19. Figure 3.9 below shows that the duality gap

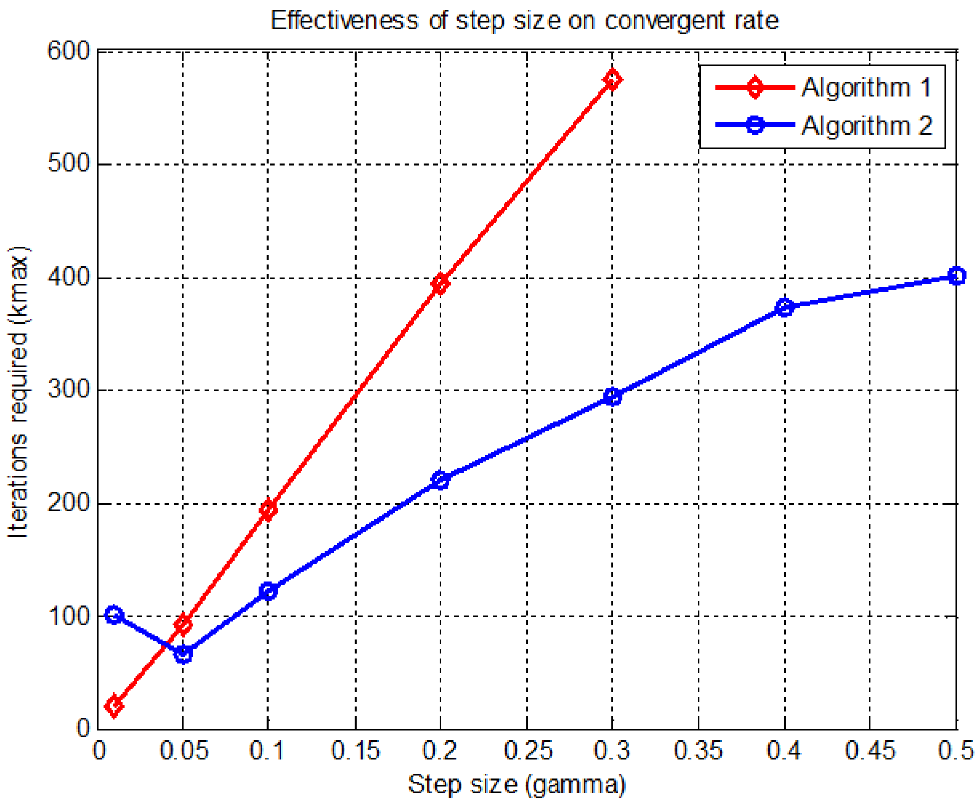

Fig. 3.8. Effect of step size selection on the convergence rate of algorithms 1 and 2

is reduced to 2.23 when the number of tones increases from 16 to 32. Because the number of subchannels is doubled, the bandwidth of each channel must be divided by 2 to make the comparison fair.

3.4.2 Scenario 2: Networks with 3 users

Figure 3.10 and 3.11 illustrate the results of running Algorithm 3 for the network of 3 users. This procedure is more time consuming, but we can observe that it also converges. The optimal dual solution are $\boldsymbol{\lambda}^* = [3.9738, 3.9802, 3.9701]$ and $\boldsymbol{\nu}^* = [0\ 0\ 0]$. The optimal WSR is 30.64 bps/Hz.

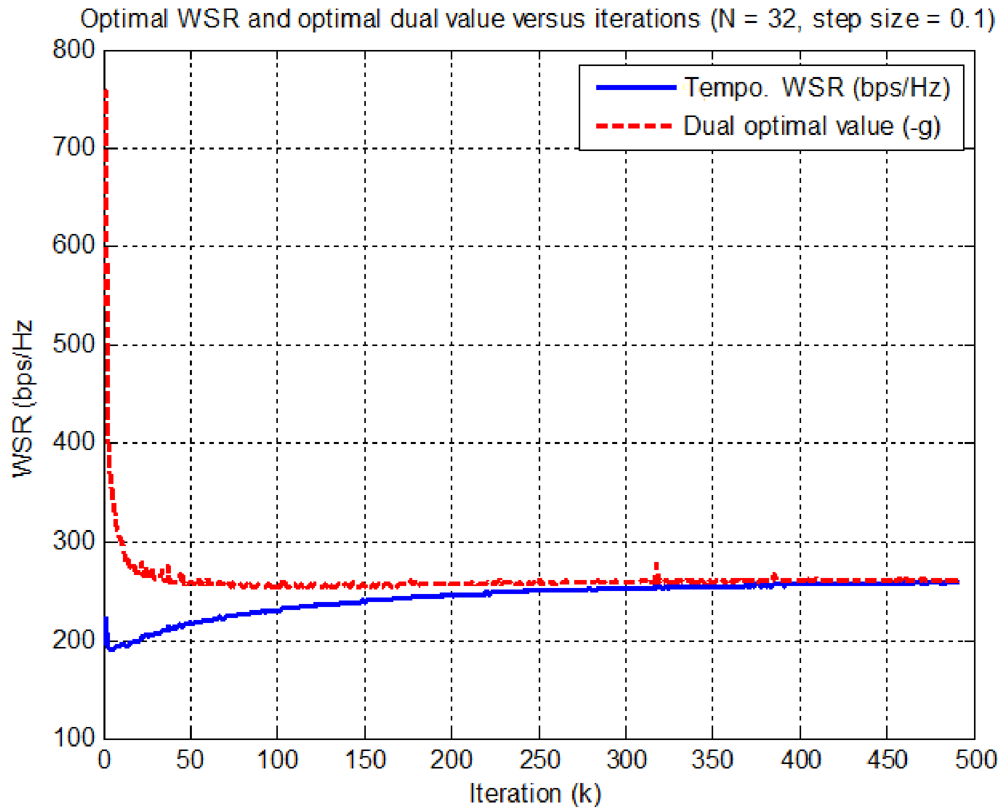

Fig. 3.9. Duality gap approaches zero when the number of tones increases to 32

3.5 Summary

In this chapter we investigate in detail the algorithms to jointly optimize the power allocation at the physical layer and scheduling at the MAC layer of wireless OFDM cooperative networks. Although there are several research papers that have explored these issues before, the most important results of this project is investigating the convergence of joint optimization algorithms for the power allocation, scheduling, and relay selection strategy in a specific model of cooperative networks, both by mathematical analysis and by computer simulation. It has been shown that the joint optimization problem can be solved successfully by using the dual optimization method for multicarrier systems, proposed in [1]. Here, we only consider the case in which the channel state information is error free. Indeed, we need to show that if

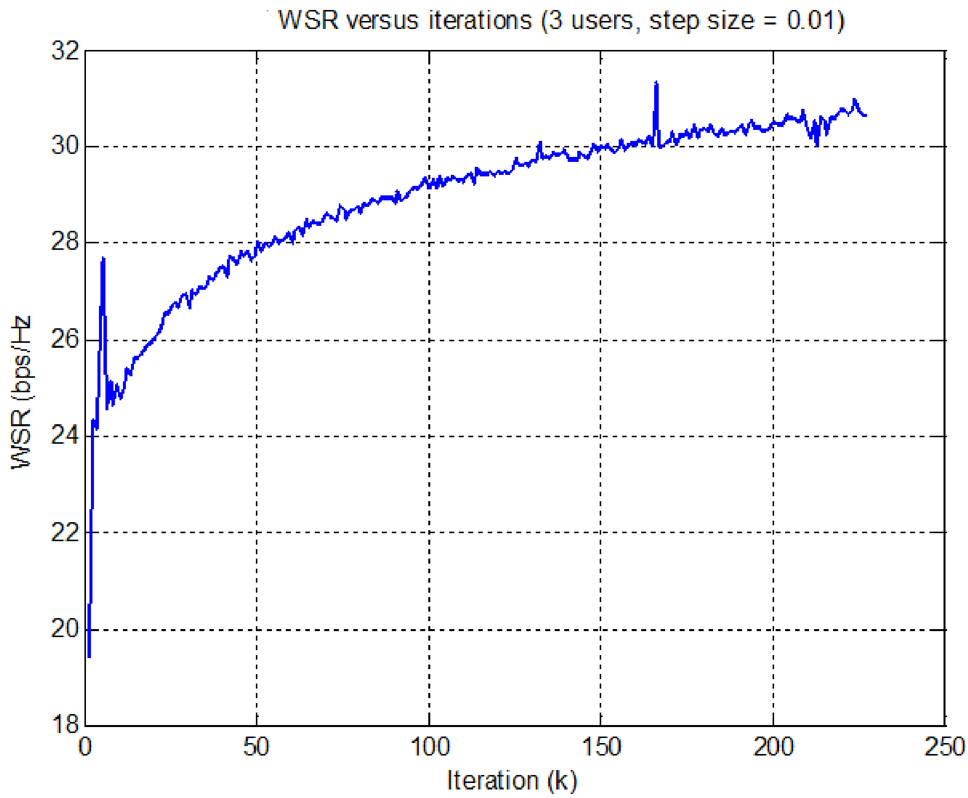

Fig. 3.10. Convergence of Algorithm 3 in condition of known CDI

the CSI has errors and the estimation error has zero mean, then Algorithm 2 still converges to the optimal solution with probability 1. Verifying this statement and further investigation of this problem with other conditions of the channel are the purposes of the next chapter.

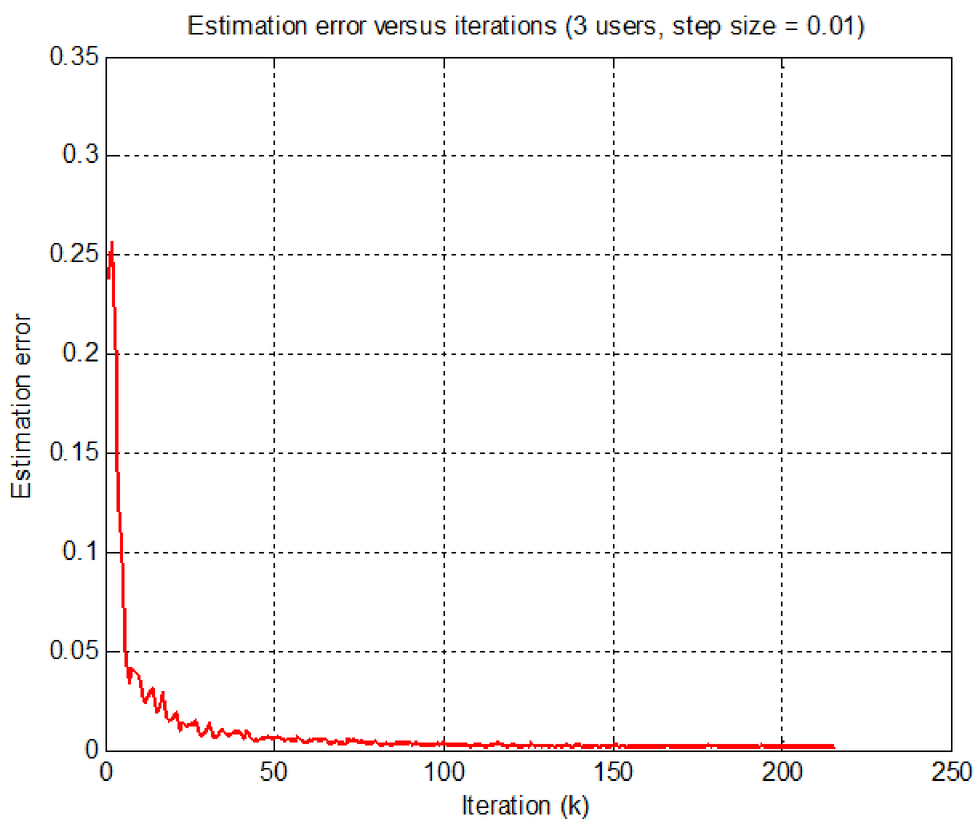

Fig. 3.11. Estimation error of Algorithm 3 for the network of 3 users

4. JOINT OPTIMIZATION OF POWER ALLOCATION AND RELAY SELECTION STRATEGY IN WIRELESS OFDM NETWORKS WITH IMPERFECT CSI

In the previous chapter, we proposed the detailed algorithms for joint optimization of power allocation, scheduling, and relay selection in cooperative networks. However, we only examined the ideal case in which the channel state information (CSI) can be fed back to the base station without error.

Here, we extend the result by considering the case in which CSI is imperfectly fed back to the base station. This work has already been published in [42]. Since the algorithm of [42] uses the instantaneous feedback of CSI at each iteration, if an error is introduced by the CSI, the resulting solution from the algorithm may no longer be optimal. Here, we address the conditions that cause the proposed algorithm to converge to the optimal solution and examine the impact of channel estimation errors on the solution. We find that the stochastic subgradients must be unbiased to achieve the exact convergence, and we also evaluate the impact of the CSI error.

4.1 System Model and Problem Formulation

We consider again the system model described in previous chapter (Fig. 3.1), with one modification. The channel state information (CSI) is now assumed to be imperfectly fed back, i.e., there is an error between the real channel gains and the channel gains which are fed back to the BS. We denote by $\tilde{\mathbf{H}}^{(n)} = \{h_{ij}^{(n)}\}_{i=1,\ldots,K;\,j=1,\ldots,K;\,n=1,\ldots,N}$ the channel gain matrices with estimation errors, and denote by $\widetilde{\mathbf{P}^{(n)}})$ and $\tilde{\boldsymbol{\tau}}^{(n)}$ the resultant optimal allocation matrix and scheduling vector when $\mathbf{H}^{(n)}$ is replaced by $\tilde{\mathbf{H}}^{(n)}$.

