Wireless network information flow: a deterministic approach

Amir Salman Avestimehr

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2008-128
http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-128.html

October 2, 2008

Wireless Network Information Flow: A Deterministic Approach

by

Amir S Avestimehr

KARSH (Sharif University of Technology) 2003
M.Sc. (University of California, Berkeley) 2005

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Philosophy

in

Engineering—Electrical Engineering and Computer Sciences

and the Designated Emphasis

in

Communication, Computation, and Statistics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor David Tse
Professor Kannan Ramchandran
Professor Sourav Chatterjee

Fall 2008

The dissertation of Amir S Avestimehr is approved:

Chair	Date
	Date
	Date

University of California, Berkeley

Fall 2008

Wireless Network Information Flow: A Deterministic Approach

by

Amir S Avestimehr

Abstract

Wireless Network Information Flow: A Deterministic Approach

by

Amir S Avestimehr

Doctor of Philosophy in Engineering—Electrical Engineering and Computer Sciences

and the Designated Emphasis

in

Communication, Computation, and Statistics

University of California, Berkeley

Professor David Tse

In communications, the multiuser Gaussian channel model is commonly used to capture fundamental features of a wireless channel. Over the past couple of decades, study of multiuser Gaussian networks has been an active area of research for many scientists. However, due to the complexity of the Gaussian model, except for the simplest networks such as the one-to-many Gaussian broadcast channel and the many-to-one Gaussian multiple access channel, the capacity region of most Gaussian networks is still unknown. For example, even the capacity of a three node Gaussian relay network, in which a point to point communication is assisted by one helper (relay), has been open for more than 30 years.

To make further progress, we present a linear finite-field *deterministic* channel model which is analytically simpler than the Gaussian model but still captures two key wireless channels: broadcast and superposition. The noiseless nature of this model allows us to focus on the interaction between signals transmitted from different nodes of the network rather than background noise of the links.

Then, we consider a model for a wireless relay network with nodes connected by such

deterministic channels, and present an exact characterization of the end-to-end capacity when there is a single source and a single destination and an arbitrary number of relay nodes. This result is a natural generalization of the celebrated max-flow min-cut theorem for wireline networks. We also characterize the multicast capacity of linear finite-field deterministic relay networks when one source is multicasting the same information to multiple destinations, with the help of arbitrary number of relays.

Next, we use the insights obtained from the analysis of the deterministic model and present an achievable rate for general Gaussian relay networks. We show that the achievable rate is within a constant number of bits from the information-theoretic cut-set upper bound on the capacity of these networks. This constant depends on the number of nodes in the network, but not the values of the channel gains. Therefore, we uniformly characterize the capacity of Gaussian relay networks within a constant number of bits, for all channel parameters. For example, we approximate the unknown capacity of the three node Gaussian relay channel within one bit/sec/Hz.

Finally, we illustrate that the proposed deterministic approach is a general tool and can be applied to other problems in wireless network information theory. In particular we demonstrate its application to make progress in two other problems: two-way relay channel and relaying with side information.

Professor David Tse　　　　Date

Acknowledgements

Towards this truly exciting accomplishment in my life, my most sincere gratitude goes to my advisor, Professor David Tse for constant and generous support and guidance during my education at Berkeley. Beside his thoughtful ideas and exceptional knowledge of the field I would like to mostly appreciate his uniquely outstanding style of research. His mind provoking questions and thoughtful discussions, have greatly influenced my thought processes during the completion of my doctoral thesis. He has taught me the true meaning of scientific research. Also his creative approach to teaching and giving engaging presentations was particularly enlightening.

I would also like to profusely thank Professor Suhas Diggavi for sharing his knowledge and experience with me. I had one of my most amazing research experiences with him in the last three years. Since the beginning of this adventurous period of time, I have learned truly innovative research approaches as well as many information theoretic concepts from him. His positive thinking and encouraging words in difficult times gave me the confidence to master challenging research problems.

I would also like to express my gratitude to Professor Kannan Ramchandran for his great comments and suggestions since the beginning of my doctoral studies. His discussions and comments have greatly improved the quality of my work. I would like to also extend my gratitude to Professor Sourav Chatterjee for his comments and suggestions on my thesis.

I would like to also thank my fellows and colleagues in the wireless foundations, specially Krish Eswaran, Lenny Grokop, Bobak Nazer, Vinod Prabhakaran and Anand Sarwate. Also special thanks to my friend Arash Jamshidi for his wonderful comments and suggestions.

Finally, I would like to extend my deepest gratitude, love and affection to my beloved

family, for loving me, believing in me and wishing the best for me. I owe all my achievements to their pure hearts; such pure hearts that any thing they dream for me comes true.

To my beloved family, Parvaneh, Mehdi, Sahar

and

my love, Laleh

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Wireless communication is one of the most vibrant areas in the communication field today. Over the past decade we have witnessed quite a few successful solutions in the wireless industry, for example second-generation (2G) and third-generation (3G) digital wireless standards with more than half a billion subscribers worldwide. As history indicates, information theory has played a significant role in these achievements by providing elegant engineering insights for several key problems arising in these systems. So far, most of these problems have been in the context of a point to point communication system. This is mainly due to a centralized infrastructure deployed in current systems, such as cellular networks.

Looking ahead, we note that the next generation of wireless communication systems will be increasingly based on new principles such as cooperation between different network entities for efficient use of resources, and interference management strategies for coexistence of different wireless systems. Clearly, wireless communication systems are evolving from a centralized architecture to a distributed one. As a result, we need to study new information theoretical problems arising in *multiuser* communication systems.

Two main distinguishing features of wireless communication are:

- first, the *broadcast* nature of wireless communication; wireless users communicate over the air and signals from any one transmitter is heard by multiple nodes with possibly different signal strengths.

- second, the *superposition* nature; a wireless node receives signals from multiple simultaneously transmitting nodes, with the received signals all superimposed on top of each other.

Because of these two effects, links in a wireless network are never isolated but instead interact in seemingly complex ways. On the one hand this facilitates the spread of information among users in a network, on the other hand it can be harmful by creating interference among users. This is in direct contrast to wireline networks, where transmitter-receiver pairs can often be thought of as isolated point-to-point links, i.e., inducing a communication graph. While there has been significant progress in understanding network flow over wired networks [1; 2; 3; 4; 5], not much is known for wireless networks.

In communication, the linear additive Gaussian channel model is commonly used to capture fundamental features of a wireless channel. Over the past couple of decades, study of multiuser Gaussian networks has been an active area of research for many scientists. However, due to the complexity of the Gaussian model, except for the simplest networks such as the one-to-many Gaussian broadcast channel and the many-to-one Gaussian multiple access channel, the capacity region of most Gaussian networks is still unknown. For example, even the capacity of a three node Gaussian relay network, in which a point to point communication is assisted by one helper (relay), has been open for more than 30 years.

So, given the current state of knowledge, how can we proceed?

In this dissertation we propose a *deterministic* approach to this problem. We present a new deterministic channel model which is analytically simpler than the Gaussian model but yet still captures the two key features of wireless communication of broadcast and superposition. The motivation to study such a model is that in contrast to fixed point-to-point channels where noise is the only source of uncertainty, in multiuser communication, the signal interactions are a critical source of uncertainty. Therefore, for a first level of understanding, our focus is on such signal interactions rather than the received noise. One way to interpret this is that it captures the interference-limited rather than the noise-limited regime.

Our goal is to utilize the deterministic model to find ”near optimal” communication schemes for the Gaussian network, and hence approximate its capacity. Our approximation of interested, sandwiches the capacity in such a way that the approximation error does not depend on network channel gains and signal-to-noise ratio (SNR) of operation. In this sense, we seek a ”uniform” approximation of the capacity. Since the achievable rates grow with SNR, and the constant of our approximation is independent of it, we can see that for moderate SNR regimes, this approximation could be interesting. Moreover, the constants in the approximation are worst case bounds, and on the average, the characterization is much tighter. Another advantage of this approach is that we can now approximately characterize arbitrary wireless networks rather than specific networks. Moreover, depending on the regime of operation, perhaps this approximate characterization might be enough for engineering practice.

1.2 Background

In this dissertation we look at the unicast and multicast scenarios in wireless networks. In the unicast scenario, one source wants to communicate to a single destination with the help of other other nodes in the network, called relays. Similarly, in the multicast scenario the source wants to transmit the same message to multiple destinations. Since in these scenarios, all destination nodes are interested in the same message, there is no real interference in the network. Therefore we can focus on the cooperative aspect of wireless networks, which also makes the problem substantially easier than a general multi-source multi-destination problem. This will be used as a first step towards the understanding of more complex network topologies.

The 3-node relay channel was first introduced in 1971 by Vander Meulen [6] and the most general strategies for this network were developed by Cover and El Gamal [7]. There has also been a significant effort by researchers to generalize these ideas to arbitrary multi-relay networks. An early attempt was done in the Ph.D. Thesis of Aref [8] where a max-flow min-cut result was established to characterize the unicast capacity of a deterministic broadcast relay network which had *no multiple-access interference*. This was an early precursor to network coding which established the multicast capacity of wireline networks, a deterministic capacitated graph which had no broadcast or multiple-access interference [1; 2; 3]. These two ideas were combined in [9], which established a max-flow min-cut characterization for multicast flows for "Aref networks" which had general (deterministic) broadcast with no multiple-access interference. Unfortunately such complete characterizations are not known for arbitrary (even deterministic) networks with both broadcast and multiple-access interference. One notable exception is the work [10] which takes a scalar deterministic linear finite field model and uses probabilistic erasures to model channel failures. For this model using results of erasure broadcast networks [11], they established the unicast capacity through a max-flow min-cut characterization. Our deterministic model circumvents this need to introduce probabilistic erasures by constructing vector interactions modeling signal scales which seems to capture the essence of noisy (Gaussian) relay networks.

There has also been a rich body of literature in directly tackling the noisy relay network capacity characterization. In [12] the "diamond" network of parallel relay channel with no direct link between the source and the destination was examined. Xie and Kumar generalized the decode-forward encoding scheme for a network of multiple relays [13]. Gastpar and Vetterli established the asymptotic capacity of a single sender, single receiver network as the number of relay nodes increases [14]. Kramer et. al. [15], Reznik et. al.[16], Khojastepour et. al. [17], Laneman, Tse and Wornell [18], Mitra and Sabharwal [19], Sendonaris et. al. [20; 21], El Gamal and Zahedi [22], Nosratinia and Hedayat [23], Yuksel and Erkip [24], and many other authors have also addressed different aspects of relaying and cooperation in wireless networks in recent years.

Though there have been many interesting and important ideas developed in these papers, the capacity characterization of Gaussian relay networks is still unresolved. In fact even a performance guarantee, such as establishing how far these schemes are from an upper bound is unknown, and hence the approximation guarantees for these schemes is unclear. As we will see in Chapter 3, several of the strategies do not yield an approximation guarantee for general networks.

1.3 Contributions of this dissertation

We summarize our main contributions below, which are more precisely stated in Chapter 4. We first develop a linear deterministic model which incorporates signal scale interaction as well as the broadcast and superposition nature of wireless medium. We establish the connection of such a model to simple multiuser Gaussian networks in Chapter 2, which also suggests a constant-bit approximate characterization of such networks based on insights from the linear deterministic model. In fact this model suggests achievable strategies to explore in the noisy (Gaussian) relay networks as seen in Chapter 3 where we apply this philosophy to progressively complex networks. In fact, these examples demonstrate that several known strategies can be arbitrary far away from the optimality.

Given the utility of this deterministic approach, in Chapter 5 we examine arbitrary deterministic signal interaction model (not necessarily linear) and establish an achievable rate for an arbitrary network with such interaction (with broadcast and multiple-access). For the special case of linear deterministic models, this achievable rate matches an upper bound to the capacity, therefore the complete characterization is possible. The analysis for arbitrary deterministic functions requires the notion of *message typicality* which gives us a tool needed for the approximate characterization of Gaussian wireless relay network capacity.

The examination of the deterministic network relay network motivates the introduction of a simple coding strategy for general Gaussian relay networks. In this scheme each relay first quantizes the received signal at the noise level, then randomly maps it to a Gaussian codeword and transmits it. In Chapter 6 we use the insights of the deterministic result to demonstrate that we can achieve a rate that is guaranteed to be within a constant gap from the information-theoretic cut-set upper bound on capacity. This constant depends on the topological parameters of the network (number of nodes in the network), but not on the values of the channel gains. Therefore, we get a uniformly good approximation of the capacity of Gaussian relay networks, uniform over all values of the channel gains. Moreover in Chapter 7, we show that this scheme is robust to the knowledge of the channel at the relays, and therefore is applicable to a compound relay network where the gains come from a class of channels. Therefore, as long as the network can support a given rate, we can achieve it without the relays knowledge of the channel gains.

In Chapter 7, we establish several other extensions to our results.

1. Compound relay network
2. Frequency selective relay network
3. Half-duplex relay network
4. Quasi-static fading relay network (underspread regime)
5. Low rate capacity approximation of Gaussian relay network

In Chapter 8 we demonstrate a more precise connection between different channel models considered in this paper. In particular we illustrate in what sense these models are close to each other.

In Chapter 9, we further discuss applications of the deterministic approach to other problems in wireless network information theory. We look at two different problems; two-way relay channel and relaying with side information, and illustrate how to use the deterministic model to find a uniformly near optimal communication scheme for each problem. We end the dissertation with final notes and discussions.

Parts of this dissertation are published in [25; 26; 27; 28; 29; 30; 31].

Chapter 2

Deterministic modeling of wireless channel

2.1 Introduction

In this dissertation we consider a relay network represented by a general directed network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of vertices representing the communication nodes in the relay network, and $\mathcal{E}$ is the set of annotated links between nodes, which describe the contribution to the signal interaction. The network is not assumed to be simple and in general loops are allowed.

We consider both unicast and multicast communication problem scenarios. Therefore a special node $S \in \mathcal{V}$ is considered the source of the message and wants to simultaneously transmit its message to all destination nodes in the set $\mathcal{D}$. If $\mathcal{D}$ contains only one node we have a unicast scenario, otherwise a multicast scenario where all nodes in $\mathcal{D}$ are interested in the same message from the source. All other nodes in the network facilitate communication between S and $\mathcal{D}$. The relationship between the received signal at a node and the transmitted signals from its neighbors is described by the channel model.

The multiuser Gaussian channel model is the standard one used in modeling the fundamental features of a wireless channel: signal strength, broadcast and superposition. The main goal in this dissertation is to get a uniform approximation of the capacity of Gaussian relay networks. To accomplish this goal, there are two main steps: first to find a ”good” relaying scheme, second to analyze the performance of this scheme and demonstrate that it achieves an approximate characterization of the capacity of Gaussian relay network for all channel gains. However, due to the complexity of the Gaussian model, both steps are quite challenging, since the model accounts for both signal interaction as well as noise.

As discussed in the introduction, our approach is to introduce and analyze a simpler linear finite-field deterministic channel model that is closely connected to the Gaussian model. The simplicity of this model allows us to make progress and get insights into Gaussian relay networks. Furthermore, we also develop new proof techniques that can also be utilized in noisy (Gaussian) relay networks.

The goal of this chapter is to introduce the linear deterministic model and illustrate how we can deterministically model three key features of a wireless channel: signal strength, broadcast and superposition.

2.2 Modeling signal strength

Consider the *real* scalar Gaussian model for point to point link,

$$y = hx + z \tag{2.1}$$

where $z \sim \mathcal{N}(0,1)$. There is also an average power constraint $E[|x|^2] \leq 1$ at the transmitter. The transmit power and noise power are both normalized to be equal to 1 and the signal-to-noise ratio (SNR) is captured in terms of channel gains. So h is a *fixed* real number representing the channel gain (signal strength), and

$$|h| = \sqrt{\mathsf{SNR}} \tag{2.2}$$

It is well known that the capacity of this point-to-point channel is

$$C_{\text{AWGN}} = \frac{1}{2}\log\left(1 + \mathsf{SNR}\right) \tag{2.3}$$

To get an intuitive understanding of this capacity formula let us write the received signal in equation (2.1), y, in terms of the binary expansions of x and z. For simplicity assume h, x and z are positive real numbers, then we have

$$y = 2^{\frac{1}{2}\log \mathsf{SNR}} \sum_{i=1}^{\infty} x(i)2^{-i} + \sum_{i=-\infty}^{\infty} z(i)2^{-i} \tag{2.4}$$

To simplify the effect of background noise assume it has a peak power equal to 1. Then we can write

$$y = 2^{\frac{1}{2}\log \mathsf{SNR}} \sum_{i=1}^{\infty} x(i)2^{-i} + \sum_{i=1}^{\infty} z(i)2^{-i} \tag{2.5}$$

or,

$$y \approx 2^{n} \sum_{i=1}^{n} x(i)2^{-i} + \sum_{i=1}^{\infty} \left(x(i+n) + z(i)\right) 2^{-i} \tag{2.6}$$

where $n = \lceil \frac{1}{2}\log \mathsf{SNR} \rceil^{+}$. Therefore if we just ignore the 1 bit of the carry-over from the second summation ($\sum_{i=1}^{\infty} \left(x(i+n) + z(i)\right) 2^{-i}$) to the first summation ($2^n \sum_{i=1}^{n} x(i)2^{-i}$) we can intuitively model a point-to-point Gaussian channel as a pipe that truncates the transmitted signal and only passes the bits that are above the noise level. Therefore think of transmitted signal x as a sequence of bits at different signal levels, with the highest signal level in x being the most significant bit (MSB) and the lowest level being the least significant bit (LSB). In this simplified model the receiver can see the n most significant bits of x without any noise and the rest are not seen at all. Clearly there is a correspondence between n and SNR in dB scale,

$$n \leftrightarrow \lceil \frac{1}{2} \log \mathsf{SNR} \rceil^{+} \tag{2.7}$$

As we notice in this simplified model there is no background noise any more and hence it is a *deterministic model*. Pictorially the deterministic model corresponding to the AWGN channel is shown in Figure 2.1. In this figure, at the transmitter there are several small circles. Each circle represents a signal level and a binary digit can be put for transmission at each signal level. Depending on n, which represents the channel gain in dB scale, the transmitted bits at the first n signal levels will be received clearly at the destination. However the bits at other signal levels will not go through the channel.

These signal levels can potentially be created by using a multi-level lattice code in the AWGN channel [32]. Then the first n levels in the deterministic model represent those levels (in the lattice chain) that are above noise level, and the remaining are the ones that are below noise level. Therefore, if we think of the transmit signal, $\mathbf{x}$, as a binary vector of length q, then the deterministic channel delivers only its first n bits to the destination. We can algebraically write this input-output relationship by shifting $\mathbf{x}$ down by $q - n$ elements or more precisely

$$\mathbf{y} = \mathbf{S}^{\mathbf{q-n}}\mathbf{x} \tag{2.8}$$

where $\mathbf{x}$ and $\mathbf{y}$ are binary vectors of length q denoting transmit and received signals respec-

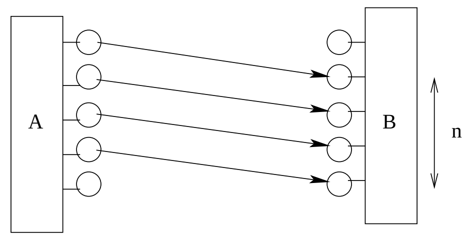


Figure 2.1: Pictorial representation of the deterministic model for point-to-point channel.

tively and $\mathbf{S}$ is the $q \times q$ shift matrix,

$$\mathbf{S} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \tag{2.9}$$

Clearly the capacity of this deterministic point-to-point channel is n, where

$$n = \lceil \frac{1}{2} \log \mathsf{SNR} \rceil^+ \tag{2.10}$$

It is interesting to note that this is a within $\frac{1}{2}$-bit approximation of the capacity of the AWGN channel[1]. In the case of complex Gaussian channel we set $n = \lceil \log \mathsf{SNR} \rceil^+$ and we get a within 1-bit approximation of the capacity.

[1]Note that this connection is only in the capacity without a formal connection in coding scheme or a direct translation of the capacity.

2.3 Modeling broadcast

Based on the intuition obtained so far, it is straightforward to think of a deterministic model for a broadcast scenario. Consider the real scalar Gaussian broadcast channel. Assume there are only two receivers. The received SNR at receiver i is denoted by SNR_i for $i = 1, 2$. Without loss of generality assume $\mathsf{SNR}_2 \leq \mathsf{SNR}_1$. Consider the binary expansion of the transmitted signal, x. Then we can deterministically model the Gaussian broadcast channel as the following:

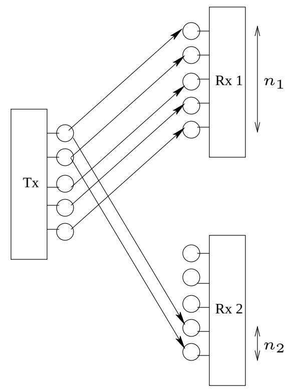
- Receiver 2 (weak user) receives only the first n_2 bits in the binary expansion of x. Those bits are the ones that arrive above the noise level.

- Receiver 1 (strong user) receives the first n_1 ($n_1 > n_2$) bits in the binary expansion of x. Clearly these bits contain what receiver 1 gets.

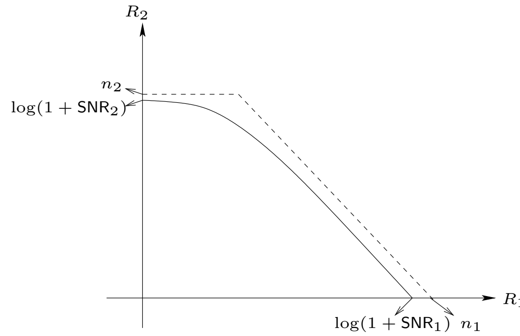
The deterministic model in some sense abstracts away the use of superposition coding and successive interference cancellation decoding in the Gaussian broadcast channel. Therefore the first n_2 levels in the deterministic model represent the cloud center that is decoded by both users, and the remaining $n_1 - n_2$ levels represent the cloud detail that is decoded only by the strong user (after decoding the cloud center and canceling it from the received signal).

Pictorially the deterministic model for a Gaussian broadcast channel is shown in figure 2.2 (a). In this particular example $n_1 = 5$ and $n_2 = 2$, therefore both users receive the first two most significant bits of the transmitted signal. However user 1 (strong user) receives additional three bits from the next three signal levels of the transmitted signal. There is also the same correspondence between n and channel gains in dB: $n_i \leftrightarrow \lceil \log \mathsf{SNR}_i \rceil^+$, for $i = 1, 2$.

To analytically demonstrate how closely we are modeling the Gaussian BC channel, the capacity region of Gaussian BC channel and deterministic BC channel are shown in



(a) Pictorial representation of the deterministic model for Gaussian BC



(b) Capacity region of Gaussian BC channel (solid line). Capacity region of deterministic BC channel (dashed line)

Figure 2.2: Pictorial representation of the deterministic model for Gaussian BC is shown in (a). Capacity region of Gaussian and deterministic BC are shown in (b).

Figure 2.2 (b). As it is seen their capacity regions are very close to each other. In fact it is easy to verify that for all SNR's these regions are always within one bit per user of each other (*i.e.* if a pair (R_1, R_2) is in the capacity region of the deterministic BC then there is a pair within one bit per component of (R_1, R_2) that is in the capacity region of the Gaussian BC)[2]. However, this is only the worst case gap and in a typical case that SNR_1 and SNR_2 are very different the gap is much smaller than one bit.

[2]A cautionary note is that as in the point-to-point case the connection is not formed in the coding scheme but just in capacity regions.

2.4 Modeling superposition

Consider a superposition scenario in which two users are simultaneously transmitting to a node. In the Gaussian model the received signal can be written as

$$y = h_1 x_1 + h_2 x_2 + z. \tag{2.11}$$

To intuitively see what happens in superposition in the Gaussian model, we again write the received signal, y, in terms of the binary expansions of x_1, x_2 and z. Assume x_1, x_2 and z are all real numbers smaller than one, and also the channel gains are

$$h_i = \sqrt{\mathsf{SNR}_i}, \quad i = 1, 2 \tag{2.12}$$

Without loss of generality assume $\mathsf{SNR}_2 < \mathsf{SNR}_1$. Then we have

$$y = 2^{\frac{1}{2}\log \mathsf{SNR}_1} \sum_{i=1}^{\infty} x_1(i) 2^{-i} + 2^{\frac{1}{2}\log \mathsf{SNR}_2} \sum_{i=1}^{\infty} x_2(i) 2^{-i} + \sum_{i=-\infty}^{\infty} z(i) 2^{-i} \tag{2.13}$$

To simplify the effect of background noise assume it has a peak power equal to 1. Then we can write

$$y = 2^{\frac{1}{2}\log \mathsf{SNR}_1} \sum_{i=1}^{\infty} x_1(i) 2^{-i} + 2^{\frac{1}{2}\log \mathsf{SNR}_2} \sum_{i=1}^{\infty} x_2(i) 2^{-i} + \sum_{i=1}^{\infty} z(i) 2^{-i} \tag{2.14}$$

or,

$$\begin{aligned} y \;\approx\; & 2^{n_1} \sum_{i=1}^{n_1-n_2} x_1(i) 2^{-i} + 2^{n_2} \sum_{i=1}^{n_2} \left(x_1(i+n_1-n_2) + x_2(i) \right) 2^{-i} \\ & + \sum_{i=1}^{\infty} \left(x_1(i+n_1) + x_2(i+n_2) + z(i) \right) 2^{-i} \end{aligned} \tag{2.15}$$

where $n_i = \lceil \frac{1}{2} \log \mathsf{SNR}_i \rceil^+$ for $i = 1, 2$. Therefore based on the intuition obtained from the point-to-point and broadcast AWGN channels, we can approximately model this as the following:

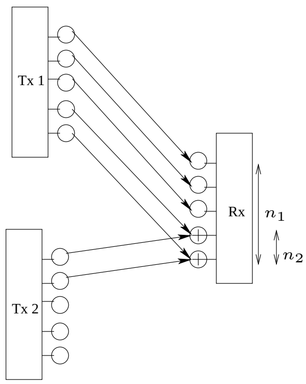
- That part of x_1 that is above SNR_2 ($x_1(i)$, $1 \leq i \leq n_1 - n_2$) is received clearly without any interaction from x_2.
- The remaining part of x_1 that is above noise level ($x_1(i)$, $n_1 - n_2 < i \leq n_1$) and that part of x_2 that is above noise level ($x_1(i)$, $1 \leq i \leq n_2$) interact with each other and are received without any noise.
- Those parts of x_1 and x_2 that are below noise level are truncated and not received at all.

The key point is how to model the interaction between the bits that are received at the same signal level. In our deterministic model we ignore the carry-overs of the real addition and we model the interaction by the modulo 2 sum of the bits that are arrived at the same signal level. Pictorially the deterministic model for a Gaussian MAC channel is shown in figure 2.3 (a). Analogous to the deterministic model for the point-to-point channel, we can write

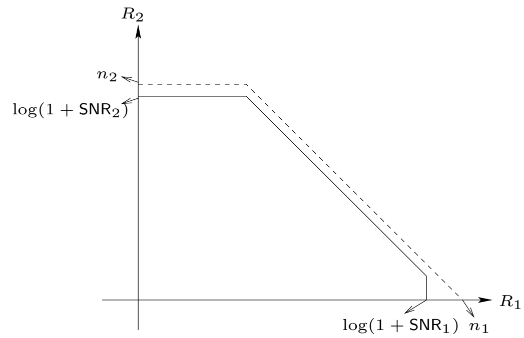
$$\mathbf{y} = \mathbf{S^{q-n_1}x_1} \oplus \mathbf{S^{q-n_2}x_2} \tag{2.16}$$

where the summation is in $\mathbb{F}_2$ (modulo 2). Here $\mathbf{x_i}$ ($i = 1, 2$) and $\mathbf{y}$ are binary vectors of length q denoting transmit and received signals respectively and $\mathbf{S}$ is a $q \times q$ shift matrix. There is also the same relationship between n_i's and the channel gain in dB: $n_i \leftrightarrow \lceil \log \mathsf{SNR}_i \rceil^+$, for $i = 1, 2$. Note that if one wants to make a connection between the deterministic model and real Gaussian MAC channel (rather than complex) a factor of $\frac{1}{2}$ is necessary.

Now compared to simple point-to-point case we now have interaction between the bits that receive at the same signal level at the receiver. However, we limit the receiver to observe only the modulo 2 summation of those bits that arrive at the same signal level. In



(a) Pictorial representation of the deterministic MAC.



(b) Capacity region of Gaussian MAC. (solid line). Capacity region of deterministic MAC.(dashed line)

Figure 2.3: Pictorial representation of the deterministic MAC is shown in (a). Capacity region of Gaussian and deterministic MACs are shown in (b).

some sense this way of modeling interaction is similar to the collision model. In the collision model if two packets arrive simultaneously at a receiver, both are dropped; similarly here if two bits arrive simultaneously at the same signal level the receiver gets only their modulo 2 sum, which means it can not figure out any of them. On the other hand, unlike in the simplistic collision model where the entire packet is lost when there is collision, the most significant bits of the stronger user remain intact. This is reminiscent of the familiar *capture* phenomenon in CDMA systems: the strongest user can be heard even when multiple users simultaneously transmit.

Now we can apply this model to Gaussian multiple access channel (MAC), in which

$$y = h_1 x_1 + h_2 x_2 + z \tag{2.17}$$

where $z \sim \mathcal{CN}(0,1)$. There is also an average power constraint equal to 1 at both transmitters. A natural question is how close is the capacity region of the deterministic model to that of the actual Gaussian model. Without loss of generality assume $\mathsf{SNR}_2 < \mathsf{SNR}_1$. The capacity region of this channel is well-known to be the set of non-negative pairs (R_1, R_2) satisfying

$$R_i \leq \log(1 + \mathsf{SNR}_i), \quad i = 1, 2 \tag{2.18}$$

$$R_1 + R_2 \leq \log(1 + \mathsf{SNR}_1 + \mathsf{SNR}_2) \tag{2.19}$$

This region is plotted with solid line in figure 2.3 (b).

It is easy to verify that the capacity region of the deterministic MAC is the set of non-negative pairs (R_1, R_2) satisfying

$$R_2 \leq n_2 \tag{2.20}$$

$$R_1 + R_2 \leq n_1 \tag{2.21}$$

where $n_i = \log \mathsf{SNR}_i$ for $i = 1, 2$. This region is plotted with dashed line in figure 2.3 (b). In this deterministic model the "carry-over" from one level to the next that would happen with real addition is ignored. However as we notice still the capacity region is very close to the capacity region of the Gaussian model. In fact it is easy to verify that they are within one bit per user of each other (*i.e.* if a pair (R_1, R_2) is in the capacity region of the deterministic MAC then there is a pair within one bit per component of (R_1, R_2) that is in the capacity region of the Gaussian MAC). The intuitive explanation for this is that in real addition once two bounded signals are added together the magnitude increases however, it can only become as large as twice the maximum size of individual ones. Therefore the cardinality size of summation is increased by at most one bit. On the other hand in finite-field addition there is no magnitude associated with signals and the summation is still in the same field size as the individual signals. So the gap between Gaussian and deterministic model for two user MAC is intuitively this one bit of cardinality increase. Similar to the broadcast example, this is only the worst case gap and when the channel gains are different it is much smaller than one bit.

Now we define the linear finite-field deterministic model.

2.5 Linear finite-field deterministic model

In this paper we consider a relay network represented by a general directed network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of vertices representing the communication nodes in the relay network, and $\mathcal{E}$ is the set of annotated between nodes, which describe the contribution to the signal interaction. The network is not assumed to be simple and in general loops are allowed.

In the linear finite-field deterministic model the communication link from node i to node j has a non-negative integer gain[3] $n_{(i,j)}$ associated with it. This number models the channel gain in a corresponding Gaussian setting. At each time t, node i transmits a vector $\mathbf{x}_i[t] \in \mathbb{F}_p^q$ and receives a vector $\mathbf{y}_i[t] \in \mathbb{F}_p^q$ where $q = \max_{i,j}(n_{(i,j)})$ and p is a positive integer indicating the field size. The received signal at each node is a deterministic function of the transmitted signals at the other nodes, with the following input-output relation: if the nodes in the network transmit $\mathbf{x}_1[t], \mathbf{x}_2[t], \ldots \mathbf{x}_N[t]$ then the received signal at node j, $1 \leq j \leq N$ is:

$$\mathbf{y}_j[t] = \sum_{i \in \mathcal{N}_j} \mathbf{S}^{q-n_{i,j}} \mathbf{x}_i[t] \tag{2.22}$$

where the summations and the multiplications are in $\mathbb{F}_p$. In this paper the field size is assumed to be two, $p = 2$, unless it is stated otherwise.

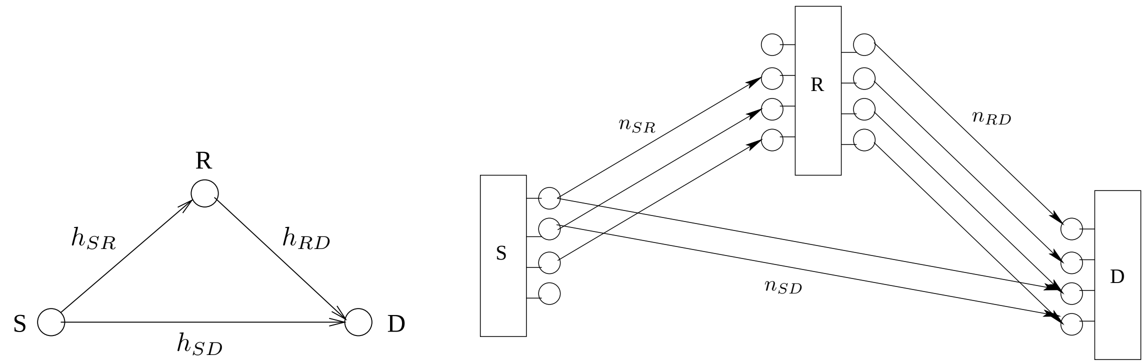
[3] Some channels may have zero gain.

Chapter 3

Motivation of our approach

3.1 Introduction

In this chapter we motivate and illustrate our approach. We look at three simple relay networks and illustrate how the analysis of these networks under the simpler linear finite-field deterministic model enables us to conjecture a near optimal relaying scheme for the Gaussian case and using this insight to provably approximate the capacity of these networks under the Gaussian model within a constant number of bits. We progress from the relay channel where several strategies yield uniform approximation to more complicated networks where progressively we see that several "simple" strategies in the literature fail to achieve a constant gap. Using the deterministic model we can whittle down the potentially successful strategies. In fact we can show that the set of strategies that yield a universal approximation shrink as we progress to more complex networks. This illustrates the power of the deterministic model to provide insights into transmission techniques for the noisy networks.



(a) The Gaussian relay channel

(b) The linear finite-field deterministic relay channel

Figure 3.1: The relay channel: (a) Gaussian model, (b) Linear finite-field deterministic model

3.2 One relay network

We start by looking at the simplest Gaussian relay network with only one relay as shown in figure 3.1 (a). We examine whether it is possible to approximate its capacity uniformly (uniform over all channel gains). To answer this question positively we need to find a relaying protocol that achieves a rate close to an upper bound on the capacity for all channel parameters. To find such a scheme we use the linear finite-field deterministic model to gain insight. The corresponding linear finite-field deterministic model of this relay channel with channel gains denoted by n_{SR}, n_{SD} and n_{RD} is shown in Figure 3.1 (b). It is easy to see that the capacity of this deterministic relay channel, C^d_{relay}, is smaller than both the maximum number of bits that can be broadcasted from the relay, and the maximum number of bits that the destination can receive. Therefore.

$$C^d_{relay} \leq \min\left(\max(n_{SR}, n_{SD}), \max(n_{RD}, n_{SD})\right) \tag{3.1}$$

$$= \begin{cases} n_{SD}, & \text{if } n_{SD} > \min\left(n_{SR}, n_{RD}\right); \\ \min\left(n_{SR}, n_{RD}\right), & \text{otherwise.} \end{cases} \tag{3.2}$$

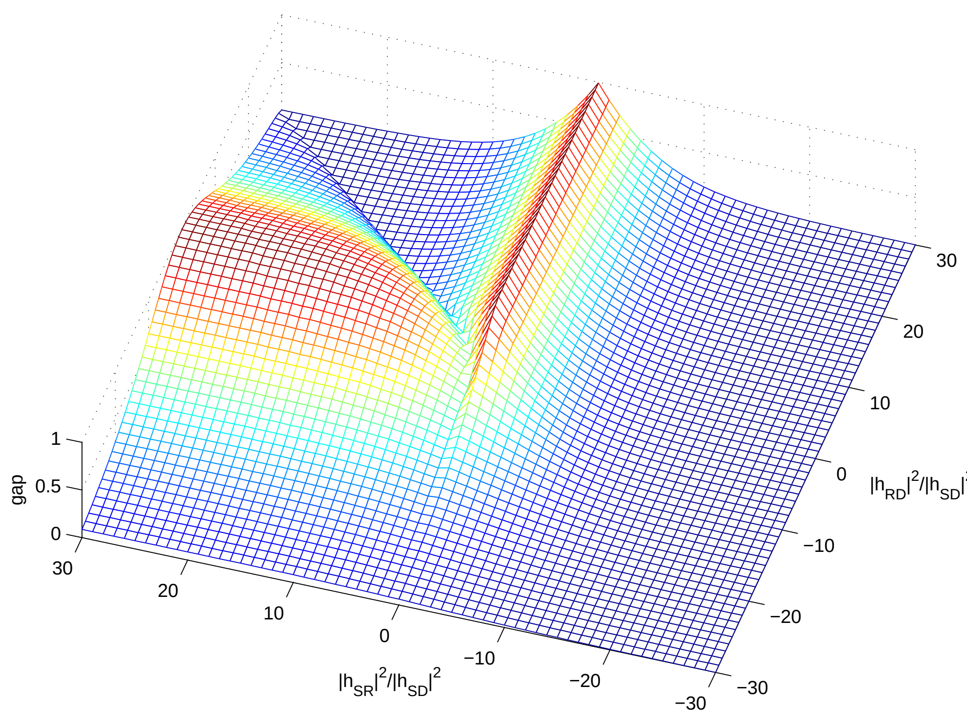


Figure 3.2: The gap between cut-set upper bound and achievable rate of decode-forward scheme in the Gaussian relay channel for different channel gains (in dB scale).

This bound simply upper bounds the capacity by the maximum number of bits that can can be sent from one side of a cut in the network (containing the source) to the other side of the cut (containing the destination), assuming that the nodes on each side of the cut can fully collaborate with each other, hence it is called the *cut-set upper* bound.

Note that equation (3.2) naturally implies a capacity-achieving scheme for this deterministic relay network: if the direct link is better than any of the links to/from the relay then the relay is silent, otherwise it helps the source by decoding its message and sending innovative bits. This suggests a decode-and-forward scheme for the original Gaussian relay channel. The question is: how does it perform? In the following theorem we show that for one-relay network the decode-forward scheme achieves within one bit/sec/Hz of the capacity for all channel parameters.

Theorem 3.2.1. *Decode-forward relaying protocol achieves within 1 bit/sec/Hz of the capacity of the one-relay Gaussian network, for all channel gains.*

Proof. See Appendix A.1. □

Therefore we showed that the maximum gap between decode-forward achievable rate and the cut-set upper bound on the capacity of Gaussian relay network is at most one bit. However we should point out that even this 1-bit gap is too conservative in many parameter values. In fact the gap would be at the maximum value only if two of the channel gains are exactly the same. Since in a wireless scenario the channel gains differ significantly this happens very rarely. In figure 3.2 the gap between the achievable rate of decode-forward scheme and the cut-set upper bound is plotted for different channel gains. In this figure x and y axis are respectively representing the channel gains from relay to destination and source to relay normalized by the gain of the direct link (source to destination) in dB scale. The z axis shows the value of the gap (in bits/sec/Hz). There are two main points that one should note in this figure: first that the gap is at most one bit which is consistent with what we showed in this section. Second, on the average the gap is much less than one bit.

