Distributed Coordination in Network Information Theory

Brett E. Schein
Distributed Coordination in Network Information Theory

by

Brett E. Schein

B.S. E.E.C.S., University of California, Berkeley (1994)
S.M. E.E.C.S., Massachusetts Institute of Technology (1996)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2001

© Massachusetts Institute of Technology 2001. All rights reserved.

Author .................................................................

Department of Electrical Engineering and Computer Science

August 20, 2001

Certified by .............................................................

Robert Gallager
Professor of Electrical Engineering
Thesis Supervisor

Accepted by ............................................................

Arthur C. Smith
Chairman, Department Committee on Graduate Students
Distributed Coordination in Network Information Theory

by

Brett E. Schein

Submitted to the Department of Electrical Engineering and Computer Science on August 20, 2001, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering

Abstract

Constructing a large-scale wireless data network is spectacularly expensive. It is therefore important to understand how to efficiently utilize the physical infrastructure and available battery power, which are expensive system resources. Unfortunately, we currently understand very little about efficient communication in a distributed environment.

In distributed wireless networks, there appears to be an interesting and complex trade-off between trying to take advantage of independent noisy signals at different relays and closely coordinating relay transmissions to a receiver. Designing the right structure for efficient communication, by choice of source transmission codebook and relay terminal processing, is the important and difficult problem on which we focus.

We use an information theoretic framework to study several very simple multiple terminal networks, focusing exclusively on single source, single destination networks where communication must take place through intermediate nodes. Our goal is to determine how much data we can get reliably from source to destination, placing no importance on delay or computational complexity. The core problem then involves distributed detection at the intermediate nodes and coordination in relaying information to the destination.

Thesis Supervisor: Robert Gallager
Title: Professor of Electrical Engineering


Acknowledgments

I would like to thank Professor Robert Gallager, my thesis advisor, for agreeing to work with me despite planning on retirement and advising twelve other students at the time. His sincere interest in his students as well as his approach to teaching and research are truly inspiring.

My thesis committee members, Professors Sanjoy Mitter and Greg Wornell, were always available and eager to meet with me. I have benefited from their wisdom in matters technical and otherwise.

I would like to thank Professors Dave Forney, Amos Lapidoth, and Eytan Modiano, as well as ex-Professor and future co-worker Mitch Trott, for their advice, encouragement, and friendship. I hope I can repay them in-kind in the future.

My fellow students helped create a tremendously rewarding experience here at LIDS. I would especially like to thank Ibrahim Abou-Faycal, Randy Berry, Sae-Young Chung, Stark Draper, Thierry Klein, Patrick Maurer, and Sekhar Tatikonda for letting me pester them relentlessly with technical problems disguised as coffee breaks.

I would like to thank my parents for their love and support throughout my life. They started with the foundation and have never stopped adding to the structure.

For my wife, Joan, my best attempt to thank her here would be insufficient. I thank the world every day that we get to share our lives together.

This research was supported in part by MIT Lincoln Laboratory and by Draper Laboratory.
Contents

1 Introduction 13
  1.1 The Network Model and Specific Examples .................. 17
  1.2 Background ................................................. 22

2 Gaussian Parallel Relay Networks 29
  2.1 Gaussian Broadcast and Multiaccess Channels ............... 29
  2.2 Converses for Gaussian Parallel Relay Networks .......... 39
  2.3 Coding for Gaussian Parallel Relay Networks .......... 52
    2.3.1 Coding to the Relays .................................. 52
    2.3.2 Forwarding Observations Without Decoding at the Relays . 60
  2.4 On Degraded Channels ..................................... 68
  2.5 Focusing on Symmetric Gaussian Networks .................. 72
    2.5.1 Forwarding Observations Without Decoding at the Relays; Low
          SNR Effects ............................................ 72
    2.5.2 Combining the Two Main Approaches ................... 81
    2.5.3 An Idea That Does Not Work — Using a Gauss-Markov Structure
          for the Input ............................................ 83
    2.5.4 Applying Block Quantization and Independent Messages .. 84
    2.5.5 Difficulty in Applying the Entropy Power Inequality for Gaussian
          Parallel Relay Networks ................................ 86

3 Discrete Memoryless Parallel Relay Networks 89
  3.1 Discrete Broadcast and Multiaccess Channels ............... 92
  3.2 Converses for Discrete Memoryless Networks ............... 93
3.3 Coding for Discrete Symmetric Networks
3.3.1 Coding to the Relays
3.3.2 Binning to Communicate Observations Without Decoding at the Relays
3.3.3 Block Quantization and Binning of the Relay Observations
3.3.4 Reproducing a Sufficient Statistic for Estimating $X^n$ Based on $(Y_1^n, Y_2^n)$

3.4 Capacity Equals End-to-End Mutual Information

3.5 Coding for Highly Asymmetric Networks
3.5.1 Coding to the Relays
3.5.2 Binning to Communicate Observations Without Decoding at the Relays
3.5.3 Capacity of the Highly Asymmetric PBSC Network with Feedback to Relay 2

3.6 Capacity Problem for Highly Asymmetric Networks

A Extended Gaussian multiaccess converse

B Coding to Relays for an Asymmetric Network

C A Lemma About Randomly Tossed Balls

D Several As Yet Fruitless Ideas for Conveses
D.1 Conjecture for Symmetric Discrete Networks
D.2 Find a network where we can reasonably guess $C_{net}$

E Summary of Main Graphical Results
List of Figures

1-1 Simple network of wireless terminals .................................. 14
1-2 The parallel relay network .............................................. 18
1-3 The parallel relay network consists of two well-known channels .... 19
1-4 The Gaussian parallel relay network .................................. 19
1-5 The PBSC network with noiseless relay channels ..................... 20
1-6 The Gaussian relay channel ............................................ 24

2-1 The Gaussian parallel relay network .................................. 29
2-2 The four SNR’s parametrizing the Gaussian network ............... 30
2-3 The Gaussian broadcast channel ....................................... 31
2-4 The Gaussian multiaccess channel .................................... 32
2-5 Broadcast cut-set bound .................................................. 41
2-6 Multiaccess cut-set bound ............................................... 43
2-7 Pure cross cut-set bound .................................................. 44
2-8 Cross cut-set bound ....................................................... 45
2-9 Evaluation of converses for a symmetric Gaussian network, \( S_1 = S_2 = 2 \) ........................................... 51
2-10 Comparing the broadcast and extended multiaccess capacity regions . 55
2-11 A highly asymmetric network regime — weak cross-cut .......... 57
2-12 Logically equivalent network when both relays access both observations 60
2-13 Evaluation of converses and the two basic achievability schemes for a symmetric Gaussian network, \( S_1 = S_2 = 2 \) ........................................... 63
2-14 Amplifying observations in the asymmetric Gaussian network .... 65
2-15 The physically degraded Gaussian relay channel ................. 69
2-16 The stochastically degraded Gaussian relay channel ............... 69
2-17 Idealized model for feedback for the Gaussian relay channel .... 71