The basic notation derived from the previous chapter is indicated below.

- $\mathcal{K} = \{0, 1, 2, ..., K\}$ is the set of user nodes, where user 0 is the base station.
- $\mathcal{N} = \{1, 2, ..., N\}$ is the set of tones.
- $h_{ij}^{(n)}$ is the channel gain between user i and user j over the n^{th} tone.
- $\tau_i^{(n)}$ is the time interval for user i over the n^{th} tone, and $\boldsymbol{\tau} = \{\tau_i^{(n)}\}_{i=1,...,K;\, n=1,...,N}$.
- $r_i^{(n)}$ is the node that acts as the relay for user i over the n^{th} tone. The relay protocol is described as follows. In the first half of the interval $\tau_i^{(n)}$, user i transmits its own data, the relay listens; in the remaining half of the interval, the relay decodes and forwards the information received in the first half of the interval to the BS.
- $P_{ij}^{(n)}$ is the power consumed by user i to send the n^{th} tone to user j, and $\mathbf{P} = \{P_{ij}^{(n)}\}_{i=1,...,K;\, j=1,...,K;\, n=1,...,N}$.

Our goal is to select a cooperation strategy that maximizes the weighted sum rate (WSR) of the network, while keeping the resources (power and time) constrained and also satisfying the fairness condition in this network.

Problem

$$\underset{\substack{P_{ij}^{(n)},\tau_i^{(n)},r_i^{(n)} \\ i=1,...,K;\, j=1,2,...K;\, n=1,2...N}}{\text{maximize}} \qquad WSR = \sum_{i=1}^{K} w_i C_i \tag{4.1}$$

$$\text{subject to:} \quad \sum_{i=1}^{K} \tau_i^{(n)} \leq 1 \quad for\ i = 1, 2, ..., N \tag{4.2}$$

$$\overline{P}_i \leq p_{i,\max} \quad (i = 1, 2, ..., K) \tag{4.3}$$

$$C_i \geq R_i \quad (i = 1, 2, ..., K). \tag{4.4}$$

4.2 Summary of Solutions and Algorithms

As established in the last chapter, the power allocation and scheduling solutions are given in (3.38) and (3.39) if the CSI is fed back perfectly. However, with imperfect

CSI, we don't know the real channel gains $h_{ij}^{(n)}$. Instead, we use $\tilde{h}_{ij}^{(n)}$ in place of $h_{ij}^{(n)}$. Hence, the power allocation and scheduling solutions in this case are given by

$$\tilde{P}_{ji}^{(n)*} = \frac{(w_i + \nu_i)}{\tilde{\lambda}_j \ln 2} - \left(\rho_{j0}\left|\tilde{h}_{j0}^{(n)}\right|^2\right)^{-1} (i = 1, 2, .., K;\ j = 1, r_i) \tag{4.5}$$

and

$$\tilde{\tau}_i^{(n)*} = \begin{cases} 1, & if\ \ i = \arg\max_k \left(E\left[f_k\left(\tilde{P}_{kk}^{(n)*}, \tilde{P}_{r_k k}^{(n)*}\right)\right]\right) \\ 0, & otherwise \end{cases} \tag{4.6}$$

To compute the optimal solution $\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*$, we use the subgradient projection method [38]. Since the full channel distribution information is not available, we adaptively use the current values of $P_{ij}^{(n)*}$ and $\tau_i^{(n)*}$ to compute the subgradients, rather than using the expectation of all these values [43] to obtain

$$\tilde{g}_{\lambda_i}(t) = \sum_{n=1}^{N}\left(\frac{1}{2}\tilde{\tau}_i^{(n)*}\tilde{P}_{ii}^{(n)*} + \frac{1}{2}\tilde{\tau}_j^{(n)*}\tilde{P}_{ij}^{(n)*}\right) - P_{i,max} \tag{4.7}$$

and

$$\tilde{g}_{\nu_i}(t) = R_i - \left[\frac{1}{2}\sum_{n=1}^{N}\left(\tilde{\tau}_i^{(n)*}\sum_{j\in\{i,r_i\}}\log_2\left|1 + \rho_{j0}\left|\tilde{h}_{j0}^{(n)}\right|^2\tilde{P}_{ji}^{(n)*}\right|\right)\right] \tag{4.8}$$

where j is such that $r_j^{(n)} = i$.

The joint optimization algorithm proposed in [43] can be summarized as

1. Choose an initial relay selection strategy for each tone of the OFDMA system.

2. Solve each per-tone optimization problem:

a. Initial condition: Choose $\tilde{\boldsymbol{\lambda}}(0)$ *and* $\tilde{\boldsymbol{\nu}}(0)$*, and then use* (4.5) *and* (4.6) *for i=1,...,K; j =1,2,...,K; n=1,...,N.*

b. Iteration: At step t + 1, choose step size $\gamma(t) = \gamma/t$*, where* γ *is some positive constant. Compute* $\tilde{g}_{\lambda_i}(t)$ *and* $\tilde{g}_{\nu_i}(t)$ *for i = 1,2,...,K using* (4.7) *and* (4.8)*. Update* $\tilde{\lambda}_i(t+1)$ *and* $\tilde{\nu}_i(t+1)$ *for i = 1,2,...,K by using the following formulas given by*

$$\begin{aligned} \tilde{\lambda}_i(t+1) &= \left[\tilde{\lambda}_i(t) + \gamma(t)\tilde{g}_{\lambda_i}(t)\right]^+ \\ \tilde{\nu}_i(t+1) &= \left[\tilde{\nu}_i(t) + \gamma(t)\tilde{g}_{\nu_i}(t)\right]^+ \end{aligned} \tag{4.9}$$

where $[x]^+ = max\{x, 0\}$.

c. Termination: Using ϵ as the maximum error allowed, stop the algorithm whenever $\left\|\tilde{\boldsymbol{\lambda}}(t+1) - \tilde{\boldsymbol{\lambda}}(t)\right\| < \epsilon$ and $\|\tilde{\boldsymbol{\nu}}(t+1) - \tilde{\boldsymbol{\nu}}(t)\| < \epsilon$ for $i = 1, 2, ..., K$. Otherwise, increase t by 1 and repeat Step 2b.

d. Store the optimal solution, and then change to another relay selection strategy. Repeat steps 2a through 2c. If the optimal value is better than the previous value, store the new optimal solution. Otherwise, discard it. Repeat Step 2d until there is no remaining strategy to consider.

3. Repeat Step 2 for the other tones until all tones have been processed.

4.3 Convergence Analysis

The convergence analysis for the case of perfect CSI has been examined in [42] (propositions 3 and 4). We can rewrite the stochastic subgradients as

$$\tilde{g}_{\lambda_i}(t) = g_{\lambda_i}(t) + \Delta_{\lambda_i}(t) + \tilde{\delta}_{\lambda_i}(t) \tag{4.10}$$

and

$$\tilde{g}_{\nu_i}(t) = g_{\nu_i}(t) + \Delta_{\nu_i}(t) + \tilde{\delta}_{\nu_i}(t) \tag{4.11}$$

where $\Delta_{\lambda_i}(t)$ and $\Delta_{\nu_i}(t)$ are the means of estimation errors ($i = 1, 2, ..., K$), which are given by

$$\begin{aligned} \Delta_{\lambda_i}(t) &\triangleq E\left[\tilde{g}_{\lambda_i}(t)\right] - g_{\lambda_i}(t) \\ &= \sum_{n=1}^{N} \frac{1}{2} E\left[\left(\tilde{\tau}_i^{(n)}\tilde{P}_{ii}^{(n)} + \tilde{\tau}_j^{(n)}\tilde{P}_{ij}^{(n)}\right) - \left(\tau_i^{(n)*}P_{ii}^{(n)*} + \tau_j^{(n)*}P_{ij}^{(n)*}\right)\right] \end{aligned} \tag{4.12}$$

and

$$\begin{aligned} \Delta_{\nu_i}(t) \triangleq E\left[\tilde{g}_{\nu_i}(t)\right] - g_{\nu_i}(t) = E&\left[\frac{1}{2}\sum_{n=1}^{N}\left(\tau_i^{(n)*}\sum_{j\in\{i,r_i\}}\log_2\left|1 + \rho_{j0}\left|h_{j0}^{(n)}\right|^2 P_{ji}^{(n)*}\right|\right)\right] \\ &- E\left[\frac{1}{2}\sum_{n=1}^{N}\left(\tilde{\tau}_i^{(n)}\sum_{j\in\{i,r_i\}}\log_2\left|1 + \rho_{j0}\left|h_{j0}^{(n)}\right|^2 \tilde{P}_{ji}^{(n)}\right|\right)\right] \end{aligned} \tag{4.13}$$

The quantities $\tilde{\delta}_{\lambda_i}(t)$ and $\tilde{\delta}_{\nu_i}(t)$ are given by

$$\tilde{\delta}_{\lambda_i}(t) \triangleq \tilde{g}_{\lambda_i}(t) - E\left[\tilde{g}_{\lambda_i}(t)\right] = \sum_{n=1}^{N} \frac{1}{2}\left[\tilde{\tau}_i^{(n)}\tilde{P}_{ii}^{(n)} + \tilde{\tau}_j^{(n)}\tilde{P}_{ij}^{(n)}\right] - \sum_{n=1}^{N}\frac{1}{2}E\left[\tilde{\tau}_i^{(n)}\tilde{P}_{ii}^{(n)} + \tilde{\tau}_j^{(n)}\tilde{P}_{ij}^{(n)}\right] \tag{4.14}$$

and

$$\begin{aligned}\tilde{\delta}_{\nu_i}(t) \triangleq \tilde{g}_{\nu_i}(t) - E\left[\tilde{g}_{\nu_i}(t)\right] = E\left[\frac{1}{2}\sum_{n=1}^{N}\left(\tilde{\tau}_i^{(n)}\sum_{j\in\{i,r_i\}}\log_2\left|1+\rho_{j0}\left|h_{j0}^{(n)}\right|^2\tilde{P}_{ji}^{(n)}\right|\right)\right] \\ - \frac{1}{2}\sum_{n=1}^{N}\left(\tilde{\tau}_i^{(n)}\sum_{j\in\{i,r_i\}}\log_2\left|1+\rho_{j0}\left|h_{j0}^{(n)}\right|^2\tilde{P}_{ji}^{(n)}\right|\right)\end{aligned} \tag{4.15}$$

For convenience, we use the notation described by the equations

$$\Delta_{\boldsymbol{\lambda}}(t) = \left[\Delta_{\lambda_1}(t), \Delta_{\lambda_2}(t), ..., \Delta_{\lambda_K}(t)\right]^T$$

and

$$\Delta_{\boldsymbol{\nu}}(t) = \left[\Delta_{\nu_1}(t), \Delta_{\nu_2}(t), ..., \Delta_{\nu_K}(t)\right]^T$$

Theorem 4.3.1 *If $E\left[\Delta_{\boldsymbol{\lambda}}(t)\right] = 0$ and $E\left[\Delta_{\boldsymbol{\nu}}(t)\right] = 0$ for every t and if the step size $\gamma(t)$ satisfies the convergence conditions (in [43]), then the proposed algorithm converges with probability one to the optimal solution.*

Proof. We can write $\tilde{\lambda}_i(t+1) = \tilde{\lambda}_i(t) + \gamma(t)\left[g_{\lambda_i}(t) + \Delta_{\lambda_i}(t) + \tilde{\delta}_{\lambda_i}(t)\right] + C$, where C is a non-negative term to guarantee that $\tilde{\lambda}_i(t+1) \geq 0$. Now, we can easily show that

$$\begin{aligned}\|\tilde{\boldsymbol{\lambda}}(t+1) - \boldsymbol{\lambda}^*\|^2 \leq \|\tilde{\boldsymbol{\lambda}}(t) - \boldsymbol{\lambda}^*\|^2 - 2\gamma(t)\langle\left[\tilde{\boldsymbol{\lambda}}(t) - \boldsymbol{\lambda}^*\right], \left[g_\lambda(t) + \Delta_\lambda(t) + \tilde{\delta}_\lambda(t)\right]\rangle \\ + \left[\gamma(t)\right]^2 \|g_\lambda(t) + \Delta_\lambda(t) + \tilde{\delta}_\lambda(t)\|^2\end{aligned} \tag{4.16}$$

Noticing that $E\left[\Delta_{\boldsymbol{\lambda}}(t)\right] = \mathbf{0}$, $E\left[\tilde{\delta}_{\boldsymbol{\lambda}}(t)\right] = \mathbf{0}$, and each term in (4.16) is bounded, we obtain

$$E\left[\|\tilde{\boldsymbol{\lambda}}(t+1) - \boldsymbol{\lambda}^*\|^2\right] \leq \|\tilde{\boldsymbol{\lambda}}(t) - \boldsymbol{\lambda}^*\|^2 - 2\gamma(t)\left\langle\tilde{\boldsymbol{\lambda}}(t) - \boldsymbol{\lambda}^*, g_{\boldsymbol{\lambda}}(t)\right\rangle + 0\left(\left[\gamma(t)\right]^2\right).$$