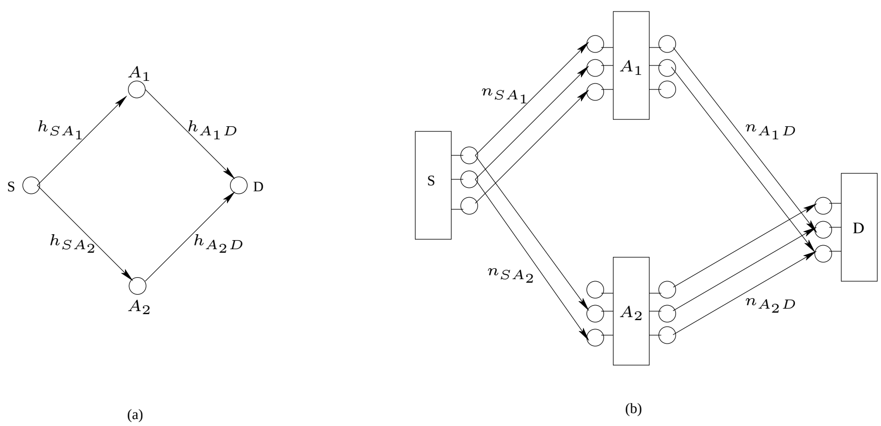


Figure 3.3: Diamond network with two relays: (a) Gaussian model, (b) Linear finite-field deterministic model

Note that the deterministic network in Figure 3.1 (b), suggests that several other relaying strategies are also optimal. For example doing a compress and forwarding will also achieve the cut-set bound. Moreover a "network coding" strategy of sending the sum (or linear combination) of the received bits will also be optimal as long as the destination receives linearly independent combinations. All these schemes can also be translated to the Gaussian case and can be shown to be uniformly approximate strategies. Therefore for the simple relay channel there are many successful candidate strategies. As we will see, this set shrinks as we go to larger relay networks.

3.3 Diamond network

Now consider the diamond Gaussian relay network, with two relays, as shown in Figure 3.3 (a). Brett Schein introduced this network in his Ph.D. thesis [12] and investigated its capacity. However the capacity of this network is still an open problem. We examine whether it is possible to uniformly approximate its capacity.

First we build the corresponding linear finite field deterministic model for this relay network as shown in Figure 3.3 (b). To investigate its capacity first we relax the interactions

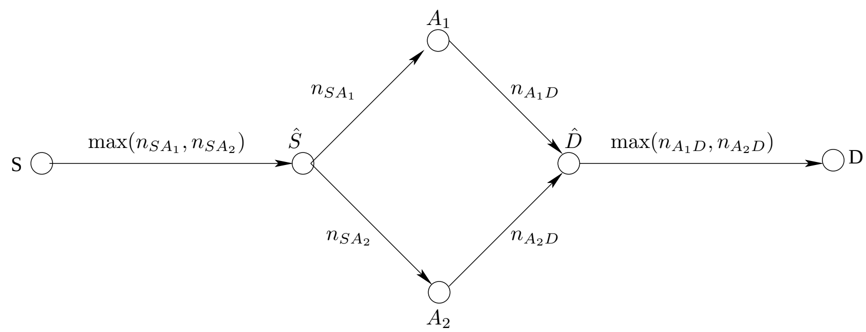


Figure 3.4: Wireline diamond network

between incoming links at each node and create the wireline network shown in Figure 3.4. In this network there are two other links added, which are from S to $\hat{S}$ and from $\hat{D}$ to D. Since the capacities of these links are respectively equal to the maximum number of bits that can be sent by the source and maximum number of bits that can be received by the destination in the original linear finite-field deterministic network, the capacity of the wireline diamond network cannot be smaller than the capacity of the linear finite-field deterministic diamond network. Now by the max-flow min-cut theorem we know that the capacity $C^{w}_{diamond}$ of the wireline diamond network is equal to the value of its minimum cut. Hence

$$\begin{aligned} C^{d}_{\text{diamond}} &\leq C^{w}_{\text{diamond}} \\ &= \min\left\{\max(n_{SA_1}, n_{SA_2}), \max(n_{A_1D}, n_{A_2D}), n_{SA_1} + n_{A_2D}, n_{SA_2} + n_{A_1D}\right\} \end{aligned} \quad (3.3)$$

As we will show in Section 5, this upper bound is in fact the cut-set upper bound on the capacity of the deterministic diamond network.

Now, we know that the capacity of the wireline diamond network is achieved by a routing solution. It is not also difficult to see that we can indeed mimic this routing solution in the linear finite-field deterministic diamond network and send the same amount of information through non-interfering links from source to relays and then from relays to destination.

Therefore the capacity of the deterministic diamond network is equal to its cut-set upper bound.

A natural analogy of this routing scheme for the Gaussian network is the following partial decode-and-forward strategy:

1. The source broadcasts two messages, m_1 and m_2, at rate R_1 and R_2 to relays A_1 and A_2.

2. Each relay A_i decodes message m_i, $i = 1, 2$.

3. Then A_1 and A_2 re-encode the messages and transmit them via the MAC channel to the destination.

Clearly at the end the destination can decode both m_1 and m_2 if (R_1, R_2) is inside the capacity region of the BC from source to relays as well as the capacity region of the MAC from relays to the destination. In the following theorem we show that for the two-relay diamond network partial decode-forward scheme achieves within one bit/sec/Hz of the capacity for all channel parameters.

Theorem 3.3.1. *Partial decode-forward relaying protocol achieves within 1 bit/sec/Hz of the capacity of the two-relay diamond Gaussian network, for all channel gains.*

Proof. See Appendix A.2. □

We can also use the linear finite-field deterministic model to understand why other simple protocols such as decode-forward and amplify-forward are not universally approximate strategies for the diamond relay network.

For example consider the linear-finite field deterministic diamond network shown in Figure 3.5 (a). Clearly the cut-set upper bound on the capacity of this network is 3 bits/unit time. In a decode-forward scheme, all participating relays should be able to decode the message. Therefore the maximum rate of the message broadcasted from the source can at most be 2 bits/unit time. Also, if we ignore relay A_2 and only use the stronger relay, still it is not possible to send information more at a rate more than 1 bit/unit time. As a result we cannot achieve the capacity of this network by using a decode-forward strategy.

Now we can use this deterministic diamond network example to illustrate that in the Gaussian diamond network the gap between the achievable rate of the decode-forward and amplify-forward schemes and the cut-set upper bound can be arbitrary large. Consider the corresponding Gaussian network of this example as shown in figure 3.5 (b). Assume a is a large real number. The cut-set upper bound is approximately,

$$\overline{C} \approx 3 \log a \tag{3.4}$$

Now clearly the achievable rate of the decode-forward strategy is upper bounded by

$$R_{DF} \leq 2 \log a \tag{3.5}$$

Therefore, as a gets larger, the gap between the achievable rate of decode-forward strategy and the cut-set upper bound (3.4) increases.

Now let us look at the amplify-forward scheme. Although this scheme does not require all relays to decode the entire message, it can be quite sub-optimal if relays inject significant noise into the system. We use the deterministic model to intuitively see this effect. In a deterministic network, the amplify-forward operation can be simply modeled by shifting bits up and down at each node. However, once the bits are shifted up the newly created LSB's represent the amplified bits of the noise and we model them by random bits. Now, consider the example shown in Figure 3.5 (a). We notice that to achieve a rate of 3 from the source to the destination, the bit at the lowest signal level of the source's signal should go through A_1 while the remaining two are going through A_2. Now if A_2 is doing amplify-forward, it will have two choices: to either forward the received signal without amplifying it, or to amplify the received signal to have three signal levels in magnitude and forward it.

The effective networks under these two strategies are respectively shown in figure 3.5 (c) and 3.5 (d). In the first case, since the total rate going through the MAC from A_1 and A_2 to D is less than two, the overall achievable rate cannot exceed two. In the second case, however, the inefficiency of amplify-forward strategy comes from the fact that A_2 is transmitting pure noise on its lowest signal level. As a result, it is corrupting the bit transmitted by A_1 and reducing the total achievable rate again to two bits/unit time. Therefore, for this channel realization, amplify-forward scheme does not achieve the capacity. This intuition can again be made more rigorous for the Gaussian case to show that amplify and forward is not a universally-approximate strategy for the diamond network.

In the diamond network it can be shown that though decode-forward and amplify-forward relaying strategies fail, other strategies such as partial decode-forward, compress-forward as well as quantize-map (the main strategy analyzed in this dissertation for general networks) are still potential universally-approximate strategies. Hence the set of possible strategies that are always universally-approximate for any network shrinks.

3.4 A four relay network

Now we consider a more complicated relay network with four relays, as shown in Figure 3.6. As the first step lets find the optimal relaying strategy for the corresponding linear finite field deterministic model. Consider an example of a linear finite field deterministic relay network shown in Figure 3.7 (a). It is easy to see that the cut-set upper bound on the capacity of this relay network is 5. Now consider the following relaying strategy,

- Source broadcasts $\mathbf{b} = [b_1, \ldots, b_5]^t$
- Relay A_1 decodes b_3, b_4, b_5 and relay A_2 decodes b_1, b_2
- Relay A_1 and A_2 respectively send $\mathbf{x}_{A_1} = [b_3, b_4, b_5, 0, 0]^t$ and $\mathbf{x}_{A_2} = [b_1, b_2, 0, 0, 0]^t$

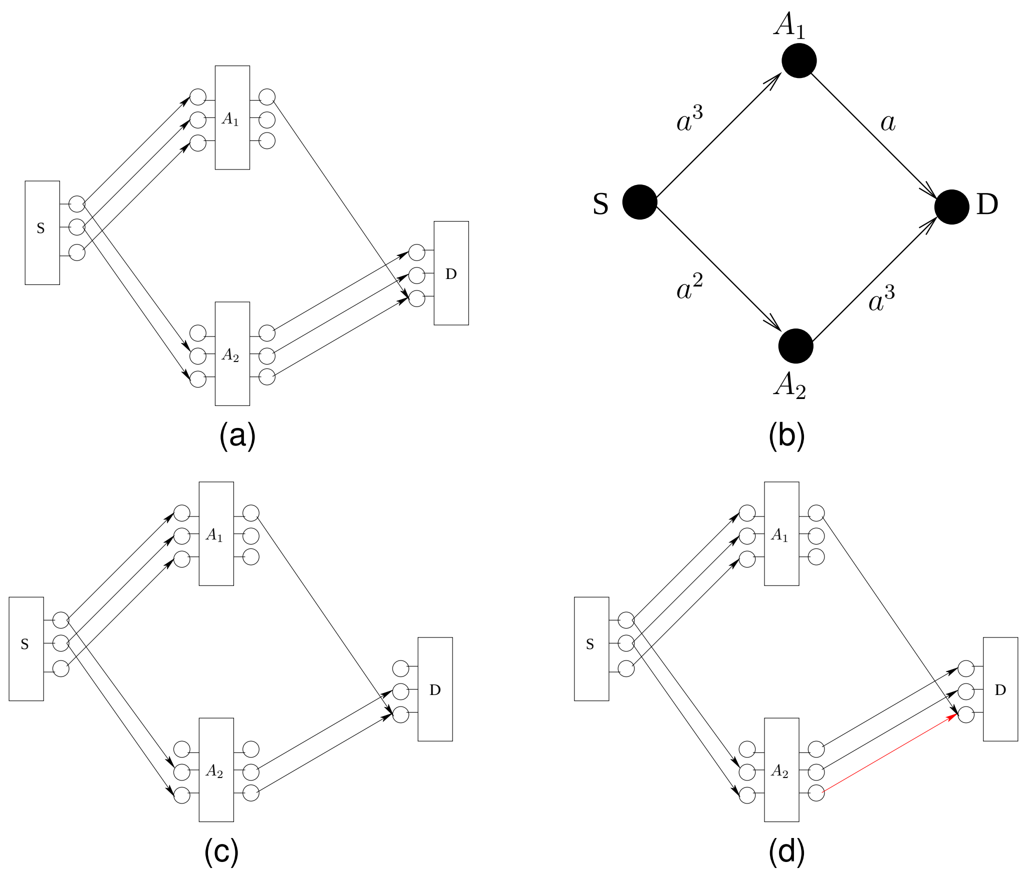
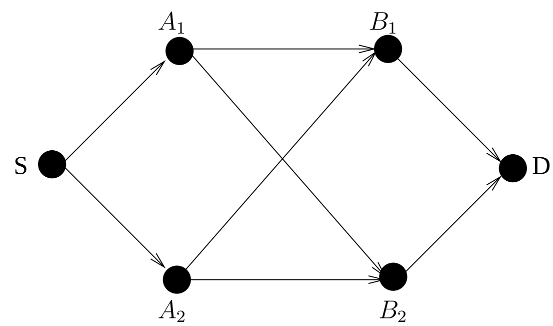


Figure 3.5: An example of the linear finite field deterministic diamond network is shown in (a). The corresponding Gaussian network is shown in (b). The effective network when R_2 just forwards the received signal is shown in (c). The effective network when R_2 amplifies the received signal to shift it up one signal level and then forward the message is shown in (d).



(a)

Figure 3.6: A two layer relay network with four relays.

- Relay B_2 decodes b_1, b_2, b_3 and sends $\mathbf{x}_{B_2} = [b_1, b_2, b_3, 0, 0]^t$
- Relay B_1 receives $\mathbf{y}_{B_1} = [0, 0, b_3, b_4 \oplus b_1, b_5 \oplus b_2]^t$ and forwards the last two equations, $\mathbf{x}_{B_1} = [b_4 \oplus b_1, b_5 \oplus b_2, 0, 0, 0]^t$
- The destination gets $\mathbf{y}_D = [b_1, b_2, b_3, b_4 \oplus b_1, b_5 \oplus b_2]^t$ and is able to decode all five bits.

Clearly with this scheme we can achieve the cut-set upper bound for this particular example. As one can note, in this optimal scheme the relay B_1 is not decoding or partially decoding a message, it is forwarding the last two LSB's. One may wonder if this is necessary, or in another words is any choice of partial decode-forward strategy suboptimal in this example? To answer this question. note that any partial decode-forward scheme can be visualized as different flows of information going from S to D that do not get mixed in the network. Now since all transmit signal levels of A_1 and A_2 are interfering with each other, it is not possible to get a rate of more than 3 bits/unit time by any partial decode-forward scheme in this example and hence it is always suboptimal.

The optimal scheme that we demonstrated above may look like a compress-forward strategy for Gaussian networks (described in [15] section V). But, as we will now show in fact a simple compress-forward strategy with Gaussian auxiliary random variables can in

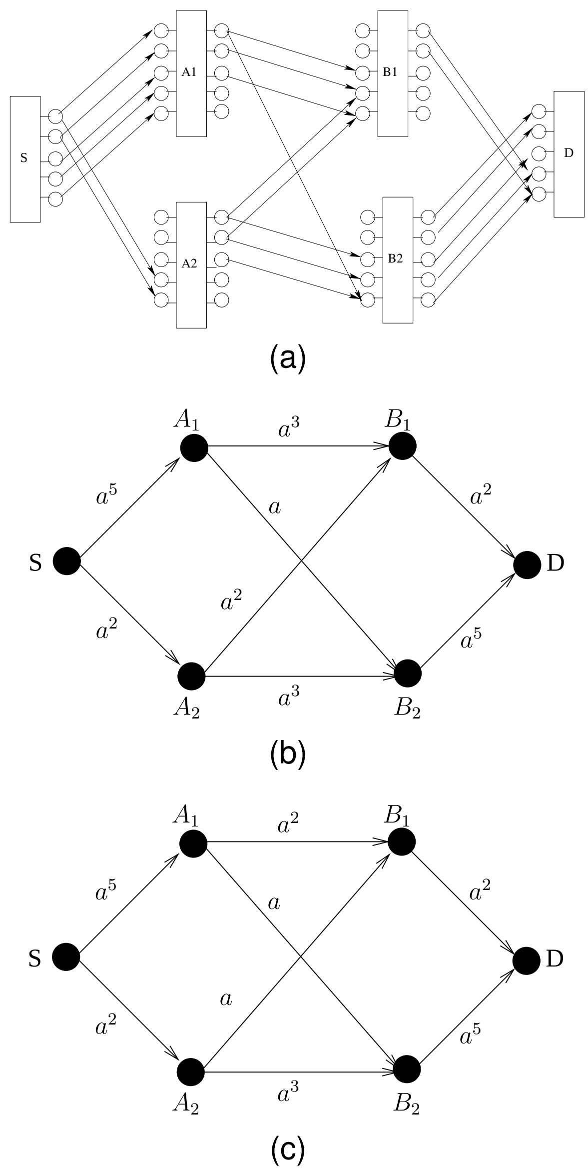


Figure 3.7: An example of a four relay linear finite filed deterministic relay network is shown in (a). The corresponding Gaussian relay network is shown in (b). The effective Gaussian network for compress-forward strategy is shown in (c).

general be far from the cut-set upper bound. So the corresponding scheme for Gaussian relay networks is not a simple compress-forward strategy.

Consider the example shown in Figure 3.7 (b). For large values of a, cut-set upper bound on the capacity of this relay network is approximately

$$\overline{C} \approx 5 \log a \tag{3.6}$$

The achievable rate of the compress-forward scheme is characterized in Theorem 3 ([15] page 9), which is in the form of a mutual information maximization over auxiliary random variables $U_{\mathcal{T}}$ and $\hat{Y}_{\mathcal{T}}$. Even though this is written in single-letter form, since there is no cardinality bounds, the rate optimization is still an infinite dimensional optimization problem. However, to simplify this problem further, assume that auxiliary random variables $U_{\mathcal{T}}$ are set to zero, and $\hat{Y}_{\mathcal{T}}$ are restricted to have a Gaussian distribution, which leads to a finite dimensional problem.

The scheme is such that the Wyner$-$Ziv source-coding region of each layer must intersect the channel-coding region of the next layer. As a result by looking at layer $\{B_1, B_2\}$ we note that node B_1 should compress its received signal to a Gaussian random variable with variance a^2. In another words, just quantize the received signal with distortion a. Therefore the effective network will look like the one shown in Figure 3.7 (c). Note that now the cut-set upper bound of this new network is approximately, $\overline{C}' \approx 4 \log a$.

As a result, with this compress-forward scheme, it is not possible to get a rate more than $4 \log a$. As a increases the gap between the achievable rate of compress-forward strategy and the cut-set upper bound increases. Therefore the simple Gaussian compress-forward strategy fails to be universally-approximate for this network.

Therefore the set of relaying strategies that can be universally approximate for general noisy (Gaussian) relay networks has shrunk progressively through our examples. We devote the rest of the paper to generalizing the steps we took for each of the examples. As we will show, in the deterministic relay network the received signal at each signal level is just an equation of the message sent by the source, and the optimal strategy is to simply shuffle these received equations at each relay and forward them. This insight leads to a natural strategy for noisy (Gaussian) relay networks that we will analyze. The strategy for each relay is to (vector) quantize the received signal reference to a distortion of the noise power and then map these bits uniformly to a transmit Gaussian codeword. The main result of our paper is to show that such a scheme is indeed universally approximate for arbitrary noisy (Gaussian) relay networks for both single unicast and multicast information flows.

Chapter 4

Main results

4.1 Introduction

In this section we precisely state the main results of the paper and briefly discuss their implications. All the results we develop are lower bounds to the achievable rate for single unicast or multicast information flow over a relay network. The capacity of a relay network, C, is defined as the supremum of all achievable rates of reliable communication from the source to the destination. Similarly, the multicast capacity of relay network is defined as the maximum rate that the source can send the same information simultaneously to all destinations.

For any network, there is a natural information-theoretic cut-set bound [33], which upper bounds the reliable transmission rate R. Applied to the relay network, we have the cut-set upper bound $\overline{C}$ on its capacity:

$$\overline{C} = \max_{p(\{\mathbf{X}_j\}_{j \in \mathcal{V}})} \min_{\Omega \in \Lambda_D} I(\mathbf{Y}_{\Omega^c}; \mathbf{X}_\Omega | \mathbf{X}_{\Omega^c}) \tag{4.1}$$

where $\Lambda_D = \{\Omega : S \in \Omega, D \in \Omega^c\}$ is all source-destination cuts (partitions).

4.2 Deterministic networks

4.2.1 Linear finite-field deterministic relay network

Applying the cut-set bound to the linear finite field deterministic relay network defined in Section 2.5, (2.22), we get:

$$\overline{C} = \max_{p(\{\mathbf{X}_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} I(\mathbf{Y}_{\Omega^c};\mathbf{X}_\Omega|\mathbf{X}_{\Omega^c}) \tag{4.2}$$

$$\stackrel{(a)}{=} \max_{p(\{\mathbf{X}_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}) \tag{4.3}$$

$$\stackrel{(b)}{=} \min_{\Omega\in\Lambda_D} \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) \tag{4.4}$$

where $\Lambda_D = \{\Omega : S \in \Omega, D \in \Omega^c\}$ is all source-destination cuts (partitions) and $\mathbf{G}_{\Omega,\Omega^c}$ is the transfer matrix associated with that cut, *i.e.,* the matrix relating the vector of all the inputs at the nodes in Ω to the vector of all the outputs in Ω^c induced by (2.22). Step (a) follows since we are dealing with deterministic networks and step (b) follows since in a linear finite-field model all cut values (*i.e.* $H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c})$) are simultaneously optimized by independent and uniform distribution of $\{x_i\}_{i\in\mathcal{V}}$ and the optimum value of each cut Ω is logarithm of the size of the range space of the transfer matrix $\mathbf{G}_{\Omega,\Omega^c}$ associated with that cut.

The following are our main results for linear finite-field deterministic relay networks,

Theorem 4.2.1. *Given a linear finite-field relay network (with broadcast and multiple access), the capacity C of such a relay network is given by,*

$$C = \min_{\Omega\in\Lambda_D} \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}). \tag{4.5}$$

Theorem 4.2.2. *Given a linear finite-field relay network (with broadcast and multiple ac-*

cess), the multicast capacity C of such a relay network is given by,

$$C = \min_{D\in\mathcal{D}} \min_{\Omega\in\Lambda_D} \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}). \tag{4.6}$$

Note that the results in Theorems 4.2.1 and 4.2.2, applies to networks with arbitrary topology and could have cycles (or feedback loops). For a single source-destination pair the result in Theorem 4.2.1 generalizes the classical max-flow min-cut theorem for wireline networks and for multicast, the result in Theorem 4.2.2 generalizes the network coding result in [1] where in both these earlier results, the communication links are orthogonal, *i.e.* no broadcast or multiple access interference. Moreover, as we will see in the proof, the encoding functions at the relay nodes (for the linear finite-field model) could be restricted to linear functions to obtain the result in Theorem 4.2.1.

4.2.2 General deterministic relay network

In the general deterministic model the received vector signal $\mathbf{y}_j$ at node $j \in \mathcal{V}$ at time t is given by

$$\mathbf{y}_j[t] = \mathbf{g}_j(\{\mathbf{x}_i[t]\}_{i\in\mathcal{N}_j}), \tag{4.7}$$

where we define the input neighbors $\mathcal{N}_j$ of j as the set of nodes whose transmissions affect j, and can be formally defined as $\mathcal{N}_j = \{i : (i,j) \in \mathcal{E}\}$. Note that this implies a deterministic multiple access channel for node j and a deterministic broadcast channel for the transmitting nodes.

The following are our main results for arbitary networks with general deterministic interaction models.

Theorem 4.2.3. *Given an arbitrary relay network with general deterministic signal interaction model (with broadcast and multiple access), we can achieve all rates R up to,*

$$\max_{\prod_{i\in\mathcal{V}} p(\mathbf{X}_i)} \min_{\Omega\in\Lambda_D} H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}). \tag{4.8}$$

This theorem easily extends to the multicast case, where we want to simultaneously transmit one message from S to all destinations in the set $D \in \mathcal{D}$:

Theorem 4.2.4. *Given an arbitrary relay network with general deterministic signal interaction model (with broadcast and multiple access), we can achieve all rates R from S multicasting to all destinations $D \in \mathcal{D}$ up to,*

$$\max_{\prod_{i\in\mathcal{V}} p(\mathbf{X}_i)} \min_{D\in\mathcal{D}} \min_{\Omega\in\Lambda_D} H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}). \tag{4.9}$$

This achievability result in Theorem 4.2.3 extends the results in [9] where only deterministic broadcast network (with no interference) were considered. Note that when we compare (4.8) to the cut-set upper bound in (4.3), we see that the difference is in the maximizing set *i.e.,* we are only able to achieve independent (product) distributions whereas the cut-set optimization is over any arbitrary distribution. In particular, if the network and the deterministic functions are such that the cut-set is optimized by the product distribution, then we would have matching upper and lower bounds. This indeed happens when we consider the linear finite-field model. Hence, Theorems 4.2.1 and 4.2.2 are just corollaries of Theorems 4.2.3 and 4.2.4.

4.3 Gaussian relay networks

In the Gaussian model the signals get attenuated by complex gains and added together with Gaussian noise at each receiver (the Gaussian noises at different receivers being independent of each other.). More formally the received signal $\mathbf{y}_j$ at node $j \in \mathcal{V}$ and time t is given by

$$\mathbf{y}_j[t] = \sum_{i \in \mathcal{N}_j} \mathbf{H}_{ij}\mathbf{x}_i[t] + \mathbf{z}_j[t] \tag{4.10}$$

where $\mathbf{H}_{ij}$ is a complex matrix where element represents the channel gain from a transmitting antenna in node i to a receiving antenna in node j, and $\mathcal{N}_j$ is the set of nodes that are neighbors of j in $\mathcal{G}$ (*i.e.* all nodes that have a nonzero channel gain to j). Furthermore, we assume there is an average power constraint equal to 1 at each transmit antenna. Also $\mathbf{z}_j$, representing the channel noise, is modeled as complex normal (Gaussian) random vector, and hence the name Gaussian signal interaction model.

Other than the complex Gaussian model, in some cases we also look at the real Gaussian model. This model is the same as the complex one except the channel inputs, channel gains, and channel noises are restricted to be real numbers.

The following is our main result for noisy (Gaussian) relay networks which is proved in Chapter 6. This is perhaps the main result of the dissertation as it applies to wireless networks with realistic channel models and gives a universally-approximate characterization.

Theorem 4.3.1. *Given a Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which could have multiple transmit and receive antennas, we can achieve all rates R up to $\overline{C} - \kappa$. Therefore the capacity of this network satisfies*

$$\overline{C} - \kappa \leq C \leq \overline{C}, \tag{4.11}$$

where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in equation (4.1), and κ is a constant and is upper bounded by $5\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

The gap (κ) holds for all values of the channel gains and is relevant particularly in the high rate regime. This constant gap result is a far stronger result than the degree of freedom result, not only because it is non-asymptotic but also because it is uniform in the many channel SNR's. This is also the first constant gap approximation of the capacity of Gaussian relay networks. As shown in Section IV, the gap between the achievable rate of well known relaying schemes and the cut-set upper bound in general depends on the channel parameters and can become arbitrarily large. Analogous to the results for deterministic networks, the result in Theorem 4.3.1 applies to an network with arbitrary topology and could have cycles.

4.4 Extensions

We have also developed several extensions of the main results and these extensions are all proved in Chapter 7.

4.4.1 Compound relay network

The result in Theorem 4.3.1 can be extended to compound relay networks where we allow each channel gain $h_{i,j}$ to be from a set $\mathcal{H}_{i,j}$, and the particular chosen values are unknown to the source node S, the relays and the destination. A communication rate R is achievable if there exist a scheme such that for any channel gain realizations, still the source can communicate to the destination at rate R, without the knowledge of the channel realizations at the source, the relays and the destination. In this case we can obtain the following result which is proved in Section 7.2.

Theorem 4.4.1. *Given a compound Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the capacity C_{cn} satisfies*

$$\overline{C}_{cn} - \kappa \leq C_{cn} \leq \overline{C}_{cn} \tag{4.12}$$

Where $\overline{C}_{cn}$ is the cut-set upper bound on the compound capacity of $\mathcal{G}$ as described below

$$\overline{C}_{cn} = \max_{p(\{\mathbf{X}_i\}_{j\in\mathcal{V}})} \inf_{h\in\mathcal{H}} \min_{\Omega\in\Lambda_D} I(\mathbf{Y}_{\Omega^c}; \mathbf{X}_\Omega | \mathbf{X}_{\Omega^c}) \tag{4.13}$$

And κ is a constant and is upper bounded by $6\sum_{i=1}^{|\mathcal{V}|}\max(M_i,N_i)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

The implication of this result is two-fold. One is that we can develop strategies that are robust to channel uncertainties, which attains the compound channel rate supported by the network without relays explicitly knowing the channels. Secondly, this might be important in characterizing the diversity-multiplexing trade-off for fading relay network, since the compound framework gives a connection to the outage probability of the rate supported by the network.

4.4.2 Half-duplex relay network

In practical implementation of wireless networks an important consideration is the half-duplex constraint. This constraint implies that a node can not transmit and receive at the same time on the same frequency band. In that context, all the results stated above are applicable to full-duplex radios, which are capable of transmitting and receiving at the same time. A natural question is whether these results can be extended to radios with half-duplex constraint. We partially answer this question by approximately characterizing the capacity for any network with *fixed* duplexing times (transmission scheduling). This does not cover strategies that adapt the duplexing time to the situation. Here is our main result for half-duplex Gaussian relay networks

Theorem 4.4.2. *Given a Gaussian relay network with half-duplex constraint, $\mathcal{G}=(\mathcal{V},\mathcal{E})$, the capacity, C_{hd}, satisfies*

$$\overline{C}_{hd}-\kappa\leq C_{hd}\leq\overline{C}_{hd} \tag{4.14}$$

Where $\overline{C}_{hd}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ and is given by

$$C_{hd}\leq\overline{C}_{hd}=\max_{\substack{p(\{\mathbf{X}_j^m\}_{j\in\mathcal{V},m\in\{1,\dots,M\}})\\ t_m:\ 0\leq t_m\leq 1,\ \sum_{m=1}^{M}t_m=1}}\ \min_{\Omega\in\Lambda_D}\sum_{m=1}^{M}t_m I(\mathbf{Y}_{\Omega^c}^m;\mathbf{X}_\Omega^m|\mathbf{X}_{\Omega^c}^m) \tag{4.15}$$

where $m \in \{1, 2, \ldots, M\}$ denotes the operation mode of the network, defined as a valid partitioning of the nodes of the network into two sets of "sender" nodes and "receiver" nodes. For each node i, the transmit and the receives signal at mode m and at time t are respectively shown by $\mathbf{x}_i^m[t]$ and $\mathbf{y}_i^m[t]$. Also t_m defines the portion of the time that network will operate in state m, as the network use goes to infinity. Also κ is a constant and is upper bounded by $5\sum_{i=1}^{|\mathcal{V}|}\max(M_i, N_i)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

Note that in Theorem 4.4.2 we can optimize duplexing times (i.e. t_m's) to increase the achievable rate. It is an open question whether optimizing the duplexing time can capture all possible rates achievable by using adaptive strategies.

4.4.3 Frequency selective relay network

We also extend the result in Theorem 4.3.1 to frequency selective channels between nodes. For this case the result can be stated as follows

Theorem 4.4.3. *Given a frequency selective Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with F different frequency bands. The capacity of this network, C, satisfies*

$$\overline{C} - \kappa \leq C \leq \overline{C} \tag{4.16}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in equation (4.1), and κ is a constant and is upper bounded by $5\sum_{i=1}^{|\mathcal{V}|}\max(M_i, N_i)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

As we will discuss in Section 7.3, this can be implemented in particular by using OFDM and appropriate spectrum shaping or allocation.

4.4.4 Fading relay network

For time varying channels where the variation is slow in comparison to block length needed for a static channel (underspread regime) we can develop the approximate ergodic capacity of relay networks:

Theorem 4.4.4. *Given a fast fading quasi-static fading Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the ergodic capacity $C_{ergodic}$ satisfies*

$$\mathcal{E}_{h_{ij}}\left[\overline{C}(\{h_{ij}\})\right] - \kappa \leq C_{ergodic} \leq \mathcal{E}_{h_{ij}}\left[\overline{C}(\{h_{ij}\})\right] \tag{4.17}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity, as described in equation (4.1), and the expectation is taken over the channel gain distribution, and κ is a constant and is upper bounded by $5\sum_{i=1}^{|\mathcal{V}|}\max(M_i, N_i)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

4.4.5 Low rate capacity approximation of Gaussian relay network

Finally, we explore a multiplicative instead of additive approximation to capacity and show that such an approximate can also be universally obtained.

Theorem 4.4.5. *Given a Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the capacity C satisfies*

$$\lambda\overline{C} \leq C \leq \overline{C} \tag{4.18}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity, as described in equation (4.1), and λ is a constant and is lower bounded by $\frac{1}{2d(d+1)}$ and d is the maximum degree of nodes in $\mathcal{G}$.

Note that this kind of approximation might be of interest in a low data rate regime, where a constant gap approximation of the capacity may not be interesting any more.

4.5 Proof program

In Chapters 5-7 we formally prove these main results. The main proof program consists of first proving Theorem 4.2.3 and the corresponding multicast result. This immediately yields Theorems 4.2.1 and 4.2.2 which are a direct consequence of these results. The insight from these results suggest the quantize-map strategy for noisy (Gaussian) relay networks. We use this insight as well as proof ideas generated for the deterministic analysis to obtain the universally-approximate capacity characterization for Gaussian relay networks in Chapter 6. In both cases we illustrate the proof by going through an example which then is generalized.

Chapter 5

Deterministic relay networks

5.1 Introduction

In this chapter we focus on noiseless deterministic relay networks. Theorems 4.2.3 and 4.2.4 are our main result for deterministic relay networks and the rest of this chapter is devoted to proving it. First we focus on networks that have a layered structure, i.e. all paths from the source to the destination have equal lengths. With this special structure we get a major simplification: a sequence of messages can each be encoded into a block of symbols and the blocks do not interact with each other as they pass through the relay nodes in the network. The proof of the result for layered network is similar in style to the random coding argument in Ahlswede et. al. [1]. We do this in sections 5.2 and 5.3, first for the linear finite-field model and then for the general deterministic model. Next, we extend the result to an arbitrary network by expanding the network over time in such a way that while source encodes the message over multiple blocks, the relays operations are memoryless over different communication blocks. Since the time-expanded network is layered and we can apply our result in the first step to it. To complete the proof of the result, we need to establish a connection between the cut values of the time-expanded network and those of the original network. We do this using sub-modularity properties of entropy in Section 5.4[1].

5.2 Layered networks: linear finite-field deterministic model

The network given in Figure 5.1 is an example of a *layered* network where the number of "hops" for each path from S to D is equal to 3 in this case[2].

In this section we give the encoding scheme for the layered linear finite-field deterministic relay networks in Section 5.2.1. In Section 5.2.2 we illustrate the proof techniques on a simple linear unicast relay network example. In Section 5.2.3 we prove main Theorems 4.2.1 and 4.2.2 for layered networks.

5.2.1 Encoding for layered linear finite-field deterministic relay network

We have a single source S with a sequence of messages $w_k \in \{1, 2, \ldots, 2^{TR}\}$, $k = 1, 2, \ldots$. Each message is encoded by the source S into a signal over T transmission times (symbols), giving an overall transmission rate of R. Each relay operates over blocks of time T symbols, and uses a mapping $f_j : \mathcal{Y}_j^T \to \mathcal{X}_j^T$ its received symbols from the previous block of T symbols to transmit signals in the next block. For the model (2.22), we will use linear

[1]The concept of time-expanded network is also used in [1], but the use there is to handle cycles. Our main use is to handle interaction between messages transmitted at different times, an issue that only arises when there is superposition of signals at nodes.

[2]Note that in the equal path network we do not have "self-interference" since all path-lengths from S to D in terms of "hops" are equal, though as we will see in the analysis that can easily be taken care of. However we do allow for self-interference in the model and we choose to handle such loops, and more generally cyclic networks, through time-expansion as will be seen in Section 5.4.

mappings $f_j(\cdot)$, *i.e.,*

$$\mathbf{x}_j = \mathbf{F}_j \mathbf{y}_j, \tag{5.1}$$

where $\mathbf{F}_j$ is chosen uniformly randomly over all matrices in $\mathbb{F}_p^{qT \times qT}$. Each relay does the encoding prescribed by (5.1). Given the knowledge of all the encoding functions $\mathbf{F}_j$ at the relays, the decoder $D \in \mathcal{D}$, attempts to decode each message w_k sent by the source.

Now suppose message w_k is sent by the source in block k, then since each relay j operates only on block of lengths T and we have a layered structure, the signals received at block k at any relay pertain to only message w_{k-l_j} where l_j is the path length from source to relay j. As a result the key simplification that occurs for layered networks is that the messages do not get mixed with each other.

Now, given the knowledge of all the encoding functions $\mathbf{F}_j$ at the relays and signals received over block $k + l_D$, the decoder $D \in \mathcal{D}$, attempts to decode the message w_k sent by the source.

5.2.2 Proof illustration

In order to illustrate the proof ideas of Theorem (4.2.3) we examine the network shown in Figure 5.1. We will analyze this network first for linear deterministic model and then we use the same example to illustrate the ideas for general deterministic functions in Section 5.3.2.

Since we have a layered network, without loss of generality consider the message $w = w_1$ transmitted by the source at block $k = 1$. At node j the signals pertaining to this message are received by the relays at block l_j. For notational simplicity we will drop the block numbers associated with the transmitted and received signals for this analysis.

Now, since we have a deterministic network, the message w will be mistaken for another message w' is if the received signal $\mathbf{y}_D(w)$ under w, is the same as that would have been received under w'. This leads to a notion of *distinguishability*, which is that messages

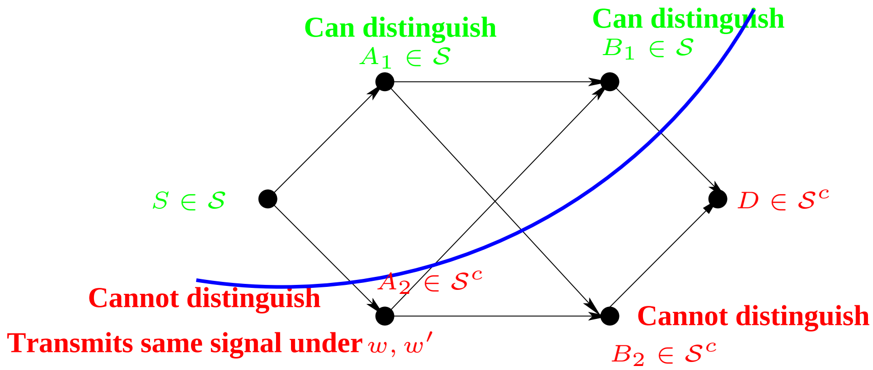


Figure 5.1: An example of layered relay network. Nodes on the left hand side of the cut can distinguish between messages w and w', while nodes on the right hand side can not.

w, w' are distinguishable at any node j if $\mathbf{y}_j(w) \neq \mathbf{y}_j(w')$.