<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-18</td>
<td>Additive white Gaussian noise channel</td>
</tr>
<tr>
<td>2-19</td>
<td>Asymptotic behavior of burst-optimized approach</td>
</tr>
<tr>
<td>3-1</td>
<td>The asymmetric PBSC network with noiseless relay channels</td>
</tr>
<tr>
<td>3-2</td>
<td>Evaluation of converse results for typical symmetric and asymmetric PBSC networks</td>
</tr>
<tr>
<td>3-3</td>
<td>Evaluation of the two basic achievability schemes for the symmetric PBSC network</td>
</tr>
<tr>
<td>3-4</td>
<td>Evaluation of the backoff scheme for the symmetric PBSC network</td>
</tr>
<tr>
<td>3-5</td>
<td>Parallel relay network with noiseless relay channels</td>
</tr>
<tr>
<td>3-6</td>
<td>Evaluation of the reduced $H(Y_1,Y_2)$ scheme for the symmetric PBSC network</td>
</tr>
<tr>
<td>3-7</td>
<td>Artificial broadcast channel used in the simultaneously combined approach</td>
</tr>
<tr>
<td>3-8</td>
<td>Evaluation of using broadcast codes to combine coding to the relays and communicating observations for the symmetric PBSC network</td>
</tr>
<tr>
<td>3-9</td>
<td>Artificially expanded output alphabet</td>
</tr>
<tr>
<td>3-10</td>
<td>Block quantization and binning approach for a PBSC network</td>
</tr>
<tr>
<td>3-11</td>
<td>Evaluation of the block quantization and binning scheme for the symmetric PBSC network, $C_1 = C_2 = 0.5$</td>
</tr>
<tr>
<td>3-12</td>
<td>Mathematical trick to prove (3.217)</td>
</tr>
<tr>
<td>3-13</td>
<td>Showing channel $B$ is a stochastically degraded version of channel $A$ for the symmetric PBSC broadcast channel</td>
</tr>
<tr>
<td>3-14</td>
<td>Parallel relay network with noiseless relay channels</td>
</tr>
<tr>
<td>3-15</td>
<td>Creating a point to point channel based on superletters</td>
</tr>
<tr>
<td>3-16</td>
<td>No continuity of flow for mutual information</td>
</tr>
<tr>
<td>3-17</td>
<td>The highly asymmetric PBSC network</td>
</tr>
<tr>
<td>3-18</td>
<td>Evaluation of results for the highly asymmetric PBSC network, $C_2 = 0.5172$</td>
</tr>
<tr>
<td>3-19</td>
<td>The highly asymmetric PBSC network with feedback to Relay 2</td>
</tr>
<tr>
<td>3-20</td>
<td>Series of two channels — an end-to-end mutual information problem</td>
</tr>
<tr>
<td>C-1</td>
<td>General form of the function $g_{f,k}(v)$</td>
</tr>
<tr>
<td>E-1</td>
<td>Evaluation of results for a symmetric Gaussian network, $S_1 = S_2 = 2$</td>
</tr>
</tbody>
</table>
E-2 Evaluation of results for the symmetric PBSC network, $C_1 = C_2 = 0.5$ 224
E-3 Evaluation of the block quantization and binning scheme for the symmetric PBSC network, $C_1 = C_2 = 0.5$ 225
E-4 Evaluation of results for the highly asymmetric PBSC network, $C_2 = 0.5$ 226
Chapter 1

Introduction

Wireless communication networks have by now evolved well beyond simple voice-based cellular telephony. So called “2.5,” third, and fourth generation cellular phone systems are currently being designed and standardized with data transfer capabilities as a dominant feature. It is a widely held belief that wireless access to the Internet is desired by the general consumer (both business and personal). This belief is supported by the wildly successful yet primitive data services of NTT DoCoMo’s I-mode phones in Japan and by the rapid adoption of wireless text messaging in Europe. Many companies and technologies will compete to provide wireless data services in one form or another, and whether the traditional cellular phone architecture will dominate is still to be determined by a plethora of start-up companies, traditional infrastructure and service provider companies, and the all-powerful governmental and market forces.

Constructing a large-scale wireless data network is spectacularly expensive. It is therefore important to understand how to efficiently utilize the physical infrastructure and available battery energy. Unfortunately, we currently understand very little about efficient communication in a distributed multiple user environment.

In this thesis, we are motivated by networks where power is of primary importance, such as cellular phone and ad hoc (e.g., packet radio) networks. In such situations, terminals may cooperate with each other in sending information through the network. Consider a cellular phone network in a rural environment, where the cell size is on the order of a square mile. In order to save power, it may make sense for a cellular phone
CHAPTER 1. INTRODUCTION

located far from its base station to transmit its message to an intermediate phone, which can in turn forward the message to the base station. In a military setting, such as a packet radio network without base stations, packet radios have no choice but to help each other send messages through the network. When there are several terminals available for relaying information, it becomes difficult to understand how to design the system most efficiently.

Intuitively, there are three physical phenomena which we would like to exploit for an efficient network design. We will articulate all three in the context of Figure 1-1. In this figure, there is a distant source that wants to transmit information to the base station. There are two wireless terminals located between the source and the base station, labeled Relay 1 and Relay 2. The two relays have no information of their own to send.

![Figure 1-1: Simple network of wireless terminals](image)

First, received power decreases rapidly with distance, generally as a power law — \(d^{-\alpha}\), for some \(\alpha \geq 2\) determined by the environment. One simple but effective way for the terminals to cooperate is to use intermediate nodes as relays. Referring to Figure 1-1, assume that Relay 1 is about halfway between the source and base station. Instead of broadcasting its message directly to the base station, the source can instead broadcast its message to Relay 1. Relay 1 can decode the message and then transmit it to the base station. If received power decays with distance as \(d^{-2}\), then this requires roughly half the total power compared with transmitting directly from source to destination — requiring one quarter from the source and one quarter from Relay 1. This type of communication is often called multi-hopped communication since a message is “hopped” between intermediate terminals. Exploiting this can save battery power and increase the physical coverage area of a base station.
Second, since the multipath characteristics differ between physical locations, different terminals generally receive independently faded versions of a source signal. Additionally, even in the absence of time-varying fading, the internal noise of each receiver is independent. Consequently, there is statistical diversity amongst the various receivers in a distributed network. In Figure 1-1, Relay 1 and Relay 2 each have an independent observation of the signal broadcast from the source. In particular, the versions of the source signal received at Relay 1 and Relay 2 are independently faded. In a slowly fading environment, if \( p \) is the probability that a particular relay experiences a deep fade, then the probability that both relays experience deep fades simultaneously is \( p^2 \). This “spatial diversity” can be exploited in practice by observing that a source message can be relayed to the base station if either relay successfully decodes it (e.g., see [30] for work along these lines). If the terminals are stationary, then there is no time variation in the strength of the source signal received at each relay. However, each relay still observes the source signal with independent additive noise (possibly with other distorting effects, which we neglect). In Figure 1-1, there are two independent noisy observations of the source signal within the network (plus a third extremely weak observation at the base station, which we neglect). Two independent noisy observations are better than one, and we would like to take advantage of the multiple observations in the network. However, these two relay observations are available at different points, i.e., no single entity has access to both observations. If the source is sending codewords, and if a single relay observation is not good enough to reliably decode the source codeword, then the two relays must coordinate in processing their observations so that the base station can ultimately decode the source codeword. We are then left with a difficult distributed detection problem. We will return to this point in a moment.