A similar inequality can be obtained for $\tilde{\boldsymbol{\nu}}$. Adding these two inequalities together, we have

$$E\left[\|\tilde{\boldsymbol{\lambda}}(t+1)-\boldsymbol{\lambda}^*\|^2+\|\tilde{\boldsymbol{\nu}}(t+1)-\boldsymbol{\nu}^*\|^2\right] \leq \|\tilde{\boldsymbol{\lambda}}(t)-\boldsymbol{\lambda}^*\|^2+|\tilde{\boldsymbol{\nu}}(t)-\boldsymbol{\nu}^*\|^2$$
$$-2\gamma(t)\left[\left\langle\tilde{\boldsymbol{\lambda}}(t)-\boldsymbol{\lambda}^*, g_{\boldsymbol{\lambda}}(t)\right\rangle+\left\langle\tilde{\boldsymbol{\lambda}}(t)-\boldsymbol{\lambda}^*, g_{\boldsymbol{\lambda}}(t)\right\rangle\right]+0\left([\gamma(t)]^2\right).$$

From this point, the proof is similar to the proof of Theorem 3.3.1. ■

Theorem 4.3.2 *If $E\left[\Delta_{\lambda_i}(t)\right] \neq 0$ or $E\left[\Delta_{\nu_i}(t)\right] \neq 0$ for some i and if the step size $\gamma(t)$ satisfies the convergence condition (in [43]), then $(\tilde{\boldsymbol{\lambda}}(t), \tilde{\boldsymbol{\nu}}(t)) : t=1,2,...$ approaches some neighborhood $\Re(d)$ around the optimal solution with probability one, where $\Re(d)$ is defined as*

$$\Re(d) \triangleq \left\{(\boldsymbol{\lambda}, \boldsymbol{\nu}): \begin{array}{l} d|g_{\lambda_i}(\boldsymbol{\lambda}, \boldsymbol{\nu})| \leq \bar{\Delta}_{\lambda_i}\ for\ some\ i \\ d|g_{\nu_i}(\boldsymbol{\lambda}, \boldsymbol{\nu})| \leq \bar{\Delta}_{\nu_i}\ for\ some\ i \\ 0 \leq d \leq 1 \end{array}\right\} \tag{4.17}$$

where $\bar{\Delta}_{\lambda_i} = \limsup_{t\to\infty} E\left[\Delta_{\lambda_i}(t)\right]$ and $\bar{\Delta}_{\nu_i} = \limsup_{t\to\infty} E\left[\Delta_{\nu_i}(t)\right]$.

Proof. Define $V(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \triangleq \|\tilde{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^*\|^2+\|\tilde{\boldsymbol{\nu}}-\boldsymbol{\nu}^*\|^2$. Consider the region $\Re(d)\bigcup B_\epsilon$, where $B_\epsilon = \left\{\left(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}\right) : V\left(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}\right) \leq \epsilon\right\}$. Assume that $\left(\tilde{\boldsymbol{\lambda}}(t), \tilde{\boldsymbol{\nu}}(t)\right) \notin \Re(d)\bigcup B_\epsilon$, then $\limsup_{t\to\infty} E\left[\Delta_{\lambda_i}(t)\right] < d|g_{\lambda_i}(\boldsymbol{\lambda}, \boldsymbol{\nu})|$. Using (4.16) with the fact that $E\left[\tilde{\boldsymbol{\delta}}_\lambda(t)\right]=0$, we obtain

$$E\left[\|\tilde{\boldsymbol{\lambda}}(t+1)-\boldsymbol{\lambda}^*\|^2\right] \leq \|\tilde{\boldsymbol{\lambda}}(t)-\boldsymbol{\lambda}^*\|^2-2\gamma(t)(1-d)\left\langle\tilde{\boldsymbol{\lambda}}(t)-\boldsymbol{\lambda}^*, g_{\boldsymbol{\lambda}}(t)\right\rangle+0\left([\gamma(t)]^2\right)$$

Finally, by applying the same method for $\tilde{\boldsymbol{\nu}}$ and adding the two resultant inequalities together, we obtain the inequality

$$E\left[\|\tilde{\boldsymbol{\lambda}}(t+1)-\boldsymbol{\lambda}^*\|^2+\|\tilde{\boldsymbol{\nu}}(t+1)-\boldsymbol{\nu}^*\|^2\right] \leq \|\tilde{\boldsymbol{\lambda}}(t)-\boldsymbol{\lambda}^*\|^2+|\tilde{\boldsymbol{\nu}}(t)$$
$$-\boldsymbol{\nu}^*\|^2-2\gamma(t)(1-d)\left[\left\langle\tilde{\boldsymbol{\lambda}}(t)-\boldsymbol{\lambda}^*, g_{\boldsymbol{\lambda}}(t)\right\rangle+\left\langle\tilde{\boldsymbol{\lambda}}(t)-\boldsymbol{\lambda}^*, g_{\boldsymbol{\lambda}}(t)\right\rangle\right]+0\left([\gamma(t)]^2\right)$$

By the same argument as in the proof of Theorem 3.3.1, the points $\{(\tilde{\boldsymbol{\lambda}}(t), \tilde{\boldsymbol{\nu}}(t)) : t=1,2,...\}$ reside in $\Re(d)\bigcup B_\epsilon$ almost surely. By letting $\epsilon \to \infty$, we complete the proof. ■

The effect of channel estimation error on the accuracy of the solutions is assessed with the following theorem.

Theorem 4.3.3 *The neighborhood $\Re(d)$ described in Theorem 4.3.2 is inner bounded by* $\{(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) : |g_{\lambda_i}(\boldsymbol{\lambda}, \boldsymbol{\nu})| = \bar{\Delta}_{\lambda_i} \, and \, |g_{\nu_i}(\boldsymbol{\lambda}, \boldsymbol{\nu})| = \bar{\Delta}_{\nu_i} \, for \, all \, i\}$.

Proof. From Theorem 4.3.2, there exists $d \in [0, 1)$ such that $E\left[\Delta_{\lambda_i}(t)\right] \leq d|g_{\lambda_i}(\boldsymbol{\lambda}, \boldsymbol{\nu})| \leq \bar{\Delta}_{\lambda_i}$, for $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \Re(d)$. Hence, we can write: $\frac{E\left[\Delta_{\lambda_i}(t)\right]}{|g_{\lambda_i}(\boldsymbol{\lambda}, \boldsymbol{\nu})|} \leq d < 1$. This holds for any $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \Re(d)$. As $n \to \infty$, $|g_{\lambda_i}(\boldsymbol{\lambda}, \boldsymbol{\nu})|$ decreases (since $(\boldsymbol{\lambda}(t), \boldsymbol{\nu}(t))$ moves toward the optimal point). Hence, d approaches 1, and we conclude that $\Re(d)$ is inner bounded by points satisfying $|g_{\lambda_i}(\boldsymbol{\lambda}, \boldsymbol{\nu})| = \bar{\Delta}_{\lambda_i}$. Similarly, we can show $|g_{\nu_i}(\boldsymbol{\lambda}, \boldsymbol{\nu})| = \bar{\Delta}_{\nu_i}$. ■

4.4 Numerical Results

In this section, a network of five users is simulated to evaluate the impact of channel estimation error. The network topology is shown in Fig. 4.1. The users are assumed to have the same priority, i.e., the weight vector is chosen to be [1 1 1 1 1]. The minimum rate and the maximum transmitted power of each user are 1 Mbps and 20 dB (normalized by the average noise power N_0W), respectively, where W represents the equivalent noise bandwidth and N_0 represents the single-sided noise spectral density. The number of OFDM tones is N = 16. The channel for each tone is assumed to be i.i.d. Rayleigh fading. The SNRs for 5 users depend on the distance from the users to the BS.

Fig. 4.2 compares the results when running the proposed algorithm for two channel conditions, that is, perfect CSI (dashed curve) and imperfect CSI with unbiased subgradients (solid curve). We observe that the algorithm converges to the same optimal solution in each case, although with different convergence rates. With the larger number of users (more than 5 users), the algorithm still works, but the complexity increases because an exhaustive search over all permutations of the set of users is required. However, this problem can be solved if we eliminate some obviously bad

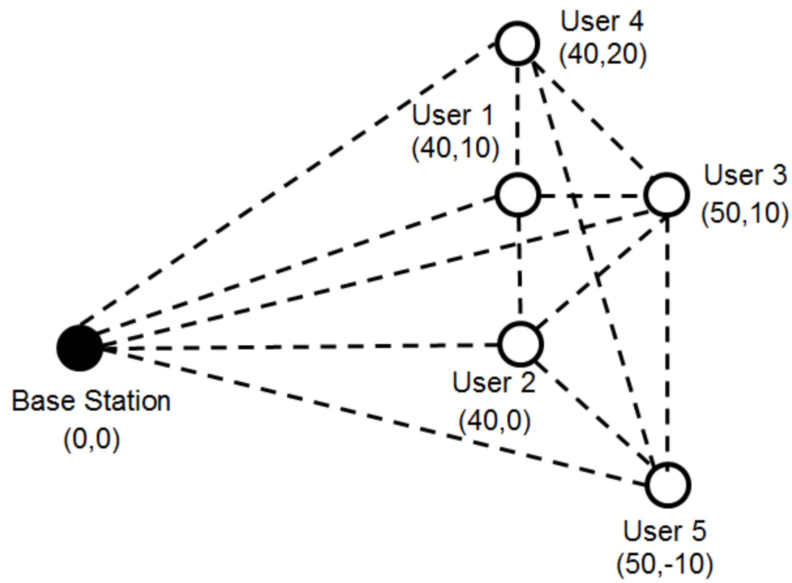

Fig. 4.1. Simulated network topology

relay strategies before searching. In Fig. 4.3, we consider the imperfect CSI condition with biased stochastic subgradients. Fig. 4.3 shows that the proposed algorithm still converges, but to a different solution. Here, we consider two cases, that is, $\Delta_{\lambda_i}(t)$ and $\Delta_{\nu_i}(t)$ are both Gaussian distributed $N(0.5, 0.01)$ (case 1) or $N(0.75, 0.01)$ (case 2). We see that the difference between the convergence point and the real optimal WSR value is larger for larger channel estimation errors. Note that in this case the WSR (suboptimal solution for the primal problem) is less than the optimal solution, but the algorithm may converge to a greater value than the optimal solution because of the convexity of the Lagrangian dual function.

4.5 Conclusion

In this chapter the convergence of the algorithms to jointly optimize the power allocation at the physical layer and scheduling at the MAC layer of wireless OFDM

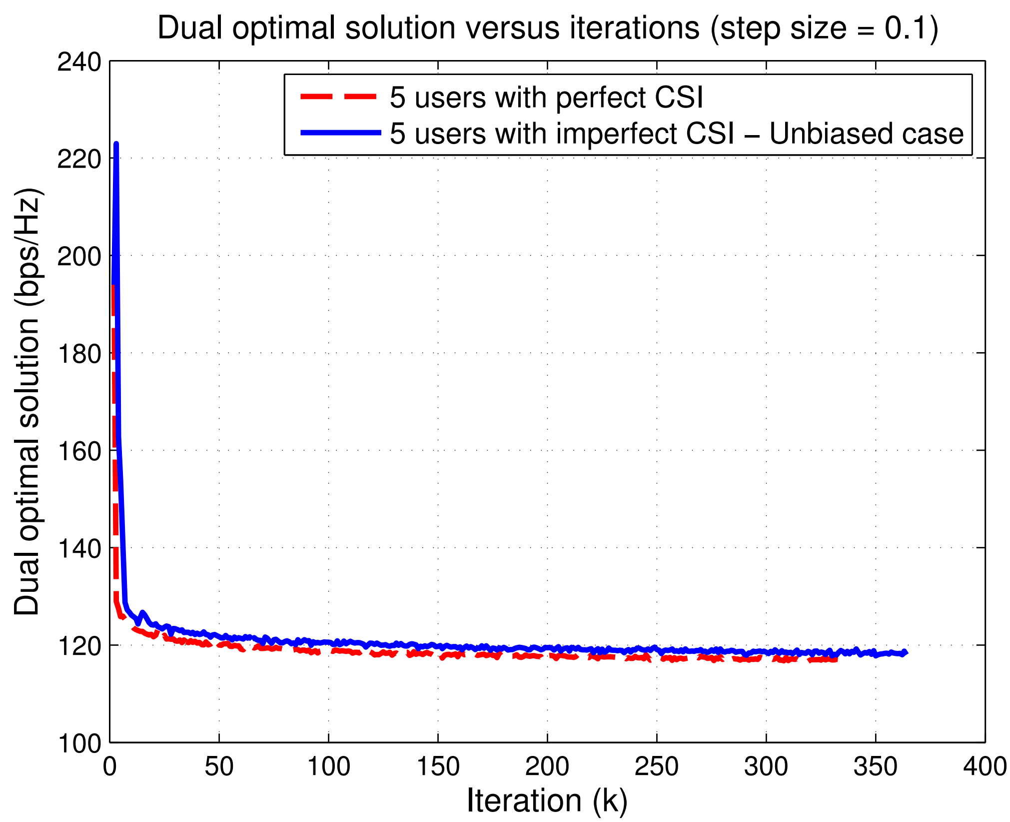

Fig. 4.2. Convergence of the algorithm with imperfect CSI (unbiased case)

cooperative networks has been investigated when the channel estimation is imperfect. Using the optimization framework proposed in [44], the joint optimization problem can be solved successfully even if the CSI is reported to the BS in error. More specifically, if the CSI has errors and the estimation error has zero mean, then the proposed algorithm still converges to the optimal solution with probability 1. Otherwise, if the estimation error has nonzero mean, then the proposed algorithm converges to a point in some neighborhood of the optimal solution.

