CAPACITY BOUNDS AND SCALING LAWS OF WIRELESS RELAY NETWORKS

by

Bo Wang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

ARIZONA STATE UNIVERSITY

December 2006

CAPACITY BOUNDS AND SCALING LAWS OF WIRELESS RELAY NETWORKS

by

Bo Wang

has been approved

August 2006

APPROVED:

_____, Chair

_____

_____

_____

_____

Supervisory Committee

ACCEPTED:

_____

Department Chair

_____

Dean, Division of Graduate Studies

ABSTRACT

Driven by the fast-growing demand on wireless applications, cooperative relaying has recently received much attention. In this dissertation, cooperative relaying in MIMO systems and sensor networks is studied extensively, with focus on capacity bounds and scaling laws.

The first thrust is on the capacity bounds of MIMO relay channels. First, a Gaussian MIMO relay channel model with fixed channel conditions is considered, and the corresponding upper bounds and lower bounds on the capacity are derived, which can be evaluated numerically. Next, an upper bound and a lower bound on the ergodic capacity are found when the channel links undergo Rayleigh fading. It is shown that the upper bound can meet the lower bound under certain conditions, indicating that the capacity can be characterized. Sufficient conditions for achieving the ergodic capacity are also investigated.

The second thrust considers a wireless sensory relay network model that consists of one source node, one destination node and relay nodes. Different from existing work, the focus here is on the more realistic case where the relay nodes have no *a priori* knowledge of channel state information (CSI) for both the backward channels and the forward channels. An amplify-and-forward (AF) with network training relaying scheme is devised for cooperative relaying. The achievable rates and the scaling laws are characterized accordingly, in the joint asymptotic regime of the number of relay nodes n, the channel coherence interval L, and the bandwidth W. Due to frequency-selective fading, power allocation across the frequency subbands at relay nodes plays a critical role in the achievable rates. It is revealed that the scaling law can be characterized, if L/W is bounded below and W is sub-linear in n.

The third thrust is on the tradeoff between sensing and communication in a sensor network, where all nodes may have sensed data. Power allocation policies for the sensed signal and the relayed signal are investigated for both event-driven sensor networks and sensor networks in a two-dimensional random field.

ACKNOWLEDGMENTS

I dedicate this dissertation to my parents and my sister, who have permanently supported me. Their love, devotion and patience have been vital for the completion of this dissertation. I owe them more than anyone else. I am deeply indebted to my advisor, Dr. Junshan Zhang, who continually supported my PhD study, and furthered my professional and personal development throughout the procedure of this work. He taught me the methodologies of conducting research, and enlightened me with his view on every aspect of life. He is not only a great advisor with deep vision but also and most importantly a kind person. I learned from him to believe in my work and myself. I am grateful to Dr. Antonia Papandreou-Suppappola, Dr. Darryl Morrell, Dr. Tolga Duman, Dr. Joseph Hui, and Dr. Cihan Tepedelenlioglu, for taking the time being my committee supervisors, and sharing with me their expertise. Their useful suggestions greatly helped to improve the technical content of this dissertation. I am thankful to a group of my friends at ASU, who have helped me in numerous ways to finish this dissertation. Some of them are Ming Hu, Dong Zheng, Yuming Zhu, Ning He and Qian Ma. Last, but definitely not least, I want to thank Ye Jiang, for providing me with spiritual support and accompanying me on this unforgettable journey.

TABLE OF CONTENTS

Page

Page

Page

Page

LIST OF FIGURES

CHAPTER 1

Introduction and Motivation

The age of ubiquitous communications and computing is fast approaching. Emerging wireless technologies, such as WiMAX and sensor networks, are changing the way we manage, search and access information. For example, WiMAX is expected to provide high-throughput connections, cellular backhaul, and high-speed connectivity. In particular, the use of multiple antenna at wireless transmitters and receivers has been identified as an enabling technique for high-rate broadband connections over wireless links. Along a different avenue, sensor networks have found great potential in many applications, including environmental monitoring, homeland security and battlefield wireless networks. Therefore, there is an urgent need to develop theories and methodologies that can resolve a number of open design issues and help propel significant advances in these areas.

Thus motivated, this study is targeted at understanding the fundamental limits of three kinds of wireless relay networks, including three-terminal MIMO relay channels, wideband relay networks and sensor networks over a random field. The fundamental limits to be pursued will provide a foundation for practical system design and serve as benchmarks on the performance of practical coding an modulation schemes. Accordingly, the study is organized into the following three thrusts: 1) exploring capacity bounds of MIMO relay

channels; 2) characterizing scaling laws of wideband relay networks; and 3) investigating the tradeoff between sensing and communication of sensor networks in a random field.

1.1. MIMO Relay Channel

Recently, the use of multiple antennas at wireless transmitters and receivers has been identified as an enabling technique for high-rate multimedia transmissions over wireless channels. Although the point-to-point MIMO channel is relatively well-understood (particularly the information-theoretic aspects therein), the general area of multi-user MIMO communications, i.e., the field of communications involving a network of many users using multiple antennas, is still at a stage of its infancy and poses a rich set of challenges to the research community.

A MIMO relay channel consists of three terminals: one source node, one destination node and one relay node, where each node are equipped with multiple transmit/receive antennas. We consider MIMO relay channels because it is a building block for wireless networks. For example, for transmissions from a base station (access point) to users, relay stations can be exploited to relay messages for end users. The motivation for using relay stations can be simply put as follows: 1) In a cellular network, direct transmissions between the base station and users close to the cell boundary can be very expensive in terms of the transmission power required to ensure reliable communications; and 2) existing RF technologies typically can accommodate only one or two antennas at the user end, indicating that current wireless systems cannot fully benefit from promising space-time techniques. By making use of relay stations (which can accommodate multiple antennas) to relay the

message, the channel is effectively converted into a *MIMO relay channel*. Another example is to utilize relay nodes for cooperative communications in ad hoc networks, where the nodes close to the active transmitter and the receiver can relay data packets from the transmitter to the receiver. In a nutshell, relaying in wireless networks is garnering much attention (see, e.g. [33, 55, 24, 47, 21, 32, 50]). Notably, some interesting cooperative schemes have recently been investigated (see [27, 28, 38, 39] and the references therein).

The MIMO relay channel can be viewed as a hybrid of a broadcast channel (BC) and a multi-access channel (MAC). For the relay node, a key assumption is that it works in a full-duplex mode. In a MIMO relay channel, both BC and MAC parts are vector channels, making it non-trivial to characterize the bounds on the channel capacity. Indeed, because the vector broadcast channel is not degraded in general, the corresponding channel capacity region is challenging to characterize. Notably, this problem has been solved by very recent work [53]. It has also been studied in [45] by using the duality technique, which establishes a dual relationship between the Gaussian MAC vector channel and the BC vector channel. The sum rate of the Gaussian vector BC channel has been studied in [46, 57, 4]. The capacity region of the MAC channel is relatively better understood (see, e.g. [54, 5, 37, 58]).

We first consider capacity bounds on the Gaussian MIMO relay channel with fixed channel gains. We derive upper bounds and lower bounds, and discuss the corresponding codebook structures. We note that the upper bound for a general relay channel in [6] involves maximization over the joint (multi-dimensional) distribution of the codebooks at the source node and the relay node, and in general the corresponding characterization for vector channels is highly non-trivial. In this chapter, we derive an upper bound involving maximization over two covariance matrices and one scalar parameter ρ. Loosely speaking,

parameter ρ "captures" the cooperation between the transmitted signals from the source node and the relay node. Moreover, using ρ leads to a simplified upper bound and enables us to solve the maximization problem by convex programming. Next, we give a lower bound by finding the maximum between the capacity for the direct link channel (from the source node to the destination node) and that for the cascaded channel (from the source node to the relay node and from the relay node to the destination node). We present algorithms to compute the bounds accordingly.

Next, we generalize the study to a more interesting case – the Rayleigh fading case. We focus on the ergodic capacity of the MIMO relay channel, assuming receiver CSI only. It is somewhat surprising that the upper bound can meet the lower bound under certain conditions (not necessarily degradedness), indicating that the ergodic capacity can be characterized exactly. Thus motivated, we investigate conditions for capacity achievability. In particular, we identify sufficient conditions to achieve the ergodic capacity when all nodes have the same number of antennas; and our intuition for this finding is as follows: The source node and the relay node can function as a "virtual" transmit antenna array when the relay node is located close to the source node, thus making it possible to achieve the capacity.

We note that the findings on the ergodic capacity point to independent coding strategies at the source node and the relay node. Such independence of coding strategies is due to the channel uncertainty (randomness) at the transmitters. Compared to the direct link, the relay channel offers a significant capacity gain, thanks to the multi-access gain and the broadcast gain. We note that both the multi-access gain and the broadcast gain benefit from the full duplexity at the relay node. Needless to say, a key step to reap the capacity gain

is to develop coding strategies for the cooperative multi-access channel and the cooperative broadcast channel. We also provide numerical examples to illustrate the upper bound and the lower bound on the ergodic capacity. Motivated by the capacity gain by using the relay node, we finally discuss the utility of the MIMO relay channel in cooperative communications in ad hoc networks.

1.2. Wideband Relay Networks

A sensor network is often designed to carry out a set of high-level data acquisition and information processing tasks. The raw sensor measurements are obtained and then need to be transported to a data collecting/fusion center to be analyzed, which requires reliable networking capabilities in disadvantaged wireless environments. There are many sensor network models developed for different applications: multi-hop sensor networks, many-to-one sensor networks, sensory relay networks - to name a few.

In the second thrust, we consider power constrained wireless relay networks in the wideband regime, because of their ability of overlay with other legacy networks and the advantage that the use of a larger bandwidth can offer significant power savings. In sensor networks, most sensor devices have limited power supply (e.g., battery supply), and lower power consumption of sensor devices is of critical importance. Wideband communications has recently garnered much attention (e.g., ultra-wide-band (UWB) systems) [56]. Indeed, UWB radios are expected to be inexpensive and low-power, and are ideal for sensor network applications. It is of great interest to investigate the fundamental limits in such systems, such as the capacity and energy efficiency. Thus motivated, we focus on the relay networks

in which each node is power constrained and the frequency bandwidth for communication may be large. Following [44], "the wideband regime" here encompasses all scenarios where the information bits transmitted per receive dimension are small. Two key challenges in power constrained wideband relay networks are as follows: 1) coherent relaying may not be feasible due to the low SNR and the corresponding cooperative relaying strategies should be pursued; and 2) when frequency bandwidth is large, it remains open what power allocation policies across the subbands work well at relay nodes.

As depicted in Fig. 1.1, a wireless relay network consists of one source-destination pair and n relay nodes. When the measurement data at the source node are sent to the destination node, many intermediate sensors can serve as relay nodes to cooperatively relay the information. (This model is well motivated by event-driven sensor networks, in which the data gathered by the source node can be relayed by other sensors in a cooperative manner.) The cooperation among relay nodes can yield diversity gain to enhance the data transport capacity between the source node and the destination node. As is standard (e.g., [14, 15, 8, 2, 29, 48]), we consider a two-hop relaying strategy, in which each relay node first receives the signal from the source node in the first hop and then forwards a processed signal to the destination node in the second hop. We refer the links between the source node and the relay nodes as the backward channels, composing the first hop; and those between the relay nodes and the destination node as the forward channels, corresponding to the second hop. We assume that every link experiences Rayleigh fading. We note that multiple source-relay-destination links offer multiple spatial routing paths for information relaying, pointing to potential cooperative diversity gain.

Clearly, The key to reap the inherent cooperative diversity gain in such networks is to

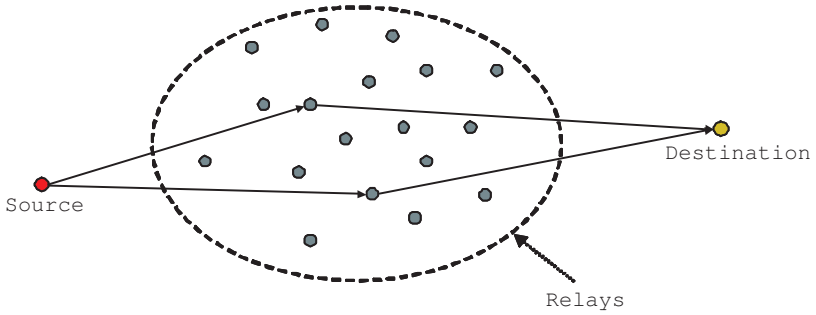

Figure 1.1. A wireless relay network model

devise good relaying strategies. Recently, "decode-and-forward (DF)" and "amplify-and-forward (AF)" have emerged as two popular candidates for two-hop information relaying. In the decode-and-forward strategy, the relay nodes first extract the information bits from the received signals, re-encode it, and then transmit the new encoded bits in the second hop. For instance, one DF scheme can be to select the most reliable source-relay-destination path to transmit the signal. In general, the optimal DF strategy is highly non-trivial and the complexity increases significantly as the number of relay nodes grows. Alternatively, simple amplify-and-forward relaying can be used to relay information and has low complexity. Intuitively speaking, to fully reap the cooperative diversity gain, a good AF relaying scheme should reduce the phase-offset across different source-relay-destination links as much as possible, and this phase-aligning operation should enable coherent combining of transmitted signals from relay nodes at the destination node. The focus of this study is to investigate the "right" relaying strategies for achieving high throughput in wireless relay networks in the wideband regime. In particular, we attempt to address the following critical issues:

- When would "amplify-and-forward" work well in relay networks in the wideband regime? What is the "right" amplify-and-forward scheme to close the gap between

coherent relaying and non-coherent relaying?

- What is the "right" power allocation strategy at the relay nodes that would maximize the scaling order of the achievable rates of relay networks in the wideband regime?

- When would it be possible to achieve the same scaling laws as in coherent relay networks?

Recall that the wideband regime encompasses all scenarios where the information bits transmitted per receive dimension are small. One key challenge in relay networks in the wideband regime is that coherent communications may not be feasible due to the low SNR. To obtain a clear understanding of the impact of low SNR on cooperative relaying, we make a more realistic assumption that at relay nodes there is no *a priori* knowledge of channel state information (CSI) for both the backward channels and the forward channels, and this is one of key features distinguishing our model from the existing ones. As a result, each relay node needs to estimate the channel conditions prior to the data transmission, in order to carry out the phase-alignment as in coherent relaying. In particular, the channel estimation of the forward links is more challenging because it is a multi-access channel. A naive approach is to carry out channel estimation of the forward links separately, which would lead to a significant reduction of the throughput and is clearly not energy-efficient. To overcome this difficulty, we study the following strategy for AF relaying. As depicted in Fig. 3.1, at the beginning of the first hop transmission, the source node first broadcasts common pilot symbols, followed by its data transmission. Based on the received signals corresponding to the pilot symbols, each relay nodes estimates its own backward channel condition. Then at the beginning of the second hop, the destination node (not the relay

nodes) sends out pilot symbols (We assume that the forward links are reciprocal in both directions in each subband.) Then, each relay node estimates its forward channel condition. After the channel estimation is done, each relay node carries out phase alignment by using its channel estimates of both the backward channel and the forward channel, amplifies the received data signal under the given power constraints, and forwards the processed signal to the destination node. For convenience, we refer this strategy as amplify-and-forward (AF) with *network training*. We emphasize that the above channel estimation approach makes use of the broadcast nature of wireless transmissions, and is an energy-efficient candidate for network training. For instance, this network training approach can be applied by all source-destination pairs in a collocated network [11].

The main focus in the second thrust is on characterizing the achievable rates and scaling laws by using AF with network training in power-constrained relay networks in the wideband regime. Simply put, the scaling law here is concerned with the order of the achievable rates as the number of the relay nodes n grows, i.e., the scaling behavior of the throughput (and capacity) as $n \to \infty$. Let ρ denote the SNR per link. It is understood that $\rho \to 0$ in the wideband regime. Furthermore, it is clear that the coherence interval of the fading channel, denoted as L, plays a key role in channel estimation. To obtain a clear understanding of the continuum between coherent relaying and non-coherent relaying, we examine the achievable rates and scaling laws in the joint asymptotic regime of the number of relay nodes n, the channel coherence interval L, the SNR per link ρ (or the bandwidth W). We note that although there is no physical connection among these parameters, the achievable rates depend on how they scale together, rather than on each of them in isolation [62, 40].

Since a wideband channel can be decomposed into multiple orthogonal narrowband channels, each with flat fading, we first investigate achievable rates and scaling laws by using AF relaying with network training in narrowband relay networks in the low SNR regime [61]. Our findings show that when ρL (which is proportional to the transmission energy per fading block) is bounded below, AF with network training achieves the same scaling law as in the coherent relay networks where the CSI is known a priori. We then generalize the study to power-constrained wideband relay networks in which frequency-selective fading is inevitable. We model the wideband relay networks as a set of multiple narrowband relay networks under an overall power constraint. We examine the achievable rates and scaling laws by using AF with network training, under two power allocation policies, respectively. Specifically, we first assume that an equal power allocation policy across the frequency subbands[1] is applied at each relay node. Then we examine the optimal power allocation policy for wideband relay networks, where each relay node is allowed to allocate its power across the subchannels and the fading blocks. In general, one would expect a higher achievable rate by using bursty flashing signals (more generally, water-filling techniques). Somewhat surprising, our findings show that the achievable rates by using AF with network training have the same scaling order as those given by the equal power allocation policy, and both policies can achieve the scaling law under the condition that L/W is bounded below and that W is sub-linear in n. That is to say, the equal power allocation policy at relay nodes is order optimal in the sense of achieving the scaling law asymptotically. Roughly speaking, the condition that W is sub-linear in n is equivalent to that the degree of node multiplexing per unit bandwidth (n/W) grows as n increases.

[1]Throughout, we use the notation of subchannel and subband interchangeably.

In related work, the capacity of wireless networks and the capacity of relay channels have been studied extensively in the past few years (see, [20, 18, 14, 15, 2, 8, 10, 51, 21, 25, 28, 36, 35, 43, 49] and the references therein). In their seminal work [20], Gupta and Kumar investigate the throughput scaling for dense many-to-many multi-hop network models, and they prove that an aggregated throughput of $O(\sqrt{n})$ can be achieved if nodes can be arbitrarily located. Grossglauser and Tse study the impact of mobility on the throughput in mobile ad hoc networks [18]. They show an $O(n)$ throughput can be achieved by using two-hop relaying and allowing possibly unbounded delay. Recently, wireless relay networks has received much attention. In [14, 15], Gastpar and Vetterli show that the asymptotic capacity of wireless relay networks behaves like $\log(n)$ under the relay traffic pattern, assuming that power allocation can be carried out across the relay nodes. Bölcskei, Nabar, Oyman and Paulraj find in [2] that the throughput of wireless relay networks, where each node is equipped with M antennas, scales as $M \log(n)$. Recent work [8] by Dana and Hassibi reveals that the power efficiency of the AF scheme scales at least by a factor of $\sqrt{n}$. Very recent works [36, 35] by Oyman and Paulraj find that a higher power efficiency of order n can be achieved by using bursty transmission. Dousse, Franceschetti and Thiran discover in [10] that the per-node throughput remains constant as the size of the wireless network increases, if a "small" fraction of nodes are allowed to be disconnected.

1.3. Communication vs. Sensing in Sensor Networks

In sensor networks, the data at the sensor nodes are transported through the network and then collected by the sink. A key assumption in this study is that some sensor nodes

cannot reach the sink directly, due to the limited transmit power. In this study, a two-hop relaying transmission pattern is introduced to help the data transport. Simply put, the nodes that cannot reach the sink broadcast their data in the first hop, and in the second hop the relay nodes combine the received data and their own sensed data and then forward to the sink. We use the mutual information between the physical sources and the received signal at the sink as the performance metric to characterize the information fidelity. Indeed, the mutual information between the input signal and the output signal is inherently related to the minimum mean square error (MMSE) estimation [9], and higher mutual information indicates more precise estimation of the input signal at the receiver.

Based on the observability of the sources, we consider two important cases. First, we study an event-driven sensor network where there is one single physical source to be monitored and the source can be observed "directly" by the nodes, in the sense that the observations are additive noisy versions of the physical source (the source variable plus noise). This model is applicable to the cases where the physical signals can "propagate" through the space. Next, we consider a senor network model over a two-dimensional (2-D) random field, where the physical sources cannot be observed directly by the sensor nodes. Instead, the network is modeled as a two-dimensional random field, in the sense that the observations and the sources follow a joint probability model, determined by the distribution of the sources and the conditional distribution of the observations given the state. The observation process can then be characterized by the transition probability corresponding to a "sensing channel" with the input (the sources) and the output (the observations).

For the event-driven sensor network, we focus on the case where there is one "critical" node that is closest to the event. The observation from the critical node is relayed to the

sink through the two-hop communication channel. The event-driven sensor network model under consideration is similar to the wireless relay network model in Chapter 3, except that in the event-driven sensor network model, each node has its own data to transmit; whereas in the wireless relay network model, the relay nodes do not have their own data to send. Motivated by the studies for the wireless relay networks [14, 52], we propose an amplify-and-forward with linear combination (AFLC) scheme, in which each relay node transmits a linear combination of the received signal from the critical node and its own signals. It is then critical to study the weights (power allocation) for the two types of signals, which in fact points to the tradeoff between sensing and communication. To characterize this tradeoff, we first show that in an additive Gaussian noise channel the mutual information is an increasing function of the SNR (defined as the ratio of the variance of the received information-bearing signal to that of received noise), which indicates that maximizing the overall SNR at the sink leads to maximum mutual information. Then we propose an AFLC scheme with optimal power allocation in the sense that it can maximizes the output SNR at each relay node. This scheme in turn maximizes the mutual information between the source variable and the received signal at the sink.

Next, we generalize the study to the sensor network model over a 2-D random field. We note that at each relay node, there are in fact two observations of the sources, and thus the relay nodes can first estimate the source variables (this can be viewed as a data fusion process) and then forward the estimates to the sink. We assume that minimum mean square error (MMSE) estimation is applied at the relay nodes to estimate the source variables based on the received signals and the sensed signals. Interestingly, MMSE provides a mechanism to perform data fusion and thus achieve the desired tradeoff of sensing and

communication. When the observations and sources are jointly Gaussian distributed, the estimates of the sources can be obtained explicitly, which enables us to characterize mutual information. We then focus on two cases. We first consider the case where there is one node whose data needs to be relayed by the other nodes. We find the MMSE estimator for each relay node and characterize the mutual information. We then consider a case where clusters can be formed in the network, and there is only one node in each cluster (referred to as the cluster head) which is responsible for the communication with the sink. We characterize the mutual information for two scenarios: I) all the cluster heads are allowed to transmit, and II) only some cluster heads are allowed to transmit.

It is also of interest to study the communication from the sink to the nodes, i.e., the down-link transmission when the sink needs to send querying messages to the nodes. Often times some nodes may not be able to obtain successfully the messages directly from the sink. Again, for simplicity, we assume that a two-hop transmission pattern is applied for data transport, where the nodes that have obtained the messages directly from the sink in the first hop can function as relay nodes by forwarding the messages to the rest of the network in the second hop. Note that in this case there can be multiple receivers in the second hop, and thus an amplify-and-forward scheme may not function well due to the lack of coherent relaying. We then propose to use a decode-and-forward (DF) scheme. Simply put, the sink first broadcasts the message at a fixed rate in the first hop. In the second hop, all nodes that successfully decode the message in the first hop then retransmit the same message using the same codebook as by the sink. We show that as the size of the network increases, the probability of failed decoding at a node decreases.

1.4. Outline of the Dissertation

The dissertation continues as follows. We study in Chapter 2 the capacity bounds for MIMO relay channel. We first derive an upper bounds and lower bounds on the capacity of the Gaussian MIMO relay channel with fixed channel gains. We then present an upper bound and a lower bound on the ergodic capacity and provide numerical results for different SNR cases. We then discuss sufficient conditions for achieving the ergodic capacity. Finally, a potential application of the relay channel in cooperative communications in ad hoc networks is presented.

In Chapter 3, we start with the models for narrowband relay networks in the low SNR regime and power-constrained wideband relay networks. We obtain an equivalent source-to-destination channel model for narrowband relay networks using AF with network training, and examine the scaling order of the achievable rates in different scenarios. We then characterize the conditions for achieving the scaling laws accordingly. We generalize the study to power-constrained wideband relay networks. We analyze the achievable rates and the scaling order corresponding to the equal power allocation policy at the relay nodes, and compare the scaling behavior with that by using the optimal power allocation policy at the relay nodes. Numerical examples are also provided to illustrate the achievable rates and capacity bounds.

In Chapter 4, we start with an event-driven sensor network model and obtain the equivalent source-to-destination channel model. We characterize the tradeoff between sensing and communication by carrying out the power allocation for the sensed signal and the communication signal. We then turn to a sensor network over a 2-D random field. Two

cases are investigated for a Gaussian field. We also examine the outage probability of the downlink transmission from the sink to the nodes

We conclude our work in Chapter 5 and present some discussion for future work.

CHAPTER 2

On the Capacity of MIMO Relay Channel

In this chapter, we study the capacity of MIMO relay channels. We first consider the Gaussian MIMO relay channel with fixed channel conditions, and derive upper bounds and lower bounds that can be obtained numerically by convex programming. We present algorithms to compute the bounds. Next, we generalize the study to the Rayleigh fading case. We find an upper bound and a lower bound on the ergodic capacity. It is somewhat surprising that the upper bound can meet the lower bound under certain regularity conditions (not necessarily degradedness), and therefore the capacity can be characterized exactly; previously this has been proven only for the degraded Gaussian relay channel. We investigate sufficient conditions for achieving the ergodic capacity; and in particular, for the case where all nodes have the same number of antennas, the capacity can be achieved under certain SNR conditions. Numerical results are also provided to illustrate the bounds on the ergodic capacity of the MIMO relay channel over Rayleigh fading. Lastly, we present a potential application of the MIMO relay channel for cooperative communications in ad hoc networks.

Throughout this chapter, we use $\mathbb{E}$ to denote the expectation operator (in some cases, subscripts are used to specify the random variable); "†" stands for the conjugate transpose;

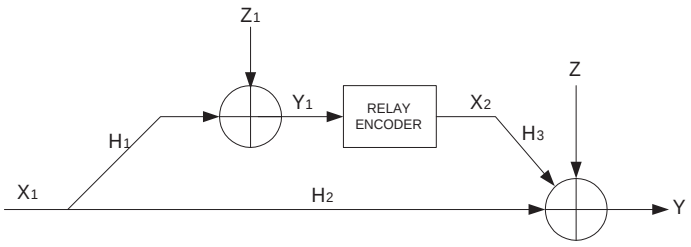

Figure 2.1. MIMO relay channel

$\mathbf{I}_M$ denotes an identity matrix; $\mathbf{0}$ is an all-zero matrix of proper dimensions; the distribution of a circularly-symmetric complex Gaussian vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted as $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; $\preceq$, $\prec$, $\succ$ and $\succeq$ are used in the matrix positive (semi-)definite ordering sense [22].

2.1. System Model

Consider a general MIMO relay channel as illustrated in Fig. 2.1, where the received signals at the relay and destination nodes can be written as

$$\begin{cases} \mathbf{Y}_1 &= \sqrt{\eta_1}\mathbf{H}_1\mathbf{X}_1 + \mathbf{Z}_1 \\ \mathbf{Y} &= \sqrt{\eta_2}\mathbf{H}_2\mathbf{X}_1 + \sqrt{\eta_3}\mathbf{H}_3\mathbf{X}_2 + \mathbf{Z}\ , \end{cases} \tag{2.1}$$

where

- $\mathbf{X}_1$, $\mathbf{X}_2$ are $M_1 \times 1$ and $M_2 \times 1$ transmitted signals from the source node and the relay node; the power constraints on the transmit signals are $\mathbb{E}(\mathbf{X}_1^\dagger \mathbf{X}_1) \leq M_1$ and $\mathbb{E}(\mathbf{X}_2^\dagger \mathbf{X}_2) \leq M_2$;

- $\mathbf{Y}$ and $\mathbf{Y}_1$ are $N \times 1$ and $N_1 \times 1$ received signals at the destination node and the relay node. We assume that

  - The relay node has two sets of antennas, one for reception and the other for transmission. That is, the relay node operates in the full-duplex mode;
  - Since the relay node has full knowledge of what to transmit therein, it can cancel out the interference from its own transmit antennas at its receive antennas;

- $\mathbf{H}_1$, $\mathbf{H}_2$ and $\mathbf{H}_3$ are $N_1 \times M_1, N \times M_1$ and $N \times M_2$ channel gain matrices, as depicted in Fig. 2.1. In what follows, we consider two scenarios for the channel matrices:
  - All the channel matrices are fixed and known at both the transmitters and the receivers;
  - All the channel matrices are random and independent, where the entries of each matrix are i.i.d complex Gaussian variables with zero mean, independent real and imaginary parts, each with variance 1/2, and they are available at the corresponding receivers only (i.e., receiver CSI only);

- η_1, η_2 and η_3 are parameters related to the signal to noise ratio (SNR) [30]:

$$\eta_1 = \frac{\mathrm{SNR}_1}{M_1}, \quad \eta_2 = \frac{\mathrm{SNR}_2}{M_1}, \quad \eta_3 = \frac{\mathrm{SNR}_3}{M_2}, \tag{2.2}$$

  where SNR_1 and SNR_2 are the normalized power ratios of $\mathbf{X}_1$ to the noise (after fading) at each receiver antenna of the relay node and the destination node, and SNR_3 is the normalized power ratio of $\mathbf{X}_2$ to the noise at each antenna of the destination node;

- $\mathbf{Z}$ and $\mathbf{Z}_1$ are independent $N \times 1$ and $N_1 \times 1$ circularly-symmetric complex Gaussian noise vectors with distributions $\mathcal{CN}(\mathbf{0},\ \mathbf{I}_N)$ and $\mathcal{CN}(\mathbf{0},\ \mathbf{I}_{N_1})$, and are uncorrelated to $\mathbf{X}_1$ and $\mathbf{X}_2$.

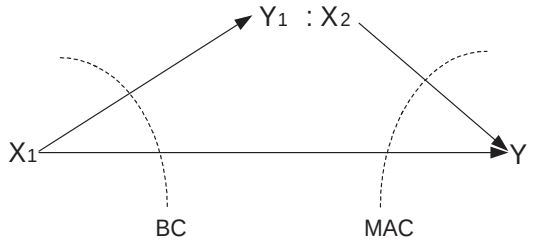

Figure 2.2. The relay channel: max-flow min-cut

2.2. Capacity Bounds: The Fixed Channel Case

In the following, we derive upper bounds and lower bounds on the capacity of the Gaussian MIMO relay channel with fixed channel gains.

2.2.1. Upper Bounds and Lower Bounds

Recall from [6] that the channel capacity of a general Gaussian relay channel is upper bounded by

$$C^G \leq \max_{p(\mathbf{x}_1,\mathbf{x}_2)} \min\left(\mathbf{I}(\mathbf{X}_1;\mathbf{Y},\mathbf{Y}_1|\mathbf{X}_2),\ \mathbf{I}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}) \right) , \tag{2.3}$$

where the first term in $\min(\cdot,\cdot)$ can be treated as the sum rate from the source node to the relay node and the destination node, corresponding to a BC part; and the second term can be viewed as the sum rate from the source node and the relay node to the destination node, corresponding to a MAC part. Indeed, (2.3) has an interesting max-flow min-cut interpretation [7], as illustrated in Fig. 2.2. Roughly speaking, the rate of the information flow transmitted on the relay channel is constrained by the bottle-neck corresponding to either the first cut (BC) or the second one (MAC).

Without loss of generality, let $\mathbf{X}_1$ and $\mathbf{X}_2$ be random vectors with zero-mean and

covariance matrices $\mathbf{\Sigma}_{ij}$, defined as $\mathbf{\Sigma}_{ij} = \mathbb{E}[\mathbf{X}_i \mathbf{X}_j^\dagger]$ for $i, j = 1, 2$. Throughout, we assume that $\det(\mathbf{\Sigma}_{22}) > 0$ and $\det(\mathbf{\Sigma}_{11}) > 0$. Define $\mathbf{A} \triangleq \mathbf{\Sigma}_{11}^{-\frac{1}{2}} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-\frac{1}{2}}$. First, we need the following lemmas (the proof of Lemma 2.1 is relegated to Appendix A).

Lemma 2.1 *There exists $\rho \in [0, 1]$ such that*

$$\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger \preceq (1 - \rho^2)\mathbf{I}_{M_1}, \tag{2.4}$$

and the equality can be achieved by a matrix $\mathbf{\Sigma}_{12} \in \mathcal{C}^{M_1 \times M_2}$ when $M_1 \leq M_2$.

Intuitively speaking, Lemma 2.1 reveals that for any given $\mathbf{\Sigma}_{11}$ and $\mathbf{\Sigma}_{22}$, we can find codebooks for $\mathbf{X}_1$ and $\mathbf{X}_2$ such that the covariance matrix $\mathbf{\Sigma}_{12}$ satisfies the following inequality:

$$\mathbf{0} \preceq \mathbf{\Sigma}_{11}^{-\frac{1}{2}} (\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21}) \mathbf{\Sigma}_{11}^{-\frac{1}{2}} \preceq (1 - \rho^2)\mathbf{I}_{M_1}.$$

Furthermore, if $M_1 \leq M_2$, then for any given $\rho \in [0, 1]$, there exists $\mathbf{\Sigma}_{12}$ such that the equality is achieved.

Lemma 2.2 *For any two complex random vectors $\mathbf{V}$ and $\mathbf{W}$, $\forall a > 0$, we have that*

$$\mathbb{E}(\mathbf{V}\mathbf{W}^\dagger + \mathbf{W}\mathbf{V}^\dagger) \preceq \mathbb{E}(\frac{1}{a}\mathbf{V}\mathbf{V}^\dagger + a\mathbf{W}\mathbf{W}^\dagger) \quad , \tag{2.5}$$

the equality is achieved if $\mathbf{V} = a\mathbf{W}$.