The probability of error at decoder D can be upper bounded using the union bound as,

$$P_e \leq 2^{RT}\mathbb{P}\left\{w \to w'\right\} = 2^{RT}\mathbb{P}\left\{\mathbf{y}_D(w) = \mathbf{y}_D(w')\right\}. \tag{5.2}$$

Since channels are deterministic, this event is random only due to the randomness in the encoder map. Therefore, the probability of this event depends on the probability that we choose such an encoder map. Now, we can write,

$$\mathbb{P}\left\{w \to w'\right\} = \sum_{\Omega \in \Lambda_D} \underbrace{\mathbb{P}\left\{\text{Nodes in } \Omega \text{ can distinguish } w, w' \text{ and nodes in } \Omega^c \text{ cannot}\right\}}_{\mathcal{P}} \tag{5.3}$$

since the events that correspond to occurrence of the distinguishability sets $\Omega \in \Lambda_D$ are disjoint. Let us examine one term in the summation in (5.3). For the cut $\Omega = \{S, A_1, B_1\}$, a necessary condition for the distinguishability set to be this cut is that $\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')$, along with $\mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w')$ and $\mathbf{y}_D(w) = \mathbf{y}_D(w')$. Now we have

$$\begin{aligned}
\mathcal{P} &= \mathbb{P}\{\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w'), \mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w'), \mathbf{y}_D(w) = \mathbf{y}_D(w'), \\
&\quad , \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \mathbf{y}_{B_1}(w) \neq \mathbf{y}_{B_1}(w')\} && (5.4) \\
&= \mathbb{P}\{\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \times \mathbb{P}\{\mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w'), \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w') | \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \\
&\quad \times \mathbb{P}\{\mathbf{y}_D(w) = \mathbf{y}_D(w'), \mathbf{y}_{B_1}(w) \neq \mathbf{y}_{B_1}(w') | \mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w'), \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \\
&\quad , \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} && (5.5) \\
&\leq \mathbb{P}\{\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \times \mathbb{P}\{\mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w') | \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \\
&\quad \times \mathbb{P}\{\mathbf{y}_D(w) = \mathbf{y}_D(w') | \mathbf{y}_{B_1}(w) \neq \mathbf{y}_{B_1}(w'), \mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w'), \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \\
&\quad , \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} && (5.6) \\
&= \mathbb{P}\{\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \times \mathbb{P}\{\mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w') | \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \\
&\quad \times \mathbb{P}\{\mathbf{y}_D(w) = \mathbf{y}_D(w') | \mathbf{y}_{B_1}(w) \neq \mathbf{y}_{B_1}(w'), \mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w')\} && (5.7)
\end{aligned}$$

where the last step is true since there is an independent random mapping at each node and we have the following markov structure in the network

$$X_S \to (Y_{A_1}, Y_{A_2}) \to (Y_{B_1}, Y_{B_2}) \to Y_D \tag{5.8}$$

Now since the source does a random linear mapping of the message onto $\mathbf{x}_S(w)$, the probability that $\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')$ is given by,

$$\mathbb{P}\{\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} = \mathbb{P}\left\{(\mathbf{I}_T \otimes \mathbf{G}_{S,A_2})(\mathbf{x}_S(w) - \mathbf{x}_S(w')) = \mathbf{0}\right\} = p^{-T\text{rank}(\mathbf{G}_{S,A_2})}, \tag{5.9}$$

since the random mapping given in (5.1) induces independent uniformly distributed $\mathbf{x}_S(w)$, $\mathbf{x}_S(w')$. Here, $\otimes$ is the Kronecker matrix product[3]. Now, in order to analyze the second

[3]If A is an m-by-n matrix and B is a p-by-q matrix, then the Kronecker product $A \otimes B$ is the mp-by-nq

probability, we see that since $\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')$, $\mathbf{x}_{A_2}(w) = \mathbf{x}_{A_2}(w')$, *i.e.,* the *same* signal is sent under both w, w'. Also if $\mathbf{y}_{A_2}(w) \neq \mathbf{y}_{A_2}(w')$, then the random mapping given in (5.1) induces independent uniformly distributed $\mathbf{x}_{A_1}(w)$, $\mathbf{x}_{A_1}(w')$ Therefore, we get

$$\begin{aligned}&\mathbb{P}\{\mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w') | \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \\ &\quad = \mathbb{P}\left\{(\mathbf{I}_T \otimes \mathbf{G}_{A_1,B_2})(\mathbf{x}_{A_1}(w) - \mathbf{x}_{A_1}(w')) = \mathbf{0}\right\} = p^{-T\text{rank}(\mathbf{G}_{A_1,B_2})}. \quad (5.10)\end{aligned}$$

Similarly we get,

$$\begin{aligned}&\mathbb{P}\{\mathbf{y}_D(w) = \mathbf{y}_D(w') | \mathbf{y}_{B_1}(w) \neq \mathbf{y}_{B_1}(w'), \mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w')\} \\ &\quad = \mathbb{P}\left\{(\mathbf{I}_T \otimes \mathbf{G}_{B_1,D})(\mathbf{x}_{B_1}(w) - \mathbf{x}_{B_1}(w')) = \mathbf{0}\right\} \\ &\quad = p^{-T\text{rank}(\mathbf{G}_{B_1,D})}. \quad (5.11)\end{aligned}$$

Putting these together, since all three would need to occur, we see that in (5.3), for the network in Figure 5.1, we have,

$$\begin{aligned}\mathcal{P} &\leq p^{-T\text{rank}(\mathbf{G}_{S,A_2})} p^{-T\text{rank}(\mathbf{G}_{A_1,B_2})} p^{-T\text{rank}(\mathbf{G}_{B_1,D})} \\ &= p^{-T\{\text{rank}(\mathbf{G}_{S,A_2}) + \text{rank}(\mathbf{G}_{A_1,B_2}) + \text{rank}(\mathbf{G}_{B_1,D})\}}. \quad (5.12)\end{aligned}$$

Note that since,

$$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{G}_{S,A_2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{A_1,B_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_{B_1,D} \end{bmatrix},$$

the upper bound for $\mathcal{P}$ in (5.12) is exactly $2^{-T\text{rank}(\mathbf{G}_{\Omega,\Omega^c})}$. Therefore, by substituting this

block matrix $A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$

back into (5.3) and (5.2), we see that

$$P_e \leq 2^{RT}|\Lambda_D|p^{-T\min_{\Omega\in\Lambda_D}\text{rank}(\mathbf{G}_{\Omega,\Omega^c})}, \tag{5.13}$$

which can be made as small as desired if $R < \min_{\Omega\in\Lambda_D}\text{rank}(\mathbf{G}_{\Omega,\Omega^c})\log p$, which is the result claimed in Theorem 4.2.1.

Now we will prove the general result for layered linear finite-field deterministic relay networks.

5.2.3 Proof of main Theorems 4.2.1 and 4.2.2 for layered networks

In this section we prove main Theorems 4.2.1 and 4.2.2 for layered networks. Since we have a layered network, without loss of generality consider the message $w = w_1$ transmitted by the source at block $k = 1$. At node j the signals pertaining to this message are received by the relays at block l_j. We analyze a l_D-layer network, each layer is a MIMO subnetwork. Therefore, as in the analysis of (5.3), we see that

$$P_e^{(D)} \leq 2^{RT}\sum_{\Omega\in\Lambda_D}\underbrace{\mathbb{P}\left\{\text{Nodes in }\Omega\text{ can distinguish }w,w'\text{ and nodes in }\Omega^c\text{ cannot}\right\}}_{\mathcal{P}} \tag{5.14}$$

We define $\mathbf{G}_{\Omega,\Omega^c}$ as the transfer matrix associated with the nodes in Ω to the nodes in Ω^c. Note that since we have a layered network this transfer matrix breaks up into block diagonal elements corresponding to each of the l_D layers of the network. More precisely, we can create $d = l_D$ disjoint sub-networks of nodes corresponding to each layer of the network, with the set of nodes $\beta_l(\Omega)$, which are at distance $l-1$ from S and are in Ω, on one side and the set of nodes $\gamma_l(\Omega)$, which are at distance l from S that are in Ω^c, on the other side, for $l = 1, \ldots, l_D$.

Each node $i \in \beta_l(\Omega)$ sees a signal related to $w = w_1$ in block $l_i = l - 1$, and therefore waits to receive this block and then does a random mapping to $\mathbf{x}_i(w) \in \mathbb{F}_p^{qT}$ The random mapping is done as in (5.1), by choosing a random matrix $\mathbf{F}_i$ of size $Tq \times Tq$ and creating

$$\mathbf{x}_i(w) = \mathbf{F}_i \mathbf{y}_i(w) \tag{5.15}$$

The received signals in the nodes $j \in \gamma_l(\Omega)$ are linear transformations of the transmitted signals from nodes $\mathcal{T}_l = \{u : (u, v) \in \mathcal{E}, v \in \gamma_l(\Omega)\}$. That is, its output depends not only on the transmitters in $\beta_l(\Omega)$, but also other transmitters at distance $l - 1$ from S that are part of Ω^c. Since all the receivers in $\gamma_l(\Omega)$ are at distance l from S, they form the receivers of the layer l. Since we are focusing on one cut Ω, to simplify the notation we drop the parameter in Ω in $\beta_l(\Omega)$ and $\gamma_l(\Omega)$, and simply denote then by β_l and γ_l. Now similar to Section 5.2.2 we can write

$$\begin{aligned}
\mathcal{P} &= \mathbb{P}\{\mathbf{y}_{\gamma_l}(w) = \mathbf{y}_{\gamma_l}(w'), \mathbf{y}_{\beta_l}(w) \neq \mathbf{y}_{\beta_l}(w'), l = 1, \ldots, l_D\} && (5.16)\\
&= \prod_{l=1}^{l_D} \mathbb{P}\{\mathbf{y}_{\gamma_l}(w) = \mathbf{y}_{\gamma_l}(w'), \mathbf{y}_{\beta_l}(w) \neq \mathbf{y}_{\beta_l}(w') | \mathbf{y}_{\gamma_j}(w) = \mathbf{y}_{\gamma_j}(w'), \mathbf{y}_{\beta_j}(w) \neq \mathbf{y}_{\beta_j}(w'),\\
&\qquad , j = 1, \ldots, l-1\} && (5.17)\\
&\leq \prod_{l=1}^{l_D} \mathbb{P}\{\mathbf{y}_{\gamma_l}(w) = \mathbf{y}_{\gamma_l}(w') | \mathbf{y}_{\beta_l}(w) \neq \mathbf{y}_{\beta_l}(w'), \mathbf{y}_{\gamma_j}(w) = \mathbf{y}_{\gamma_j}(w'), \mathbf{y}_{\beta_j}(w) \neq \mathbf{y}_{\beta_j}(w'),\\
&\qquad , j = 1, \ldots, l-1\} && (5.18)\\
&= \prod_{l=1}^{l_D} \mathbb{P}\{\mathbf{y}_{\gamma_l}(w) = \mathbf{y}_{\gamma_l}(w') | \mathbf{y}_{\beta_l}(w) \neq \mathbf{y}_{\beta_l}(w'), \mathbf{y}_{\gamma_{l-1}}(w) = \mathbf{y}_{\gamma_{l-1}}(w')\} && (5.19)
\end{aligned}$$

where the last step is true since there is an independent random mapping at each node and we have a markovian layered structure in the network.

Now note that as in the example network of Section 5.2.2, for all the transmitting nodes in γ_{l-1} which cannot distinguish between w, w' the transmitted signal would be the same under both w and w'. Therefore, in order to calculate the probability that nodes in γ_l cannot distinguish between w, w' or that $\mathbf{y}_{\gamma_l}(w) - \mathbf{y}_{\gamma_l}(w') = \mathbf{0}$, we see that

$$\mathbf{y}_{\gamma_l}(w) - \mathbf{y}_{\gamma_l}(w') = \tilde{\mathbf{G}}_l \left[\mathbf{x}_{\beta_l}(w) - \mathbf{x}_{\beta_l}(w')\right], \;\; l = 1, \ldots, l_D \tag{5.20}$$

Due to the time-invariant channel conditions we see that $\tilde{\mathbf{G}}_l = \mathbf{I}_T \otimes \mathbf{G}_l$, where $\otimes$ is the Kronecker product.

Now, note that if the distinct signals $\mathbf{y}_i(w), \mathbf{y}_i(w')$ received at the nodes $i \in \beta_l$ could be *jointly uniformly and independently* mapped to the transmitted signals $\mathbf{u}_l(w), \mathbf{u}_l(w')$, then we could say that the probability of this occurrence is $\frac{\text{size of null space}}{\text{size of whole space}}$. Clearly this is given by,

$$\begin{aligned} \mathbb{P}\{\mathbf{y}_{\gamma_l}(w) = \mathbf{y}_{\gamma_l}(w') | \mathbf{y}_{\beta_l}(w) \neq \mathbf{y}_{\beta_l}(w'), \mathbf{y}_{\gamma_{l-1}}(w) = \mathbf{y}_{\gamma_{l-1}}(w')\} &= p^{-\text{rank}(\tilde{\mathbf{G}}_l)} \\ &= p^{-T\text{rank}(\mathbf{G}_l)} \end{aligned} \tag{5.21}$$

Not only the signals $\mathbf{y}_i(w)$ are uniformly randomly mapped *individually* at each node $i \in \beta_l$, the overall map across all nodes in β_l is also uniform, and hence the probability given in (5.21) is the correct one. Therefore we get

$$\mathcal{P} \leq p^{-T \sum_{l=1}^{d} \text{rank}(\mathbf{G}_l)}. \tag{5.22}$$

Now the probability of mistaking w for w' at receiver $D \in \mathcal{D}$ is therefore

$$\begin{aligned} \mathbb{P}\{w \to w'\} &\leq \sum_{\Omega \in \Lambda_D} p^{-T \sum_{l=1}^{d(\Omega)} \text{rank}(\mathbf{G}_l(\Omega))} \\ &\leq 2^{|\mathcal{V}|} p^{-T \min_{\Omega \in \Lambda} \text{rank}(\mathbf{G}_{\Omega, \Omega^c})}, \end{aligned}$$

where we have used $|\Lambda_D| \leq 2^{|\mathcal{V}|}$. Note that we have used the fact that since $\mathbf{G}_{\Omega,\Omega^c}$ was block diagonal, with blocks, $\mathbf{G}_l(\Omega)$, we see that $\sum_{l=1}^{d(\Omega)} \text{rank}(\mathbf{G}_l(\Omega)) = \text{rank}(\mathbf{G}_{\Omega,\Omega^c})$. If we declare an error if *any* receiver $D \in \mathcal{D}$ makes an error, we see that since we have 2^{RT} messages, from the union bound we can drive the error probability to zero if we have,

$$R < \min_{D\in\mathcal{D}} \min_{\Omega\in\Lambda_D} \text{rank}(\mathbf{G}_{\Omega,\Omega^c}) \log p. \tag{5.23}$$

Since as seen in Section 4.2.2, the cut-set is also identical to the expression in (5.23), we have proved the following result.

Theorem 5.2.1. *Given a layered (equal path) linear finite-field relay network (with broadcast and multiple access), the multicast capacity C of such a relay network is given by,*

$$C = \min_{D\in\mathcal{D}} \min_{\Omega\in\Lambda_D} \text{rank}(\mathbf{G}_{\Omega,\Omega^c}) \log p, \tag{5.24}$$

5.3 Layered networks: general deterministic model

In this section we prove main theorems 4.2.3 and 4.2.4 for layered networks. We first generalize the encoding scheme to accommodate arbitrary deterministic functions of (4.7) in Section 5.4.1. We then illustrate the ingredients of the proof using the same example as in Section 5.2.2. Then we prove the result for layered networks in Section 5.3.3.

5.3.1 Encoding for layered general deterministic relay network

We have a single source S with a sequence of messages $w_k \in \{1, 2, \ldots, 2^{TR}\}$, $k = 1, 2, \ldots$. Each message is encoded by the source S into a signal over T transmission times (symbols), giving an overall transmission rate of R. We will use strong (robust) typicality as defined in [34]. The notion of joint typicality is naturally extended from Definition 5.3.1.

Definition 5.3.1. *We define* $\underline{x} \in T_\delta$ *if*

$$|\nu_{\underline{x}}(x) - p(x)| \leq \delta p(x),$$

where $\nu_{\underline{x}}(x) = \frac{1}{T}|\{t : x_t = x\}|$, *is the empirical frequency.*

Each relay operates over blocks of time T symbols, and uses a mapping $f_j : \mathcal{Y}_j^T \rightarrow \mathcal{X}_j^T$ its received symbols from the previous block of T symbols to transmit signals in the next block. In particular, block k of T received symbols is denoted by $\mathbf{y}_j^{(k)} = \{y[(k-1)T+1], \dots, y[kT]\}$ and the transmit symbols by $\mathbf{x}_j^{(k)}$. Choose some product distribution $\prod_{i \in \mathcal{V}} p(x_i)$. At the source S, map each of the indices in $w_k \in \{1, 2, \dots, 2^{TR}\}$, choose $f_S(w_k)$ onto a sequence uniformly drawn from $T_\delta(X_S)$, which is the typical set of sequences in $\mathcal{X}_S^T$. At any relay node j choose f_j to map each typical sequence in $\mathcal{Y}_j^T$ *i.e.,* $T_\delta(Y_j)$ onto typical set of transmit sequences *i.e.,* $T_\delta(X_j)$, as

$$\mathbf{x}_j^{(k)} = f_j(\mathbf{y}_j^{(k-1)}), \tag{5.25}$$

where f_j is chosen to map uniformly randomly each sequence in $T_\delta(Y_j)$ onto $T_\delta(X_j)$. Each relay does the encoding prescribed by (5.25). Now, given the knowledge of all the encoding functions f_j at the relays and signals received over block $k + l_D$, the decoder $D \in \mathcal{D}$, attempts to decode the message w_k sent by the source.

5.3.2 Proof illustration

Now, we illustrate the ideas behind the proof of Theorem 4.2.3 for layered networks using the same example as in Section 5.2.2, which was done for the linear deterministic model. Since we are dealing with deterministic networks, the logic upto (5.3) in Section 5.2.2 remains the same. We will again illustrate the ideas using the cut $\Omega = \{S, A_1, B_1\}$. As in Section 5.2.2, we can write

$$\begin{aligned}
\mathcal{P} &= \mathbb{P}\{\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w'), \mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w'), \mathbf{y}_D(w) = \mathbf{y}_D(w'), \\
&\quad , \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \mathbf{y}_{B_1}(w) \neq \mathbf{y}_{B_1}(w')\} \quad (5.26) \\
&= \mathbb{P}\{\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \times \mathbb{P}\{\mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w'), \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w') | \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \\
&\quad \times \mathbb{P}\{\mathbf{y}_D(w) = \mathbf{y}_D(w'), \mathbf{y}_{B_1}(w) \neq \mathbf{y}_{B_1}(w') | \mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w'), \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \\
&\quad , \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \quad (5.27) \\
&\leq \mathbb{P}\{\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \times \mathbb{P}\{\mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w') | \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \\
&\quad \times \mathbb{P}\{\mathbf{y}_D(w) = \mathbf{y}_D(w') | \mathbf{y}_{B_1}(w) \neq \mathbf{y}_{B_1}(w'), \mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w'), \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \\
&\quad , \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \quad (5.28) \\
&= \mathbb{P}\{\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \times \mathbb{P}\{\mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w') | \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \\
&\quad \times \mathbb{P}\{\mathbf{y}_D(w) = \mathbf{y}_D(w') | \mathbf{y}_{B_1}(w) \neq \mathbf{y}_{B_1}(w'), \mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w')\} \quad (5.29)
\end{aligned}$$

where the last step is true since there is an independent random mapping at each node and we have a markovian layered structure in the network.

Notice that as in Section 5.2.2, we are suppressing the block numbers associated with the received signals. It is clear that for $w = w_1$, the block numbers associated with $\mathbf{y}_{A_2}, \mathbf{y}_{B_2}, \mathbf{y}_D$ are $1, 2, 3$ respectively.

Note that since $\mathbf{y}_j \in T_\delta(Y_j)$ with high probability, we can focus only on the typical received signals. Let us first examine the probability that $\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')$. Since S can distinguish between w, w', it maps these sub-messages independently to two transmit signals $\mathbf{x}_S(w), \mathbf{x}_S(w') \in T_\delta(X_S)$, hence we can see that this probability is,

$$\mathbb{P}\{\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} = \mathbb{P}\left\{(\mathbf{x}_S(w'), \mathbf{y}_{A_2}(w)) \in T_\delta(X_S, Y_{A_2})\right\} = 2^{-TI(X_S;Y_{A_2})}. \quad (5.30)$$

Now, in order to analyze the probability the second probability, as seen in the linear model analysis, we see that since $\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')$, $\mathbf{x}_{A_2}(w) = \mathbf{x}_{A_2}(w')$, *i.e.,* the *same* signal is sent under both w, w'. Therefore, since naturally $(\mathbf{x}_{A_2}(w), \mathbf{y}_{B_2}(w)) \in T_\delta(X_{A_2}, Y_{B_2})$, obviously, $(\mathbf{x}_{A_2}(w'), \mathbf{y}_{B_2}(w)) \in T_\delta(X_{A_2}, Y_{B_2})$ as well. Therefore, under w', we already have $\mathbf{x}_{A_2}(w')$ to be jointly typical with the signal that is received under w. However, since A_1 can distinguish between w, w', it will map the transmit sequence $\mathbf{x}_{A_1}(w')$ to a sequence which is independent of $\mathbf{x}_{A_1}(w)$ transmitted under w. Since an error occurs when $(\mathbf{x}_{A_1}(w'), \mathbf{x}_{A_2}(w'), \mathbf{y}_{B_2}(w)) \in T_\delta(X_{A_1}, X_{A_2}, Y_{B_2})$, and since A_2 cannot distinguish between w, w', we also have $\mathbf{x}_{A_2}(w) = \mathbf{x}_{A_2}(w')$, we require that $(\mathbf{x}_{A_1}, \mathbf{x}_{A_2}, \mathbf{y}_{B_2})$ generated like $p(\mathbf{x}_{A_1})p(\mathbf{x}_{A_2}, \mathbf{y}_{B_2})$ behaves like a jointly typical sequence. Therefore, this probability is given by,

$$\begin{aligned}
&\mathbb{P}\{\mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w') | \mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w'), \mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')\} \\
&= \mathbb{P}\left\{(\mathbf{x}_{A_1}(w'), \mathbf{x}_{A_2}(w), \mathbf{y}_{B_2}(w)) \in T_\delta(X_{A_1}, X_{A_2} Y_{B_2})\right\} \doteq \\
&2^{-TI(X_{A_1};Y_{B_2},X_{A_2})} \stackrel{(a)}{=} 2^{-TI(X_{A_1};Y_{B_2}|X_{A_2})},
\end{aligned} \tag{5.31}$$

where $\doteq$ indicates exponential equality (where we neglect subexponential constants), and (a) follows since we have generated the mappings f_j independently, it induces an independent distribution on X_{A_1}, X_{A_2}. Another way to see this is that the probability of (5.31) is given by $\frac{|T_\delta(\mathbf{X}_{A_1}|\mathbf{x}_{A_2},\mathbf{y}_{B_2})|}{|T_\delta(\mathbf{X}_{A_1})|}$, which by using properties of (robustly) typical sequences [34] yields the same expression as in (5.31). Note that the calculation in (5.31) is similar to one of the error event calculations in a multiple access channel,

Using a similar logic we can write,

$$\begin{aligned}
&\mathbb{P}\{\mathbf{y}_D(w) = \mathbf{y}_D(w') | \mathbf{y}_{B_1}(w) \neq \mathbf{y}_{B_1}(w'), \mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w')\} \\
&= \mathbb{P}\left\{(\mathbf{x}_{B_1}(w'), \mathbf{x}_{B_2}(w), \mathbf{y}_D(w)) \in T_\delta(X_{B_1}, X_{B_2} Y_D)\right\} \doteq \\
&2^{-TI(X_{B_1};Y_D,X_{B_2})} \stackrel{(a)}{=} 2^{-TI(X_{B_1};Y_D|X_{B_2})}.
\end{aligned} \tag{5.32}$$

Therefore, putting (5.30)–(5.32) together as done in (5.12) we get

$$\mathcal{P} \leq 2^{-T\{I(X_S;Y_{A_2})+I(X_{A_1};Y_{B_2}|X_{A_2})+I(X_{B_1};Y_D|X_{B_2})\}}$$

Note that for this example, due to the Markovian structure of the network we can see that[4] $I(Y_{\Omega^c};X_\Omega|X_{\Omega^c}) = I(X_S;Y_{A_2}) + I(X_{A_1};Y_{B_2}|X_{A_2}) + I(X_{B_1};Y_D|X_{B_2})$, hence as in (5.13) we get that,

$$P_e \leq 2^{RT}|\Lambda_D|2^{-T\min_{\Omega\in\Lambda_D} I(Y_{\Omega^c};X_\Omega|X_{\Omega^c})}, \tag{5.33}$$

and hence the error probability can be made arbitrarily small if $R < \min_{\Omega\in\Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c})$.

5.3.3 Proof of main Theorems 4.2.3 and 4.2.4 for layered networks

As in the example illustrating the proof in Section 5.3.2, the logic of the proof in the general deterministic functions follows that of the linear model quite closely. In particular, as in Section 5.2 we can define the bi-partite network associated with a cut Ω. Instead of a transfer matrix $\mathbf{G}_{\Omega,\Omega^c}(\cdot)$ associated with the cut, we have a transfer function $\tilde{\mathbf{G}}_\Omega$. Since we are still dealing with a layered network, as in the linear model case, this transfer function breaks up into components corresponding to each of the l_D layers of the network. More precisely, we can create $d = l_D$ disjoint sub-networks of nodes corresponding to each layer of the network, with the set of nodes $\beta_l(\Omega)$, which are at distance $l-1$ from S and are in Ω, on one side and the set of nodes $\gamma_l(\Omega)$, which are at distance l from S that are in Ω^c, on the other side, for $l = 1, \ldots, l_D$. Each of these clusters have a transfer function $\mathbf{G}_l(\cdot), l = 1, \ldots, l_D$ associated with them.

[4]Note that though in the encoding scheme there is a dependence between $X_{A_1}, X_{A_2}, X_{B_1}, X_{B_2}$ and X_S, in the single-letter form of the mutual information, under a product distribution, $X_{A_1}, X_{A_2}, X_{B_1}, X_{B_2}, X_S$ are independent of each other. Therefore for example, Y_{B_2} is independent of X_{B_2} leading to $H(Y_{B_2}|X_{A_2}, X_{B_2}) = H(Y_{B_2}|X_{A_2})$. Using this argument for the cut-set expression $I(Y_{\Omega^c};X_\Omega|X_{\Omega^c})$, we get the expansion.

As in the linear model, each node $i \in \beta_l(\Omega)$ sees a signal related to $w = w_1$ in block $l_i = l - 1$, and therefore waits to receive this block and then does a mapping using the general encoding function given in (5.25) as

$$\mathbf{x}_j^{(k)}(w) = f_j^{(k)}(\mathbf{y}_j^{(k-1)}(w)). \tag{5.34}$$

The received signals in the nodes $j \in \gamma_l(\Omega)$ are deterministic transformations of the transmitted signals from nodes $\mathcal{T}_l = \{u : (u, v) \in \mathcal{E}, v \in \gamma_l(\Omega)\}$. As in the linear model analysis of Section 5.2, the dependence is on all the transmitting signals at distance $l - 1$ from the source, not just the ones in $\beta_l(\Omega)$. Since all the receivers in $\gamma_l(\Omega)$ are at distance l from S, they form the receivers of the layer l. Now similar to Section 5.3.2 we can write

$$\begin{aligned}
\mathcal{P} &= \mathbb{P}\{\mathbf{y}_{\gamma_l}(w) = \mathbf{y}_{\gamma_l}(w'), \mathbf{y}_{\beta_l}(w) \neq \mathbf{y}_{\beta_l}(w'), l = 1, \ldots, l_D\} & (5.35)\\
&= \prod_{l=1}^{l_D} \mathbb{P}\{\mathbf{y}_{\gamma_l}(w) = \mathbf{y}_{\gamma_l}(w'), \mathbf{y}_{\beta_l}(w) \neq \mathbf{y}_{\beta_l}(w') | \mathbf{y}_{\gamma_j}(w) = \mathbf{y}_{\gamma_j}(w'), \mathbf{y}_{\beta_j}(w) \neq \mathbf{y}_{\beta_j}(w'), \\
&\quad , j = 1, \ldots, l-1\} & (5.36)\\
&\leq \prod_{l=1}^{l_D} \mathbb{P}\{\mathbf{y}_{\gamma_l}(w) = \mathbf{y}_{\gamma_l}(w') | \mathbf{y}_{\beta_l}(w) \neq \mathbf{y}_{\beta_l}(w'), \mathbf{y}_{\gamma_j}(w) = \mathbf{y}_{\gamma_j}(w'), \mathbf{y}_{\beta_j}(w) \neq \mathbf{y}_{\beta_j}(w'), \\
&\quad , j = 1, \ldots, l-1\} & (5.37)\\
&= \prod_{l=1}^{l_D} \mathbb{P}\{\mathbf{y}_{\gamma_l}(w) = \mathbf{y}_{\gamma_l}(w') | \mathbf{y}_{\beta_l}(w) \neq \mathbf{y}_{\beta_l}(w'), \mathbf{y}_{\gamma_{l-1}}(w) = \mathbf{y}_{\gamma_{l-1}}(w')\} & (5.38)
\end{aligned}$$

Note that as in the example network of Section 5.3.2, for all the transmitting nodes in γ_{l-1} which cannot distinguish between w, w' the transmitted signal would be the same under both w and w'. Therefore, all the nodes in γ_{l-1} cannot distinguish between w, w' and therefore

$$\mathbf{x}_j(w) = \mathbf{x}_j(w'), \ \ j \in \gamma_{l-1}.$$

Hence it is clear that since $(\{\mathbf{x}_j(w)\}_{j\in\gamma_{l-1}}, \mathbf{y}_{\gamma_l}(w)) \in T_\delta$, we have that

$$(\{\mathbf{x}_j(w')\}_{j\in\gamma_{l-1}}, \mathbf{y}_{\gamma_l}(w)) \in T_\delta.$$

Therefore, just as in Section 5.3.2, we see that the probability that

$$\begin{aligned}
&\mathbb{P}\{\mathbf{y}_{\gamma_l}(w) = \mathbf{y}_{\gamma_l}(w')|\mathbf{y}_{\beta_l}(w) \neq \mathbf{y}_{\beta_l}(w'), \mathbf{y}_{\gamma_{l-1}}(w) = \mathbf{y}_{\gamma_{l-1}}(w')\} \\
&= \mathbb{P}\left\{(\mathbf{x}_{\beta_l}(w'), \mathbf{x}_{\gamma_{l-1}}(w), \mathbf{y}_{\gamma_l}(w)) \in T_\delta(X_{\beta_l}, X_{\gamma_{l-1}}, Y_{\gamma_l})\right\} \\
&\doteq 2^{-TI(X_{\beta_l};Y_{\gamma_l}|X_{\gamma_{l-1}})}.
\end{aligned} \tag{5.39}$$

Therefore we get

$$\mathcal{P} \leq \prod_{l=1}^{d} 2^{-TI(X_{\beta_l};Y_{\gamma_l}|X_{\gamma_{l-1}})} = 2^{-T\sum_{l=1}^{d} H(Y_{\gamma_l}|X_{\gamma_{l-1}})}. \tag{5.40}$$

Note that due to the Markovian nature of the layered network, we have

$$\sum_{l=1}^{d} H(Y_{\gamma_l}|X_{\gamma_{l-1}}) = H(Y_{\Omega^c}|X_{\Omega^c}) \tag{5.41}$$

From this point onwards the proof closely follows the steps as in the linear model from (5.22) onwards. Similarly in multicast scenario we declare an error if *any* receiver $D \in \mathcal{D}$ makes an error, we see that since we have 2^{RT} messages, from the union bound we can drive the error probability to zero if we have,

$$R < \max_{\prod_{i\in\mathcal{V}} p(x_i)} \min_{D\in\mathcal{D}} \min_{\Omega\in\Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}). \tag{5.42}$$

Therefore we have proved the following result.

Theorem 5.3.2. *Given a layered (equal path) general deterministic relay network (with broadcast and multiple access), we can achieve any rate R from S multicasting to all destinations $D \in \mathcal{D}$, with R satisfying:*

$$R < \max_{\prod_{i\in\mathcal{V}} p(x_i)} \min_{D\in\mathcal{D}} \min_{\Omega\in\Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) \tag{5.43}$$

5.4 Arbitrary networks

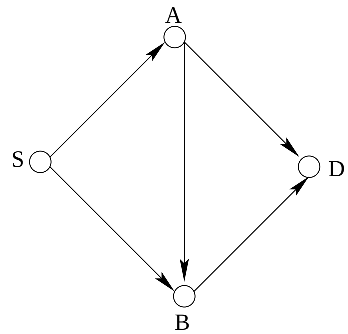
First we formally describe the encoding strategy:

5.4.1 Encoding for general deterministic relay network

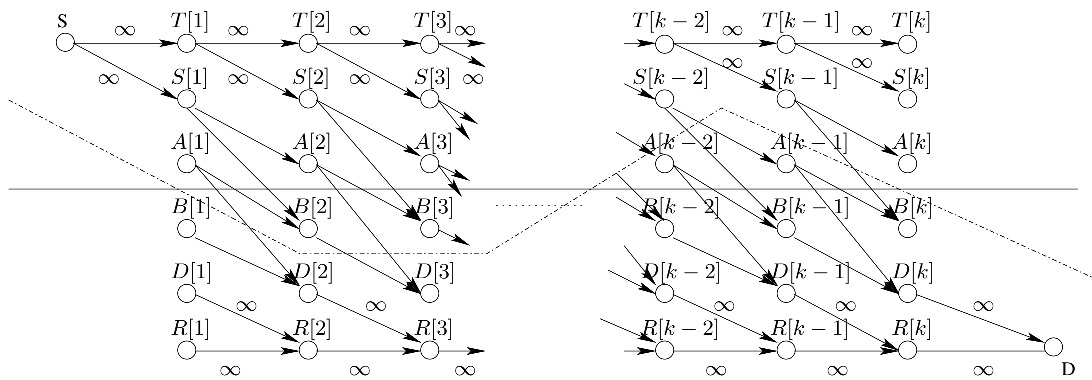
We have a single source S with message $W \in \{1, 2, \ldots, 2^{TKR}\}$ which is encoded by the source S into a signal over KT transmission times (symbols), giving an overall transmission rate of R. Each relay operates over blocks of time T symbols, and uses a mapping $f_j^{[t]} : \mathcal{Y}_j^T \rightarrow \mathcal{X}_j^T$ its received symbols from the previous block of T symbols to transmit signals in the next block. In particular, block k of T received symbols is denoted by $\mathbf{y}_j^{(k)} = \{y^{[(k-1)T+1]}, \ldots, y^{[kT]}\}$ and the transmit symbols by $\mathbf{x}_j^{(k)}$. Choose some product distribution $\prod_{i\in\mathcal{V}} p(x_i)$. At the source S, map each of the indices in $W \in \{1, 2, \ldots, 2^{TKR}\}$ choose $f_S^{(k)}(W)$ onto a sequence uniformly drawn from $T_\delta(X_S)$, which is the typical set of sequences in $\mathcal{X}_S^T$. At any relay node j choose $f_j^{(k)}$ to map each typical sequence in $\mathcal{Y}_j^T$ *i.e.,* $T_\delta(Y_j)$ onto typical set of transmit sequences *i.e.,* $T_\delta(X_j)$, as

$$\mathbf{x}_j^{(k)} = f_j^{(k)}(\mathbf{y}_j^{(k-1)}), \tag{5.44}$$

where $f_j^{(k)}$ is chosen to map uniformly randomly each sequence in $T_\delta(Y_j)$ onto $T_\delta(X_j)$ and is done independently for each block k. Each relay does the encoding prescribed by (5.44). Given the knowledge of all the encoding functions $f_j^{(k)}$ at the relays and signals received



(a) An example of general deterministic network



(b) Unfolded deterministic network. An example of steady cuts and wiggling cuts are respectively shown by solid and dotted lines.

Figure 5.2: An example of a general deterministic network with un equal paths from S to D is shown in (a). The corresponding unfolded network is shown in (b).

over $K + |\mathcal{V}| - 2$ blocks, the decoder $D \in \mathcal{D}$, attempts to decode the message W sent by the source.

Given the proof for layered networks with equal path lengths, we are ready to tackle the proof of Theorem 4.2.3 and Theorem 4.2.4 for general relay networks. The ingredients are developed below. First is that we can explicitly represent our relaying scheme by unfolded the network over time to create a layered deterministic network. The idea is to unfold the network to K stages such that i-th stage is representing what happens in the network during $(i-1)T$ to $iT-1$ symbol times. For example in figure 5.2 (a) a network with unequal paths from S to D is shown. Figure 5.2(b) shows the unfolded form of this network. As we notice each node $v \in \mathcal{V}$ is appearing at stage $1 \leq i \leq K$ as $v[i]$. There are additional nodes: $T[i]$'s and $R[i]$'s. These nodes are just virtual transmitters and receivers that are put to buffer and synchronize the network. Since all communication links connected to these nodes ($T[i]$'s and $R[i]$'s) are modelled as wireline links without any capacity limit they would not impose any constraint on the network. One should notice that in general there must be an infinite capacity link between the same node and itself appearing at different times however, since in the relaying scheme that we described above, the relays are limited to have a finite memory T, these links are omitted [5]. Now we show the following lemma,

Lemma 5.4.1. *Assume $\mathcal{G}$ is a general deterministic network and $\mathcal{G}_{unf}^{(K)}$ is a network obtained by unfolding $\mathcal{G}$ over K time steps (as shown in figure 5.2). Then the following communication rate is achievable in $\mathcal{G}$:*

$$R < \frac{1}{K} \max_{\prod_{i \in \mathcal{V}} p(x_i)} \min_{\Omega_{unf} \in \Lambda_D} H(Y_{\Omega_{unf}^c} | X_{\Omega_{unf}^c}) \tag{5.45}$$

where the minimum is taken over all cuts Ω_{unf} in $\mathcal{G}_{unf}^{(K)}$.

Proof. By unfolding $\mathcal{G}$ we get an acyclic layered deterministic network. Therefore by

[5] This idea was introduced for graphs in [1] to handle cycles in a graph

theorem 5.3.2 we can achieve the rate

$$R_{\text{unf}} < \max_{\prod_{i\in\mathcal{V}} p(x_i)} \min_{\Omega_{\text{unf}}\in\Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) \tag{5.46}$$

in the time-expanded graph. Since it takes K steps to translate and achievable scheme in the time-expanded graph to an achievable scheme in the original graph, then the Lemma is proved. □

Note that the achievability scheme that we used to prove Lemma 5.4.1 was obtained by applying the encoding scheme described in section 5.3.1 to the network that is unfolded over K blocks. This translates to the encoding scheme defined in Section 5.4.1 for a general deterministic relay network.

5.4.2 Proof of main Theorems 4.2.3 and 4.2.4

If we look at different cuts in the time-expanded graph we notice that there are two types of cuts. One type separates the nodes at different stages identically. An example of such a steady cut is drawn with solid line in figure 5.2 (b) which separates $\{S, A\}$ from $\{B, D\}$ at all stages. Clearly each steady cut in the time-expanded graph corresponds to a cut in the original graph and moreover its value is K times the value of the corresponding cut in the original network. However there is another type of cut which does not behave identically at different stages. An example of such a wiggling cut is drawn with dotted line in figure 5.2 (b). There is no correspondence between these cuts and the cuts in the original network.

Now comparing Lemma 5.4.1 to the main theorem 4.2.3 we want to prove, we notice that in this Lemma the achievable rate is found by taking the minimum of cut-values over all cuts in the time-expanded graph (steady and wiggling ones). However in theorem 4.2.3 we want to prove that we can achieve a rate by taking the minimum of cut-values over only the cuts in the original graph or similarly over the steady cuts in the time-expanded network. So a natural question is that in a time-expanded network does it make any difference if we take the minimum of cut-values over only steady cuts rather than all cuts ? Quite interestingly we show in the following Lemma that asymptotically as $K \to \infty$ this difference (normalized by $1/K$) vanishes.

Lemma 5.4.2. *Consider a general deterministic network, $\mathcal{G}$. Assume a product distribution on $\{x_i\}_{i\in\mathcal{V}}$, $p(\{x_i\}_{i\in\mathcal{V}}) = \prod_{i\in\mathcal{V}} p(x_i)$. Now in the time-expanded graph, $\mathcal{G}_{unf}^{(K)}$, assume that for each node $i \in \mathcal{V}$, $\{x_i[t]\}_{1\leq t\leq K}$ are distributed i.i.d. according to $p(x_i)$ in the original network. Also for any $1 \leq t_1, t_2 \leq K$ and $i \neq j$, $x_i[t_1]$ is independent of $x_j[t_2]$. Then for any cut Ω_{unf} on the unfolded graph we have,*

$$(K - L + 1) \min_{\Omega\in\Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) \leq H(Y_{\Omega^c_{unf}}|X_{\Omega^c_{unf}}) \tag{5.47}$$

where $L = 2^{|\mathcal{V}|-2}$.

Proof. See Appendix A.3. $\square$

Now since for any product distribution

$$\min_{\Omega_{\text{unf}}\in\Lambda_D} H(Y_{\Omega^c_{\text{unf}}}|X_{\Omega^c_{\text{unf}}}) \leq K \min_{\Omega\in\Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) \tag{5.48}$$

we have an immediate corollary of this lemma

Corollary 5.4.3. *Assume $\mathcal{G}$ is a general deterministic network and $\mathcal{G}_{unf}^{(K)}$ is a network obtained by unfolding $\mathcal{G}$ over K time steps then*

$$\begin{aligned} &\lim_{K\to\infty} \tfrac{1}{K} \max_{\prod_{i\in\mathcal{V}} p(x_i)} \min_{\Omega_{unf}\in\Lambda_D} H(Y_{\Omega^c_{unf}}|X_{\Omega^c_{unf}}) \\ &= \max_{\prod_{i\in\mathcal{V}} p(x_i)} \min_{\Omega\in\Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) \end{aligned} \tag{5.49}$$

Now by Lemma 5.4.1 and corollary 5.4.3, the proof of main Theorem 4.2.3 is complete.