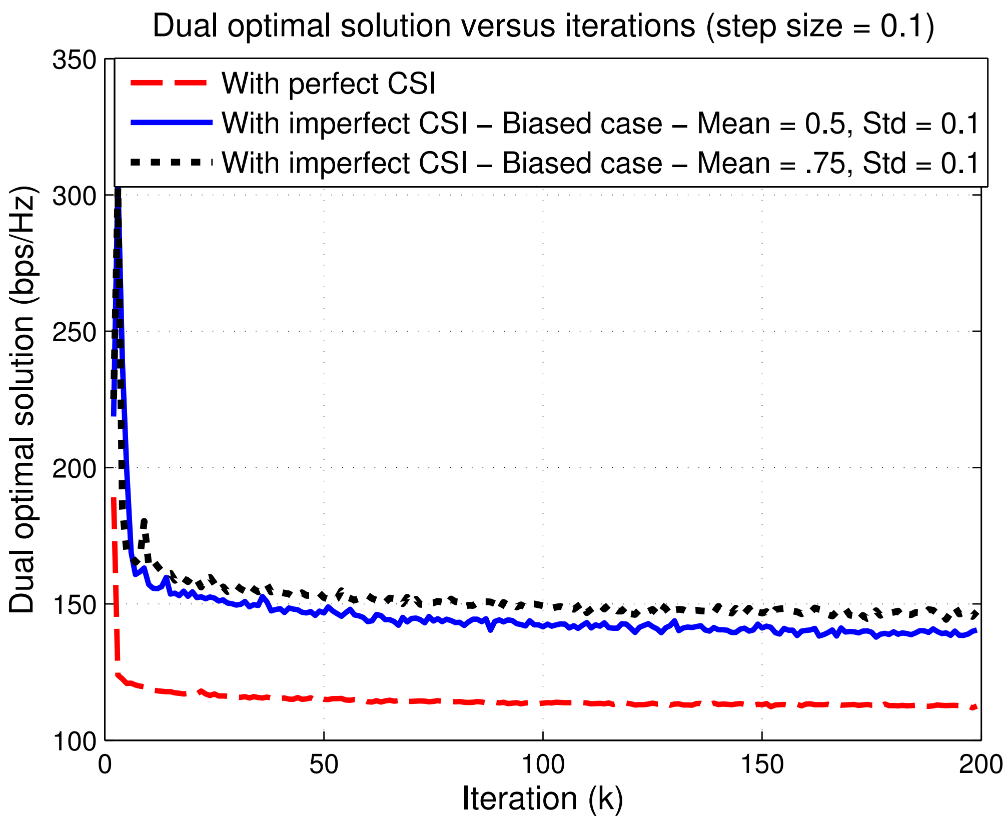

Fig. 4.3. Convergence of the algorithm with imperfect CSI (biased case)

5. OPTIMIZATION OF POWER ALLOCATION AND INTEGER COEFFICIENTS OF RELAY FUNCTIONS IN COMPUTE-AND-FORWARD RELAY NETWORKS

5.1 Compute-and-Forward

Of central importance in designing reliable communication schemes over wireless relay networks is the question of how to deal with the interference from other source nodes and the additive noise at the destination nodes. In the aforementioned Decode-and-Forward scheme, the relays can completely remove noise by decoding the original codewords before forwarding them to the destination. However, this strategy still suffers from interference if there are multiple transmitters in the network. Another approach is to try to use the interference-reducing techniques available in MIMO channels, in which the interactions between interference signals can be exploited. Amplify-and-Forward and Compress-and-Forward are two schemes that fall into this category. Noise is not removed in the Amplify-and-Forward scheme, and is just removed partially in the Compress-and-Forward scheme. Hence, it can be amplified and accumulated at the destination.

The natural question is how to handle both interference and noise, and one ingenious answer came from Nazer et al. in [20]. The idea is based on the network coding principle, that means the relays try to recover a noiseless linear function of the codewords sent by the source nodes, instead of decoding each codeword separately. In this way, we can not only harness the interactions introduced by the channel if the computed functions are well designed, but also remove noise completely at the relays. Finally, the decoded functions are combined and inverted at the destination to recover the original messages. This strategy is called Compute-and-Forward, or physical-layer network coding, or analog network coding, because of its similarity to network coding.

It's required that the linear functions of the transmitted codewords that the relays are going to decode must also be a codeword. Hence, a structured coding must be used instead of random coding. In [20], Nazer proposed the nested lattice codes, which have the desired property mentioned above. A lattice code consists of a lattice and a shaping domain. A lattice is a set of linear combinations of a finite number of independent vectors, with integer coefficients. There are an infinite number of lattice points. However, the set of lattice points which stay inside a bounded shaping region forms a lattice codebook. In particular, if the shaping region is created from another lattice, then the resulting code is call nested lattice code.

The maximum achievable capacity of Compute-and-Forward is also derived in [20]. However, there is a technical difficulty when implementing this relaying scheme. The lattice property holds only for integer linear combinations of codewords, while the functions computed by the channel have real (or complex) coefficients in general. When the relays try to decode these functions, some quantization errors can be introduced. To overcome this problem, it's suggested to scale the received signal so that it's close to an integer combination. The larger the scale factor is, the smaller quantization error that the system suffers. But a larger scale factor also results in the amplification of noise, which degrades the performance of the system. This tradeoff between small quantization error and large noise amplification raises a question of how to optimize the scale factors. In practice, the scale factors depend on the transmit power of the source nodes. Hence, we can solve an optimization problem of power allocation in networks and the integer coefficients of the computational functions to get the maximum achievable rate. This is the scope of this chapter.

In this chapter, we consider the typical Compute-and-Forward scheme that is described in [20] with K source nodes and K relay nodes. An algorithm to optimize the power allocation and the integer coefficients is proposed. The remainder of this chapter is organized as follows. Section 5.2 summarizes the background theory related to lattice coding. The target system model, as well as the corresponding optimization problem, is described in Section 5.3. In Section 5.4, an algorithm for solving that optimization problem is proposed. The numerical results to support the analysis are presented in Section 5.5. Finally, Section 5.6 contains concluding remarks.

5.2 Lattices and Lattice Coding

As mentioned above, interesting links were found recently between lattices and coding schemes for wireless relay networks, especially for the Compute-and-Forward strategy. Lattice codes can help to achieve the capacity of Gaussian point-to-point channels as shown in [45]. Good lattices tend to be "perfect" in all aspects as the dimension goes to infinity [46]. In this section, we introduce some basic definitions and main figures of merit of lattices for the further study of lattices in the area of Gaussian network information theory.

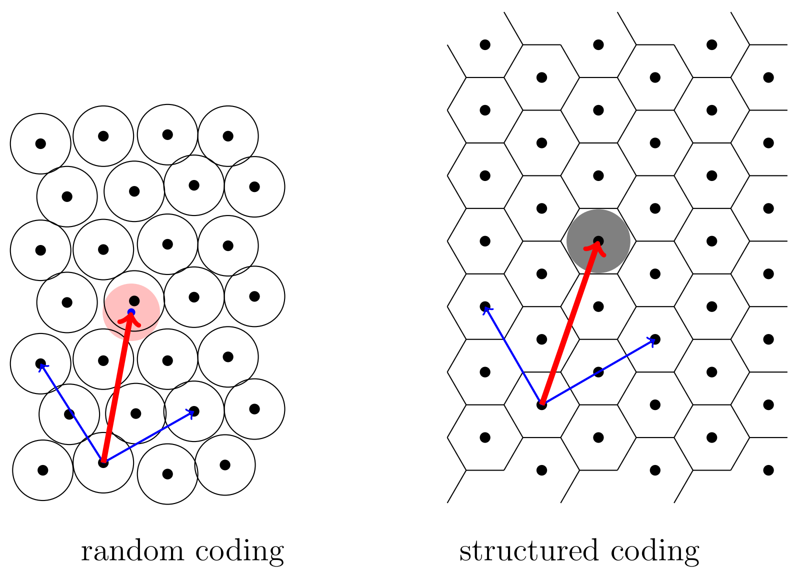

Fig. 5.1. Random coding v.s. structured coding

Definition 5.2.1 *An n-dimensional lattice $\mathcal{C}$ is defined by a set of n basis vectors $g_1, g_2, ..., g_n \in \mathbb{R}^n$. The lattice $\mathcal{C}$ is composed of all integral combinations of the basis vectors, i.e.,*

$$\Lambda = \{\lambda = \mathbf{G}\mathbf{i} : \mathbf{i} \in \mathbb{Z}^n\} \tag{5.1}$$

where $\mathbb{Z} = \{0, \pm1, \pm2, ...\}$ and the $n \times n$ generator matrix $\mathbf{G}$ is given by $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_n]$.

Definition 5.2.2 *(Nearest neighbor quantizer and Voronoi region). The nearest neighbor quantizer $Q(\cdot)$ associated with $\mathcal{C}$ is defined by*

$$Q(\mathbf{x}) = \arg\min_{\boldsymbol{\lambda} \in \Lambda} \|\mathbf{x} - \boldsymbol{\lambda}\| \tag{5.2}$$

where $\|\cdot\|$ denotes the Euclidean norm. The fundamental Voronoi region of Λ, denoted by $\mathcal{V}$, is a set of points in $\mathbb{R}^n$ closest to the zero codeword, i.e., $\mathcal{V}_0 = \{\mathbf{x} : Q(\mathbf{x}) = \mathbf{0}\}$. The Voronoi region associated with each $\boldsymbol{\lambda} \in \Lambda$ is the set of points $\mathbf{x}$ such that $Q(\mathbf{x}) = \boldsymbol{\lambda}$.

According to the definition of the Voronoi region, every $\mathbf{x} \in \mathbb{R}^n$ can be uniquely expressed as $\mathbf{x} = \boldsymbol{\lambda} + \mathbf{r}$ with $\boldsymbol{\lambda} \in \Lambda, \mathbf{r} \in \mathcal{V}$.

Definition 5.2.3 *(Modulo lattice operation). The modulo lattice operation with respect to a lattice Λ is defined as,*

$$\mathbf{x}\, mod\, \Lambda = \mathbf{x} - Q(\mathbf{x}) \tag{5.3}$$

Definition 5.2.4 *(Second moment of a lattice). The second moment σ_Λ^2 of the lattice Λ is defined as*

$$\sigma_\Lambda^2 = \frac{1}{n} E\|\mathbf{U}\|^2 = \frac{1}{n} . \frac{\int_\mathcal{V} \|\mathbf{x}\|^2 d\mathbf{x}}{V} \tag{5.4}$$

where $\mathbf{U}$ is a random vector uniformly distributed over $\mathcal{V}$ and $V \triangleq V(\Lambda) = |\mathcal{V}|$.

Definition 5.2.5 *(Normalized second moment). The normalized second moment of Λ is defined as*

$$G(\Lambda) V^{2/n} = \sigma_\Lambda^2 1n . \sigma_\Lambda^2 \int_\mathcal{V} \|\mathbf{x}\|^2 d\mathbf{x} V^{1+2/n} \tag{5.5}$$

The minimum possible value of $G(\Lambda_n)$ over all lattices in $\mathbb{R}^n$ is denoted by G_n. It's shown by Poltyrev [47] that $G_n \geq G_n^* > \frac{1}{2\pi e}$, where G_n^* is the normalized second moment of an n-dimensional sphere and $\frac{1}{2\pi e}$ is the normalized second moment of an infinite-dimensional sphere. A result in [47] states that there exists a sequence of lattices Λ_n with

$$\lim_{n\to\infty} G_n = \frac{1}{2\pi e}$$

i.e., there exists a sequence of "good" lattices Λ_n^* whose Voronoi region $\mathcal{V}$ approaches a sphere in the sense that $G(\Lambda_n^*) = G_n \to G_n^* \to \frac{1}{2\pi e}$ as $n \to \infty$. We say that such lattices are good for quantization [48].

Definition 5.2.6 *(Nested Lattices, nesting ratio and coset leaders)*

- *A pair of n-dimensional lattices* (Λ_1, Λ_2) *is called nested if* $\Lambda_2 \subset \Lambda_1$, *i.e., there exists corresponding generator matrices* $\mathbf{G}_1$ *and* $\mathbf{G}_2$ *such that* $\mathbf{G}_2 = \mathbf{G}_1.\mathbf{J}$, *where* $\mathbf{J}$ *is an* $n \times n$ *integer matrix whose determinant is greater than one.*

- *Denote the Voronoi regions of* Λ_1 *and* Λ_2 *as* $\mathcal{V}_1$ *and* $\mathcal{V}_2$, *and their volumes as* V_1 *and* V_2, *respectively. The nesting ratio of the above nested lattices is defined as*

$$\sqrt[n]{det\mathbf{J}} = \sqrt[n]{\frac{V_2}{V_1}}$$

- *The points of the set* $C = \Lambda_1 \, mod \, \Lambda_2 \triangleq \Lambda_1 \bigcap \mathcal{V}_2$ *are called the coset leaders of* Λ_2 *relative to* Λ_1.