Third, each relay in Figure 1-1 acts as a separate transmitter to the base station. Current cellular systems are designed to minimize multiaccess interference at the base station by orthogonalizing the transmissions to the base station. This is commonly done by transmitting in different time or frequency slots, or by transforming relay signals using nearly orthogonal direct-sequence spreading codes. For the network communication problem we have been discussing, on the other hand, we would ideally like the two relays to act as a single transmitter with multiple antennas — we could
CHAPTER 1. INTRODUCTION

then take advantage of various transmit diversity techniques. When the terminals are stationary, the ideal transmit diversity technique involves coherent combining of signals at the base station. To illustrate, assume Relay 1 transmits a signal which is received at the base station with energy $E$ (if Relay 2 were silent). Assume Relay 2 transmits a scaled version of the same signal, also received with energy $E$ (if Relay 1 were silent). If the base station can coherently combine the two signals, then the net received signal has energy $4E$ rather than $2E$. This technique is sometimes referred to as beam-forming, and sometimes as using the two transmitters as a phased array. The result is an effective increase in received power at the base station. In a wireless environment, such coherent combining would require synchronism between the RF carriers rather than transmission symbols, thus minimally requiring some form of feedback or common and precise timing information. This level of precision may be difficult to achieve in a real, distributed system, but it is important to investigate its potential for improving both real-world systems and theoretical understanding of networks.

There appears to be an interesting and complex trade-off between trying to exploit the second and third phenomena simultaneously, namely, taking advantage of independent relay observations and closely coordinating relay transmissions to the base station. The effective power boost at the base station increases as the relay transmissions become more closely related. On the other hand, the more closely related the relay transmissions, the less independent information about the source they can communicate to the base station. This is a basic trade-off, but the communication problem is significantly more complicated. Coordinating relay transmissions for an effective power boost at the base station requires some form of common information at the two relays. The source must get its information to the relays by broadcasting its signal over parallel, noisy paths. Designing the right structure for efficient communication, by choice of source transmission codebook and relay terminal processing, is the important and difficult problem.

To study the problem, we must define a framework in which to ask relevant questions. Many factors may be taken into account when defining “system performance,” such as implementation cost, communication protocol complexity, end-to-end delay, power efficiency, statistical variation due to channel fading and noise, etc. There are
complicated trade-offs involved, and, speaking generally, it is often impossible to get concrete results without greatly simplifying the problem. A principle that works well is to isolate the issues we would like to understand, using simple mathematical models that incorporate all of the essential structure of the problem while eliminating unnecessary complexity. We follow this principle with our information theoretic approach in this thesis.

Even after adopting simple models for noisy channels, multiple terminal networks remain poorly understood. The information theory community has studied two basic multiple user channels, the broadcast channel and the multiaccess channel, and we know the capacity regions of large classes of both types. To gain insight into more general networks, it makes sense to begin by putting these two basic channels together. In particular, we will introduce the parallel relay network, comprised of a broadcast channel followed by a multiaccess channel. Surprisingly, little is known about what happens when these two well-studied channels are combined. We study several versions of these simple multiple terminal networks, focusing exclusively on single source, single destination problems where communication must take place through intermediate nodes. Our goal is to determine how much data we can get reliably from source to destination, placing no importance on delay or computational complexity. The core problem then involves distributed detection at the intermediate nodes and coordination in relaying information to the destination. In this thesis, we will derive several interesting and reasonably straightforward results that yield some insight into the problem. Unfortunately, we will also demonstrate that, beyond the straightforward results, it is a very difficult problem.

1.1 The Network Model and Specific Examples

The general system we will study is the discrete-time communication network pictured in Figure 1-2. We call this the parallel relay network. It consists of an input terminal with input $X$, two relay terminals which observe $Y_1$ and $Y_2$ and transmit $W_3$ and $W_4$, and an ultimate decoder which receives $Y$. The input, relay observation, and relay transmission alphabets are denoted by $\mathcal{X}$, $\mathcal{Y}_1$ or $\mathcal{Y}_2$, and $\mathcal{W}_1$ or $\mathcal{W}_2$, re-
spective. The input terminal is connected to the relays by a memoryless broadcast channel with conditional transition probability denoted by $p(y_1, y_2 | x)$. The relays are connected to the decoder by a memoryless multiaccess channel with conditional transition probability denoted by $p(y | w_3, w_4)$. The only source of extrinsic information is the input terminal, and the only purpose of the relays is to help get information from the input terminal to the decoder. Each relay is an arbitrarily complex processor whose transmission signal is a function, possibly a random function, of its observation signal. For convenience, we define the absolute time bases appropriately at each terminal so that $W_{3,k}$ and $W_{4,k}$ are functions of $Y_{1,k}^{k-1}$ and $Y_{2,k}^{k-1}$, respectively. Our goal in this thesis is to determine the capacity of the network, defined as the supremum of reliably achievable communication rates from the input terminal to the decoder.

The parallel relay network is equivalent to the series connection of two well-studied multiterminal channels, along with intermediate relays (processors), as pictured in Figure 1-3. If the relays were not present, that is, if the broadcast channel outputs were directly connected to the multiaccess channel inputs, then the network would logically degenerate to a memoryless single-user channel. With the relays present, however, the problem is difficult. On the left side is a memoryless broadcast channel, and on the right side is a memoryless multiaccess channel. Much is known about both of these channels, in the sense that we know the capacity region of a rather general class of memoryless broadcast channels, and we know the capacity region of memoryless multiaccess channels. Unfortunately, even after completing the work of this thesis, we do not completely understand what happens when we connect

---

1 We think of the discrete-time model as being derived from a continuous-time system. Implicitly, then, there is an arbitrary time delay between terminals due to propagation delay and projection onto signal space basis functions. The only requirement we need is for the relays to synchronize their transmissions to the decoder, allowing for arbitrary and possibly different delays between “inputs” and “outputs” at the relays.
1.1. THE NETWORK MODEL AND SPECIFIC EXAMPLES

Figure 1-3: The parallel relay network consists of two well-known channels

the two together with intermediate relays, yielding the parallel relay network. The capacity regions of the broadcast and multiaccess channels, by definition, concern independent streams of information to and from the users. In the parallel relay network, the information is bifurcated on the broadcast channel, processed at separate points by the relay terminals, and reconverged on the multiaccess channel. The bifurcation, distributed processing, and reconvergence of dependent information is the core problem that we will face in this thesis.

We will consider both Gaussian and discrete-alphabet parallel relay networks. The Gaussian version is pictured in Figure 1-4. If $x_k, w_{3,k},$ and $w_{4,k}$ are the three terminal inputs at time $k,$ then $Y_{1,k} = x_k + Z_{1,k},$ $Y_{2,k} = x_k + Z_{2,k},$ and $Y_k = w_{3,k} + w_{4,k} + Z_k.$ The three noise processes, $\{Z_{1,k}\}, \{Z_{2,k}\},$ and $\{Z_k\},$ are independent. Additionally, each noise process is i.i.d. Finally, $Z_{1,k}, Z_{2,k},$ and $Z_k$ are zero-mean Gaussian random variables with variances $N_{Z_1}, N_{Z_2},$ and $N_Z,$ respectively.

Figure 1-4: The Gaussian parallel relay network

For this Gaussian network, we further assume that all three terminals have average
power constraints. In the limit as \( n \to \infty \),

\[
\frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \{ X_k^2 \} \leq P_X, \quad \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \{ W_{3,k}^2 \} \leq P_{W_3}, \quad \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \{ W_{4,k}^2 \} \leq P_{W_4}.
\]

The Gaussian parallel relay network is effectively parametrized by four signal to noise ratios (SNR’s): \( S_1 = \frac{P_X}{N_{Z_1}} \), \( S_2 = \frac{P_X}{N_{Z_2}} \), \( S_3 = \frac{P_{W_3}}{N_Z} \), and \( S_4 = \frac{P_{W_4}}{N_Z} \).