5.5 Summary

In this chapter we focused on noiseless deterministic relay networks and we proved Theorems 4.2.3 and 4.2.4, which are lower bounds to the achievable rate for single unicast or multicast information flow over general deterministic relay networks. We proved this result by first focusing on networks that have a layered structure. Next, we extended the result to an arbitrary network by considering its time-expanded representation and establishing a connection between the cut values of the time-expanded network and those of the original network. As a corollary, this result yields the complete characterization of the unicast and multicast capacity of linear finite-field deterministic relay networks.

Chapter 6

Gaussian relay networks

6.1 Introduction

So far, we have focused on noiseless deterministic relay networks. As we discussed in Chapters 2 and 3, our linear finite field deterministic model in some sense captures the high SNR behavior approximation of the Gaussian model, therefore we expect to be able to lift up the intuitions and the results obtained so far and translate them to approximate results for the noisy Gaussian relay networks.

Theorem 4.3.1 is our main result for Gaussian relay networks and the rest of this chapter is devoted to prove it. Similar to the deterministic case, first we focus on networks that have a layered structure that the messages do not get mixed in the network. The proof of the result for layered network is done in section 6.2. Next, we extend the result to an arbitrary network by expanding the network over time, as done in Chapter 5. Since the time-expanded network is layered and we can apply our result in the first step to it. To complete the proof of the result, we need to establish a connection between the cut values of the time-expanded network and those of the original network. We do this using submodularity properties of entropy.

We first prove the theorem for the single antenna case, then at the end we extend it to a multiple antenna scenario.

6.2 Layered Gaussian relay networks

In this section we prove main theorem 4.3.1 for a special case of layered networks, where all paths from the source to the destination in $\mathcal{G}$ have equal length. We start by describing the relaying strategy.

6.2.1 Encoding for layered Gaussian relay networks

We have a single source S with a sequence of messages $w_k \in \{1, 2, \dots, 2^{TR}\}$, $k = 1, 2, \dots$. Each message is encoded by the source S into a signal over T transmission times (symbols), giving an overall transmission rate of R.

At each node we create a random Gaussian codebook. Source randomly maps each message to one of its Gaussian codewords and sends it in T transmission times. Now each relay operates over blocks of time T symbols. In particular block k of T received symbols at node j is denoted by $\mathbf{y}_j^{(k)} = \{y_j[(k-1)T+1], \dots, y_j[kT]\}$ and the transmit symbols by $\mathbf{x}_j^{(k)}$. Now the relaying strategy is the following: each received sequence $\mathbf{y}_j^{(k)}$ at node j is first quantized into $\hat{\mathbf{y}}_j^{(k)}$ which is then randomly mapped into a Gaussian codeword $\mathbf{x}_j^{(k)}$ using a random mapping function $f_j(\hat{\mathbf{y}}_j^{(k)})$. For quantization, we use an optimal Gaussian vector quantizer with distortion equal to 1 (which is equal to the noise variance of the channel).

Now, given the knowledge of all the encoding functions f_j at the relays and signals received over block $k + l_D$, the decoder $D \in \mathcal{D}$, attempts to decode the message w_k sent by the source.

6.2.2 Proof illustration

Consider the encoding-decoding strategy as described in section 6.2.1. Our goal is to show that, using this strategy, all rates described in the theorem are achievable. Similar to the deterministic case, we use the *distinguishability* argument.

In the deterministic model each message is mapped to a deterministic sequence of transmit codewords through the network. The destination can not distinguish between two messages if and only if its received signal under these two messages are identical. If so, there would be a partition of nodes in the network such that the nodes on one side of the cut can distinguish between these two messages and the rest can not. This naturally corresponds to a cut separating the source and the destination in the network and the probability that this happens can be related to the cut-value. However, in the noisy case, the difference from the previous analysis is that each message is potentially mapped to a *set* of possible transmit sequences. The particular transmit sequence chosen depends on the noise realization, which can be considered "typical". Pictorially it means that there is some fuzziness around the sequence of transmit codewords associated with each message. Hence, two messages will still be distinguishable at a node if the fuzzy received signal associated with them are not overlapping. This has two different consequences: first, there will be more possibilities of confusing a message at the destination. Second, if a node can not distinguish between two different messages, it will not necessarily transmit the same sequence under those different messages. Intuitively, if we can somehow bound the list sizes corresponding to different messages, we will be able to bound the effect of this extra randomness and achieve a communicate rate close to the cut-set bound.

In order to illustrate the proof ideas of Theorem (4.3.1) we examine the network shown in Figure 5.1.

Assume a message w is transmitted by the source. Once the destination receives $\mathbf{y}_D$, quantizes it to get $\hat{\mathbf{y}}_D$. Then, it will decode the message by finding the unique message that is jointly typical with $\hat{\mathbf{y}}_D$ (the precise definition of typicality will be given later). An error occurs if either w is not jointly typical with $\hat{\mathbf{y}}_D$ or there is another message w' such that $\hat{\mathbf{y}}_D$ is jointly typical with *both* w, w'.

Now for the relay network, a natural way to define whether a message w is typical with a received sequence is whether we have a "plausible" transmit sequence[1] under w which is jointly typical with the received sequence. More formally, we have the following definitions.

At the source node, since there is only one transmit codeword associated with each message, the set of transmitted sequences that are typical with a message w are

$$\mathcal{X}_S(w) = \{\mathbf{x}_S(w)\} \tag{6.1}$$

Now inductively we define

Definition 6.2.1. *At each node i, we define $(\hat{\mathbf{y}}_i, w) \in T_\delta$ if*

$$(\hat{\mathbf{y}}_i, \{\mathbf{x}_j\}_{j \in In(i)}) \in T_\delta \text{ for some } \mathbf{x}_j \in \mathcal{X}_j(w),\ \forall j \in In(i) \tag{6.2}$$

where $In(i)$ is defined as the set of nodes with signals incident on node i.

Definition 6.2.2. *At each node i, we define the set of received sequences that are typical with a message w as,*

$$\mathcal{Y}_i(w) = \{\hat{\mathbf{y}}_i : (\hat{\mathbf{y}}_i, w) \in T_\delta\}, \tag{6.3}$$

Finally we define,

Definition 6.2.3. *At each node i, we define the set of transmitted sequences that are typical*

[1] Plausibility essentially means that the transmit sequence is a member of the typical set of possible transmit sequences under w.

with a message w as,

$$\mathcal{X}_i(w) = \{\mathbf{x}_i : \mathbf{x}_i = f_i(\hat{\mathbf{y}}_i), \hat{\mathbf{y}}_i \in \mathcal{Y}_i(w)\}, \tag{6.4}$$

which defines the "typical" transmit set associated with a message w.

Note here that since $\mathbf{x}_i = f_i(\hat{\mathbf{y}}_i)$, then naturally $(\mathbf{x}_i, \hat{\mathbf{y}}_i) \in T_\delta$.

Therefore, we see that if a message w is typical with a received sequence, we have a sequence of typical transmit sequences in the network that are jointly typical with the w and the received sequence at the destination.

Now note the following important observation,

Observation Note that if node i cannot distinguish between two messages w, w', this means that the signal received at node i, $\hat{\mathbf{y}}_i$ is such that $(\hat{\mathbf{y}}_i, w) \in T_\delta$ and $(\hat{\mathbf{y}}_i, w') \in T_\delta$. Therefore we see that

$$\hat{\mathbf{y}}_i \in \mathcal{Y}_i(w) \cap \mathcal{Y}_i(w'). \tag{6.5}$$

Due to the mapping $\mathbf{x}_i = f_i(\hat{\mathbf{y}}_i)$, we therefore see that $\mathbf{x}_i \in \mathcal{X}_i(w) \cap \mathcal{X}_i(w')$. Therefore, there exists a sequence under w' which is the same as that transmitted under w and could therefore have been potentially transmitted under w'.

Now, assuming a message w is transmitted by the source, an error occurs at the destination if either w is not jointly typical with $\hat{\mathbf{y}}_D$, or there is another message w' such that $\hat{\mathbf{y}}_D$ is jointly typical with *both* w, w'. By the law of large numbers, the probability of the first event becomes arbitrarily small as communication block length, T, goes to infinity. So we just need to analyze the probability of the second event. To do so, we evaluate the probability that $\hat{\mathbf{y}}_D$ is jointly typical with both w and w', where w' is another message independent of w. Then we use union bound over all w''s to bound the probability of the second event.

Based on our earlier observation, if $\hat{\mathbf{y}}_D$ is jointly typical with w, w', then there must be a typical transmit sequence $\mathbf{x}'_{\mathcal{V}} = (\mathbf{x}'_S, \mathbf{x}'_{A_1}, \mathbf{x}'_{A_2}, \mathbf{x}'_{B_1}, \mathbf{x}'_{B_2})$ under w' such that,

$$(\hat{\mathbf{y}}_D, \mathbf{x}'_{B_1}, \mathbf{x}'_{B_2}) \in T_\delta \tag{6.6}$$

This means that the destination thinks this is a plausible sequence. Now for any such sequence there is a natural cut, Ω, in $\mathcal{G}$ such that the nodes on the right hand side of the cut (*i.e.* in Ω) can tell $\mathbf{x}'_{\mathcal{V}}$ is not a plausible sequence, and those on the left hand side of the cut (*i.e.* in Ω^c) can not. Therefore we can write

$$\mathbb{P}\left\{w \to w'\right\} = \mathbb{P}\left\{(\hat{\mathbf{y}}_D, w') \in T_\delta\right\} \leq \sum_{\mathbf{x}'_{\mathcal{V}} \in \mathcal{X}_{\mathcal{V}}(w')} \sum_{\Omega \in \Lambda_D} \underbrace{\mathbb{P}\left\{\text{Nodes in } \Omega \text{ tell } \mathbf{x}'_{\mathcal{V}} \text{ is not plausible \& nodes in } \Omega^c \text{ tell } \mathbf{x}'_{\mathcal{V}} \text{ is plausible}\right\}}_{\mathcal{P}} \tag{6.7}$$

For now, assume that the cut is $\Omega = \{S, A_1, B_1\}$, as shown in figure 5.1. Since A_2, B_2 and D think $\mathbf{x}'_{\mathcal{V}}$ is a plausible sequence, we have

$$\begin{aligned} (\hat{\mathbf{y}}_{A_2}, \mathbf{x}'_S) &\in T_\delta & (6.8)\\ (\hat{\mathbf{y}}_{B_2}, \mathbf{x}'_{A_1}, \mathbf{x}'_{A_2}) &\in T_\delta & (6.9)\\ (\hat{\mathbf{y}}_D, \mathbf{x}'_{B_1}, \mathbf{x}'_{B_2}) &\in T_\delta & (6.10) \end{aligned}$$

Then we have

$$
\begin{aligned}
\mathcal{P} &= \mathbb{P}\{(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta,(\hat{\mathbf{y}}_{B_2},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\in T_\delta,(\hat{\mathbf{y}}_D,\mathbf{x}'_{B_1},\mathbf{x}'_{B_2})\in T_\delta,(\hat{\mathbf{y}}_{A_1},\mathbf{x}'_S)\notin T_\delta,\\
&\quad ,(\hat{\mathbf{y}}_{B_1},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\notin T_\delta\} \qquad (6.11)\\
&= \mathbb{P}\{(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta\}\times\mathbb{P}\{(\hat{\mathbf{y}}_{B_2},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\in T_\delta,(\hat{\mathbf{y}}_{A_1},\mathbf{x}'_S)\notin T_\delta|(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta\}\\
&\quad\times\mathbb{P}\{(\hat{\mathbf{y}}_D,\mathbf{x}'_{B_1},\mathbf{x}'_{B_2})\in T_\delta,(\hat{\mathbf{y}}_{B_1},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\notin T_\delta|(\hat{\mathbf{y}}_{B_2},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\in T_\delta,(\hat{\mathbf{y}}_{A_1},\mathbf{x}'_S)\notin T_\delta,\\
&\quad ,(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta\} \qquad (6.12)\\
&\le \mathbb{P}\{(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta\}\times\mathbb{P}\{(\hat{\mathbf{y}}_{B_2},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\in T_\delta|(\hat{\mathbf{y}}_{A_1},\mathbf{x}'_S)\notin T_\delta(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta\}\\
&\quad\times\mathbb{P}\{(\hat{\mathbf{y}}_D,\mathbf{x}'_{B_1},\mathbf{x}'_{B_2})\in T_\delta|(\hat{\mathbf{y}}_{B_1},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\notin T_\delta,(\hat{\mathbf{y}}_{B_2},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\in T_\delta,\\
&\quad ,(\hat{\mathbf{y}}_{A_1},\mathbf{x}'_S)\notin T_\delta,(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta\} \qquad (6.13)\\
&= \mathbb{P}\{(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta\}\times\mathbb{P}\{(\hat{\mathbf{y}}_{B_2},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\in T_\delta|(\hat{\mathbf{y}}_{A_1},\mathbf{x}'_S)\notin T_\delta,(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta\}\\
&\quad\times\mathbb{P}\{(\hat{\mathbf{y}}_D,\mathbf{x}'_{B_1},\mathbf{x}'_{B_2})\in T_\delta|(\hat{\mathbf{y}}_{B_1},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\notin T_\delta,(\hat{\mathbf{y}}_{B_2},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\in T_\delta\} \qquad (6.14)
\end{aligned}
$$

where the last step is true since there is an independent random mapping at each node and we have the following markov structure in the network

$$
X_S \rightarrow (Y_{A_1},Y_{A_2}) \rightarrow (Y_{B_1},Y_{B_2}) \rightarrow Y_D \qquad (6.15)
$$

For any such sequence $\mathbf{x}'_{\mathcal{V}}$, since w is independent of w', we have

$$
\mathbb{P}\left\{(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta\right\} \le 2^{-TI(X_S;Y_{A_2})} \qquad (6.16)
$$

Now, for the layer (A_1,A_2), we condition on a particular sequence $\mathbf{x}_{A_2}$ to have been transmitted by A_2. If $\mathbf{x}'_{A_2}=\mathbf{x}_{A_2}$, since $\mathbf{x}'_{A_1}$ is chosen independent of $\mathbf{x}_{A_1}$ we have,

$$
\mathbb{P}\{(\hat{\mathbf{y}}_{B_2},\mathbf{x}'_{A_1},\mathbf{x}'_{A_2})\in T_\delta|(\hat{\mathbf{y}}_{A_1},\mathbf{x}'_S)\notin T_\delta,(\hat{\mathbf{y}}_{A_2},\mathbf{x}'_S)\in T_\delta,\mathbf{x}'_{A_2}=\mathbf{x}_{A_2}\} \le 2^{-TI(\hat{Y}_{B_2};X_{A_1}|X_{A_2})}, \qquad (6.17)
$$

and similarly If $\mathbf{x}'_{A_2} \neq \mathbf{x}_{A_2}$, since $\mathbf{x}'_{A_1}, \mathbf{x}'_{A_2}$ are chosen independent of $\mathbf{x}_{A_1}, \mathbf{x}_{A_2}$ we have,

$$\begin{aligned} \mathbb{P}\{(\hat{\mathbf{y}}_{B_2}, \mathbf{x}'_{A_1}, \mathbf{x}'_{A_2}) \in T_\delta | (\hat{\mathbf{y}}_{A_1}, \mathbf{x}'_S) \notin T_\delta, (\hat{\mathbf{y}}_{A_2}, \mathbf{x}'_S) \in T_\delta, \mathbf{x}'_{A_2} \neq \mathbf{x}_{A_2}\} \\ \leq 2^{-TI(\hat{Y}_{B_2}; X_{A_1}, X_{A_2})} \quad (6.18) \\ \leq 2^{-TI(\hat{y}_{B_2}; X_{A_1} | X_{A_2})} \quad (6.19) \end{aligned}$$

Therefore in any case,

$$\mathbb{P}\{(\hat{\mathbf{y}}_{B_2}, \mathbf{x}'_{A_1}, \mathbf{x}'_{A_2}) \in T_\delta | (\hat{\mathbf{y}}_{A_1}, \mathbf{x}'_S) \notin T_\delta, (\hat{\mathbf{y}}_{A_2}, \mathbf{x}'_S) \in T_\delta\} \leq 2^{-TI(\hat{Y}_{B_2}; X_{A_1} | X_{A_2})}, \quad (6.20)$$

Similarly we can show that,

$$\mathbb{P}\{(\hat{\mathbf{y}}_D, \mathbf{x}'_{B_1}, \mathbf{x}'_{B_2}) \in T_\delta | (\hat{\mathbf{y}}_{B_1}, \mathbf{x}'_{A_1}, \mathbf{x}'_{A_2}) \notin T_\delta, (\hat{\mathbf{y}}_{B_2}, \mathbf{x}'_{A_1}, \mathbf{x}'_{A_2}) \in T_\delta\} \leq 2^{-TI(\hat{Y}_D; X_{B_1} | X_{B_2})}, \quad (6.21)$$

Therefore for any typical sequence $\mathbf{x}'_{\mathcal{V}}$, we have

$$\begin{aligned} \mathcal{P} &\leq 2^{-TI(X_S; Y_{A_2})} \times 2^{-TI(\hat{Y}_{B_2}; X_{A_1} | X_{A_2})} \times 2^{-TI(\hat{Y}_D; X_{B_1} | X_{B_2})} \\ &= 2^{-TI(X_\Omega; \hat{Y}_{\Omega^c} | X_{\Omega^c})} \quad (6.22) \end{aligned}$$

Now, by summing over all possible $\mathbf{x}'_{\mathcal{V}}$'s and cuts, the probability of confusing w with w' can be bounded by

$$\mathbb{P}\{w \to w'\} \leq |\mathcal{X}_{\mathcal{V}}(w')| \sum_{\Omega} 2^{-TI\left(X_\Omega; \hat{Y}_{\Omega^c} | X_{\Omega^c}\right)} \quad (6.23)$$

In the next section, we make these arguments precise, and by bounding $|\mathcal{X}_{\mathcal{V}}(w')|$ we prove our main theorem 4.3.1 for networks with a layered structure.

6.2.3 Proof of main Theorem 4.3.1 for layered networks

In this section we extend the idea from section 6.2.2 and analyze a l_D-layer network, $\mathcal{G}$.

Based on the proof strategy illustrated in section 6.2.2, we proceed with the error probability analysis of our scheme that was described in section 6.2.1. Assume message w is being transmitted. Then we have

$$\mathcal{P}_{\text{error}} = \mathbb{P}\left\{(\hat{\mathbf{y}}_D, w) \notin T_\delta\right\} + \mathbb{P}\left\{\exists w' \neq w \text{ s.t.}(\hat{\mathbf{y}}_D, w') \in T_\delta \;\&\; (\hat{\mathbf{y}}_D, w) \in T_\delta\right\} \tag{6.24}$$

By law of large numbers, the probability of the first event becomes arbitrarily small as communication block length, T, goes to infinity. So to bound the probability of error, we just need to analyze the probability that $\hat{\mathbf{y}}_D$ is jointly typical with *both* w, w', for a message w' independent of w. We denote this event by $w \to w'$.

Now if $\hat{\mathbf{y}}_D$ is jointly typical with w', then there must be a typical transmit sequence $\mathbf{x}'_{\mathcal{V}} \in \mathcal{X}_{\mathcal{V}}(w')$ under w' such that $(\hat{\mathbf{y}}_D, \mathbf{x}'_{\mathcal{V}}) \in T_\delta$. This means that the destination thinks this is a plausible sequence. Therefore, there is a natural source-destinationcut, Ω, in $\mathcal{G}$ such that the nodes on the right hand side of the cut (*i.e.* in Ω) can tell $\mathbf{x}'_{\mathcal{V}}$ is not a plausible sequence, and those on the left hand side of the cut (*i.e.* in Ω^c) can not. Since we are dealing with a layered network, each cut decomposes the network to $d = l_D$ disjoint sub-networks , such that at each layer we have the set of nodes $\beta_l(\Omega)$, which are at distance $l-1$ from S and are in Ω, on one side and the set of nodes $\gamma_l(\Omega)$, which are at distance l from S that are in Ω^c, on the other side, for $l = 1, \ldots, l_D$. Therefore, by definition we have

$$(\hat{\mathbf{y}}_{\gamma_l(\Omega)}, \mathbf{x}'_{\beta_l(\Omega)}, \mathbf{x}'_{\gamma_{l-1}(\Omega)}) \in T_\delta, \quad l = 1, \ldots, l_D \tag{6.25}$$

Therefore, we can write

$$
\begin{aligned}
\mathbb{P}\{w \to w'\} &= \mathbb{P}\{(\hat{\mathbf{y}}_D, w') \in T_\delta\} &(6.26)\\
&\le \sum_{\mathbf{x}'_{\mathcal{V}} \in \mathcal{X}_{\mathcal{V}}(w')} \mathbb{P}\left\{(\hat{\mathbf{y}}_D, \mathbf{x}'_{\gamma_{l_D-1}}) \in T_\delta\right\} &(6.27)\\
&= \sum_{\mathbf{x}'_{\mathcal{V}} \in \mathcal{X}_{\mathcal{V}}(w')} \sum_{\Omega \in \Lambda_D} \mathbb{P}\{(\hat{\mathbf{y}}_{\gamma_l(\Omega)}, \mathbf{x}'_{\beta_l(\Omega)}, \mathbf{x}'_{\gamma_{l-1}(\Omega)}) \in T_\delta,\\
&\quad ,(\hat{\mathbf{y}}_{\beta_l(\Omega)}, \mathbf{x}'_{\beta_{l-1}(\Omega)}, \mathbf{x}'_{\gamma_{l-2}(\Omega)}) \notin T_\delta \ , l = 1, \ldots, l_D\} &(6.28)\\
&= \sum_{\mathbf{x}'_{\mathcal{V}} \in \mathcal{X}_{\mathcal{V}}(w')} \sum_{\Omega \in \Lambda_D} \prod_{l=1}^{l_D} \mathbb{P}\{(\hat{\mathbf{y}}_{\gamma_l(\Omega)}, \mathbf{x}'_{\beta_l(\Omega)}, \mathbf{x}'_{\gamma_{l-1}(\Omega)}) \in T_\delta,\\
&\quad ,(\hat{\mathbf{y}}_{\beta_l(\Omega)}, \mathbf{x}'_{\beta_{l-1}(\Omega)}, \mathbf{x}'_{\gamma_{l-2}(\Omega)}) \notin T_\delta | (\hat{\mathbf{y}}_{\gamma_j(\Omega)}, \mathbf{x}'_{\beta_j(\Omega)}, \mathbf{x}'_{\gamma_{j-1}(\Omega)}) \in T_\delta,\\
&\quad ,(\hat{\mathbf{y}}_{\beta_j(\Omega)}, \mathbf{x}'_{\beta_{j-1}(\Omega)}, \mathbf{x}'_{\gamma_{j-2}(\Omega)}) \notin T_\delta, j = 1, \ldots, l-1\} &(6.29)\\
&\le \sum_{\mathbf{x}'_{\mathcal{V}} \in \mathcal{X}_{\mathcal{V}}(w')} \sum_{\Omega \in \Lambda_D} \prod_{l=1}^{l_D} \mathbb{P}\{(\hat{\mathbf{y}}_{\gamma_l(\Omega)}, \mathbf{x}'_{\beta_l(\Omega)}, \mathbf{x}'_{\gamma_{l-1}(\Omega)}) \in T_\delta |\\
&\quad (\hat{\mathbf{y}}_{\beta_l(\Omega)}, \mathbf{x}'_{\beta_{l-1}(\Omega)}, \mathbf{x}'_{\gamma_{l-2}(\Omega)}) \notin T_\delta, (\hat{\mathbf{y}}_{\gamma_j(\Omega)}, \mathbf{x}'_{\beta_j(\Omega)}, \mathbf{x}'_{\gamma_{j-1}(\Omega)}) \in T_\delta,\\
&\quad ,(\hat{\mathbf{y}}_{\beta_j(\Omega)}, \mathbf{x}'_{\beta_{j-1}(\Omega)}, \mathbf{x}'_{\gamma_{j-2}(\Omega)}) \notin T_\delta, j = 1, \ldots, l-1\} &(6.30)\\
&= \sum_{\mathbf{x}'_{\mathcal{V}} \in \mathcal{X}_{\mathcal{V}}(w')} \sum_{\Omega \in \Lambda_D} \prod_{l=1}^{l_D} \mathbb{P}\{(\hat{\mathbf{y}}_{\gamma_l(\Omega)}, \mathbf{x}'_{\beta_l(\Omega)}, \mathbf{x}'_{\gamma_{l-1}(\Omega)}) \in T_\delta |\\
&\quad (\hat{\mathbf{y}}_{\beta_l(\Omega)}, \mathbf{x}'_{\beta_{l-1}(\Omega)}, \mathbf{x}'_{\gamma_{l-2}(\Omega)}) \notin T_\delta, (\hat{\mathbf{y}}_{\gamma_{l-1}(\Omega)}, \mathbf{x}'_{\beta_{l-1}(\Omega)}, \mathbf{x}'_{\gamma_{l-2}(\Omega)}) \in T_\delta\} &(6.31)
\end{aligned}
$$

where the last step is true since there is an independent random mapping at each node and we have a markovian layered structure in the network. Now, conditioned on a particular $\mathbf{x}_{\gamma_{l-1}(\Omega)} \in \mathcal{X}_{\gamma_{l-1}(\Omega)}(w)$, we have two situations, either $\mathbf{x}'_{\gamma_{l-1}(\Omega)} = \mathbf{x}_{\gamma_{l-1}(\Omega)}$, or not. If $\mathbf{x}'_{\gamma_{l-1}(\Omega)} = \mathbf{x}_{\gamma_{l-1}(\Omega)}$, since $\mathbf{x}'_{\beta_l(\Omega)}$ is chosen independent of $\mathbf{x}_{\beta_l(\Omega)}$ (because of the random encoding function at each node), we have

$$\mathbb{P}\{(\hat{\mathbf{y}}_{\gamma_l(\Omega)}, \mathbf{x}'_{\beta_l(\Omega)}, \mathbf{x}'_{\gamma_{l-1}(\Omega)}) \in T_\delta | (\hat{\mathbf{y}}_{\beta_l(\Omega)}, \mathbf{x}'_{\beta_{l-1}(\Omega)}, \mathbf{x}'_{\gamma_{l-2}(\Omega)}) \notin T_\delta, \mathbf{x}'_{\gamma_{l-1}(\Omega)} = \mathbf{x}_{\gamma_{l-1}(\Omega)}, \\ , (\hat{\mathbf{y}}_{\gamma_{l-1}(\Omega)}, \mathbf{x}'_{\beta_{l-1}(\Omega)}, \mathbf{x}'_{\gamma_{l-2}(\Omega)}) \in T_\delta\} \leq 2^{-TI\left(X_{\beta_l(\Omega)}; \hat{Y}_{\gamma_l(\Omega)} | X_{\gamma_{l-1}(\Omega)}\right)} \tag{6.32}$$

On the other hand if $\mathbf{x}'_{\gamma_{l-1}(\Omega)} \neq \mathbf{x}_{\gamma_{l-1}(\Omega)}$, since $\mathbf{x}'_{\gamma_{l-1}(\Omega)}$ and $\mathbf{x}'_{\beta_l(\Omega)}$ are respectively chosen independent of $\mathbf{x}_{\gamma_{l-1}(\Omega)}$ and $\mathbf{x}_{\beta_l(\Omega)}$ (because of the random encoding function at each node), we have

$$\begin{aligned} \mathbb{P}\{(\hat{\mathbf{y}}_{\gamma_l(\Omega)}, \mathbf{x}'_{\beta_l(\Omega)}, \mathbf{x}'_{\gamma_{l-1}(\Omega)}) \in T_\delta | (\hat{\mathbf{y}}_{\beta_l(\Omega)}, \mathbf{x}'_{\beta_{l-1}(\Omega)}, \mathbf{x}'_{\gamma_{l-2}(\Omega)}) \notin T_\delta, \mathbf{x}'_{\gamma_{l-1}(\Omega)} \neq \mathbf{x}_{\gamma_{l-1}(\Omega)}, \\ , (\hat{\mathbf{y}}_{\gamma_{l-1}(\Omega)}, \mathbf{x}'_{\beta_{l-1}(\Omega)}, \mathbf{x}'_{\gamma_{l-2}(\Omega)}) \in T_\delta\} &\leq 2^{-TI\left(X_{\gamma_{l-1}(\Omega)}, X_{\beta_l(\Omega)}; \hat{Y}_{\gamma_l(\Omega)}\right)} && (6.33) \\ &\leq 2^{-TI\left(X_{\beta_l(\Omega)}; \hat{Y}_{\gamma_l(\Omega)} | X_{\gamma_{l-1}(\Omega)}\right)} && (6.34) \end{aligned}$$

Now by equations (6.31), (6.32), (6.34), we get

$$\begin{aligned} \mathbb{P}\{w \to w'\} &\leq |\mathcal{X}_{\mathcal{V}}(w')| \sum_{\Omega} \prod_{l=1}^{l_D} 2^{-TI\left(X_{\beta_l(\Omega)}; \hat{Y}_{\gamma_l(\Omega)} | X_{\gamma_{l-1}(\Omega)}\right)} && (6.35) \\ &= |\mathcal{X}_{\mathcal{V}}(w')| \sum_{\Omega} 2^{-TI\left(X_\Omega; \hat{Y}_{\Omega^c} | X_{\Omega^c}\right)} && (6.36) \end{aligned}$$

As the last ingredient of the proof, we state the following lemma:

Lemma 6.2.4. *Consider a layered Gaussian relay network, $\mathcal{G}$, then,*

$$|\mathcal{X}_{\mathcal{V}}(w')| \leq 2^{T\kappa_1} \tag{6.37}$$

where $\kappa_1 = |\mathcal{V}|$ is a constant depending on the total number of nodes in $\mathcal{G}$.

Proof. See Appendix A.4. □

Therefore, by (6.36) and lemma 6.2.4, we have the following,

Lemma 6.2.5. *Given a Gaussian relay network $\mathcal{G}$ with a layered structure, all rates R satisfying the following condition are achievable,*

$$R < \min_{\Omega \in \Lambda_D} I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}) - \kappa_1 \tag{6.38}$$

where X_i, $i \in \mathcal{V}$, are iid with complex normal (Gaussian) distribution, and $\kappa_1 = |\mathcal{V}|$ is a constant depending on the total number of nodes in $\mathcal{G}$.

To prove our main theorem 4.3.1 for layered networks, we state the following lemma,

Lemma 6.2.6. *Given a Gaussian relay network $\mathcal{G}$, then*

$$\overline{C} - \min_{\Omega \in \Lambda_D} I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}) < \kappa_2 \tag{6.39}$$

where X_i, $i \in \mathcal{V}$, are iid with complex normal (Gaussian) distribution, $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in equation (4.1), and $\kappa_2 = 2|\mathcal{V}|$.

Proof. See Appendix A.5. □

Now by lemma 6.2.5 and lemma 6.2.6, we have the following main result

Theorem 6.2.7. *Given a Gaussian relay network $\mathcal{G}$ with a layered structure and single antenna at each node, all rates R satisfying the following condition are achievable,*

$$R < \overline{C} - \kappa_{Lay} \tag{6.40}$$

where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in equation (4.1), and $\kappa_{Lay} = \kappa_1 + \kappa_2 = 3|\mathcal{V}|$ is a constant depending on the total number of nodes in $\mathcal{G}$ (denoted by $|\mathcal{V}|$).

6.3 General Gaussian relay networks

Given the proof for layered networks with equal path lengths, we are ready to tackle the proof of Theorem 4.3.1 for general Gaussian relay networks.

First we formally describe the encoding strategy:

6.3.0.1 Encoding for general Gaussian relay network

We have a single source S with a sequence of messages $w_k \in \{1, 2, \ldots, 2^{KTR}\}$, $k = 1, 2, \ldots$. Each message is encoded by the source S into a signal over KT transmission times (symbols), giving an overall transmission rate of R.

At each node we create a random Gaussian codebook. Source randomly maps each message to one of its Gaussian codewords and sends it in KT transmission times. Still, each relay operates over blocks of time T symbols. In particular each received sequence $\mathbf{y}_i^{(k)}$ at node i is quantized into $\hat{\mathbf{y}}_i^{(k)}$ which is then randomly mapped into a Gaussian codeword $\mathbf{x}_i^{(k)}$ using a random mapping function $f_i(\hat{\mathbf{y}}_i^{(k)})$. Given the knowledge of all the encoding functions at the relays and signals received over $K + |V| - 2$ blocks, the decoder D, attempts to decode the message W sent by the source.

Note that the general achievability scheme that we use here is similar to the one described in Section 6.2.1 for layered networks, except now the message $W \in \{1, \ldots, 2^{KRT}\}$ is encoded by the source S into a signal over KT transmission times (symbols), whilel each relay operates over blocks of time T symbols.

The ingredients of the proof are developed below. First similar to the deterministic case (Section 5.4), we use time expansion idea to explicitly represent our relaying scheme. Now we state the following lemma which is a corollary of Theorem 6.2.7.

Lemma 6.3.1. *Given a Gaussian relay network, $\mathcal{G}$, all rates R satisfying the following* *condition are achievable,*

$$R < \frac{1}{K} \min_{\Omega_{unf} \in \Lambda_D} I(Y_{\Omega_{unf}^c}; X_{\Omega_{unf}} | X_{\Omega_{unf}^c}) - \kappa_1 \tag{6.41}$$

where $\mathcal{G}_{unf}^{(K)}$ is the time expanded graph associated with $\mathcal{G}$, random variables $\{X_i[t]\}_{1 \leq t \leq K}$, $i \in \mathcal{V}$ are iid with complex normal (Gaussian) distribution, and $\kappa_1 = 3|\mathcal{V}|$.

Proof. By unfolding $\mathcal{G}$ we get an acyclic network such that all the paths from the source to the destination have equal length. Therefore, by theorem 6.2.7, all rates R_{unf}, satisfying the following condition are achievable in the time-expanded graph

$$R_{\text{unf}} < \min_{\Omega_{\text{unf}} \in \Lambda_D} I(Y_{\Omega_{\text{unf}}^c}; X_{\Omega_{\text{unf}}} | X_{\Omega_{\text{unf}}^c}) - \kappa_{\text{unf}} \tag{6.42}$$

where $\{X_i[t]\}_{1 \leq t \leq K}, i \in \mathcal{V}$ are iid with complex normal (Gaussian) distribution, and $\kappa_{\text{unf}} = 3K|\mathcal{V}|$. Since it takes K steps to translate and achievable scheme in the time-expanded graph to an achievable scheme in the original graph, and $\kappa_1 = \frac{1}{K}\kappa_{\text{unf}} = 3|\mathcal{V}|$, then the Lemma is proved. □

Similar to the deterministic case, in the following lemma we show that the minimum cut value of the time expanded graph (normalized by $1/K$) approaches to the minimum cut value of the original graph as $K \to \infty$.

Lemma 6.3.2. *Consider a Gaussian relay network, $\mathcal{G}$. Then for any cut Ω_{unf} on the unfolded graph we have,*

$$(K - L + 1) \min_{\Omega \in \Lambda_D} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}) \leq I(Y_{\Omega_{unf}^c}; X_{\Omega_{unf}} | X_{\Omega_{unf}^c}) \tag{6.43}$$

where $L = 2^{|\mathcal{V}|-2}$, $X_{i \in \mathcal{V}}$ are iid with complex normal (Gaussian) distribution, and $\{X_i[t]\}$, $1 \leq t \leq K,\ i \in \mathcal{V}$ are also iid with complex normal (Gaussian) distribution.

Proof. See Appendix A.6. □

Hence, by lemma 6.3.1 and lemma 6.3.2 we have the following lemma,

Lemma 6.3.3. *Given a Gaussian relay network $\mathcal{G}$, all rates R satisfying the following condition are achievable,*

$$R < \min_{\Omega \in \Lambda_D} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}) - \kappa_1 \tag{6.44}$$

where X_i, $i \in \mathcal{V}$, are i.i.d. with complex normal (Gaussian) distribution, and $\kappa_1 = 3|\mathcal{V}|$.

Now by lemma 6.2.6 we know that,

$$\begin{aligned} \overline{C} - \min_{\Omega \in \Lambda_D} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}) &\leq \overline{C} - \min_{\Omega \in \Lambda_D} I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}) \\ &\leq 2|\mathcal{V}| \end{aligned} \tag{6.45}$$

where X_i, $i \in \mathcal{V}$, are iid with complex normal (Gaussian) distribution.

Therefore, by lemma 6.3.3 and inequality (6.45) all rates up to $\overline{C} - |\mathcal{V}|(3+2) = \overline{C} - 5|\mathcal{V}|$ are achieved and the proof of our main theorem 4.3.1 is complete.

To prove Theorem 4.3.1 for the multicast scenario, we just need to note that if all relays will perform exactly the same strategy then by our theorem, each destination, $D \in \mathcal{D}$, will be able to decode the message with low error probability as long as the rate of the message satisfies

$$R < \overline{C}_D - \kappa \tag{6.46}$$

where $\kappa < 5|\mathcal{V}|$ is a constant and $\overline{C}_D = \max_{p(\{x_j\}_{j \in \mathcal{V}})} \min_{\Omega \in \Lambda_D} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c})$ is the cut-set upper bound on the capacity from the source to D. Therefore as long as $R < \min_D \overline{C}_D - \kappa$, all destinations can decode the message and hence the theorem is proved when we have single antennas at each node.

In the case that we have multiple antennas at each node, the achievability strategy remains the same, except now each node receives a vector of observations from different antennas. We will first quantize the received signal of each antenna at noise level and then map it to another transmit codeword. The error probability analysis is exactly the same as before. However, the gap between the achievable rate and the cut-set bound will be larger. We can upper bound the gap by assuming that we have a network with at most *i.e.* $\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i)$ virtual nodes (each correspond to an antenna). Therefore from our previous analysis we know that the gap is at most $5\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i)$ and the theorem is proved.

6.4 Summary

In this chapter we proved Theorem 4.3.1, which is a lower bound to the achievable rate for information flow over noisy Gaussian relay networks. Similar to the deterministic case, first we proved if for networks that have a layered structure, then extended it to an arbitrary network by considering its time-expanded representation. We proved that the gap between this achievable rate and the cut-set upper bound is uniformly bounded by a constant that is independent of the channel gains. Hence, established a uniform approximation result for the capacity of Gaussian relay networks. This is the first constant gap approximation of the capacity of Gaussian relay networks.

Chapter 7

Extensions of our main result

7.1 Introduction

In this chapter we extend our main result for Gaussian relay networks (Theorem 4.3.1) to the following scenarios:

1. Compound relay network
2. Frequency selective relay network
3. Half-duplex relay network
4. Quasi-static fading relay network (underspread regime)
5. Low rate capacity approximation of Gaussian relay network

We first motivate each extension and then prove our result for that extension.

7.2 Compound relay network

The relaying strategy that we proposed for general Gaussian relay networks does not require any channel information at the relays, relays just quantize at noise level and forward through a random mapping. The approximation gap also does not depend on the channel gain values. As a result our main result for Gaussian relay networks (Theorem 4.3.1) can be extended to compound relay networks where we allow each channel gain $h_{i,j}$ to be from a set $\mathcal{H}_{i,j}$, and the particular chosen values are unknown to the source node S, the relays and the destination node D. A communication rate R is achievable if there exist a scheme such that for any channel gain realizations, still the source can communicate to the destination at rate R, without the knowledge of the channel realizations at the source, the relays and the destination. For completeness, we restate our result for compound relay network (Theorem 4.4.1) and then prove it.

Theorem 7.2.1. *Given a compound Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the capacity C_{cn} satisfies*

$$\overline{C}_{cn} - \kappa \leq C_{cn} \leq \overline{C}_{cn} \tag{7.1}$$

Where $\overline{C}_{cn}$ is the cut-set upper bound on the compound capacity of $\mathcal{G}$ as described below

$$\overline{C}_{cn} = \max_{p(\{\mathbf{X}_i\}_{j\in\mathcal{V}})} \inf_{h\in\mathcal{H}} \min_{\Omega\in\Lambda_D} I(\mathbf{Y}_{\Omega^c}; \mathbf{X}_\Omega|\mathbf{X}_{\Omega^c}) \tag{7.2}$$

And κ is a constant and is upper bounded by $6\sum_{i=1}^{|\mathcal{V}|}\max(M_i, N_i)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

Proof outline: We sketch the proof for the case that nodes have single antenna, its extension to the multiple antenna scenario is straightforward. As we mentioned earlier, the relaying strategy that we used in main Theorem 4.3.1, does not require any channel information. However, if all channel gains are known at the final destination, all rates within a constant gap to the cut-set upper bound are achievable. Now we first evaluate how much we lose if the final destination only knows a quantized version of the channel gains. In particular assume that each channel gain is bounded $|h_{ij}| \in [h_{\min}, h_{\max}]$, and final destination only knows the channel gain values quantized at $\frac{1}{\mathsf{SNR}}$ level so that overall with signal it is at noise level. Then since there is a transmit power constraint equal to one at each node, the effect of this channel uncertainty can be mimicked by adding a Gaussian noise of variance 1 at each relay node (or reducing all channel SNR's of the links 3dB), which will result in a reduction of at most $|\mathcal{V}|$ bits from the cut-set upper bound. Therefore with access to only quantized channel gains, we will lose at most $|\mathcal{V}|$ more bits, which means the gap between the achievable rate and the cut-set bound is at most $6|\mathcal{V}|$.