The proof follows simply from the fact that the covariance matrix of random vector $(\mathbf{V} - a\mathbf{W})$ is always positive semi-definite. That is,

$$\mathrm{Cov}(\mathbf{V} - a\mathbf{W}) = \mathbb{E}\left(\mathbf{V}\mathbf{V}^\dagger + a^2 \mathbf{W}\mathbf{W}^\dagger\right) - a\mathbb{E}\left(\mathbf{V}\mathbf{W}^\dagger + \mathbf{W}\mathbf{V}^\dagger\right) \succeq \mathbf{0}\,.$$

Let $\mathbf{V} = \sqrt{\eta_2}\mathbf{H}_2\mathbf{\Sigma}_{11}^{\frac{1}{2}}\mathbf{A}\mathbf{\Sigma}_{22}^{-\frac{1}{2}}\mathbf{X}_2$ and $\mathbf{W} = \sqrt{\eta_3}\mathbf{H}_3\mathbf{X}_2$. It follows that

$$\begin{aligned}\mathbb{E}(\mathbf{V}\mathbf{W}^\dagger + \mathbf{W}\mathbf{V}^\dagger) &= \sqrt{\eta_2\eta_3}\mathbf{H}_2\mathbf{\Sigma}_{12}\mathbf{H}_3^\dagger + \sqrt{\eta_2\eta_3}(\mathbf{H}_2\mathbf{\Sigma}_{12}\mathbf{H}_3^\dagger)^\dagger \\ \mathbb{E}(\mathbf{V}\mathbf{V}^\dagger + \mathbf{W}\mathbf{W}^\dagger) &= \eta_2\mathbf{H}_2\mathbf{\Sigma}_{11}^{\frac{1}{2}}\mathbf{A}\mathbf{A}^\dagger\mathbf{\Sigma}_{11}^{\frac{1}{2}}\mathbf{H}_2^\dagger + \eta_3\mathbf{H}_3\mathbf{\Sigma}_{22}\mathbf{H}_3^\dagger \ .\end{aligned}$$

If signals $\mathbf{X}_1$ and $\mathbf{X}_2$ are chosen such that $\mathbf{A}\mathbf{A}^\dagger = \rho^2\mathbf{I}_{M_1}$, then

$$\mathbb{E}(\mathbf{V}\mathbf{V}^\dagger + \mathbf{W}\mathbf{W}^\dagger) = \rho^2\eta_2\mathbf{H}_2\mathbf{\Sigma}_{11}\mathbf{H}_2^\dagger + \eta_3\mathbf{H}_3\mathbf{\Sigma}_{22}\mathbf{H}_3^\dagger \ .$$

Applying Lemma 2.2, for $\forall a > 0$, we have that

$$\sqrt{\eta_2\eta_3}\mathbf{H}_2\mathbf{\Sigma}_{12}\mathbf{H}_3^\dagger + \sqrt{\eta_2\eta_3}(\mathbf{H}_2\mathbf{\Sigma}_{12}\mathbf{H}_3^\dagger)^\dagger \preceq \frac{\rho^2}{a}\eta_2\mathbf{H}_2\mathbf{\Sigma}_{11}\mathbf{H}_2^\dagger + a\eta_3\mathbf{H}_3\mathbf{\Sigma}_{22}\mathbf{H}_3^\dagger \ . \tag{2.6}$$

It is clear that the optimal distribution $p(\mathbf{x}_1, \mathbf{x}_2)$ in (2.3) is Gaussian [41, 6] (see, e.g., Step (b) in page 10 and Step (a) in page 12). Observe that if Gaussian codebooks are applied, the maximization problem on the right hand side of (2.3) would be with respected to three covariance matrices $\mathbf{\Sigma}_{11}$, $\mathbf{\Sigma}_{22}$ and $\mathbf{\Sigma}_{12}$; and this is non-convex and highly non-trivial in general. In Lemma 2.1 and Lemma 2.2, we use one parameter ρ to capture the correlation between $\mathbf{X}_1$ and $\mathbf{X}_2$ (in contrast to the cross-covariance matrix $\mathbf{\Sigma}_{12}$), and this enables us to solve the optimization by convex programming techniques (in Appendix A, we present the definition for ρ). In what follows, we present our results for the upper bound on the capacity of Gaussian MIMO relay channels.

Theorem 2.1 ***[Fixed Channel Case]*** *An upper bound on the capacity of the MIMO relay channel is given by*

$$C^G \leq C^G_{upper} = \max_{0\leq\rho\leq 1,\ \mathbf{\Sigma}_{11},\ \mathbf{\Sigma}_{22}} \min\ (C_1^G\ ,\ C_2^G)\ , \tag{2.7}$$

where $\mathrm{tr}(\mathbf{\Sigma}_{11}) \leq M_1$, $\mathrm{tr}(\mathbf{\Sigma}_{22}) \leq M_2$; C_1^G *and* C_2^G *are given by*

$$C_1^G \triangleq \log \left[\det \left(\mathbf{I}_{M_1} + \left(1-\rho^2\right) \begin{bmatrix} \sqrt{\eta_1}\mathbf{H}_1 \\ \sqrt{\eta_2}\mathbf{H}_2 \end{bmatrix} \mathbf{\Sigma}_{11} \begin{bmatrix} \sqrt{\eta_1}\mathbf{H}_1 \\ \sqrt{\eta_2}\mathbf{H}_2 \end{bmatrix}^{\dagger} \right) \right], \tag{2.8}$$

$$C_2^G \triangleq \inf_{a>0} \log \left[\det \left(\mathbf{I}_N + \left(\eta_2 + \frac{\rho^2}{a}\sqrt{\eta_2\eta_3} \right) \mathbf{H}_2 \mathbf{\Sigma}_{11} \mathbf{H}_2^{\dagger} + \left(\eta_3 + a\sqrt{\eta_2\eta_3} \right) \mathbf{H}_3 \mathbf{\Sigma}_{22} \mathbf{H}_3^{\dagger} \right) \right] \tag{2.9}$$

[**Remarks:**] To find the upper bound, we need to characterize the optimal input covariance matrices $\mathbf{\Sigma}_{11}$ and $\mathbf{\Sigma}_{22}$. Define

$$f_0(\mathbf{\Sigma}_{11},\ \mathbf{\Sigma}_{22},\ \rho) \triangleq \min(C_1^G\ ,\ C_2^G)$$

It can be seen that C_1^G is concave in $\mathbf{\Sigma}_{11}$ and that C_2^G is concave in $(\mathbf{\Sigma}_{11},\ \mathbf{\Sigma}_{22})$. It follows that $\min(C_1^G\ ,\ C_2^G)$ is concave in $(\mathbf{\Sigma}_{11},\ \mathbf{\Sigma}_{22})$[3]. More precisely, for a given ρ, the upper bound is concave in $(\mathbf{\Sigma}_{11},\ \mathbf{\Sigma}_{22})$; and it can be found by convex programming. In what follows, we present an algorithm to compute the upper bound.

Algorithm I

1. Carry out a quantization of interval $[0, 1]$, and denote the corresponding set of (ascending) values as $\{\rho_1, \rho_2, ..., \rho_m\}$;

2. For a given ρ_i, apply convex programming to find the optimal value of $f_0(\cdot, \cdot)$ and corresponding optimal $\mathbf{\Sigma}_{11}$ and $\mathbf{\Sigma}_{22}$, by solving the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & f_0(\mathbf{\Sigma}_{11},\ \mathbf{\Sigma}_{22}) \\ \text{subject to} \quad & \text{tr}(\mathbf{\Sigma}_{11}) - M_1 \leq 0,\ \text{tr}(\mathbf{\Sigma}_{22}) - M_2 \leq 0 \\ & \mathbf{\Sigma}_{11} \succ \mathbf{0},\ \mathbf{\Sigma}_{22} \succ \mathbf{0}\ ; \end{aligned}$$

3. Pick another ρ_i and repeat Step 2. Go to the next step when the set $\{\rho_1, \rho_2, ..., \rho_m\}$ is exhausted;

4. Compare all the values of $f_0(\cdot, \cdot)$ associated with $\{\rho_1, \rho_2, ..., \rho_m\}$, and find the largest and identify the corresponding optimal parameters ρ_k^*, $\mathbf{\Sigma}_{11}$ and $\mathbf{\Sigma}_{22}$;

5. Quantize $[\rho_{k-1}^*, \rho_{k+1}^*]$ and go to Step 2 using $\{\rho_{k-1}^* + \Delta\rho_1, \rho_{k-1}^* + \Delta\rho_2, ..., \rho_{k-1}^* + \Delta\rho_m\}$ for a new search;

6. Compare the refined results from the new search with the old ones. If the error requirement ϵ is met, end the procedures; otherwise, go to previous steps for another new search.

A few more words on Algorithm I. Since there is no a priori information about ρ in the initialization step, the quantization is equal-span. After the first iteration, we choose

the "best guess" of ρ (namely ρ_k^*), and then refine the search around it. To guarantee the convergency to the optimal point, the quantization level m should be reasonably large (e.g., $m \geq 10$).

In the above, we use ρ to capture the correlation between signals transmitted from the source node and the relay node. Now, we discuss the structure of the corresponding codebooks. We can rewrite the transmitted signal $\mathbf{X}_1$ as

$$\mathbf{X}_1 = \widetilde{\mathbf{X}}_{10} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{X}_2 \ ,$$

with $\widetilde{\mathbf{X}}_{10} \triangleq \mathbf{X}_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{X}_2$. Observe that $\widetilde{\mathbf{X}}_{10}$ is independent of $\mathbf{X}_2$. Intuitively speaking, vector signal $\mathbf{X}_1$ can be decomposed into two orthogonal parts, which are independent of each other. The second part, $\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{X}_2$, stands for the projection of $\mathbf{X}_1$ onto the direction of $\mathbf{X}_2$. Thus, $\mathbf{X}_1 = \widetilde{\mathbf{X}}_{10} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{X}_2$ can be viewed as a generalization of [6, Theorem 5].

For the special case where all the channel coefficients are scalars (denoted as h_1, h_2 and h_3 respectively), we have that $\Sigma_{11} = \Sigma_{22} = 1$. Accordingly, we have that

$$C_1^G = \log\left[1 + (1-\rho^2)(\eta_1|h_1|^2 + \eta_2|h_2|^2)\right],$$

and

$$\begin{aligned} & \log\left[1 + \eta_2|h_2|^2 + \eta_3|h_3|^2 + \sqrt{\eta_2\eta_3}(\frac{\rho^2}{a}|h_2|^2 + a|h_3|^2)\right] \\ \geq \ & \log\left[1 + \eta_2|h_2|^2 + \eta_3|h_3|^2 + 2\rho\sqrt{\eta_2\eta_3}|h_2||h_3|\right] \ . \end{aligned} \tag{2.10}$$

It follows that

$$C_2^G = \log\left[1 + \eta_2|h_2|^2 + \eta_3|h_3|^2 + 2\rho\sqrt{\eta_2\eta_3}|h_2||h_3|\right] \ , \tag{2.11}$$

which boils down to a result in [12].

Consider channel models where the relay node may or may not be used to aid the transmissions. If not used, the channel becomes a point-to-point Gaussian MIMO channel. On the other hand, if the relay node is used to aid the transmission and the destination node treats the information directly from the source node as noise, the channel boils down to a cascaded channel. We have the following lower bound by finding the maximum between the information rates for the two channel models.

Theorem 2.2 ***[Fixed Channel case]*** *A lower bound on the capacity of the Gaussian MIMO relay channel is given by*

$$C^G \geq C^G_{lower} = \max(C^G_d,\ \min(C^G_3,\ C^G_4)), \tag{2.12}$$

where

$$C^G_d \triangleq \max_{\mathbf{\Sigma}_{11}} \log\left[\det\left(\mathbf{I}_N + \eta_2 \mathbf{H}_2 \mathbf{\Sigma}_{11} \mathbf{H}_2^\dagger\right)\right] \tag{2.13}$$

$$C^G_3 \triangleq \max_{\mathbf{\Sigma}_{11}} \log\left[\det\left(\mathbf{I}_{N_1} + \eta_1 \mathbf{H}_1 \mathbf{\Sigma}_{11} \mathbf{H}_1^\dagger\right)\right] \tag{2.14}$$

$$C^G_4 \triangleq \max_{\mathbf{\Sigma}_{22}} \log\left[\det\left(\mathbf{I}_N + \eta_3 \mathbf{H}_3 \mathbf{\Sigma}_{22} \mathbf{H}_3^\dagger \left(\mathbf{I}_N + \eta_2 \mathbf{H}_2 \mathbf{\Sigma}^*_{11} \mathbf{H}_2^\dagger\right)^{-1}\right)\right] \tag{2.15}$$

with

$$\mathbf{\Sigma}^*_{11} \triangleq \arg\max_{\mathbf{\Sigma}_{11} \succ 0} \log\left[\det\left(\mathbf{I}_{N_1} + \eta_1 \mathbf{H}_1 \mathbf{\Sigma}_{11} \mathbf{H}_1^\dagger\right)\right].$$

We outline the procedure to compute the lower bound as follows:

Algorithm II

1. Use the water-filling technique to find C^G_d [41];
2. Use the water-filling technique to find C^G_3 and the corresponding optimal $\mathbf{\Sigma}^*_{11}$;
3. Substitute $\mathbf{\Sigma}^*_{11}$ into (2.15) to find C^G_4 by using the water-filling technique.

In Appendix B, we provide another lower bound by using the fact that the following rate is achievable for any given distribution $p(\mathbf{x}_1, \mathbf{x}_2)$ [6]:

$$R = \min\left(\mathbf{I}(\mathbf{X}_1; \mathbf{Y}_1|\mathbf{X}_2),\ \mathbf{I}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y})\right)\ . \tag{2.16}$$

We note that this lower bound does not admit to a closed-form solution, but it may yield a tighter bound.

2.2.2. Proofs of Theorems 2.1 and 2.2

2.2.2.1. *Proof of Theorem 2.1*

Proof: Define $\mathbf{H}_{12} \triangleq \begin{bmatrix} \sqrt{\eta_1}\mathbf{H}_1 \\ \sqrt{\eta_2}\mathbf{H}_2 \end{bmatrix}$. Given $\mathbf{X}_2 = \mathbf{x}_2$, we can rewrite the channel model as follows:

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y} \end{bmatrix} = \mathbf{H}_{12}\mathbf{X}_1 + \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z} + \sqrt{\eta_3}\mathbf{H}_3\mathbf{x}_2 \end{bmatrix}\ . \tag{2.17}$$

The sum rate of the corresponding BC channel is given by

$$
\begin{aligned}
& \mathbf{I}(\mathbf{X_1}; \mathbf{Y_1}, \mathbf{Y}|\mathbf{X_2}) \\
= \;& h(\mathbf{Y}_1, \mathbf{Y}|\mathbf{X}_2) - h(\mathbf{Y}_1, \mathbf{Y}|\mathbf{X}_1, \mathbf{X}_2) \\
\overset{(a)}{=} \;& \mathbb{E}_{\mathbf{x}_2}\left[h(\mathbf{Y}_1, \mathbf{Y}|\mathbf{X}_2 = \mathbf{x}_2)\right] - h(\mathbf{Z}_1, \mathbf{Z}) \\
= \;& \mathbb{E}_{\mathbf{x}_2}\left[h(\mathbf{Y}_1, \mathbf{Y}|\mathbf{X}_2 = \mathbf{x}_2)\right] - \log((\pi e)^{N_1+N}\det(\mathbf{I}_{N_1+N})) \\
\overset{(b)}{\leq} \;& \mathbb{E}_{\mathbf{x}_2}\left[\log\left((\pi e)^{N_1+N}\det\left(\mathrm{Cov}\left(\left[\begin{array}{c}\mathbf{Y}\\ \\ \mathbf{Y}_1\end{array}\right]\middle|\mathbf{X}_2 = \mathbf{x}_2\right)\right)\right)\right] - \log((\pi e)^{N_1+N}) \\
\overset{(c)}{=} \;& \mathbb{E}_{\mathbf{x}_2}\left[\log\left(\det\left(\mathbf{I} + \mathbf{H}_{12}\boldsymbol{\Sigma}_{\mathbf{X}_1|\mathbf{X}_2=\mathbf{x}_2}\mathbf{H}_{12}^{\dagger}\right)\right)\right] \\
\overset{(d)}{=} \;& \log\left[\det\left(\mathbf{I} + \mathbf{H}_{12}\boldsymbol{\Sigma}_{11}^{\frac{1}{2}}\left(\mathbf{I} - \boldsymbol{\Sigma}_{11}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{22}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{12}^{\dagger}\boldsymbol{\Sigma}_{11}^{-\frac{1}{2}}\right)\boldsymbol{\Sigma}_{11}^{\frac{1}{2}}\mathbf{H}_{12}^{\dagger}\right)\right] \qquad (2.18) \\
= \;& \log\left[\det\left(\mathbf{I} + \mathbf{H}_{12}\boldsymbol{\Sigma}_{11}^{\frac{1}{2}}\left(\mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}\right)\boldsymbol{\Sigma}_{11}^{\frac{1}{2}}\mathbf{H}_{12}^{\dagger}\right)\right] \\
\overset{(e)}{\leq} \;& \log\left[\det\left(\mathbf{I} + \left(1-\rho^2\right)\mathbf{H}_{12}\boldsymbol{\Sigma}_{11}\mathbf{H}_{12}^{\dagger}\right)\right], \qquad (2.19)
\end{aligned}
$$

where

(a) follows from the definition of conditional entropy;

(b) from the fact that circularly-symmetric complex Gaussian distribution maximizes entropy [41];

(c) from the fact that

$$
\mathrm{Cov}\left(\left[\begin{array}{c}\mathbf{Y}\\ \\ \mathbf{Y}_1\end{array}\right]\middle|\mathbf{X}_2 = \mathbf{x}_2\right) = \mathrm{Cov}(\mathbf{H}_{12}\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) + \mathrm{Cov}\left(\left[\begin{array}{c}\mathbf{Z}_1\\ \\ \mathbf{Z} + \sqrt{\eta_3}\mathbf{H}_3\mathbf{x}_2\end{array}\right]\right),
$$

which is based on that given $\mathbf{X}_2 = \mathbf{x}_2$, the vector $\left[\begin{array}{c}\mathbf{Y}\\ \\ \mathbf{Y}_1\end{array}\right]$ is the sum of two independent circularly-symmetric complex Gaussian vectors;

(d) is because that $\mathbf{\Sigma}_{\mathbf{X}_1|\mathbf{X}_2=\mathbf{x}_2}$ is the conditional covariance matrix of $\mathbf{X}_1$ given that $\mathbf{X}_2 = \mathbf{x}_2$, and

$$\begin{aligned}\mathbf{\Sigma}_{\mathbf{X}_1|\mathbf{X}_2=\mathbf{x}_2} &= \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}^{\dagger} \\ &= \mathbf{\Sigma}_{11}^{\frac{1}{2}}\left(\mathbf{I} - \mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-\frac{1}{2}}\mathbf{\Sigma}_{22}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}^{\dagger}\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\right)\mathbf{\Sigma}_{11}^{\frac{1}{2}} ,\end{aligned}$$

which is independent of $\mathbf{x}_2$;

(e) from the proof below:

Applying Lemma 2.1, we have that

$$\mathbf{0} \preceq \mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger} \preceq (1-\rho^2)\mathbf{I}_{M_1} .$$

Then, by [22, p.470],

$$\begin{aligned}\mathbf{H}_{12}\mathbf{\Sigma}_{11}^{\frac{1}{2}}\left(\mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}\right)\mathbf{\Sigma}_{11}^{\frac{1}{2}}\mathbf{H}_{12}^{\dagger} &\preceq \mathbf{H}_{12}\mathbf{\Sigma}_{11}^{\frac{1}{2}}\left(1-\rho^2\right)\mathbf{I}_{M_1}\mathbf{\Sigma}_{11}^{\frac{1}{2}}\mathbf{H}_{12}^{\dagger} \\ &= \left(1-\rho^2\right)\mathbf{H}_{12}\mathbf{\Sigma}_{11}\mathbf{H}_{12}^{\dagger} .\end{aligned}$$

It follows that

$$\mathbf{I} + \mathbf{H}_{12}\mathbf{\Sigma}_{11}^{\frac{1}{2}}\left(\mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}\right)\mathbf{\Sigma}_{11}^{\frac{1}{2}}\mathbf{H}_{12}^{\dagger} \preceq \mathbf{I} + \left(1-\rho^2\right)\mathbf{H}_{12}\mathbf{\Sigma}_{11}\mathbf{H}_{12}^{\dagger} .$$

Observing that both sides in the above expression are positive definite, we have that

$$0 < \det\left(\mathbf{I} + \mathbf{H}_{12}\mathbf{\Sigma}_{11}^{\frac{1}{2}}\left(\mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}\right)\mathbf{\Sigma}_{11}^{\frac{1}{2}}\mathbf{H}_{12}^{\dagger}\right) \leq \det\left(\mathbf{I} + \left(1-\rho^2\right)\mathbf{H}_{12}\mathbf{\Sigma}_{11}\mathbf{H}_{12}^{\dagger}\right) .$$

We conclude that

$$\log\left[\det\left(\mathbf{I} + \mathbf{H}_{12}\mathbf{\Sigma}_{11}^{\frac{1}{2}}\left(\mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}\right)\mathbf{\Sigma}_{11}^{\frac{1}{2}}\mathbf{H}_{12}^{\dagger}\right)\right] \leq \log\left[\det\left(\mathbf{I} + \left(1-\rho^2\right)\mathbf{H}_{12}\mathbf{\Sigma}_{11}\mathbf{H}_{12}^{\dagger}\right)\right] .$$

In summary, we have shown that

$$\mathbf{I}(\mathbf{X_1}; \mathbf{Y_1}, \mathbf{Y}|\mathbf{X_2}) \leq \log\left[\det\left(\mathbf{I} + \left(1-\rho^2\right)\mathbf{H}_{12}\mathbf{\Sigma}_{11}\mathbf{H}_{12}^{\dagger}\right)\right] . \tag{2.20}$$

Next, we turn the attention to the sum rate of the MAC part. Observe that

$$
\begin{aligned}
& \mathbf{I}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}) \\
= \;& h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}_1, \mathbf{X}_2) \\
= \;& h\left(\left[\sqrt{\eta_2}\mathbf{H}_2, \sqrt{\eta_3}\mathbf{H}_3 \right] \left[\begin{array}{c} \mathbf{X}_1 \\ \\ \mathbf{X}_2 \end{array} \right] + \mathbf{Z} \right) - h(\mathbf{Z}) \\
= \;& h\left(\left[\sqrt{\eta_2}\mathbf{H}_2, \sqrt{\eta_3}\mathbf{H}_3 \right] \left[\begin{array}{c} \mathbf{X}_1 \\ \\ \mathbf{X}_2 \end{array} \right] + \mathbf{Z} \right) - \log[(\pi e)^N \det(\mathbf{I}_N)] \\
\overset{(a)}{\leq} \;& \log \left[(\pi e)^N \det \left(\mathrm{Cov} \left(\left[\sqrt{\eta_2}\mathbf{H}_2, \sqrt{\eta_3}\mathbf{H}_3 \right] \left[\begin{array}{c} \mathbf{X}_1 \\ \\ \mathbf{X}_2 \end{array} \right] + \mathbf{Z} \right) \right) \right] - \log((\pi e)^N) \\
= \;& \log \left[\det \left(\mathbf{I} + \left[\sqrt{\eta_2}\mathbf{H}_2, \sqrt{\eta_3}\mathbf{H}_3 \right] \left[\begin{array}{cc} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{array} \right] \left[\begin{array}{c} \sqrt{\eta_2}\mathbf{H}_2^\dagger \\ \\ \sqrt{\eta_3}\mathbf{H}_3^\dagger \end{array} \right] \right) \right] \\
= \;& \log \left[\det \left(\mathbf{I} + \eta_2 \mathbf{H}_2 \mathbf{\Sigma}_{11} \mathbf{H}_2^\dagger + \eta_3 \mathbf{H}_3 \mathbf{\Sigma}_{22} \mathbf{H}_3^\dagger + \sqrt{\eta_2\eta_3}\mathbf{H}_2 \mathbf{\Sigma}_{12} \mathbf{H}_3^\dagger + \sqrt{\eta_2\eta_3}\mathbf{H}_3 \mathbf{\Sigma}_{21} \mathbf{H}_2^\dagger \right) \right] \qquad (2.21) \\
\overset{(b)}{\leq} \;& \inf_{a>0} \log \left[\det \left(\mathbf{I} + \eta_2 \mathbf{H}_2 \mathbf{\Sigma}_{11} \mathbf{H}_2^\dagger + \eta_3 \mathbf{H}_3 \mathbf{\Sigma}_{22} \mathbf{H}_3^\dagger + \frac{\rho}{a} \sqrt{\eta_2\eta_3}\mathbf{H}_2 \mathbf{\Sigma}_{11} \mathbf{H}_2^\dagger + a\sqrt{\eta_2\eta_3}\mathbf{H}_3 \mathbf{\Sigma}_{22} \mathbf{H}_3^\dagger \right) \right] \\
= \;& \inf_{a>0} \log \left[\det \left(\mathbf{I} + (\eta_2 + \frac{\rho}{a}\sqrt{\eta_2\eta_3}) \mathbf{H}_2 \mathbf{\Sigma}_{11} \mathbf{H}_2^\dagger + (\eta_3 + a\sqrt{\eta_2\eta_3}) \mathbf{H}_3 \mathbf{\Sigma}_{22} \mathbf{H}_3^\dagger \right) \right], \qquad (2.22)
\end{aligned}
$$

where

(a) from the fact that the circularly-symmetric complex Gaussian distribution maximizes entropy [41] as mentioned before;

(b) can be shown as follows: First, by Lemma 2.2, we have that for $\forall\ a > 0$,

$$
\sqrt{\eta_2\eta_3}\mathbf{H}_2 \mathbf{\Sigma}_{12} \mathbf{H}_3^\dagger + \sqrt{\eta_2\eta_3}(\mathbf{H}_2 \mathbf{\Sigma}_{12} \mathbf{H}_3^\dagger)^\dagger \preceq \frac{\rho}{a}\eta_2 \mathbf{H}_2 \mathbf{\Sigma}_{11} \mathbf{H}_2^\dagger + a\eta_3 \mathbf{H}_3 \mathbf{\Sigma}_{22} \mathbf{H}_3^\dagger \ .
$$

Next, following the same procedures above, it is easy to show that

$$\begin{aligned} & \log\left[\det\left(\mathbf{I}+\eta_2\mathbf{H}_2\mathbf{\Sigma}_{11}\mathbf{H}_2^{\dagger}+\eta_3\mathbf{H}_3\mathbf{\Sigma}_{22}\mathbf{H}_3^{\dagger}+\sqrt{\eta_2\eta_3}\mathbf{H}_2\mathbf{\Sigma}_{12}\mathbf{H}_3^{\dagger}+\sqrt{\eta_2\eta_3}\mathbf{H}_3\mathbf{\Sigma}_{21}\mathbf{H}_2^{\dagger}\right)\right] \\ \leq & \log\left[\det\left(\mathbf{I}+\eta_2\mathbf{H}_2\mathbf{\Sigma}_{11}\mathbf{H}_2^{\dagger}+\eta_3\mathbf{H}_3\mathbf{\Sigma}_{22}\mathbf{H}_3^{\dagger}+\frac{\rho^2}{a}\sqrt{\eta_2\eta_3}\mathbf{H}_2\mathbf{\Sigma}_{11}\mathbf{H}_2^{\dagger}+a\sqrt{\eta_2\eta_3}\mathbf{H}_3\mathbf{\Sigma}_{22}\mathbf{H}_3^{\dagger}\right)\right] . \end{aligned}$$

Lastly, taking the infimum on both sides yields (2.22). ∎

2.2.2.2. *Proof of Theorem 2.2*

Proof: If there is no relay node, the channel is a point-to-point MIMO channel, and the corresponding channel capacity is given by

$$C_d^G \triangleq \max_{\mathbf{\Sigma}_{11}} \log\left[\det\left(\mathbf{I}_N+\eta_2\mathbf{H}_2\mathbf{\Sigma}_{11}\mathbf{H}_2^{\dagger}\right)\right]. \tag{2.23}$$

Consider the case where the destination node treats the signal from the source node as noise. In this case, the source node optimizes its transmission only for the source-relay link, and the relay node optimizes the transmission corresponding to the relay-destination link. That is, the channel boils down to be a cascaded channel. The following information rate is achievable for the source-relay link:

$$C_3^G = \max_{\mathbf{\Sigma}_{11}} \log\left[\det\left(\mathbf{I}_{N_1}+\eta_1\mathbf{H}_1\mathbf{\Sigma}_{11}\mathbf{H}_1^{\dagger}\right)\right], \tag{2.24}$$

and the corresponding optimal covariance matrix is $\mathbf{\Sigma}_{11}^*$. For the relay-destination link, the received signal at the destination is distorted by both noise and the signal from the source node. Since the (optimal) signal from the source is also Gaussian, with covariance matrix $\mathbf{\Sigma}_{11}^*$, the covariance matrix for noise plus the source signal is $\mathbf{I}_N+\eta_2\mathbf{H}_2\mathbf{\Sigma}_{11}^*\mathbf{H}_2^{\dagger}$. Therefore,

the achievable information rate for the relay-destination link is given by

$$\begin{aligned} C_4^G &= \max_{\mathbf{\Sigma}_{22}} \log \left[\det \left(\left(\mathbf{I}_N + \eta_2 \mathbf{H}_2 \mathbf{\Sigma}_{11}^* \mathbf{H}_2^\dagger + \eta_3 \mathbf{H}_3 \mathbf{\Sigma}_{22} \mathbf{H}_3^\dagger \right) \left(\mathbf{I}_N + \eta_2 \mathbf{H}_2 \mathbf{\Sigma}_{11}^* \mathbf{H}_2^\dagger \right)^{-1} \right) \right] \\ &= \max_{\mathbf{\Sigma}_{22}} \log \left[\det \left(\mathbf{I}_N + \eta_3 \mathbf{H}_3 \mathbf{\Sigma}_{22} \mathbf{H}_3^\dagger \left(\mathbf{I}_N + \eta_2 \mathbf{H}_2 \mathbf{\Sigma}_{11}^* \mathbf{H}_2^\dagger \right)^{-1} \right) \right]. \end{aligned} \tag{2.25}$$

∎

The lower bound follows by combining (2.23) with (2.24)and (2.25).

2.3. Capacity Bounds: The Rayleigh Fading Case

Now consider channel models where all the channel gain matrices are random, and suppose that the channel gains are known at the corresponding receivers only. In this scenario, we study the ergodic capacity of the MIMO relay channel with receiver CSI only. Simply put, the ergodic capacity is the highest achievable data rate by coding the transmission symbols over infinitely many blocks [60, p.11].

Theorem 2.3 ***[Rayleigh Fading Case]*** *An upper bound on the ergodic capacity of the MIMO relay channel is given by*

$$C^R \leq C_{upper}^R = \min(C_1^R,\ C_2^R)\ , \tag{2.26}$$

with

$$C_1^R \triangleq \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_{M_1} + \eta_1 \mathbf{H}_1^\dagger \mathbf{H}_1 + \eta_2 \mathbf{H}_2^\dagger \mathbf{H}_2 \right) \right] \tag{2.27}$$

$$C_2^R \triangleq \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + \eta_2 \mathbf{H}_2 \mathbf{H}_2^\dagger + \eta_3 \mathbf{H}_3 \mathbf{H}_3^\dagger \right) \right], \tag{2.28}$$

where the expectations are taken with respect to channel matrices $\mathbf{H}_1$, $\mathbf{H}_2$ *and* $\mathbf{H}_3$.

Proof: Because of fading, the channel matrices are now random. Then it follows that

$$C^R \leq C^R_{upper} = \max_{p(\mathbf{x}_1,\mathbf{x}_2)} \min(\mathbb{E}_{\mathbf{H}}\mathbf{I}(\mathbf{X}_1;\mathbf{Y},\mathbf{Y}_1|\mathbf{X}_2,\mathbf{H}_1,\mathbf{H}_2),\ \mathbb{E}_{\mathbf{H}}\mathbf{I}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}|\mathbf{H}_2,\mathbf{H}_3)), \quad (2.29)$$

where the expectations are taken with respect to corresponding channel coefficients.

In the proof for Theorem 2.1, we have shown that

$$C^G_{upper} = \max_{\mathbf{\Sigma}_{11},\mathbf{\Sigma}_{22},\mathbf{\Sigma}_{12}} \min\left(C'_1, C'_2\right), \quad (2.30)$$

where

$$C'_1 \triangleq \log\left[\det\left(\mathbf{I} + \mathbf{H}_{12}\left(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}^{\dagger}\right)\mathbf{H}_{12}^{\dagger}\right)\right] \quad (2.31)$$

$$C'_2 \triangleq \log\left[\det\left(\mathbf{I} + \mathbf{H}_{23}\begin{bmatrix}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22}\end{bmatrix}\mathbf{H}_{23}^{\dagger}\right)\right] \quad (2.32)$$

with $\mathbf{H}_{23} \triangleq \left[\sqrt{\eta_2}\mathbf{H}_2, \sqrt{\eta_3}\mathbf{H}_3\right]$.

Observe the power constraints on input signals: $\mathbb{E}(\mathbf{X}_1^{\dagger}\mathbf{X}_1) = \mathrm{tr}(\mathbf{\Sigma}_{11}) \leq M_1$ and $\mathbb{E}(\mathbf{X}_2^{\dagger}\mathbf{X}_2) = \mathrm{tr}(\mathbf{\Sigma}_{22}) \leq M_2$. Then it follows that

$$\mathrm{tr}\left(\begin{bmatrix}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22}\end{bmatrix}\right) = \mathrm{tr}(\mathbf{\Sigma}_{11}) + \mathrm{tr}(\mathbf{\Sigma}_{22}) \leq M_1 + M_2.$$

Along the same lines of [26], we conclude that the optimal signal covariance matrices which maximize $\mathbb{E}_{\mathbf{H}}(C'_2)$ are identity matrices, i.e., $\mathbf{\Sigma}_{11} = \mathbf{I}_{M_1}$, $\mathbf{\Sigma}_{22} = \mathbf{I}_{M_2}$ and $\mathbf{\Sigma}_{12} = \mathbf{0}$. Therefore,

$$\mathbb{E}_{\mathbf{H}}(C'_2) \leq \mathbb{E}_{\mathbf{H}} \log\left[\det\left(\mathbf{I} + \mathbf{H}_{23}\mathbf{H}_{23}^{\dagger}\right)\right] = \mathbb{E}_{\mathbf{H}} \log\left[\det\left(\mathbf{I}_N + \eta_2\mathbf{H}_2\mathbf{H}_2^{\dagger} + \eta_3\mathbf{H}_3\mathbf{H}_3^{\dagger}\right)\right]. \quad (2.33)$$

Note that $\mathbf{H}_{12}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}^{\dagger}\mathbf{H}_{12}^{\dagger}$ is non-negative with probability 1 [19, Theorem 3.2]. Along the same line of the proof of Theorem 2.1, we have that

$$\mathbb{E}_{\mathbf{H}}(C'_1) \leq \mathbb{E}_{\mathbf{H}} \log\left[\det\left(\mathbf{I} + \mathbf{H}_{12}\mathbf{\Sigma}_{11}\mathbf{H}_{12}^{\dagger}\right)\right].$$

Furthermore, by [41, Theorem 1], $\mathbf{\Sigma}_{11} = \mathbf{I}_{M_1}$ can also maximize the RHS of the above. We conclude that the covariance matrices that maximize $\mathbb{E}_{\mathbf{H}}(C_2')$ can also maximize $\mathbb{E}_{\mathbf{H}}(C_1')$, i.e.,

$$\mathbb{E}_{\mathbf{H}}(C_1') \leq \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I} + \mathbf{H}_{12} \mathbf{H}_{12}^{\dagger} \right) \right] = \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + \eta_1 \mathbf{H}_1^{\dagger} \mathbf{H}_1 + \eta_2 \mathbf{H}_2^{\dagger} \mathbf{H}_2 \right) \right]. \quad (2.34)$$

In a nutshell, the mutual information rates in (2.29) are maximized by choosing $\mathbf{X}_1$ and $\mathbf{X}_2$ to be independent circular-symmetric complex Gaussian vectors with $\mathbf{\Sigma}_{11} = \mathbf{I}_{M_1}$, $\mathbf{\Sigma}_{22} = \mathbf{I}_{M_2}$ and $\mathbf{\Sigma}_{12} = \mathbf{0}$. ∎

In what follows, we present a lower bound on the ergodic capacity.