Roughly speaking, a nested lattice code is a lattice code whose bounding region is the Voronoi region of a sublattice.

Definition 5.2.7 *The coding rate of the nested lattice code is defined as*

$$R = \frac{1}{n} log \|\mathcal{C}\| = \frac{1}{n} log \frac{V_2}{V_1}. \tag{5.6}$$

Finally, we present some properties of the modulo operation.

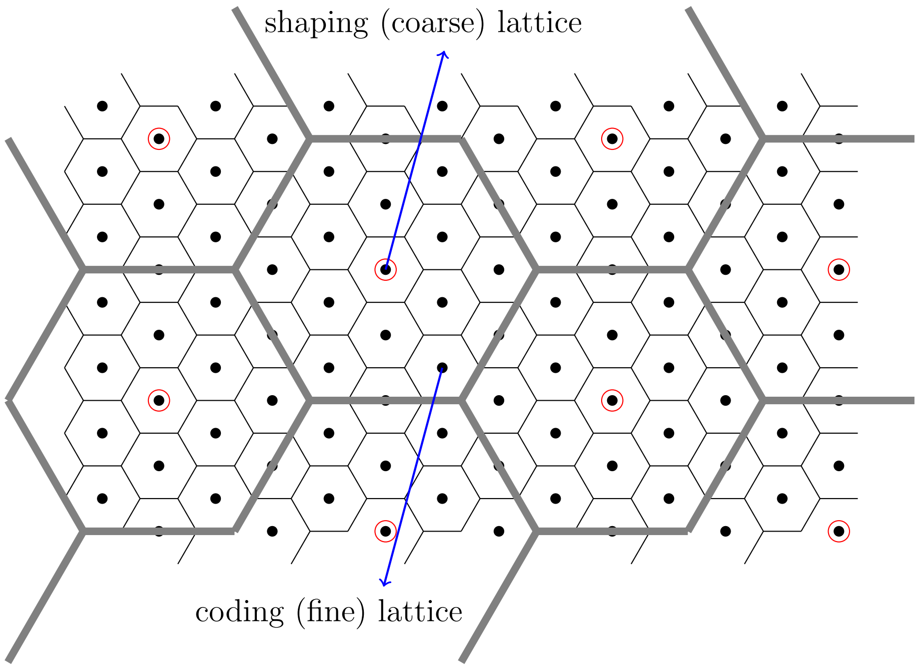

Fig. 5.2. Nested lattice code

Lemma 5.2.1

$$
\begin{aligned}
&(i) \quad (\mathbf{x}\, mod\, \Lambda + \mathbf{y})\, mod\, \Lambda = (\mathbf{x} + \mathbf{y})\, mod\, \Lambda, \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n. \\
&(ii) \quad [k(\mathbf{x}\, mod\, \Lambda)]\, mod\, \Lambda = (k\mathbf{x})\, mod\, \Lambda, \forall k \in \mathbb{Z}, \mathbf{x} \in \mathbb{R}^n. \\
&(iii) \quad \gamma(\mathbf{x}\, mod\, \Lambda) = (\gamma \mathbf{x})\, mod\, (\gamma\Lambda), \forall \gamma \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n.
\end{aligned} \tag{5.7}
$$

5.3 System Model

A typical communication scenario where the Compute-and-Forward strategy with nested lattice codes has proven beneficial is described in Fig. 5.2. In this model, there are K transmitters which are sending K independent messages to a destination with the support of K relays. Each relay tries to decode an integer linear function of the transmitted messages, and then forwards the decoded function to the destination via an error-free channel. The destination will combine all those functions and converted them to K original messages separately.

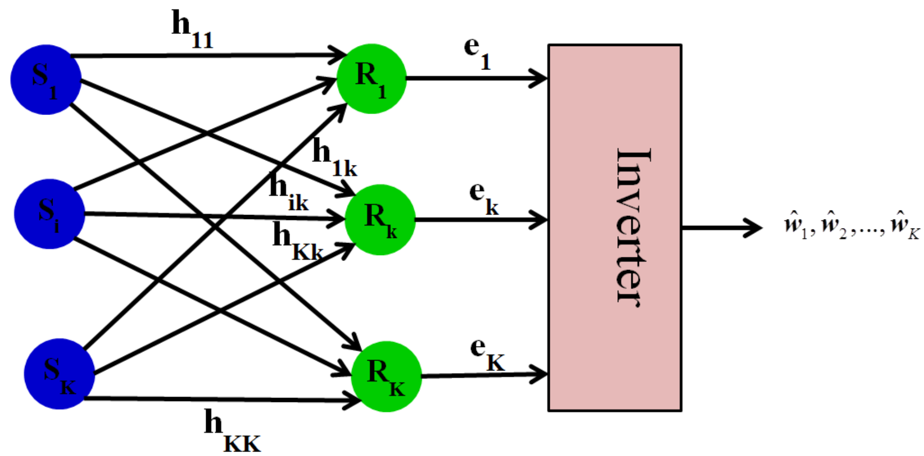

Fig. 5.3. System model

Suppose that we have an n-dimensional lattice Λ with second moment $\sigma_\Lambda^2 = P$ and the Voronoi region $\mathfrak{V}$. We choose a fine lattice Λ_C such that the codebook $\mathfrak{C} = \Lambda_C \bigcap \mathfrak{V}$ consists of 2^{nR} codewords. Let $P_{i,max}$ be the maximum power that the i^{th} transmitter can transmit. Assume that the i^{th} transmitter wants to transmit the codeword $\mathbf{w}_i$ to the destination.

Encoding: To transmit the message $\mathbf{w}_i$, the i^{th} transmitter sends the following signal

$$\mathbf{x}_i = \sqrt(P_i/P)([\mathbf{w}_i - \mathbf{u}_i]\, mod\, \Lambda) \tag{5.8}$$

where P_i is the transmitted power of the i^{th} transmitter and $\mathbf{u}_i$ is a dithering vector, which is uniformly distributed in the Voronoi region of Λ and independent of $\mathbf{w}_i$. It has been shown [49] that $\mathbf{x}_i$ is also independent of $\mathbf{w}_i$ and uniformly distributed over the Voronoi region of Λ.

The received signal at the k^{th} relay is given by

$$\mathbf{y}_k = \sum_{i=1}^{K} h_{ik}\mathbf{x}_i + \mathbf{n}_k \tag{5.9}$$

Decoding: the k^{th} relay tries to decode correctly the following integer combination $\mathbf{e}_i = \sum_{i=1}^{K} m_{ik}\mathbf{w}_i$ from the received signal, where m_{ik} are some integers predesigned by the relay. To do this, the relay computes the following quantity

$$\begin{aligned}
\mathbf{y}_k^{'} &= (\alpha_k \mathbf{y}_k + \sum_{i=1}^{K} m_{ik}\mathbf{u}_i)\, mod\, \Lambda \\
&= [\alpha_k(\sum_{i=1}^{K} h_{ik}\mathbf{x}_i + \mathbf{n}_k) + \sum_{i=1}^{K} (\frac{m_{ik}}{\sqrt{P_i/P}}\mathbf{x}_i - \frac{m_{ik}}{\sqrt{P_i/P}}\mathbf{x}_i + m_{ik}\mathbf{u}_i)]\, mod\, \Lambda \\
&= \left[\sum_{i=1}^{K}\left(\alpha_k h_{ik} - \frac{m_{ik}}{\sqrt{P_i/P}}\right)\mathbf{x}_i + \sum_{i=1}^{K}([m_{ik}(\mathbf{w}_i - \mathbf{u}_i)\, mod\, \Lambda] + m_{ik}\mathbf{u}_i) + \alpha_k \mathbf{n}_k\right] mod\, \Lambda \\
&= \left[\sum_{i=1}^{K} m_{ik}\mathbf{w}_i + \sum_{i=1}^{K}\left(\alpha_k h_{ik} - \frac{m_{ik}}{\sqrt{P_i/P}}\right)\mathbf{x}_i + \alpha_k \mathbf{n}_k\right] mod\, \Lambda \\
&= \left(\sum_{i=1}^{K} m_{ik}\mathbf{w}_i + \mathbf{n}_k^{'}\right) mod\, \Lambda
\end{aligned} \tag{5.10}$$

where $\alpha_k \in \mathbb{R}$ is some factor which is determined later. Here, we use the properties (i) and (ii) in (5.7) to do the algebra. The vector $\mathbf{n}_k^{'}$ is the equivalent noise, which is given from (5.10).

$$\mathbf{n}_k^{'} \triangleq \left[\alpha_k \mathbf{n}_k + \sum_{i=1}^{K}\left(\alpha_k h_{ik} - \frac{m_i}{\sqrt{P_i/P}}\right)\mathbf{x}_i\right] mod\, \Lambda \tag{5.11}$$

The relay now can decode the lattice point $\mathbf{e}_k = \sum_{i=1}^{K} m_{ik}\mathbf{x}_i$ from $\mathbf{y}_k^{'}$. Finally, the destination collects all decoded functions $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_K$ from the relays and solves for the original codewords $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K$ from the following equation

$$\mathbf{e} = \mathbf{M}.\mathbf{w} \tag{5.12}$$

where $\mathbf{e} = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_K]^T, \mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K]^T$ and $\mathbf{M} = [m_{ik}], 1 \leq i \leq K, 1 \leq k \leq K$.

5.3.1 Rate Analysis

The following theorem has been introduced in [20].

Theorem 5.3.1 *For real-valued AWGN networks with channel coefficient vectors* $\mathbf{h}_m \in \mathbb{R}^K$ *and equation coefficient vectors* $\mathbf{a}_m \in \mathbb{Z}^K$*, the following computation rate is achievable at the* m^{th} *relay node*

$$\mathcal{R}(\mathbf{h}_m, \mathbf{a}_m) = \max_{\alpha_m \in \mathbb{R}} \left[\frac{1}{2} log^+ \left(\frac{P}{\alpha_m^2 + P\|\alpha_m \mathbf{h}_m - \mathbf{a}_m\|^2} \right) \right] \tag{5.13}$$

This is maximized by choosing α_m *to be the MMSE coefficient* $\alpha_m = \frac{P\mathbf{h}_m^T \mathbf{a}_m}{1+P\|\mathbf{h}_m\|^2}$*, which results in a computation rate of*

$$\mathcal{R}(\mathbf{h}_m, \mathbf{a}_m) = \frac{1}{2} \overset{+}{\log} \left[\left(\|\mathbf{a}_m\|^2 - \frac{P(\mathbf{h}_m^T \mathbf{a}_m)^2}{1 + P\|\mathbf{h}_m\|^2} \right)^{-1} \right] \tag{5.14}$$

Proof. See [20]. ■

In the above theorem, the equal power constraint $\|\mathbf{x}_m\|^2 \leq nP$ is assumed. Now, suppose that each source node can be assigned a maximum transmit power given by $P_{i,max}$. We can incorporate this asymmetric power constraint by scaling the channel coefficients properly. Thus, by applying the Theorem 5.3.1 for each relay and adding them together, the sum rate that can be achieved for the considered network can be found as

$$SR = \frac{1}{2} \sum_{k=1}^{K} \overset{+}{\log} \left[\left(\|\mathbf{m}_k\|^2 - \frac{P((\beta \circ \mathbf{h}_k)^T \mathbf{m}_k)^2}{N_k + P\|(\beta \circ \mathbf{h}_k)\|^2} \right)^{-1} \right] \tag{5.15}$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_K]^T$.

5.3.2 Problem Formulation

The joint optimization of power allocation and integer coefficients can be stated as

Problem 2

$$\begin{aligned} \text{maximize} \quad & SR = \frac{1}{2} \sum_{k=1}^{K} \overset{+}{\log} \left[\left(\|\mathbf{m}_k\|^2 - \frac{P((\boldsymbol{\beta} \circ \mathbf{h}_k)^T \mathbf{m}_k)^2}{N_k + P\|(\boldsymbol{\beta} \circ \mathbf{h}_k)\|^2} \right)^{-1} \right] \\ \text{subject to} \quad & 0 \leq \beta_k \leq \sqrt{(P_{k,max}/P)}, \quad for\, k = 1, 2, ..., K \\ \text{and} \quad & |det(\mathbf{m})| \geq 1 \quad and \quad \mathbf{m} \in \mathbb{Z}^{k \times k} \quad \text{for}\, k = 1, 2, ..., K \end{aligned} \tag{5.16}$$

This is an Mixed Integer Nonlinear Programming (MINLP) problem. Moreover, this problem is non-convex. The methods to solve this kind of problems have been introduced in [50]. However, the complexity of the algorithms is quite high and needs to be considered carefully. Another implementable approach is to solve this problem iteratively. If we fix the power factors $\boldsymbol{\beta}$ and try to optimize the integer coefficients $\mathbf{m}$, we get a Mixed Integer Quadratic Programming (MIQP) problem. If we fix the integer coefficients $\mathbf{m}$, then the power allocation can be optimized by using the Lagrange method, although it's a non-convex problem. In the next section, the algorithms for the sum rate optimization problem will be analyzed.