Furthermore, as shown in [35] there exists a universal decoder for this finite group of channel sets. Hence we can use this decoder at the final destination and decode the message as if we knew the channel gains quantized at the noise level, for all rates up to

$$R < \max_{p(\{x_i\}_{j\in\mathcal{V}})} \inf_{\hat{h}\in\hat{\mathcal{H}}} \min_{\Omega\in\Lambda_D} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}) \tag{7.3}$$

where $\hat{\mathcal{H}}$ is representing the quantized state space. Now as we showed earlier, if we restrict the channels to be quantized at noise level the cut-set upper bound changes at most by $|\mathcal{V}|$, therefore

$$\overline{C}_{cn} - |\mathcal{V}| \leq \max_{p(\{x_i\}_{j\in\mathcal{V}})} \inf_{\hat{h}\in\hat{\mathcal{H}}} \min_{\Omega\in\Lambda_D} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}) \tag{7.4}$$

Therefore from equations (7.3) and (7.4) all rates up to $\overline{C}_{cn} - 6|\mathcal{V}|$ are achievable and the proof can be completed.

Now by using the ideas in [36] and [37], we believe that an infinite state universal decoder can also be analysed to give "completely oblivious to channel" results. □

7.3 Frequency selective Gaussian relay network

In this section we generalize our main result to the case that the channels are frequency selective. Since one can present a frequency selective channel as a MIMO link, where each antenna is operating at a different frequency band[1], this extension is a just straight forward corollary of the case that nodes have multiple antennas. For completeness, we restate our result for frequency selective relay network (Theorem 4.4.3).

Theorem 7.3.1. *Given a frequency selective Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with F different frequency bands. The capacity of this network, C, satisfies*

$$\overline{C} - \kappa \le C \le \overline{C} \tag{7.5}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in equation (4.1), and κ is a constant and is upper bounded by $5\sum_{i=1}^{|\mathcal{V}|}\max(M_i, N_i)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

7.4 Half duplex relay network (fixed scheduling)

One of the practical constraints on wireless networks is that the transceivers can not transmit and receive at the same time on the same frequency band, known as the half-duplex constraint. As a result of this constraint, the achievable rate of the network will in general be lower. In this section we study the capacity of wireless relay networks under the half-duplex constraint. The model that we use to study this problem is the same as [38]. In this model the network has finite modes of operation. Each mode of operation (or state of the network), denoted by $m \in \{1, 2, \dots, M\}$, is defined as a valid partitioning of the nodes of the network into two sets of "sender" nodes and "receiver" nodes such that there is no active link that arrives at a sender node[2]. For each node i, the transmit and the receive signal at mode m are respectively shown by x_i^m and y_i^m. Also t_m defines the portion of the time that network will operate in state m, as the network use goes to infinity. The cut-set upper

[1] This can be implemented in particular by using OFDM and appropriate spectrum shaping or allocation.

[2] Active link is defined as a link which is departing from the set of sender nodes

bound on the capacity of the Gaussian relay network with half-duplex constraint, C_{hd}, is shown to be [38]:

$$C_{hd} \leq \overline{C}_{hd} = \max_{\substack{p(\{x_j^m\}_{j\in\mathcal{V}, m\in\{1,\ldots,M\}}) \\ t_m:\ 0\leq t_m\leq 1,\ \sum_{m=1}^{M} t_m=1}} \min_{\Omega\in\Lambda_D} \sum_{m=1}^{M} t_m I(Y_{\Omega^c}^m; X_\Omega^m | X_{\Omega^c}^m) \tag{7.6}$$

For completeness, we restate our result for half-duplex relay network (Theorem 4.4.2) and then prove it.

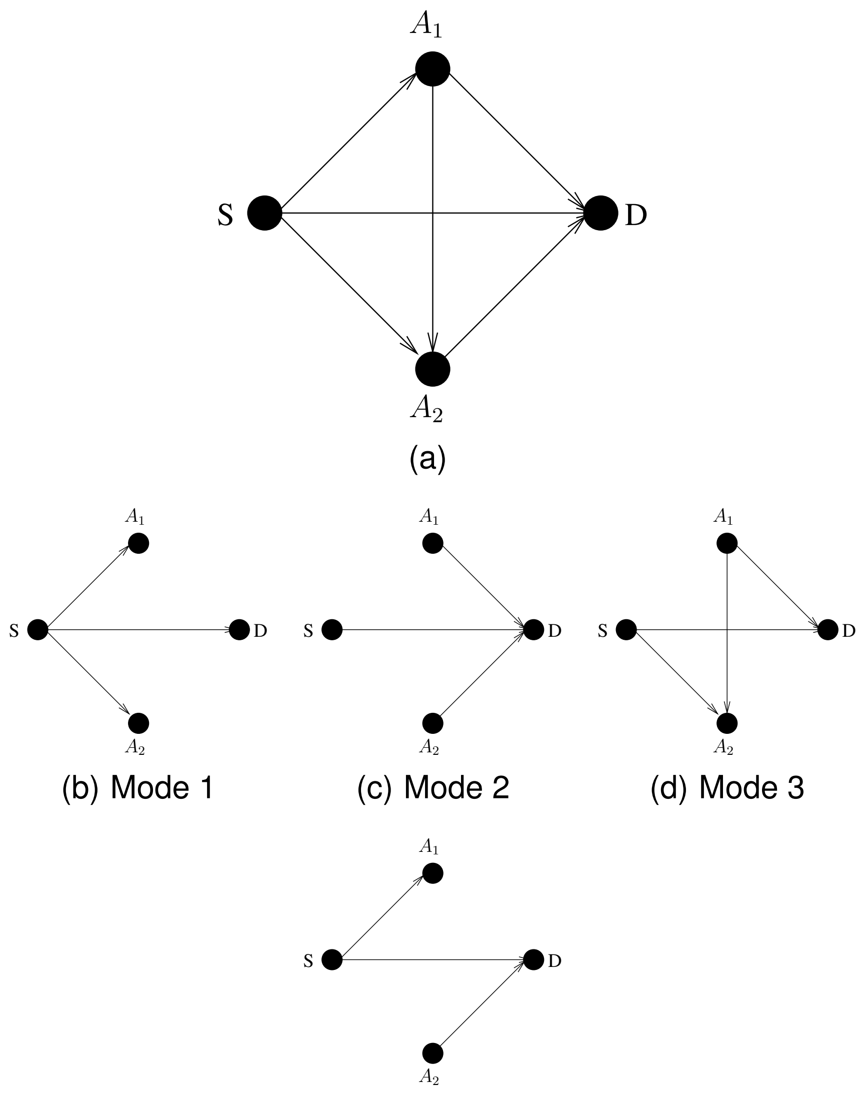
Theorem 7.4.1. *Given a Gaussian relay network with half-duplex constraint, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the capacity, C_{hd}, satisfies*

$$\overline{C}_{hd} - \kappa \leq C_{hd} \leq \overline{C}_{hd} \tag{7.7}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in equation (7.6), and κ is a constant and is upper bounded by $5\sum_{i=1}^{|\mathcal{V}|}\max(M_i, N_i)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

Proof. We prove the result for the case that nodes have single antenna, its extension to the multiple antenna scenario is straightforward. Since each relay can be either in a transmit or receive mode, we have a total of $M = 2^{|\mathcal{V}|-2}$ number of modes. An example of a network with two relay and all four modes of half-duplex operation of the relays are shown in Figure 7.1.

Now consider the t_i's that maximize $\overline{C}_{hd}$ in (4.15). Assume that they are rational numbers (otherwise look at the sequence of rational numbers approaching them) and set W to be the LCM (least common divisor) of the denominators. Now increase the bandwidth of system by W and allocate Wt_i of bandwidth to mode i, $i = 1, \ldots, M$. Now each mode is running at a different frequency band, therefore as shown in Figure 7.2 we can combine all these modes and create a frequency selective relay network. Since the links are orthogonal to each other, still the cut-set upper bound on the capacity of this frequency selective relay network (in bits/sec/Hz) is the same as (4.15). Now by theorem 4.4.3 we know that our



(a)

(b) Mode 1

(c) Mode 2

(d) Mode 3

(e) Mode 4

Figure 7.1: An example of a relay network with two relays is shown in (a). All four modes of half-duplex operation of the relays are shown in $(b) - (e)$.

quantize-map-forward scheme achieves, within a constant gap, κ, of $\overline{C}_{hd}$ for all channel gains. In this relaying scheme, at each block, each relay transmits a signal that is only a function of its received signal in the previous block and hence does not have memory over different blocks. Now we will translate this scheme to a scheme in the original network that modes are just at different times (not different frequency bands). The idea is that we can expand exactly communication block of the frequency selective network into W blocks of the original network and allocating Wt_i of these blocks to mode i. Then in the Wt_i blocks that are allocated to mode i, all relays do exactly what they do in frequency band i. This is pictorially described in Figure 7.3 for the network of Figure 7.2. This figure shows how one communication block of the frequency selective network (a) is expanded over W blocks of the the original half-duplex network (b). Now since the transmitted signal at each frequency band is only a function of the data received in the previous block of the frequency selective network, the ordering of the modes inside the W blocks of the original network is not important at all. Therefore with this strategy we can achieve within a constant gap, κ, of the cut-set bound of the half-duplex relay network and the proof is complete.

The main difference between this strategy and our original strategy for full duplex networks is that now the relays are required to have a much larger memory. As a matter of fact, in the full duplex scenario the relays had only memory over one block (what they sent was only a function of the previous block). However for the half-duplex scenario the relays are required to have a memory over W blocks and clearly W can be arbitrary large.

□

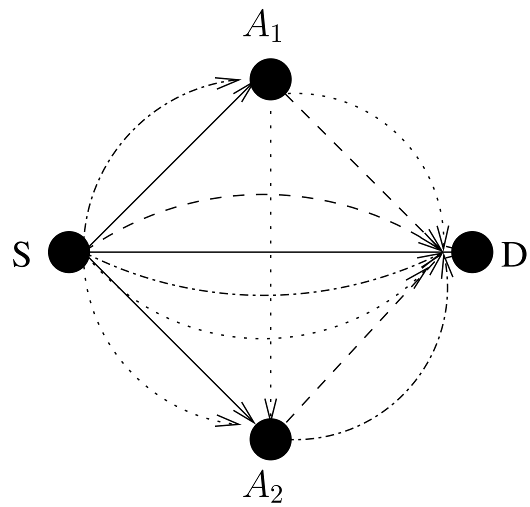


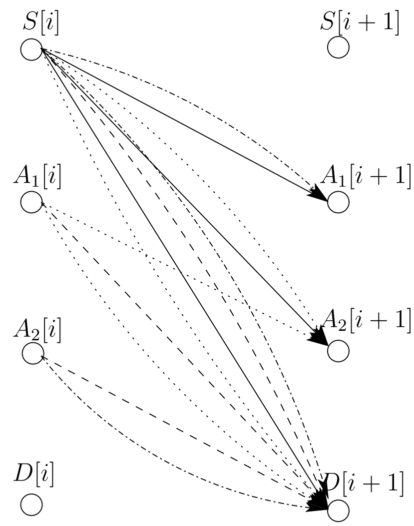
Figure 7.2: Combination of all half-duplex modes of the network shown in figure 7.1. Each mode operates at a different frequency band.

7.5 Quasi-static fading relay network (underspread regime)

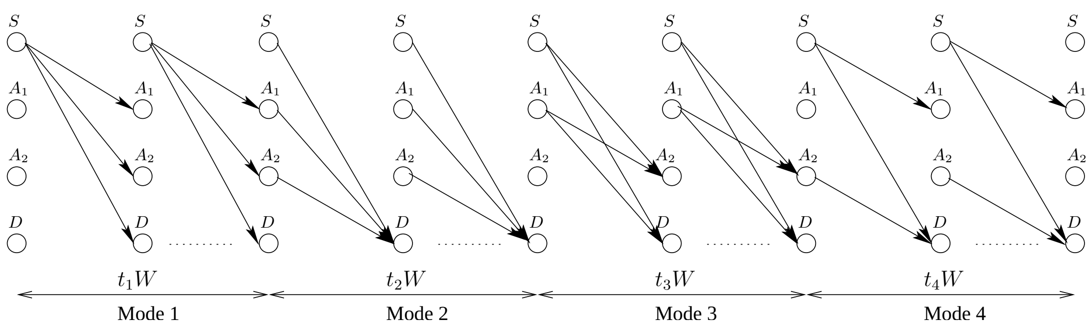
In a wireless environment channel gains are not fixed and change over time. In this section we consider a typical scenario in which although the channel gains are changing, they can be considered time invariant over a long time scale (for example during the transmission of a block). This happens when the coherence time of the channel (T_c) is much larger than the delay spread (T_d). Here the delay spread is the largest extent of the unequal path lengths, which is in some sense corresponding to inter-symbol interference. Now, depending on how fast the channel gains are changing compared to the delay requirements, we have two different regimes: fast fading or slow fading scenarios. We consider each case separately.

7.5.1 Fast fading

In the fast fading scenario the channel gains are changing much faster compared to the delay requirement of the application (*i.e.* coherence time of the channel, T_c, is much smaller than



(a) One communication block of the frequency selective half-duplex network



(b) Expanding over W blocks of the original network

Figure 7.3: One communication block of the frequency selective network (a) is expanded over W blocks of the original half-duplex network (b).

the delay requirements). Therefore, we can interleave data and encode it over different coherence time periods. In this scenario, ergodic capacity of the network is the relevant capacity measure to look at. For completeness, we restate our result for fast fading relay networks (Theorem 4.4.4) and then prove it.

Theorem 7.5.1. *Given a fast fading quasi-static fading Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the ergodic capacity $C_{ergodic}$ satisfies*

$$\mathcal{E}_{h_{ij}}\left[\overline{C}(\{h_{ij}\})\right] - \kappa \leq C_{ergodic} \leq \mathcal{E}_{h_{ij}}\left[\overline{C}(\{h_{ij}\})\right] \tag{7.8}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity, as described in equation (4.1), and the expectation is taken over the channel gain distribution, and κ is a constant and is upper bounded by $5\sum_{i=1}^{|\mathcal{V}|}\max(M_i, N_i)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

Proof. We prove the result for the case that nodes have single antenna, its extension to the multiple antenna scenario is straightforward. Upper bound is just the cut-set upper bound. For the achievability note that the relaying strategy that we proposed for general wireless relay networks does not depend on the channel realization, relays just quantize at noise level and forward through a random mapping. The approximation gap also does not depend on the channel parameters. As a result by coding data over L different channel realizations the following rate is achievable

$$\frac{1}{L}\sum_{l=1}^{L}\left(\overline{C}(\{h_{ij}\}^l) - \kappa\right) \tag{7.9}$$

Now as $L \to \infty$,

$$\frac{1}{L}\sum_{l=1}^{L}\overline{C}(\{h_{ij}\}^l) \to \mathcal{E}_{h_{ij}}\left[\overline{C}\right] \tag{7.10}$$

and the theorem is proved. □

7.5.2 Slow fading

In a slow fading scenario the delay requirement does not allow us to interleave data and encode it over different coherence time periods. We assume that there is no channel gain information available at the source, therefore there is no definite capacity and for a fixed target rate R we should look at the outage probability,

$$\mathcal{P}_{out}(R) = \mathbb{P}\left\{C(\{h_{ij}\}) < R\right\} \tag{7.11}$$

where the probability is calculated over the distribution of the channel gains and the ϵ-outage capacity is defined as

$$C_\epsilon = \mathcal{P}_{out}^{-1}(\epsilon) \tag{7.12}$$

Here is our result to approximate the outage probability

Theorem 7.5.2. *Given a slow fading quasi-static fading Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the outage probability, $\mathcal{P}_{out}(R)$ satisfies*

$$\mathbb{P}\left\{\overline{C}(\{h_{ij}\}) < R\right\} \leq \mathcal{P}_{out}(R) \leq \mathbb{P}\left\{\overline{C}(\{h_{ij}\}) < R + \kappa\right\} \tag{7.13}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity, as described in equation (4.1), and the probability is calculated over the distribution of the channel gains, and κ is a constant and is upper bounded by $5\sum_{i=1}^{|\mathcal{V}|}\max(M_i, N_i)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

Proof. Lower bound is just based on the cut-set upper bound on the capacity. For the upper bound we use the compound network result. Therefore, based on Theorem 4.4.1 we know that as long as $\overline{C}(\{h_{ij}\}) - \kappa < R$ there will not be an outage. $\square$

7.6 Low rate capacity approximation of Gaussian relay network

In a low data rate regime, a constant gap approximation of the capacity may not be interesting any more. A more useful kind of approximation in this regime would be a universal multiplicative approximation (instead of additive), where the multiplicative factor does not depend on the channel gains in the network. For completeness, we restate our result for multiplicative approximation (Theorem 4.4.5) and then prove it.

Theorem 7.6.1. *Given a Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the capacity C satisfies*

$$\lambda \overline{C} \leq C \leq \overline{C} \tag{7.14}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity, as described in equation (4.1), and λ is a constant and is lower bounded by $\frac{1}{2d(d+1)}$ and d is the maximum degree of nodes in $\mathcal{G}$.

Proof. First we use a time division scheme and make all links in the network orthogonal to each other. By Vizing's theorem[3] any simple undirected graph can be edge colored with at most $d+1$ colors, where d is the maximum degree of nodes in $\mathcal{G}$. Since our graph $\mathcal{G}$ is a directed graph we need at most $2(d+1)$ colors. Therefore we can generate $2(d+1)$ time slots and assign the slots to directed graphs such that at any node all the links are orthogonal to each other. Therefore each link is used a $\frac{1}{2(d+1)}$ fraction of the time. We further impose the constraint that each of these links is used a total $\frac{1}{2d(d+1)}$ of the time but with d times more power. Now by coding we can convert each links $h_{i,j}$ into a noise free link with capacity

$$c_{i,j} = \frac{1}{2d(d+1)} \log(1 + d|h_{i,j}|^2) \tag{7.15}$$

[3] For example see [39] p.153

By Ford-Fulkerson theorem we know that the capacity of this network is

$$C_{\text{orthogonal}} = \min_{\Omega} \sum_{i,j:i\in\Omega,j\in\Omega^c} c_{i,j} \tag{7.16}$$

And this rate is achievable in the original Gaussian relay network. Now we will prove that

$$C_{\text{orthogonal}} \geq \frac{1}{2d(d+1)}\overline{C} \tag{7.17}$$

To show this assume in the orthogonal network each node transmit the same signal on its outgoing links, and also each node takes the summation of all incoming links (normalized by $\frac{1}{\sqrt{d}}$) and denote it as the received signal. Then the received signal at each node is j is

$$\begin{aligned} y_j[t] &= \frac{1}{\sqrt{d}}\sum_{i=1}^{d}\left(h_{ij}\sqrt{d}x_i[t] + z_{ij}[t]\right) && (7.18)\\ &= \sum_{i=1}^{d} h_{ij}x_i[t] + \tilde{z}_j[t] && (7.19)\end{aligned}$$

where

$$\tilde{z}_j[t] = \frac{\sum_{i=1}^{d} z_{ij}[t]}{\sqrt{d}} \sim \mathcal{CN}(0,1) \tag{7.20}$$

Therefore we get a network which is statically similar to the original non-orthogonal network, however each time-slot is only a $\frac{1}{d(d+1)}$ fraction of the time slots in the original network. Therefore without this restriction the cut-set of the orthogonal network can only increase. Hence

$$C_{\text{orthogonal}} \geq \frac{1}{2d(d+1)}\overline{C} \tag{7.21}$$

□

Chapter 8

Connections between models

8.1 Introduction

So far we have used the insights obtained from the deterministic channel model to be able to approximated the capacity of the Gaussian relay network. In this section we investigate whether the capacity of a relay network under these models are close to each other in some sense. First, we show that the generalized degrees of freedom of a Gaussian relay network and its corresponding linear finite-field deterministic relay network are the same. This verifies that the deterministic model is properly capturing the high SNR behavior of Gaussian networks. Next we investigate whether there is a non-asymptotic connection between these two models. As we illustrate in an example, the finite field operations of the linear finite field deterministic model does not allow a non-asymptotic connection between the two models, hence the capacity of a relay network under these two models can be far from each other. Motivated by this observation we propose another deterministic channel model, called the *truncated deterministic model*, such that the capacity of any Gaussian relay network and its corresponding truncated deterministic relay network are within a constant gap of each other, uniformly for all channel gains.

8.2 Connections between the linear finite field deterministic model and the Gaussian model

In this section we show the connection between the generalized degrees of freedom of a Gaussian relay network and the capacity of its corresponding linear finite field deterministic relay network. The generalized degrees of freedom was first defined in [40] for 2×2 interference channel. We first define a natural generalization of this for Gaussian relay networks.

Consider a Gaussian wireless relay network as defined in section 4.3. The conventional degrees of freedom characterizes how the capacity of this network grows as P, the individual average power constraint at nodes, increases. More formally

$$d = \lim_{P \to \infty} \frac{C}{\log P} \tag{8.1}$$

Note that in this formulation all channel gains h_{ij}'s are fixed and only P increases. It is easy to show that for any Gaussian relay network

$$d = \begin{cases} 1 & \text{if there is a path with non zero gains from } S \text{ to } D \text{ ,} \\ 0 & \text{otherwise.} \end{cases} \tag{8.2}$$

Therefore for relay networks degrees of freedom is a very coarse quantity of capacity, and only reveals whether there is a path with nonzero gains from the source to the destination. An intuitive explanation for this is the following: in this formulation while the channel gains are fixed the transmit power is increasing. As a result the ratio between the signal to noise ratio of different links in dB scale goes to one. Hence for sufficiently large P, signal to noise ratio of all links are almost the same (in dB scale) and as a result a pure routing solution achieves the maximum degrees of freedom.

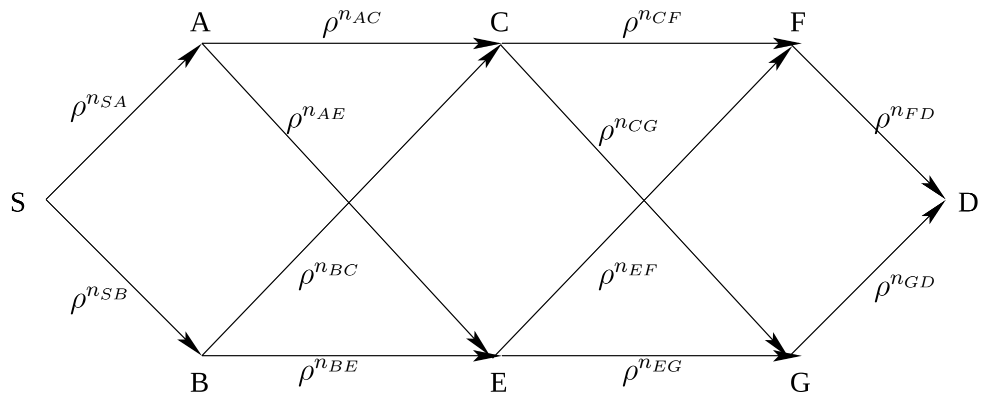


Figure 8.1: A three layer relay network.

Based on this observation we formulate the generalized degrees of freedom in such a way that while SNR of all links are increasing, their ratio in dB scale is fixed. In another words we look at a Gaussian relay network where all channel gains are in the form of $h_{ij} = \rho^{n_{ij}}$ for fixed integers $n_{ij} \in \mathcal{N} \cup \{0\}$ and a variable $\rho \in \mathcal{R}^+$. An example of such network is shown in figure 8.1. Now we define the generalized degrees of freedom of this network as the following

$$d(\{n_{ij}\}) = \lim_{\rho \to \infty} \frac{C(\rho, \{n_{ij}\})}{\log \rho} \tag{8.3}$$

The main result of this section is the following

Theorem 8.2.1. *Consider a gaussian network where all channel gains are in the form $\rho^{n_{ij}}$ where n_{ij}'s are non negative fixed integers and $\rho \in \mathcal{R}^+$ is a variable. There is also an average power constraint equal to 1 is at each node. Then the unicast capacity of this network from S to D satisfies:*

$$\lim_{\rho \to \infty} \frac{C}{\log \rho} = \min_{\Omega} \operatorname{rank}(\mathbf{H}_\Omega) \tag{8.4}$$

where the minimum is taken over all cuts, Ω, in the corresponding linear finite field deterministic relay network and $\mathbf{H}_\Omega$ is the transfer matrix associated with that cut and rank is *evaluated in $\mathcal{R}$.*

Proof. See Appendix A.8. □

8.3 Non asymptotic connection between the linear finite field deterministic model and the Gaussian model

In section 8.2 we illustrated an asymptotic connection between the linear finite field deterministic model and the Gaussian model in terms of the generalized degrees of freedom. Now, a natural question is whether there is a non-asymptotic (constant gap) connection between the capacity of the linear finite field deterministic model and the Gaussian model? Unfortunately, in this section we answer this question negatively. We show that the finite field operations of the linear finite field deterministic model does not allow a non-asymptotic connection between the two models

To show this we just need to look at a MIMO channel and show that the capacity of this system under the Gaussian model and the corresponding deterministic model can be very far from each other for some channel gain values. Consider a 2×2 MIMO real Gaussian channel with channel gain values as shown in Figure 8.2 (a), where k is an integer larger than 2. The channel gain parameters of the corresponding linear finite field deterministic model are:

$$n_{11} = \lceil \log_2 h_{11} \rceil^+ = \lceil \log_2 (2^k - 2^{k-2}) \rceil^+ = k \tag{8.5}$$

$$n_{12} = n_{21} = n_{22} = \lceil \log_2 2^k \rceil^+ = k \tag{8.6}$$

Now lets compare the capacity of the MIMO channel under these two models. The capacity

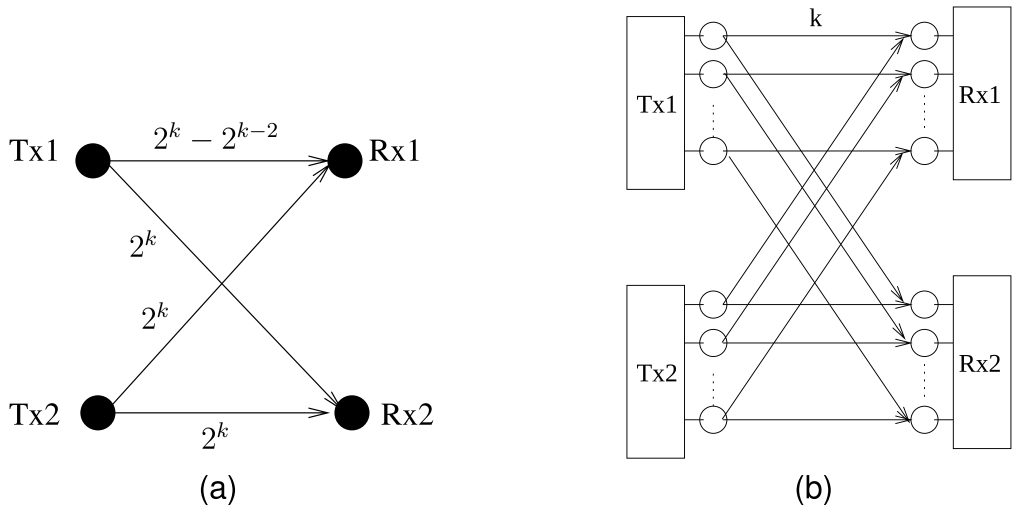


Figure 8.2: An example of a 2×2 Gaussian MIMO channel is shown in (a). The corresponding linear finite field deterministic MIMO channel is shown in (b).

of the Gaussian MIMO channel with equal power allocation is

$$C_{\text{Gaussian}} = \frac{1}{2}\log\left(\det\left(I + HH^t\right)\right) \tag{8.7}$$

where

$$H = \begin{pmatrix} 2^k - 2^{k-2} & 2^k \\ 2^k & 2^k \end{pmatrix} \tag{8.8}$$

Therefore for large k we have,

$$C_{\text{Gaussian}} = \frac{1}{2}\log\left(|\det\begin{pmatrix} (2^k - 2^{k-2})^2 + 2^{2k} + 1 & 2^{2k+1} - 2^{2k-2} \\ 2^{2k+1} - 2^{2k-2} & 2^{2k+1} + 1 \end{pmatrix}|\right) \tag{8.9}$$

$$= \frac{1}{2}\log\left(|1 + 2^{2k+1} + 9 \times 2^{2k-4} - 31 \times 2^{4k-4}||\right) \tag{8.10}$$

$$\approx 2k - 2 \tag{8.11}$$

However the capacity of the corresponding linear finite field deterministic MIMO is simply

$$C_{\text{LFF}} = \text{rank}\begin{pmatrix} I_k & I_k \\ I_k & I_k \end{pmatrix} = k \tag{8.12}$$

$$\tag{8.13}$$

Therefore by comparing (8.11) and (8.12), for large k we have

$$C_{\text{Gaussian}} - C_{\text{LFF}} \approx k - 2 \tag{8.14}$$

which goes to infinity as k increases.

Motivated by this observation we propose another deterministic channel model, called the truncated deterministic model, such that the capacity of any Gaussian relay network and its corresponding truncated deterministic relay network are within a constant gap of each other, uniformly for all channel gains.

8.4 Truncated deterministic model

The truncated deterministic model is a subclass of the deterministic model in which the received signal $\mathbf{y}_j$ at node $j \in \mathcal{V}$ and time t is given by

$$\mathbf{y}_j[t] = \left[\sum_{i \in \mathcal{N}_j} \mathbf{H}_{ij} x_i[t] \right] \tag{8.15}$$

where $\mathbf{H}_{ij}$ is a complex matrix where element represents the channel gain from a transmitting antenna in node i to a receiving antenna in node j, and $\mathcal{N}_j$ is the set of nodes that are neighbors of j in $\mathcal{G}$. Also the function [.] is defined as the following and is applied component wise to a vector,

$$[x_R + jx_I] = [x_R] + j[x_I] \tag{8.16}$$

where for any real number, $[x]$ represents its closest integer . Furthermore, we assume there is an average power constraint equal to 1 at each transmitter.

Clearly any Gaussian model has a corresponding truncated deterministic model. Compared to the Gaussian model, there is no channel noise in the truncated model. However, its effect is partially captured by the truncation of the received signal.

8.5 Connection between the truncated deterministic model and the Gaussian model

In this section we establish a more concrete connection between the truncated deterministic model and the Gaussian model. We show that the capacity of any relay network under these two models are within a constant gap of each other, where the constant does not depend on the channel parameters. Note that this implies a functional connection between the reliable transmission rates, but *not* any operational connection in terms of coding strategies between the two models.

Theorem 8.5.1. *Assume $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a Gaussian relay network with real gains and real Gaussian noise. The capacity of this relay network, $C_{Gaussian}$, and the capacity of the corresponding truncated deterministic model, $C_{Truncated}$, satisfy the following relationship*

$$|C_{Gaussian} - C_{Truncated}| \leq 13|\mathcal{V}| \tag{8.17}$$

To prove this Theorem first we need the following lemma,

Lemma 8.5.2. *Let G be the channel gains matrix of a $m \times n$ MIMO system. Assume that there is an average power constraint equal to one at each node. Then for any input distribution $P_{\mathbf{X}}$,*

$$|I(\mathbf{X}; G\mathbf{X} + Z) - I(\mathbf{X}; [G\mathbf{X}])| \leq 8n \tag{8.18}$$

where $Z = [z_1, \ldots, z_n]$ is a vector of n i.i.d. $\mathcal{N}(0,1)$ random variables.

Proof. Look at appendix A.7. □

Now we prove Theorem 8.5.1.

Proof. (proof of Theorem 8.5.1)
First note that the value of any cut in the network is the same as the mutual information of a MIMO system. Therefore from Lemma 8.5.2 we have

$$|\overline{C}_{\text{Gaussian}} - \overline{C}_{\text{Truncated}}| \leq 8|\mathcal{V}| \tag{8.19}$$

Now pick i.i.d normal distribution for $\{X_i\}_{i \in \mathcal{V}}$. Now by applying Theorem 4.2.3 to the truncated deterministic relay network

$$C_{\text{Truncated}} \geq \min_{\Omega \in \Lambda_D} I(Y_{\Omega^c}^{\text{truncated}}; X_\Omega | X_{\Omega^c}) \tag{8.20}$$

Now by Lemma 6.2.6 and Lemma 8.5.2 we know have the following

$$\min_{\Omega \in \Lambda_D} I(Y_{\Omega^c}^{\text{truncated}}; X_\Omega | X_{\Omega^c}) \geq I(Y_{\Omega^c}^{\text{Gaussian}}; X_\Omega | X_{\Omega^c}) - 8|\mathcal{V}| \tag{8.21}$$

$$\geq \overline{C}_{\text{Gaussian}} - 10|\mathcal{V}| \tag{8.22}$$

Then from equations (8.19) and (8.22) we have

$$\overline{C}_{\text{Gaussian}} - 10|\mathcal{V}| \leq C_{\text{Truncated}} \leq \overline{C}_{\text{Gaussian}} + 8|\mathcal{V}| \tag{8.23}$$

Also from main Theorem 4.3.1 we know that

$$\overline{C}_{\text{Gaussian}} - 5|\mathcal{V}| \leq C_{\text{Gaussian}} \leq \overline{C}_{\text{Gaussian}} \tag{8.24}$$

Therefore

$$|C_{\text{Gaussian}} - C_{\text{Truncated}}| \leq 13|\mathcal{V} \tag{8.25}$$

□

Chapter 9

Other applications of the deterministic approach

9.1 Introduction

So far we have considered the application of our deterministic approach to relay networks. However, it is a general approach that can be applied to other problems in wireless network information theory. The simplifications of the linear finite field deterministic model will allow us to focus more on the interaction between users and get insights in interference limited scenarios. In this section we discuss a few other applications of the deterministic approach. The first example is a variation on the relay channel problem. We study the capacity of the full-duplex bidirectional (two-way) relay channel, where a relay node is supporting the exchange of information between two nodes. We use the linear finite-field deterministic model to find a near optimal good transmission strategy for the relay. We analyze the achievable rate region of the proposed scheme and show that the scheme achieves to within 3 bits/sec/Hz the cut-set bound for all values of channel gains.

In the second example we consider transmission of a Gaussian source over a Gaussian relay channel, where the relay terminal has access to a correlated side information. Here there are two complications in the problem: complicated channel model and complicated source model. For the source model, again we can apply the linear finite-field deterministic channel model and simplify it. Quite interestingly, we also think of the dual of our deterministic channel model for sources. So we first propose a binary-expansion source model for Gaussian sources. Then we apply both the deterministic channel model together with the binary expansion source model to make progress in this problem. In particular we show that a simple cooperation scheme is uniformly near optimal for all range of channel gains and correlation factors.

9.2 Approximate capacity of the two-way relay channel

9.2.1 Introduction

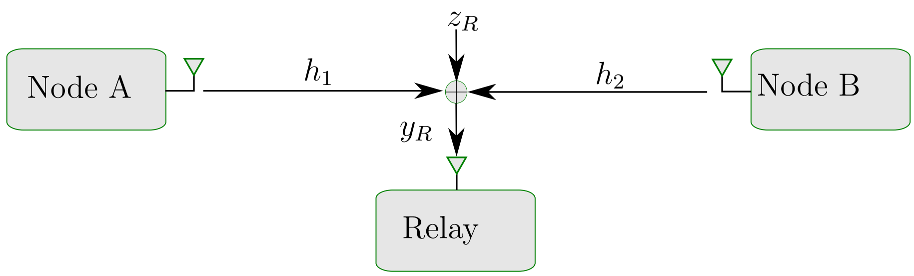
Bidirectional or two-way communication between two nodes was first studied by Shannon himself in [41]. Nowadays the two-way communication where an additional node acting as a relay is supporting the exchange of information between the two nodes is attracting increasing attention. Some achievable rate regions for the two-way relay channel using different strategies at the relay, such as decode-and-forward, compress-and-forward and amplify-and-forward, have been analyzed in [42]. Network coding techniques have been proposed by [43; 44; 45] (and others) in order to improve the transmission rate. While inferior to traditional routing at low signal-to-noise-ratios (SNR), it was shown in [46] that network coding achieves twice the rate of routing at high SNR. Similarly, in [47] the half-duplex two-way relay channel where the channel gains are all equal to one is investigated. It was shown that a combination of a decode-and-forward strategy using lattice codes and a joint decoding strategy is asymptotically optimal. Our work represents an alternative approach, however for the full-duplex case. Furthermore, here we analyze the general case, where the channel gains are all different (in general) and channel reciprocity is not assumed. The capacity region of the so called broadcast two-way half-duplex relay channel, i.e. assuming that the communication takes places in two hops and the relay is decoding the received messages completely, was recently characterized in [48].

The main focus is, however, so far on the one-way relay channel. Such cooperative communication schemes are particularly important when reliable communication can not be guaranteed by using a conventional point-to-point connection. Cooperation between two source nodes for communication to a common receiver was proposed in [49]. There, a non-cooperative phase is followed by a cooperative one and it is shown that this strategy outperforms non-cooperative strategies. Cooperation by using distributed space-time coding techniques in networks has been analyzed in [50; 51]. Relaying can be expected to be adopted in current and future wireless systems, as it has been introduced in the 802.16j (WiMAX) standard.

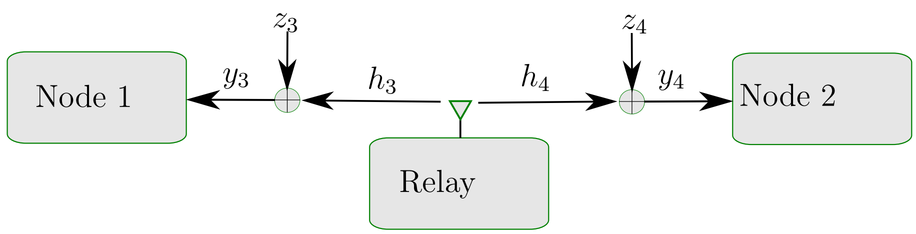
In this section, we study the capacity of the full-duplex two-way relay channel, which, to the best of our knowledge, is not known in general. Similar to the general relay network, we show that our scheme can achieve to within 3 bits/sec/Hz of the capacity for all channel parameter values.

9.2.2 System model

The system model of the two-way full-duplex relay channel is shown in Fig. 9.1. Communication takes place simultaneously from the relay to the nodes and vice versa. As can be observed from Fig. 9.1, channel reciprocity is not assumed here. Thus, in general h_1, which is the channel parameter describing the link from node A to the relay, is different from h_3, the channel describing the link from the relay to node A (and similarly for h_2 and h_4).



(a) Communication to the relay



(b) Communication from the relay

Figure 9.1: Bidirectional relaying

The received signal at the relay is given by (cf.Fig. 9.1(a))

$$y_R = h_1 x_A + h_2 x_B + z_R, \tag{9.1}$$

where x_A and x_B are the signals transmitted from node A and node B, respectively. The variable z_R describes the additive Gaussian noise at the relay. Without loss of generality, we assume that $\mathbb{E}\left[|x_A|^2\right] = \mathbb{E}\left[|x_B|^2\right] = \mathbb{E}\left[|z_R|^2\right] = 1$. The received signals at the nodes are given by (cf. Fig. 9.1(b))

$$\begin{aligned} y_A &= h_3 x_R + z_2 \\ y_B &= h_4 x_R + z_3. \end{aligned} \tag{9.2}$$

The variables z_2 and z_3 are the unit variance additive Gaussian noises at node A and node B, respectively.