For discrete-alphabet versions, we will focus primarily on the parallel binary symmetric channel (PBSC) network, pictured in Figure 1-5. The link from input \( X \) to Relay 1 observation \( Y_1 \) is a binary symmetric channel (BSC) with crossover probabil-

![](image.png)

**Figure 1-5:** The PBSC network with noiseless relay channels

ity \( \alpha_1 \). Similarly, the link from input \( X \) to Relay 2 observation \( Y_2 \) is an independent BSC with crossover probability \( \alpha_2 \). As implied by the figure, the relay observations \( Y_1 \) and \( Y_2 \) are conditionally independent given the input \( X \) (i.e., \( Y_1 \to X \to Y_2 \) form a Markov chain). The multiaccess side of the PBSC is highly degenerate. It consists of a pair of non-interfering, noiseless binary channels which can be used at a symbol rate different from the input symbol rate. In particular, for each \( n \) input uses, denoted \( X^n \), we can transmit \( nR_3 \) binary digits, denoted \( W_3 \), noiselessly from Relay 1 to the receiver. Similarly, we can transmit \( nR_4 \) binary digits, denoted \( W_4 \), noiselessly from Relay 2 to the receiver. Defining exactly when a new bit can be transmitted on the noiseless links relative to the number of symbols observed thus far can be handled easily but is irrelevant, and therefore we will not provide the details. We study this network primarily because of its simplicity, and we will provide additional motivation for studying this particular network when we discuss our results in Chapter 3. It does not appear that we learn anything new when studying a network with such a degenerate multiaccess channel paired with a Gaussian broadcast channel. Therefore, though the techniques we use can be applied, we do not explicitly present results for that network.
1.1. THE NETWORK MODEL AND SPECIFIC EXAMPLES

We introduce several other illustrative discrete-alphabet networks as issues arise. In all of the networks we consider, the broadcast channel observations are conditionally independent given the input, $p(y_1, y_2|x) = p(y_1|x)p(y_2|x)$. This is a natural and simplifying assumption for our purposes. There are some problems of a similar flavor in the literature where this channel structure is necessary to reach any reasonable conclusions. We will remark upon these later in Section 1.2.

For much of our work, we study networks which are completely symmetric top to bottom. That is, for the Gaussian network of Figure 1-4, we assume that the relay observation noises have the same power, $N_{Z_1} = N_{Z_2}$, and that the relays each have the same power constraint, $P_{W_3} = P_{W_4}$. We refer to this as the symmetric Gaussian network. For the PBSC network of Figure 1-5, we assume that the BSC’s have the same crossover probability, $\alpha_3 = \alpha_2$, and that the noiseless links each have the same rate, $R_3 = R_4$. We refer to this as the symmetric PBSC network. Generically, we refer to such networks as symmetric networks. At times we consider networks where the broadcast channel is symmetric while the multiaccess channel is not. Generically, we refer to these as networks with a symmetric broadcast channel. Studying symmetric networks has two advantages. First, numerical results are easier to present. For the symmetric Gaussian and PBSC networks, we reduce the parametrization from four dimensions down to two. Second, we avoid some complexity when thinking about the broadcast side of the network — since the relay observations $Y_1$ and $Y_2$ are stochastically identical, any message that can be decoded reliably by one relay can also be decoded by the other. This simplifies coding methods where part or all of the input message is decoded reliably at the relays. Asymmetric networks, and in particular, networks where the broadcast channel is asymmetric, appear to only obfuscate the fundamental issues. However, we do study several asymmetric networks both for completeness and to gain additional insight.

We will present numerical results for various networks by fixing the parameters of the broadcast side while varying those of the multiaccess side. For the symmetric Gaussian network, we will fix the input SNR, $S_1 = \frac{P_1}{N_{Z_1}} = \frac{P_2}{N_{Z_2}}$, while sweeping the multiaccess SNR, $S_3 = \frac{P_{W_3}}{N_Z} = \frac{P_{W_4}}{N_Z}$. For the symmetric PBSC network, we will fix the BSC crossover probabilities $\alpha = \alpha_1 = \alpha_2$ while sweeping the binary link rates $R_3 = R_4$. We can thereby observe how coding strategies perform in various regimes.
1.2 Background

Related distributed detection problems have been studied in various guises for quite some time, all with limited success. The fundamental problem invariably arises from having multiple observations of a source of information at different points within a network. In most manifestations of the problem, these observations must be compressed individually and sent to a final destination. At that final destination, a decision must be made about the source, such as an estimation of the true source output or a guess whether a particular event occurred. We make this explicit for the parallel relay network of Figure 1-2. There are two distributed observations, \( Y_1 \) and \( Y_2 \). Each relay can observe and therefore incorporate information only from its own observation. At the final destination, the decoder decides upon a likely input, \( X \). For these distributed problems, it has proved extremely difficult to understand how best to compress and communicate the relevant observations to the decision maker. Much of the work on these problems eliminate the noisy communication problem altogether by assuming noiseless links (with limited capacity) between the distributed terminals and the decision maker.

Tenney and Sandell introduced a hypothesis testing formulation [39]. This problem and several variants are referred to as decentralized detection in the decision theory literature. The problem has been studied with both a Bayesian and Neyman-Pearson formulation. Tsitsiklis provides a somewhat recent survey of the relevant literature [41]. All formulations we have seen assume each terminal (sensor) sends one of a finite number of messages noiselessly to the destination. Most of the available results identify structural properties of optimal policies ([40] and, e.g., [27]), rather than quantifying performance.

For the decentralized detection problem, some stochastic structure appears necessary for tractability. In particular, one representative formulation is NP-complete. However, the problem becomes computationally tractable if we assume conditional independence amongst the observations (conditioned on the source input) [42]. We often make this assumption when studying our communication problem in this thesis.

Berger, Zhang, and Viswanathan introduced a rate-distortion formulation [7]. This problem is referred to as the CEO problem in the information theory literature.
1.2. BACKGROUND

The discrete alphabet problem was studied in [7], while Viswanathan and Berger and then Oohama studied a Gaussian version [44], [33]. For this problem, there are again a multitude of terminals (sensors) with noisy observations of the source. All three cited works assume conditional independence amongst the sensor observations (conditioned on the source input). Each terminal sends a bitstream of limited rate noiselessly to the destination. Essentially all of the available results are asymptotic as the number of sensors grows to infinity.

One important difference between the CEO problem and our parallel relay communication problem is that, in the CEO problem, the source is fixed by nature (and memoryless in the literature). For our communication problem, on the other hand, we design the source to fit the network. That is, we design coding schemes that will work well for communicating through the network. From an engineer’s perspective, this is advantageous: we are given a set of network resources and constraints, and we decide how best to use them. From an analytical perspective, this is not necessarily advantageous: for a given coding scheme, we must prove that no other scheme could perform better.

One would hope that the sensor processing approaches appropriate for the CEO problem would yield analogous approaches appropriate for communication. Indeed, our coding theorems based on block quantization are analogous in some ways to results in the CEO problem literature. Unfortunately, for the CEO problem, proving the optimality of the approaches (and even the validity in [33]) requires the number of sensors to grow to infinity. This scenario is of limited interest here. Additionally, there is a difference in objectives between these two problems. For the CEO problem, we wish to minimize distortion in an estimate of the source. For the communication problem, we wish to maximize the reliably achievable communication rate. We will comment on this difference of objectives at the end of this section.