Theorem 2.4 ***[Rayleigh Fading Case]*** *A lower bound on the ergodic capacity of the MIMO relay channel is given by*

$$C^R \geq C^R_{lower} = \max(C^R_d, \ \min(C^R_3, \ C^R_2)) \ , \quad (2.35)$$

with

$$C^R_d \triangleq \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N \ + \eta_2 \mathbf{H}_2 \mathbf{H}_2^{\dagger} \right) \right] \quad (2.36)$$

$$C^R_3 \triangleq \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_{N_1} + \eta_1 \mathbf{H}_1 \mathbf{H}_1^{\dagger} \right) \right] \ , \quad (2.37)$$

where the expectations are taken with respect to corresponding channel matrices.

Proof: Based on [6, Section VI], the following rate is achievable by using block Markov coding

$$R = \max_{p(\mathbf{x}_1, \mathbf{x}_2)} \min(\mathbf{I}(\mathbf{X}_1; \mathbf{Y}_1 | \mathbf{X}_2) \ , \ \mathbf{I}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y})), \quad (2.38)$$

Since the receivers have full CSI, it follows that

$$\mathbf{I}(\mathbf{X}_1;\mathbf{Y}_1|\mathbf{X}_2) = \mathbb{E}_{\mathbf{H}}[\mathbf{I}(\mathbf{X}_1;\mathbf{Y}_1|\mathbf{X}_2,\mathbf{H}_1)] \quad (2.39)$$

$$\mathbf{I}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}\ ) = \mathbb{E}_{\mathbf{H}}[\mathbf{I}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}|\mathbf{H}_2,\mathbf{H}_3)]. \quad (2.40)$$

where the expectations are taken with respect to the corresponding channel matrices. Note that

$$\mathbf{I}(\mathbf{X}_1;\mathbf{Y}_1|\mathbf{X}_2,\mathbf{H}_1) \leq \log\left[\det\left(\mathbf{I}+\mathbf{H}_1\left(\mathbf{\Sigma}_{11}-\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}^{\dagger}\right)\mathbf{H}_1^{\dagger}\right)\right].$$

Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be independent circularly-symmetric complex Gaussian random vectors with $\mathbf{\Sigma}_{11}=\mathbf{I}_{M_1}$, $\mathbf{\Sigma}_{22}=\mathbf{I}_{M_2}$ and $\mathbf{\Sigma}_{12}=\mathbf{0}$. Then, the lower bound in (2.35) follows along the same line of the proof of Theorem 2.3.

∎

[**Remarks**:] Interestingly, Theorems 2.3 and 2.4 reveal that in a Rayleigh fading channel, the corresponding codebooks at the source node and the relay node are independent, i.e., $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent. (As shown below, the upper bound and the lower bound can "converge" under certain conditions, indicating that the ergodic capacity of the MIMO relay channel can be characterized exactly). In contrast, $\mathbf{X}_1$ and $\mathbf{X}_2$ are correlated for the fixed channel cases. Indeed, the capacity of the MAC part is achieved when the source node and the relay node have "complete" cooperation for the fixed channel case [6]. Our intuition for this finding is as follows: Since we consider the relay channel with receiver CSI only, the transmitters have no knowledge about the channel realizations. As a result, the optimal codebooks at the source node and the relay node are independent, due to the channel uncertainty at the transmitters. Intuitively speaking, *it is the uncertainty (randomness) of* $\mathbf{H}_1$ *and* $\mathbf{H}_2$ *at the transmitters that makes* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *independent.*

Recall that in a single-user MIMO channel with receiver CSI only, the capacity is achieved when the power allocation across transmit antennas is equal and the signals are independent [41, Theorem 1]. If we treated the relay node and the source node as an antenna-clustering transmitter [13], the optimal signaling would indicate independent signals across transmit antennas.

It is clear that the communications between the source node and the destination node can be improved by using relaying (see, e.g., Case III in Section 2.3.1). From Equations (2.28) and (2.27), the capacity gain for the MAC part is due to the multi-access gain; and the capacity gain for the BC part originates from the broadcast gain. Needless to say, a key to reap the capacity gains is to develop coding strategies for the cooperative multi-access channel and the cooperative broadcast channel therein.

2.3.1. Numerical Examples

We now illustrate via numerical examples the bounds in Theorems 2.3 and 2.4. For the sake of clarity, we consider the case where the numbers of antennas at all the transmitters and receivers are equal (e.g., in ad hoc networks, all the nodes are equipped with identical RF devices). We study the upper bound and the lower bound with different SNR parameters (namely η_1, η_2 and η_3). The SNR parameters play a key role in the upper bound and the lower bound. In what follows, we study three cases with different SNR parameters. The number of the antennas is assumed to be 2 in all cases, and $\eta_2 = SNR_2/2$, with SNR_2 being the SNR for the direct link.

[**Case I:**] In this case, $\eta_1 = \eta_2 = \eta_3$; this models the scenario where the source node, the relay node and the destination node are separated by equal distances. Fig. 2.3 depicts the

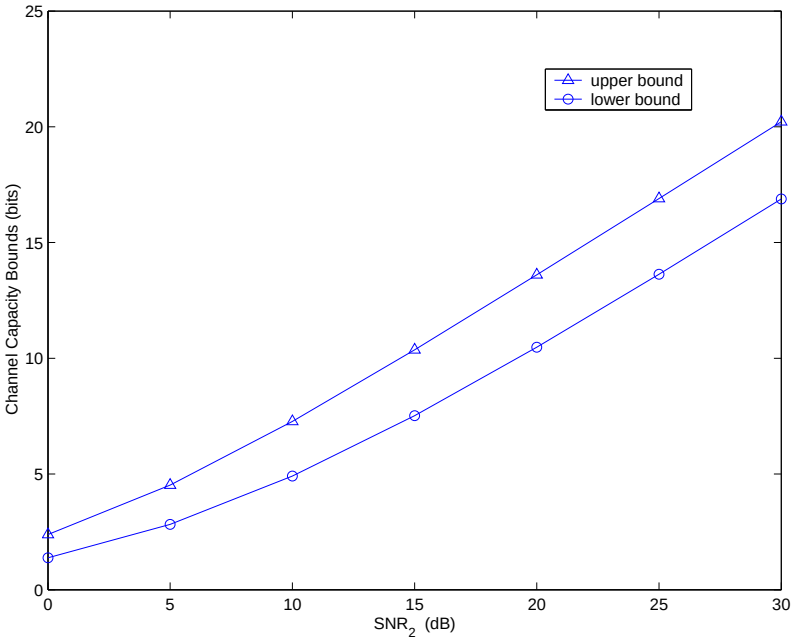

Figure 2.3. Capacity bounds vs. SNR_2 for the case $\eta_1 = \eta_2 = \eta_3$

upper bound and the lower bound.

[**Case II:**] In this case, $\eta_1 = \eta_2$ and $\eta_3 = 10\eta_2$, which "captures" that the relay node is closer to the destination node than to the source node. Comparing Fig. 2.4 with Fig. 2.3, we observe that the upper bound and the lower bound for Case I and Case II are the same. We will elaborate further on this in Section 2.4 (see Lemma 2.5).

[**Case III:**] In this case, $\eta_2 = \eta_3$ and $\eta_1 = 10\eta_2$, which "captures" a scenario that the relay node is closer to the source node than to the destination node. Surprisingly, as shown in Fig. 2.5, the upper bound and the lower bound "converge". That is to say, the ergodic capacity of the MIMO relay channel over Rayleigh fading can be characterized under this SNR condition. We will discuss in Section 2.4 sufficient conditions under which the ergodic capacity can be achieved.

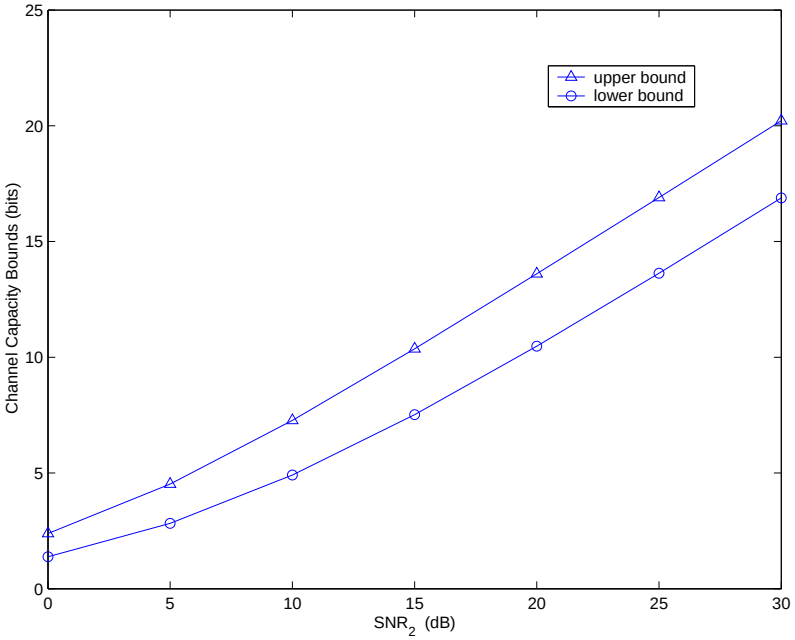

Figure 2.4. Capacity bounds vs. SNR_2 for the case $\eta_1 = \eta_2$ and $\eta_3 = 10\eta_1$

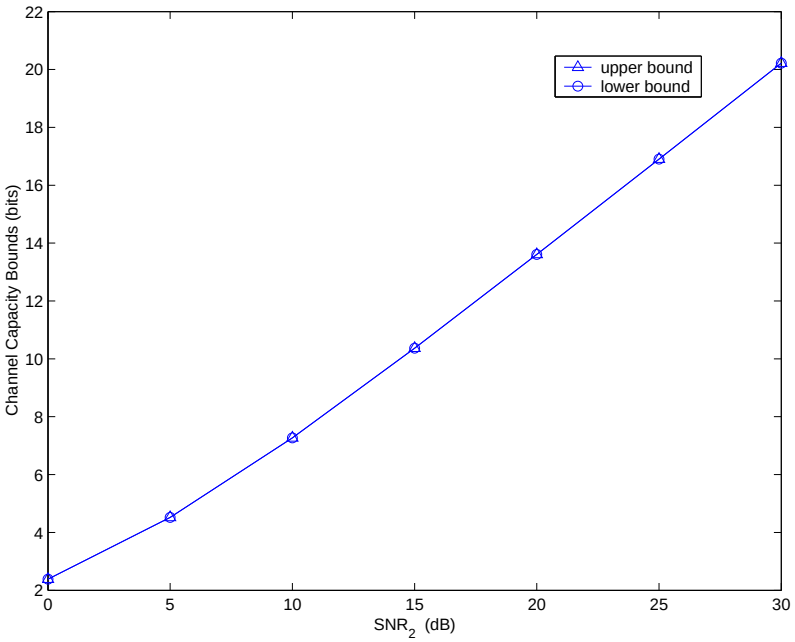

Figure 2.5. Capacity bounds vs. SNR_2 for the case $\eta_2 = \eta_3$ and $\eta_1 = 10\eta_2$

2.4. Discussions on Capacity Achievability

The Gaussian MIMO relay channel with fixed channel gains can be viewed as a vector generalization of the classical Gaussian relay channel in [6], and is not degraded in general. Characterizing the corresponding capacity remains open. In the following, we turn our attention to the fading case. Specifically, we investigate in Section 2.4.1 sufficient conditions that give exact characterization of the ergodic capacity. Since the ergodic capacity involves expectations with respect to random matrices and does not admit an "explicit" expression, we study in Section 2.4.2 the high SNR regime and use approximations to identify SNR conditions for achieving the capacity; in Section 2.4.3, we examine the scale case, for which we derive the explicit conditions for capacity achievability and the explicit capacity expression.

2.4.1. Regularity Conditions for Capacity Achievability

In Section 2.3, we have presented a lower bound and an upper bound on the ergodic capacity of the MIMO relay channel and also provide numerical examples for different SNR cases. Interestingly, the upper bound and the lower bound in Fig. 2.5 "converge", which indicates that under certain regularity conditions, the ergodic capacity of the MIMO relay channel can be characterized exactly. Indeed, we observe that in (2.26) and (2.35), there is a common term C_2^R. If C_2^R is smaller than both C_1^R and C_3^R, the ergodic capacity of the MIMO relay channel is given by C_2^R. We state this important observation in the following proposition.

Proposition 2.1 *If $C_1^R \geq C_2^R$ and $C_3^R \geq C_2^R$, then the ergodic capacity of the MIMO relay channel is given by C_2^R.*

In what follows, we study the conditions in Proposition 2.1 in terms of SNR parameters. For tractability, we assume that all the numbers of antennas are equal, i.e., $M_1 = M_2 = N_1 = N$ (for simplicity, we use N to denote the number of the antennas). We first need the following lemma.

Lemma 2.3 *The upper bound on the ergodic capacity of the MIMO relay channel is* C_2^R, *if* $\eta_1 \geq \eta_3$.

Proof: Under the assumption on equal numbers of antennas, it follows that $\mathbf{H}_i^\dagger \mathbf{H}_i$ and $\mathbf{H}_i \mathbf{H}_i^\dagger$ have identical probability distributions, for $i = 1, 2, 3$. Therefore,

$$\begin{aligned} C_1^R &= \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + \eta_1 \mathbf{H}_1 \mathbf{H}_1^\dagger + \eta_2 \mathbf{H}_2 \mathbf{H}_2^\dagger \right) \right] \\ &= \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + \eta_1 \mathbf{H}_3 \mathbf{H}_3^\dagger + \eta_2 \mathbf{H}_2 \mathbf{H}_2^\dagger \right) \right]. \end{aligned}$$

The second equality follows the fact that $\mathbf{H}_1 \mathbf{H}_1^\dagger$, $\mathbf{H}_2 \mathbf{H}_2^\dagger$ and $\mathbf{H}_3 \mathbf{H}_3^\dagger$ are i.i.d.Wishart matrices, and accordingly $(\mathbf{H}_1 \mathbf{H}_1^\dagger, \mathbf{H}_2 \mathbf{H}_2^\dagger)$ and $(\mathbf{H}_3 \mathbf{H}_3^\dagger, \mathbf{H}_2 \mathbf{H}_2^\dagger)$ follow identically probability laws. Furthermore, since C_1^R increases monotonically with η_1, we have that $C_1^R \geq C_2^R$ if $\eta_1 \geq \eta_3$.

∎

Lemma 2.3 gives the conditions under which the upper bound is C_2^R. It remains to examine when C_2^R would meet the lower bound, that is, $C_2^R \leq C_3^R$(it is easy to show that C_2^R is always greater than C_d^R), and thus the ergodic capacity can be characterized. In general, it is non-trivial to determine in terms of η_1, η_2 and η_3 if $C_2^R \leq C_3^R$ (except the scalar case), because the ergodic capacity expressions involve expectations with respect to random matrices. In light of this fact, we first find an upper bound on C_2^R and compare it with C_3^R. The following lemma provides an upper bound on C_2^R.

Lemma 2.4 *For any* $\eta_2 \geq 0$ *and* $\eta_3 \geq 0$, $C_2^R \leq \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + \frac{\eta_2+\eta_3}{2} \left(\mathbf{H}_2 \mathbf{H}_2^\dagger + \mathbf{H}_3 \mathbf{H}_3^\dagger \right) \right) \right]$.

Proof: First, we rewrite C_2^R as

$$C_2^R = \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + [\mathbf{H}_2, \mathbf{H}_3] \begin{bmatrix} \eta_2 \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \eta_3 \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{H}_2^\dagger \\ \mathbf{H}_3^\dagger \end{bmatrix} \right) \right].$$

Along the same line of the proof of [41, Theorem 1], it can be shown that if the total power is kept constant, i.i.d. input signals with the equal power allocation can maximize the mutual information. It then follows that

$$\begin{aligned} C_2^R &\leq \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + [\mathbf{H}_2, \mathbf{H}_3] \begin{bmatrix} \frac{\eta_2 N + \eta_3 N}{2N} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \frac{\eta_2 N + \eta_3 N}{2N} \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{H}_2^\dagger \\ \mathbf{H}_3^\dagger \end{bmatrix} \right) \right] \\ &= \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + \frac{\eta_2 + \eta_3}{2} [\mathbf{H}_2, \mathbf{H}_3] \begin{bmatrix} \mathbf{H}_2^\dagger \\ \mathbf{H}_3^\dagger \end{bmatrix} \right) \right]. \end{aligned}$$

It is clear that the equality can be achieved if $\eta_2 = \eta_3$. ∎

In Section 2.4.2, we will derive sufficient conditions for capacity achievability by combining Proposition 2.1, Lemmas 2.3 and 2.4, building on which we elaborate further on the numerical results exhibited in Case III in Section 2.3.1.

Next, we present in Lemma 2.5 the conditions under which the upper bound and the lower bound diverge. This sheds light on the existence of the gap between the upper bound and the lower bound in Fig. 2.3 and Fig. 2.4.

Lemma 2.5 *If* $\eta_1 = \eta_2$ *and* $\eta_1 \leq \eta_3$, *then the upper bound on the ergodic capacity of the MIMO relay channel is* C_1^R, *and the lower bound is* C_3^R.

Proof: The first statement directly follows Lemma 2.3, and it remains to show the second one. We have that

$$\begin{aligned} C_2^R &= \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + \eta_2 \mathbf{H}_2 \mathbf{H}_2^\dagger + \eta_3 \mathbf{H}_3 \mathbf{H}_3^\dagger \right) \right] \\ &\overset{(a)}{\geq} \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + \eta_2 \mathbf{H}_2 \mathbf{H}_2^\dagger \right) \right] \\ &\overset{(b)}{=} \mathbb{E}_{\mathbf{H}} \log \left[\det \left(\mathbf{I}_N + \eta_1 \mathbf{H}_1 \mathbf{H}_1^\dagger \right) \right], \end{aligned}$$

where (a) follows that $\mathbf{H}_3 \mathbf{H}_3^\dagger \succ \mathbf{0}$ with probability 1 [19, Theorem 3.2]; and (b) follows from the facts that $\eta_1 = \eta_2$, and $\mathbf{H}_1 \mathbf{H}_1^\dagger$ and $\mathbf{H}_2 \mathbf{H}_2^\dagger$ are i.i.d. random matrices, as aforementioned. ∎

Next, we examine two cases where the capacity-achieving conditions can be expressed in explicit form.

2.4.2. Achievability of Ergodic Capacity: The High SNR Regime

In the high SNR regime [34, Proposition 2], i.e., η_1, η_2 and η_3 are large, C_3^R can be approximated as

$$C_3^R \approx N \log_2(\eta_1) + \frac{1}{\ln 2} \left(\sum_{j=1}^{N} \sum_{p=1}^{N-j} \frac{1}{p} - \gamma N \right), \tag{2.41}$$

where $\gamma \approx 0.57721566$ is Euler's constant. The same approximation can be applied to the upper bound on C_2^R in Lemma 2.4; i.e.,

$$C_2^R \leq N \log_2(\frac{\eta_2 + \eta_3}{2}) + \frac{1}{\ln 2} \left(\sum_{j=1}^{N} \sum_{p=1}^{2N-j} \frac{1}{p} - \gamma N \right). \tag{2.42}$$

Comparing the RHS of (2.41) with that of (2.42), we note that $C_2^R \leq C_3^R$ if the following condition holds:

$$N \log_2\left(\frac{2\eta_1}{\eta_2+\eta_3}\right) \geq \frac{1}{\ln 2} \sum_{j=1}^{N} \sum_{p=N+1-j}^{2N-j} \frac{1}{p}, \tag{2.43}$$

or equivalently,

$$\boxed{\frac{\eta_1}{\eta_2+\eta_3} \geq 2^{\frac{q}{N}-1}, \quad \text{where} \quad q \triangleq \frac{1}{\ln 2} \sum_{j=1}^{N} \sum_{p=N+1-j}^{2N-j} \frac{1}{p} \quad . \tag{2.44}}$$

Combining (2.44) with the conditions in Lemma 2.3, we have that the ergodic capacity of the MIMO relay channel is C_2^R.

Recall that in Section 2.3.1, Case III reveals a somewhat surprising result that the upper bound meets the lower bound (in Case III, $\eta_2 = \eta_3$ and $N = 2$). Based on (2.44) and Lemma 2.3, it follows that if $\eta_1 \geq 3.2113\eta_2$ (which also implies $\eta_1 > \eta_3$), then $C_2^R < C_3^R$ and $C_2^R < C_1^R$. That is to say, the threshold value for this case is 3.2113, at which the upper bound and the lower bound "converge". Accordingly, in Fig. 2.5, where $\eta_1 = 10\eta_2$ and $\eta_2 = \eta_3$, the upper bound meets the lower bound and the ergodic capacity is achieved. To elaborate further on this, we present two different SNR parameters (cf. Case III in Section 2.3.1). As shown in Fig. 2.6, when $\eta_1 = 3\eta_2$, the upper bound is very close to the lower bound. Fig. 2.7 shows that the upper bound "meets" to the lower bound perfectly when $\eta_1 = 3.5\eta_2$, indicating that the capacity can be characterized exactly.

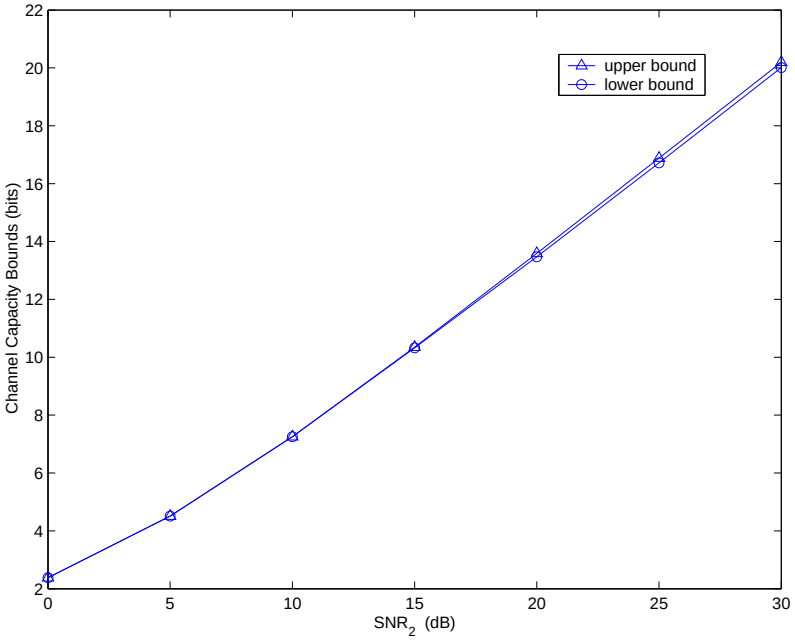

Figure 2.6. Capacity bounds vs. SNR_2 for the case of $\eta_2 = \eta_3$ and $\eta_1 = 3\eta_2$

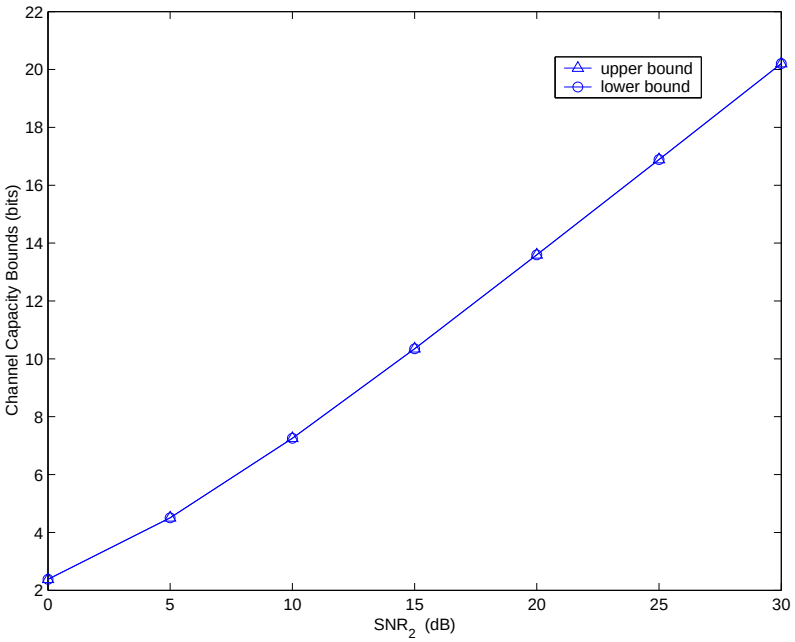

Figure 2.7. Capacity bounds vs. SNR_2 for the case of $\eta_2 = \eta_3$ and $\eta_1 = 3.5\eta_2$

2.4.3. Achievability of Ergodic Capacity: $N = 1$

If all the numbers of the antennas are 1, the MIMO relay channel boils down to a scalar relay channel. In this case,

$$\begin{aligned} C_2^R &= \mathbb{E}_{\mathbf{h}} \log \left[1 + \eta_2 |h_2|^2 + \eta_3 |h_3|^2\right], \\ C_3^R &= \mathbb{E}_{\mathbf{h}} \log \left[1 + \eta_1 |h_1|^2\right]. \end{aligned}$$

Since $|h_1|^2$ has χ^2 distribution with freedom of 2 (namely, exponential distribution), and $\mathbb{E}(|h_1|^2) = 1$, $X \triangleq \eta_1 |h_1|^2$ has the probability density function

$$p_X(x) = \frac{1}{\eta_1} \exp\left(-\frac{x}{\eta_1}\right), \quad x \geq 0.$$

We can find C_3^R by

$$C_3^R = \frac{1}{\ln 2} \int_0^\infty \ln(1+x) \frac{1}{\eta_1} \exp\left(-\frac{x}{\eta_1}\right) dx. \tag{2.45}$$

Using the integral in [17, p.568], we have that

$$C_3^R = -\frac{1}{\ln 2} e^{\frac{1}{\eta_1}} \operatorname{Ei}(-\frac{1}{\eta_1}), \tag{2.46}$$

where Ei(.) is the exponential integral function in [17, p.875].

Next, consider the random variable $Y \triangleq \eta_2 |h_2|^2 + \eta_3 |h_3|^2$. Since $\mathbb{E}(|h_2|^2) = \mathbb{E}(|h_3|^2) = 1$, we have that

$$\begin{aligned} p_Y(y) &= \int_0^y \frac{1}{\eta_2 \eta_3} \exp(-\frac{y-t}{\eta_3}) \exp(-\frac{t}{\eta_2}) dt \\ &= \frac{1}{\eta_2 - \eta_3} \left(e^{-\frac{y}{\eta_2}} - e^{-\frac{y}{\eta_3}}\right). \end{aligned} \tag{2.47}$$

Hence, C_2^R can be computed by

$$C_2^R = \frac{1}{\ln 2} \int_0^\infty \ln(1+y) \frac{1}{\eta_2 - \eta_3} \left(e^{-\frac{y}{\eta_2}} - e^{-\frac{y}{\eta_3}}\right) dy \tag{2.48}$$

$$= \frac{1}{\ln 2} \frac{1}{\eta_2 - \eta_3} \left[\eta_3 e^{\frac{1}{\eta_3}} \operatorname{Ei}(-\frac{1}{\eta_3}) - \eta_2 e^{\frac{1}{\eta_2}} \operatorname{Ei}(-\frac{1}{\eta_2})\right] \tag{2.49}$$

Combining (2.46) with (2.49), we conclude that if

$$\frac{1}{\eta_2-\eta_3}\left[\eta_3 e^{\frac{1}{\eta_3}}\,\mathrm{Ei}(-\frac{1}{\eta_3})-\eta_2 e^{\frac{1}{\eta_2}}\,\mathrm{Ei}(-\frac{1}{\eta_2})\right]\leq e^{\frac{1}{\eta_1}}\,\mathrm{Ei}(-\frac{1}{\eta_1}), \tag{2.50}$$

and $\eta_1 \geq \eta_3$, then the ergodic capacity of the relay channel is given by

$$\boxed{C_2^R=\frac{1}{\ln 2}\frac{1}{\eta_2-\eta_3}\left[\eta_3 e^{\frac{1}{\eta_3}}\,\mathrm{Ei}(-\frac{1}{\eta_3})-\eta_2 e^{\frac{1}{\eta_2}}\,\mathrm{Ei}(-\frac{1}{\eta_2})\right].}$$

2.5. An Application of MIMO Relay Channels in Cooperative Communications in Ad Hoc Networks

In what follows, we explore the utility of the MIMO relay channel for cooperative communications in ad hoc networks. We consider ad hoc networks using the IEEE 802.11 CSMA/CA standard. In such a context, the medium access control protocol uses the RTS (request-to-send) / CTS (clear-to-send) handshake to set up a communication link. More specifically, as shown in Fig. 2.8, source node S transmits a RTS packet to request the channel and destination node D replies with a CTS packet. If the RTS/CTS dialogue is successful, S and D begin their data communication, whereas all other nodes that hear either the RTS packet or the CTS packet are kept silent for a specified duration.

A key observation is that the silent node (node R in Fig. 2.8 within the shaded area) can be exploited to relay information. We use the capacity results on MIMO relay channel to characterize the performance gain therein over the direct transmission without relaying. The union of the transmission region of S and D is the so-called RTS/CTS reserved floor; R would have been kept silent while S communicates with D as RTS/CTS dialogue dictates; d is the distance between S and D. Note that if R lies in the shaded area depicted in Fig. 2.8,

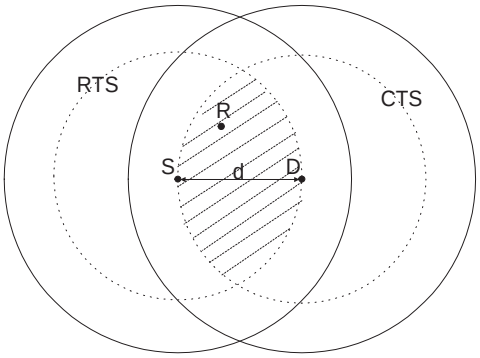

Figure 2.8. A sketch of RTS/CTS dialogue

it would have a shorter distance to both S and D than d. Thus, during the communication between S and D, R can function as a relay station to aid the data transmission. Thus motivated, we call this shaded area a "relay region", because any silent node within such a region can act as a relay node. Moreover, if R is equipped with multiple antennas, S, R and D form a MIMO relay channel.

Consider the Rayleigh fading channel. Define the relative gain of the capacity by using node R to relay data as $g \triangleq (C^R - C_d^R)/C_d^R$, where C^R is the relay channel capacity, and C_d^R is the channel capacity corresponding to the direct link given in Theorem 2.4. Applying Theorem 2.3 and Theorem 2.4, we have that

$$\frac{C_{lower}^R - C_d^R}{C_d^R} \leq \ g \leq \ \frac{C_{upper}^R - C_d^R}{C_d^R} \ . \tag{2.51}$$

Observe that η_1 and η_3 depend on the coordinates of the relay node. Without loss of generality, let the coordinates of S, D and R be $(-\frac{d}{2}, 0)$, $(\frac{d}{2}, 0)$ and (α, β). It follows that

$$\eta_1 = \eta_{10} \left[d/\sqrt{(\alpha + \frac{d}{2})^2 + \beta^2} \right]^n \quad \text{and} \quad \eta_3 = \eta_{30} \left[d/\sqrt{(\alpha - \frac{d}{2})^2 + \beta^2} \right]^n \ ,$$

where n is the path loss parameter in wireless links; and η_{10} and η_{30} are constants.

Since η_1 and η_3 change while the relay station moves, C_{lower} and C_{upper} are functions of (α, β) accordingly. Assume that R occurs with equal probability at any position within the shaded area (defined as $\mathcal{A}$) depicted in Fig. 2.8. Then it follows that

$$\begin{aligned}
\bar{C}^R_{upper} &= \frac{1}{S}\int\int_{(\alpha,\ \beta)\in\mathcal{A}_1} C^R_1(\alpha,\beta)\ d\alpha\, d\beta + \frac{1}{S}\int\int_{(\alpha,\ \beta)\in\mathcal{A}_2} C^R_2(\alpha,\beta)\ d\alpha\, d\beta, \\
\bar{C}^R_{lower} &= \frac{1}{S}\int\int_{(\alpha,\ \beta)\in\mathcal{A}_3} C^R_d(\alpha,\beta)\ d\alpha\, d\beta + \frac{1}{S}\int\int_{(\alpha,\ \beta)\in\mathcal{A}_4} \min(C^R_3(\alpha,\beta),\ C^R_2(\alpha,\beta))\ d\alpha\, d\beta\ ,
\end{aligned}$$

where S is the area of $\mathcal{A}$, given by $S = \left(\frac{2}{3}\pi - \frac{\sqrt{3}}{2}\right) d^2$; and $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$ are areas defined as

$$\begin{aligned}
\mathcal{A}_1 &\triangleq \left\{(\alpha,\beta)\,\middle|\, C^R_1 \le C^R_2\right\}\bigcap\mathcal{A}\ , \qquad \mathcal{A}_2 \triangleq \left\{(\alpha,\beta)\,\middle|\, C^R_1 > C^R_2\right\}\bigcap\mathcal{A}, \\
\mathcal{A}_3 &\triangleq \left\{(\alpha,\beta)\,\middle|\, C^R_d \ge \min(C^R_3,\ C^R_2)\right\}\bigcap\mathcal{A}\ , \quad \mathcal{A}_4 \triangleq \left\{(\alpha,\beta)\,\middle|\, C^R_d < \min(C^R_3,\ C^R_2)\right\}\bigcap\mathcal{A}\ .
\end{aligned}$$

By using the lower bound and upper bound on the ergodic capacity of the relay channel, we have the corresponding lower and upper bounds on average relaying gain $\bar{g}$:

$$\frac{\bar{C}^R_{lower} - C^R_d}{C^R_d} \ \le\ \bar{g} \le \frac{\bar{C}^R_{upper} - C^R_d}{C^R_d}. \tag{2.52}$$

CHAPTER 3

Achievable Rates and Scaling Laws of Power-constrained Wireless Sensory Relay Networks

3.1. System Model

In this section, we present the model for narrowband relay networks in the low SNR regime and that for power constrained wideband relay networks.

3.1.1. Narrowband Relay Networks in the Low SNR Regime

We impose the following assumptions for the narrowband relay network model.