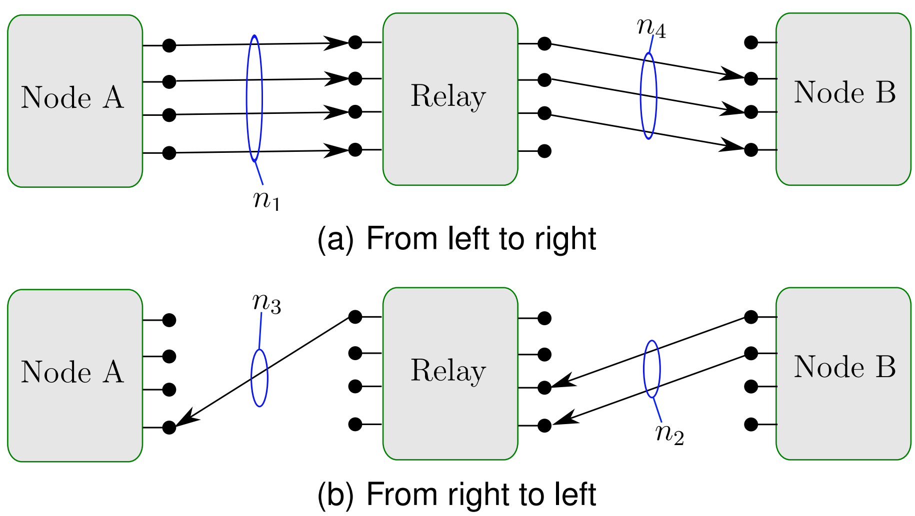


Figure 9.2: Deterministic model for bidirectional relaying

9.2.3 Deterministic two-way relay channel

We start our analysis by considering the deterministic model of the two-way relay channel as shown in Fig. 9.2.

The following theorem is our main result for the deterministic two-way relay network.

Theorem 9.2.1. *The capacity region of the bi-directional linear finite field deterministic relay network is:*

$$R_{AB} \leq \min(n_1, n_4) \tag{9.3}$$

$$R_{BA} \leq \min(n_2, n_3) \tag{9.4}$$

Furthermore, the cut-set bound is achievable with a simple shift-and-forward strategy at the relay.

In the rest of the section, we give a sketch of the proof. We use an algebraic approach to solve the problem of finding the optimal strategy. In the deterministic model assume that node A and B sends $\mathbf{x}_A$ and $\mathbf{x}_B \in \mathbb{F}_2^q$, respectively, where $q = \max(n_1, n_2, n_3, n_4)$. The received signal at the relay is then given by

$$\mathbf{y}_R = \mathbf{S}^{q-n_1}\mathbf{x}_A + \mathbf{S}^{q-n_2}\mathbf{x}_B. \tag{9.5}$$

Now consider a linear coding strategy at the relay. Therefore, it is going to send

$$\mathbf{x}_R = \mathbf{G}\mathbf{y}_R = \mathbf{G}(\mathbf{S}^{q-n_1}\mathbf{x}_A + \mathbf{S}^{q-n_2}\mathbf{x}_B), \tag{9.6}$$

where $\mathbf{G}$ is an arbitrary $q \times q$ generating matrix that is a design choice.

The received signal at node A is thus given by

$$\mathbf{y}_A = \mathbf{S}^{q-n_3}\mathbf{x}_R = \mathbf{S}^{q-n_3}\mathbf{G}(\mathbf{S}^{q-n_1}\mathbf{x}_A + \mathbf{S}^{q-n_2}\mathbf{x}_B) \tag{9.7}$$

while node B receives

$$\mathbf{y}_B = \mathbf{S}^{q-n_4}\mathbf{x}_R = \mathbf{S}^{q-n_4}\mathbf{G}(\mathbf{S}^{q-n_1}\mathbf{x}_A + \mathbf{S}^{q-n_2}\mathbf{x}_B). \tag{9.8}$$

Since node A and node B respectively know their own signals $\mathbf{x}_A$ and $\mathbf{x}_B$, they can cancel it from their received signal. Hence effectively they receive

$$\mathbf{y}'_A = \mathbf{S}^{q-n_3}\mathbf{G}\mathbf{S}^{q-n_2}\mathbf{x}_B \tag{9.9}$$

$$\mathbf{y}'_B = \mathbf{S}^{q-n_4}\mathbf{G}\mathbf{S}^{q-n_1}\mathbf{x}_A \tag{9.10}$$

The question is, whether we can find a matrix $\mathbf{G}$, such that the rates $R_{AB} = \min(n_1, n_4)$ and $R_{BA} = \min(n_2, n_3)$ in (9.3) and (9.4) are achievable. By obtaining such a matrix, we would also gain insights how the processing at the relay should be done in an optimal way.

Now we state the following lemma, whose prove is omitted due to the lack of space.

Lemma 9.2.2. *It is possible to convert the network in Fig. 9.2 into one of the following two cases without changing the cut-set bound.*

1. $n_1 = n_4$ *and* $n_2 \leq n_3$
2. $n_2 = n_3$ *and* $n_1 \leq n_4$

Therefore, by Lemma 9.2.2 and symmetry we only need to study this case:

$$n_1 = n_4 \text{ and } n_2 \leq n_3 \tag{9.11}$$

It turns out, that we are indeed able to construct a matrix $\mathbf{G}$, such that the cut-set bound is achievable. The generating matrices $\mathbf{G}$ for the individual cases are given as follows.

1. $q = n_1$

$$\mathbf{G} = \begin{bmatrix} \mathbf{0}_{n_2\times(q-n_2)} & \mathbf{I}_{n_2} \\ \mathbf{I}_{q-n_2} & \mathbf{0}_{(q-n_2)\times n_2} \end{bmatrix} \tag{9.12}$$

2. $q = n_3$

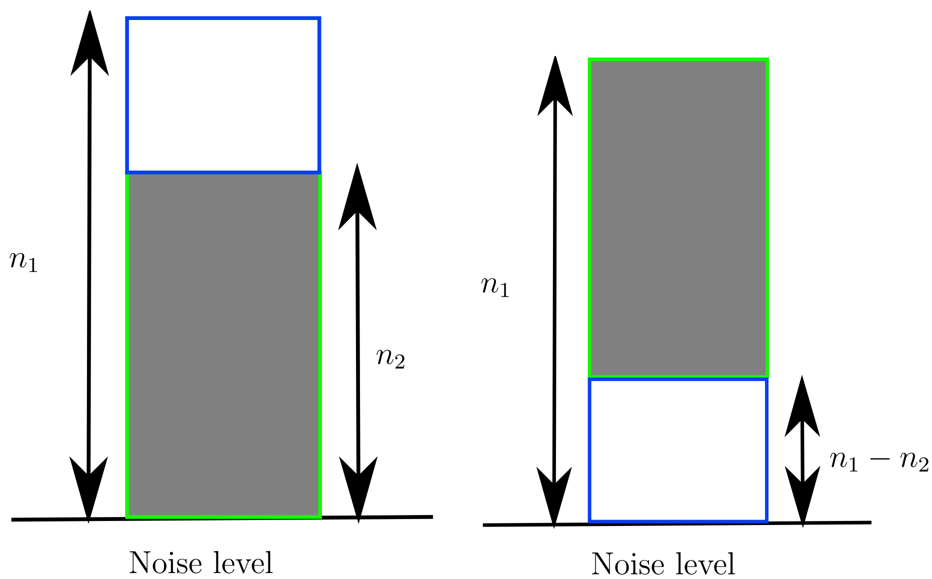
(a) $n_2 \leq n_1$

$$\mathbf{G} = \begin{bmatrix} \mathbf{0}_{n_1\times(q-n_1)} & \mathbf{I}_{n_1} \\ \mathbf{0}_{q-n_1} & \mathbf{0}_{(q-n_1)\times n_1} \end{bmatrix} \tag{9.13}$$

(b) $n_2 > n_1$

$$\mathbf{G} = \begin{bmatrix} \mathbf{0}_{n_1\times(q-n_1)} & \mathbf{I}_{n_1} \\ \mathbf{I}_{q-n_1} & \mathbf{0}_{(q-n_1)\times n_1} \end{bmatrix} \tag{9.14}$$

In the following we give interpretations of the different generating matrices $\mathbf{G}$ in (9.12), (9.13), and (9.14) for the three cases.



(a) Received signal at Relay (b) Transmit signal from Relay

Figure 9.3: Signal levels at relay: Receive phase and transmit phase

9.2.3.1 Interpretation of the case $n_1 = q$

We start with the generating matrix $\mathbf{G}$ in (9.12). The interpretation of this operation for the deterministic case is the following. The relay receives $n_1 = q$ signal levels as shown in Fig. 9.3(a). The last n_2 contain information from both node A and node B (gray area in Fig. 9.3(a)) and the other (top) signal levels are only information from A (white area in Fig. 9.3(a)). The relay is now creating a codeword, which has the last n_2 received signal levels at highest level (gray area in Fig. 9.3(b)) and the remaining bits of A at lower signal levels (white area in Fig. 9.3(a)).

The interpretation of the scheme for Gaussian is the following. At first the relay decodes a part of the message, namely $\mathbf{x}_A^{(1)}$, received from node A that has arrived above the signal level of node B and subtracts it from the overall received signal. The remaining part (lowest n_2 levels) of the overall received signal at the relay is just the summation of signals from both the node A and node B. The argumentation here is that the relay can not decode this summation and thus it quantize it. The interesting part is now that the relay creates the transmit signal by using a superposition code [52]. The cloud center of this superposition code is the quantized signal, while the bin index is the information $\mathbf{x}_A^{(1)}$ it has decoded from node A.

9.2.3.2 Interpretation of the case $n_3 = q$

We start with the case $n_2 \leq n_1$. Here, the relay receives n_1 signal levels. The relay then simply shifts the received signal up and forwards it. The corresponding scheme for Gaussian is thus amplify-forward. As an alternative approach, we could also use a similar superposition strategy as in the case of $n_1 = q$. However, as we will show later on, the simple amplify-and-forward strategy is enough in order to achieve to within 3 bits the capacity for all channel parameter values.

The case with $n_2 > n_1$ is analogous to the case with $n_1 = q$. Here the relay receives n_2 signal levels. The last n_1 bits contain information for both node A and B and the rest is just the information for node A. The interpretation of the scheme for Gaussian is very similar to the scheme for $n_1 = q$ and thus omitted.

9.2.4 Gaussian Two-way relay channel

In this section, we use the insights obtained from studying the deterministic two-way relay channel to find near-optimal relaying strategies in the Gaussian case as defined in section 9.2.2. It follows our main result for the Gaussian two-way relay channel and the rest of this section is devoted to proving it.

Theorem 9.2.3. *Consider a Gaussian two-way relay channel as defined in section 9.2.2 with unit average noise and transmit power at each node. The capacity of this system* *satisfies*

$$\bar{C}_{AB} - 3 \leq C_{AB} \leq \bar{C}_{AB} \tag{9.15}$$

and

$$\bar{C}_{BA} - 3 \leq C_{BA} \leq \bar{C}_{BA}, \tag{9.16}$$

where $\bar{C}_{AB} = \log(1+\min(|h_1|^2, |h_4|^2))$ *and* $\bar{C}_{BA} = \log(1+\min(|h_2|^2, |h_3|^2))$ *is the cut-set upper bound on the capacity of the transmission from A to B and B to A, respectively [33].*

Since Lemma 9.2.2 holds also for the Gaussian case, we again need to study only the case that $|h_1|^2 = |h_4|^2$ and $|h_2|^2 \leq |h_3|^2$. Now we discuss the achievability strategy:

9.2.4.1 Achievability strategy

In general, the transmit signals from node A, node B and the relay are given by

$$\mathbf{x}_A = \sqrt{\alpha_A}\mathbf{x}_A^{(1)} + \sqrt{1-\alpha_A}\mathbf{x}_A^{(2)} \tag{9.17}$$

$$\mathbf{x}_B = \sqrt{\alpha_B}\mathbf{x}_B^{(1)} + \sqrt{1-\alpha_B}\mathbf{x}_B^{(2)} \tag{9.18}$$

$$\mathbf{x}_R = \sqrt{\alpha_R}\mathbf{x}_R^{(1)} + \sqrt{1-\alpha_R}\mathbf{x}_R^{(2)}. \tag{9.19}$$

where $\mathbf{x}_A^{(1)}$, $\mathbf{x}_A^{(2)}$, $\mathbf{x}_B^{(1)}$, $\mathbf{x}_B^{(2)}$, $\mathbf{x}_R^{(1)}$, and $\mathbf{x}_R^{(2)}$ are codewords chosen from a random Gaussian codebook of size $2^{nR_{AB}^{(1)}}$, $2^{nR_{AB}^{(2)}}$, $2^{nR_{BA}^{(1)}}$, $2^{nR_{BA}^{(2)}}$, $2^{nR_R^{(1)}}$, and $2^{nR_R^{(2)}}$, respectively. At node A (and similarly for node B) we have two messages $\mathbf{m}_A^{(1)}$ and $\mathbf{m}_A^{(2)}$ of size $2^{nR_{AB}^{(1)}}$ and $2^{nR_{AB}^{(2)}}$ that are mapped to $x_A^{(1)}$ and $x_A^{(2)}$, respectively. The relay signaling strategy depends on the channel gains and will be specificized later for each case. The choice of α_A, α_B, and α_R depend on the magnitude of the channel gains $|h_1|$, $|h_2|$, $|h_3|$, and $|h_4|$.

9.2.4.2 $|h_1|^2 \geq |h_3|^2$

Following the insights gained from the deterministic model, for $|h_1|^2 \geq |h_3|^2$ we set $\alpha_B = 0$ and $R_{BA}^{(1)} = 0$. The transmit signal at node B then reduces to

$$\mathbf{x}_B = \mathbf{x}_B^{(2)}. \tag{9.20}$$

Thus, the receive signal at the relay is given by

$$\mathbf{y}_R = \left(\sqrt{\alpha_A}\mathbf{x}_A^{(1)} + \sqrt{1-\alpha_A}\mathbf{x}_A^{(2)}\right) h_1 + h_2\mathbf{x}_B + \mathbf{z}_R \tag{9.21}$$

α_A is chosen such that the received signal of $\mathbf{x}_A^{(2)}$ and $\mathbf{x}_B$ are at the same scale. Thus, the following expression has to hold

$$\sqrt{1-\alpha_A}h_1 = h_2, \tag{9.22}$$

which gives

$$1 - \alpha_A = \left(\frac{h_2}{h_1}\right)^2. \tag{9.23}$$

Form $\mathbf{y}_R$, the relay first decodes $\mathbf{x}_A^{(1)}$ (i.e. $\mathbf{m}_A^{(1)}$) by treating the remaining received signals $\mathbf{x}_A^{(2)}$ and $\mathbf{x}_B$ as noise. This can be done with low error probability as long as

$$R_{AB}^{(1)} \leq \log\left(1 + \frac{\alpha_A|h_1|^2}{1 + (1-\alpha_A)\,|h_1|^2 + |h_2|^2}\right) = \log\left(1 + \frac{|h_1|^2 - |h_2|^2}{1 + 2|h_2|^2}\right) \tag{9.24}$$

Then the relay maps the decoded $\mathbf{x}_A^{(1)}$ to another codeword $\mathbf{x}_R^{(1)}$ of size $2^{nR_R^{(1)}}$ with $R_R^{(1)} = R_{AB}^{(1)}$. If the above expression is fulfilled, the relay can decode the signal $\mathbf{x}_A^{(1)}$ and cancel it from the received signal in (9.21). Thus, we have

$$\tilde{\mathbf{y}}_R = \sqrt{1-\alpha_A}\mathbf{x}_A^{(2)} h_1 + h_2 \mathbf{x}_B + \mathbf{z}_R \tag{9.25}$$

As suggested in the deterministic model, $\tilde{\mathbf{y}}_R$ is not decoded. Rather, a quantization is performed. The relay uses an optimal vector quantizer of size $2^{nR_R^{(2)}}$ and maps the quantization index to a codeword $\mathbf{x}_R^{(2)}$. Then the relay transmits (9.19), where

$$\alpha_R = \frac{\alpha_A}{2|h_2|^2 + 1}. \tag{9.26}$$

Now nodes A and B first attempt to decode $\mathbf{x}_R^{(2)}$. Since node A knows $\mathbf{x}_R^{(1)}$ it can cancel it from the received signal, however node B is treating $\mathbf{x}_R^{(1)}$ as noise. The decoding of $\mathbf{x}_R^{(2)}$ can be done with low error probability as long as

$$R_R^{(2)} \leq \min\left(\log\left(1 + \frac{|h_1|^2(1-\alpha_R)}{|h_1|^2\alpha_R + 1}\right), \log\left(1 + |h_3|^2(1-\alpha_R)\right)\right) \tag{9.27}$$

$$= \min\left(\log\left(\frac{|h_1|^2 + 1}{|h_1|^2\alpha_R + 1}\right), \log\left(1 + |h_3|^2(1-\alpha_R)\right)\right) \tag{9.28}$$

The second expression is obtained due to node A. By assuming that node A knows the strategy of relay and the codebook it has used, it can reconstruct $\mathbf{x}_R^{(1)}$ perfectly, since it contains only its own message. Using interference cancellation results in a interference free channel. The first expression is obtained due to node B which observes part of the signal from the relay, i.e. $\mathbf{x}_R^{(1)}$, as additional noise. Then node B cancels $\mathbf{x}_R^{(2)}$ from its received signal and attempts to decode $\mathbf{x}_R^{(1)}$. This can be done with low error probability if

$$R_R^{(1)} \leq \log\left(1 + \alpha_R |h_1|^2\right). \tag{9.29}$$

Now that nodes A and B have decoded $\mathbf{x}_R^{(1)}$, they can create

$$\tilde{\mathbf{y}}_R^Q = \beta\tilde{\mathbf{y}}_R + \mathbf{z}_Q = \beta\left(\sqrt{1-\alpha_A}\mathbf{x}_A^{(2)}h_1 + h_2\mathbf{x}_B + \mathbf{z}_R\right) + \mathbf{z}_Q \tag{9.30}$$

$$\overset{(9.22)}{=} \beta\left(h_2\left(\mathbf{x}_A^{(2)} + \mathbf{x}_B\right) + \mathbf{z}_R\right) + \mathbf{z}_Q \tag{9.31}$$

where

$$\beta = (1 - D/\sigma^2_{\tilde{\mathbf{y}}_R}) \tag{9.32}$$

and $\mathbf{z}_Q$ is due to the quantization noise with variance

$$\sigma_Q^2 = D(1 - D/\sigma^2_{\tilde{\mathbf{y}}_R}): \tag{9.33}$$

Thus, the distortion D in our case has to fulfill

$$D = 2^{-R_R^{(2)}}\sigma^2_{\tilde{y}_R} = \min\left(\frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}, \frac{1}{1+|h_3|^2(1-\alpha_R)}\right)\left(2|h_2|^2+1\right) \tag{9.34}$$

Assuming that the nodes are able to cancel the own message from $\tilde{\mathbf{y}}_R^Q$, they can decode each others codeword with low error probability if

$$R_{BA} \leq \min\left(\log\left(1 + \frac{|h_2|^2\left(1 - \frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}\right)}{1 + \frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}2|h_2|^2}\right), \log\left(1 + \frac{|h_2|^2\left(1 - \frac{1}{1+|h_3|^2(1-\alpha_R)}\right)}{1 + \frac{2|h_2|^2}{1+|h_3|^2(1-\alpha_R)}}\right)\right) \tag{9.35}$$

and

$$R_{AB}^{(2)} \leq \min\left(\log\left(1+\frac{|h_2|^2\left(1-\frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}\right)}{1+\frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}2|h_2|^2}\right), \log\left(1+\frac{|h_2|^2\left(1-\frac{1}{1+|h_3|^2(1-\alpha_R)}\right)}{1+\frac{2|h_2|^2}{1+|h_3|^2(1-\alpha_R)}}\right)\right) \tag{9.36}$$

Therefore the rate in (9.35) and

$$\begin{aligned} R_{AB} \overset{(9.24),(9.36)}{\leq} & \log\left(1+\frac{|h_1|^2-|h_2|^2}{1+2|h_2|^2}\right)+\min\{\log\left(1+\frac{|h_2|^2\left(1-\frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}\right)}{1+\frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}2|h_2|^2}\right), \\ & , \log\left(1+\frac{|h_2|^2\left(1-\frac{1}{1+|h_3|^2(1-\alpha_R)}\right)}{1+\frac{2|h_2|^2}{1+|h_3|^2(1-\alpha_R)}}\right)\} \end{aligned} \tag{9.37}$$

are achievable.

With some algebra, we can show that

$$\begin{aligned} & \min\left(\log\left(1+\frac{|h_2|^2\left(1-\frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}\right)}{1+\frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}2|h_2|^2}\right), \log\left(1+\frac{|h_2|^2\left(1-\frac{1}{1+|h_3|^2(1-\alpha_R)}\right)}{1+\frac{2|h_2|^2}{1+|h_3|^2(1-\alpha_R)}}\right)\right) \\ & \geq \log\left(1+|h_2|^2\right)-\log(3) \end{aligned} \tag{9.38}$$

and

$$\begin{aligned} & \log\left(1+\frac{|h_1|^2-|h_2|^2}{1+2|h_2|^2}\right)+ \\ & \quad +\min\left(\log\left(1+\frac{|h_2|^2\left(1-\frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}\right)}{1+\frac{\alpha_R|h_1|^2+1}{|h_1|^2+1}2|h_2|^2}\right), \log\left(1+\frac{|h_2|^2\left(1-\frac{1}{1+|h_3|^2(1-\alpha_R)}\right)}{1+\frac{2|h_2|^2}{1+|h_3|^2(1-\alpha_R)}}\right)\right) \\ & \geq \log\left(1+|h_1|^2\right)-\max(2,3). \end{aligned} \tag{9.39}$$

Thus, we are at most 3 bits away from the cut-set bound.

9.2.4.3 Case $|h_1|^2 < |h_3|^2$

- Amplify-and-forward: $|h_2|^2 < |h_1|^2$

  With $\alpha_A = \alpha_B = 0$, the transmit signals from node A and node B reduce to $\mathbf{x}_A = \mathbf{x}_A^{(2)}$ and $\mathbf{x}_B = \mathbf{x}_B^{(2)}$ chosen from a random Gaussian codebook of size $2^{nR_{AB}}$ and $2^{nR_{BA}}$, respectively. Thus, the received signal at the relay is given by

$$\mathbf{y}_R = h_1\mathbf{x}_A + h_2\mathbf{x}_B + \mathbf{z}_R \tag{9.40}$$

  Using a amplify and forward strategy, the transmit signal at the relay is thus given by

$$\mathbf{x}_R = \frac{1}{\sqrt{|h_1|^2 + |h_2|^2 + 1}}\mathbf{y}_R. \tag{9.41}$$

  Using (9.2), the received signals at the nodes are given by

$$\mathbf{y}_A = \frac{h_3}{\sqrt{|h_1|^2 + |h_2|^2 + 1}}\left(h_1\mathbf{x}_A + h_2\mathbf{x}_B + \mathbf{z}_R\right) + \mathbf{z}_A \tag{9.42}$$

$$\mathbf{y}_B = \frac{h_4}{\sqrt{|h_1|^2 + |h_2|^2 + 1}}\left(h_1\mathbf{x}_A + h_2\mathbf{x}_B + \mathbf{z}_R\right) + \mathbf{z}_B. \tag{9.43}$$

  Assuming that the nodes are able to decode their own messages successfully and cancel it from the received signal. Therefore the nodes can decode each other signals with low error probability as long as

$$R_{AB} \leq \log\left(1 + \frac{|h_1|^2|h_4|^2}{|h_4|^2 + |h_1|^2 + |h_2|^2 + 1}\right) \tag{9.44}$$

$$R_{BA} \leq \log\left(1 + \frac{|h_2|^2|h_3|^2}{|h_3|^2 + |h_1|^2 + |h_2|^2 + 1}\right). \tag{9.45}$$

With some algebra, we can show that

$$\log\left(1+\frac{|h_1|^2|h_4|^2}{|h_4|^2+|h_1|^2+|h_2|^2+1}\right) \geq \log\left(1+|h_1|^2\right)-\log(3) \tag{9.46}$$

$$\log\left(1+\frac{|h_2|^2|h_3|^2}{|h_3|^2+|h_1|^2+|h_2|^2+1}\right) \geq \log\left(1+|h_2|^2\right)-\log(3) \tag{9.47}$$

Thus, we are at most within $\log(3)$ bits away from the cut-set bound, which is strictly better than we aimed for.

- $|h_2|^2 > |h_1|^2$:

The following derivations are very similar to the case $|h_1|^2 \geq |h_3|^2$ with slight differences. First of all, $\alpha_A = 0$ and α_B and α_R are now given by

$$\alpha_B = 1-\frac{|h_1|^2}{|h_2|^2} \text{ and } \alpha_R = \frac{\alpha_B|h_2|^2}{|h_3|^2\left(2|h_1|^2+1\right)}. \tag{9.48}$$

While we had a $\min$-operator in the case $|h_1|^2 > |h_3|^2$, here it can be shown that $|h_1|^2 > {}^{|h_3|^2}/_{\left(\alpha_R|h_3|^2+1\right)}$ is never fulfilled in this case. Thus, we have to consider only $|h_1|^2 \leq {}^{|h_3|^2}/_{\left(|h_3|^2\alpha_R+1\right)}$ and the $\min$-operator is obsolete. Therefore the nodes can decode each other signals with low error probability as long as

$$R_{AB} \leq \log\left(1+\frac{|h_1|^2\left(1-\frac{1}{1+|h_1|^2(1-\alpha_R)}\right)}{1+\frac{2|h_1|^2}{1+|h_1|^2(1-\alpha_R)}}\right) \tag{9.49}$$

and

$$R_{BA} \leq \log\left(\frac{1+|h_1|^2+|h_2|^2}{1+2|h_1|^2}\right)+R_{AB}. \tag{9.50}$$

With some algebra, we can show that

$$\log\left(1+\frac{|h_1|^2\left(1-\frac{1}{1+|h_1|^2(1-\alpha_R)}\right)}{1+\frac{2|h_1|^2}{1+|h_1|^2(1-\alpha_R)}}\right)\geq \log\left(1+|h_1|^2\right)-\log(3) \tag{9.51}$$

and

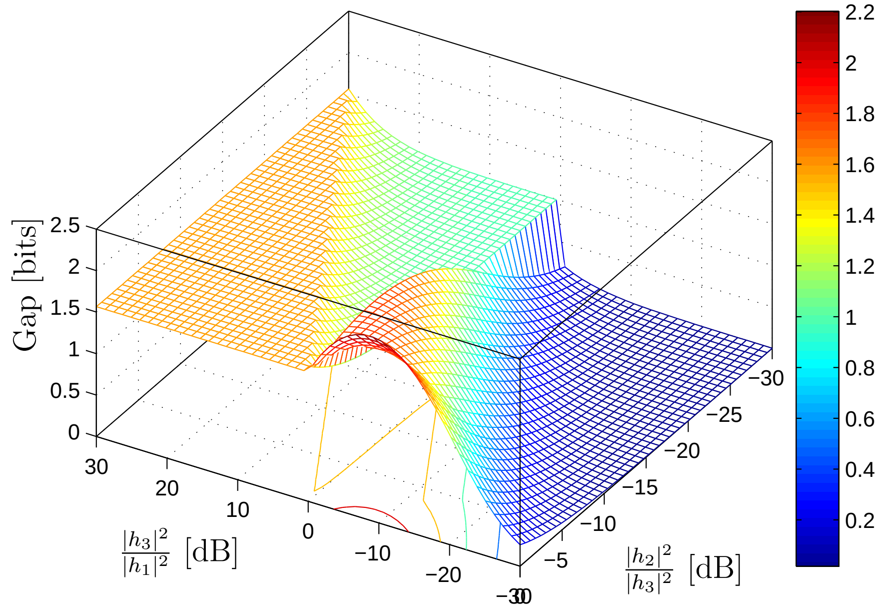
$$\log\left(\frac{1+|h_1|^2+|h_2|^2}{1+2|h_1|^2}\right)+R_{AB}\geq \log\left(1+|h_2|^2\right)-3 \tag{9.52}$$

Thus, we are at most 3 bits away from the cut-set bound.

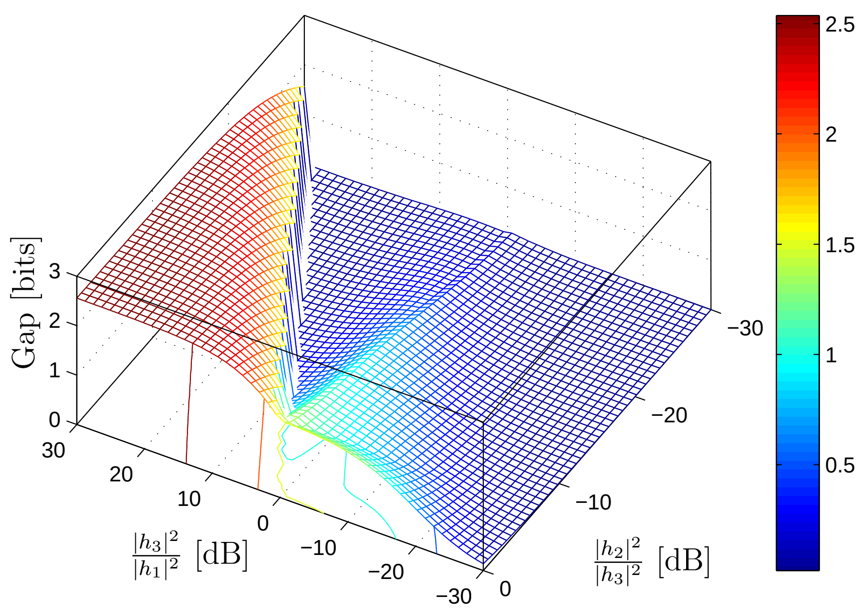
9.2.5 Illustration

In Fig. 9.4(a) and 9.4(b), the gap between the rates R_{AB} and R_{BA} and the corresponding cut-set upper bound is plotted for different channel gains, respectively.

The x-coordinate is representing the ratio of the channel gain from the relay to node A (i.e. h_3) to the reverse direction, i.e. from node A to the relay (i.e. h_1), in dB scale. On the y-coordinate we have the ratio of the channel and from the node B to the relay (i.e. h_2) to the reverse direction, i.e. from the relay to node B (i.e. $h_4 = h_1$), in dB scale. The ordinate shows the gap in bits. From the simulations, we observe that the gap is in general less than 3 bits, which verifies our theoretical results. We also observe that for a certain region, the gap is less than 1 bit. This region is especially large for R_{BA}. In the plot, we normalized the channel gain h_1 to 20 dB higher than the noise variance. Interestingly, it turns out that the gap is further reduced by shrinking the channel gain h_1 (not shown here).



(a) Gap for R_{AB}



(b) Gap for R_{BA}

Figure 9.4: Gap to the cut-set upper bound

9.3 Deterministic binary-expansion model for Gaussian sources

So far we have considered the binary-expansion deterministic model for channels. However, one can think of the dual of this model for sources. In this section we investigate this for Gaussian sources. Then in the next section we demonstrate an application of this model to the cooperative relaying with side information problem.

Assume u and v are two correlated Gaussian sources with mean zero and covariance matrix

$$\mathrm{Cov}\,[uv] = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{9.53}$$

Then we can write

$$v = \rho u + z \tag{9.54}$$

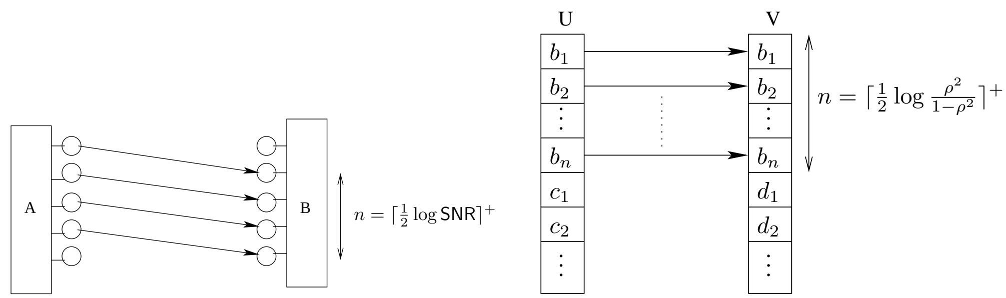
where $z \sim \mathcal{N}(0, 1-\rho^2)$ is independent of u. Therefore we can relate u and v through a Gaussian channel with

$$\mathsf{SNR} = \frac{\rho^2}{1-\rho^2} \tag{9.55}$$

The deterministic linear finite field model for this Gaussian channel is as shown in Figure 9.5(a) with channel strength

$$n = \lceil \frac{1}{2} \log \mathsf{SNR} \rceil^+ = \lceil \frac{1}{2} \log \frac{\rho^2}{1-\rho^2} \rceil^+ \tag{9.56}$$

Now we can create a deterministic binary-expansion model for these two Gaussian sources. Assume u and v are positive numbers and consider their binary expansion. In a hight correlation regime that $|\rho|$ is very close to one, those bits of u and v that are above the signal level of w are more or less the same. Therefore we can build the deterministic binary-expansion model for these two Gaussian sources u and v by representing each source as a



(a) Channel model (b) Source model

Figure 9.5: The deterministic linear finite filed model for point-to-point channel is shown in (a). The deterministic binary-expansion model for two sources is shown in (b)

sequence of bits, denoted by U and V. Then the correlation between the sources determines the number of first MSB's that are the same between them. This is pictorially shown in Figure 9.5(b).

Similar to the squared error distortion measure, we define the following measure for the distortion between two deterministic sources U and V,

$$d(U, V) = \sum_{i=1}^{q} (U(i) - V(i))^2 2^{-2(i-1)} \tag{9.57}$$

where $U(i)$ and $V(i)$ are respectively the i-th bit in U and V respectively. To verify that it is a proper distance measure we just need to show that for any three binary-expansion sources U, V and W

$$d(U, V) + d(V, W) \geq d(U, W) \tag{9.58}$$

To show this it is sufficient to note that at each level we have

$$(U(i) - V(i))^2 + (V(i) - W(i))^2 \geq (U(i) - W(i))^2 \tag{9.59}$$

Clearly this model can be extended to model M sources. In this case the correlation between source i and j is modeled with n_{ij} that represents the number of matching MSB's between them.

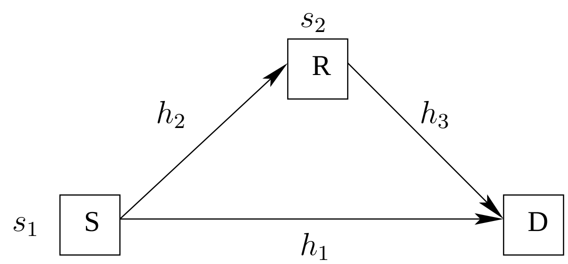
9.4 Cooperative relaying with side information

In this section we consider transmission of a Gaussian source over a Gaussian relay channel, where the relay terminal has access to correlated side information. In [53] authors studied this problem and proposed several cooperative joint source-channel coding strategies. However, still the best cooperative source channel coding strategy is not known. Here we use the linear finite field deterministic channel model together with the binary-expansion source model to make progress in this problem. In particular we show that a simple scheme is uniformly near optimal for all range of channel gains and correlation factors.

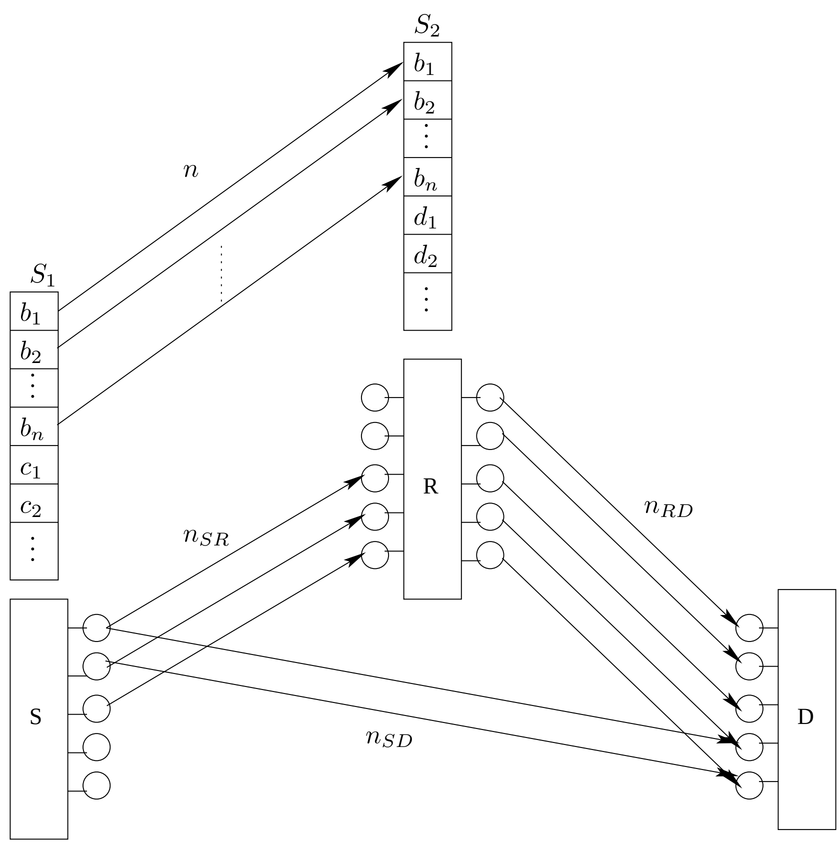
The system model for this problem is shown in figure 9.6(a). We assume that two zero-mean jointly Gaussian sources s_1 and s_2 generate the i.i.d. sequence $\{s_{1,k}, s_{2,k}\}_{k=1}^{\infty}$. The sequences $s_{1,k}$ and $s_{2,k}$ are available at the source and relay encoders, respectively. The covariance matrix of the sources is

$$\text{Cov}\left[s_1 s_2\right] = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{9.60}$$

The destination is interested in estimating s_1 and the problem is to find the minimum possible average distortion and the scheme that achieves it. Here we use our deterministic channel and source models to approximate the best possible distortion within a factor of 6 and also find a simple near optimal cooperative strategy.



(a) Gaussian model



(b) Deterministic model

Figure 9.6: The Gaussian model and the binary expansion model for cooperative relaying with side information are respectively shown in (a) and (b).

9.4.1 Analysis of the deterministic problem

The linear finite field deterministic channel model together with the binary expansion source model for this problem is shown in Figure 9.6(a). To minimize the distortion, we should find a scheme that provides the maximum possible number of MSB's of S_1 to the destination. From the cut-set upper bound we know that the maximum number of MSB's of S_1 that the destination can recover is upper bounded by

$$\overline{C}_{\text{with side information}} = \min\left(n + \max(n_{SD}, n_{SR}), \max(n_{SD}, n_{RD})\right) \tag{9.61}$$

Now there is a simple strategy to achieve this upper bound:

- Relay uses its correlated observation to send r_R number of bits at its top signal levels to the destination, where r_R is

$$r_R = \min\left(n + \max(n_{SD}, n_{SR}), \max(n_{SD}, n_{RD})\right) - \min\left(\max(n_{SD}, n_{SR}), \max(n_{SD}, n_{RD})\right) \tag{9.62}$$

- Once we remove the relay's r_R signal levels, the effective gain from the relay to the destination is $n_{RD} - r_R$. Now we know that the capacity of this effective relay channel is equal to the cut-set upper bound and is achieved by a simple decode-forward protocol. Therefore the source can use the relay to send r_S number of bits to the destination, where r_S is

$$r_S = \min\left(\max(n_{SD}, n_{SR}), \max(n_{SD}, n_{RD} - r_R)\right) \tag{9.63}$$

Now the total number of MSB's of S_1 that is sent to the destination is equal to $r_S + r_R$. Now we show that $r_S + r_R$ is equal to the original cut-set upper bound shown in equation (9.61).

Table 9.1: Achievable rate of the proposed scheme for the cooperative relaying with side information problem in the deterministic case.

Cases	r_R	r_S	$r_S + r_R$
$n_{RD} \leq n_{SD}$	0	n_{SD}	n_{SD}
$n_{SR} \leq n_{SD} \leq n_{RD}$	$\min(n, n_{RD} - n_{SD})$	n_{SD}	$\min(n + n_{SD}, n_{RD})$
$n_{SD} \leq n_{RD} \leq n_{SR}$	0	n_{RD}	n_{RD}
$n_{SD} \leq n_{SR} \leq n_{RD}$	$\min(n_{RD} - n_{SR})$	n_{SR}	$n_{SR} + \min(n_{RD} - n_{SR})$

All possible cases are summarized in Table 9.1. Since $r_S + r_R$ in the last column of Table 9.1 is always equal to $\overline{C}_{\text{with side information}}$, therefore in the deterministic case the cut-set upper bound is achievable. Furthermore our deterministic scheme suggests a natural scheme for the Gaussian problem. In the next section we show that, quite interestingly, by using this scheme the destination will be able to estimate S_1 within a factor of six of the best possible distortion, uniformly for all channel gains and correlation values.