More directly related to our communication problem, Van der Meulen introduced the relay channel [17], [18]. Cover and El Gamal significantly extended those results [13]. For simplicity of presentation, we picture the Gaussian version of the relay channel in Figure 1-6, though the literature has focused almost exclusively on discrete alphabet versions. Because of its similarity to our parallel relay network, we will define the relay channel problem explicitly.
Chapter 1. Introduction

The relay channel consists of an input terminal with input $X$, a relay terminal which observes $Y_1$ and transmits $W_3$, and an ultimate decoder which receives $Y$. The only source of extrinsic information is the input terminal, and the only purpose of the relay is to help get information from the input terminal to the receiver. The relay is an arbitrarily complex processor whose transmission at time $k$, $W_{3,k}$, is a function, possibly a random function, of its available observations, $Y_1^{k-1}$. In other words, $W_{3,k}$ is a function of $Y_1^{k-1}$ but cannot depend directly on $Y_{1,k}$\(^2\). In the Gaussian case, we also assume average power constraints at the input terminal and at the relay. If $x_k$ and $w_{3,k}$ are the input and relay transmissions at time $k$, respectively, then $Y_{1,k} = x_k + Z_{1,k}$ and $Y_k = x_k + w_{3,k} + Z_k$. The two noise processes, $\{Z_{1,k}\}$ and $\{Z_k\}$, are independent. Additionally, both noise processes are i.i.d. Finally, $Z_{1,k}$ and $Z_k$ are zero-mean Gaussian random variables with variance $N_{Z_1}$ and $N_Z$, respectively. The goal in this problem is to determine the capacity of the relay channel, defined as the supremum of reliably achievable communication rates from the input terminal to the receiver. The only essential difference between the general relay channel and the one described here is that in general, there is a given conditional joint distribution, $p(y_1, y \mid x, w_3)$, for the pair of observations given the pair of inputs. Thus in general, both random observations at time $k$, $Y_{1,k}$ and $Y_k$, are stochastically related to both inputs at time $k$, $X_k$ and $W_{3,k}$, via $p(y_1, y \mid x, w_3)$. Once again, $W_{3,k}$ is a (possibly random) function of $Y_1^{k-1}$ but cannot depend directly on $Y_{1,k}$.

The capacity of the relay channel is still unknown in all but degenerate cases ([13], [23], [4], [28], and [43]). All of the later literature address relay channels where

\(^2\)We have to be a bit more careful about the discrete-time model for the relay channel. The important requirement is that the input and relay terminals can synchronize their transmissions to the decoder. We define time origins and hence indices so that $Y_k$ is a stochastic function of $X_k$ and $W_{3,k}$. Additionally, $Y_{1,k}$ is a stochastic function of $X_k$. Due to propagation delay and relay processing, it is not physically reasonable to then assume that $W_{3,k}$ can depend on $Y_{1,k}$.
the channel outputs, $Y_{1,k}$, $Y_k$, or both, are deterministic functions of the channel inputs $X_k$ and $W_{3,k}$. In the earlier literature, Cover and El Gamal found the capacity only of the physically degraded relay channel (and the reversely physically degraded relay channel, which is defined similarly and which we will not address). Related to the concept of physical degradedness, there is a somewhat related concept of stochastic degradedness. We will differentiate these concepts after presenting some of our results, which will provide an appropriate context for the discussion. The difference between these two classes of channels is absolutely central to the network communication problem — both for the relay channel and for our parallel relay network. We believe that the stochastically degraded class is the appropriate one to study.

Apart from the literature cited above, the only other progress on the relay channel is apparently provided by Zhang [53]. The contribution of this work is an unusual converse statement for a specific type of relay channel. Zhang considers a discrete alphabet and a stochastically degraded relay channel with a noiseless binary link from the relay to the receiver (as described for the PBSC network of Figure 1-5). Unfortunately, Zhang’s proof of the converse statement is incomprehensible to us, and thus we cannot gain any insight from that work.

One primary difference between the relay channel and our parallel relay network is that in the relay channel, there is a direct path from the input transmission, $X$, to the receiver observation, $Y$, as well as to the relay observation, $Y_1$. However, the relay requires a one unit delay before it can incorporate its observation $Y_1$ into its transmission $W_3$. This appears to necessitate more complicated coding procedures to communicate efficiently through the network. In the work of Cover and El Gamal on the relay channel, this complication takes the form of superposition coding in the achievability theorems. These more complicated approaches apparently need to be exploited for general mesh networks, but we believe that these more general networks can be studied more easily once we understand communication through the parallel relay network. In particular, we believe that essentially similar fundamental issues are involved in determining the capacity of the two network topologies.

In addition to the relay channel, there is a significant amount of tangentially related work in the information theory literature. We have found several source coding results helpful in proving achievability theorems based on random coding.
arguments. When we generate an ensemble of codewords \( \{X^n\} \) using a single-letter distribution \( p(x) \), it is sometimes useful to think of \( X \) as an i.i.d. source with distribution \( p(x) \). Similarly, it is sometimes useful to think of the relay observations \( (Y_1, Y_2) \) as induced “observation sources” with single-letter, joint distribution \( p(y_1, y_2) = \sum_x p(x) \cdot p(y_1, y_2|x) \). For instance, viewing the pair of relay observations \( (Y_1, Y_2) \) as an induced i.i.d. observation source, we can apply the distributed source coding technique of Slepian and Wolf [38] to send the induced “source” to the destination over noiseless binary channels. We will provide the details later in the thesis. We have found a similar application for the source coding with side information results independently discovered by Wyner [49] and by Ahlswede and Körner [5]. Though we will focus on discrete networks with a pair of noiseless binary links to the destination, we could also consider discrete networks with a general multiaccess channel connecting the relays to the destination. In this case, the analogous source coding problem would be that posed by Cover, El Gamal, and Salehi [14].

There is additional relevant work which does not quite fit into the communication framework we study. Han considered a set of correlated, memoryless sources connected to input terminals, all with the same destination. The input terminals and the destination are connected to each other by an arbitrary mesh of memoryless, noisy point-to-point links, and processors (i.e., terminals) capable of decoding and reencoding messages are located at the end of every link. Han assumes there are no cycles in the topology of the interconnecting network [25]. He found necessary and sufficient conditions on the capacity of interconnecting links for reproducing all sources reliably at the destination. However, he does not consider any multiterminal channels. In particular, there are no broadcast or multiaccess channels. Rather, if a terminal is connected to several other terminals, it has an independent noisy link between each (with a separate input for each link). Also, the work of Ahlswede, Cai, Li, and Yeung [2] contains a single source, multiple destination formulation using a similar model, in their case with noiseless interconnecting links of various rates.

Finally, the work on point-to-point, indirect rate-distortion problems appears tangentially related. The problem was introduced with mean-square distortion by Dobrushin and Tsybakov [19], independently generalized by Sakrison [36] and by Wolf and Ziv [47], and later revisited with a different perspective by Witsenhausen [45].
1.2. BACKGROUND

These problems involve rate distortion of a source that is input into a noisy channel. The goal in this problem is to quantize the source, but we can only observe and operate on the output of the noisy channel. The gist of the results is that in several cases of interest, it is optimal to first estimate the source based on the channel output, and then to compress the source estimator as if it were the true source. When the source and channel are memoryless, the estimator has a very simple form. It can be decomposed symbol by symbol. The resulting source estimator is then itself i.i.d. For our purposes, we consider a source that is a set of codewords rather than an i.i.d. memoryless input. When these codewords cannot be reliably decoded from the output of the channel, the problem becomes complicated. It is unclear whether we can use the work in this area for anything more than intuition.