- there are n relay nodes between the source node and the destination node. There is no direct link between the source node and the destination node. In the first hop, each relay node listens to the transmissions from the source node, and then forwards the processed signal to the destination node in the second hop;
- the communication bandwidth is B (Hz);
- the ambient noise at the relay nodes and the destination node has power spectral density N_0;

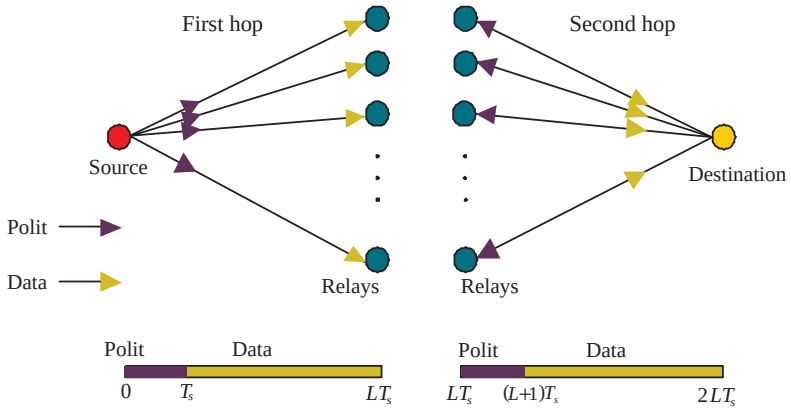

Figure 3.1. Amplify-and-forward with network training

- all nodes have an average transmit power constraint P (watts) in one fading block;
- all channels experience i.i.d. frequency-flat block fading. Denote the backward channel coefficients as $\{\alpha_i\}$, and the forward channel coefficients as $\{\beta_i\}$. We assume that $\{\alpha_i\}$ and $\{\beta_i\}$ are complex Gaussian random variables, with zero mean and unit variance;
- the backward channels have a coherence interval of L symbol periods, so do the forward channels. Throughout, we assume that $B = 1/T_s$. Accordingly, the channel coherence time is LT_s (seconds);
- within one fading block, the forward channel conditions are symmetric in both directions, that is, the channel condition from the relay node to the destination node is the same as that from the destination node to the relay node.

Without loss of generality, we assume that the average receive SNR for each backward (forward) link is P/BN_0, i.e, $\rho = P/(BN_0)$. It is understood that $\rho \to 0$ in the wideband regime. On the other hand, the coherence interval L points to how fast the channel changes. Clearly, a larger L is "favorable" for more accurate channel estimation, and we would expect

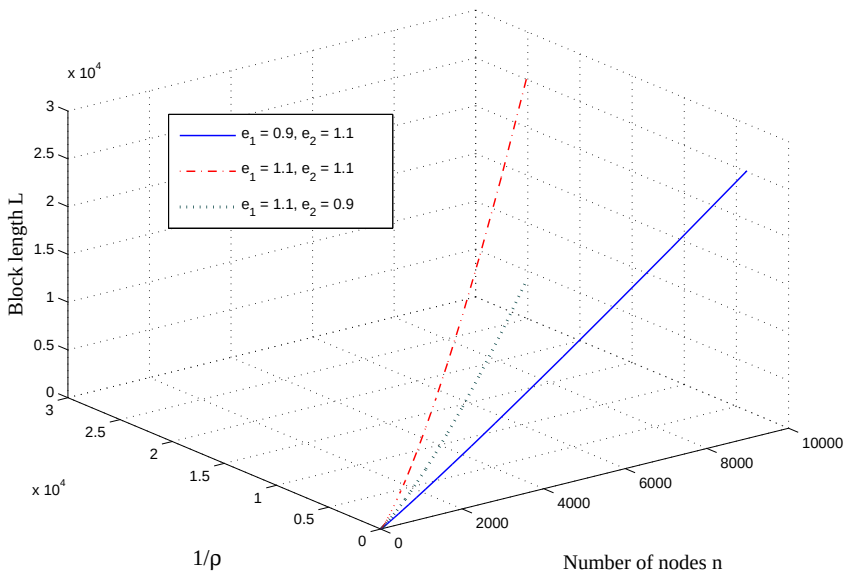

Figure 3.2. Asymptotic regime of $(n, 1/\rho, L)$

that when L exceeds a certain threshold, the achievable rates of the relay networks can scale the same as in coherent relaying case where a priori knowledge of CSI is available at relay nodes. In contrast, a low SNR makes channel estimation more challenging.

To obtain a clear understanding of the continuum between coherent relaying and non-coherent relaying, we let ρ and L scale with n. We investigate the achievable rates in the joint asymptotic regime of n, L and ρ. Accordingly, we can obtain a set of achievable rates corresponding to different values of ρ and L. Define

$$e_1 \triangleq \frac{\log(\frac{1}{\rho})}{\log n} = -\frac{\log \rho}{\log n} \quad \text{and} \quad e_2 \triangleq \frac{\log L}{\log n}. \tag{3.1}$$

Remarks: Because the focus of this study is on the scaling law, i.e., the scaling order of the capacity, it is natural to define e_1 and e_2 on the logarithmic scale. Although there is no physical connection among these three parameters, the achievable rates and their scaling behavior depend on the relation among these three key parameters, rather than on

each of them alone. Simply put, the asymptotic regime is along a curve determined by (n, e_1, e_2) in the three-dimensional space of $(n, 1/\rho, L)$, as illustrated in Fig. 3.2. Indeed, we show in Section 3.3 that the simultaneous scaling of the three parameters enables the characterization of the achievable rates and their scaling in terms of how they are related through the exponents.

3.1.2. Amplify-and-Forward (AF) with Network Training

As aforementioned, the channel estimation at relay nodes is carried out in the network training phase (see Fig. 3.1). More specifically, the source node first broadcasts common pilot symbols to relay nodes at the beginning of the first hop transmission, followed by its data transmission. Then the destination node broadcasts common pilot symbols at the beginning of the second hop. Based on the received signals corresponding to the pilot symbols, each relay nodes estimates its backward channel condition and its forward channel condition, respectively. After the channel estimation is done, each relay node carries out phase alignment by using its channel estimates, amplifies the received data signal under the given power constraints, and forwards the processed signal to the destination node.

3.1.2.1. *Channel Estimation via Network Training*

We assume that the minimum mean-square error (MMSE) estimation method is applied to estimate $\{\alpha_i\}$ and $\{\beta_i\}$. Let E_{total} denote the total energy for one hop transmission over L symbols, and E_{tr} be the energy for the training with $E_{tr} = \eta E_{total}$. Since the estimate

error of the MMSE estimator depends only on E_{tr} [42, p. 599], we further assume that the training can be done using one pilot sample (symbol). We assume that $L \gg 1$.

Consider the channel estimation for the ith channel in the first hop:

$$Y_i^p = \sqrt{E_{tr}}\alpha_i + Z_i \ , \tag{3.2}$$

where Y_i^p is the received pilot symbol, and $Z_i \sim \mathcal{CN}(0,\ N_0)$ is the noise at the ith relay node. Let $\tilde{\alpha_i}$ be the estimate error and $\hat{\alpha}_i$ be the channel estimator at the ith relay node., i.e., $\tilde{\alpha_i} \triangleq \alpha_i - \hat{\alpha}_i$. By the orthogonality principle of the MMSE estimation, the estimate error and the channel estimator are uncorrelated. Since all channel coefficients are complex Gaussian, the estimate error is in fact *independent* from the channel estimator. Recall that $T_s = 1/B$, and thus $E_{tr} = \eta LP/B$. We conclude that the variance of $\hat{\alpha}_i$ is

$$\mathrm{var}(\tilde{\alpha_i}) = \frac{N_0}{\frac{\eta LP}{B} + N_0} = \frac{1}{\eta\rho L + 1}. \tag{3.3}$$

Similarly, we have that the variance of $\hat{\beta}_i$ is

$$\mathrm{var}(\tilde{\beta_i}) = \frac{1}{\eta\rho L + 1}. \tag{3.4}$$

3.1.2.2. *Amplify-and-Forward at Relay Nodes*

In what follows, we establish the channel and signal models corresponding to AF with network training. Recall that the network training is done via using one pilot symbol in the first hop and another one in the second hop. During the data transmission period in the first hop, the information signal received by the ith relay node is given by

$$Y_i(t) = \alpha_i X(t) + Z_i(t), \quad \text{for} \quad T_s \le t \le LT_s. \tag{3.5}$$

where $X(t)$is the transmitted signal from the source node, and $Z_i(t)$ is complex Gaussian noise at the ith relay node. In the second hop, the ith relay node applies AF with network training to relay the received signal $Y_i(t)$. In particular, it carries out phase alignment via multiplying $Y_i(t)$ with $\frac{\hat{\alpha_i}^*\hat{\beta_i}^*}{|\hat{\alpha_i}||\hat{\beta_i}|}$. The corresponding transmitted signal at the ith relay node is given by

$$S_i(t) = A_i[\alpha_i X(t - LT_s) + Z_i(t - LT_s)] \ , \quad \text{for} \quad (L+1)T_s \leq t \leq 2LT_s \tag{3.6}$$

where A_i is the amplification factor to meet the power constraint for each relay node, and is given by

$$A_i = \frac{\hat{\alpha_i}^*\hat{\beta_i}^*}{|\hat{\alpha_i}||\hat{\beta_i}|}\sqrt{\frac{P}{|\alpha_i|^2\frac{P(1-\eta)L}{L-1} + BN_0}} \ , \tag{3.7}$$

The destination node collects signals from n relay nodes, and the received signal is given as below:

$$\begin{aligned} Y(t) &= \sum_{i=1}^{n} \beta_i S_i(t) + Z(t) \\ &= \sum_{i=1}^{n} \beta_i A_i \alpha_i X(t - LT_s) + \sum_{i=1}^{n} \beta_i A_i Z_i(t - LT_s) + Z(t), \text{ for } (L+1)T_s \leq t \leq 2LT_s \end{aligned} \tag{3.8}$$

where $Z(t)$ is complex Gaussian noise at the destination node. Moreover, as $\rho \to 0$, α_i, $BN_0 \gg |\alpha_i|^2\frac{P(1-\eta)L}{L-1}$ with probability 1, and hence in the low SNR regime,

$$A_i \approx \frac{\hat{\alpha_i}^*\hat{\beta_i}^*}{|\hat{\alpha_i}||\hat{\beta_i}|}\sqrt{\rho} \ . \tag{3.9}$$

3.1.3. Power-Constrained Wideband Relay Networks

In wideband relay networks, frequency-selective fading becomes inevitable when the frequency bandwidth increases. To study the scaling laws of wideband relay networks, we

investigate an equivalent model where each wideband channel is decomposed into a set of frequency-flat sub-channels. Let the bandwidth of each subchannel in wideband relay networks be B and the total number of subchannels is K. Then, the overall frequency bandwidth $W = KB$. Define

$$e_3 \triangleq \frac{\log K}{\log n}. \tag{3.10}$$

Since B is fixed, we have that the total frequency bandwidth $W = \Theta(K) = \Theta(n^{e_3})$. Again, we are interested in how the three parameters n, L and W jointly determine the achievable rates; and the above setting is used to obtain a clear understanding how the simultaneous scaling of the three key parameters impacts the scaling laws.

We have a few words on the exponents e_1 and e_3. Recall that

$$e_1 \triangleq \frac{\log(\frac{1}{\rho})}{\log n} = \frac{-\log \frac{P}{N_o} + \log B}{\log n}.$$

It is clear that e_1 is intimately related to e_3. Indeed, the "low SNR" in the narrowband relay network can be viewed as the result of a large bandwidth and fixed power. On the other hand, it should be cautioned that the wideband relay network model is not a simple aggregation of many narrowband relay network models, because power allocation across the subchannels at relay nodes does play an important role to determine the capacity. We elaborate further in Section 3.4 the power allocation across the subchannels.

Due to frequency-selective fading, the backward (forward) links may have different channel statistics in different frequency subbands. To take this into account, we assume that for the kth subchannel,

- let $\{\alpha_{i,k}\}$ and $\{\beta_{i,k}\}$ denote the i.i.d. Rayleigh fading coefficients corresponding to the backward link and the forward link, respectively. Assume that $\alpha_{i,k}$ and $\beta_{i,k}$ have

zero mean and common variance σ_k^2;

- assume that all relay nodes have an average power constraint p_k within any fading block (for the kth subchannel);
- assume that the source node allocates its average power uniformly across the subchannels within one fading block, namely, P/K for each subchannel.
- denote the portion of energy allocated for training as η_k, which could be different across the subchannels.

We assume uniform power allocation at the source node, mainly because the optimal power allocation at the source node requires the knowledge of the channel gains of all subchannels and also depends on $\{p_k\}$'s. It is clear that the optimization of $\{p_k\}$'s is highly non-trivial and it is unlikely to be available in practical systems.

Letting X_k be the transmitted signal in the kth subband from the source node, we have that the received signal at the destination node in the kth subband is given by (cf. (3.8))

$$Y(t) = \sum_{i=1}^{n} \beta_{i,k} A_{i,k} \alpha_{i,k} X_k(t - LT_s) + \sum_{i=1}^{n} \beta_{i,k} A_{i,k} Z_{i,k}(t - LT_s) + Z_k(t) \ , \tag{3.11}$$

for $(L+1)T_s \leq t \leq 2LT_s$. The amplification factor in the kth subband at the ith relay node $A_{i,k}$ is given as

$$A_{i,k} \approx \frac{\hat{\alpha}_{i,k}^* \hat{\beta}_{i,k}^*}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|} \sqrt{\frac{p_k}{BN_0}}. \tag{3.12}$$

Along the same line as in the study for the narrowband relay network model, we have

that for the kth subchannel, the variance of $\hat{\alpha}_{i,k}$ and $\hat{\beta}_{i,k}$ is

$$\mathrm{var}(\tilde{\alpha}_{i,k}) = \frac{\sigma_k^2 W N_0}{\eta_k \sigma_k^2 P L + W N_0}, \tag{3.13}$$

$$\mathrm{var}(\tilde{\beta}_{i,k}) = \frac{\sigma_k^2 W N_0}{\eta_k \sigma_k^2 P L + W N_0}. \tag{3.14}$$

3.2. Summary of Main Results on Scaling Laws and Achievable Rates

For convenience, we introduce the following natation for asymptotic order relationship. For two sequences $\{f(n)\}$ and $\{g(n)\}$, we write

$$f(n) \overset{\bullet}{\geq} g(n) \tag{3.15}$$

if

$$\liminf_{n\to\infty} \frac{f(n)}{g(n)} \geq a$$

where $a > 0$ is some positive constant, and vice verse. Intuitively speaking, $f(n) \overset{\bullet}{\geq} g(n)$ indicates that $f(n)$ grows at least at the same rate as $g(n)$.

In the following, we summarize the main results on the scaling laws in the joint asymptotic regimes.

Theorem 3.1 ***[Narrowband Relay Networks]*** *If there exists $\zeta \geq 0$ such that $\rho L \overset{\bullet}{\geq} n^{\zeta}$, then as $n \to \infty$, the capacity of the narrowband relay networks in the low SNR regime scales as*

$$C = \begin{cases} \Theta(\log(n\rho)) & \text{if } \liminf_{n\to\infty} \log(n\rho) > 0 \\ \mathrm{O}(1) & \text{if } \limsup_{n\to\infty} \log(n\rho) \leq 0 \end{cases}. \tag{3.16}$$

Theorem 3.2 ***[Wideband Relay Networks]*** *a). If there exist $\delta \geq 0$ and $0 < \epsilon < 1$ such that $\frac{L}{W} \overset{\bullet}{\geq} n^{\delta}$ and $W \overset{\bullet}{\leq} n^{1-\epsilon}$, then as $n \to \infty$, the capacity of power-constrained wideband relay networks scales as*

$$C^{WB} = \Theta\left(W \log\left(\frac{n}{W}\right)\right). \tag{3.17}$$

b). The equal power allocation policy at relay nodes can achieve the same scaling order as that by the optimal power allocation policy at relay nodes.

Remarks on scaling laws: Theorem 3.1 provides a sufficient condition to achieve the scaling law of the narrowband relay network model in the joint asymptotic regime of n, L and ρ. Roughly speaking, the scaling law can be characterized by (3.16) provided that the transmission energy per fading block (proportional to ρL) is bounded below as the number of relay nodes n grows. That is to say, "almost" coherent relaying can be made possible based on good channel estimation, provided that the amount of energy available within each fading block is at least constant. Observe that $\log(n\rho) = (1 - e_1)\log(n)$. It turns out that if ρ is bounded below, the capacity of narrowband relay networks is of $\Theta(\log(n))$ (cf. [14]).

Theorem 3.2 provides a sufficient condition to achieve the scaling law for power-constrained wideband relay networks in the joint asymptotic regime of n, L and W. Since $L/W = LT_sP/K$, we can view L/W as the (normalized) energy per fading block in each subband, and n/W as the degree of node diversity per unit bandwidth. Intuitively speaking, the above condition reveals that if the energy per fading block in each subband, proportional to L/W, is bounded below and W is sub-linear in n, then the scaling law can be characterized by (3.17). Note that under the above conditions, (3.17) indicates that the

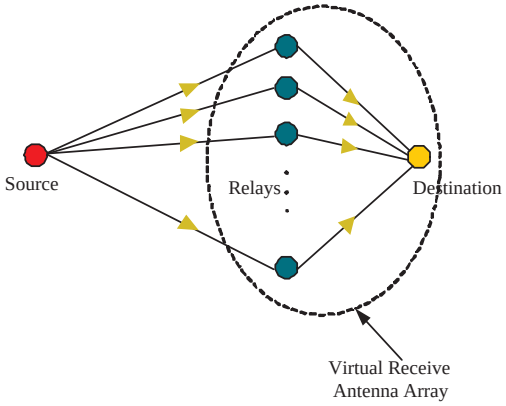

Figure 3.3. W is sublinear in n

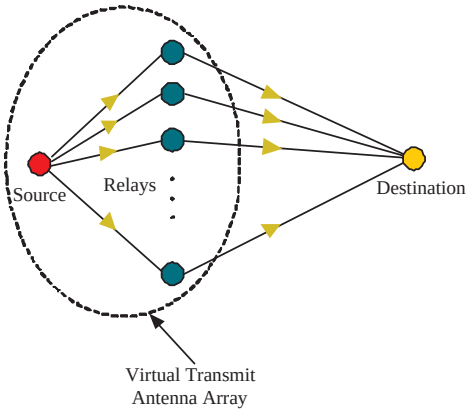

Figure 3.4. W is suplinear in n

capacity of wideband relay networks is of $\Theta(n^{1-e_3}\log(n))$.

The conditions in Theorem 3.2 suggest that when W is sublinear in n, i.e., $W \overset{\bullet}{<} n$, the wideband relay networks using AF with network training yields achievable rates of the same scaling order as that of the Cut-Set upper bound. The underlying rationale is that the scaling law for the wideband relay networks using AF relaying is limited by the node diversity order n. In general, in a power-constrained wideband relay network using AF relaying, a larger bandwidth would lead to a smaller amplification factor for relaying, which in turn may reduce the power of the desired signal embedded in the amplified signal from each relay node, and therefore the equivalent overall SNR at the destination node may or may not increase, depending on how W and n scale together. It can be shown that the power of the aggregated noise from the relay nodes is of $\Theta(n)$ because every relay node has an average power constraint P, and the power of the ambient noise at the destination node is of $\Theta(W)$. In the case when W is sublinear in n, the aggregated "amplified" noise from all the relay nodes asymptotically dominates the ambient noise at the destination node, and all forward links are virtually "noise-free". Such "noise-free" links effectively convert all

the relay nodes and the destination node into an antenna array of n receive antennas, as shown in Fig. 3.3. Observe that the relay nodes carry out phase alignment, and as a result, the equivalent virtual antenna array effectively conducts maximal-ratio combining (MRC). Thus, the relay network functions as if it were a single-input-multiple-output (SIMO) system with optimal decoding at the receiver. We note that the upper bound on the capacity of the relay network can be obtained by using the Cut-Set Theorem which corresponds to the capacity of a SIMO system when receive CSI is available. Then it can be expected that under the "noise-free" condition mentioned above, the scaling law can be achieved.

In contrast, when W is suplinear in n, the ambient noise at the destination node asymptotically dominates the aggregated "amplified" noise from all relay nodes. More specifically, it turns out that in this case the smaller amplification factor at the relay nodes would decrease the overall SNR at the destination node as n and W grow together, indicating that the achievable rate of AF does not scale well. We also note that in this case the source node and all relay nodes virtually function as a "noisy" antenna array of n transmit antennas, as depicted in Fig. 3.4; and the virtual antenna array is "noisy" in the sense that the transmitted signals from the relay nodes consist of the desired signals and the amplified noise.

In summary, the scaling law can be achieved if W is sublinear in n, i.e., $W \overset{\bullet}{<} n$. When W is suplinear in n, i.e., $W \overset{\bullet}{>} n$, there exits a gap between the scaling order of the upper bound and that of the achievable rates. For the case when W is linear in n, the achievable rate would depend on other parameters. We elaborate further on this in Sections 3.3.5 and 3.4.4.

The second part of Theorem 3.2 indicates that AF with network training, under simple

equal power allocation across the frequency subbands at the relay nodes, can achieve the scaling law. Our intuition is as follows. When the AF relaying strategy is applied at a relay node, it "amplifies" both the signals and the received noise simultaneously, regardless of what power allocation scheme is applied. As a result, the optimal power allocation scheme would achieve the scaling order under the same conditions as the equal power allocation one. We elaborate further on this in Section 3.4.3.

Remarks on achievable rates: We have also investigated the achievable rates and their scaling behavior when the conditions for achieving scaling laws in Theorems 3.1 and 3.2 are not satisfied. In particular, for the narrowband relay network model, when $e_1 > e_2$, the corresponding achievable rates are summarized in Proposition 3.2. For the wideband relay network model, when $e_3 > e_2$, the corresponding achievable rates under equal power allocation at relay nodes can be found in Section 3.4.2. We show in Sections 3.3 and 3.4 that in both cases the estimate error cannot be made small and that the corresponding scaling orders lie in between the scaling order for non-coherent relaying and that for coherent relaying.

3.3. Narrowband Relay Networks in the Low SNR Regime

3.3.1. Equivalent Source-to-Destination Channel Model

In the following, we first study the equivalent source-to-destination model for the narrowband relay networks. Recall that in the first hop, the information data signal received by the ith relay node is given by

$$Y_i(t) = \alpha_i X(t) + Z_i(t) \ , \quad \text{for} \quad T_s \leq t \leq LT_s. \tag{3.18}$$

In the second hop, the ith relay node applies AF with network training to relay the received signal $Y_i(t)$. In particular, it carries out phase alignment, and the corresponding transmitted signal at the ith relay node is given by

$$S_i(t) = A_i[\alpha_i X(t - LT_s) + Z_i(t - LT_s)] \ , \quad \text{for} \quad (L+1)T_s \leq t \leq 2LT_s, \tag{3.19}$$

where A_i is the amplification factor to meet the power constraint for each relay node, and is given by

$$A_i = \frac{\hat{\alpha_i}^* \hat{\beta_i}^*}{|\hat{\alpha_i}||\hat{\beta_i}|} \sqrt{\frac{P}{|\alpha_i|^2 \frac{P(1-\eta)L}{L-1} + BN_0}} \ . \tag{3.20}$$

The destination node collects signals from n relay nodes, and the received signal is given as:

$$\begin{aligned} Y(t) &= \sum_{i=1}^{n} \beta_i S_i(t) + Z(t) \\ &= \sum_{i=1}^{n} \beta_i A_i \alpha_i X(t - LT_s) + \sum_{i=1}^{n} \beta_i A_i Z_i(t - LT_s) + Z(t), \text{ for } (L+1)T_s \leq t \leq 2LT_s \end{aligned} \tag{3.21}$$

After some algebra, we can rewrite (3.21) as

$$\begin{aligned} Y(t) = \quad & \sum_i \sqrt{\rho} |\hat{\alpha_i}||\hat{\beta_i}| X(t - LT_s) \\ & + \underbrace{\sum_i \sqrt{\rho} \left(\frac{|\hat{\alpha_i}|^2 \tilde{\beta_i} \hat{\beta_i}^* + |\hat{\beta_i}|^2 \tilde{\alpha_i} \hat{\alpha_i}^* + \tilde{\alpha_i} \hat{\alpha_i}^* \tilde{\beta_i} \hat{\beta_i}^*}{|\hat{\alpha_i}||\hat{\beta_i}|} \right) X(t - LT_s)}_{\Omega_1(t)} \\ & + \underbrace{\sum_i \sqrt{\rho} \left(\frac{\hat{\alpha_i}^* |\hat{\beta_i}|^2 + \hat{\alpha_i}^* \tilde{\beta_i} \hat{\beta_i}^*}{|\hat{\alpha_i}||\hat{\beta_i}|} \right) Z_i(t - LT_s)}_{\Omega_2(t)} \\ & + Z(t), \quad \text{for} \quad (L+1)T_s \leq t \leq 2LT_s \ . \end{aligned} \tag{3.22}$$

In the above signal model, the received signal at the destination node consists of

- the information-bearing signal, $\sum_i \sqrt{\rho} |\hat{\alpha_i}||\hat{\beta_i}| X(t - LT_s)$,

- the signal due to the estimate error, $\Omega_1(t)$,
- the aggregated amplified noise from all relay nodes, $\Omega_2(t)$,
- and the noise at the destination node, $Z(t)$.

We observe that in the above model the information is embedded in the received signal with unknown parameters, because the destination node does not have the knowledge of the instantaneous CSI of the backward and forward channels. Define $H \triangleq \sqrt{\rho}\sum_i |\hat{\alpha_i}||\hat{\beta}_i|$. We can rewrite the equivalent source-to-destination model as

$$Y(t) = \mathbb{E}[H]X(t - LT_s) + \Omega_1(t) + \Omega_2(t) + \Omega_3(t) + Z(t), \tag{3.23}$$

where $\Omega_3(t) \triangleq (H - \mathbb{E}[H])X(t - LT_s)$. In this equivalent source-to-destination channel model (3.23), it is clear that $\mathbb{E}[H]$ can be easily made available at the destination node, because it is determined by the channel statistics only. Furthermore, both $\Omega_1(t)$ and $\Omega_3(t)$ are signal-dependent, and $\Omega_2(t)$ is non-Gaussian. This equivalent model is then a deterministic channel with non-Gaussian signal-dependent noise, provided that $\Omega_1(t)+\Omega_2(t)+\Omega_3(t)+Z(t)$ is treated as the aggregated noise. In general, it is non-trivial to characterize the exact capacity of the channel model (3.23).

In what follows, we turn to characterize an achievable rate for (3.23) instead, which serves as a lower bound. Observe that the channel side information is imperfect in the channel model in (3.23), and the mutual information between the input and the output depends heavily on the variance of the channel estimate error. Along the lines of [31], an achievable rate is given by (we relegate the detailed derivation to the Appendix C)

$$C \geq B\frac{L-1}{L}\log\left(1 + \frac{\mathbb{E}\left[|\mathbb{E}[H]X(t - LT_s)|^2\right]}{\mathbb{E}[|\Omega_1(t)|^2] + \mathbb{E}[|\Omega_2(t)|^2] + \mathbb{E}[|\Omega_3(t)|^2] + WN_0}\right). \tag{3.24}$$

The RHS of (3.24) is an achievable rate for the equivalent model in (3.23). Interestingly, it can be obtained by treating $\Omega_1(t) + \Omega_2(t) + \Omega_3(t) + Z(t)$ as white Gaussian noise with the same variance. Note that $\Omega_2(t) + Z(t)$ is a dominating term in the equivalent noise due to the low SNR. As noted in Section 3.2, if $\Omega_2(t)$ is dominant, it can be shown that the relay networks using AF with network training act as if it were a single-input-multiple-output (SIMO) system.

For notational convenience, define

$$S \triangleq \frac{\mathbb{E}\left[|\mathbb{E}[H]X(t-LT_s)|^2\right]}{\mathbb{E}[|\Omega_1(t)|^2] + \mathbb{E}[|\Omega_2(t)|^2] + \mathbb{E}[|\Omega_3(t)|^2] + WN_0} \tag{3.25}$$

as the equivalent receive SNR corresponding to the equivalent model in (3.23). It remains to characterize the power of the desired signal term and the noise terms in (3.23). After some algebra (the detailed derivations have been relegated to Appendix D), we have that

$$S = \frac{n\frac{\pi^2}{16}\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2 \frac{(1-\eta)L}{L-1}P}{\left(1 - \frac{\pi^2}{16}\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2\right)\frac{(1-\eta)L}{L-1}P + BN_0 + \frac{(BN_0)^2}{nP}} \ . \tag{3.26}$$

Assuming that the coherence interval $L \geq 2$, we have that

$$0 \leq \left(1 - \frac{\pi^2}{16}\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2\right)\frac{(1-\eta)L}{L-1}P \leq 2P \ . \tag{3.27}$$

Since the receive SNR per link ρ, is small, we have that

$$\left(1 - \frac{\pi^2}{16}\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2\right)\frac{(1-\eta)L}{L-1}P = \mathrm{o}\left(\mathrm{BN}_0 + \frac{(\mathrm{BN}_0)^2}{\mathrm{nP}}\right) \ . \tag{3.28}$$

Therefore, we conclude that when $n \to \infty$,

$$S = \Theta\left(\frac{n\frac{\pi^2}{16}\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2 \frac{(1-\eta)L}{L-1}\rho}{1+\frac{1}{n\rho}}\right) \ . \tag{3.29}$$

Then, the achievable rate of narrowband relay networks using AF with network training at relay nodes is given by

$$R = B\frac{L-1}{L}\log\left(1+S\right) \tag{3.30}$$

$$= \Theta\left(\log\left(1+\frac{n\frac{\pi^2}{16}\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2\frac{(1-\eta)L}{L-1}\rho}{1+\frac{1}{n\rho}}\right)\right). \tag{3.31}$$

It is clear from (3.29) that S depends on n, L and ρ through $n\rho$ and ρL. Since $\rho L = LPT_s/N_0$, we note that ρL can be regarded as the normalized transmission energy per fading block, and $\eta\rho L$ corresponds to the normalized energy for training per fading block. Observe $\rho n = nPT_s/N_0$, where nPT_s can be viewed as the total energy across the relay nodes to relay one symbol.

3.3.2. An Upper Bound on the Capacity

We next present an upper bound on the capacity of narrowband relay networks in the low SNR regime. By the Cut-Set Theorem in [7], the capacity of the relay network is upper bounded by the information rate corresponding to the "broadcast cut", regardless of the transmission schemes. That is,

$$C < 2B\log\left(1+\frac{\sum_i|\alpha_i|^2P}{BN_0}\right). \tag{3.32}$$

Define $C_{upper} \triangleq 2B\log\left(1+\frac{\sum_i|\alpha_i|^2P}{BN_0}\right)$. By the Law of Large Numbers, $\sum_i|\alpha_i|^2$ grows linearly in n as $n\to\infty$, it follows that C_{upper} is $\Theta(\log(1+n\rho))$. Accordingly, we have the following lemma.

Lemma 3.1 *As $n \to \infty$, the upper bound on the capacity of narrowband relay networks, C_{upper}, scales as*

$$C_{upper} = \begin{cases} \Theta(\log(n\rho)) & \text{if } \liminf_{n\to\infty} \log(n\rho) > 0 \\ \mathrm{O}(1) & \text{if } \limsup_{n\to\infty} \log(n\rho) \leq 0 \end{cases} . \tag{3.33}$$

3.3.3. The Perfect CSI Case

Needless to say, the highest possible achievable rate of the AF strategy occurs when there is perfect CSI at the relay nodes, i.e., the ith relay node has a priori knowledge on α_i and β_i. The corresponding achievable rate is given by [8]

$$R_{pft} = B \log \left(1 + \frac{\frac{\pi^2}{16} n\rho}{1 + \frac{1}{n\rho}} \right) . \tag{3.34}$$

Note that the above result can also be obtained by letting $\mathrm{var}(\tilde{\alpha}) = 0$, $\mathrm{var}(\tilde{\beta}) = 0$ and $L \to \infty$ while we prove (3.31).

We have the following lemma on R_{pft}.

Lemma 3.2 *If perfect CSI on the backward channels and the forward channels is available at the corresponding relay node, then as $n \to \infty$, the achievable rate by using AF scales as*

$$R_{pft} = \begin{cases} \Theta(\log(n\rho)) & \text{if } \liminf_{n\to\infty} \log(n\rho) > 0 \\ \mathrm{O}(1) & \text{if } \limsup_{n\to\infty} \log(n\rho) \leq 0 \end{cases} . \tag{3.35}$$

Note that in the perfect CSI case, the achievable rate (a lower bound on the capacity) has the same scaling order as the upper bound, thus yielding the scaling law. In what follows, we investigate the achievable rates of narrowband relay network in the low SNR regime, where there is no a priori knowledge on the CSI at the relay nodes.

3.3.4. Optimal Energy Allocation for Channel Estimation

As aforementioned, $E_{tr} = \eta E_{total}$ is allocated for training. It is clear that $\eta^* < 1$, and that the larger η is, the more accurate the channel estimation would be, at the cost of the energy for data transmission. That is to say, there is a trade-off between the energy for training and that for data transmission. It is easy to see that the optimal η is the one that maximizes the equivalent receive SNR in (3.29). Hence, taking derivative of S with respect to η, we have that the optimal η is given by

$$\eta^* = \frac{-3 + \sqrt{8\rho L + 9}}{2\rho L} \ , \tag{3.36}$$

where $\rho L = PLT_s/N_0$, as aforementioned, can be viewed as the total energy available within each coherence time, normalized with respect to N_0. Since $1/\rho$ and L can grow unbounded as $n \to \infty$, ρL can range from 0 to ∞. We investigate the achievable rates for the following three cases:

- I) ρL grows unbounded (equivalently $e_1 < e_2$);
- II)ρL approaches zero (equivalently $e_1 > e_2$);
- III) ρL is of $\Theta(1)$ (equivalently $e_1 = e_2$).

3.3.5. The Achievable Rates of Narrowband Relay Networks in the Low SNR Regime

3.3.5.1. *The Case with $e_1 < e_2$*

In this case, the coherence interval L grows faster than $1/\rho$. As a result, the energy per fading block ρL increases as n grows, indicating that there exists $\epsilon > 0$, such that

$$\rho L \overset{\bullet}{\geq} n^{\epsilon}. \tag{3.37}$$

It can be shown that

$$\eta^* = \Theta\left(n^{\frac{e_1-e_2}{2}}\right) \qquad \text{and} \qquad \eta^*\rho L = \Theta\left(n^{\frac{e_2-e_1}{2}}\right). \tag{3.38}$$

(3.38) reveals that when $e_1 < e_2$, the energy for training within each fading block also grows as $n \to \infty$, even though $\eta^* \to 0$. Indeed, $\eta\rho L = \Theta(\sqrt{\rho L})$, and both the total energy and the energy for training per fading block grows unbounded. Since the MMSE estimation depends on $\eta\rho L$ only, it follows that channel estimation can be accurate in this scenario, thereby enabling coherent relaying.

In light of the results in (3.35) for the perfect CSI case, we expect that the achievable rates are different for different values of e_1. Indeed, if $e_1 < 1$, it can be shown that

$$\begin{aligned} \lim_{n\to\infty} S &= \lim_{n\to\infty} n\frac{\pi^2}{16}\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2 \frac{(1-\eta)L}{L-1}\rho \\ &= \Theta\left(n^{1-e_1}\right), \end{aligned} \tag{3.39}$$

indicating that the achievable rate of the relay network is

$$R = \Theta\left(\log(n\rho)\right). \tag{3.40}$$

Interestingly, the achievable rate in the case where $e_1 < 1$ can scale the same as in the perfect CSI case. Note that if ρ is bounded below, the achievable rate is of $\Theta(\log(n))$ (cf. [14]).