5.4 Optimization of Compute-and-Forward Relay Network

5.4.1 Iterative Solution

By observing the objective function, we can easily see that the integer coefficient vector $\mathbf{m}_k$ only involves in the k^{th} term of the objective function (that is the maximum rate at the k^{th} relay). Hence, if the power factor $\boldsymbol{\beta}$ is fixed, then the optimization of the integer coefficients can be done separately at each relay. For instance, at the k^{th} relay, we have the following problem to solve

$$\begin{aligned} &maximize \quad R_k = \overset{+}{\log}\left[\left(\|\mathbf{m}_k\|^2 - \frac{P((\mathbf{c}_k)^T\mathbf{m}_k)^2}{N_k + P\|(\mathbf{c}_k)\|^2}\right)^{-1}\right] \\ &subject\,to \quad \mathbf{m}_k \in \mathbb{Z}^k \end{aligned} \tag{5.17}$$

where $\mathbf{c}_k = \boldsymbol{\beta} \circ \mathbf{h}_k$ is a constant vector.

Now, assume that we can obtain the optimal solution f_k^* for Problem (5.17). The maximum sum rate of the network can be found from the following problem

$$\begin{aligned} &\text{maximize} \qquad SR = \frac{1}{2}\sum_{k=1}^{K}\overset{+}{\log}\left[\left(\|\mathbf{m}_k\|^2 - \frac{P((\boldsymbol{\beta}\circ\mathbf{h}_k)^T\mathbf{m}_k)^2}{N_k + P\|(\boldsymbol{\beta}\circ\mathbf{h}_k)\|^2}\right)^{-1}\right] \\ &\text{subject to} \qquad 0 \le \beta_k \le \sqrt{(P_{k,max}/P)}, \quad for\, k = 1, 2, ..., K \end{aligned} \tag{5.18}$$

Problem (5.17) can be reduced to a MIQP problem because the logarithm function is an increasing function. In fact, Problem (5.17) can be stated equivalently as

$$\begin{aligned} \text{minimize} \quad & g_k = N_k\|\mathbf{m}_k\|^2 + P\|\mathbf{c}_k\|^2\|\mathbf{m}_k\|^2 - P(\mathbf{c}_k^T\mathbf{m}_k)^2 \\ \text{subject to} \quad & \mathbf{m}_k \in \mathbb{Z}^k \quad \textit{and} \quad 0 < g_k < N_k + P\|\mathbf{c}_k\|^2 \end{aligned} \tag{5.19}$$

Problem (5.18) is a non-convex and nonlinear optimization problem. However, we are able to solve it by global optimization theory, which is shown later.

a. Integer coefficients sub-problem

Let's solve the Problem (5.17). First, we notice that the first inequality constraint of (5.19) is always satisfied because $(\mathbf{c}_k^T\mathbf{m}_k)^2 \leq \|\mathbf{c}_k\|^2\|\mathbf{m}_k\|^2$ (by Schwartz's inequality). The second inequality constraint of (5.19) is to guarantee that the expression inside the $log^+(\cdot)$ function is greater than 1. If there is no feasible $\mathbf{m}_k$ for this constraint, then the optimal value of the Problem (5.17) will be zero and $\mathbf{m}_k$ can be any integer vector.

Because $g_k \geq 0$, $\forall \mathbf{m}_k$, the quadratic form g_k is semi-definite, so the Problem (5.19) can be solved easily by the well-known methods for MIQP such as cutting plane or branch-and-bound. For example, we can use branch-and-bound algorithm in [51].

b. Power allocation sub-problem

Consider the power allocation subproblem (5.18).

$$\begin{aligned} \text{maximize} \quad & SR = \frac{1}{2}\sum_{k=1}^{K} \overset{+}{\log}\left[\left(\|\mathbf{m}_k\|^2 - \frac{P((\circ\mathbf{h}_k)^T\mathbf{m}_k)^2}{N_k + P\|(\boldsymbol{\beta}\circ\mathbf{h}_k)\|^2}\right)^{-1}\right] \\ \text{subject to} \quad & 0 \leq \beta_k \leq \sqrt{(P_{k,max}/P)}, \quad for\, k = 1, 2, ..., K \end{aligned}$$

This problem is a nonconvex optimization problem. However, the objective function and the constraint functions are continuous and differentiable almost everywhere. Therefore, we can use one of the well-studied iterative algorithms for the nonconvex optimization problem. Fig. 5.4 shows the sum rate of our model as a function of the power allocation factors $\boldsymbol{\beta}$ where the integer coefficients $\mathbf{M}$ are fixed.

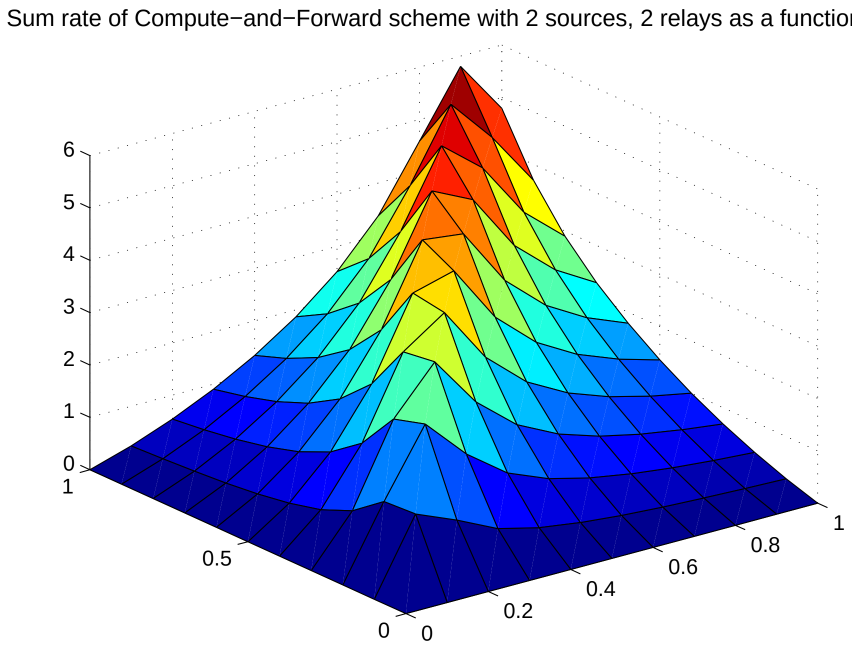

Fig. 5.4. Objective function of Problem (5.18)

5.4.2 Algorithm

In this section, we summarize the iterative algorithm to jointly optimize the power allocation and the integer coefficients of the computational functions of Compute-and-Forward relay networks. This algorithm is described as follows.

Algorithm 5.1.

1. Initialization: Set $i = 1$ *and choose the initial vector* $\boldsymbol{\beta}^{(0)}$.

2. Set $\boldsymbol{\beta} = \boldsymbol{\beta}^{(i-1)}$ *and solve the integer coefficient problem* (5.19) *for each relay* $k = 1, 2, ..., K$ *to get the solution at the* i^{th} *iteration* $\mathbf{M}^{(i)}$.

3. Set $\mathbf{M} = \mathbf{M}^{(i)}$ *and solve the power allocation problem* (5.18) *to get the solution* $\boldsymbol{\beta}^{(i)}$. *Compute the sum rate at the* i^{th} *iteration* $SR^{(i)}$.

4. Increase the iteration index $i = i + 1$ *and repeat Step 2.*

5. The algorithm will terminate whenever $\|\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}^{(i-1)}\| < \epsilon$ *and* $\|SR^{(i)} - SR^{(i-1)}\| < \epsilon_0$, *where* ϵ *and* ϵ_0 *are the predefined maximum errors.*

In most cases, this algorithm converges to the optimal solution of the problem. However, this convergence is not guaranteed in all cases. The temporary solution may bounce back and forth between some finite number of suboptimal solutions as we'll show by the numerical results.

5.5 Numerical Results

In this section, some numerical results are provided to evaluate the proposed algorithm. For simplicity, we assume that the noise signals at the receivers have the same powers $N_1 = N_2 = ... = N_K = N$. We also assume that each transmitter has a maximum transmitted power of 10dBW. The performances of the network are measured by the sum rates of all receivers at different levels of SNR.

We consider two network topologies; one with two sources and two relays and the other with 3 sources and 3 relays. The channel coefficients between the sources and the relays are modeled as independent zero-mean Gaussian random variables, whose variances depend on the strength of the corresponding links. Typically, we express the strength of each link by its Euclidean distance in the network topology. We normalize the distance so that the distance $d = 1$ corresponds to the channel gain variance $\sigma^2 = 0\,dBW$. The network topologies are shown in Fig. 5.5. As mentioned in previous sections, the full channel state information (CSI) is available in all of the simulation scenarios. To solve the MIQP sub-problem in Step 2 of the proposed algorithm, we use the TOMLAB optimization tool [52].

Fig. 5.6 compares the results when running the proposed algorithm for 2-user 2-relay networks in different channel conditions. For better channel conditions, the optimal sum rate is higher. (The blue curve corresponds to the best channel condition and the pink curve corresponds to the worst channel condition.) In this case (2 users and 2 relays), the objective function and the constraints are rational functions and polynomials of second order, so we have a quadratic rational optimization problem.

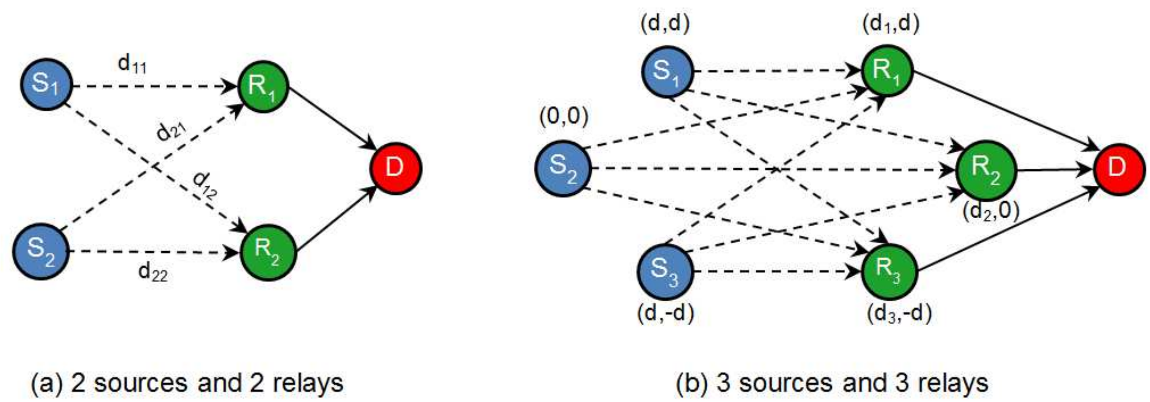

Fig. 5.5. Simulated network topology

The proposed algorithm converges very quickly to the optimal solution after only a few iterations.

Fig. 5.7 shows the respective results for 3-user, 3-relay networks in different channel conditions. Again, the numerical results are consistent with the theoretical conjecture. However, the convergence of the algorithm is not guaranteed in this case. Here, we can only get to some sub-optimal solution of the problem. Fig. 5.8 shows the sum rate of the network versus the iterations for a special case.

Finally, Fig. 5.8 compares the sum rate achieved by the Compute-and-Forward strategy with the sum rate achieved when each relay decodes the message from its corresponding source, independent of the other relays. Other source nodes are considered as interference sources. We observe that the Compute-and-Forward scheme is dominating the non-relaying scheme.