More generally, we have been unable to generate converse results for our communication problem by adapting approaches that work for source coding or for rate-distortion problems. Though we may opt to randomly generate codewords with i.i.d. symbols, a converse theorem must apply to every codebook. Converse theorems for source coding problems invariably hinge upon the temporal structure of the source, most often memorylessness. There is no such useful structure for a general communication codebook. For example, assume we have a binary input alphabet, \( \mathcal{X} = \{0, 1\} \), and that we are attempting to communicate at rates approaching one bit per input channel use. Intuition suggests that the input, \( X^n \), must appear in some sense like an i.i.d. source equally likely 0 or 1 (for otherwise \( \frac{1}{n} I(X^n; Y^n) \leq \frac{1}{n} H(X^n) < 1 \)). We can try to gain insight by considering what prevents us from reliably transmitting such a source through the network (or conversely, by determining what requirements we need from the network in order to successfully do so). Ultimately, however, we have been unable to capitalize on this intuition in the form of a converse theorem.

In addition to the difference of temporal structure, there is also a difference of objectives between source coding problems and communication problems. Here we briefly explore the connection between these two objectives. Let us assume that at the input, we use an i.i.d. source, \( X \), with single-letter distribution \( p(x) \). Denote the rate-distortion function for this i.i.d. source (with Hamming distortion) by \( R_X(d) \), where \( d \) is the distortion. Denote by \( \hat{X}(Y^n) \) any estimator of \( X^n \) based on the receiver observation \( Y^n \). Assume the average probability of symbol error for this estimator equals \( \hat{d} \).
Then the rate-distortion theorem implies \( \frac{1}{n} I(X^n; Y^n) \geq \frac{1}{n} I(X^n; \hat{X}(Y^n)) \geq R_X(\hat{d}) \) [21, Th. 9.2.1]. We can then derive a random coding argument based on superletters that proves we can communicate reliably at rate \( R_{\text{ach}} = \frac{1}{n} I(X^n; Y^n) \geq R_X(\hat{d}) \). The problem of minimizing distortion of an i.i.d. source through a network thus provides an achievable communication rate. However, there are two important issues. First, the rate-distortion function provides a lower bound to the mutual information between input and output, and this lower bound may not be tight. Therefore, having found the scheme that minimizes distortion, we may actually be achieving a larger mutual information, \( \frac{1}{n} I(X^n; Y^n) \), without being able to quantify it. Second, the scheme that minimizes distortion through the network may not maximize the mutual information between input and output. That is, there may be another scheme that yields a larger distortion as well as a larger mutual information. Thus, despite our intuition, there is not necessarily a close connection between minimizing distortion and maximizing mutual information, even when assuming an i.i.d. input, \( X \). These are some of the difficulties we have encountered when attempting to derive converse theorems for our communication problem based on source coding or rate-distortion results.
Chapter 2

Gaussian Parallel Relay Networks

We study the Gaussian parallel relay network of Figure 2-1. This network is parametrized by four signal to noise ratios: \( S_1 = \frac{p_1}{N_1} \), \( S_2 = \frac{p_2}{N_2} \), \( S_3 = \frac{p_3}{N_3} \), and \( S_4 = \frac{p_4}{N_4} \), which we mark pictorially in Figure 2-2 for easy reference. We proceed by presenting upper bounds to the network capacity, or converse results, followed by achievable coding results. For much of the work, we focus on symmetric networks, where \( S_1 = S_2 \) and \( S_3 = S_4 \). For these symmetric networks, we refer to \( S_1 \) as the broadcast SNR and to \( S_3 \) as the multiaccess SNR. Some of the achievability results are significantly more difficult to describe for asymmetric networks, but they are fundamentally similar in nature.

2.1 Gaussian Broadcast and Multiaccess Channels

We first review the capacity regions of the two-user Gaussian broadcast and multiaccess channels. We then extend the results for the Gaussian multiaccess channel
to the case where one user knows the message of the other. We independently re-
proved this latter result for the Gaussian case, originally proved by Prelov [35] for
both discrete and Gaussian multiaccess channels. Prelov derived mutual information
inequalities which define the capacity region for the discrete multiaccess channel. He
then restated without proof the logarithmic form of the inequalities for the Gaussian
channel. The derivation of the converse is primarily a non-trivial exercise in manipula-
tion of second-order statistics. More importantly, however, the simple derivation of
the achievability argument yields significant insight into our parallel relay problem,
and therefore we will present the proof in full. We will need these three sets of results
to formulate a simple coding technique for the Gaussian parallel relay network.

Cover introduced broadcast channels [12]. Bergmans established the capacity-
achieving argument for memoryless, degraded broadcast channels [8]. Gallager estab-
lished the converse for discrete, memoryless, degraded broadcast channels [22], while
Bergmans established the converse for Gaussian memoryless broadcast channels [9].

The memoryless Gaussian broadcast channel is pictured in Figure 2-3. There is
one sender with input $X$ and two receivers with observations $Y_1$ and $Y_2$. If $x_k$
is the input at time $k$, then $Y_{1,k} = x_k + Z_{1,k}$ and $Y_{2,k} = x_k + Z_{2,k}$. The two noise
processes, $\{Z_{1,k}\}$ and $\{Z_{2,k}\}$, are independent. Additionally, both noise processes are
i.i.d. Finally, $Z_{1,k}$ and $Z_{2,k}$ are zero-mean Gaussian random variables with variance
$N_{Z_1}$ and $N_{Z_2}$, respectively. We denote a Gaussian random variable with mean $m$
and variance $v$ by $\mathcal{N}(m, v)$, and thus with this notation, $Z_{1,k} \sim \mathcal{N}(0, N_{Z_1})$. Finally, the
input terminal has an average power constraint $P_X$. That is, in the limit as $n \to \infty$,

$$
\frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \{ X_k^2 \} \leq P_X.
$$

(2.1)
Assume without loss of generality that \( N_{Z_2} \geq N_{Z_1} \), i.e., Decoder 2 has a noisier observation than Decoder 1.

![Diagram of the Gaussian broadcast channel]

The goal is to reliably transmit independent message streams to each decoder. Denoting by \( R_{\text{top}} \) the rate to Decoder 1 and by \( R_{\text{bot}} \) the rate to Decoder 2, the capacity region turns out to be the set of pairs \( \{ R_{\text{top}}, R_{\text{bot}} \} \) such that, for some \( \alpha \in [0, 1] \),

\[
R_{\text{top}} \leq \frac{1}{2} \log_2 \left( 1 + \frac{\alpha P_X}{N_{Z_1}} \right) = \frac{1}{2} \log_2 (1 + \alpha \cdot S_1), \tag{2.2}
\]

\[
R_{\text{bot}} \leq \frac{1}{2} \log_2 \left( 1 + \frac{(1-\alpha) P_X}{\alpha P_X + N_{Z_2}} \right) = \frac{1}{2} \log_2 \left( 1 + \frac{(1-\alpha) S_2}{\alpha S_2 + 1} \right). \tag{2.3}
\]

Furthermore, Decoder 1 can reliably decode the message stream for Decoder 2. There is a straightforward random coding argument which proves the achievability of these boundary rate pairs for a given \( \alpha \). Split the sender’s power in two, devoting a fraction \( \alpha \) for \( R_{\text{top}} \) and the remaining fraction \( (1-\alpha) \) for \( R_{\text{bot}} \). Choose \( 2^{nR_{\text{top}}} \) length \( n \) codewords for Decoder 1, generating each symbol of each codeword randomly according to \( \mathcal{N}(0, \alpha \cdot P_X) \). Similarly, generate \( 2^{nR_{\text{bot}}} \) length \( n \) codewords for Decoder 2, generating each symbol of each codeword randomly according to \( \mathcal{N}(0, (1-\alpha) \cdot P_X) \). The signal \( X \) transmitted at the input is the algebraic sum of the codewords corresponding to each message.