In contrast, if $e_1 \geq 1$, we have that $\rho \leq \frac{1}{n}$ as $n \to \infty$, i.e., the SNR goes to zero faster than $\frac{1}{n}$. Accordingly,

$$\begin{aligned} \lim_{n\to\infty} S &= \lim_{n\to\infty} \frac{n\frac{\pi^2}{16}\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2 \frac{(1-\eta)L}{L-1}\rho}{\frac{1}{n\rho}} \\ &= \Theta\left(n^{2-2e_1}\right) . \end{aligned} \tag{3.41}$$

It follows that the achievable rate for this case is

$$R = \mathrm{O}(1) . \tag{3.42}$$

Our intuition for this case is as follows: even if coherent relaying is possible in this case, the achievable rate would grow slower than $\log n$ because total power across the relay nodes $n\rho \overset{\bullet}{\leq} 1$.

The following proposition summarizes the results on the achievable rate by using AF with network training for the case $e_1 < e_2$.

Proposition 3.1 *If there exists $\epsilon > 0$, such that $\rho L \overset{\bullet}{\geq} n^{\epsilon}$, then as $n \to \infty$, the achievable rate by using AF with network training scales as*

$$R = \begin{cases} \Theta(\log(n\rho)) & \textit{if} \quad e_1 < 1 \\ \mathrm{O}(1) & \textit{if} \quad e_1 \geq 1 \end{cases} . \tag{3.43}$$

3.3.5.2. *The Case with $e_1 > e_2$*

In this scenario, $1/\rho$ grows faster than L. Accordingly, ρL approaches 0 as n goes to infinity. Along the same lines as above, we study the optimal energy allocation and the

achievable rates for different values of e_1. It can be shown that the optimal energy allocation for using AF with network training is given by

$$\begin{aligned}\lim_{n\to\infty} \eta^* &= \lim_{n\to\infty} \left(\frac{-3+3\sqrt{1+\frac{8\rho L}{9}}}{2\rho L} \right) \\ &= \frac{2}{3}\,, \end{aligned} \tag{3.44}$$

and

$$\eta^* \rho L = \Theta\left(n^{e_2-e_1}\right). \tag{3.45}$$

We observe that in this case $\eta\rho L = \Theta(\rho \mathrm{L})$, whereas $\eta\rho L = \Theta(\sqrt{\rho \mathrm{L}})$ for the case with $e_1 < e_2$. We note that in this case the total normalized energy per fading block goes to zero, so as the training energy.

When $e_1 < 1$, we have that

$$\begin{aligned}\lim_{n\to\infty} S &= \lim_{n\to\infty} n\frac{\pi^2}{16}\left(\frac{2}{3}n^{e_2-e_1}\right)^2 \frac{(1-\frac{2}{3})L}{L-1}\rho \\ &= \Theta\left(n^{1+2e_2-3e_1}\right)\,. \end{aligned} \tag{3.46}$$

In contrast, if $e_1 \geq 1$, then

$$\begin{aligned}\lim_{n\to\infty} S &= \lim_{n\to\infty} \frac{n\frac{\pi^2}{16}\left(\frac{2}{3}n^{e_2-e_1}\right)^2 \frac{(1-\frac{2}{3})L}{L-1}\rho}{\frac{1}{n\rho}} \\ &= \Theta\left(n^{2+2e_2-4e_1}\right)\,. \end{aligned} \tag{3.47}$$

In summary, we have the following proposition.

Proposition 3.2 *If $\rho L \to 0$ as $n \to \infty$, then the achievable rate by using AF with network training scales as*

- *If* $e_1 < 1$, *then*

$$R = \begin{cases} \Theta\left(\log(n^{1+2e_2-3e_1})\right) & for \quad e_1 < \frac{1+2e_2}{3} \\ \mathrm{O}(1) & for \quad e_1 \geq \frac{1+2e_2}{3} \end{cases} ; \tag{3.48}$$

- *if* $e_1 \geq 1$, *then*

$$R = \begin{cases} \Theta\left(\log(n^{2+2e_2-4e_1})\right) & for \quad e_1 < \frac{1+e_2}{2} \\ \mathrm{O}(1) & for \quad e_1 \geq \frac{1+e_2}{2} \end{cases} . \tag{3.49}$$

3.3.5.3. *The Case with* $e_1 = e_2$

This case is equivalent to $\rho L = \Theta(1)$. It can be shown that $\eta^* \rho L = \Theta(1)$, i.e.,, ρL and $\eta^* \rho L$ are bounded. Along the same line as above, we have that when $e_1 < 1$,

$$\begin{aligned} \lim_{n\to\infty} S &= \lim_{n\to\infty} n\frac{\pi^2}{16}\left(\frac{\eta^*\rho L}{\eta^*\rho L + 1}\right)^2 \frac{(1-\eta^*)L}{L-1}\rho \\ &= \Theta\left(n^{1-e_1}\right) . \end{aligned} \tag{3.50}$$

For the case $e_1 \geq 1$, it can be shown that

$$\begin{aligned} \lim_{n\to\infty} S &= \lim_{n\to\infty} \frac{n\frac{\pi^2}{16}\left(\frac{\eta^*\rho L}{\eta^*\rho L+1}\right)^2 \frac{(1-\eta^*)L}{L-1}\rho}{\frac{1}{n\rho}} \\ &= \Theta\left(n^{2-2e_1}\right) . \end{aligned} \tag{3.51}$$

Correspondingly, we have the following proposition.

Proposition 3.3 *If the normalized energy per fading block* ρL *is bounded, then as* $n \to \infty$, *the achievable rate by using AF with network training scales as*

$$R = \begin{cases} \Theta(\log(n\rho)) & if \quad e_1 < 1 \\ \mathrm{O}(1) & if \quad e_1 \geq 1 \end{cases} . \tag{3.52}$$

Remarks: Summarizing the above three cases in terms of e_1 and e_2, we conclude that if $e_1 \leq e_2$, the achievable rates have the same scaling order as the upper bound in (3.1), thereby yielding the scaling law in Theorem 3.1. More specifically, as $n \to \infty$, if there exist $\zeta \geq 0$ and $a > 0$, such that $\rho L \overset{\bullet}{\geq} an^{\zeta}$, then as $n \to \infty$, the capacity of the relay networks scales as

$$C = \begin{cases} \Theta(\log(n\rho)) & \text{if} \quad \lim\limits_{n\to\infty} \inf \log(n\rho) > 0 \\ \mathrm{O}(1) & \text{if} \quad \lim\limits_{n\to\infty} \sup \log(n\rho) \leq 0 \end{cases}. \tag{3.53}$$

3.4. Power-Constrained Wideband Relay Networks

Since a frequency-selective fading channel can be decomposed as multiple parallel frequency-flat fading subchannels [42], we first examine the equivalent source-to-destination channel model corresponding to the kth subband in wideband relay networks. Along the same line of Section 4.1, the received signal in equivalent model for the kth subchannel is given by

$$Y_k(t) = \mathbb{E}[H_k]X_k(t - LT_s) + \Omega_{1k}(t) + \Omega_{2k}(t) + \Omega_{3k}(t) + Z_k(t), \tag{3.54}$$

where

- $\mathbb{E}[H_k]X_k(t - LT_s)$ is the received signal used for information extracting, and $\mathbb{E}[H_k]$ denotes the equivalent channel gain;
- $\Omega_{1k}(t)$ is due to the estimate error;
- $\Omega_{2k}(t)$ corresponds to the aggregated amplified noise at the relay nodes;
- $\Omega_{3k}(t)$ stands for the received signal weighted by unknown channel strength.

In what follows, we characterize the achievable rate of each equivalent subchannel in (3.54), along the lines for narrowband relay network models. As aforementioned, the achievable rate can be obtained by treating $\Omega_{1k}(t) + \Omega_{2k}(t) + \Omega_{3k}(t) + Z_k(t)$ as equivalent Gaussian noise of the same power. Since the SNR per link in each subband is small, $\Omega_{2k}(t) + Z_k(t)$ is a dominating term in the equivalent noise. In particular, when $\Omega_{2k}(t)$ dominates, the relay network using AF with network training acts as if it were a single-input-multiple-output (SIMO) system.

For convenience, define

$$S_k \triangleq \frac{\mathbb{E}\left[|\mathbb{E}[H_k]X_k(t-LT_s)|^2\right]}{\mathbb{E}[|\Omega_{1k}(t)|^2] + \mathbb{E}[|\Omega_{2k}(t)|^2] + \mathbb{E}[|\Omega_{3k}(t)|^2] + WN_0} \tag{3.55}$$

as the equivalent overall SNR corresponding to the equivalent model in (3.54) Next, we examine the power of the terms in (3.54) (the derivation is relegated to Appendix D). After some algebra, the equivalent SNR in the equivalent network model for the kth subband can be shown to be

$$S_k = \frac{n\sigma_k^4 \frac{\pi^2}{16} \left(\frac{\eta_k \sigma_k^2 PL}{\eta_k \sigma_k^2 PL + WN_0}\right)^2 \frac{(1-\eta_k)L}{L-1}\frac{P}{K}}{\sigma_k^4 \left(1 - \frac{\pi^2}{16}\left(\frac{\eta_k \sigma_k^2 PL}{\eta_k \sigma_k^2 PL + WN_0}\right)^2\right)\frac{(1-\eta_k)L}{L-1}\frac{P}{K} + \sigma_k^2 BN_0 + \frac{(BN_0)^2}{np_k}} \ . \tag{3.56}$$

Note that as $K \to \infty$, $P/K \ll BN_0$. It follows that in the joint asymptotic regime of n, L and W,

$$S_k = \Theta\left(\frac{n\sigma_k^4 \frac{\pi^2}{16} \left(\frac{\eta_k \sigma_k^2 PL}{\eta_k \sigma_k^2 PL + WN_0}\right)^2 \frac{(1-\eta_k)L}{L-1}\frac{P}{K}}{\sigma_k^2 BN_0 + \frac{(BN_0)^2}{np_k}}\right) \ . \tag{3.57}$$

The optimal η_k maximizing S_k is given by

$$\eta_k^* = \frac{-3 + \sqrt{\frac{8\sigma_k^2 PL}{WN_0} + 9}}{\frac{2\sigma_k^2 PL}{WN_0}} \ . \tag{3.58}$$

We emphasize that η_k^* can be different across different subbands, simply because their channel statistics are different.

In a nutshell, the achievable rate by using AF with network training in the power-constrained wideband relay network can be given as

$$R^{WB} = \sum_{k=1}^{K} \frac{B(L-1)}{L} \log\left(1 + S_k\right) \ . \tag{3.59}$$

3.4.1. An Upper Bound on the Capacity

Similar to the narrowband case, by the Cut-Set Theorem, one upper bound on the capacity of the wideband relay networks is given by the information rate through the "broadcast-cut", i.e.,

$$\begin{aligned} C^{WB} &< \sum_{k=1}^{K} \left[2B \log\left(1 + \frac{\sum_i |\alpha_{i,k}|^2 \frac{P}{K}}{BN_0}\right)\right] & (3.60)\\ &\overset{(a)}{\leq} 2W \log\left(1 + \frac{1}{K} \frac{\sum_k \sum_i |\alpha_{i,k}|^2 P}{WN_0}\right) \ , & (3.61) \end{aligned}$$

where (a) follows Jensen's Inequality. For convenience, define

$$C_{upper}^{WB} \triangleq 2W \log\left(1 + \frac{1}{K} \frac{\sum_k \sum_i |\alpha_{i,k}|^2 P}{WN_0}\right) \ . \tag{3.62}$$

By the Laws of Large Numbers, we have that $\sum_k \sum_i |\alpha_{i,k}|^2$ increases linearly with respect to n and K in the asymptotic regime. Then it follows that

$$C_{upper}^{WB} = \Theta\left(W \log\left(1 + \frac{nP}{WN_0}\right)\right) \ . \tag{3.63}$$

Accordingly, we have the following lemma.

Lemma 3.3 *As $n \to \infty$, the scaling order of C_{upper}^{WB} is given by*

$$C_{upper}^{WB} = \begin{cases} \Theta(W \log(\frac{n}{W})) & \textit{if} \quad \liminf_{n\to\infty} \log(n/W) \geq 0 \\ \Theta(n) & \textit{if} \quad \limsup_{n\to\infty} \log(n/W) < 0 \end{cases} . \tag{3.64}$$

In the next section, we first study the network model where a simple power allocation policy, namely the equal power allocation policy, is applied at relay nodes.

3.4.2. Achievable Rates with the Equal Power Allocation Policy at Relay Nodes

When the equal power allocation policy is applied at relay nodes, the relaying power for the kth subband p_k is simply P/K. From (3.54), the SNR corresponding to the kth equivalent subchannel is

$$\begin{aligned} S_{Eq,k} &\triangleq \frac{n\sigma_k^4 \frac{\pi^2}{16} \left(\frac{\eta_k \sigma_k^2 PL}{\eta_k \sigma_k^2 PL + WN_0}\right)^2 \frac{(1-\eta_k)L}{L-1} \frac{P}{K}}{\sigma_k^2 BN_0 + \frac{(BN_0)^2}{n\frac{P}{K}}} \\ &= \frac{n\sigma_k^4 \frac{\pi^2}{16} \left(\frac{\eta_k \sigma_k^2 PL}{\eta_k \sigma_k^2 PL + WN_0}\right)^2 \frac{(1-\eta_k)L}{L-1} P}{\sigma_k^2 WN_0 + \frac{(WN_0)^2}{nP}} . \end{aligned} \tag{3.65}$$

Following the same line as in the study on the narrowband relay network model, we next investigate the scaling behavior of $S_{Eq,k}$ in three scenarios in terms of e_3 and e_2.

- If $e_3 < e_2$, then the optimal ratio of the energy allocation for training is given by

$$\lim_{n\to\infty} \eta_k^* = 0 . \tag{3.66}$$

Observing that

$$\frac{\eta_k^* PL}{WN_0} = \Theta\left(n^{\frac{e_2 - e_3}{2}}\right) , \tag{3.67}$$

we conclude that the equivalent SNR scales as

$$S_{Eq,k} = \begin{cases} \Theta\left(n^{1-e_3}\right) & \text{if} \quad e_3 < 1 \\ \Theta\left(n^{2-2e_3}\right) & \text{if} \quad e_3 \geq 1 \end{cases} . \tag{3.68}$$

- If $e_3 > e_2$, then the optimal energy allocation for training is given by

$$\lim_{n\to\infty} \eta_k^* = \frac{2}{3}. \tag{3.69}$$

Since

$$\frac{\eta_k^* PL}{WN_0} = \Theta\left(n^{e_2-e_3}\right) , \tag{3.70}$$

we conclude that the equivalent SNR scales as

$$S_{Eq,k} = \begin{cases} \Theta\left(n^{1+2e_2-3e_3}\right) & \text{if} \quad e_3 < 1 \\ \Theta\left(n^{2+2e_2-4e_3}\right) & \text{if} \quad e_3 \geq 1 \end{cases} . \tag{3.71}$$

- If $e_3 = e_2$, then L/W is bounded, i.e., $L/W = \Theta(1)$. It follows that

$$\frac{\eta_k^* PL}{WN_0} = \Theta(1) . \tag{3.72}$$

Then, the equivalent SNR scales as

$$S_{Eq,k} = \begin{cases} \Theta\left(n^{1-e_3}\right) & \text{if} \quad e_3 < 1 \\ \Theta\left(n^{2-2e_3}\right) & \text{if} \quad e_3 \geq 1 \end{cases} . \tag{3.73}$$

We observe that the scaling behavior of $S_{Eq,k}$ in (3.68), (3.71) and (3.73) satisfies that

$$S_{Eq,k} = \Theta(S_{Eq,l}), \quad k \neq l. \tag{3.74}$$

That is to say, in the asymptotic regime, when the equal power allocation policy is applied at relay nodes, the equivalent receive SNRs in different subchannels have the same scaling

behavior. Combining (3.74) with (3.59), the achievable rate of wideband relay networks using the equal power allocation policy at relay nodes is given by

$$R_{Eq}^{WB} = \sum_{k=1}^{K} \frac{B(L-1)}{L} \log\left(1 + S_{Eq,k}\right) \ . \tag{3.75}$$

$$= \Theta\left(W \log\left(1 + S_{Eq,k}\right)\right) \ . \tag{3.76}$$

Summarizing the above, we conclude that

- if $e_3 \leq e_2$, then

$$R_{Eq}^{WB} = \begin{cases} \Theta\left(W \log(\frac{n}{W})\right) & \text{for} \quad e_3 \leq 1 \\ \Theta\left(\frac{n^2}{W}\right) & \text{for} \quad e_3 > 1 \end{cases} ; \tag{3.77}$$

- if $e_3 > e_2$ and $e_3 < 1$, then

$$R_{Eq}^{WB} = \begin{cases} \Theta\left(W \log\left(\frac{nL^2}{W^3}\right)\right) & \text{for} \quad e_3 < \frac{1+2e_2}{3} \\ \Theta\left(\frac{nL^2}{W^2}\right) & \text{for} \quad e_3 \geq \frac{1+2e_2}{3} \end{cases} ; \tag{3.78}$$

- if $e_3 > e_2$ and $e_3 \geq 1$, then

$$R_{Eq}^{WB} = \begin{cases} \Theta\left(W \log\left(\frac{n^2L^2}{W^4}\right)\right) & \text{for} \quad e_3 < \frac{1+e_2}{2} \\ \Theta\left(\frac{n^2L^2}{W^3}\right) & \text{for} \quad e_3 \geq \frac{1+e_2}{2} \end{cases} . \tag{3.79}$$

3.4.3. Achievable Rates with the Optimal Power Allocation Policy at Relay Nodes

Under equal power allocation, the relay nodes allocate power uniformly across the subbands within one fading block and there is no power allocation across fading blocks. A natural question to ask is that "If more flexibility is available, i.e., if each relay node is able to allocate its power across different subbands (in the frequency-domain) based

on the channel statistics and across the fading blocks based on the instantaneous channel conditions (in the time-domain), would this outperform the equal power allocation policy?" In general, one would expect a higher achievable rate, and this is often the case by using bursty flashing signals (more generally, water-filling techniques). Thus motivated, we next examine the optimal power allocation policies, where each relay node allocates its power across the fading blocks and across the subbands, while keeping an overall long-term power constraint across the subbands.

Let $p_{i,k,m}$ denote the relaying power of the ith relay node in the kth subband and in the mth fading block[1]. Assume that an average power constraint for the ith relay node is

$$\frac{1}{M}\sum_{m=1}^{M}\sum_{k=1}^{K} p_{i,k,m} \leq P \ , \tag{3.80}$$

where M is sufficiently large. Worth noting is that there is no power allocation across the relay nodes is allowed (cf. [14]).

Next, we characterize the equivalent source-to-destination model, assuming that relay nodes are allowed to carry out power allocation under the constraint in (3.80). More specifically, in the kth subband and the mth fading block, the equivalent model is given by

$$Y_{k,m}(t) = \mathbb{E}[H_{k,m}]X_k(t - LT_s) + \Omega_{1k,m}(t) + \Omega_{2k,m}(t) + \Omega_{3k,m}(t) + Z_k(t) \ , \tag{3.81}$$

The power of the desired signal is given by

$$\begin{aligned}\mathbb{E}\left[|\mathbb{E}[H_{k,m}]X_k(t - LT_s)|^2\right] &= \frac{\pi^2}{16}\frac{\sigma_k^4}{BN_0}\left(\sum_{i=1}^{n}\sqrt{p_{i,k,m}}\right)^2\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL + WN_0}\right)^2 \\ &\quad \cdot\frac{(1-\eta_k)L}{L-1}\frac{P}{K} \ .\end{aligned} \tag{3.82}$$

[1]Note that each relay node transmits only during the second hop. Without loss of generality, we assume that the power constraint is imposed for the transmission duration for each relay node.

The power of the noise terms in the kth subband is given by

$$\mathbb{E}[|\Omega_{1k,m}(t)|^2] = \frac{\sigma_k^4}{BN_0}\left(\sum_{i=1}^{n} p_{i,k,m}\right)\left(1-\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL+WN_0}\right)^2\right)\frac{(1-\eta_k)L}{L-1}\frac{P}{K}\ , \quad (3.83)$$

$$\mathbb{E}[|\Omega_{2k,m}(t)|^2] = \sigma_k^2\left(\sum_{i=1}^{n} p_{i,k,m}\right)\ , \quad (3.84)$$

$$\mathbb{E}[|\Omega_{3k,m}(t)|^2] = \frac{\sigma_k^4}{BN_0}\left(\sum_{i=1}^{n} p_{i,k,m}\right)\left(\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL+WN_0}\right)^2 - \frac{\pi^2}{16}\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL+WN_0}\right)^2\right) \cdot\frac{(1-\eta_k)L}{L-1}\frac{P}{K}\ . \quad (3.85)$$

For convenience, define the corresponding SNR at the kth subband and the mth fading block as

$$\begin{aligned} S_{k,m} &\triangleq \frac{\mathbb{E}\left[\left|\mathbb{E}[H_{k,m}]X_{k,m}(t-LT_s)\right|^2\right]}{\mathbb{E}[|\Omega_{1k,m}(t)|^2]+\mathbb{E}[|\Omega_{2k,m}(t)|^2]+\mathbb{E}[|\Omega_{3k,m}(t)|^2]+WN_0} \\ &= \frac{\sigma_k^4\frac{\pi^2}{16}n\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{p_{i,k,m}}\right)^2\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL+WN_0}\right)^2\frac{(1-\eta_k)L}{L-1}\frac{P}{K}}{\sigma_k^2\left(\frac{1}{n}\sum_{i=1}^{n}p_{i,k,m}\right)BN_0+\frac{(BN_0)^2}{n}}\ . \end{aligned} \quad (3.86)$$

Next we study the optimal power allocation across the fading blocks and across the subbands for relay nodes. We note that the optimal power allocation should exploit the degrees of freedom in both the time domain and the frequency domain, and is the solution to the following optimization problem:

$$\begin{aligned} \max_{\{p_{i,k,m}\}} \quad & \frac{1}{M}\sum_{m=1}^{M}\sum_{k=1}^{K}\frac{B(L-1)}{L}\log\left(1+S_{k,m}\right) \\ \text{subject to} \quad & \frac{1}{M}\sum_{m=1}^{M}\sum_{k=1}^{K}p_{i,k,m}\le P, \ \text{ and } \ p_{i,k,m}\ge 0. \end{aligned} \quad (3.87)$$

In general, the optimal power allocation involves water-filling non-linear functions such as $x(\cdot)^+$. Needless to say, it is highly non-trivial to characterize the exact solution.

Let R^{WB^*} be the achievable rate corresponding to the optimal power allocation policy. In what follows, we use a "sandwich" argument instead to characterize the scaling order of

R^{WB*}; that is, we first find an upper bound and a lower bound on R^{WB*}, and then show that these two bounds have the same scaling order under the same conditions.

[An Upper Bound on R^{WB*}]:

First, by Cauchy-Schwarz's Inequality, we have that

$$\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{p_{i,k,m}}\right)^2 \leq \frac{1}{n}\sum_{i=1}^{n}p_{i,k,m}, \tag{3.88}$$

where the equality can be achieved if and only if $p_{i,k,m} = p_{j,k,m}$, for $i \neq j; i, j = 1, 2, ..., n$.

For convenience, define σ^2_{max} and σ^2_{min} as the maximum and minimum of $\{\sigma^2_1, \sigma^2_2, ..., \sigma^2_K\}$, respectively; and define η_{max} and η_{min} as the maximum and minimum of $\{\eta_1, \eta_2, ..., \eta_K\}$, respectively. Using (3.88), we have that

$$S_{k,m} \leq S'_{k,m}\ , \tag{3.89}$$

where

$$S'_{k,m} \triangleq \frac{\sigma^4_{max}\frac{\pi^2}{16}n\left(\frac{1}{n}\sum_{i=1}^{n}p_{i,k,m}\right)\left(\frac{\eta_{max}\sigma^2_{max}PL}{\eta_{min}\sigma^2_{min}PL+WN_0}\right)^2\frac{(1-\eta_{min})L}{L-1}\frac{P}{K}}{\sigma^2_{min}\left(\frac{1}{n}\sum_{i=1}^{n}p_{i,k,m}\right)BN_0 + \frac{(BN_0)^2}{n}}\ . \tag{3.90}$$

We need the following lemma.

Lemma 3.4 *The function $f(x_1,\ x_2, ..., x_n)$ is concave in $[x_1,\ x_2, ..., x_n]$, where*

$$f(x_1,\ x_2, ..., x_n) = \log\left(1 + \frac{C\sum_{i=1}^{n}x_i}{D_1\sum_{i=1}^{n}x_i + D_2}\right), \tag{3.91}$$

with $x_i \geq 0$, $i = 1, 2, ..., n$; C, D_1 and D_2 being positive constants.

Proof: The proof simply follows the facts that

- *log* $\left(1 + \frac{C}{D_1}x\right)$ is concave and increasing in x with $x \geq 0$;

- $\frac{x}{x+D_2/D_1}$ is concave and increasing in x with $x \geq 0$;
- $\sum_i x_i$ is concave in vector $[x_1,\ x_2, ..., x_n]$ with $x_i \geq 0$.

By the properties of the function composition that preserves the concavity [3, p.84], we conclude the proof of Lemma 3.4. ∎

Applying Lemma 3.4, we have that $S'_{k,m}$ is concave in vector $[p_{1,k,m},\ p_{2,k,m}, ..., p_{n,k,m}]$. Then, by Jensen's Inequality, it follows that

$$\frac{1}{MK}\sum_{m=1}^{M}\sum_{k=1}^{K}\log\left(1+S'_{k,m}\right)$$

$$\leq \log\left(1+\frac{\sigma_{max}^4\frac{\pi^2}{16}n\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{MK}\sum_{k=1}^{K}\sum_{m=1}^{M}p_{i,k,m}\right)\left(\frac{\eta_{max}\sigma_{max}^2PL}{\eta_{min}\sigma_{min}^2PL+WN_0}\right)^2\frac{(1-\eta_{min})L}{L-1}\frac{P}{K}}{\sigma_{min}^2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{MK}\sum_{k=1}^{K}\sum_{m=1}^{M}p_{i,k,m}\right)BN_0+\frac{(BN_0)^2}{n}}\right) \quad (3.92)$$

$$= \log\left(1+\frac{\sigma_{max}^4\frac{\pi^2}{16}n\left(\frac{\eta_{max}\sigma_{max}^2PL}{\eta_{min}\sigma_{min}^2PL+WN_0}\right)^2\frac{(1-\eta_{min})L}{L-1}P}{\sigma_{min}^2WN_0+\frac{(WN_0)^2}{nP}}\right). \quad (3.93)$$

We note that the RHS of (3.93) does not depend on $p_{i,k,m}$, which indicates that the LHS of (3.92) can be maximized by setting $p_{i,k,m} = P/K$, $1 \leq i \leq n$, $1 \leq k \leq K$, $1 \leq m \leq M$. Then it follows that the achievable rate, R^{WB*}, is upper bounded by

$$\begin{aligned} R^{WB*} &\leq \max_{\{p_{i,k,m}\}}\frac{KB(L-1)}{MKL}\sum_{m=1}^{M}\sum_{k=1}^{K}\log\left(1+S'_{k,m}\right) \\ &= \frac{W(L-1)}{L}\log\left(1+\frac{\sigma_{max}^4\frac{\pi^2}{16}n\left(\frac{\eta_{max}\sigma_{max}^2PL}{\eta_{min}\sigma_{min}^2PL+WN_0}\right)^2\frac{(1-\eta_{min})L}{L-1}P}{\sigma_{min}^2WN_0+\frac{(WN_0)^2}{nP}}\right). \quad (3.94) \end{aligned}$$

For convenience, define

$$R_{upper}^{WB\ *} \triangleq W\log\left(1+\frac{\sigma_{max}^4\frac{\pi^2}{16}n\left(\frac{\eta_{max}\sigma_{max}^2PL}{\eta_{min}\sigma_{min}^2PL+WN_0}\right)^2\frac{(1-\eta_{min})L}{L-1}P}{\sigma_{min}^2WN_0+\frac{(WN_0)^2}{nP}}\right). \quad (3.95)$$

[A Lower Bound on R^{WB*}]:

Since R^{WB*} is the maximum achievable rate when the relay node can allocation its power across the fading blocks and across the subchannels, any other achievable rate serves as a lower bound on R^{WB*}. Then it follows that

$$R^{WB*} \geq R_{Eq}^{WB} \ . \tag{3.96}$$

Observe that the scaling order of R_{Eq}^{WB} in (3.76) is in fact the same as that of ${R_{upper}^{WB}}^{*}$ in (3.95) (the proof is relegated to Appendix E), we conclude that in the asymptotic regime of n, L and W,

$$R_{Eq}^{WB} = \Theta({R_{upper}^{WB}}^{*}) \ . \tag{3.97}$$

Somewhat surprisingly, as shown in (3.97), the scaling order of the "optimal" achievable rate R^{WB*} is exactly the same as that of R_{Eq}^{WB}. That is to say, even if the relay nodes can optimally allocate the power across the fading blocks and among the subbands, the scaling order of the corresponding achievable rate is the same as that by applying the equal power allocation policy at the relay nodes.

3.4.4. Scaling Laws and Discussions

We have a few important observations are in order. First, the findings in Section 3.4.2 reveal that if $e_2 \geq e_3$ and $e_3 < 1$, then the achievable rate by using AF with network training, under the equal power allocation policy at relay nodes, has the same scaling order as the upper bound given by Lemma 3.3. That is to say, the scaling laws can be characterized

accordingly. More specifically, as $n \to \infty$, if there exist $\delta \geq 0$, and $0 < \epsilon < 1$, such that $\frac{L}{W} \overset{\bullet}{\geq} n^{\delta}$ and $\frac{n}{W} \overset{\bullet}{\geq} n^{\epsilon}$, then as $n \to \infty$, the capacity of wideband relay networks scales as

$$C^{WB} = \Theta\left(W \log\left(\frac{n}{W}\right)\right). \tag{3.98}$$

Under this sufficient condition, assuming equal power allocation across the subbands, AF with network training can yield an achievable rate of the same scaling order as that of the upper bound, thus enabling us to characterize the scaling law. Furthermore, combing the results in Sections 3.4.2 and 3.4.3, we conclude that the conditions for achieving the scaling law under optimal power allocation are the same as that under equal power allocation. In a nutshell, in power-constrained wideband relay networks using AF with network training, the equal power allocation policy at relay nodes can result in the same scaling order as the optimal power allocation policy at relay nodes. However, it should be cautioned that from the point of view of exact achievable rates (not only the scaling orders), the wideband relay networks are not a simple collection of narrowband relay networks with equal power allocation.

We observe that in general, in a power-constrained wideband relay network where AF is employed, a larger bandwidth corresponds to a smaller amplification factor for relaying, which in turn may reduce the power of the desired signal embedded in the amplified signal from each relay node, and therefore the equivalent overall SNR at the destination node may or may not increase, depending on how W and n scale together. As shown in the above sections, the power of the aggregated "amplified" noise from the relay nodes is of $\Theta(n)$, and the power of the ambient noise at the destination node is of $\Theta(W)$. When W is sublinear in n, the "amplified" noise from all the relay nodes dominates the ambient

noise at the destination node. As a result, all forward links are virtually "noise-free". Such "noise-free" links effectively convert all the relay nodes and the destination node into an antenna array of n receive antennas. Furthermore, observe that the relay nodes carry out phase alignment, and the equivalent virtual antenna array effectively conducts maximal-ratio combining (MRC). Thus, the relay network functions as if it were a single-input-multiple-output system (SIMO) with optimal decoding at the receiver. Recall that the upper bound in Lemma 3.3, given by the Cut-Set Theorem, assumes that an optimal joint decoding is conducted by all receivers, i.e., the upper bound corresponds to the capacity of a SIMO system. It is now not surprising that under the "noise-free" conditions mentioned above, the lower bound can meet the upper bound.

3.5. Numerical Examples

We now illustrate via numerical examples the achievable rates and upper bounds in Theorem 3.1, Theorem 3.2, Lemma 4.1 and Lemma 5.1. The key parameters used in the examples are listed as below:

- transmission power of the source node and the relay nodes: $P = 1$ dBm;
- bandwidth of each subband: $B = 10$ kHz;
- noise power spectral density: $N_0 = 10^{-10}$ watts/Hz;
- the number of relay nodes n ranges from 10^2 to 10^4;
- the SNR per link in narrowband relay networks: $\rho = n^{-0.7}$; the number of subbands in wideband relay networks: $K = n^{0.7}$; and coherence interval: $L = n^{1.5}$. That is,

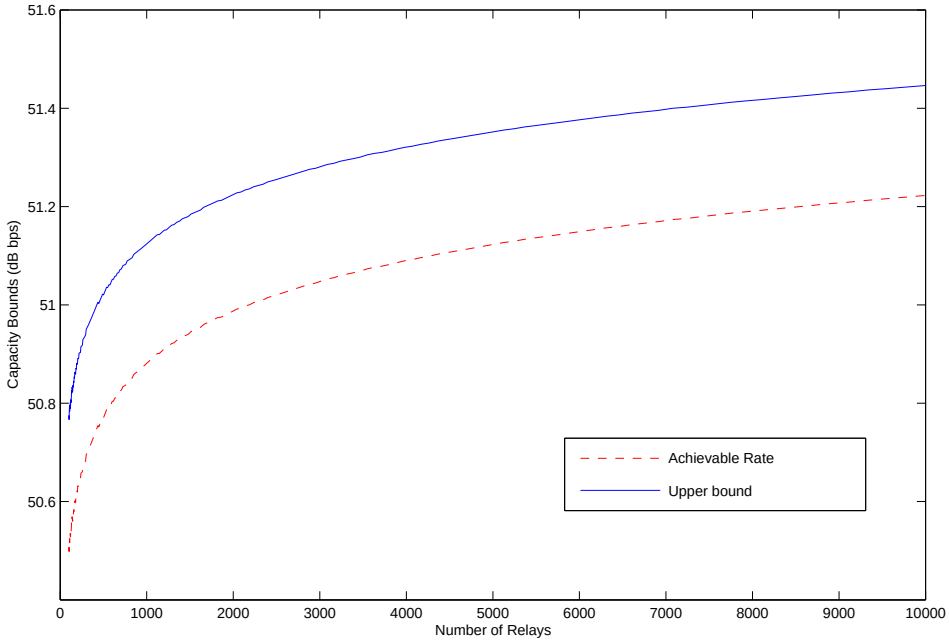

Figure 3.5. Narrowband relay networks: achievable rates and upper bounds

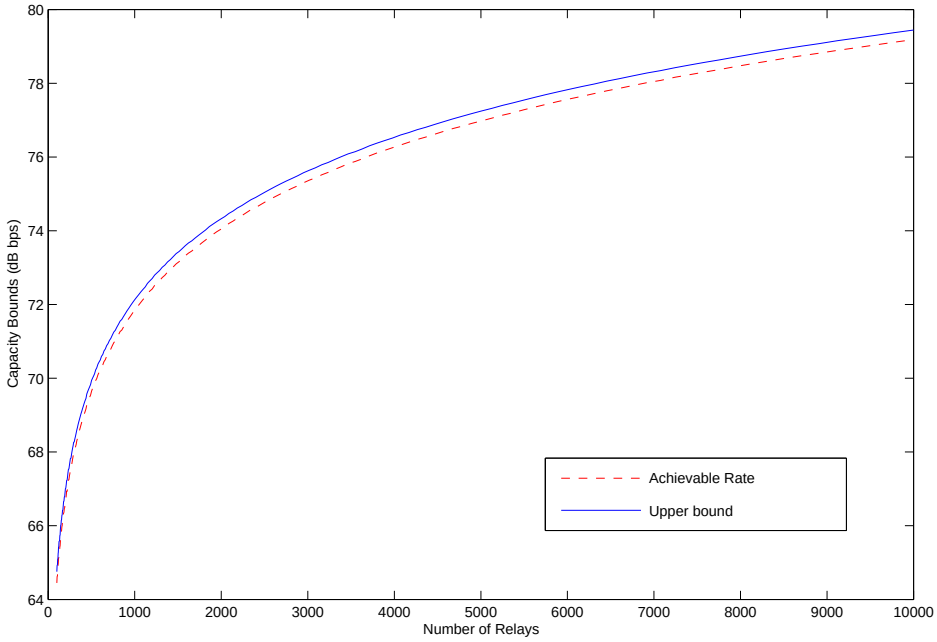

Figure 3.6. Wideband relay networks: achievable rates and upper bounds

$e_1 = e_3 = 0.7$ and $e_2 = 1.5$, which meet the scaling-law-achieving conditions given in Theorem 3.1 and Theorem 3.2;

- the unit of all capacity bounds is dB bits/second.