5.6 Conclusion

In this chapter, we study a multiple-access relay network, which is equipped with the Compute-and-Forward relay strategy. This strategy enables the relays to decode linear integer combinations of the codewords sent from the source nodes using the

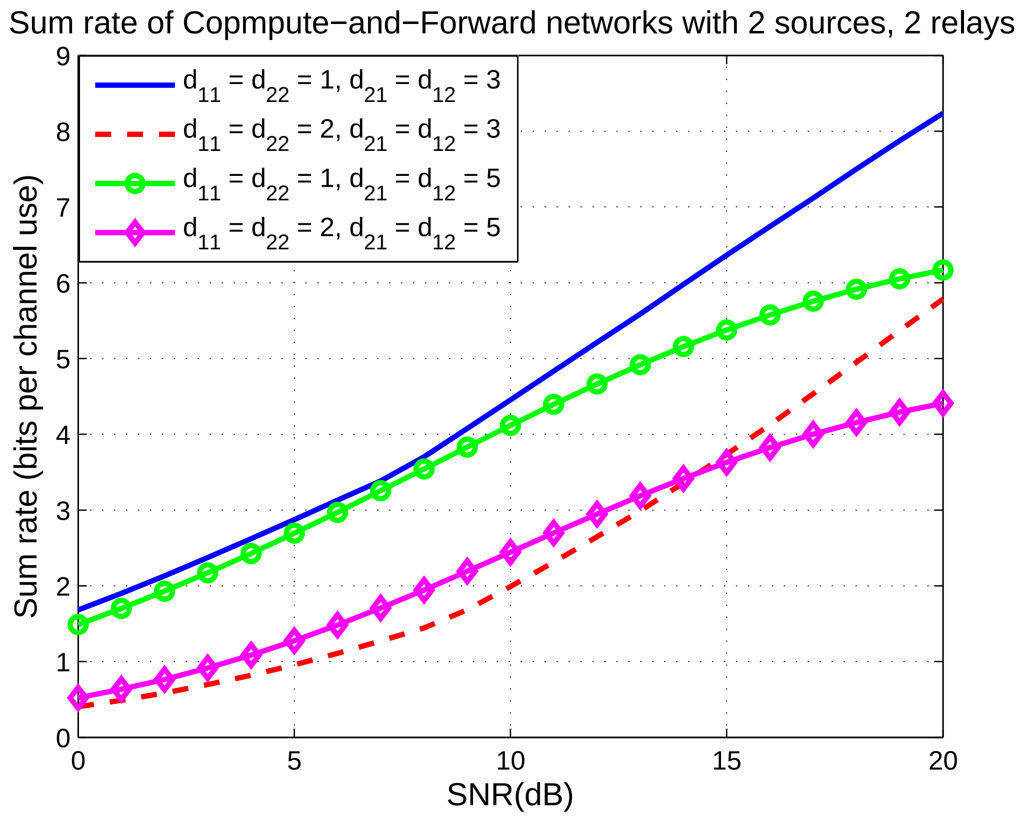

Fig. 5.6. Optimal sum rate of 2x2 Compute-and-Forward relay network

noisy linear combination provided by the channel. The destination finally collects sufficient combinations from the relays and solves a system of linear equations to recover the original messages. In the basic Compute-and-Forward strategy, the relays are free to select the linear equations they want to recover. However, because the channel coefficients are not integers, there is always a quantization error when the relays try to decode their linear integer combination. This error can be reduced if we scale the transmitted signal so that the noisy combination of them is close to a linear integer combination. The error is typically decreased when we increase the scaling factor, but doing that also degrades the performance of the network due to the amplification of noise. Moreover, we cannot increase the scaling arbitrarily because of the power limitation. To solve this problem, a joint optimization problem of the

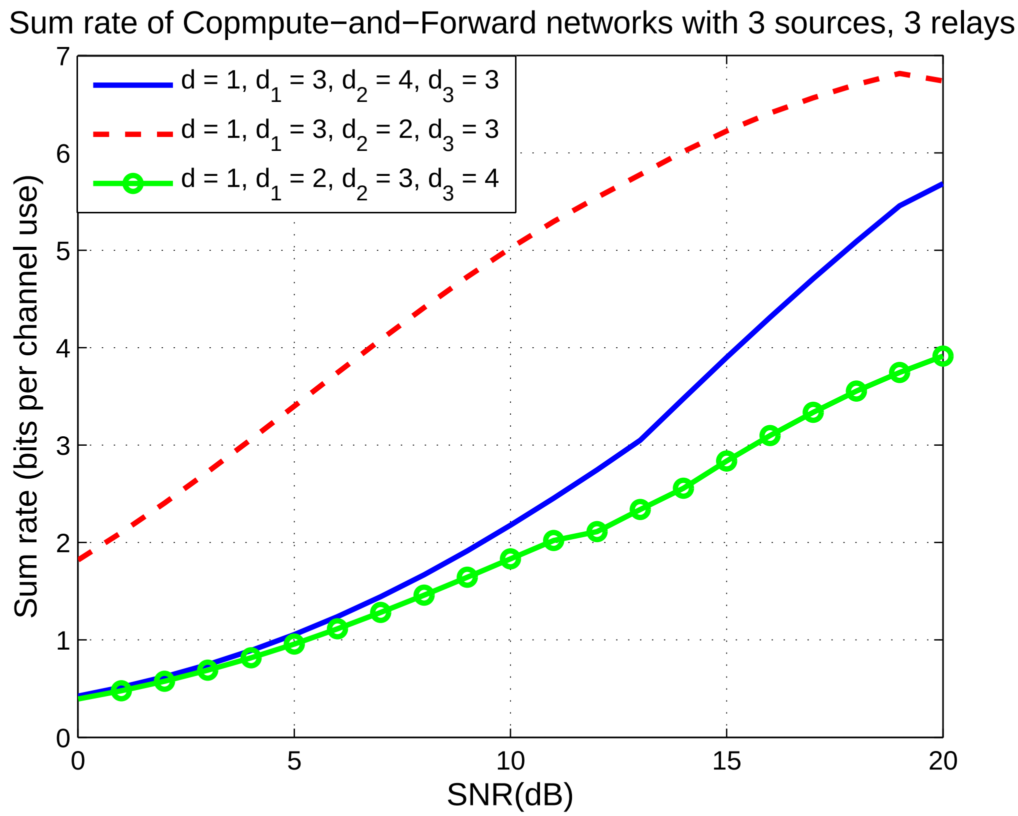

Fig. 5.7. Optimal sum rate of 3x3 Compute-and-Forward relay network

integer coefficients for the relay and power allocation for the sources is presented in this chapter. Although this optimization is nonconvex, nonlinear, and has mixed integer and non-integer variables, it can be solved iteratively and numerically. This interesting idea is also confirmed by the numerical results.

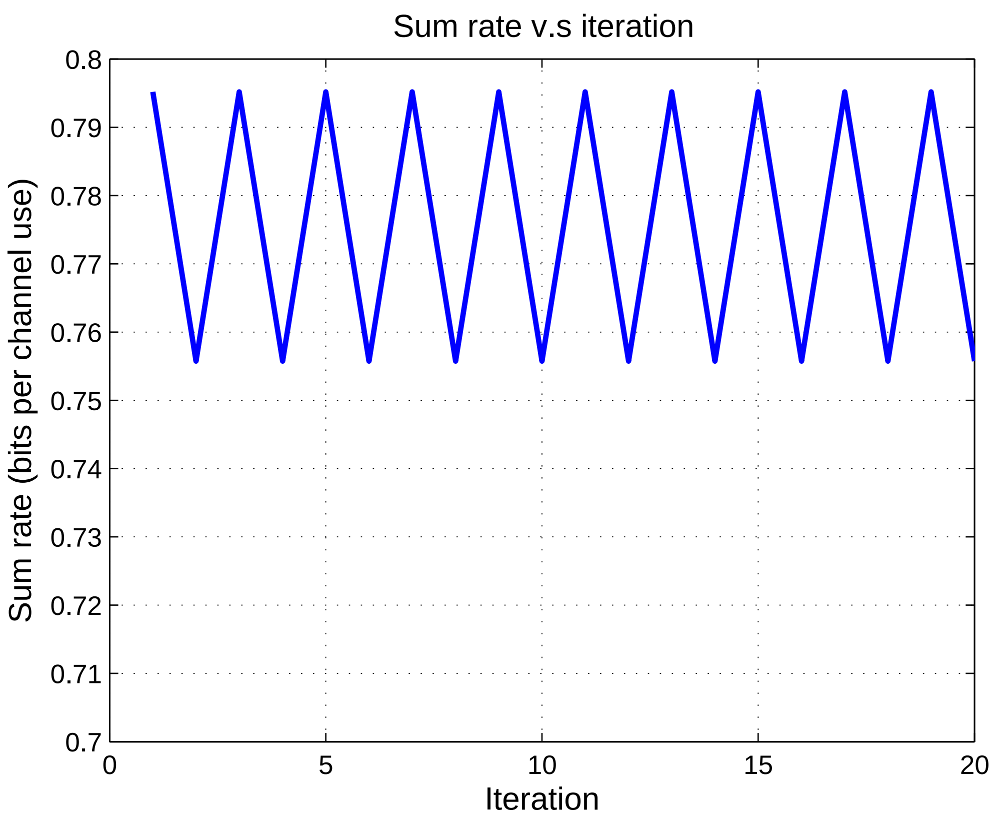

Fig. 5.8. A divergent case of the algorithm

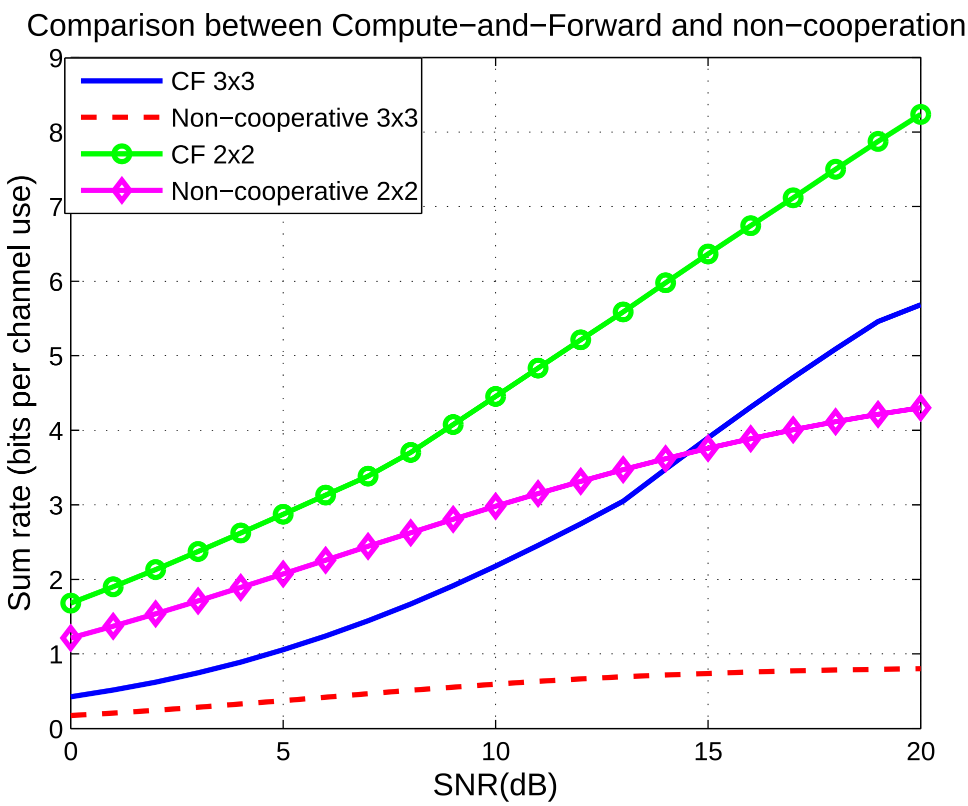

Fig. 5.9. Performance of Compute-and-Forward

6. SUMMARY AND FURTHER WORK

Minimizing the capacity loss that results from distributing the radio resources to the relay transmissions is an important and challenging problem that needs to be solved to accelerate the development of relay-assisted wireless networks. Many optimization problems in cooperative networks have been investigated to maximize the performance of the networks (in terms of capacity, throughput, BER, etc.) subject to the constraints on the resources available in the system, such as time, bandwidth, power, etc. There has been significant progress in solving these problems. This is because of the advance of convex optimization tools, which have enabled useful approaches for solving the problems.

However, while the development of optimization in relay networks has been ongoing for years, most of the optimization problems that have been solved so far have only considered the involved factors separately. In these cases, the formulated optimization problems are either a convex problem or a problem that can be converted easily to a convex problem. What is needed in current cooperative network research is an optimization involving a combination of all of the affecting factors, and this is the main purpose of this current work.

Recently, specific advances in non-convex optimization theory continue to enable new techniques which are capable of solving complicated optimization problems. This work describes one such technique and applies that to solve a bigger optimization problem for wireless cooperative networks.

6.1 Summary of the Results

The first and important result that has been achieved so far is the solution for the joint optimization of power allocation, time scheduling, and relay selection strategy, subject to the constraints of maximum transmission power and minimum required rate for each user. The main method that is used to obtain this result is the dual method for non-convex optimization in multi-carrier systems, proposed by Yu et al. [1]. By applying this result, the original problem is transfered to a dual problem, and the structure of the problem is exploited to subdivide it into smaller and simpler subproblems. Finally each subproblem is solved using the KKT conditions [37].

In addition to providing a solution to the problem, in this work a set of numerical algorithms to obtain the optimal solution are also proposed. The first algorithm is based on the assumption that the channel gain distribution is known, and we can take the expectation of the channel gains in order to compute the optimal solution. The second algorithm is designed for real-time situations, when the channel distribution is not known beforehand, but the base station can get the channel state information from the feedback channel. In this case, we can replace the expectation by the immediate value of the channel gains to obtain a real-time algorithm.

Both algorithms have been verified in terms of mathematical analysis and numerical simulation. An understanding of the convergence of these algorithms is very important for the implementation in practice. The convergence of both algorithms is proved by using the stochastic optimization theory [38]. The important result on convergence is that all the algorithms will converge with probability 1 if the adaptive step size is chosen to satisfy some basic conditions. However, the location of the convergence point will depend on the availability of the channel state information.

- If the channel distribution is known, or if the channel state information is fed back perfectly, then the proposed algorithms will converge with probability 1 to the optimal solution of the joint optimization problem.

- If the channel state information is not fed back correctly, then there are two possibilities: if the channel estimation error satisfies the unbiased condition, then the real-time algorithm still converges to the optimal solution with probability 1. However, if the channel estimation error is biased, then the convergence point is not the same as the optimal point of the problem. It's also proved that the convergence point will stay in some bounded neighborhood of the optimal point with probability 1.