9.4.2 Approximating the Gaussian problem

The model in the Gaussian case is shown in Figure 9.6(b). As usual we assume that the noise power of each link, as well as the transmit power of each node is normalized to one. We further consider the real Gaussian model. Similar to previous section, we can still use the cut-set bound to derive a lower bound on the minimum achievable distortion for estimating s_1 at the destination. As shown in [53], this is given by

$$D_{\min} \geq \min_{0 \leq \xi \leq 1} \max \left((1-\rho^2)(1 + (1-\xi^2)(|h_1|^2 + |h_2|^2))^{-1}, (1 + |h_1|^2 + |h_3|^2 + 2\xi |h_1||h_3|)^{-1} \right) \tag{9.64}$$

where ξ is the correlation between the source and the relay codewords. We can further lower bound this by

$$D_{\min} \geq \underline{D} = \max \left((1-\rho^2)(1 + |h_1|^2 + |h_2|^2)^{-1}, (1 + (|h_1| + |h_3|)^2)^{-1} \right) \tag{9.65}$$

Now the optimal scheme that we found in previous section for the deterministic problem, naturally suggests us a protocol for the Gaussian case. The protocol is described as the following:

- The relay uses a fraction λ for transmitting a quantized version of s_2 to the destination and the remaining to perform decode-forward.
- The source uses the relay (with the remaining $1 - \lambda$ fraction of its power) to provide further information to the destination.
- The destination first decodes the quantized s_2, treating the decode-forward codeword as noise. Then it combines the side information received from the relay and the information received from the decode-forward codeword to obtain a reproduction of the source.

A similar scheme has also been studied in [54], called Hybrid Joint Source-Channel Decode-and-Forward. Here we will show that in fact with this protocol we can achieve a distortion that is within a factor of 6 of the cut-set lower bound (9.65) for all channel gains and correlation values.

Pictorially this scheme converts the system to the one shown in Figure 9.7(a). As shown in this figure, the relay uses a fraction of this figure to provide information about s_2 and the remaining to assist the source. Since we know that a decode-forward protocol achieves within 0.5 bit of the cut-set upper bound of the relay channel, therefore by coding we can convert the system to the one shown in Figure 9.7(b). Here there are two noiseless links with rate R_1 and R_2 are available from the source and the relay to the destination. The values of R_1 and R_2 are

$$\begin{aligned} R_1 &= \overline{\overline{C}}_{\text{relay}}(h_1, h_2, \sqrt{1-\lambda}h_3) - 0.5 \\ &= \min\left(\frac{1}{2}\log\left(1 + |h_1|^2 + |h_2|^2\right), \frac{1}{2}\log\left(1 + (|h_1| + \sqrt{1-\lambda}|h_3|)^2\right)\right) - 0.5 \qquad (9.66) \\ R_2 &= \frac{1}{2}\log\left(1 + \frac{\lambda|h_3|^2}{1 + (1-\lambda)|h_3|^2 + |h_1|^2}\right) \qquad (9.67) \end{aligned}$$

Now from the intuition obtained from the deterministic case we can predict a near optimal value of λ as the following:

$$\lambda^* = \begin{cases} 0, & \text{if } |h_3| \neq \max(|h_1|, |h_2|, |h_3|); \\ 1 - \frac{|h_2|^2}{|h_3|^2}, & \text{otherwise.} \end{cases} \tag{9.68}$$

Now the problem is similar to the one helper problem studied in [55]. As shown in [55], the achievable squared error distortion of estimating s_1 at the destination with the help of the relay is

$$D_h(R_1, R_2) = 2^{-2R_1}(1 - \rho^2 + \rho^2 2^{-2R_2}) \tag{9.69}$$

Now we analyze the performance of our strategy. Here is the main result

Theorem 9.4.1. *Consider the cooperative source-channel relaying strategy that is described above, with power allocation that is described in equation (9.68). With this scheme the destination will be able to estimate s_1 with a distortion that is uniformly within a factor of 6 of the minimum possible distortion, for all channel gains and correlation values.*

Proof. First note that if $\rho^2 \leq \frac{1}{2}$ then

$$\underline{D} = \max\left((1-\rho^2)(1 + |h_1|^2 + |h_2|^2)^{-1}, (1 + (|h_1| + |h_3|)^2)^{-1}\right) \tag{9.70}$$

$$\geq \frac{1}{2}\max\left((1 + |h_1|^2 + |h_2|^2)^{-1}, (1 + (|h_1| + |h_3|)^2)^{-1}\right) \tag{9.71}$$

Therefore by setting $\lambda = 0$ we get

$$D_h(R_1, R_2) = 2 \times \max\left((1 + |h_1|^2 + |h_2|^2)^{-1}, (1 + (|h_1| + |h_3|)^2)^{-1}\right) \tag{9.72}$$

$$\leq 4\underline{D} \tag{9.73}$$

This means that if the correlation between s_1 and s_2 is small enough ($\rho^2 \leq 0.5$), by ignoring

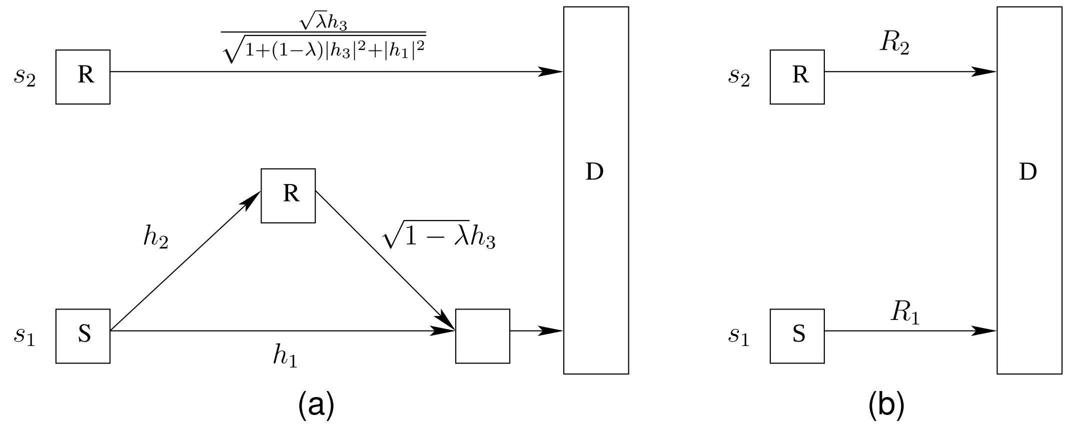


Figure 9.7: Pictorial representation of the protocol for the Gaussian cooperative relaying with side information is shown in (a). By coding we can make the channels noiseless and convert the system to the one shown in (b).

s_2 and using the relay only for decode-forward protocol we don't loose more than a factor of 4. Now assume $\rho^2 \geq 0.5$ and consider the following cases:

1. If $|h_3| \neq \max(|h_1|, |h_2|, |h_3|)$. In this case

$$D_h(R_1, R_2) = 2 \times \max\left((1+|h_1|^2+|h_2|^2)^{-1}, (1+(|h_1|+|h_3|)^2)^{-1}\right) \tag{9.74}$$

and

$$\underline{D} = \max\left((1-\rho^2)(1+|h_1|^2+|h_2|^2)^{-1}, (1+(|h_1|+|h_3|)^2)^{-1}\right) \tag{9.75}$$

We have two possibilities:

- If $|h_3| \leq |h_2|$. In this case we have

$$\begin{aligned}
(1+(|h_1|+|h_3|)^2)^{-1} &\geq (1+(|h_1|+|h_2|)^2)^{-1} && (9.76)\\
&\geq \frac{1}{2}(1+|h_1|^2+|h_2|^2)^{-1} && (9.77)\\
&\geq (1-\rho^2)(1+|h_1|^2+|h_2|^2)^{-1} && (9.78)
\end{aligned}$$

where the last step is true since $\rho^2 \geq 0.5$. Therefore in this case

$$\underline{D} = (1 + (|h_1| + |h_3|)^2)^{-1} \tag{9.79}$$

Also since $2(1 + (|h_1| + |h_3|)^2)^{-1} \geq (1 + |h_1|^2 + |h_2|^2)^{-1}$,

$$D_h(R_1, R_2) \leq 4(1 + (|h_1| + |h_3|)^2)^{-1} \leq 4\underline{D} \tag{9.80}$$

Therefore again our scheme is within a factor of 4 of the best possible performance.

- If $|h_3| \leq |h_1|$. In this case we have

$$D_h(R_1, R_2) = 4(1 + |h_1|^2 + |h_2|^2)^{-1} \tag{9.81}$$

And

$$\begin{aligned} \underline{D} &= \max\left((1-\rho^2)(1 + |h_1|^2 + |h_2|^2)^{-1}, (1 + (|h_1| + |h_3|)^2)^{-1}\right) && (9.82)\\ &\geq (1 + (|h_1| + |h_3|)^2)^{-1} && (9.83)\\ &\geq (1 + 4|h_1|^2)^{-1} && (9.84) \end{aligned}$$

Now if the source uses only the direct link we have

$$R_1 = \frac{1}{2}\log(1 + |h_1|^2) \tag{9.85}$$

Therefore

$$D_h(R_1, R_2) = (1 + |h_1|^2)^{-1} \leq 4(1 + 4|h_1|^2)^{-1} \leq \underline{D} \tag{9.86}$$

Therefore again our scheme is within a factor of 4 of the best possible performance.

2. If $|h_3| = \max(|h_1|, |h_2|, |h_3|)$. Then $\lambda = 1 - \frac{|h_2|^2}{|h_3|^2}$ and

$$R_1 = \frac{1}{2}\log\left(1 + |h_1|^2 + |h_2|^2\right) - 0.5 \tag{9.87}$$

$$R_2 = \frac{1}{2}\log\left(1 + \frac{|h_3|^2 - |h_2|^2}{1 + |h_1|^2 + |h_2|^2}\right) = \frac{1}{2}\log\left(\frac{1 + |h_1|^2 + |h_3|^2}{1 + |h_1|^2 + |h_2|^2}\right) \tag{9.88}$$

Therefore

$$D_h(R_1, R_2) = 2 \times (1 + |h_1|^2 + |h_2|^2)^{-1}\left(1 - \rho^2 + \rho^2 \frac{1 + |h_1|^2 + |h_2|^2}{1 + |h_1|^2 + |h_3|^2}\right) \tag{9.89}$$

$$= 2 \times (1 + |h_1|^2 + |h_2|^2)^{-1}(1 - \rho^2)\left(1 + \frac{\rho^2}{1 - \rho^2}\frac{1 + |h_1|^2 + |h_2|^2}{1 + |h_1|^2 + |h_3|^2}\right) \tag{9.90}$$

Remember that

$$\underline{D} = \max\left((1 - \rho^2)(1 + |h_1|^2 + |h_2|^2)^{-1}, (1 + (|h_1| + |h_3|)^2)^{-1}\right) \tag{9.91}$$

Now if

$$(1 - \rho^2)(1 + |h_1|^2 + |h_2|^2)^{-1} \geq (1 + (|h_1| + |h_3|)^2)^{-1} \tag{9.92}$$

Then

$$\underline{D} = (1 - \rho^2)(1 + |h_1|^2 + |h_2|^2)^{-1} \tag{9.93}$$

Also

$$1+\frac{\rho^2}{1-\rho^2}\frac{1+|h_1|^2+|h_2|^2}{1+|h_1|^2+|h_3|^2} \leq 1+\frac{2\rho^2}{1-\rho^2}\frac{1+|h_1|^2+|h_2|^2}{1+(|h_1|+|h_3|)^2} \tag{9.94}$$

$$\leq 1+\frac{2}{1-\rho^2}\frac{1+|h_1|^2+|h_2|^2}{1+(|h_1|+|h_3|)^2} \tag{9.95}$$

$$\leq 3 \tag{9.96}$$

Therefore

$$D_h(R_1,R_2) \leq 6\underline{D} \tag{9.97}$$

Therefore in this case our scheme is within a factor of 6 of the best possible performance.

In the other case, if

$$(1-\rho^2)(1+|h_1|^2+|h_2|^2)^{-1} \leq (1+(|h_1|+|h_3|)^2)^{-1} \tag{9.98}$$

Then

$$\underline{D} = (1+(|h_1|+|h_3|)^2)^{-1} \tag{9.99}$$

And

$$D_h(R_1,R_2) = 2\times(1+|h_1|^2+|h_2|^2)^{-1}(1-\rho^2)\left(1+\frac{\rho^2}{1-\rho^2}\frac{1+|h_1|^2+|h_2|^2}{1+|h_1|^2+|h_3|^2}\right) \tag{9.100}$$

$$\leq 2\times(1+|h_1|^2+|h_2|^2)^{-1}(1-\rho^2)\left(\frac{\rho^2}{1-\rho^2}\frac{1+|h_1|^2+|h_2|^2}{1+|h_1|^2+|h_3|^2}\right)+2\underline{D} \tag{9.101}$$

$$= \frac{2\rho^2}{1+|h_1|^2+|h_3|^2}+2\underline{D} \tag{9.102}$$

$$\leq \frac{2}{1+|h_1|^2+|h_3|^2}+2\underline{D} \tag{9.103}$$

$$\leq \frac{4}{1+(|h_1|+|h_3|)^2}+2\underline{D} \tag{9.104}$$

$$= 6\underline{D} \tag{9.105}$$

Therefore again in this case our scheme is within a factor of 6 of the best possible performance. □

Chapter 10

Conclusions

In this thesis we proposed a new deterministic approach to make progress in problems in wireless network information theory. The main idea was to develop a simpler deterministic channel model that allows us to focus on the interactions between users signals rather than the background noise of the systems. So far, our main focus has been on its application to unicast and multicast problems, and in particular we have been able to find a uniformly approximate characterization of the unicast/multicast capacity of Gaussian relay networks. This is the first constant gap approximation of the capacity of Gaussian relay networks.

As we demonstrated in Chapter 9, the proposed deterministic approach can also be applied to other problems in wireless network information theory. In particular we demonstrated its application to two other problems, one was a multi-session communication problem which was an extension of the relay channel. The other problem involved a combination of source and channel coding. Hence, we also developed a dual of our binary expansion channel model for Gaussian sources and applied it to make progress in the problem.

As discussed so far, our deterministic model has the potential to be both an effective tool in engineering as well as a powerful tool in information theory. On the one hand, the simplicity of this model allows engineers to obtain intuitive insights into complicated wireless networks. On the other hand, the inherent connections between this model and the Gaussian model enables information theorists to demonstrate new concrete theoretical results in Gaussian networks. I believe, these will have an impact on the design of future wireless communication systems.

Bibliography

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, IT-46(4):1204–1216, July 2000.

[2] S.-Y.R. Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 43(2):371–81, February 2003.

[3] R. Koetter and M. Médard. An algebraic approach to network coding. *IEEE/ACM Trans. Netw.*, 11(5):782–795, 2003.

[4] R. W. Yeung, S.-Y. Li, and N. Cai. *Network Coding Theory (Foundations and Trends(R) in Communications and Information Theory)*. Now Publishers Inc., Hanover, MA, USA, 2006.

[5] C. Fragouli and E. Soljanin. Network coding fundamentals. *Foundations and Trends in Networking*, 2(1):1–133, 2007.

[6] E. C. van der Meulen. Three-terminal communication channels. *Ad. Appl. Pmb.*, 3:120–154, September 1971.

[7] T.M. Cover and A. El Gamal. Capacity theorems for the relay channel. *IEEE Transactions on Info. Theory*, 25(5):572–584, September 1979.

[8] M. Aref. Information flow in relay networks. *PhD. Thesis, Stanford University*, October 1980.

[9] N. Ratnakar and G. Kramer. The multicast capacity of deterministic relay networks with no interference. *IEEE Transactions on Information Theory*, 52(6):2425–2432, June 2006.

[10] P. Gupta, S. Bhadra, and S. Shakkottai. On network coding for interference networks. *International Symposium on Information Theory (ISIT), Seattle*, July 2006.

[11] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros. Capacity of wireless erasure networks. *IEEE Transactions on Information Theory*, 52(3):789–804, March 2006.

[12] B. Schein. Distributed coordination in network information theory. *Massachusetts Institute of Technology*, 2001.

[13] L. L. Xie and P. R. Kumar. A network information theory for wireless communication: Scaling laws and optimal operation. *IEEE Transactions on Information Theory*, IT-50(5):748–767, May 2004.

[14] M. Gastpar and M. Vetterli. On the capacity of large gaussian relay networks. *IEEE Transactions on Information Theory*, 51(3):765–779, March 2005.

[15] G. Kramer, M. Gastpar, and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Transactions on Information Theory*, 51(9):3037–3063, September 2005.

[16] A. Reznik, S. R. Kulkarni, and S. Verdu. Degraded gaussian multirelay channel: Capacity and optimal power allocation. *IEEE Transactions on Information Theory*, 50(12):3037–3046, December 2004.

[17] M. A. Khojastepour, A. Sabharwal, and B. Aazhang. Lower bounds on the capacity of gaussian relay channel. *Proc 38th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ*, pages 597–602, March 2004.

[18] D. N. C. Tse J. N. Laneman and G. Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transactions on Information Theory*, IT-50(12):3062–3080, December 2004.

[19] U. Mitra and A. Sabharwal. Complexity constrained sensor networks: achievable rates for two relay networks and generalizations. *Information Processing in Sensor Networks Symposium, Berkeley, CA*, April 2004.

[20] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversitypart i: System description. *IEEE Transactions on Communications*, 51(11):1927–1938, November 2003.

[21] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversitypart ii: Implementation aspects and performance analysis. *IEEE Transactions on Communications*, 51(11):1939–1948, November 2003.

[22] A. El Gamal and S. Zahedi. Capacity of a class of relay channels with orthogonal components. *IEEE Transactions on Information Theory*, IT-51(5):1815–1817, May 2005.

[23] A. Nosratinia and A. Hedayat. Cooperative communication in wireless networks. *IEEE Communications Magazine*, 42(10):74–80, October 2004.

[24] M. Yuksel and E. Erkip. Multiple-antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective. *IEEE Transactions on Information Theory*, 53(10):3371–3393, October 2007.

[25] S. Avestimehr, S. Diggavi, and D. Tse. A deterministic model for wireless relay networks and its capacity. *IEEE Information Theory Workshop (ITW), Bergen, Norway*, July 2007.

[26] S. Avestimehr, S. Diggavi, and D. Tse. A deterministic approach to wireless relay networks. *45th Allerton Conf. On Comm., Control, and Computing 2007, Monticello, Illinois, USA, September 26- 28*, 2007.

[27] S. Avestimehr, S. Diggavi, and D. Tse. Wireless network information flow. *45th Allerton Conf. On Comm., Control, and Computing 2007, Monticello, Illinois, USA, September 26- 28*, 2007.

[28] S. Avestimehr, S. Diggavi, and D. Tse. Information flow over compound wireless relay networks. *International Zurich seminar on communications (IZS), Zurich*, March 2008.

[29] S. Avestimehr, S. Diggavi, and D. Tse. Approximate capacity of gaussian relay networks. *International Symposium on Information Theory (ISIT), Toronto, Canada*, July 2008.

[30] S. Avestimehr, S. Diggavi, and D. Tse. Approximate characterization of capacity in gaussian relay networks. *IEEE International Wireless Communications and Mobile Computing Conference (IWCMC), Crete*, August 2008.

[31] S. Avestimehr, A. Sezgin, and D. Tse. Approximate capacity of the two-way relay channel: A deterministic approach. *Preprint*.

[32] G. D. Forney and G. Ungerboeck. Modulation and coding for linear gaussian channels. *IEEE Transactions on Information Theory*, 44(6):2384–2415, October 1998.

[33] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing, 2nd edition, 2006.

[34] A. Orlitsky and J. Roche. Coding for computing. *IEEE Transactions on Information Theory*, 47(3):903–917, March 2001.

[35] M. Feder and A. Lapidoth. Universal decoding for channels with memory. *IEEE Transactions on Information Theory*, 44(9):1726–1745, September 1998.

[36] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *The Annals of Mathematical Statistics*, 30(4):1229–1241, December 1959.

[37] W. L. Root and P. P. Varaiya. Capacity of classes of gaussian channels. *SIAM Journal on Applied Mathematics*, 16(6):1350–1393, November 1968.

[38] M. A. Khojastepour, A. Sabharwal, and B. Aazhang. Bounds on achievable rates for general multi-terminal networks with practical constraints. *In Proc. of 2nd International Workshop on Information Processing (IPSN*, pages 146–161, 2003.

[39] B. Bollobas. *Modern Graph Theory*. Graduate Texts in Mathematics, vol. 184, Springer, New York, 1998.

[40] R. Etkin, D. Tse, and H. Wang. Gaussian interference channel capacity to within one bit. *submitted to IEEE Transactions on Information Theory. Also available at http://www.eecs.berkeley.edu/∼dtse/interference.pdf*, Feb. 2007.

[41] C. E. Shannon. Two-way communication channels. *Proc. 4th Berkeley Symp. Mathematical Statistics Probability, Berkeley, CA*, pages 611–644, 1961.

[42] B. Rankov and A. Wittneben. Achievable rate regions for the two-way relay channel. *ISIT, Seattle, USA*, July 9-14, 2006.

[43] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft. XORs in the Air: Practical Wireless Network Coding. *ACM SIGCOMM, Pisa, Italy*, September, 11-15 2006.

[44] C. Hausl and J. Hagenauer. Iterative network and channel decoding for the two-way relay channel. *Proc. of IEEE ICC 2006, Istanbul, Turkey*, June 2006.

[45] I.-J. Baik and S.-Y. Chung. Networking coding for two-way relay channels using lattices. *Proc. of IEEE ICC 2008, Beijing, China*, pages 3898–3902, May 2007.

[46] S. Katti, S. Gollakota, and D. Katabi. Embracing Wireless Interference: Analog Network Coding. *ACM SIGCOMM, Kyoto, Japan*, August 27-31, 2007.

[47] K. Narayanan, M. P. Wilson, and A. Sprintson. Joint physical layer coding and network coding for bi-directional relaying. *Proc. of Allerton Conference on Communication, Control and Computing*, 2007.

[48] T.J. Oechtering, C. Schnurr, I. Bjelakovic, and H. Boche. Broadcast Capacity Region of Two-Phase Bidirectional Relaying. *IEEE Transactions on Information Theory*, 54(1):454–458, January 2008.

[49] A. Sendonaris, E. Erkip, and B. Aazhang. User Cooperation Diversity–Part I: System Description. *IEEE Transactions on Communications*, 51(11):1927–1938, November 2003.

[50] J.N. Laneman and G.W. Wornell. Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks. *IEEE Transactions on Info. Theory*, 49(10):2415–2425, October 2003.

[51] R.U. Nabar, H. Bölcskei, and F.W. Kneubühler. Fading relay channels: Performance limits and space-time signal design. *IEEE Journal on Selec. Areas in Commun.*, 22(6):1099–1109, Aug. 2004.

[52] T.M. Cover. Broadcast channels. *IEEE Trans. on Information Theory*, 18(1):2–14, 1972.

[53] D. Gunduz, C.T.K. Ng, E. Erkip, and A.J. Goldsmith. Source transmission over relay channel with correlated relay side information. *International Symposium on Information Theory (ISIT), Nice, France*, June 2007.

[54] D. Gunduz, E. Erkip, A.J. Goldsmith, and H.V. Poor. Lossy source transmission over the relay channel. *preprint*.

[55] Y. Oohama. Gaussian multiterminal source coding. *IEEE Transactions on Information Theory*, 43(11):1912–1923, November 1997.

[56] N.J.A. Harvey, R. Kleinberg, and A.R. Lehman. On the capacity of information networks. *IEEE Transactions on Information Theory*, 52(6):2445–2464, June 2006.

Appendix A

Proofs

A.1 Proof of Theorem 3.2.1

If $|h_{SR}| < |h_{SD}|$ then the relay is ignored and a communication rate equal to $R = \log(1 + |h_{SD}|^2)$ is achievable. If $|h_{SR}| > |h_{SD}|$ the problem becomes more interesting. In this case we can think of a decode-forward scheme as described in [7]. Then by using a block-Markov encoding scheme the following communication rate is achievable:

$$R = \min\left(\log\left(1 + |h_{SR}|^2\right), \log\left(1 + |h_{SD}|^2 + |h_{RD}|^2\right)\right) \tag{A.1}$$

Therefore overall the following rate is always achievable:

$$R_{\text{DF}} = \max\{\log(1 + |h_{SD}|^2), \min\left(\log\left(1 + |h_{SR}|^2\right), \log\left(1 + |h_{SD}|^2 + |h_{RD}|^2\right)\right)\}$$

Now we compare this achievable rate with the cut-set upper bound on the capacity of the Gaussian relay network,

$$C \leq \overline{C} = \max_{|\rho| \leq 1} \min\{\log\left(1 + (1 - \rho^2)(|h_{SD}|^2 + |h_{SR}|^2)\right), \log\left(1 + |h_{SD}|^2 + |h_{RD}|^2 + 2\rho|h_{SD}||h_{RD}|\right)\} \tag{A.2}$$

Note that if $|h_{SR}| > |h_{SD}|$ then

$$R_{DF} = \min\left(\log\left(1+|h_{SR}|^2\right), \log\left(1+|h_{SD}|^2+|h_{RD}|^2\right)\right) \tag{A.3}$$

and for all $|\rho| \leq 1$ we have

$$\log\left(1+(1-\rho^2)(|h_{SD}|^2+|h_{SR}|^2)\right) \leq \log\left(1+|h_{SR}|^2\right)+1 \tag{A.4}$$

$$\log\left(1+|h_{SD}|^2+|h_{RD}|^2+2\rho|h_{SD}||h_{RD}|\right) \leq \log\left(1+|h_{SD}|^2+|h_{RD}|^2\right)+1 \tag{A.5}$$

Hence

$$R_{\text{DF}} \geq \overline{C}_{\text{relay}} - 1 \tag{A.6}$$

Also if $|h_{SR}| > |h_{SD}|$,

$$R_{DF} = \log(1+|h_{SD}|^2) \tag{A.7}$$

and

$$\log\left(1+(1-\rho^2)(|h_{SD}|^2+|h_{SR}|^2)\right) \leq \log\left(1+|h_{SD}|^2\right)+1 \tag{A.8}$$

therefore again,

$$R_{\text{DF}} \geq \overline{C}_{\text{relay}} - 1 \tag{A.9}$$

A.2 Proof of Theorem 3.3.1

The cut-set upper bound on the capacity of diamond network is:

$$\begin{aligned} C_{\text{diamond}} &\leq \overline{C} \leq \min\{\log\left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right) \\ &\quad ,\log\left(1+(|h_{A_1D}|+|h_{A_2D}|)^2\right) \\ &\quad ,\log(1+|h_{SA_1}|^2)+\log(1+|h_{A_2D}|^2) \\ &\quad ,\log(1+|h_{SA_2}|^2)+\log(1+|h_{A_1D}|^2)\} \end{aligned} \tag{A.10}$$

Without loss of generality assume

$$|h_{SA_1}| \geq |h_{SA_2}| \tag{A.11}$$

Then we have the following cases:

1. $|h_{SA_1}| \leq |h_{A_1D}|$:
   In this case
   $$R_{PDF} \geq \log(1+|h_{SA_1}|^2) \geq \overline{C}-1 \tag{A.12}$$

2. $|h_{SA_1}| > |h_{A_1D}|$:
   Let $\alpha = \frac{|h_{A_1D}|^2}{|h_{SA_1}|^2}$ then
   $$\begin{aligned} R_{PDF} &= \log(1+|h_{A_1D}|^2)+\min\left\{\log\left(1+\frac{(1-\alpha)\,|h_{SA_2}|^2}{\alpha|h_{SA_2}|^2+1}\right),\log\left(1+\frac{|h_{A_2D}|^2}{1+|h_{A_1D}|^2}\right)\right\} \\ &= \min\left\{\log\left(\frac{(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2+1}\right),\log\left(1+|h_{A_1D}|^2+|h_{A_2D}|^2\right)\right\} \end{aligned} \tag{A.13}$$

   Now if

   $$\log\left(\frac{(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2+1}\right) \geq \log\left(1+|h_{A_1D}|^2+|h_{A_2D}|^2\right) \tag{A.14}$$

we have

$$R_{PDF} = \log\left(1 + |h_{A_1D}|^2 + |h_{A_2D}|^2\right) \quad (A.15)$$
$$\geq \log\left(1 + (|h_{A_1D}| + |h_{A_2D}|)^2\right) - 1 \quad (A.16)$$
$$\geq \overline{C} - 1 \quad (A.17)$$

therefore the achievable rate of partial decode-forward scheme is within one bit of the cut-set bound. So we just need to look at the case that

$$R_{PDF} = \log\left(\frac{(1 + |h_{SA_2}|^2)(1 + |h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2 + 1}\right) \quad (A.18)$$

In this case consider two possibilities:

- $\alpha|h_{SA_2}|^2 \leq 1$:

  In this case we have

$$R_{PDF} = \log\left(\frac{(1 + |h_{SA_2}|^2)(1 + |h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2 + 1}\right) \quad (A.19)$$
$$\geq \log\left(\frac{(1 + |h_{SA_2}|^2)(1 + |h_{A_1D}|^2)}{2}\right) \quad (A.20)$$
$$= \log(1 + |h_{SA_2}|^2) + \log(1 + |h_{A_1D}|^2) - 1 \quad (A.21)$$
$$\geq \overline{C} - 1 \quad (A.22)$$

- $\alpha|h_{SA_2}|^2 \geq 1$:

  In this case we are going to show that

$$R_{PDF} = \log\left(\frac{(1 + |h_{SA_2}|^2)(1 + |h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2 + 1}\right) \quad (A.23)$$
$$\geq \log\left(1 + |h_{SA_1}|^2 + |h_{SA_2}|^2\right) - 1 \quad (A.24)$$
$$\geq \overline{C} - 1 \quad (A.25)$$

To show this we just need to prove

$$\frac{(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2+1} \geq \frac{1}{2}\left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right) \tag{A.26}$$

By replacing $\alpha = \frac{|h_{A_1D}|^2}{|h_{SA_1}|^2}$, we get

$$2|h_{SA_1}|^2(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2) \geq \left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right)\left(|h_{SA_1}|^2+|h_{SA_2}|^2|h_{A_1D}|^2\right) \tag{A.27}$$

But note that

$$\begin{aligned}
&2|h_{SA_1}|^2(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2) - \left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right)\left(|h_{SA_1}|^2+|h_{SA_2}|^2|h_{A_1D}|^2\right)\\
&= |h_{SA_1}|^2 + |h_{SA_1}|^2|h_{A_1D}|^2 + (|h_{SA_1}|^2|h_{SA_2}|^2 - |h_{SA_2}|^4|h_{A_1D}|^2)+\\
&\quad + (|h_{SA_1}|^2|h_{A_1D}|^2 - |h_{SA_2}|^2|h_{A_1D}|^2) + (|h_{SA_1}|^2|h_{SA_2}|^2|h_{A_1D}|^2 - |h_{SA_1}|^4) && \text{(A.28)}\\
&= |h_{SA_1}|^2 + |h_{SA_1}|^2|h_{A_1D}|^2 + |h_{SA_2}|^2(|h_{SA_1}|^2 - |h_{SA_2}|^2|h_{A_1D}|^2)+\\
&\quad + |h_{A_1D}|^2(|h_{SA_1}|^2 - |h_{SA_2}|^2) + |h_{SA_1}|^2(|h_{SA_2}|^2|h_{A_1D}|^2 - |h_{SA_1}|^2) && \text{(A.29)}\\
&= |h_{SA_1}|^2 + |h_{SA_1}|^2|h_{A_1D}|^2 + (|h_{SA_1}|^2 - |h_{SA_2}|^2)(|h_{SA_2}|^2|h_{A_1D}|^2 - |h_{SA_1}|^2 + |h_{A_1D}|^2) && \text{(A.30)}\\
&\geq 0 && \text{(A.31)}
\end{aligned}$$

Where the last step is true since

$$|h_{SA_1}|^2 \geq |h_{SA_2}|^2 \tag{A.32}$$

$$|h_{SA_2}|^2|h_{A_1D}|^2 \geq |h_{SA_2}|^2 \quad (\text{since } \alpha|h_{SA_2}|^2 \geq 1) \tag{A.33}$$

A.3 Proof of Lemma 5.4.2

First note that any cut in the unfolded graph, Ω_{unf}, partitions the nodes at each stage $1 \leq i \leq K$ to $\mathcal{U}_i$ (on the left of the cut) and $\mathcal{V}_i$ (on the right of the cut). If at one stage $S[i] \in \mathcal{V}_i$ or $D[i] \in \mathcal{U}_i$ then the cut passes through one of the infinite capacity edges (capacity Kq)

and hence Lemma 5.4.2 is obviously proved. Therefore without loss of generality assume that $S[i] \in \mathcal{U}_i$ and $D[i] \in \mathcal{V}_i$ for all $1 \leq i \leq K$. Now since for each $i \in \mathcal{V}$, $\{x_i[t]\}_{1 \leq t \leq K}$ are i.i.d distributed we can write[1]

$$H(Y_{\Omega_{\text{unf}}^c} | X_{\Omega_{\text{unf}}^c}) = \sum_{i=1}^{K-1} H(Y_{\mathcal{V}_{i+1}} | X_{\mathcal{V}_i}) \tag{A.34}$$

For simplification we define

$$\psi(\mathcal{V}_1, \mathcal{V}_2) \triangleq H(Y_{\mathcal{V}_2} | X_{\mathcal{V}_1}) \tag{A.35}$$

Now we show the following lemmas

Lemma A.3.1. *The $\tilde{\mathcal{V}}_i$'s defined in Lemma A.6.1 satisfy,*

$$\tilde{\mathcal{V}}_l \subseteq \tilde{\mathcal{V}}_{l-1} \subseteq \cdots \subseteq \tilde{\mathcal{V}}_1 \tag{A.36}$$

Proof. Proof is clear. □

Lemma A.3.2. *Let $\mathcal{V}_1, \ldots, \mathcal{V}_l$ be l non identical subsets of $\mathcal{V} - \{S\}$ such that $D \in \mathcal{V}_i$ for all $1 \leq i \leq l$. Also assume that $\tilde{\mathcal{V}}_1, \ldots, \tilde{\mathcal{V}}_l$ are as defined in lemma A.3.5. Then for any $v \in \mathcal{V}$ we have*

$$|\{i | v \in \mathcal{V}_i\}| = |\{j | v \in \tilde{\mathcal{V}}_j\}| \tag{A.37}$$

Proof. This lemma just states that for each $v \in V$ the number of times that v appears in V_i's is equal to the number of times that v appears in $\tilde{V}_i$'s. To prove it assume that v appears in V_i's is n. Then clearly

$$v \in \tilde{V}_j, \quad j = 1, \ldots, n \tag{A.38}$$

[1] As in Section 5.3.2, under the product distribution the mutual information expression of the cut-set breaks into a summation.

Now for any $j > n$ any element that appears in each $\tilde{V}_j$ must appear in at least j of V_i's and since v only appears in n of V_i's therefore,

$$v \notin \tilde{V}_j, \quad j > n \tag{A.39}$$

therefore

$$|\{i|v \in V_i\}| = |\{j|v \in \tilde{V}_j\}| = n \tag{A.40}$$

□

Lemma A.3.3. *Let* $\mathcal{V}_1, \ldots, \mathcal{V}_l$ *be* l *non identical subsets of* $\mathcal{V} - \{S\}$ *such that* $D \in \mathcal{V}_i$ *for all* $1 \leq i \leq l$. *Also assume a product distribution on* $X_i,\ i \in \mathcal{V}$. *Then*

$$H(X_{\mathcal{V}_1}) + \cdots + H(X_{\mathcal{V}_l}) = H(X_{\tilde{\mathcal{V}}_1}) + \cdots + H(X_{\tilde{\mathcal{V}}_l}) \tag{A.41}$$

where $\tilde{\mathcal{V}}_i$*'s are defined in Lemma A.3.5 and* $H(.)$ *is just the binary entropy function.*

Proof. For any $v \in V$ define

$$n_v = |\{i|v \in V_i\}| \tag{A.42}$$

and

$$\hat{n}_v = |\{j|v \in \tilde{V}_j\}| \tag{A.43}$$

Now since $X_i,\ i \in V$ are independent of each other we have

$$H(X_{V_1}) + \cdots + H(X_{V_l}) = \sum_{v \in V} n_v H(X_v) \tag{A.44}$$

and

$$H(X_{\tilde{V}_1}) + \cdots + H(X_{\tilde{V}_k}) = \sum_{v \in V} \hat{n}_v H(X_v) \tag{A.45}$$

By lemma A.3.2 we know that $n_v = \hat{n}_v$ for all $v \in V$ hence the lemma is proved. □

The following Lemma is just a straight forward generalization of submodularity to more than two sets (see also [56], Theorem 5 where this result is applied to the entropy function which is submodular).

Lemma A.3.4. *Let $\mathcal{V}_1, \dots, \mathcal{V}_k$ be a collection of sets. Assume that $\xi(\cdot)$ is a submodular function. Then,*

$$\xi(\mathcal{V}_1) + \cdots + \xi(\mathcal{V}_k) \geq \xi(\tilde{\mathcal{V}}_1) + \cdots + \xi(\tilde{\mathcal{V}}_k) \tag{A.46}$$

where $\tilde{\mathcal{V}}_i$'s are defined in Lemma A.3.5.

Lemma A.3.5. *Let $\mathcal{V}_1, \dots, \mathcal{V}_l$ be l non identical subsets of $\mathcal{V} - \{S\}$ such that $D \in \mathcal{V}_i$ for all $1 \leq i \leq l$. Also assume a product distribution on $x_i,\ i \in \mathcal{V}$. Then*

$$\psi(\mathcal{V}_1, \mathcal{V}_2) + \cdots + \psi(\mathcal{V}_{l-1}, \mathcal{V}_l) + \psi(\mathcal{V}_l, \mathcal{V}_1) \geq \sum_{i=1}^{l} \psi(\tilde{\mathcal{V}}_i, \tilde{\mathcal{V}}_i) \tag{A.47}$$

where for $k = 1, \dots, l$,

$$\tilde{\mathcal{V}}_k = \bigcup_{\{i_1,\dots,i_k\} \subseteq \{1,\dots,l\}} (\mathcal{V}_{i_1} \cap \cdots \cap \mathcal{V}_{i_k}) \tag{A.48}$$

or in another words each $\tilde{\mathcal{V}}_j$ is the union of $\binom{l}{j}$ sets such that each set is intersect of j of $\mathcal{V}_i$'s.