Fig. 3.5 depicts the lower bounds and the upper bounds on the capacity for a narrowband relay network model. Worth noting is that "dB bits/second" is used as the unit, because the main goal is to study the scaling order. It can be seen that when the number of relay nodes exceeds some threshold (2500 in this example), the two curves have the same slope, i.e., the lower bound and the upper bound have the same scaling order, corroborating the results in Theorem 3.1.

Fig. 3.6 illustrates the lower bound and the upper bound on the capacity for a power-constrained wideband relay network. It can be seen that the upper bounds and the lower bounds become "parallel" when n exceeds approximately 3000, indicating that the lower bounds and the upper bounds have the same scaling order asymptotically.

3.6. Conclusion

In this chapter, we study the achievable rates and scaling laws of large-scale wireless relay networks in the wideband regime. Assuming that relay nodes have no *a priori* knowledge of channel state information (CSI) for both the backward channels and the forward channels, we examine the achievable rates in the joint asymptotic regime of the number of relay nodes n, the channel coherence interval L, and the SNR per link ρ, or the system bandwidth W. We first study narrowband relay networks in the low SNR regime. We study a relaying scheme, namely amplify-and-forward (AF) with network training, in which

the source node and the destination node broadcast training symbols and each relay node carries out channel estimation and then uses AF relaying to relay information. We provide an equivalent source-to-destination channel model, and characterize the corresponding achievable rate. Our findings show that when ρL, proportional to the transmission energy in each fading block, is bounded below, the achievable rate has the same scaling order as in coherent relaying, thus enabling us to characterize the scaling law of the relay networks in the low SNR regime.

We then generalize the study to power-constrained wideband relay networks, where frequency-selective fading is taken into account. In particular, we examine the scaling behavior of the achievable rates corresponding to two power allocation policies applied at relay nodes, namely, a simple equal power allocation policy and the optimal power allocation policy. We identify the conditions under which the scaling law of the wideband relay networks can be achieved by both power allocation policies. Somewhat surprising, our findings indicate that these two power allocation policies result in achievable rates of the same scaling behavior, and the scaling law can be characterized under the condition that L/W, proportional to the energy per fading block per subband, is bounded below and that W is sub-linear in n.

Throughout this chapter, we have focused on the achievable rates using amplify-and-forward (AF) with network training. Another popular candidate for information relaying is decode-and-forward (DF). In the DF strategy, the relay nodes first extract the information bits from the received signals, re-encode it, and then transmit the new encoded bits in the second hop. In general, the optimal DF strategy is highly non-trivial and corresponding complexity can increase significantly as the number of relay nodes grows.

CHAPTER 4

Communication-Sensing Tradeoff in Wireless Sensor Networks

4.1. System Model

As depicted in Fig. 4.1, n sensors are used to monitor some physical phenomena (sources). We assume that the phenomena can be characterized by a sequence of discrete time random vectors $\{(s_0(t), s_2(t), ...s_{M-1}(t)), t = 1, 2, ...\}$, where t is the time index [16]. Assume that the random process is stationary and the vectors are temporally i.i.d.. The observations generated at the sensor nodes are denoted by $\{x_0(t), x_1(t), ...x_{n-1}(t)\}$ and the received signal at the sink is represented by $y(t)$. Assume that the observations are memoryless in the sense that the observations at time t only depend on the physical sources at time t, and that $y(t)$ is also memoryless, indicating the statistics are invariant in t. For the simplicity of notation, let $\{s_0, s_2, ...s_{M-1}\}$ denote the random vector of the physical sources at one snapshot, and let $\{x_0, x_1, ...x_{n-1}\}$ and y stand for the corresponding observations and received signal at the sink, respectively.

A central metric of concern here is the fidelity of the reconstructed process with respect to the physical sources. In this chapter, we focus on characterizing the mutual information

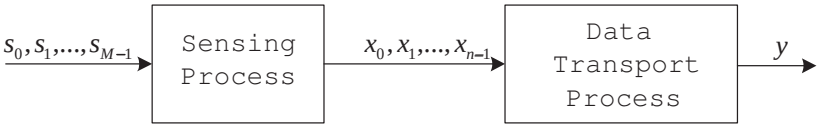

Figure 4.1. A sensor network model

between $\{s_0, s_2, ...s_{M-1}\}$ and y. Indeed, it is shown in [9] that the minimum-mean-square-error (MMSE) is inherently related to the mutual information in an additive Gaussian noise channel.

Due to limited transmit power, some sensor nodes cannot reach the sink directly, and thus their data need to be relayed by other nodes (namely relay nodes). More specifically, data from the nodes that cannot reach the sink directly arrive at the relay nodes in the first hop, and then the relay nodes process the received data and their own sensed data, after which the processed data are transmitted to the sink in the second hop.

Note that the mutual information of interest is given by

$$I(s_0, s_1, ...s_{M-1}; y) = h(s_0, s_1, ...s_{M-1}) - h(s_0, s_1, ...s_{M-1}|y), \tag{4.1}$$

which boils down to minimizing $h(s_0, s_1, ...s_{M-1}|y)$. The conditional entropy $h(s_0, s_1, ...s_{M-1}|y)$ depends on two processes: the sensing process (from $\{s_0, s_1, ...s_{M-1}\}$ to $\{x_0, x_1, ...x_{n-1}\}$) and the transport process (from $\{x_0, x_1, ...x_{n-1}\}$ to y). Note that at each relay node there are two kinds of signals, i.e., its own sensed signal and the received signal. A key step here is to investigate the joint processing of these two signals. We show that this boils down to the problem of balancing the tradeoff between sensing and communication.

For ease of exposition, we consider the cases where there is only one physical source s to be monitored. Depending on the observability of the physical source, we study accordingly

two important models. Specifically, we start with a basic network model, in which the observation at each node is an additive noisy version of the physical source (namely the source variable plus noise) [1]. Then we generalize the case to the sensor network over a two-dimensional random field, where the physical source and the observations follow some joint distribution [15].

4.2. Event-driven Sensor Relay Networks

As illustrated in Fig. 4.2, when an even occurs, all sensor nodes generate data and then the data are transmitted to the sink through the network. We assume that there are n sensor nodes, each with average power constraint p. Let node 0 denote the "critical node" that is closest to the event.[1] Denote the channel gain corresponding to the link from the 0th node to the ith node as α_i, and the channel gain corresponding to the link from the ith node to the sink as β_i. Assume that $\{\alpha_i\}$ and $\{\beta_i\}$ keep constant during the data transmission, and α_i and β_i are available at the ith relay node. For simplicity, we assume that $\{\alpha_i\}$ and $\{\beta_i\}$ are real numbers, which can be made true by doing phase compensation at the corresponding relay node [52].

4.2.1. Additive Noisy Signal Model

The observation of s at the ith sensor node is

$$x_i = s + n_i, \quad i = 0, 1, ..., n-1, \tag{4.2}$$

[1]There can be a number of "critical nodes" in the network. Our results presented here can be generalized to the case where multiple critical nodes exist.

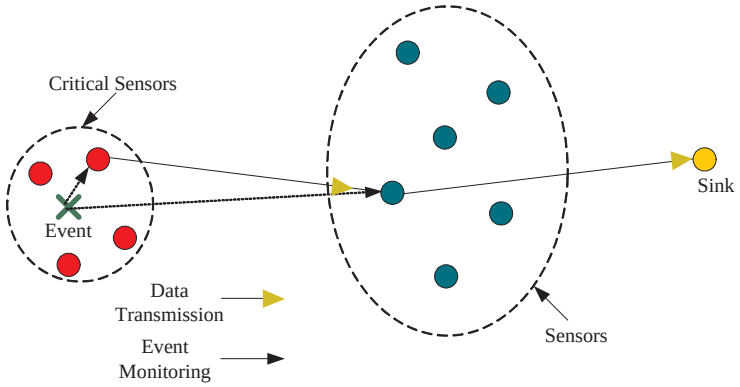

Figure 4.2. Event-driven sensor networks

where s is a random variable with zero mean and variance p_s. n_i denotes the noise at the ith sensor node and is complex Gaussian with zero mean and variance σ_i^2.

The received signal at the ith relay node is given by

$$y_i = \alpha_i x_0 + w_i, \quad i = 1, ..., n-1, \tag{4.3}$$

where w_i is complex Gaussian with zero mean and variance N. After processing the received signal y_i and its own sensed signal x_i jointly, the ith node transmits x_i', and the received signal at the sink is given by

$$y = \sum_{i=1}^{n-1} \beta_i x_i' + w, \tag{4.4}$$

where w is complex Gaussian with zero mean and variance N, and β_i is the channel gain corresponding to the link from the ith node to the sink.

4.2.2. Amplify and Forward with Linear Combination (AFLC)

Given the above signal model, it is natural to ask:

- How each relay node should process jointly the received signal and its own sensed signal?

We note that the received signal (4.3) and the sensed signal (4.2) at the i relay node contain the same variable s. In this study, we focus on a scheme where the ith node combines the received signal and its sensed signal using different weights that meet the power constraint. We refer to this scheme as "amplify-and-forward with linear combining"(AFLC), where the generated signal $x_i'(t)$ at the ith node is given by

$$\begin{aligned} x_i' &= A_i y_i + \bar{A}_i x_i \\ &= (\bar{A}_i + A_i\alpha_i)s + \bar{A}_i n_i + A_i\alpha_i n_0 + A_i w_i, \end{aligned} \tag{4.5}$$

where A_i and $\bar{A}$ are the amplification factors for the received signal and the sensed signal at the ith node, respectively. Accordingly, the received signal at the sink is given by

$$\begin{aligned} y = &\sum_{i=1}^{n-1} \beta_i \left(\bar{A}_i + A_i\alpha_i\right) s \\ &+ \sum_{i=1}^{n-1} \beta_i \left(\bar{A}_i n_i + A_i\alpha_i n_0 + A_i w_i\right) + w. \end{aligned} \tag{4.6}$$

Note the noise n_0, n_i, w_i and w are independent. Furthermore, recall that the channel conditions are fixed during the communication. Therefore, (4.6) corresponds to an additive Gaussian noise channel model, and the corresponding overall SNR is given by

$$SNR_{overall} \triangleq \frac{\left|\sum_{i=1}^{n-1} \beta_i \left(\bar{A}_i + A_i\alpha_i\right)\right|^2 p_s}{\mathrm{var}\left(\sum_{i=1}^{n-1} \beta_i \left(\bar{A}_i n_i + A_i\alpha_i n_0 + A_i w_i\right) + w\right)}. \tag{4.7}$$

Needless to say, a key step in maximizing the mutual information $I(s;y)$ is to find the optimal $\{A_i, \bar{A}_i\}$. Next, we show that the mutual information is an increasing function of

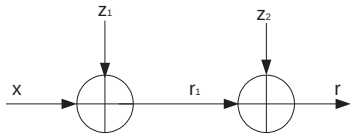

Figure 4.3. A degraded Gaussian noise channel

the SNR. To this end, consider an additive Gaussian noise channel illustrated in Fig. 4.3,

$$r_1 = x + z_1 \tag{4.8}$$

$$r = x + z_1 + z_2, \tag{4.9}$$

where z_1 and z_2 are independent complex Gaussian, and x is input signal with a power constraint. Clearly, the SNR for the first channel (from x to r_1) is higher than that for the cascaded channel (from x to r). Observe that $x \to r_1 \to r$ forms a Markov chain. By the data processing inequality [7], we have that $I(x; r_1) > I(x; r)$.

In a nutshell, maximizing mutual information $I(s; y)$ over $\{A_i, \bar{A}_i\}$ is equivalent to maximizing $SNR_{overall}$ in (4.7).

4.2.3. Optimal Power Allocation for AFLC

For convenience, define the *output* SNR ρ_i corresponding to (4.5) as

$$\rho_i \triangleq \frac{\left(\bar{A}_i + A_i\alpha_i\right)^2 p_s}{\mathrm{var}\left(\bar{A}_i n_i(t) + A_i\alpha_i n_0(t) + A_i w_i(t)\right)}. \tag{4.10}$$

Rewrite the signal x_i' in (4.5) as

$$x_i' = \sqrt{\frac{p}{1+\rho_i}}(\sqrt{\rho_i}s' + n_i'), \tag{4.11}$$

where n_i' is complex Gaussian with zero mean and unit variance, and $s' = s/\sqrt{p_s}$ is the power-normalized version of s. Accordingly, the overall SNR at the sink is given by

$$SNR_{overall} = \frac{\left(\sum_{i=1}^{n-1} \beta_i \sqrt{\rho_i/(1+\rho_i)}\right)^2}{\sum_{i=1}^{n-1} \beta_i^2/(1+\rho_i) + N/p}. \tag{4.12}$$

It can be shown that the overall SNR is an increasing function of each ρ_i for $i = 1, 2, ..., n-1$. Thus, we conclude that maximizing the output SNR ρ_i at the ith node locally is necessary for maximizing the overall SNR ρ at the sink.

Next, we turn to the power allocation policy. Recall that

$$\rho_i = \frac{\left(\bar{A}_i^2 + 2\alpha_i \bar{A}_i A_i + A_i^2 \alpha_i^2\right) p_s}{\bar{A}_i^2 \sigma_i^2 + A_i^2 \alpha_i^2 \sigma_0^2 + A_i^2 N} \tag{4.13}$$

$$= \frac{p_s}{\sigma_i^2}\left[1 + 2\alpha_i \left(\frac{\eta_i + a_i}{\eta_i^2 + b_i}\right)\right], \tag{4.14}$$

where

$$a_i \triangleq \frac{1}{2\alpha_i}\left(\alpha_i^2 - \frac{\alpha_i^2 \sigma_0^2 + N}{\sigma_i^2}\right) \tag{4.15}$$

$$b_i \triangleq \frac{\alpha_i^2 \sigma_0^2 + N}{\sigma_i^2} \tag{4.16}$$

$$\eta_i \triangleq \frac{A_i}{\bar{A}_i} \tag{4.17}$$

Let A_i^* and $\bar{A}_i^*$ correspond to the optimal power allocation at the ith node. After some algebra, we have that

$$A_i^* = \sqrt{\frac{p}{\eta_i^{*2}(p_s + \sigma_i^2) + \alpha_i^2 p + N}}, \tag{4.18}$$

$$\bar{A}_i^* = \eta_i^* \sqrt{\frac{p}{\eta^{*2}(p_s + \sigma_i^2) + \alpha_i^2 p + N}}, \tag{4.19}$$

where $\eta_i^* = \sqrt{a_i^2 + b_i} - a_i$.

4.2.4. Discussion

Case I: If the sensed data gathered at the relay nodes are much noisier than the received signal from the 0th node, then it follows that

$$\eta_i^* \rightarrow 0 \tag{4.20}$$

$$A_i^* \rightarrow \sqrt{\frac{p}{\alpha_i^2 p + N}}, \tag{4.21}$$

$$\bar{A}_i^* \rightarrow 0. \tag{4.22}$$

It is clear that in this case, the sensor network boils down to a relay network studied in [49, 52, 14].

Case II: Another interesting case is that the channel condition corresponding to the link from the critical node to the ith relay node is poor, i.e. $\alpha_i \rightarrow 0$. It turns out that the received data which is transmitted from the critical node is noisier than the sensed data at the sensor nodes. In this case, we have that

$$\lim_{\alpha_i \rightarrow 0} \eta_i^* \alpha_i = \sqrt{\frac{1}{4}(\alpha_i^2 - b)^2 + \alpha_i^2 b} - \frac{1}{2}(\alpha_i^2 - b) \tag{4.23}$$

$$= \frac{N}{\sigma_i^2}, \tag{4.24}$$

or equivalently,

$$\lim_{\alpha_i \rightarrow 0} \eta_i^* = \Theta\left(\frac{1}{\alpha_i}\right). \tag{4.25}$$

Therefore, as $\alpha_i \rightarrow 0$, the amplification factors for relaying information and transmitting sensed data are given by

$$\lim_{\alpha_i \rightarrow 0} A_i^* = 0, \tag{4.26}$$

$$\lim_{\alpha_i \rightarrow 0} \bar{A}_i^* = \sqrt{\frac{p}{p_s + \sigma_i^2}} \tag{4.27}$$

In this scenario, as the signals transmitted from the 0th node become noisy, the nodes tend to allocating full power for their own sensed signals.

4.3. Sensor Networks over a Two-Dimensional Random Field

Consider a sensor network over a two-dimensional random field. Let $p(x_0, x_1, ...x_{n-1}|s)$ denote the conditional pdf of the observations $\{x_0, x_1, ...x_{n-1}\}$, where $\{x_0, x_1, ...x_{n-1}\}$ are not necessarily the additive noisy version of s, as studied in the event-driven sensor network model.

4.3.1. Estimate and Forward

Note that at the ith relay node there are in fact two "observations" of the physical source s, namely the sensed signal (denoted as x_i), and the received signal from the senders in the first hop (denoted as y_i). In this scenario, the ith relay node can first estimate s, and then transmits the estimate $\hat{s}_i$ to the sink. Assume that the minimum mean-square error (MMSE) estimator is used at each relay node. Then, it follows that $\hat{s}_i$ is the conditional expectation of s given x_i and y_i,

$$\hat{s}_i = \mathbb{E}(s|x_i, y_i) \tag{4.28}$$

$$= \int sp(s|x_i, y_i)ds. \tag{4.29}$$

Let C_i denote the set containing the indices of the nodes that transmit signals to the ith node in the first hop. Let $\{x_j\}_{j\in C_i}$ stand for the transmitted signals from the nodes in C_i to the ith node in the first hop, and let $p(y_i|\{x_j\}_{j\in C_i})$ represent the transition probability corresponding to the communication channel from the nodes in C_i to the ith node.

It is clear that to characterize $\hat{s}_i$, we need $p(s|x_i, y_i)$ that is given by

$$p(s|x_i, y_i) = \frac{\int p(s, x_i, y_i, \{x_j\}_{j\in C_i}) d\{x_j\}_{j\in C_i}}{\int\int p(s, x_i, y_i, \{x_j\}_{j\in C_i}) d\{x_j\}_{j\in C_i} ds}, \tag{4.30}$$

where $p(s, x_i, y_i, \{x_j\}_{j\in C_i})$ can be obtained by

$$\begin{aligned} p(s, x_i, y_i, \{x_j\}_{j\in C_i}) &= p(x_i|s, y_i, \{x_j\}_{j\in C_i}) p(s, y_i, \{x_j\}_{j\in C_i}) \\ &\overset{(a)}{=} p(x_i|s) p(y_i|s, \{x_j\}_{j\in C_i}) p(s, \{x_j\}_{j\in C_i}) \\ &\overset{(b)}{=} p(x_i|s) p(y_i|\{x_j\}_{j\in C_i}) p(\{x_j\}_{j\in C_i}|s), \end{aligned} \tag{4.31}$$

where (a) follows that $p(x_i|s, y_i, \{x_j\}_{j\in C_i}) = p(x_i|s)$, which corresponds to the sensing process at the ith node; and (b) follows that $p(y_i|s, \{x_j\}_{j\in C_i}) = p(y_i|\{x_j\}_{j\in C_i})$, which points to the data transmission from the nodes in C_i to the ith relay node.

Generally speaking, it may not be possible to find a closed form for (4.30) due to the high-dimensional integrals involved. In what follows, we focus on the Gaussian field case, where $\{s, x_0, ..., x_{n-1}\}$ are jointly Gaussian.

4.4. Gaussian Sensor Networks in 2-D Random Field

4.4.1. System Model

We now consider an interesting model, where s and $x_0, x_1, ...x_{n-1}$ are jointly complex Gaussian. In this case, a key observation is that the observation vector $X \triangleq [x_0, x_1, ...x_{n-1}]'$ can be rewritten as

$$X = \frac{1}{p_s} \operatorname{cov}(X, s) s + \Omega, \tag{4.32}$$

$$\operatorname{cov}(\Omega) = \operatorname{cov}(X) - \frac{1}{p_s} \operatorname{cov}(X, s) \operatorname{cov}(s, X), \tag{4.33}$$

with $\mathrm{cov}(X, s) \triangleq \mathbb{E}[X^{\dagger} s]$. Ω is independent of s and zero-mean. That is, the observations can be expressed as the additive noisy versions of the source variable s. Furthermore, under the assumption that the communication channels are additive Gaussian noise channels, the two observations of the physical source at each relay node are both additive noisy versions of s. We note that to carry out MMSE estimation, each relay node needs to know the corresponding covariance between the observations and the source.

In what follows, we focus on two cases. For the first case, we assume that there is only one node as the sender in the first hop, and the other nodes act as relay nodes; we call this model sensory relay networks. In the second case, we consider a clustered sensor network, where the nodes can be made into clusters. In each cluster, there is a cluster head that can communicate with the sink, and the rest of the nodes send data to the cluster head only. Furthermore, we assume that there is no inter-cluster interference.

In what follows, we examine the use of the MMSE estimation at the relay node/cluster head and characterize the corresponding mutual information between the source and the sink.

4.4.2. Case I: Sensory Relay Networks

Consider a sensory relay network (see Fig. 4.4), where there is only one node (denoted as the 0th node) that cannot reach the sink directly. The channel gain corresponding to the link from the 0th node to the ith node is α_i, and that from the ith node to the sink is β_i. Assume that α_i and β_i are available at the ith node.

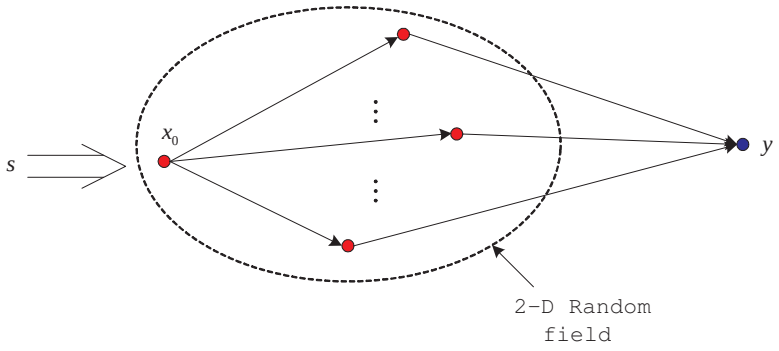

Figure 4.4. A sensory relay network in a 2-D random field

Recall that the sensed signal at the ith node can be rewritten as

$$x_i = \frac{\xi_{x_i s}}{p_s} s + \omega_i, \tag{4.34}$$

where $i = 0, 1, ...n$. ω_i is complex Gaussian, independent of s, and with zero mean and variance $\sigma_i^2 = p - |\xi_{x_i s}|^2/p_s$, where $\xi_{x_i s} \triangleq \mathbb{E}[x_i^* s]$. At the ith relay node, the received signal y_i from the 0th node is

$$\begin{aligned} y_i &= \alpha_i x_0 + z_i & (4.35)\\ &= \frac{\alpha_i \xi_{x_0 s}}{p_s} s + \alpha_i \omega_0 + z_i, & (4.36) \end{aligned}$$

where z_i is complex Gaussian with zero mean and variance N, and independent of s and $w_i,\ i = 1, 2, ...n-1$.

Rewrite (4.34) and (4.36) in a vector form:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \xi_{x_i s}/p_s \\ \alpha_i \xi_{x_0 s}/p_s \end{bmatrix} s + \begin{bmatrix} \omega_i \\ \alpha_i \omega_0 + z_i \end{bmatrix}, \tag{4.37}$$

The MMSE estimator at the ith relay node is given by

$$\begin{aligned} \hat{s}_i &= c_1 x_i + c_2 y_i & (4.38)\\ &= \frac{1}{p_s}(c_1 \xi_{x_i s} + c_2 \alpha_i \xi_{x_0 s}) s + c_1 \omega_i + c_2(\alpha_i \omega_0 + z_i) & (4.39) \end{aligned}$$

where c_1 and c_2 can by found by using the orthogonal principle of the MMSE estimator, and given by

$$c_1 = \frac{|\alpha_i|^2\sigma_0^2\xi^*_{x_is} + N\xi^*_{x_is}}{D} \tag{4.40}$$

$$c_2 = \frac{\alpha_i^*\xi^*_{x_0s}}{D} \tag{4.41}$$

with

$$D \triangleq \frac{1}{p_s}\left(|\alpha_i|^2|\xi_{x_0s}|^2\sigma_i^2 + |\alpha_i|^2|\xi_{x_is}|^2\sigma_0^2 + N|\xi_{x_is}|^2\right) + |\alpha_i|^2\sigma_0^2\sigma_i^2 + N\sigma_i^2. \tag{4.42}$$

The estimate $\hat{s}_i$ is amplified to meet with the transmit power constraint p and then transmitted to the sink. Denote the corresponding amplification factor as k_i for the signal transmitted from ith node. Combining (4.38) (4.40) and (4.41), we have that the transmitted signal from the ith node is given by

$$\begin{aligned} x_i' = & \underbrace{\frac{k_i}{Dp_s}((|\alpha_i|^2\sigma_0^2 + N)|\xi_{x_is}|^2 + |\alpha_i|^2|\xi_{x_0s}|^2)}_{h_i} s \\ & + \underbrace{\frac{k_i}{D}(|\alpha_i|^2\sigma_0^2\xi^*_{x_is} + N\xi^*_{x_is})\omega_i + \alpha_i^*\xi^*_{x_0s}(\alpha_i\omega_0 + z_i)}_{n_i}, \end{aligned} \tag{4.43}$$

where k_i is determined by $\mathrm{var}(x_i') = p$. Accordingly, the received signal at the sink is given by

$$y = \sum_{i=1}^{n-1} h_i\beta_i x_i' + \sum_{i=1}^{n-1}\beta_i n_i + z, \tag{4.44}$$

where z is complex Gaussian with zero mean and variance N. The corresponding mutual information between s and y is then

$$I(s;y) = \frac{1}{2}\log\left(1 + \frac{\left|\sum_{i=1}^{n-1} h_i\beta_i\right|^2 p_s}{\sum_{i=1}^{n-1}|\beta_i|^2\mathrm{var}(n_i) + N}\right). \tag{4.45}$$

It is clear that in (4.44) and (4.45), in order to maximize the receive SNR and thus the mutual information $I(s;y)$, the optimal strategy is to compensate the phase offset incurred by β_i by choosing proper k_i, so that coherent combining can be carried out at the sink. In a nutshell, k_i should be chosen by taking into account the transmit power constraint p and the phase compensation for β_i.

Discussion: EF vs. AFLC. Recall that in a sensory relay network over a 2-D Gaussian field, the observations $x_i, i = 0, 1, ...n-1$ can be expressed as the additive noisy versions of s. The estimate and forward strategy boils down to determining the weights for the two observation signals at each relay node to minimize the estimate error. We also note that the underlined optimal power allocation AFLC requests to maximize the output SNR at each relay node by combining the received signal and the sensed signal. It is then natural to ask that if the two schemes can yield the same achievable rates. It can be shown that for the sensory relay network that monitors one physical source, the MMSE estimator yields the weights that can also maximize the output SNR at each relay node. Therefore, in this case, EF relaying and AFLC relaying can achieve the same mutual information between s and y.

Recall that in the sensory relay network, where one Gaussian source is monitored, the sensed signal and the received signal at one relay node are both noisy versions of the physical source s. Then we can rewrite the two observations at the ith relay node as

$$x_i = h_{x_i} s + v_1 \tag{4.46}$$

$$y_i = h_{y_i} s + v_2, \tag{4.47}$$

where v_1 and v_2 are complex Gaussian with zero mean and variance $\sigma_{v_1}^2$ and $\sigma_{v_2}^2$, respectively.

For ease of exposit, assume that all variables are real.

First, consider that AFLC relaying is applied for the sensory relay network. Let A_1 and A_2 as the coefficients for x_i and y_i corresponding to AFLC scheme. The output SNR is given by

$$\begin{aligned}
\rho_i &= \frac{(A_1 h_{x_i} + A_2 h_{y_i})^2 p_s}{A_1^2 \sigma_{v_1}^2 + A_2^2 \sigma_{v_2}^2} && (4.48)\\
&= \frac{(A_1 \sigma_{v_1} \frac{h_{x_i}}{\sigma_{v_1}} + A_2 \sigma_{v_2} \frac{h_{y_i}}{\sigma_{v_2}})^2 p_s}{A_1^2 \sigma_{v_1}^2 + A_2^2 \sigma_{v_2}^2} \\
&\overset{(a)}{\leq} \frac{(A_1^2 \sigma_{v_1}^2 + A_2^2 \sigma_{v_2}^2)(h_{x_i}^2/\sigma_{v_1}^2 + h_{y_i}^2/\sigma_{v_2}^2) p_s}{A_1^2 \sigma_{v_1}^2 + A_2^2 \sigma_{v_2}^2} \\
&= \left(\frac{h_{x_i}^2}{\sigma_{v_1}^2} + \frac{h_{y_i}^2}{\sigma_{v_2}^2} \right) p_s, && (4.49)
\end{aligned}$$

where (a) follows the Cauchy-Schwarz inequality.

Next, we assume that the estimate-and-forward relaying is applied, and the MMSE estimate of s is given by

$$\hat{s}_{MMSE} = \frac{\left(\frac{h_{x_i}^2}{\sigma_{v_1}^2} + \frac{h_{y_i}^2}{\sigma_{v_2}^2} \right) s + \frac{h_{x_i}}{\sigma_{v_1}^2} v_1 + \frac{h_{y_i}}{\sigma_{v_2}^2} v_2}{1 + \frac{h_{x_i}^2}{\sigma_{v_1}^2} + \frac{h_{y_i}^2}{\sigma_{v_2}^2}}. \quad (4.50)$$

It can be shown that the EF relaying scheme yield the same output SNR as the AFLC relaying scheme.

We note that when multiple sources exist in the random field, the weights corresponding to the AFLC scheme depend on the global information on the channel conditions, because AFLC is devised to maximize the overall SNR at the sink. As a result, the optimization could not be carried out independently at each relay node. In this case, EF relaying provides a simpler approach to process the two observations at the relay nodes.

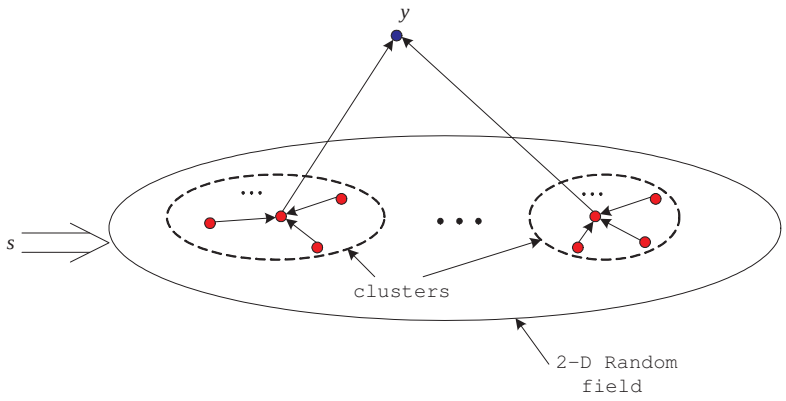

Figure 4.5. A cluster sensor network in a 2-D random field

4.4.3. Case II: Clustered Sensor Networks

In this section, we consider the setting depicted in Fig.4.5, where there are few nodes that can reach the sink directly and indeed these few nodes act as the gateways to relay the data from the other nodes. Furthermore, the sensor nodes whose data need to be relayed can form clusters, and in each cluster there is one cluster head that can communicate with the sink. For simplicity, we assume that there is no inter-cluster interference. That is, each cluster head only receives the transmitted data from the nodes in its own cluster and the interference incurred by the ongoing transmissions in other clusters can be ignored. In this scenario, each cluster head performs MMSE estimation and then transmits the estimator $\hat{s}_i$ to the sink.

Assume that there are K clusters. Let C_i denote the set of the indices corresponding to the sensor nodes (excluding the cluster head) in the ith cluster. Let $\alpha_{ij}, j \in C_i$ represent the channel gain corresponding to the link from the jth node to the ith cluster head, and β_i stand for the channel gain with respect to the link from the ith cluster head to the sink.