Numerical results are provided to confirm the correctness of the solution and the algorithms. The simulation results show that both algorithms converge, and the duality gap between the primal and the dual problems reduces to zero when the number of subcarriers becomes large. The affect of the step size and the randomness of the channel condition on the convergence rate are also illustrated by numerical results.

The results from Chapter 3 and Chapter 4 are for the Decode-and-Forward cooperative networks only. However, it's not difficult to extend the results for other cooperative schemes like Amplify-and-Forward and Compress-and-Forward by using the same approach. However, for the new cooperative scheme such as Compute-and-Forward, we need to formulate a new problem to involve the optimization of network code design. This problem has also been introduced in Chapter 5. In Compute-and-Forward relay networks, each relay tries to decode a linear combination of the codewords sent from different sources, instead of decoding each of them separately as in the Decode-and-Forward scheme. We are interested in the joint optimization of the integer coefficients of the recovered equations at each relay as well as the power allocation to each sender so that the considered network can achieve the maximum sum rate. This is a non-convex, nonlinear, mixed integer programming problem, and it is NP hard. In this dissertation, an iterative algorithm is proposed to solve that problem, and the numerical results are also introduced to confirm the idea.

6.2 Further Work and Directions

As reported in Chapter 3 and 4, a significant progress has been made to solve joint optimization problems in wireless cooperative networks, which is the main purpose of this work. With an appropriate method for solving non-convex problems, we have obtained the solution for multiple-source, one-destination relay networks using the Decode-and-Forward strategy. We also have proposed algorithms to compute the optimal solutions and proved the convergence of the algorithms in different cases.

However, to complete this research, there are still several remaining tasks.

- The complexity of the algorithms has not been investigated sufficiently. Indeed, the complexity of the real-time algorithm still increases very fast when the number of users increases because we're still using the exhaustive search when finding the best relay strategy in each iterative step. One method to improve this drawback is to derive some basic criteria so that we can remove a lot of obviously bad strategies, and hence, reduce the total complexity.

- The error between the real optimal solution and the solution obtained from the real-time algorithm in case of imperfect CSI has not been evaluated. We already bounded the convergence point inside a neighborhood of the optimal solution; however, how large this neighborhood is, and if the error is negligible are the questions that need to be answered.

In addition to the topics mentioned above, during my work toward this dissertation, I found the following issues to be both very interesting and essential for bringing cooperative communications to reality.

- As mentioned at the beginning of this report, the most recent approach for designing cooperation schemes in relay networks that has attracted the most interest from researchers is Compute-and-Forward, which exploits the interference in multi-user communications by using structured codes. Actually, the class of structured codes that were used in [20] had been introduced before and its capability to achieve the Shannon limit had also been shown in [45], where the codes are called lattice codes. The challenge is how to design practical lattice codes that are not too complicated for computation, but which still approach the capacity limit. Independent of the research on relay networks, several researchers have found some interesting results about lattice codes. Specifically, a practical type of lattice codes has been proposed by N. Sommer, M. Feder and O. Shalvi based on the similarity to LDPC codes, called LDLC (Low Density Lattice Codes) [49]. The convergence analysis and some efficient methods for decoding LDLC has also been given in [53] - [54]. Unfortunately, so far there are no results on the connection between LDLC and cooperative communication networks. A problem of designing practical LDLC or something similar is an interesting direction, and optimization theory may be useful here.

- Another interesting optimization problem in relay networks is to minimize the error of relay channel estimation. In relay channel estimation problems, we design an appropriate training data sequence and allocate the transmission power to the source node and relay nodes. Then the training data from the source node is sent to the destination node with the assistance of one or more relay nodes for which we want to estimate the channel conditions. By observing the received data from source nodes and relay nodes and knowing the transmitted data, the destination nodes can estimate the channel coefficients. This is a least-square problem and it turns out that the MSE (mean-squared error) of the estimation will depend on the designed sequence and the power allocation. By optimizing the training sequence and the power allocation, we can minimize the MSE of the estimation.

LIST OF REFERENCES

LIST OF REFERENCES

[1] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *Communications, IEEE Transactions on*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

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

[3] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *Information Theory, IEEE Transactions on*, vol. 25, no. 5, pp. 572 – 584, Sep. 1979.

[4] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. System description," *Communications, IEEE Transactions on*, vol. 51, no. 11, pp. 1927 – 1938, Nov. 2003.

[5] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part II. Implementation aspects and performance analysis," *Communications, IEEE Transactions on*, vol. 51, no. 11, pp. 1939 – 1948, Nov. 2003.

[6] J. Laneman, G. Wornell, and D. Tse, "An efficient protocol for realizing cooperative diversity in wireless networks," in *Information Theory, 2001. Proceedings. 2001 IEEE International Symposium on*, p. 294, 2001.

[7] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *Information Theory, IEEE Transactions on*, vol. 50, no. 12, pp. 3062 – 3080, Dec. 2004.

[8] T. Hunter and A. Nosratinia, "Diversity through coded cooperation," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 2, pp. 283 – 289, Feb. 2006.

[9] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *Communications Magazine, IEEE*, vol. 42, no. 10, pp. 74 – 80, Oct. 2004.

[10] W. S. K.J. Ray Liu, Ahmed K. Sadek and A. Kwasinski, *Cooperative Communications and Networking.* Cambridge University Press, 2009.

[11] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *Information Theory, IEEE Transactions on*, vol. 51, no. 9, pp. 3037 – 3063, Sep. 2005.

[12] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd. Int. Symp. Information Theory (Tsahkadsor, Armenian S.S.R.)*, pp. 23–52, Publishing House of the Hungarian Academy of Sciences, 1971.

[13] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Prob.*, vol. 2, no. 5, pp. 805–814, Oct. 1974.

[14] E. C. V. D. Meulen, "Three-terminal communication channels," *Advances in Applied Probability*, vol. 3, no. 1, pp. 120–154, Spring 1971.

[15] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *Information Theory, IEEE Transactions on*, vol. 46, no. 4, pp. 1204 –1216, Jul. 2000.

[16] R. G. Gallager, "Low-density parity-check codes," 1963.

[17] C. Li, G. Yue, M. Khojastepour, X. Wang, and M. Madihian, "LDPC-coded cooperative relay systems: performance analysis and code design," *Communications, IEEE Transactions on*, vol. 56, no. 3, pp. 485 –496, Mar. 2008.

[18] P. Razaghi and W. Yu, "Parity forwarding for multiple-relay networks," *Information Theory, IEEE Transactions on*, vol. 55, no. 1, pp. 158 –173, Jan. 2009.

[19] P. Razaghi and W. Yu, "Bilayer low-density parity-check codes for Decode-and-Forward in relay channels," *Information Theory, IEEE Transactions on*, vol. 53, no. 10, pp. 3723 –3739, Oct. 2007.

[20] B. Nazer and M. Gastpar, "Compute-and-Forward: Harnessing interference through structured codes," *Information Theory, IEEE Transactions on*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[21] B. Nazer and M. Gastpar, "Compute-and-Forward: A novel strategy for cooperative networks," in *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, pp. 69 –73, Oct. 2008.

[22] K. Phan, T. Le-Ngoc, S. Vorobyov, and C. Tellambura, "Power allocation in wireless relay networks: A geometric programming-based approach," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pp. 1 –5, Nov. 30 2008-Dec. 4 2008.

[23] K. Phan, L. B. Le, S. Vorobyov, and T. Le-Ngoc, "Centralized and distributed power allocation in multi-user wireless relay networks," in *Communications, 2009. ICC '09. IEEE International Conference on*, pp. 1 –5, Jun. 2009.

[24] W. Wang, L. Ong, and M. Motani, "Transmission schedule optimization for half-duplex multiple-relay networks," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2009. WiOPT 2009. 7th International Symposium on*, pp. 1 –9, June 2009.

[25] S. H. Kim, H. Jin, and D. K. Sung, "Optimizing the persistent scheduling in two-hop relay networks," in *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*, pp. 138 –143, Mar. 2011.

[26] Z. Ding and K. Leung, "Cross-layer routing optimization for wireless networks with cooperative diversity," in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pp. 1 –5, Sep. 2008.

[27] U. Niesen and P. Whiting, "The degrees of freedom of Compute-and-Forward," *Information Theory, IEEE Transactions on*, vol. 58, no. 8, pp. 5214–5232, Aug. 2012.

[28] M. E. Soussi, A. Zaidi, and L. Vandendorpe, "Compute-and-Forward on a multiaccess relay channel: Coding and sum-rate optimization," *CoRR*, vol. abs/1209.4483, 2012.

[29] C. E. Shannon, "A mathematical theory of communication," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, Jan. 2001.

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley and Sons, Incorporated, 1991.

[31] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: the relay case," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, pp. 1577 – 1586, 2002.

[32] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Information Theory, 2000. Proceedings. IEEE International Symposium on*, p. 22, 2000.

[33] A. Reznik, S. R. Kulkarni, and S. Verd, "Degraded Gaussian multirelay channel: Capacity and optimal power allocation," *IEEE Transactions on Information Theory*, vol. 50, pp. 3037–3046, Dec. 2004.

[34] P. Gupta and P. Kumar, "Towards an information theory of large networks: an achievable rate region," *Information Theory, IEEE Transactions on*, vol. 49, no. 8, pp. 1877 – 1894, Aug. 2003.

[35] G. Kramer, I. Marić, and R. D. Yates, "Cooperative communications," *Found. Trends Netw.*, vol. 1, no. 3, pp. 271–425, Aug. 2006.

[36] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory.*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

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

[38] D. Bertsekas, *Nonlinear Programming.* Athena Scientific, 1996.

[39] X. Guo and X.-G. Xia, "Distributed linear convolutive space-time codes for asynchronous cooperative communication networks," in *Communications, 2008. ICC '08. IEEE International Conference on*, pp. 1134 –1138, May 2008.

[40] L. Ljung, G. Pflug, and H. Walk, *Stochastic approximation and optimization of random systems.* Basel, Switzerland, Switzerland: Birkhauser Verlag, 1992.

[41] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming.* Athena Scientific, 1st ed., 1996.

[42] P. Tran and J. Lehnert, "Joint optimization of relay selection and power allocation in cooperative OFDM networks with imperfect channel estimation," in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, pp. 2076 –2080, Apr. 2012.

[43] P. Tran and J. Lehnert, "Joint optimization of power allocation and cooperation in wireless OFDM networks," in *Advanced Technologies for Communications, 2009. ATC '09. International Conference on*, pp. 289 –294, Oct. 2009.

[44] T. C.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 2, pp. 328 –339, Feb. 2007.

[45] U. Erez and R. Zamir, "Achieving 1/2 log (1+SNR) on the AWGN channel with lattice encoding and decoding," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2293 – 2314, Oct. 2004.

[46] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *Information Theory, IEEE Transactions on*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005.

[47] G. Poltyrev, "On coding without restrictions for the AWGN channel," *Information Theory, IEEE Transactions on*, vol. 40, no. 2, pp. 409–417, Mar. 1994.

[48] R. Zamir and M. Feder, "On lattice quantization noise," *Information Theory, IEEE Transactions on*, vol. 42, no. 4, pp. 1152–1159, Jul. 1996.

[49] N. Sommer, M. Feder, and O. Shalvi, "Low-density lattice codes," *Information Theory, IEEE Transactions on*, vol. 54, no. 4, pp. 1561 –1585, Apr. 2008.

[50] P. Belotti, C. Kirches, S. Leyffer, J. Linderoth, J. Luedtke, and A. Mahajan, "Mixed-integer nonlinear optimization," *Acta Numerica*, vol. 22, pp. 1–131, 4 2013.

[51] E. L. Lawler and D. E. Wood, "Branch-and-Bound methods: A survey," *Operations Research*, vol. 14, no. 4, pp. 699–719, Jul. - Aug. 1966.

[52] K. Holmstrom, "TOMLAB – An environment for solving optimization problems in MATLAB," in *Proceedings for the Nordic MATLAB Conference '97*, pp. 27–28, 1997.

[53] B. Kurkoski and J. Dauwels, "Message-passing decoding of lattices using Gaussian mixtures," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pp. 2489 –2493, Jul. 2008.

[54] D. Bickson, A. Ihler, H. Avissar, and D. Dolev, "A low density lattice decoder via non-parametric belief propagation," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pp. 439–446, 2009.

VITA

VITA

Phuong Tran was born in 1979 in Ho Chi Minh City, Vietnam. He received the B.Eng. degree and M.Eng in Electrical Engineering from Ho Chi Minh University of Technology, Ho Chi Minh City, Vietnam in 2002 and 2005, respectively. In 2007, he was awarded a Fellowship from the Vietnam Education Foundation to work to obtain the PhD degree at the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN. Since 2009, he has been a Research Assistant under the supervision of Professor James S. Lehnert of the School of Electrical and Computer Engineering, Purdue University, and served as a Teaching Assistant in the same deparment. Phuong Tran is also an IEEE Student Member since 2006. His major interests are in the area of wireless communications, especially in the areas of optimization of cooperative communication systems and network information theory.