Proof. First note that

$$\begin{aligned}
&\psi(\mathcal{V}_1, \mathcal{V}_2) + \cdots + \psi(\mathcal{V}_{l-1}, \mathcal{V}_l) + \psi(\mathcal{V}_l, \mathcal{V}_1) = \\
&H(Y_{\mathcal{V}_2}|X_{\mathcal{V}_1}) + \cdots + H(Y_{\mathcal{V}_l}|X_{\mathcal{V}_{l-1}}) + H(Y_{\mathcal{V}_1}|X_{\mathcal{V}_l}) = \\
&H(Y_{\mathcal{V}_2}, X_{\mathcal{V}_1}) + \cdots + H(Y_{\mathcal{V}_l}, X_{\mathcal{V}_{l-1}}) + H(Y_{\mathcal{V}_1}, X_{\mathcal{V}_l}) - \sum_{i=1}^{l} H(X_{\mathcal{V}_i})
\end{aligned}$$

and

$$\sum_{i=1}^{l} \psi(\tilde{\mathcal{V}}_i, \tilde{\mathcal{V}}_i) = \sum_{i=1}^{l} H(Y_{\tilde{\mathcal{V}}_i} | X_{\tilde{\mathcal{V}}_i}) \tag{A.49}$$

$$= \sum_{i=1}^{l} H(Y_{\tilde{\mathcal{V}}_i}, X_{\tilde{\mathcal{V}}_i}) - \sum_{i=1}^{l} H(X_{\tilde{\mathcal{V}}_i}) \tag{A.50}$$

Now define the set

$$\mathcal{W}_i = \{Y_{\mathcal{V}_i}, X_{\mathcal{V}_{i-1}}\}, \quad i = 1, \ldots, l \tag{A.51}$$

where $\mathcal{V}_0 = \mathcal{V}_l$. Since by lemma A.3.2 we have

$$\sum_{i=1}^{l} H(X_{\mathcal{V}_i}) = \sum_{i=1}^{l} H(X_{\tilde{\mathcal{V}}_i}) \tag{A.52}$$

we just need to prove that

$$\sum_{i=1}^{l} H(\mathcal{W}_i) \geq \sum_{i=1}^{l} H(Y_{\tilde{\mathcal{V}}_i}, X_{\tilde{\mathcal{V}}_i}) \tag{A.53}$$

Now by since entropy is a submodular function by Lemma A.3.4 (k-way submodularity) we have,

$$\sum_{i=1}^{l} H(\mathcal{W}_i) \geq \sum_{i=1}^{l} H(\tilde{\mathcal{W}}_i) \tag{A.54}$$

where

$$\tilde{\mathcal{W}}_r = \bigcup_{\{i_1,\ldots,i_r\} \subseteq \{1,\ldots,l\}} (\mathcal{W}_{i_1} \cap \cdots \cap \mathcal{W}_{i_r}), \quad r = 1, \ldots, l \tag{A.55}$$

Now for any r ($1 \leq r \leq l$) we have

$$\begin{aligned}
\tilde{\mathcal{W}}_r &= \bigcup_{\{i_1,\ldots,i_r\}\subseteq\{1,\ldots,l\}} (\mathcal{W}_{i_1} \cap \cdots \cap \mathcal{W}_{i_r}) \\
&= \bigcup_{\{i_1,\ldots,i_r\}\subseteq\{1,\ldots,l\}} (\{Y_{\mathcal{V}_{i_1}}, X_{\mathcal{V}_{i_1-1}}\} \cap \cdots \cap \{Y_{\mathcal{V}_{i_r}} X_{\mathcal{V}_{i_r-1}}\}) \\
&= \bigcup_{\{i_1,\ldots,i_r\}\subseteq\{1,\ldots,l\}} (\{Y_{\mathcal{V}_{i_1}\cap\cdots\cap\mathcal{V}_{i_r}}, X_{\mathcal{V}_{(i_1-1)}\cap\cdots\cap X_{\mathcal{V}_{(i_r-1)}}}\}) \\
&= \left\{Y_{\bigcup_{\{i_1,\ldots,i_r\}}(\mathcal{V}_{i_1}\cap\cdots\cap\mathcal{V}_{i_r})}, X_{\bigcup_{\{i_1,\ldots,i_r\}}(\mathcal{V}_{(i_1-1)}\cap\cdots\cap\mathcal{V}_{(i_r-1)})}\right\} \\
&= \{Y_{\tilde{\mathcal{V}}_r}, X_{\tilde{\mathcal{V}}_r}\}
\end{aligned}$$

Therefore by equation (A.54) we have,

$$\begin{aligned}
\sum_{i=1}^{l} H(\mathcal{W}_i) &\geq \sum_{i=1}^{l} H(\tilde{\mathcal{W}}_i) && \text{(A.56)} \\
&= \sum_{i=1}^{l} H(Y_{\tilde{\mathcal{V}}_i}, X_{\tilde{\mathcal{V}}_i}) && \text{(A.57)}
\end{aligned}$$

Hence the Lemma is proved. □

Now note that

$$H(Y_{\Omega^c_{\text{unf}}}|X_{\Omega^c_{\text{unf}}}) = \sum_{i=1}^{K-1} H(Y_{\mathcal{V}_{i+1}}|X_{\mathcal{V}_i}) = \sum_{i=1}^{K-1} \psi(\mathcal{V}_i, \mathcal{V}_{i+1}) \tag{A.58}$$

Consider the sequence of $\mathcal{V}_i$'s. Note that there are total of $L = 2^{|\mathcal{V}|-2}$ possible subsets of $\mathcal{V}$ that contain D but not S. Assume that $\mathcal{V}_s$ is the first set that is revisited. Assume that it is revisited at step $\mathcal{V}_{s+l}$. Therefore by Lemma A.3.5 we have

$$\sum_{i=1}^{l-1} \psi(\mathcal{V}_i, \mathcal{V}_{i+1}) \geq \sum_{i=1}^{l} \psi(\tilde{\mathcal{V}}_i, \tilde{\mathcal{V}}_i) \tag{A.59}$$

where $\tilde{\mathcal{V}}_i$'s are described in Lemma A.3.5. Now note that any of those $\tilde{\mathcal{V}}_i$ contains D but not S and hence it describes a cut in the original graph, therefore

$$\psi(\tilde{\mathcal{V}}_i, \tilde{\mathcal{V}}_i) \geq \min_{\Omega \in \Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) \tag{A.60}$$

Hence

$$\sum_{i=1}^{l-1} \psi(\mathcal{V}_i, \mathcal{V}_{i+1}) \geq l \min_{\Omega \in \Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) \tag{A.61}$$

which means that the value of that loop is at least length of the loop times the min-cut of the original graph. Now since in any $L-1$ time frame there is at least one loop therefore except at most a path of length $L-1$ everything can be replaced with the value of the min-cut in $\sum_{i=1}^{K-1} \psi(\mathcal{V}_i, \mathcal{V}_{i+1})$. Therefore,

$$\sum_{i=1}^{K-1} \psi(\mathcal{V}_i, \mathcal{V}_{i+1}) \geq (K-L+1) \min_{\Omega \in \Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) \tag{A.62}$$

A.4 Proof of Lemma 6.2.4

Assume message w' is transmitted. Consider a relay, R, at the first layer. Then, the total number of quantized outputs at R would be

$$2^{H(\hat{\mathbf{Y}}_R|\mathbf{X}_S)} = 2^{TI(Y_R;\hat{Y}_R|X_S)} \tag{A.63}$$

Since we are using an optimal Gaussian vector quantizer at the noise level (i.e. with distortion 1), we can write

$$\hat{Y}_R = \alpha Y_R + N, \tag{A.64}$$

where $N \sim \mathcal{CN}(0, \sigma_N^2)$ is a complex Gaussian noise independent of Y_R and

$$\alpha = \frac{\sigma_Y^2 - 1}{\sigma_Y^2}, \quad \sigma_N^2 = (1-\alpha^2)\sigma_Y^2 - 1 \tag{A.65}$$

Hence

$$\begin{aligned} I(Y_R; \hat{Y}_R | X_S) &= \log\left(1 + \frac{\alpha^2}{\sigma_N^2}\right) && \text{(A.66)} \\ &= \log(1+\alpha) \le 1 && \text{(A.67)} \end{aligned}$$

Hence the list size of R would be smaller than 2^T. Now the list of typical transmit sequences can be viewed as a tree such that at each node, due to the noise, each path will be branched to at most 2^T other typical possibilities. Therefore, the total number of typical transmit sequences would be smaller than the product of the expansion coefficient (*i.e.* 2^T) over all nodes in the graph. Or, more precisely

$$\begin{aligned} &\log\left(|\mathcal{X}_{\mathcal{V}}(w')|\right) = \log\left(|\mathcal{Y}_{\mathcal{V}}(w')|\right) && \text{(A.68)} \\ &= H(\hat{\mathbf{Y}}_{\mathcal{V}} | w') = \sum_{l=1}^{l_D} H(\hat{\mathbf{Y}}_{\gamma_l} | \hat{\mathbf{Y}}_{\gamma_{l-1}}) && \text{(A.69)} \\ &= \sum_{l=1}^{l_D} H(\hat{\mathbf{Y}}_{\gamma_l} | \mathbf{X}_{\gamma_{l-1}}) && \text{(A.70)} \\ &\le \sum_{l=1}^{l_D} T|\gamma_l| = T|\mathcal{V}| && \text{(A.71)} \end{aligned}$$

Where γ_l is the set of nodes at the l-th layer of the network. Hence,

$$|\mathcal{X}_{\mathcal{V}}(w')| \le 2^{T|\mathcal{V}|} \tag{A.72}$$

and the proof is complete.

A.5 Proof of Lemma 6.2.6

We know that the capacity of a $r \times t$ MIMO channel H, with water filling is

$$C_{wf} = \sum_{i=1}^{n} \log(1 + \tilde{Q}_{ii}\lambda_i) \tag{A.73}$$

where $n = min(r, t)$, and λ_i's are the singular values of H and $\tilde{Q}_{ii}$ is given by water filling solution satisfying

$$\sum_{i=1}^{n} \tilde{Q}_{ii} = nP \tag{A.74}$$

With equal power allocation

$$C_{ep} = \sum_{i=1}^{n} \log(1 + P\lambda_i) \tag{A.75}$$

Now note that

$$\begin{aligned}
C_{wf} - C_{ep} &= \log\left(\frac{\prod_{i=1}^{n}(1 + \tilde{Q}_{ii}\lambda_i)}{\prod_{i=1}^{n}(1 + P\lambda_i)}\right) && \text{(A.76)} \\
&\leq \log\left(\frac{\prod_{i=1}^{n}(1 + \tilde{Q}_{ii}\lambda_i)}{\prod_{i=1}^{n}\max(1, P\lambda_i)}\right) && \text{(A.77)} \\
&= \log\left(\prod_{i=1}^{n}\frac{1 + \tilde{Q}_{ii}\lambda_i}{\max(1, P\lambda_i)}\right) && \text{(A.78)} \\
&= \log\left(\prod_{i=1}^{n}\left(\frac{1}{\max(1, P\lambda_i)} + \frac{\tilde{Q}_{ii}\lambda_i}{\max(1, P\lambda_i)}\right)\right) && \text{(A.79)} \\
&\leq \log\left(\prod_{i=1}^{n}\left(1 + \frac{\tilde{Q}_{ii}\lambda_i}{P\lambda_i}\right)\right) && \text{(A.80)} \\
&= \log\left(\prod_{i=1}^{n}\left(1 + \frac{\tilde{Q}_{ii}}{P}\right)\right) && \text{(A.81)}
\end{aligned}$$

Now note that

$$\sum_{i=1}^{n}(1+\frac{\tilde{Q}_{ii}}{P}) = 2n \tag{A.82}$$

and therefore by arithmetic mean-geometric mean inequality we have

$$\prod_{i=1}^{n}\left(1+\frac{\tilde{Q}_{ii}}{P}\right) \leq \left(\frac{\sum_{i=1}^{n}(1+\frac{\tilde{Q}_{ii}}{P})}{n}\right)^{n} = 2^{n} \tag{A.83}$$

and hence

$$C_{ep} - C_{wf} \leq n \tag{A.84}$$

A.6 Proof of Lemma 6.3.2

First, we prove a lemma which is a slight generalization of Lemma A.3.5.

Lemma A.6.1. *Let $\mathcal{V}_1, \ldots, \mathcal{V}_l$ be l non identical subsets of $\mathcal{V} - \{S\}$ such that $D \in \mathcal{V}_i$ for all $1 \leq i \leq l$. Also assume a product distribution on continuous random variables $X_i,\ i \in \mathcal{V}$. Then*

$$h(Y_{\mathcal{V}_2}|X_{\mathcal{V}_1}) + \cdots + h(Y_{\mathcal{V}_l}|X_{\mathcal{V}_{l-1}}) + h(Y_{\mathcal{V}_1}|X_{\mathcal{V}_l}) \geq \sum_{i=1}^{l} H(Y_{\tilde{\mathcal{V}}_i}|X_{\tilde{\mathcal{V}}_i}) \tag{A.85}$$

where for $k = 1, \ldots, l$,

$$\tilde{\mathcal{V}}_k \quad = \bigcup_{\{i_1,\ldots,i_k\}\subseteq\{1,\ldots,l\}} (\mathcal{V}_{i_1} \cap \cdots \cap \mathcal{V}_{i_k}) \tag{A.86}$$

or in another words each $\tilde{\mathcal{V}}_j$ is the union of $\binom{l}{j}$ sets such that each set is intersect of j of $\mathcal{V}_i$'s.

Proof. First note that

$$h(Y_{\mathcal{V}_2}|X_{\mathcal{V}_1}) + \cdots + h(Y_{\mathcal{V}_l}|X_{\mathcal{V}_{l-1}}) + h(Y_{\mathcal{V}_1}|X_{\mathcal{V}_l}) =$$
$$h(Y_{\mathcal{V}_2}, X_{\mathcal{V}_1}) + \cdots + h(Y_{\mathcal{V}_l}, X_{\mathcal{V}_{l-1}}) + h(Y_{\mathcal{V}_1}, X_{\mathcal{V}_l}) - \sum_{i=1}^{l} h(X_{\mathcal{V}_i})$$

and

$$\sum_{i=1}^{l} h(Y_{\tilde{\mathcal{V}}_i}|X_{\tilde{\mathcal{V}}_i}) = \sum_{i=1}^{l} h(Y_{\tilde{\mathcal{V}}_i}, X_{\tilde{\mathcal{V}}_i}) - \sum_{i=1}^{l} h(X_{\tilde{\mathcal{V}}_i}) \tag{A.87}$$

Now define the set

$$\mathcal{W}_i = \{Y_{\mathcal{V}_i}, X_{\mathcal{V}_{i-1}}\}, \quad i = 1, \dots, l \tag{A.88}$$

where $\mathcal{V}_0 = \mathcal{V}_l$.

It is easy to show that,

$$\sum_{i=1}^{l} h(X_{\mathcal{V}_i}) = \sum_{i=1}^{l} h(X_{\tilde{\mathcal{V}}_i}) \tag{A.89}$$

Therefore, we just need to prove that

$$\sum_{i=1}^{l} h(\mathcal{W}_i) \geq \sum_{i=1}^{l} h(Y_{\tilde{\mathcal{V}}_i}, X_{\tilde{\mathcal{V}}_i}) \tag{A.90}$$

Now, since the differential entropy function is a submodular function we have,

$$\sum_{i=1}^{l} h(\mathcal{W}_i) \geq \sum_{i=1}^{l} h(\tilde{\mathcal{W}}_i) \tag{A.91}$$

where

$$\tilde{\mathcal{W}}_r = \bigcup_{\{i_1,\ldots,i_r\}\subseteq\{1,\ldots,l\}} (\mathcal{W}_{i_1} \cap \cdots \cap \mathcal{W}_{i_r}), \quad r = 1,\ldots,l \tag{A.92}$$

Now for any r ($1 \le r \le l$) we have

$$\begin{aligned}
\tilde{\mathcal{W}}_r &= \bigcup_{\{i_1,\ldots,i_r\}\subseteq\{1,\ldots,l\}} (\mathcal{W}_{i_1} \cap \cdots \cap \mathcal{W}_{i_r}) \\
&= \bigcup_{\{i_1,\ldots,i_r\}\subseteq\{1,\ldots,l\}} (\{Y_{\mathcal{V}_{i_1}}, X_{\mathcal{V}_{i_1-1}}\} \cap \cdots \cap \{Y_{\mathcal{V}_{i_r}} X_{\mathcal{V}_{i_r-1}}\}) \\
&= \bigcup_{\{i_1,\ldots,i_r\}\subseteq\{1,\ldots,l\}} (\{Y_{\mathcal{V}_{i_1}\cap\cdots\cap\mathcal{V}_{i_r}}, X_{\mathcal{V}_{(i_1-1)}\cap\cdots\cap X_{\mathcal{V}_{(i_r-1)}}}\}) \\
&= \left\{Y_{\bigcup_{\{i_1,\ldots,i_r\}}(\mathcal{V}_{i_1}\cap\cdots\cap\mathcal{V}_{i_r})}, X_{\bigcup_{\{i_1,\ldots,i_r\}}(\mathcal{V}_{(i_1-1)}\cap\cdots\cap\mathcal{V}_{(i_r-1)})}\right\} \\
&= \{Y_{\tilde{\mathcal{V}}_r}, X_{\tilde{\mathcal{V}}_r}\}
\end{aligned}$$

Therefore by equation (A.91) we have,

$$\sum_{i=1}^{l} h(\mathcal{W}_i) \geq \sum_{i=1}^{l} h(\tilde{\mathcal{W}}_i) \tag{A.93}$$

$$= \sum_{i=1}^{l} h(Y_{\tilde{\mathcal{V}}_i}, X_{\tilde{\mathcal{V}}_i}) \tag{A.94}$$

Hence the Lemma is proved.

□

Now we are ready to prove Lemma 6.3.2. First note that any cut in the unfolded graph, Ω_{unf}, partitions the nodes at each stage $1 \le i \le K$ to $\mathcal{U}_i$ (on the left of the cut) and $\mathcal{V}_i$ (on the right of the cut). If at one stage $S[i] \in \mathcal{V}_i$ or $D[i] \in \mathcal{U}_i$ then the cut passes through one of the infinite capacity edges (capacity Kq) and hence the lemma is obviously proved. Therefore without loss of generality assume that $S[i] \in \mathcal{U}_i$ and $D[i] \in \mathcal{V}_i$ for all $1 \le i \le K$.

Now since for each $i \in \mathcal{V}$, $\{x_i[t]\}_{1\le t\le K}$ are i.i.d distributed we can write

$$I(Y_{\Omega^c_{\text{unf}}}; X_{\Omega_{\text{unf}}}|X_{\Omega^c_{\text{unf}}}) = \sum_{i=1}^{K-1} I(Y_{\mathcal{V}_{i+1}}; X_{\mathcal{U}_i}|X_{\mathcal{V}_i}) \tag{A.95}$$

Consider the sequence of $\mathcal{V}_i$'s. Note that there are total of $L = 2^{|\mathcal{V}|-2}$ possible subsets of $\mathcal{V}$ that contain D but not S. Assume that $\mathcal{V}_s$ is the first set that is revisited. Assume that it is revisited at step $\mathcal{V}_{s+l}$. We have,

$$\sum_{i=s}^{s+l-1} I(Y_{\mathcal{V}_{i+1}}; X_{\mathcal{U}_i}|X_{\mathcal{V}_i}) = \sum_{i=s}^{s+l-1} h(Y_{\mathcal{V}_{i+1}}|X_{\mathcal{V}_i}) - h(Y_{\mathcal{V}_{i+1}}|X_{\mathcal{V}_i}, X_{\mathcal{U}_i}) \tag{A.96}$$

Now by Lemma A.6.1 we have

$$\sum_{i=s}^{s+l-1} h(Y_{\mathcal{V}_{i+1}}|X_{\mathcal{V}_i}) \ge \sum_{i=1}^{l} h(Y_{\tilde{\mathcal{V}}_i}|X_{\tilde{\mathcal{V}}_i}) \tag{A.97}$$

where $\tilde{\mathcal{V}}_i$'s are as described in lemma A.6.1. Next, note that $h(Y_{\mathcal{V}_{i+1}}|X_{\mathcal{V}_i}, X_{\mathcal{U}_i})$ is just the entropy of channel noises, and since for any $v \in \mathcal{V}$ we have

$$|\{i|v \in \mathcal{V}_i\}| = |\{j|v \in \tilde{\mathcal{V}}_j\}| \tag{A.98}$$

, we get

$$\sum_{i=s}^{s+l-1} h(Y_{\mathcal{V}_{i+1}}|X_{\mathcal{V}_i}, X_{\mathcal{U}_i}) = \sum_{i=1}^{l} h(Y_{\tilde{\mathcal{V}}_i}|X_{\tilde{\mathcal{V}}_i}, X^c_{\tilde{\mathcal{V}}_i}) \tag{A.99}$$

Now by putting (A.97) and (A.99) together, we get

$$\begin{aligned}\sum_{i=s}^{s+l-1} I(Y_{\mathcal{V}_{i+1}}; X_{\mathcal{U}_i}|X_{\mathcal{V}_i}) &\ge \sum_{i=1}^{l} I(Y_{\tilde{\mathcal{V}}_i}; X_{\tilde{\mathcal{V}}_i^c}|X_{\tilde{\mathcal{V}}_i}) &\text{(A.100)}\\ &\ge l \min_{\Omega\in\Lambda_D} I(Y_{\Omega^c}; X_{\Omega}|X_{\Omega^c}) &\text{(A.101)}\end{aligned}$$

Now since in any $L-1$ time frame there is at least one loop, therefore except at most a path of length $L-1$ everything in $\sum_{i=1}^{K-1} I(Y_{\mathcal{V}_{i+1}}; X_{\mathcal{U}_i}|X_{\mathcal{V}_i})$. can be replaced with the value of the min-cut. Therefore,

$$\sum_{i=1}^{K-1} I(Y_{\mathcal{V}_{i+1}}; X_{\mathcal{U}_i}|X_{\mathcal{V}_i}) \geq (K-L+1) \min_{\Omega \in \Lambda_D} I(Y_{\Omega^c}; X_\Omega|X_{\Omega^c}) \tag{A.102}$$

and hence the proof is complete.

A.7 Proof of Lemma 8.5.2

To prove this lemma we need to first prove the following two main lemmas:

Lemma A.7.1. *Let G be the channel gains matrix of a $m \times n$ MIMO system. Assume that there is an average power constraint equal to one at each node. Then for any input distribution $P_{\mathbf{X}}$,*

$$|I(\mathbf{X}; [G\mathbf{X}+Z]) - I(\mathbf{X}; [G\mathbf{X}])| \leq 7n \tag{A.103}$$

where $Z = [z_1, \ldots, z_n]$ is a vector of n i.i.d. $\mathcal{N}(0,1)$ random variables.

Lemma A.7.2. *Let G be the channel gains matrix of a $m \times n$ MIMO system. Assume that there is an average power constraint equal to one at each node. Then for any input distribution $P_{\mathbf{X}}$,*

$$|I(\mathbf{X}; G\mathbf{X}+Z) - I(\mathbf{X}; [G\mathbf{X}+Z])| \leq n \tag{A.104}$$

where $Z = [z_1, \ldots, z_n]$ is a vector of n i.i.d. $\mathcal{N}(0,1)$ random variables.

Note that the main Lemma that we want to prove in this section (Lemma 8.5.2) is just a corollary of these two lemmas. The reason is the following:

$$\left.\begin{array}{r}|I(\mathbf{X};[G\mathbf{X}+Z])-I(\mathbf{X};[G\mathbf{X}])|\leq 7n\\ |I(\mathbf{X};G\mathbf{X}+Z)-I(\mathbf{X};[G\mathbf{X}+Z])|\leq n\end{array}\right\}\Rightarrow|I(\mathbf{X};G\mathbf{X}+Z)-I(\mathbf{X};[G\mathbf{X}])|\leq 8n \tag{A.105}$$

Therefore we just need to prove Lemma A.7.1 and Lemma A.7.2. In order to prove Lemma A.7.1 we need the following lemma and its corollary.

Lemma A.7.3. *Consider integer-valued random variables X, R and S such that*

$$X \perp R \tag{A.106}$$

$$S \in \{-L,\ldots,0,\ldots,L\} \tag{A.107}$$

$$\mathbb{P}\{|R|\geq k\} \leq e^{-f(k)}, \quad \textit{for all } k\in\mathcal{Z}^+ \tag{A.108}$$

for some integer L and a function $f(.)$. Let

$$Y = X+R+S \tag{A.109}$$

Then

$$H(Y|X) \leq 2\log_2 e\left(\sum_{k=1}^{\infty} f(k)e^{-f(k)}\right)+\frac{2L+1}{2}+\mathcal{N}_f \tag{A.110}$$

$$H(X|Y) \leq \log(2L+1)+2\log_2 e\left(\sum_{k=1}^{\infty} f(k)e^{-f(k)}\right)+\frac{2L+1}{2}+\mathcal{N}_f \tag{A.111}$$

where

$$\mathcal{N}_f = \left|\{n\in\mathcal{Z}^+|e^{-f(n)}>\frac{1}{2}\}\right| \tag{A.112}$$

Proof. By definition we have

$$\begin{aligned} H(Y|X) &= H(X+R+S|X) && \text{(A.113)}\\ &= H(R+S|X) && \text{(A.114)}\\ &\leq H(R+S) && \text{(A.115)}\\ &= -\sum_k \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} && \text{(A.116)} \end{aligned}$$

Now since $-p\log p \leq \frac{1}{2}$ for $0 \leq p \leq 1$, we have

$$-\sum_{k=-L}^{L} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} \leq \frac{2L+1}{2} \qquad \text{(A.117)}$$

Now note that for $|k| > L$ we have

$$\mathbb{P}\{R+S=k\} \leq \mathbb{P}\{|R| \geq |k|-L\} \leq e^{-f(|k|-L)} \qquad \text{(A.118)}$$

Since $p\log p$ is decreasing in p for $p < \frac{1}{2}$ we have

$$\begin{aligned} &-\sum_{k=L+1}^{\infty} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\}\\ &\quad= -\sum_{\substack{k>L\\ k-L\in\mathcal{N}_f}} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} - \sum_{\substack{k>L\\ k-L\notin\mathcal{N}_f}} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\}\\ &\quad\leq \frac{\mathcal{N}_f}{2} + \sum_{k=L+1}^{\infty} e^{-f(k-L)} f(k-L)\log e && \text{(A.119)} \end{aligned}$$

and similarly

$$
\begin{aligned}
&-\sum_{k=-\infty}^{-L} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} \\
&\quad = -\sum_{\substack{k<-L \\ |k|-L\in\mathcal{N}_f}} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} - \sum_{\substack{k<-L \\ |k|-L\notin\mathcal{N}_f}} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} \\
&\quad \leq \frac{\mathcal{N}_f}{2} + \sum_{k=L+1}^{\infty} e^{-f(k-L)} f(k-L)\log e \qquad \text{(A.120)}
\end{aligned}
$$

Now by combining (A.117), (A.119) and (A.120) we get

$$
H(Y|X) \leq 2\log_2 e\left(\sum_{k=1}^{\infty} f(k)e^{-f(k)}\right) + \frac{2L+1}{2} + \mathcal{N}_f \qquad \text{(A.121)}
$$

Now we prove the second inequality

$$
\begin{aligned}
H(X|Y) &= H(X|X+R+S) && \text{(A.122)} \\
&= H(X) - I(X;X+R+S) && \text{(A.123)} \\
&= H(X) - H(X+R+S) + H(X+R+S|X) && \text{(A.124)} \\
&\leq H(X) - H(X+R+S|S) + H(Y|X) && \text{(A.125)} \\
&= H(X) - H(X+R|S) + H(Y|X) && \text{(A.126)} \\
&= H(X) - H(X+R) + I(X+R;S) + H(Y|X) && \text{(A.127)} \\
&\leq H(X) - H(X+R) + H(S) + H(Y|X) && \text{(A.128)} \\
&\leq H(X) - H(X+R) + \log(2L+1) + H(Y|X) && \text{(A.129)} \\
&\leq H(X) - H(X+R|R) + \log(2L+1) + H(Y|X) && \text{(A.130)} \\
&= H(X) - H(X) + \log(2L+1) + H(Y|X) && \text{(A.131)} \\
&= \log(2L+1) + H(Y|X) && \text{(A.132)}
\end{aligned}
$$

Therefore

$$H(X|Y) \leq \log{(2L+1)} + 2\log_2 e \left(\sum_{k=1}^{\infty} f(k)e^{-f(k)}\right) + \frac{2L+1}{2} + \mathcal{N}_f \tag{A.133}$$

□

Corollary A.7.4. *Assume v is a continuous random variable, then*

$$H([v+z]||[v]) \quad \leq \quad 7 \tag{A.134}$$

$$H([v]||[v+z]) \quad \leq \quad 7 \tag{A.135}$$

where z is a $\mathcal{N}(0,1)$ random variable independent of v and $[.]$ is the integer part of a real number.

Proof. We use lemma A.7.3 with variables

$$X \quad = \quad [v] \tag{A.136}$$

$$R \quad = \quad [z] \tag{A.137}$$

$$S \quad = \quad [\{v\} + \{z\}] \tag{A.138}$$

Then $L = 1$ and since

$$\mathbb{P}\left\{|[z]| \geq k\right\} \quad \leq \quad \mathbb{P}\left\{|[z]| - \frac{1}{2} \geq k\right\} \tag{A.139}$$

$$= \quad 2Q(k - \frac{1}{2}) \tag{A.140}$$

$$\leq \quad e^{-\frac{(k-\frac{1}{2})^2}{2}} \tag{A.141}$$

Therefore

$$f(k) = \frac{(k-\frac{1}{2})^2}{2} \tag{A.142}$$

Also since

$$e^{-\frac{(k-\frac{1}{2})^2}{2}} < \frac{1}{2}, \quad \text{for } k \geq 3 \tag{A.143}$$

we have

$$\mathcal{N}_f = \{1, 2\} \tag{A.144}$$

Now we have

$$\log(2L+1) + 2\log_2 e\left(\sum_{k=1}^{\infty} f(k)e^{-f(k)}\right) + \frac{2L+1}{2} + \mathcal{N}_f \tag{A.145}$$

$$= 2\log_2 e\left(\sum_{k=1}^{\infty} \frac{(k-\frac{1}{2})^2}{2} e^{-\frac{(k-\frac{1}{2})^2}{2}}\right) + 3.5 + \log_2 3 \tag{A.146}$$

$$\approx 6.89 < 7 \tag{A.147}$$

As a result

$$H([v+z]||[v]) \leq 7 \tag{A.148}$$

$$H([v]||[v+z]) \leq 7 \tag{A.149}$$

□

Now we prove Lemma A.7.1.

Proof. **(proof of Lemma A.7.1)**

First note that

$$I(\mathbf{X};[G\mathbf{X}]) \leq I(\mathbf{X};[G\mathbf{X}+Z]) + I(\mathbf{X};[G\mathbf{X}]|[G\mathbf{X}+Z]) \tag{A.150}$$

$$= I(\mathbf{X};[G\mathbf{X}+Z]) + H([G\mathbf{X}]|[G\mathbf{X}+Z]) \tag{A.151}$$

$$\leq I(\mathbf{X};[G\mathbf{X}+Z]) + 7n \tag{A.152}$$

where the last step is true because of Corollary A.7.4. Also

$$I(\mathbf{X};[G\mathbf{X}+Z]) \leq I(\mathbf{X};[G\mathbf{X}]) + I(\mathbf{X};[G\mathbf{X}+Z]|[G\mathbf{X}]) \tag{A.153}$$
$$\leq I(\mathbf{X};[G\mathbf{X}]) + H([G\mathbf{X}+Z]|[G\mathbf{X}]) \tag{A.154}$$
$$\leq I(\mathbf{X};[G\mathbf{X}]) + 7n \tag{A.155}$$

where the last step is true because of Corollary A.7.4. Now from equations (A.152) and (A.155) we have

$$|I(\mathbf{X};[G\mathbf{X}+Z]) - I(\mathbf{X};[G\mathbf{X}])| \leq 7n \tag{A.156}$$

□

Now we prove Lemma A.7.2

Proof. **(proof of Lemma A.7.2)**
Define the following random variables:

$$\mathbf{Y} = G\mathbf{X}+Z \tag{A.157}$$
$$\hat{\mathbf{Y}} = [G\mathbf{X}+Z] \tag{A.158}$$
$$\tilde{\mathbf{Y}} = \hat{\mathbf{Y}}+\mathbf{U} \tag{A.159}$$

where $\mathbf{U} = [U_1, \ldots, U_n]$ is a vector of n i.i.d. U$[0,1]$ random variables, independent of $\mathbf{X}$ and Z.

Now by data processing inequality we have

$$I(\mathbf{X};\mathbf{Y}) \geq I(\mathbf{X};\hat{\mathbf{Y}}) \geq I(\mathbf{X};\tilde{\mathbf{Y}}) \tag{A.160}$$

Now note that,

$$\begin{aligned}
I(\mathbf{X};\mathbf{Y}) - I(\mathbf{X};\tilde{\mathbf{Y}}) &= h(\mathbf{Y}) - h(\tilde{\mathbf{Y}}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - h(\mathbf{Y}|\mathbf{X}) && \text{(A.161)}\\
&= h(\mathbf{Y}) - h(\tilde{\mathbf{Y}}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - \frac{n}{2}\log(2\pi e) && \text{(A.162)}\\
&= h(\mathbf{Y}|\tilde{\mathbf{Y}}) - h(\tilde{\mathbf{Y}}|\mathbf{Y}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - \frac{n}{2}\log(2\pi e) && \text{(A.163)}\\
&= h(\mathbf{Y}|\tilde{\mathbf{Y}}) - h(\mathbf{U}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - \frac{n}{2}\log(2\pi e) && \text{(A.164)}\\
&= h(\mathbf{Y}|\tilde{\mathbf{Y}}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - \frac{n}{2}\log(2\pi e) && \text{(A.165)}
\end{aligned}$$

where the last step is true since $h(\mathbf{U}) = nh(U_1) = n\log 1 = 0$. Now note that

$$|y - \tilde{y}| \leq \max(|[x+z] - x|) + \max|u| = \frac{3}{2} \quad \text{(A.166)}$$

Therefore

$$\begin{aligned}
h(\mathbf{Y}|\tilde{\mathbf{Y}}) &= h(\mathbf{Y} - \tilde{\mathbf{Y}}|\tilde{\mathbf{Y}}) && \text{(A.167)}\\
&\leq n\log\max(|y - \tilde{y}|) && \text{(A.168)}\\
&= n\log\frac{3}{2} && \text{(A.169)}
\end{aligned}$$

For the second term we have,

$$\begin{aligned}
\tilde{y} &= [g_1\mathbf{x} + z] + U && \text{(A.170)}\\
&= g_1\mathbf{x} + z + \delta(g_1\mathbf{x} + z) + U && \text{(A.171)}
\end{aligned}$$

where $\delta(x) = x - [x]$ is a function representing the difference between a real number and its closest integer, and g_1 is the first row of G. Clearly $|\delta(x)| \leq \frac{1}{2}$ for all $x \in \mathcal{R}$. Therefore given X the variance of $\tilde{\mathbf{Y}}$ is bounded by

$$\begin{aligned}
\text{Var}\left[\tilde{Y}|\mathbf{X}\right] &= \text{Var}\left[z+\delta(g_1\mathbf{X}+z)+U\right] && \text{(A.172)}\\
&\leq \text{Var}\left[z\right]+\text{Var}\left[\delta(g_1\mathbf{X}+z)|\mathbf{X}\right]+2\text{Cov}\left[Z,\delta(\mathbf{X}+z)|\mathbf{X}\right]+\\
&\quad +\text{Var}\left[U\right] && \text{(A.173)}\\
&\leq \text{Var}\left[z\right]+|\max\delta(.)|^2+2\sqrt{\text{Var}\left[Z\right]\times\max\delta(.)}+\text{Var}\left[U\right] && \text{(A.174)}\\
&= 1+\frac{1}{4}+\sqrt{2}+\frac{1}{12} && \text{(A.175)}
\end{aligned}$$

Therefore

$$\begin{aligned}
h(\tilde{\mathbf{Y}}|\mathbf{X}) &\leq nh(Y|\tilde{\mathbf{Y}}) && \text{(A.176)}\\
&\leq \frac{n}{2}\log 2\pi e\text{Var}\left[Y|\tilde{\mathbf{Y}}\right] && \text{(A.177)}\\
&\leq \frac{n}{2}\log 2\pi e(1+\frac{1}{4}+\sqrt{2}+\frac{1}{12}) && \text{(A.178)}
\end{aligned}$$

Now from equation (A.165), (A.169) and (A.165) we have

$$\begin{aligned}
I(\mathbf{X};\mathbf{Y})-I(\mathbf{X};\tilde{\mathbf{Y}}) &\leq h(\mathbf{Y}|\tilde{\mathbf{Y}})+h(\tilde{\mathbf{Y}}|\mathbf{X})-\frac{n}{2}\log\left(2\pi e\right)\\
&\leq n\log\frac{3}{2}+\frac{n}{2}\log 2\pi e(1+\frac{1}{4}+\sqrt{2}+\frac{1}{12})-\frac{n}{2}\log 2\pi e\\
&\approx 0.81n<n && \text{(A.179)}
\end{aligned}$$

□

A.8 Proof of Theorem 8.2.1

Before proving this theorem, we first state and prove the following lemma

Lemma A.8.1. *Consider a Gaussian relay network, $G_{Gaussian}$, where all channel gains are in the form $\rho^{n_{ij}}$ where n_{ij}'s are non negative fixed integers and $\rho \in \mathcal{N}$ is a variable. Denote the maximum degree of nodes in this graph by η. Now consider, $G_{LFF\text{-}det}$, the linear finite field deterministic network associated with $G_{Gaussian}$. Then we have the following connection between $C_{Gaussian}(\rho)$, the capacity of the Gaussian relay network and $C_{LFF\text{-}det}(p)$, the capacity of the corresponding linear finite field deterministic relay network with finite field size p:*

$$C_{LFF\text{-}det}(\lfloor \rho/\eta \rfloor_p) \leq C_{Gaussian}(\rho) \leq C_{LFF\text{-}det}(\lceil \rho\eta \rceil_p) \tag{A.180}$$

where for any $x \in$
$mathcalR^+$, $\lfloor x \rfloor_p$ and $\lceil x \rceil_p$ are respectively the closest prime number smaller than x and larger than x.

Proof. First by Theorem 8.5.1 we know that

$$C_{\text{truncated}}(\rho) - 13|V| \leq C_{\text{Gaussian}}(\rho) \leq C_{\text{truncated}}(\rho) + 13|V| \tag{A.181}$$

Now we prove that

$$C_{\text{LFF-det}}(\lfloor \rho/\eta \rfloor_p) \leq C_{\text{truncated}}(\rho) \leq C_{\text{LFF-det}}(\lceil \rho\eta \rceil_p) \tag{A.182}$$

Here we just prove the first inequality, the second inequality can also be proved similarly. Assume we have an achievability scheme for $G_{\text{LFF-det}}$ with field size $\lfloor \frac{\rho}{\eta} \rfloor_p$. We show that there is a corresponding scheme in $G_{\text{truncated}}(\rho)$ such that all nodes transmit and receive identical signals in both linear finite field and truncated deterministic models.

Now we describe the corresponding communication scheme for the truncated,

- Source, S, will transmit $x_S = \sum_{i=1}^{q} x_S^{det}(i)\rho^{-i}$.
- Each node, v will receive a signal y_v and then,

1. Finds the q-ary representation of y_v in base ρ: $y_v = \sum_{i=1}^{q} y_v(i)\rho^{i-1}$
2. Computes the modulo of each component base $\lfloor \frac{\rho}{\eta} \rfloor_p$ to create an element in $\mathbb{F}^q_{\lfloor \rho/\eta \rfloor_p}$, *i.e.* sets, $y_v(i) = y_v(i) \mod \rho^{\lfloor \frac{\rho}{\eta} \rfloor_p}$ for $i = 1, \ldots, q$.
3. Uses the mapping described by the linear finite field deterministic communication scheme to find x_v^{det} and transmits $x_v = \sum_{i=1}^{q} x_v^{det}(i)\rho^{-i}$

- The destination also creates y_D and uses it for decoding

Now it is easy to see that since there is no carry-over between adjacent signal levels, y_D is the same in both deterministic models. Hence

$$C_{\text{LFF-det}}(\lfloor \rho/\eta \rfloor_p) \leq C_{\text{truncated}}(\rho) \tag{A.183}$$

Now by (A.181) and (A.182) the proof is complete. □

Now we prove Theorem 8.2.1. Assume ρ is integer, by lemma A.8.1 we have,

$$C_{\text{LFF-det}}(\lfloor \rho/\eta \rfloor_p) \leq C_{\text{Gaussian}}(\rho) \leq C_{\text{LFF-det}}(\lceil \rho\eta \rceil_p) \tag{A.184}$$

Now note that

$$C_{\text{LFF-det}}(\lfloor \rho/\eta \rfloor_p) = \min_{\Omega} \text{rank}(\mathbf{H}_\Omega) \log\left(\lfloor \rho/\eta \rfloor_p\right) \tag{A.185}$$

where rank is evaluated in finite field $\mathbb{F}_{\lfloor \rho/\eta \rfloor_p}$. Also

$$C_{\text{LFF-det}}(\lceil \rho\eta \rceil_p) = \min_{\Omega} \text{rank}(\mathbf{H}_\Omega) \log\left(\lceil \rho\eta \rceil_p\right) \tag{A.186}$$

where rank is evaluated in finite field $\mathbb{F}_{\lceil \rho\eta \rceil_p}$. Now first note for large enough ρ both these rank evaluations are the same and equal to the the rank of the corresponding matrix in $\mathcal{R}$.

Furthermore since

$$\lfloor \rho/\eta \rfloor_p \geq \frac{\rho}{2\eta} \tag{A.187}$$

$$\lceil \rho\eta \rceil_p \leq 2\rho\eta \tag{A.188}$$

we have

$$\lim_{\rho\to\infty} \frac{\log\left(\lfloor \rho/\eta \rfloor_p\right)}{\log \rho} = 1 \tag{A.189}$$

$$\lim_{\rho\to\infty} \frac{\log\left(\lceil \rho\eta \rceil_p\right)}{\log \rho} = 1 \tag{A.190}$$

This completes the proof.