The received signal at the cluster head in the ith cluster is then given by

$$y_i = \sum_{j \in C_i} \alpha_{ij} x_j + z_i \tag{4.51}$$

$$= \frac{1}{p_s} \sum_{j \in C_i} \alpha_{ij} \xi_{x_j s} s + \sum_{j \in C_i} \alpha_{ij} \omega_j + z_i. \tag{4.52}$$

Recall that the sensed signal at the ith cluster head is

$$x_i = \frac{\xi_{x_i s}}{p_s} s + \omega_i. \tag{4.53}$$

Then, we can rewrite (4.52) and (4.53) into a vector form,

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \xi_{x_i s}/p_s \\ \sum_{j \in C_i} \alpha_{ij} \xi_{x_j s}/p_s \end{bmatrix} s + \begin{bmatrix} \omega_i \\ \sum_{j \in C_i} \alpha_{ij} \omega_j + z_i \end{bmatrix}, \tag{4.54}$$

Following the same line as in the previous case, given x_i and y_i the MMSE estimator yields the coefficients as

$$d_1 = \frac{\sum_{j \in C_i} |\alpha_{ij} \sigma_j|^2 \xi^*_{x_i s} + N \xi^*_{x_i s}}{F} \tag{4.55}$$

$$d_2 = \frac{\sum_{j \in C_i} \alpha^*_{ij} \xi^*_{x_j s}}{F}, \tag{4.56}$$

where

$$F \triangleq \frac{1}{p_s} \left(\left| \sum_{j \in C_i} \alpha_{ij} \xi_{x_j s} \right|^2 \sigma_i^2 + |\xi_{x_i s}|^2 \sum_{j \in C_i} |\alpha_{ij} \sigma_j|^2 + N |\xi_{x_i s}|^2 \right) + \sum_{j \in C_i} |\alpha_{ij} \sigma_j|^2 \sigma_i^2 + N \sigma_i^2. \tag{4.57}$$

Accordingly, the transmitted signal from the cluster head in the ith cluster is a scaled version of the MMSE estimator, i.e.,

$$x'_i = q_i (d_1 x_i + d_2 y_i) \tag{4.58}$$

$$= \underbrace{\frac{q_i}{p_s} \left(d_1 \xi_{x_i s} + d_2 \sum_{j \in C_i} \alpha_{ij} \xi_{x_j s} \right)}_{g_i} s + \underbrace{q_i d_1 \omega_i + q_i d_2 \left(\sum_{j \in C_i} \alpha_{ij} \omega_j + z_i \right)}_{\theta_i}. \tag{4.59}$$

Then, it follows that the received signal at the sink is

$$y = \sum_{i=1}^{K} g_i \beta_i x'_i + \sum_{i=1}^{K} \beta_i \theta_i + z \tag{4.60}$$

where z is complex Gaussian with zero mean and variance N. The corresponding mutual information between s and y is then

$$\begin{aligned} & I(s;y) \\ &= \frac{1}{2}\log\left(1 + \frac{\left|\sum_{i=1}^{M} g_i\beta_i\right|^2 p_s}{\sum_{i=1}^{M}|\beta_i|^2 \mathrm{var}(\theta_i) + N}\right) \qquad (4.61) \\ &= \frac{1}{2}\log\left(1 + \frac{\frac{1}{p_s}\left|\sum_{i=1}^{M} q_i\beta_i\left(d_1\xi_{x_i s} + d_2\sum_{j\in C_i}\alpha_{ij}\xi_{x_j s}\right)\right|^2}{\sum_{i=1}^{M}|\beta_i|^2\left(|q_i d_1|^2\mathrm{var}(\omega_i) + |q_i d_2|^2\left(\sum_{j\in C_i}|\alpha_{ij}|^2\mathrm{var}(\omega_j) + N\right)\right) + N}\right), \qquad (4.62) \end{aligned}$$

In this case, we note that the phase alignment at the relay node is still a necessity to maximize the mutual information in (4.62). That is, the amplification factor q_i should be chosen according to the phase of the channel gain β_i the power constraint p. Furthermore, we observe that if $\alpha_{ij}\xi_{x_j s}$ is positive, the mutual information would be higher. To this end, the jth (for $j \in C_i$) node in the ith cluster can send $\frac{\alpha^*_{ij}\xi^*_{x_j s}}{|\alpha_{ij}\xi_{x_j s}|}x_j$, so that the phase of the transmitted signals in the ith cluster can be aligned. Indeed, this can be made possible by using network training so that the jth node can compensate the phase offset incurred by $\alpha_{ij}\xi_{x_j s}$.

Discussion: In the above studies, all the cluster heads are assumed to participate in data transport. Examining (4.62) further reveals that some cluster heads may contribute "more noise" than others, due to the inaccurate MMSE estimation and the poor channel condition β_i. Therefore, if each cluster head can be informed by the sink of whether to transmit or to keep silent, it is then of critical importance to characterize the optimal set of the participating cluster heads so as to maximize the mutual information. Simply put, we need

to solve the following optimization problem:

$$\max_{I_1,I_2,...I_M} \frac{\left|\sum_{i=1}^{M} I_i g_i \beta_i\right|^2 p_s}{\sum_{i=1}^{M} I_i |\beta_i|^2 \mathrm{var}(\theta_i) + N}$$
$$\text{subject to} \quad I_i \in \{0,1\}, \quad i = 1,2,...M \tag{4.63}$$

It is clear that $\{I_i\}$ depend on the channel conditions $\{\beta_i\}$ and the noise power $\{\mathrm{var}(\theta_i)\}$ that is determined by the MMSE estimator. That is, the global information on channel conditions and noise power at the different cluster heads are needed for the sink to determine which cluster heads are allowed to transmit.

Clearly, the above problem is non-convex. If M is small, and cluster heads can send $\{\beta_i\}$ and $\{\mathrm{var}(\theta_i)\}$ to the sink, then the sink can perform exhausted search for $\{I_i\}$ (with complexity of 2^M), thanks to the fact that $\{I_i\}$ are integers.

4.5. Downlink Transmissions

In some scenarios, it is also important to provide a mechanism for the sink to transmit to the nodes for the purpose of querying or to change the sensor settings.

Note that in this case each node is a destination and furthermore some nodes may not be in the transmission range of the sink within which the nodes can successfully decode the data directly sent by the sink, we correspondingly study a two-hop transmission scheme decode-and-forward (DF). Specifically, in the first hop, the sink broadcasts the messages to the network, the nodes that can hear the transmission would be able to decode the information correctly if their receive SNR is above certain threshold. In the second hop, those nodes that have successfully accessed the message transmit the same messages to the rest of the network. Indeed, in such a scenario, amplify-and-forward may not perform well

due to the fact that there can be multiple destination nodes in the second hop transmission, which makes it impossible to carry out coherent relaying since the phase alignment cannot be done uniformly across those potential destination nodes.

In what follows, we characterize the outage probability when a node fails decoding and examine how to diminish it.

4.5.1. Network Model

In the downlink transmission, we assume that the sink transmits the message at a fixed symbol rate R to the network, and the transmission lasts L symbols for each hop. Assume that each node has transmit power p. Let G_1 denote the set of nodes that can successfully decode the messages sent by the sink in the first hop, and G_2 stand for the rest of the nodes in the network. We assume that there is a link between any two nodes in the network, i.e., one node can always receive the signals from the sink and the other nodes in the network. Let $x^{DL}(t)$ denote the symbol transmitted from the sink. Without loss of generality, we assume that a Guassian input signaling is applied, which is known to the nodes across the network.

In the first hop, the received signal at the ith nodes is given by

$$y_i^{DL}(t) = \beta_i^{DL} x^{DL}(t) + z_i^{DL}(t),, \quad 1 \leq t \leq L, \tag{4.64}$$

where β_i^{DL} is the channel gain corresponding to the link from the sink to the ith node, and $z_i^{DL}(t)$ is complex Gaussian noise with zero mean and variance N.

At the end of the first hop, the ith node declares a success of decoding the messages if R is lower than the achievable rate corresponding to the link from the sink to the node.

That is, if

$$R < \log\left(1 + |\beta_i^{DL}|^2 \frac{p}{N}\right), \tag{4.65}$$

or equivalently,

$$|\beta_i^{DL}|^2 > (2^R - 1)\frac{N}{p}, \tag{4.66}$$

then a successful decoding is declared; otherwise if

$$R \geq \log\left(1 + |\beta_i^{DL}|^2 \frac{p}{N}\right), \tag{4.67}$$

a failure to decode is announced, and the corresponding node does not transmit in the second hop. Instead, it would keep listening to the other nodes in the second hop.

In the second hop, the nodes in G_1 transmit the recovered symbols $x^{DL}(t-L)$ to the rest of the network. The received signal at the jth ($j \in G_2$) node is given by

$$y_j^{DL}(t) = \sum_{i \in G_1} \alpha_{i,j}^{DL} x^{DL}(t-L) + z_j^{DL}(t) \tag{4.68}$$

where $L+1 \leq t \leq 2L$. $\alpha_{i,j}^{DL}$ is the channel gain corresponding to the link from the ith node in G_1 to the jth node in G_2, $z_j^{DL}(t)$ is complex Gaussian noise with zero mean and variance N. We can also rewrite the above model as

$$y_j^{DL}(t) = \sum_{i \in G_1 \cup G_2} \alpha_{i,j}^{DL} I_{\{|\beta_i^{DL}|^2 > (2^R-1)\frac{N}{p}\}} x^{DL}(t-L) + z_j^{DL}(t),\ j \in G_2, \tag{4.69}$$

where $I_{\{\}}$ is the indicator function.

4.5.2. When Channel Gains are i.i.d. Random Variables

In this case, we assume that $\{\alpha_{i,j}, 1 \leq i \leq n\}$ are i.i.d. complex random variables with zero mean and constrained variance $\sigma_{\alpha,j}^2$, and $\{\beta_i, 1 \leq i \leq n\}$ are complex i.i.d. random variables, with zero mean and constrained variance σ_β^2.

For convenience, define

$$h_j^{DL} \triangleq \frac{1}{\sqrt{n-1}} \sum_{i \in G_1 \cup G_1, i \neq j} \alpha_{i,j}^{DL} I_{\{|\beta_i^{DL}|^2 > (2^R-1)\frac{N}{p}\}}. \tag{4.70}$$

We note that by the Central Limit Theorem [59], as the size of the network $n \to \infty$, h_j^{DL} converges in distribution to a complex Gaussian random variable with zero mean and variance $\sigma^2_{h_j^{DL}} \triangleq \sigma^2_{\alpha,j} Pr\left(|\beta_i^{DL}|^2 > (2^R-1)\frac{N}{p})\right)$. Accordingly, we rewrite (4.69) as

$$y_j^{DL}(t) = \left(\frac{h_j^{DL}}{\sigma_{h_j^{DL}}}\right)\left(\sigma_{h_j^{DL}}\sqrt{n-1}x^{DL}(t-L)\right) + z_j^{DL}(t). \tag{4.71}$$

It is clear that as $n \to \infty$, the above channel model converges to a classic Rayleigh fading channel with the equivalent signal power that scales as $\Theta\left(\sigma^2_{h_j^{DL}}(n-1)\right)$.

We note that a key feature in the DF scheme is that for each node in the network, there can be two opportunities to correctly decode the messages. In fact, the probability of correct decoding for the ith node is given by

$$\begin{aligned} Pr\,(i\text{th node succeeds}) &= Pr\,(i \in G_1) \\ &+ Pr\,(i \in G_2)\, Pr\left(\left|\left(\frac{h_i^{DL}}{\sigma_{h_i^{DL}}}\right)\right|^2 > \frac{(2^R-1)N}{(n-1)\sigma^2_{h_i^{DL}}p}\right) \end{aligned} \tag{4.72}$$

As $n \to \infty$, the probability for the ith node fails to decode converges to

$$\begin{aligned} Pr\,(i\text{th node fails}) &= 1 - \exp\left(-\frac{(2^R-1)N}{\sigma^2_\beta p}\right) \\ &- \left(1-\exp\left(-\frac{(2^R-1)N}{\sigma^2_\beta p}\right)\right)\exp\left(-\frac{(2^R-1)N}{(n-1)\sigma^2_{h_i^{DL}}p}\right) \\ &= \left(1-\exp\left(-\frac{(2^R-1)N}{\sigma^2_\beta p}\right)\right)\left(1-\exp\left(-\frac{(2^R-1)N}{(n-1)\sigma^2_{h_i^{DL}}p}\right)\right) \end{aligned} \tag{4.73}$$

Indeed, we note that as $n \to \infty$, the outage probability decreases.

CHAPTER 5

Conclusion and Future Work

5.1. Contributions

Data relaying is a building block for multi-user communications in wireless networks. In this dissertation, we study three different relay models of wireless networks, including a MIMO relay channel, a wideband relay network and sensor networks over a random field.

In Chapter 2, we first study capacity bounds for the Gaussian MIMO relay channel with fixed channel gains. We derive an upper bound that involves convex optimization over two covariance matrices and *one* scalar parameter ρ. Loosely speaking, parameter ρ "captures" the cooperation between the source node and the relay node, and makes it possible to solve the maximization problem using convex programming. We give an algorithm to computer the upper bound. We also present lower bounds on the MIMO relay channel capacity and provide algorithms to compute the bounds. Next, we consider the Rayleigh fading case. We give an upper bound and a lower bound on the ergodic capacity. It is somewhat surprising that the upper bound can meet the lower bound under certain conditions (not necessarily degraded), indicating that the ergodic capacity can be characterized exactly. In particular, we identify sufficient conditions to achieve the ergodic capacity when all nodes have the

same number of antennas; and our intuition for this finding is that the source node and the relay node can function as a "virtual" transmit antenna array when the relay node is located close to the source node, thus making it possible to achieve the capacity. Then we study the sufficient conditions under which the ergodic capacity can be characterized exactly. We examine two interesting cases, namely the high SNR regime and the scalar relay channel case, and present the SNR conditions for achieving the capacity. The capacity results we obtain indicate independent coding schemes at the source node and the relay node; and our intuition is that the channel uncertainty (randomness) at transmitters results in such independence of the codebooks. We finally discuss a potential application of the MIMO relay channel in cooperative communications in ad hoc networks by using the capacity results.

In Chapter 3, we study the achievable rates and scaling laws of large-scale wireless relay networks in the wideband regime. Assuming that relay nodes have no *a priori* knowledge of channel state information (CSI) for both the backward channels and the forward channels, we examine the achievable rates in the joint asymptotic regime of the number of relay nodes n, the channel coherence interval L, and the SNR per link ρ, or the system bandwidth W. We first study narrowband relay networks in the low SNR regime. We study a relaying scheme, namely amplify-and-forward (AF) with network training, in which the source node and the destination node broadcast training symbols and each relay node carries out channel estimation and then uses AF relaying to relay information. We provide an equivalent source-to-destination channel model, and characterize the corresponding achievable rate. Our findings show that when ρL, proportional to the transmission energy in each fading block, is bounded below, the achievable rate has the same scaling order as in coherent

relaying, thus enabling us to characterize the scaling law of the relay networks in the low SNR regime.

We then generalize the study to power-constrained wideband relay networks, where frequency-selective fading is taken into account. In particular, we examine the scaling behavior of the achievable rates corresponding to two power allocation policies applied at relay nodes, namely, a simple equal power allocation policy and the optimal power allocation policy. We identify the conditions under which the scaling law of the wideband relay networks can be achieved by both power allocation policies. Somewhat surprising, our findings indicate that these two power allocation policies result in achievable rates of the same scaling order, and the scaling law can be characterized under the condition that L/W, proportional to the energy per fading block per subband, is bounded below and that W is sub-linear in n.

In Chapter 4, we study the information gathering in sensor networks, where the sensor nodes observe the physical phenomena, generate data and then transport the data through the network. We focus on the mutual information between the physical sources and the received signal at the sink. First, we consider an event-driven sensory network model, where there is one critical sensor node that is closest to the event, whose data needs to be relayed by the rest of $n-1$ relay nodes. A two-hop transmission pattern is applied for data communications from the sensor nodes to the sink. Simply put, the critical sensor broadcasts its data to the rest of the network in the first hop, and then the other nodes apply an amplify-and-forward with linear combining (AFLC) scheme to process jointly the received signal from the critical node and their own sensed data, and then retransmit in the second hop. A key issue here is the power allocation between the received signals and the

sensed signals at the nodes, so as to maximize the mutual information between the input signal (event) and the received signal at the sink. We find the corresponding optimal power allocation policy at the relay nodes boils down to maximizing the output SNR at each relay node.

We next study Gaussian sensor networks over a 2-D random field. We propose to use estimate-and-forward (EF) at each relay node for data fusion and data transport, where the relay nodes first estimate the physical source and then transmit the estimates to the sink. We investigate two network models, i.e., a sensory relay network over a 2-D filed and a cluster sensor network. For the sensory relay network model, when there is one physical source involved, the EF scheme is equivalent to the AFLC scheme. In the clustered sensor network model, we find that allowing some of the cluster heads to transmit may improve the mutual information. We also investigate the downlink transmission when the sink needs to send messages to the nodes.

5.2. Future Work

In Chapter 4, we have studied the mutual information and the corresponding power allocation policies for an event-driven sensor network model and sensor networks over a 2-D random field. Consider a more general n-node sensor network, as illustrated in Fig. 5.1, where each sensor node has data to transmit, denoted as x_i, for $i = 0, 1, ..., n-1$. A two-hop transmission pattern is assumed for communication. Let y stand for the received signal at the sink which is responsible for collecting the data from each node. Correspondingly, we are interested in characterizing the mutual information between $\{x_0, x_1, ...x_{n-1}\}$ and y.

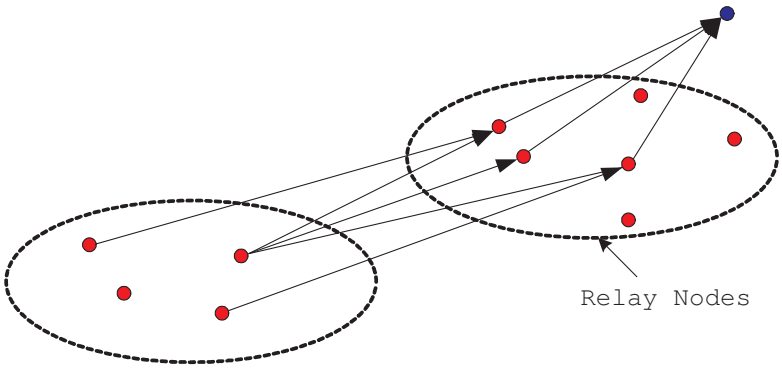

Figure 5.1. A general model of sensor networks in a 2-D random field

Assume that there are n_1 nodes that cannot reach the sink directly, and n_2 relay nodes can communicate with the sink ($n_1 + n_2 = n$). Let C_1 denote this set of n_1 nodes, and let C_2 represent this set of n_2 nodes. Let α_{ij} stand for the channel gain between jth ($j \in C_1$) node and ith ($i \in C_2$) relay node, and denote β_i, for $i \in C_2$ as the channel gain from the ith relay node to the sink. The received signal at the i-th relay node is given

$$y_i = \sum_{j \in C_1} \alpha_{ij} x_j + z_i, \quad i = 1, ..., n-1, \tag{5.1}$$

where z_i is complex Gaussian with zero mean and variance N. Applying the AFLC scheme to combine the two signals, we have that the transmitted signal at the i-th sensor is given by

$$x_i' = A y_i + \bar{A}_i x_i \tag{5.2}$$

$$= A_i \sum_{j \in C_1} \alpha_{ij} x_j + \bar{A}_i x_i + A_i z_i \tag{5.3}$$

The received signal at the sink is given by

$$y = \sum_{i=1}^{n-1} \beta_i x_i' + z \tag{5.4}$$

$$= \sum_{i \in C_2} \sum_{j \in C_1} \beta_i A_i \alpha_{ij} x_j + \sum_{i \in C_2} \beta_i \bar{A}_i x_i + \sum_{i \in C_2} \beta_i A_i z_i + z, \tag{5.5}$$

where w is complex Gaussian with zero mean and variance N.

Recall that maximizing the mutual information between the $\{x_0, x_1, ..., x_{n-1}\}$ and the received signal y leads to maximizing the overall SNR at the sink, which is defined as

$$SNR_{overall} \triangleq \frac{\mathbb{E}_{x_0,x_1,...x_{n-1}}\left[\left|\sum_{i\in C_2}\sum_{j\in C_1}\beta_i A_i \alpha_{ij} x_j + \sum_{i\in C_2}\beta_i \bar{A}_i x_i\right|^2\right]}{\sum_{i\in C_2}|\beta_i A_i|^2 N + N} \tag{5.6}$$

Assume that $\{\alpha_{ij}\}, j \in C_1,\ i \in C_2$ and $\{\beta_i\},\ i \in C_2$ are known throughout the network, then we may be able to characterize the maximum overall SNR. First, we need to derive the variance of the terms in (5.5). For convenience, define

$$U_1 \triangleq [u_1^{(1)}, u_2^{(1)}, ...u_{n_1}^{(1)}]', \qquad u_j^{(1)} \triangleq \sum_{i\in C_2}\beta_i A_i \alpha_{ij} \quad , j = 1, 2, ..., n_1 \tag{5.7}$$

$$U_2 \triangleq [u_1^{(2)}, u_2^{(2)}, ...u_{n_1}^{(2)}]', \qquad u_j^{(2)} \triangleq \beta_j \bar{A}_j \quad , j = 1, 2, ...n_2 \tag{5.8}$$

$$U_3 \triangleq [u_1^{(3)}, u_2^{(3)}, ...u_{n_1}^{(3)}]', \qquad u_j^{(3)} \triangleq \beta_j A_j \quad , j = 1, 2, ...n_2 \tag{5.9}$$

and

$$\Sigma_1 \triangleq \mathbb{E}\left[X_1 X_1^\dagger\right], \quad \Sigma_2 \triangleq \mathbb{E}\left[X_2 X_2^\dagger\right], \quad \Sigma_{12} \triangleq \mathbb{E}\left[X_1 X_2^\dagger\right], \tag{5.10}$$

where "$\dagger$" denotes the operator of transpose and conjugate. X_1 denote the vector with each element as the observation at the node in C_1 and X_2 denote the vector with each element

as the observation at the node in C_2. Then we have that

$$
\begin{aligned}
v_1 &\triangleq \mathbb{E}\left[\left|\sum_{i\in C_2}\sum_{j\in C_1}\beta_i A_i \alpha_{ij} x_j\right|^2\right] && (5.11)\\
&= U_1\Sigma_1 U_1^{\dagger} && (5.12)\\
v_2 &\triangleq \mathbb{E}\left[\left|\sum_{i\in C_2}\beta_i \bar{A}_i x_i\right|^2\right] && (5.13)\\
&= U_2\Sigma_2 U_2^{\dagger} && (5.14)\\
v_3 &\triangleq \mathbb{E}\left[\left(\sum_{i\in C_2}\sum_{j\in C_1}\beta_i A_i \alpha_{ij} x_j\right)\left(\sum_{i\in C_2}\beta_i \bar{A}_i x_i\right)^*\right] && (5.15)\\
&\quad +\mathbb{E}\left[\left(\sum_{i\in C_2}\sum_{j\in C_1}\beta_i A_i \alpha_{ij} x_j\right)^*\left(\sum_{i\in C_2}\beta_i \bar{A}_i x_i\right)\right] && (5.16)\\
&= U_1\Sigma_{12} U_2^{\dagger} + U_2\Sigma_{12}^{\dagger} U_1^{\dagger} && (5.17)
\end{aligned}
$$

In a nutshell, the problem boils down to

$$
\begin{aligned}
&\max_{\{A_i\},\{\bar{A}_i\}} \frac{v_1+v_2+v_3}{\sum_{i\in C_2}|\beta_i A_i|^2 N + N}\\
\text{subject to}\quad &\mathbb{E}\left[\left|A_i\sum_{j\in C_1}\alpha_{ij}x_j + \bar{A}_i x_i\right|^2\right] + |A_i|^2 N \le p \qquad (5.18)
\end{aligned}
$$

We have taken some initial steps for solving the above optimization problem. First, we note that the optimal power allocation depends on the global information of the channel gains and the correlation among the source signals. Second, the optimization is non-convex. It may be highly involved to obtain the closed form solution for the optimization problem. Alternatively, it may be of interest to study the scaling behavior of $SNR_{overall}$, along the line in Chapter 3. In particular, the above problem can be studied for the following three cases: 1) $n_1 = o(n_2)$, 2) $n_1 = \Theta(n_2)$ and 3) $n_2 = o(n_1)$.

In this dissertation, we have studied the capacity bounds and achievable rates by using three relaying schemes: decode-and-forward (DF) relaying in Chapter 2, amplify-

and-forward (AF) relaying in Chapter 3, and estimate-and-forward relaying in Chapter 4. By reaping the node diversity gain, the above schemes can increase the data rate of the networks. Another candidate of interest can be "compress-and-forward"(CF) relaying scheme. where the relay nodes quantize the signal received from the source nodes and then encode the quantized signals which are forwarded to the destination node. A key step in CF relaying is to optimize quantization and retransmission of the received signals at relay nodes, which depends on a certain distortion requirement for information fidelity.

REFERENCES

[1] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Information Processing in Sensor Networks (IPSN '06)*, Nashville, TN, April 2006.

[2] H. Bölcskei, R. U. Nabar, Özgü r Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay network," *IEEE Transactions on Wireless Communications*, 2006, to appear.

[3] S. Boyd and L. Vandenberghe, *Convex Opitmization*. Cambridge University Press, 2004.

[4] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, pp. 1691–1706, July 2003.

[5] R. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Transactions on Information Theory*, vol. 39, pp. 773–785, May 1993.

[6] T. M. Cover and A. A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, pp. 572–584, Sept. 1979.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons Inc., 1991.

[8] A. F. Dana and B. Hassibi, "On the power-efficiency of sensory and ad-hoc wireless networks," *IEEE Transactions on Information Theory*, June 2003, submitted.

[9] S. S. Dongning Guo and S. Verdú, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 1261–1281, April 2005.

[10] O. Dousse, M. Franceschetti, and P. Thiran, "Information theoretic bounds on the throughput scaling of wireless relay networks," in *Proceedings of Infocom*, Miami, March 2005.

[11] R. G. Gallager, "Finding parity in a simple broadcast network," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 176–180, March 1988.

[12] E. Gammal, "Lecture notes 7: Relay channel," *EEE478 Stanford University*, fall, 2002.

[13] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: Coding and antenna-clustering capacity," in *ISIT 2002*, Lausanne, Switzerland, June 2002.

[14] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proceedings of Infocom*, June 2002, pp. 1577–1586.

[15] ——, "On the capacity of large gaussian relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 765–779, March 2005.

[16] ——, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 745–754, April 2005.

[17] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*. San Diego San Francisco New York: Academic Press, 2000.

[18] M. Grossglauser and D. Tse, "Mobility increases the capacity of adhoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, Aug. 2002.

[19] A. Gupta and D. Nagar, *Matrix Variate Distributions*. Champman & Hall Inc., 2000.

[20] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 288–404, March 2000.

[21] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channel," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2020–2040, June 2005.

[22] R. A. Horn and G. R. Johnson, *Matrix Analysis*. Cambridge, United Kingdom: Cambridge University Press, 1999.

[23] S. Jayaweera and H. Poor, "Capacity of mutliple-antenna systems with both receiver and transmitter channel state information," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2697–2708, 2003.

[24] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proceedings of Allerton Conference on Communication, Control, and Computing (Allerton 2003)*, Oct. 2003, pp. 1074–1083.

[25] ——, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

[26] G. Kramer, P. Gupta, and M. Gastpar, "Information-theoretic multi-hopping for relay networks," in *2004 Int. Zurich Seminar (Zurich Switzerland)*, February 2004.

[27] J. Laneman, D. Tse, and G. W. Wornell, "An efficient protocol for realizing cooperative diversity in wireless networks," Washington, Jun 2001.

[28] ——, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, pp. 2415–2425, Oct. 2003.

[29] I. Maric and R. D. Yates, "Power and bandwidth allocation for cooperative strategies in gaussian relay networks," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, Nov. 2004, (invited).

[30] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multi-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, pp. 139–157, January 1999.

[31] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[32] R. U. Nabar, H. Bölcskei, and F.W.Kneubühler, "Fading relay channels: Performance limits and space-time signal design," *IEEE Journal on Selected Areas in Communications*, June 2004, to appear.

[33] R. U. Nabar, Özgü r Oyman, H. Bölcskei, and A. J.Paulraj, "Capacity scaling laws in MIMO wireless networks," in *Proceedings of Allerton Conference on Communication, Control, and Computing (Allerton 2003)*, Oct. 2003, pp. 378–389.

[34] Özgü r Oyman, R. U. Nabar, H. Bölcskei, and A. J.Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Transactions on Signal Porcessing*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.

[35] Özgü r Oyman and A. J. Paulraj, "Power-bandwidth tradeoff in linear mimo interference relay network," *IEEE Transactions on Wireless Communications*, Oct. 2005, submitted.

[36] ——, "Power-bandwidth tradeoff in linear multi-antenna interference relay networks," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sept. 2005.

[37] W. Rhee and J. M. Cioffi, "Ergodic capacity of multi-antenna Gaussian multiple access channels," in *35th Asilomar Conference on Signals, Systems and Computers*, Nov. 2001.

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

[39] ——, "User cooperation diversity–part II: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[40] I. E. Telatar and D. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1384–1400, July 2000.

[41] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transction on Telecommunications*, vol. 10, pp. 585–595, Nov. 1999.

[42] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[43] V. V. Veeravalli and Y. Liang, "Gaussian orthogonal relay channels: Optimal resource allocation and capacity," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 3284 – 3289, Sept. 2005.

[44] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[45] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2658–2668, Oct. 2003.

[46] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[47] B. Wang and J. Zhang, "MIMO relay channel and its application for cooperative communication in ad hoc networks," in *Proceedings of Allerton Conference on Communication, Control, and Computing (Allerton 2003)*, Oct. 2003, pp. 1556–1565.

[48] ——, "Achievable rates and scaling laws of wireless relay networks in the low snr regime," in *Proceedings of Allerton Conference on Communication, Control, and Computing (Allerton 2005)*, Sep. 2005.

[49] ——, "Throughput scaling of wideband sensory relay networks: Cooperative relaying, power allocation and achievable rates," in *Proceedings of Infocom 06*, Barcelona, Spain, Apr. 2006.

[50] B. Wang, J. Zhang, and A. Høst-Madsen, "On ergodic capacity of MIMO relay channel," in *Proceedings of Conference on Information Sciences and Systems (CISS 2004)*, March 2004, pp. 603–608.

[51] ——, "On the capacity of MIMO relay channels," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

[52] B. Wang, J. Zhang, and L. Zheng, "Achievable rates and scaling laws of power-constrained wireless relay networks," *IEEE Transactions on Information Theory*, Sep. 2006, to appear.

[53] H. Weingarten, Y. Steinberg, and S. S. (Shitz), "The capacity region of the Gaussian MIMO broadcast channel," in *ISIT2004*, Chicago, June 2004.

[54] A. Wyner, "Shannon-theorectic approach to a Gaussian cellular multiple-access channel," *IEEE Transactions on Information Theory*, vol. 41, pp. 1713–1727, November 1994.

[55] L.-L. Xie and P. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Transactions on Information Theory*, vol. 50, pp. 748–767, May 2004.

[56] L. Yang and G. Giannakis, "Ultra-wideband communications: an idea whose time has come," *IEEE Signal Processing Magazine*, vol. 21, Nov. 2004.

[57] W. Yu and J. M. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1875 – 1892, Sep. 2004.

[58] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple access channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 145–152, January 2004.

[59] J. Zhang and E. K. P. Chong, "Linear MMSE multiuser receivers: MAI conditional weak convergence and network capacity," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 2114–2122, July 2002.

[60] L. Zheng, "Diversity-multiplexing tradeoff: A comprehensive view of multiple antenna systems," Ph.D. dissertation, University of California at Berkeley, Sept. 2002.

[61] L. Zheng, D. Tse, and M. Médard, "Channel coherence in the low SNR regime," in *Proceedings of CISS*, Princeton, March 2004, pp. 520–525.

[62] ——, "Channel coherence in the low SNR regime," *IEEE Transactions on Information Theory*, Aug 2005, submitted.

APPENDIX A

PROOF OF LEMMA 2.1

Since $\mathbf{\Sigma}_{22} \succ \mathbf{0}$, it follows from [22, 7.7.2] that

$$\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{\Sigma}_{11}^{-\frac{1}{2}} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{12}^{\dagger} \mathbf{\Sigma}_{11}^{-\frac{1}{2}}$$

is positive semi-definite. Therefore, we have that

$$\mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger} \preceq \mathbf{I}_{M_1} \ ,$$

where "$\preceq$" is in terms of positive semi-definite ordering [22]. Given $\mathbf{X}_2 = \mathbf{x}_2$, the covariance matrix of $\mathbf{X}_1$,

$$\mathbf{\Sigma}_{\mathbf{X}_1|\mathbf{X}_2=\mathbf{x}_2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21}$$

is also positive semi-definite. It is easy to see that

$$\mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger} = \mathbf{\Sigma}_{11}^{-\frac{1}{2}} \mathbf{\Sigma}_{\mathbf{X}_1|\mathbf{X}_2=\mathbf{x}_2} \mathbf{\Sigma}_{11}^{-\frac{1}{2}} \succeq \mathbf{0} \ .$$

In summary, we have that

$$\mathbf{0} \preceq \mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger} \preceq \mathbf{I}_{M_1} \ .$$

Thus, by the continuity of $\mathbf{A}$ in the vector space, there always exists $\rho \in [0, 1]$ such that

$$\mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger} \preceq (1 - \rho^2)\mathbf{I}_{M_1} \ ,$$

or equivalently,

$$\mathbf{A}\mathbf{A}^{\dagger} \succeq \rho^2 \mathbf{I}_{M_1} \ . \tag{A.1}$$

The equality in (A.1) can be achieved if $M_1 \leq M_2$.

The converse of Lemma 2.1 holds by the continuity. That is to say, for a given ρ, there exists $\mathbf{\Sigma}_{12}$ such that $\mathbf{A}\mathbf{A}^{\dagger} = \rho^2 \mathbf{I}_{M_1}$.

APPENDIX B

ANOTHER LOWER BOUND FOR THE FIXED CHANNEL CASE

Appealing to [6], we can find another lower bound on the capacity for the fixed channel case. By using Block-Markov coding, the following rate is achieved for any given distribution $p(\mathbf{x}_1, \mathbf{x}_2)$:

$$R = \min\left(\mathbf{I}(\mathbf{X}_1; \mathbf{Y}_1|\mathbf{X}_2),\ \mathbf{I}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}) \right) . \tag{B.1}$$

Recall that the optimal distribution is Gaussian [7, 6]; and the transmitted signal $\mathbf{X}_1$ at the source node can be decomposed as

$$\mathbf{X}_1 \quad = \quad \widetilde{\mathbf{X}}_{10} + \mathbf{X}_{11}, \tag{B.2}$$

where for brevity, we define $\mathbf{X}_{11} \triangleq \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{X}_2$. Based on the decomposition above and Lemma 2.1, the power constraints on $\widetilde{\mathbf{X}}_{10}$ and $\mathbf{X}_{11}$ are given by

$$\mathbb{E}(\widetilde{\mathbf{X}}_{10}^{\dagger}\widetilde{\mathbf{X}}_{10}) \leq (1-\rho^2)M_1 \qquad \text{and} \qquad \mathbb{E}(\mathbf{X}_{11}^{\dagger}\mathbf{X}_{11}) \leq \rho^2 M_1. \tag{B.3}$$

We can choose the signal $\widetilde{\mathbf{X}}_{10}$ to maximize the information rate for the source-relay link because $\mathbf{I}(\mathbf{X}_1; \mathbf{Y}_1|\mathbf{X}_2) = \mathbf{I}(\widetilde{\mathbf{X}}_{10}; \mathbf{Y}_1)$. Then the corresponding solution for $\widetilde{\mathbf{X}}_{10}$ can be found by using the water-filling technique [23]. More specifically, let the SVD of $\mathbf{H}_1$ be $\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{V}_1^{\dagger}$. Accordingly, the water-filling solution for $\widetilde{\mathbf{X}}_{10}$ is given by

$$\widetilde{\mathbf{X}}_{10} = \mathbf{V}_1\mathbf{T}_{10}, \tag{B.4}$$

where $\mathbf{T}_{10}$ is a circularly-symmetric complex Gaussian vector satisfying

$$\mathbb{E}(\mathbf{T}_{10}\mathbf{T}_{10}^{\dagger}) = \mathbf{D}_{10} \tag{B.5}$$

with $\mathbf{D}_{10}$ being a diagonal matrix (see [23]).