A simple way to interpret the mathematical achievability argument is as follows. For clarity, we neglect the \( \epsilon \)'s in the following explanation. To Decoder 2, the codewords corresponding to \( R_{\text{top}} \), averaged over the ensemble, appear as additional Gaussian noise of power \( \alpha \cdot P_X \). Considering the effective single-user Gaussian channel,
then, Decoder 2 can decode its message provided $R_{bot}$ satisfies (2.3). As an aside, if the codewords for $R_{bot}$ are generated i.i.d. $X_{R_{bot}} \sim \mathcal{N}(0, (1 - \alpha) \cdot P_X)$, and the average power of codewords for $R_{top}$ is $\alpha \cdot P_X$ for each codeword symbol, then we minimize the mutual information $I(X_{R_{bot}}; Y_2)$ by assuming the codewords for $R_{top}$ look i.i.d., $X_{R_{top}} \sim \mathcal{N}(0, \alpha \cdot P_X)$ [15, pg. 263, Ex. 1]. Therefore, it is conservative to assume that the effective noise seen by Decoder 2 is i.i.d. $\mathcal{N}(0, \alpha \cdot P_X + N_{Z_2})$, and this assumption turns out to be sufficient for the Gaussian broadcast channel problem. Returning to the interpretation, since $N_{Z_1} \leq N_{Z_2}$, Decoder 1 can decode $R_{bot}$ as well. Decoder 1 subtracts off the codeword corresponding to $R_{bot}$ from its observation $Y_1$. The resulting channel from sender to Decoder 1 is a single-user Gaussian channel with input power constraint $\alpha P_X$ and noise power $N_{Z_1}$. Decoder 1 can thus decode its message provided $R_{top}$ satisfies (2.2).

We consider next the Gaussian multiaccess channel with independent messages. The capacity region for the discrete multiaccess channel was found independently by Ahlswede [1] and by Liao [31]. It can be easily extended to the Gaussian case (e.g., see Wyner [48]). An achievability argument based on superposition coding, which avoids the original time-sharing argument, was given by Carleial [11].

The Gaussian multiaccess channel is pictured in Figure 2-4. There are two users, with inputs $W_3$ and $W_4$, and one receiver, with observation $Y$. If $w_{3,k}$ and $w_{4,k}$ are the inputs at time $k$, then $Y_k = w_{3,k} + w_{4,k} + Z_k$, where $Z_k \sim \mathcal{N}(0, N_Z)$. The noise process $\{Z_k\}$ is i.i.d. The two users each have an average power constraint. In the limit as $n \to \infty$,

\[
\frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \{ W_{3,k}^2 \} \leq P_{W_3}, \quad \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \{ W_{4,k}^2 \} \leq P_{W_4}.
\]

![Figure 2-4: The Gaussian multiaccess channel](image-url)
2.1. GAUSSIAN BROADCAST AND MULTIACCESS CHANNELS

The goal is to reliably transmit independent messages from the two users to the decoder. The users have no knowledge of the other user’s message. Denoting by $R_{\text{top}}$ the rate of User 1 and by $R_{\text{bot}}$ the rate of User 2, the capacity region has been shown to be the set of rate pairs $\{R_{\text{top}}, R_{\text{bot}}\}$ such that

$$R_{\text{top}} \leq \frac{1}{2} \log_2 \left( 1 + \frac{P_{W_3}}{N} \right) = \frac{1}{2} \log_2 (1 + S_3),$$

$$R_{\text{bot}} \leq \frac{1}{2} \log_2 \left( 1 + \frac{P_{W_4}}{N} \right) = \frac{1}{2} \log_2 (1 + S_4),$$

$$R_{\text{top}} + R_{\text{bot}} \leq \frac{1}{2} \log_2 \left( 1 + \frac{P_{W_3} + P_{W_4}}{N} \right) = \frac{1}{2} \log_2 (1 + S_3 + S_4). \quad (2.4)$$

The achievability argument for these rate pairs can be interpreted in much the same way for the Gaussian multiaccess channel as for the Gaussian broadcast channel. For clarity, we neglect the ε’s in the following explanation. Consider first a corner point of the capacity region:

$$R_{\text{top}} = \frac{1}{2} \log_2 (1 + S_3), \quad (2.5)$$

$$R_{\text{bot}} = \frac{1}{2} \log_2 \left( 1 + \frac{S_4}{S_3 + 1} \right). \quad (2.6)$$

Choose $2^{nR_{\text{top}}}$ length $n$ codewords for User 1, generating each symbol of each codeword randomly according to $\mathcal{N}(0, P_{W_3})$. Similarly, choose $2^{nR_{\text{bot}}}$ length $n$ codewords for User 2, generating each symbol of each codeword randomly according to $\mathcal{N}(0, P_{W_4})$. At the decoder, first decode $R_{\text{bot}}$, treating the codewords from User 1 as additional Gaussian noise of power $P_{W_3}$. Then subtract the codeword from User 2 off of the received vector and decode $R_{\text{top}}$. To achieve any point on the boundary $(R_{\text{top}} + R_{\text{bot}}) = \frac{1}{2} \log_2 (1 + S_3 + S_4)$, we can time-share between codes achieving the two corner points.

We can also achieve any rate pair on this boundary without recourse to time-sharing, as demonstrated by Carleial [11], by using superposition coding. This is a simple extension of the same philosophy. To achieve a rate pair somewhere on this boundary, a user, say User 1, is split into two sub-users with independent messages, say User 1a and User 1b. User 1 devotes a fraction of its power to User 1a and the remaining fraction to User 1b. The codewords from the two sub-users are algebraically added and sent as the signal $W_3$. Now there are three independently generated Gaus-
sian codebooks. The receiver decodes and subtracts off the codeword from User 1a, then from User 2, and finally from User 1b. With the power split at User 1 chosen appropriately, we can achieve any point on the boundary. Both User 1a and User 1b know each other’s message, but we do not need to exploit this. Additionally, note that unlike time-sharing, superposition coding does not require block synchronization between the users.

We consider next the Gaussian multiaccess channel where one user, say User 1, knows the message of the other. We will say Message 1 is the message of User 1, and Message 2 is the message of User 2. In other words, User 1 knows Message 1 and Message 2, while User 2 knows only Message 2. For the converse, the technique we use for bounding mutual information via signal correlation was also used for the physically degraded Gaussian relay channel [13] and for the Gaussian multiaccess channel with feedback [34].