Now consider the multi-access link from the source node and the relay node to the destination node, i.e., $(\mathbf{X}_1, \mathbf{X}_2) \rightarrow \mathbf{Y}$. We can view this as a MIMO channel with $M_1 +$

M_2 transmit antennas and N receive antennas. Define $m = \min(M_1 + M_2,\ N)$. Since $\mathbf{I}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}) = \mathbf{I}(\mathbf{X}_{11}, \mathbf{X}_2; \mathbf{Y})$, we can choose the joint distribution $p(\mathbf{X}_{11}, \mathbf{X}_2)$ to maximize the sum information rate for the multi-access link. To this end, the received signal at the destination node can be written as

$$\mathbf{Y} = \begin{bmatrix} \sqrt{\eta_2}\mathbf{H}_2 & \sqrt{\eta_3}\mathbf{H}_3 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{11} \\ \mathbf{X}_2 \end{bmatrix} + \sqrt{\eta_2}\mathbf{H}_2\mathbf{V}_1\mathbf{T}_{10} + \mathbf{Z}. \tag{B.6}$$

Let $\mathbf{L}^{-1}$ be a whitening matrix for $\sqrt{\eta_2}\mathbf{H}_2\mathbf{V}_1\mathbf{T}_{10} + \mathbf{Z}_1$, and define $\mathbf{H}_w \triangleq \mathbf{L}^{-1}\begin{bmatrix} \sqrt{\eta_2}\mathbf{H}_2 & \sqrt{\eta_3}\mathbf{H}_3 \end{bmatrix}$. Let $\mathbf{U}_w\mathbf{\Lambda}_w\mathbf{V}_w^\dagger$ be the SVD of $\mathbf{H}_w$. Rewrite the transmitted signal as

$$\begin{bmatrix} \mathbf{X}_{11} \\ \mathbf{X}_2 \end{bmatrix} = \mathbf{V}_w\mathbf{T}_w, \tag{B.7}$$

where $\mathbf{T}_w$ is a circularly-symmetric complex Gaussian vector satisfying

$$\mathbb{E}(\mathbf{T}_w\mathbf{T}_w^\dagger) = \mathbf{D}_w \tag{B.8}$$

with $\mathbf{D}_w$ being a diagonal matrix (see [23]).

Then the sum information rate for the multi-access link of the MIMO relay channel is given by

$$\mathbf{I}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}) = \log\left[\det\left(\mathbf{I} + \mathbf{\Lambda}_w\mathbf{D}_w\mathbf{\Lambda}_w^\dagger\right)\right] + \log\left[\det\left(\mathbf{L}\mathbf{L}^\dagger\right)\right]. \tag{B.9}$$

This is a water-filling problem with respect to $\mathbf{D}_w$ except that each terminal has its own power constraint. Let $\mathbf{V}_{w_1}$ denote the M_1 first rows of $\mathbf{V}_w$, and $\mathbf{V}_{w_2}$ the remaining rows.

Then the power constraints are as follows:

$$\mathrm{tr}(\mathbf{V}_{w_1}\mathbf{D}_w\mathbf{V}_{w_1}^{\dagger}) = \sum_{k=1}^{m} c_{1k}[\mathbf{D}_w]_{kk} \leq \rho^2 M_1 \quad \text{(B.10)}$$

$$\mathrm{tr}(\mathbf{V}_{w_2}\mathbf{D}_w\mathbf{V}_{w_2}^{\dagger}) = \sum_{k=1}^{m} c_{2k}[\mathbf{D}_w]_{kk} \leq M_2, \quad \text{(B.11)}$$

where c_{1k} is the non-zero eigenvalue of $\mathbf{V}_{w_1}\mathbf{V}_{w_1}^{\dagger}$, and c_{2k} is the non-zero eigenvalue of $\mathbf{V}_{w_2}\mathbf{V}_{w_2}^{\dagger}$; and $[\mathbf{D}_w]_{kk}$ denotes the k^{th} element along the diagonal of $\mathbf{D}_w$.

The problem can be solved by using the Lagrange dual function:

$$\begin{aligned}\mathcal{L}(\lambda_1, \lambda_2) = & \sum_{k=1}^{m} \log(1 + [\mathbf{\Lambda}_w^{\dagger}\mathbf{\Lambda}_w]_{kk}[\mathbf{D}_w]_{kk}) \\ & + \lambda_1 \left(\sum_{k=1}^{m} c_{1k}[\mathbf{D}_w]_{kk} - \rho^2 M_1 \right) + \lambda_2 \left(\sum_{k=1}^{m} c_{2k}[\mathbf{D}_w]_{kk} - M_2 \right), \quad \text{(B.12)}\end{aligned}$$

where λ_1 and λ_2 are Lagrange multipliers.

Alternatively, we can find a suboptimal solution by first finding the standard water-filling solution subject to a total power constraint, following by scaling this to satisfy the individual power constraints. That is to say, rewrite

$$\begin{bmatrix} \mathbf{X}_{11} \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 \mathbf{V}_{w_1} \\ \alpha_2 \mathbf{V}_{w_2} \end{bmatrix} \mathbf{T}_w, \quad \text{(B.13)}$$

where α_1 and α_2 satisfy

$$\alpha_1 \mathrm{tr}[\mathbf{V}_{w_1}\mathbf{D}_w\mathbf{V}_{w_1}^{\dagger}] = \rho^2 M_1 \quad \text{(B.14)}$$

$$\alpha_2 \mathrm{tr}[\mathbf{V}_{w_2}\mathbf{D}_w\mathbf{V}_{w_2}^{\dagger}] = M_2. \quad \text{(B.15)}$$

Then, we can maximize the achievable rate with respect to parameter ρ after plugging D_{01} and D_w into (B.1).

APPENDIX C

PROOF OF (3.24)

The derivation is a modification of [31, (46)], where the author assumed the noise is Gaussian. In what follows, we show that letting the noise be non-Gaussian and signal-dependent can yield the same result.

For convenience, define

$$F \triangleq H + \sum_i \sqrt{\rho} \left(\frac{|\hat{\alpha}_i|^2 \tilde{\beta}_i \hat{\beta}_i^* + |\hat{\beta}_i|^2 \tilde{\alpha}_i \hat{\alpha}_i^* + \tilde{\alpha}_i \hat{\alpha}_i^* \tilde{\beta}_i \hat{\beta}_i^*}{|\hat{\alpha}_i||\hat{\beta}_i|} \right). \tag{C.1}$$

Then, we have the equivalent source-to-destination signal model as

$$Y(t) = FX(t - LT_s) + \Omega_2(t) + Z(t). \tag{C.2}$$

Observe that in the above equivalent signal model ,

- the channel gain F and the transmitted signal $X(t - LT_s)$ are independent, and $X(t - LT_s)$ and the noise term $\Omega_2(t) + Z(t)$ are uncorrelated;
- $\mathbb{E}[F] = \mathbb{E}[H]$;
- the noise term $\Omega_2(t) + Z(t)$ is non-Gaussian and signal-dependent, and is zero-mean and of bounded power.

Using standard vector representation for band-limited signals [7], the continuous-time channel model in (C.2) can be mapped onto a set of discrete-time subchannels. That is, the signal model in the jth dimension is given by

$$Y^{(j)} = FX^{(j)} + \Omega_2^{(j)} + Z^{(j)} \tag{C.3}$$

Next, we examine the achievable rate corresponding to the channel model in (C.3). Note that the mutual information between $X^{(j)}$ and $Y^{(j)}$ is

$$I\left(X^{(j)}; Y^{(j)}\right) = h\left(X^{(j)}\right) - h\left(X^{(j)}|Y^{(j)}\right). \tag{C.4}$$

Along the lines in [31], we fix the distribution of $X^{(j)}$ as complex Gaussian with zero mean, even though this choice may not be optimal for maximizing the mutual information. Next, we find an upper bound on $h\left(X^{(j)}|Y^{(j)}\right)$, which holds for all possible distributions for $X^{(j)}$. We have that

$$\begin{aligned} h\left(X^{(j)}|Y^{(j)}\right) &= h\left(X^{(j)} - bY^{(j)}|Y^{(j)}\right) \\ &\overset{\text{(i)}}{\leq} h\left(X^{(j)} - bY^{(j)}\right) \\ &\overset{\text{(ii)}}{\leq} \log\left(2\pi e \cdot \text{var}\left(X^{(j)} - bY^{(j)}\right)\right), \end{aligned} \tag{C.5}$$

where (i) follows the fact that conditioning always decreases entropy; (ii) follows that the entropy of a random variable of bounded variance is upper-bounded by the entropy of a Gaussian random variable with the same variance; and the above bound holds for any b. Let b be the coefficient, for which $bY^{(j)}$ is the linear minimum mean-square error(LMMSE) estimate of $X^{(j)}$ in terms of $Y^{(j)}$. That is,

$$b = \frac{\mathbb{E}\left[X^{(j)}Y^{(j)}\right]}{\text{var}\left(Y^{(j)}\right)}. \tag{C.6}$$

We note that F and $X(t-LT_s)$ are independent, and $X(t-LT_s)$ and $\Omega_2(t)+Z(t)$ are uncorrelated and zero-mean. After some algebra, we have that

$$\begin{aligned} \max_{p(X^{(j)})} I\left(X^{(j)};Y^{(j)}\right) &\geq \log\left(2\pi e \cdot \text{var}\left(X^{(j)}\right)\right) - \log\left(2\pi e \cdot \text{var}\left(X^{(j)} - bY^{(j)}\right)\right) \\ &= \log\left(1 + \frac{\mathbb{E}\left[\left|\mathbb{E}[F]X^{(j)}\right|^2\right]}{\mathbb{E}[|\Omega_1^{(j)}|^2] + \mathbb{E}[|\Omega_2^{(j)}|^2] + \mathbb{E}[|\Omega_3^{(j)}|^2] + \mathbb{E}[|Z^{(j)}|^2]}\right) \end{aligned} \tag{C.7}$$

Combing all the dimensions and taking into account the fact that the relay nodes work in half-duplex mode, we have that the lower bound on the capacity of (3.23) is given by

$$C \geq B\frac{L-1}{L}\log\left(1 + \frac{\mathbb{E}\left[\left|\mathbb{E}[H]X(t-LT_s)\right|^2\right]}{\mathbb{E}[|\Omega_1(t)|^2] + \mathbb{E}[|\Omega_2(t)|^2] + \mathbb{E}[|\Omega_3(t)|^2] + WN_0}\right), \tag{C.8}$$

where $\frac{L-1}{L}$ denotes the portion of time for information transmission.

APPENDIX D

PROOF OF (3.82), (3.83), (3.84) AND (3.85)

Recall that in (3.81),

$$H_{k,m} \triangleq \sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| , \tag{D.1}$$

$$\Omega_{1k,m}(t) \triangleq \sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \left(\frac{|\hat{\alpha}_{i,k}|^2\tilde{\beta}_{i,k}\hat{\beta}^*_{i,k} + |\hat{\beta}_{i,k}|^2\tilde{\alpha}_{i,k}\hat{\alpha}^*_{i,k} + \tilde{\alpha}_{i,k}\hat{\alpha}^*_{i,k}\tilde{\beta}_{i,k}\hat{\beta}^*_{i,k}}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|} \right) X_k(t - LT_s) , \tag{D.2}$$

$$\Omega_{2k,m}(t) \triangleq \sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \left(\frac{\hat{\alpha}^*_{i,k}|\hat{\beta}_{i,k}|^2 + \hat{\alpha}^*_{i,k}\tilde{\beta}_{i,k}\hat{\beta}^*_{i,k}}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|} \right) Z_{i,k}(t - LT_s) , \tag{D.3}$$

$$\Omega_{3k,m}(t) \triangleq (H_{k,m} - \mathbb{E}[H_{k,m}])X_k(t - LT_s) , \tag{D.4}$$

where $Z_{i,k}(t - LT_s)$ is the noise at the ith relay node within the kth subband.

first, note that the power of the signal used for decoding is given by

$$\mathbb{E}\left[|\mathbb{E}[H_{k,m}]X_k(t - LT_s)|^2\right] = |\mathbb{E}[H_{k,m}]|^2 \, \mathbb{E}[|X_k(t - LT_s)|^2] . \tag{D.5}$$

Observe that

$$\begin{aligned}
|\mathbb{E}[H_{k,m}]|^2 &= \left| \mathbb{E}\left[\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| \right] \right|^2 \\
&= \left| \sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \mathbb{E}\left[\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right|\right] \right|^2 \\
&\overset{(a)}{=} \left| \sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \mathbb{E}\left[|\hat{\alpha}_{i,k}|\right] \mathbb{E}\left[\left|\hat{\beta}_{i,k}\right|\right] \right|^2 \\
&\overset{(b)}{=} \frac{1}{BN_0} \left(\sum_i \sqrt{p_{i,k,m}} \right)^2 \left(\mathbb{E}\left[|\hat{\alpha}_{i,k}|\right]\right)^2 \left(\mathbb{E}\left[\left|\hat{\beta}_{i,k}\right|\right]\right)^2 \\
&\overset{(c)}{=} \frac{1}{BN_0} \left(\sum_i \sqrt{p_{i,k,m}} \right)^2 \left(\frac{\pi\sigma_k^2}{4} \frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL + WN_0} \right)^2
\end{aligned} \tag{D.6}$$

where

(a) follows the fact that $\hat{\alpha}_{i,k}$ and $\hat{\beta}_{i,k}$ are independent;

(b) follows that within a given subband, channel coefficients are i.i.d random variables;

(c) comes from the fact that $\mathbb{E}\left[|\hat{\alpha}_{i,k}|\right] = \mathbb{E}\left[\left|\hat{\beta}_{i,k}\right|\right]$, and that

$$\begin{aligned}\mathbb{E}\left[|\hat{\alpha}_{i,k}|\right] &= \sqrt{\frac{\pi}{4}\mathbb{E}[|\hat{\alpha}_{i,k}|^2]} \\ &= \sqrt{\frac{\pi\sigma_k^2}{4}\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL + WN_0}} \,. \end{aligned} \tag{D.7}$$

Hence, we have that

$$\mathbb{E}\left[|\mathbb{E}[H_{k,m}]X_k(t-LT_s)|^2\right] = \frac{\pi^2}{16}\frac{\sigma_k^4}{BN_0}\left(\sum_{i=1}^{n}\sqrt{p_{i,k,m}}\right)^2\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL + WN_0}\right)^2\frac{(1-\eta_k)L}{L-1}\frac{P}{K} \,. \tag{D.8}$$

Next, we characterize the power of $\Omega_{1k,m}(t)$.

$$\begin{aligned}\mathbb{E}\left[\left|\Omega_{1k,m}(t)\right|^2\right] &= \mathbb{E}\left[\left|\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}}\cdot\frac{|\hat{\alpha}_{i,k}|^2\tilde{\beta}_{i,k}\hat{\beta}_{i,k}^* + |\hat{\beta}_{i,k}|^2\tilde{\alpha}_{i,k}\hat{\alpha}_{i,k}^* + \tilde{\alpha}_{i,k}\hat{\alpha}_{i,k}^*\tilde{\beta}_{i,k}\hat{\beta}_{i,k}^*}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|}\right|^2\right] \\ &\quad\cdot\mathbb{E}[|X_k(t-LT_s)|^2] \,. \end{aligned} \tag{D.9}$$

We have that

$$
\begin{aligned}
& \mathbb{E}\left[\left|\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \cdot \frac{|\hat{\alpha}_{i,k}|^2\tilde{\beta}_{i,k}\hat{\beta}^*_{i,k} + |\hat{\beta}_{i,k}|^2\tilde{\alpha}_{i,k}\hat{\alpha}^*_{i,k} + \tilde{\alpha}_{i,k}\hat{\alpha}^*_{i,k}\tilde{\beta}_{i,k}\hat{\beta}^*_{i,k}}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|}\right|^2\right] \\
= \;& \mathbb{E}\left[\left|\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \left(\underbrace{\frac{|\hat{\alpha}_{i,k}|^2\tilde{\beta}_{i,k}\hat{\beta}^*_{i,k}}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|}}_{U_{1i,k}} + \underbrace{\frac{|\hat{\beta}_{i,k}|^2\tilde{\alpha}_{i,k}\hat{\alpha}^*_{i,k}}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|}}_{U_{2i,k}} + \underbrace{\frac{\tilde{\alpha}_{i,k}\hat{\alpha}^*_{i,k}\tilde{\beta}_{i,k}\hat{\beta}^*_{i,k}}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|}}_{U_{3i,k}}\right)\right|^2\right] \\
= \;& \frac{1}{BN_0}\sum_i p_{i,k,m}\,\mathbb{E}\left[\left|U_{1i,k}+U_{2i,k}+U_{3i,k}\right|^2\right] \\
& + \frac{1}{BN_0}\sum_i\sum_{j,j\neq i}\sqrt{p_{i,k,m}p_{j,k,m}}\,\mathbb{E}\left[\left(U_{1i,k}+U_{2i,k}+U_{3i,k}\right)\left(U^*_{1j,k}+U^*_{2j,k}+U^*_{3j,k}\right)\right] \\
\overset{(a)}{=} \;& \frac{1}{BN_0}\left(\sum_i p_{i,k,m}\left(\mathbb{E}[|U_{1i,k}|^2]+\mathbb{E}[|U_{2i,k}|^2]+\mathbb{E}[|U_{3i,k}|^2]\right)\right)+0 \\
\overset{(b)}{=} \;& \frac{1}{BN_0}\left(\sum_i p_{i,k,m}\right)\left(\frac{\sigma_k^2WW_0}{\eta_k\sigma_k^2PL+WN_0}\cdot\frac{\eta_k\sigma_k^4PL}{\eta_k\sigma_k^2PL+WN_0}\right. \\
& \qquad \left. + \frac{\sigma_k^2WN_0}{\eta_k\sigma_k^2PL+WN_0}\cdot\frac{\eta_k\sigma_k^4PL}{\eta_k\sigma_k^2PL+WN_0} + \left(\frac{\sigma_k^2WN_0}{\eta_k\sigma_k^2PL+WN_0}\right)^2\right) \\
= \;& \frac{\sigma_k^4}{BN_0}\left(\sum_i p_{i,k,m}\right)\left(1-\left(\frac{\eta_k\sigma_k^2PL}{\eta_k\sigma_k^2PL+WN_0}\right)^2\right)\,, \qquad \text{(D.10)}
\end{aligned}
$$

where

(a) follows the facts that

1. $\tilde{\alpha}_{i,k}$ and $\tilde{\beta}_{i,k}$ are zero-mean;
2. $U_{li,k}$ and $U_{l'i,k}$ ($l \neq l';\ l,l' = 1,2,3$) are orthogonal;
3. $U_{li,k}$ and $U_{l'j,k}$ ($l,l' = 1,2,3$ and $i \neq j$) are orthogonal;

(b) follows that

1. $\mathbb{E}[|U_{1i,k}|^2]$, $\mathbb{E}[|U_{1i,k}|^2]$ and $\mathbb{E}[|U_{1i,k}|^2]$ are independent of i, and thus the summation is only taken with respect to $p_{i,k,m}$;

2.

$$\begin{aligned}\mathbb{E}[|U_{1i,k}|^2] &= \mathbb{E}[|U_{2i,k}|^2] \\ &= \frac{\sigma_k^2 W W_0}{\eta_k \sigma_k^2 PL + WN_0} \cdot \frac{\eta_k \sigma_k^4 PL}{\eta_k \sigma_k^2 PL + WN_0} \qquad \text{(D.11)} \\ \mathbb{E}[|U_{3i,k}|^2] &= \left(\frac{\sigma_k^2 W W_0}{\eta_k \sigma_k^2 PL + WN_0}\right)^2 . \qquad \text{(D.12)}\end{aligned}$$

Therefore, the power of $\Omega_{1k,m}(t)$ is given by

$$\mathbb{E}[|\Omega_{1k,m}(t)|^2] = \frac{\sigma_k^4}{BN_0}\left(\sum_{i=1}^{n} p_{i,k,m}\right)\left(1 - \left(\frac{\eta_k \sigma_k^2 PL}{\eta_k \sigma_k^2 PL + WN_0}\right)^2\right)\frac{(1-\eta_k)L}{L-1}\frac{P}{K} . \quad \text{(D.13)}$$

The power of the aggregated noise $\Omega_{2k,m}(t)$ is given by

$$\begin{aligned}\mathbb{E}[|\Omega_{2k,m}(t)|^2] &= \mathbb{E}\left[\left|\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}}\left(\frac{\hat{\alpha}_{i,k}^*|\hat{\beta}_{i,k}|^2 + \hat{\alpha}_{i,k}^*\tilde{\beta}_{i,k}\hat{\beta}_{i,k}^*}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|}\right) Z_{i,k}(t - LT_s)\right|^2\right] \\ &= \mathbb{E}\left[\left|\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}}\left(\frac{\hat{\alpha}_{i,k}^*|\hat{\beta}_{i,k}|^2 + \hat{\alpha}_{i,k}^*\tilde{\beta}_{i,k}\hat{\beta}_{i,k}^*}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|}\right)\right|^2\right] BN_0 . \qquad \text{(D.14)}\end{aligned}$$

Along the same line in the proof for (D.10), we have that

$$\begin{aligned}&\mathbb{E}\left[\left|\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}}\left(\frac{\hat{\alpha}_{i,k}^*|\hat{\beta}_{i,k}|^2 + \hat{\alpha}_{i,k}^*\tilde{\beta}_{i,k}\hat{\beta}_{i,k}^*}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|}\right)\right|^2\right] \\ = \;& \mathbb{E}\left[\left|\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}}\left(\underbrace{\frac{\hat{\alpha}_{i,k}^*|\hat{\beta}_{i,k}|^2}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|}}_{V_{1i,k}} + \underbrace{\frac{\hat{\alpha}_{i,k}^*\tilde{\beta}_{i,k}\hat{\beta}_{i,k}^*}{|\hat{\alpha}_{i,k}||\hat{\beta}_{i,k}|}}_{V_{2i,k}}\right)\right|^2\right] \\ = \;& \frac{1}{BN_0}\sum_i p_{i,k,m}\,\mathbb{E}\left[\left|V_{1i,k} + V_{2i,k}\right|^2\right] \\ & + \frac{1}{BN_0}\sum_i \sum_{j,j\neq i} \sqrt{p_{i,k,m}p_{j,k,m}}\,\mathbb{E}\left[\left(V_{1i,k} + V_{2i,k}\right)\left(V_{1j,k}^* + V_{2j,k}^*\right)\right] \\ \overset{(a)}{=} \;& \frac{1}{BN_0}\sum_i p_{i,k,m}\,\left(\mathbb{E}\left[\left|V_{1i,k}\right|^2\right] + \mathbb{E}\left[\left|V_{2i,k}\right|^2\right]\right) + 0 \\ = \;& \frac{1}{BN_0}\sum_i p_{i,k,m}\,\left(\mathbb{E}\left[\left|\hat{\beta}_{i,k}\right|^2\right] + \mathbb{E}\left[\left|\tilde{\beta}_{i,k}\right|^2\right]\right) \\ = \;& \frac{\sigma_k^2}{BN_0}\left(\sum_i p_{i,k,m}\right) , \qquad \text{(D.15)}\end{aligned}$$

where

(a) follows the facts that

1. $V_{li,k}$ and $V_{l'j,k}$ $(i \neq j,\ l, l' = 1, 2)$ are independent;

2. $\mathbb{E}[V_{1i,k}] = 0$ and $\mathbb{E}[V_{2i,k}] = 0$.

As a result, we have that

$$\mathbb{E}[|\Omega_{2k,m}(t)|^2] \quad = \quad \sigma_k^2 \left(\sum_{i=1}^{n} p_{i,k,m} \right) \ . \tag{D.16}$$

$$\begin{aligned} \mathbb{E}\left[|\Omega_{3k,m}(t)|^2\right] \quad &= \quad \mathbb{E}\left[\left(\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| - \mathbb{E}\left[\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| \right] \right)^2 \right] \\ &\quad \cdot \mathbb{E}\left[|X_k(t - LT_s)|^2 \right] \ . \end{aligned} \tag{D.17}$$

Note that

$$\begin{aligned} & \mathbb{E}\left[\left(\sum_i \sqrt{\frac{p_{i,k,m}}{BN_0}} \left(\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| - \mathbb{E}\left[\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| \right] \right) \right)^2 \right] \\ = \quad & \frac{1}{BN_0} \sum_i p_{i,k,m} \mathbb{E}\left[\left(\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| - \mathbb{E}\left[\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| \right] \right)^2 \right] \\ & + \frac{1}{BN_0} \sum_i \sum_{j, j \neq i} \sqrt{p_{i,k,m} p_{j,k,m}}\ \mathbb{E}\left[\left(\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| - \mathbb{E}\left[\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| \right] \right) \left(\left|\hat{\alpha}_{j,k}\hat{\beta}_{j,k}\right| - \mathbb{E}\left[\left|\hat{\alpha}_{j,k}\hat{\beta}_{j,k}\right| \right] \right) \right] \\ \overset{(a)}{=} \quad & \frac{1}{BN_0} \sum_i p_{i,k,m} \left(\mathbb{E}\left[\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right|^2 \right] - \left(\mathbb{E}\left[\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| \right] \right)^2 \right) + 0 \\ \overset{(b)}{=} \quad & \frac{1}{BN_0} \sum_i p_{i,k,m} \left(\left(\mathbb{E}\left[|\hat{\alpha}_{i,k}|^2 \right] \right)^2 - \left(\mathbb{E}\left[|\hat{\alpha}_{i,k}| \right] \right)^4 \right) \\ = \quad & \frac{\sigma_k^4}{BN_0} \left(\sum_i p_{i,k,m} \right) \left(\left(\frac{\eta_k \sigma_k^2 PL}{\eta_k \sigma_k^2 PL + WN_0} \right)^2 - \frac{\pi^2}{16} \left(\frac{\eta_k \sigma_k^2 PL}{\eta_k \sigma_k^2 PL + WN_0} \right)^2 \right) \ , \end{aligned} \tag{D.18}$$

where

(a) follows the fact that

$$\begin{aligned}
& \mathbb{E}\left[\left(\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| - \mathbb{E}\left[\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right|\right]\right)\left(\left|\hat{\alpha}_{j,k}\hat{\beta}_{j,k}\right| - \mathbb{E}\left[\left|\hat{\alpha}_{j,k}\hat{\beta}_{j,k}\right|\right]\right)\right] \\
= \;& \mathbb{E}\left[\left(\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right| - \mathbb{E}\left[\left|\hat{\alpha}_{i,k}\hat{\beta}_{i,k}\right|\right]\right)\right] \cdot \mathbb{E}\left[\left(\left|\hat{\alpha}_{j,k}\hat{\beta}_{j,k}\right| - \mathbb{E}\left[\left|\hat{\alpha}_{j,k}\hat{\beta}_{j,k}\right|\right]\right)\right] \\
= \;& 0; \qquad \text{(D.19)}
\end{aligned}$$

(b) follows that $\hat{\alpha}_{i,k}$ and $\hat{\beta}_{i,k}$ are i.i.d. random variables.

Thus, the power of $\Omega_{3k,m}(t)$ is given by

$$\begin{aligned}
\mathbb{E}[|\Omega_{3k,m}(t)|^2] \;=\; & \frac{\sigma_k^4}{BN_0}\left(\sum_{i=1}^{n} p_{i,k,m}\right)\left(\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL + WN_0}\right)^2 - \frac{\pi^2}{16}\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL + WN_0}\right)^2\right) \\
& \cdot\frac{(1-\eta_k)L}{L-1}\frac{P}{K}\,. \qquad \text{(D.20)}
\end{aligned}$$

When $p_{i,k,m} \equiv p_k$, namely, a common power constraint in the kth subband across fading blocks is applied at each relay node, then it follows that

$$\sum_{i=1}^{n} p_{i,k,m} \;=\; np_k \qquad \text{(D.21)}$$

and

$$\left(\sum_{i=1}^{n}\sqrt{p_{i,k,m}}\right)^2 \;=\; n^2 p_k, \qquad \text{(D.22)}$$

Therefore we have that the power of the desired signal in the kth subband is given by

$$\mathbb{E}\left[\left|\mathbb{E}[H_k]X_k(t-LT_s)\right|^2\right] = n^2\sigma_k^4\frac{\pi^2}{16}\frac{p_k}{BN_0}\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL + WN_0}\right)^2\frac{(1-\eta_k)L}{L-1}\frac{P}{K}\,. \qquad \text{(D.23)}$$

The noise power in the kth subband is given by

$$\mathbb{E}[|\Omega_{1k}(t)|^2] = n\sigma_k^4\frac{p_k}{BN_0}\left(1 - \left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL + WN_0}\right)^2\right)\frac{(1-\eta_k)L}{L-1}\frac{P}{K}\,, \qquad \text{(D.24)}$$

$$\mathbb{E}[|\Omega_{2k}(t)|^2] = n\sigma_k^2 p_k\,, \qquad \text{(D.25)}$$

$$\begin{aligned}
\mathbb{E}[|\Omega_{3k}(t)|^2] = \; & n\sigma_k^4\frac{p_k}{BN_0}\left(\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL + WN_0}\right)^2 - \frac{\pi^2}{16}\left(\frac{\eta_k\sigma_k^2 PL}{\eta_k\sigma_k^2 PL + WN_0}\right)^2\right) \\
& \cdot\frac{(1-\eta_k)L}{L-1}\frac{P}{K}\,. \qquad \text{(D.26)}
\end{aligned}$$

For the narrowband relay network ($K = 1$), we have that

$$\sigma_k^2 = 1 \ , \qquad \eta_k = 1 \tag{D.27}$$

and

$$p_k = P. \tag{D.28}$$

Then it follows that the power of the signal term is given by

$$\begin{aligned} \mathbb{E}\left[|\mathbb{E}[H]X(t - LT_s)|^2\right] &= \rho\left(\mathbb{E}\left[\sum_i |\alpha_i||\beta_i|\right]\right)^2 \frac{(1-\eta)L}{L-1}P \\ &= n^2\frac{\pi^2}{16}\rho\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2 \frac{(1-\eta)L}{L-1}P \ , \end{aligned} \tag{D.29}$$

and the power of the noise terms is given by

$$\mathbb{E}[|\Omega_1(t)|^2] = n\rho\left(1 - \left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2\right)\frac{(1-\eta)L}{L-1}P \ , \tag{D.30}$$

$$\mathbb{E}[|\Omega_2(t)|^2] = \rho\mathbb{E}\left[\left|\sum_i \frac{\hat{\alpha_i}^*}{|\hat{\alpha_i}|}\left(\hat{\beta}_i + \frac{\tilde{\beta}_i\hat{\beta_i}^*}{|\hat{\beta}_i|}\right)\right|^2\right] BN_0 = nP \ ,$$

$$\mathbb{E}[|\Omega_3(t)|^2] = n\rho\left(\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2 - \frac{\pi^2}{16}\left(\frac{\eta\rho L}{\eta\rho L+1}\right)^2\right)\frac{(1-\eta)L}{L-1}P. \tag{D.31}$$

APPENDIX E

PROOF OF (3.97)

Recall that

$$R_{upper}^{WB\ *} = W\log\left(1+\frac{\sigma_{max}^4\frac{\pi^2}{16}n\left(\frac{\eta_{max}\sigma_{max}^2PL}{\eta_{min}\sigma_{min}^2PL+WN_0}\right)^2\frac{(1-\eta_{min})L}{L-1}P}{\sigma_{min}^2WN_0+\frac{(WN_0)^2}{nP}}\right)\ , \tag{E.1}$$

$$R_{Eq}^{WB} = \Theta\left(W\log\left(1+S_{Eq,k}\right)\right)\ . \tag{E.2}$$

In order to prove (3.97), it suffices to show that

$$S_{Eq,k} = \Theta(S_{k,upper})\ , \tag{E.3}$$

where

$$S_{k,upper} \triangleq \frac{\sigma_{max}^4\frac{\pi^2}{16}n\left(\frac{\eta_{max}\sigma_{max}^2PL}{\eta_{min}\sigma_{min}^2PL+WN_0}\right)^2\frac{(1-\eta_{min})L}{L-1}P}{\sigma_{min}^2WN_0+\frac{(WN_0)^2}{nP}}\ . \tag{E.4}$$

Note that $\eta_k^* = \left(-3+\sqrt{8\sigma_k^2PL/WN_0+9}\right)/(2\sigma_k^2PL/WN_0)$ is monotonically decreasing in σ_k^2. That is to say, η_{min} is corresponding to σ_{max}^2 and η_{max} is with respect to σ_{min}^2. Following the same line as in the equal power allocation policy, we have that

- If $e_3 < e_2$, then the optimal energy allocation for training is given by

$$\lim_{n\to\infty}\eta_{min} = 0\ . \tag{E.5}$$

Observing that

$$\frac{\eta_{min}PL}{WN_0} = \Theta\left(n^{\frac{e_2-e_3}{2}}\right) \quad \text{and} \quad \frac{\eta_{max}PL}{WN_0} = \Theta\left(n^{\frac{e_2-e_3}{2}}\right)\ , \tag{E.6}$$

we conclude that

$$S_{k,upper} = \begin{cases} \Theta\left(n^{1-e_3}\right) & \text{if} \quad e_3 < 1 \\ \Theta\left(n^{2-2e_3}\right) & \text{if} \quad e_3 \geq 1 \end{cases}\ . \tag{E.7}$$

- If $e_3 > e_2$, then the optimal energy allocation for training is given by

$$\lim_{n\to\infty} \eta_{min} = \frac{2}{3}. \tag{E.8}$$

Observing that

$$\frac{\eta_{min} PL}{WN_0} = \Theta\left(n^{e_2-e_3}\right) \quad \text{and} \quad \frac{\eta_{max} PL}{WN_0} = \Theta\left(n^{e_2-e_3}\right) \ , \tag{E.9}$$

we conclude that

$$S_{k,upper} = \begin{cases} \Theta\left(n^{1+2e_2-3e_3}\right) & \text{if} \quad e_3 < 1 \\ \Theta\left(n^{2+2e_2-4e_3}\right) & \text{if} \quad e_3 \geq 1 \end{cases} . \tag{E.10}$$

- If $e_3 = e_2$, namely, $L/W = \Theta(1)$, then it follows that

$$\frac{\eta_{min} PL}{WN_0} = \Theta(1) \quad \text{and} \quad \frac{\eta_{max} PL}{WN_0} = \Theta(1) \ . \tag{E.11}$$

Then it follows that

$$S_{k,upper} = \begin{cases} \Theta\left(n^{1-e_3}\right) & \text{if} \quad e_3 < 1 \\ \Theta\left(n^{2-2e_3}\right) & \text{if} \quad e_3 \geq 1 \end{cases} . \tag{E.12}$$

Clearly, the scaling behavior of $S_{k,upper}$ is the same as that of $S_{Eq,k}$. Indeed, $S_{k,upper}$ takes the same form as $S_{Eq,k}$ except the difference between the coefficients (η_k, η_{min}, η_{max}, σ_k^2, σ_{min}^2 and σ_{max}^2). As a result, the scaling order of $S_{Eq,k}$ and $S_{k,upper}$ depends only on the form and the order of L, W and n, thereby concluding the proof of (3.97).