Theorem 2.1.1. [35, (Prelov)] Let $R_{\text{top}}$ be the rate of Message 1 in bits per channel symbol, and let $R_{\text{bot}}$ be the rate of Message 2. Message 1 and Message 2 are independent. The capacity region of the two-user Gaussian multiaccess channel, where User 1 knows both Message 1 and Message 2, while User 2 knows only Message 2, is $B\{S_3, S_4\}$, where

$$B\{S_3, S_4\} = \left\{(R_{\text{top}}, R_{\text{bot}}) \mid \exists \beta \in [0, 1]: \begin{align*} R_{\text{top}} &\leq \frac{1}{2} \log_2 \left(1 + (1 - \beta)S_3\right) \\ R_{\text{top}} + R_{\text{bot}} &\leq \frac{1}{2} \log_2 \left(1 + S_3 + S_4 + 2\sqrt{\beta S_3 S_4}\right) \end{align*} \right\}. \quad (2.7)$$

Remark: For future reference, we refer to this as the extended multiaccess capacity region for the Gaussian multiaccess channel.

Proof.

Achievability: The underlying observation here is that User 1 knows both Message 1 and Message 2, while User 2 knows Message 2 only. This allows the users to cooperate in a very clean way — the information distributed in this simple network is either complete or completely absent.
2.1. GAUSSIAN BROADCAST AND MULTIACCESS CHANNELS

Since User 1 knows Message 2, it can devote a fraction of its power to helping send that message. Assume User 1 devotes a fraction $\alpha$ of its power to do so, where $\alpha \in [0, 1]$. The remaining fraction $(1 - \alpha)$ is used to send Message 1. The signal $W_3$ sent by User 1 is the algebraic sum of independent codewords, one for each message. Since User 2 has power $P_{W_2}$ and User 1 devotes power $\alpha P_{W_3}$, together they have power $\left(\sqrt{\alpha P_{W_3}} + \sqrt{P_{W_2}}\right)^2$ when they send the same codeword for Message 2. It follows immediately from the results for the Gaussian multiaccess channel with independent messages, (2.4), that the following region $\mathcal{A}\{S_3, S_4\}$ is achievable:

$$\mathcal{A}\{S_3, S_4\} =$$

$$\left\{ (R_{\text{top}}, R_{\text{bot}}) \mid \exists \alpha \in [0, 1] : \begin{align*}
R_{\text{top}} &\leq \frac{1}{2} \log_2 (1 + (1 - \alpha)S_3) \\
R_{\text{bot}} &\leq \frac{1}{2} \log_2 (1 + \alpha S_3 + S_4 + 2\sqrt{\alpha S_3 S_4}) \\
R_{\text{top}} + R_{\text{bot}} &\leq \frac{1}{2} \log_2 (1 + S_3 + S_4 + 2\sqrt{S_3 S_4})
\end{align*} \right\}$$

Next we show that the second inequality in (2.8) is superfluous. We repeat here the region $\mathcal{B}\{S_3, S_4\}$ of the theorem statement:

$$\mathcal{B}\{S_3, S_4\} =$$

$$\left\{ (R_{\text{top}}, R_{\text{bot}}) \mid \exists \beta \in [0, 1] : \begin{align*}
R_{\text{top}} &\leq \frac{1}{2} \log_2 (1 + (1 - \beta)S_3) \\
R_{\text{top}} + R_{\text{bot}} &\leq \frac{1}{2} \log_2 (1 + S_3 + S_4 + 2\sqrt{\beta S_3 S_4})
\end{align*} \right\}$$

(2.9)

Clearly $\mathcal{A}\{S_3, S_4\} \subseteq \mathcal{B}\{S_3, S_4\}$. To show the reverse inclusion, assume $(R_{\text{top}}^*, R_{\text{bot}}^*) \in \mathcal{B}\{S_3, S_4\}$, and let $\beta \in [0, 1]$ satisfy the two inequalities in (2.9). We will simply manipulate the defining inequalities of $\mathcal{B}\{S_3, S_4\}$ to upper bound $R_{\text{bot}}^*$ as a function of $R_{\text{top}}^*$.

Let $\alpha^* \in [\beta', 1]$ uniquely satisfy

$$R_{\text{top}}^* = \frac{1}{2} \log_2 (1 + (1 - \alpha^*)S_3).$$

(2.10)
Since $\alpha^* \geq \beta'$, we know

$$R_{\text{top}}^* + R_{\text{bot}}^* \leq \frac{1}{2} \log_2 \left( 1 + S_3 + S_4 + 2\sqrt{\alpha^* S_3 S_4} \right). \quad (2.11)$$

Therefore,

$$R_{\text{bot}}^* \leq \frac{1}{2} \log_2 \left( 1 + \frac{\alpha^* S_3 + S_4 + 2\sqrt{\alpha^* S_3 S_4}}{1 + (1 - \alpha^*) S_3} \right) \quad (2.12)$$

$$\leq \frac{1}{2} \log_2 \left( 1 + \alpha^* S_3 + S_4 + 2\sqrt{\alpha^* S_3 S_4} \right). \quad (2.13)$$

We have thus identified $\alpha = \alpha^*$ and found that $(R_{\text{top}}^*, R_{\text{bot}}^*) \in \mathcal{A}\{S_3, S_4\}$.

Converse: In the appendix we present the complete converse proof. Here we present only the essential idea.

In the appendix, we apply Fano’s inequality twice to relate symbolwise error probability to mutual information over blocks of letters. In particular, we show that the two communication rates, $R_{\text{top}}$ and $R_{\text{bot}}$, are essentially upper bounded by the mutual information terms:

$$R_{\text{top}} \leq \frac{1}{n} I(W_3^n; Y^n \mid W_4^n), \quad (2.14)$$

$$(R_{\text{top}} + R_{\text{bot}}) \leq \frac{1}{n} I(W_3^n, W_4^n; Y^n). \quad (2.15)$$

Next, in the appendix, we apply standard inequalities for memoryless channels to upper bound the block mutual information terms in (2.14) and (2.15) by a sum of symbolwise mutual terms:

$$I(W_3^n; Y^n \mid W_4^n) \leq \sum_{i=1}^{n} I(W_{3,i}; Y_i \mid W_{4,i}), \quad (2.16)$$

$$I(W_3^n, W_4^n; Y^n) \leq \sum_{i=1}^{n} I(W_{3,i}, W_{4,i}; Y_i). \quad (2.17)$$

Now for the essential idea behind the converse. Notice that the mutual information terms in (2.16) are maximized when $W_3$ and $W_4$ are independent. In contrast, the mutual information terms in (2.17) are maximized when $W_3$ and $W_4$ are perfectly
2.1. **GAUSSIAN BROADCAST AND MULTIACCESS CHANNELS**

...correlated. There is a trade-off in terms of the correlation between the signals \( W_3 \) and \( W_4 \). We can parametrize the correlation between \( W_3 \) and \( W_4 \) and then maximize (2.16) and (2.17) for a given correlation. Here we demonstrate the trade-off by considering only a single time index. In the appendix, using convexity appropriately, we extend the same approach to the sum of \( n \) terms in (2.16) and (2.17). We proceed by simultaneously considering the first term in each of the summations (for simplicity, we drop the time index subscript from the notation in what follows, thus \( W_3 \) and \( W_4 \) refer to \( W_{3,1} \) and \( W_{4,2} \), respectively).

Let \( W_3 \) and \( W_4 \) be random variables with energy defined as

\[
P_3 = \mathbb{E}\{ W_3^2 \} ,
\]

\[
P_4 = \mathbb{E}\{ W_4^2 \} .
\]  

(2.18)  

(2.19)

The correlation between \( W_3 \) and \( W_4 \) is defined as

\[
\rho = \frac{\mathbb{E}\{ W_3 \cdot W_4 \} }{\sqrt{P_3 \cdot P_4}}.
\]  

(2.20)

Rearranging (2.20),

\[
\rho \sqrt{P_3 \cdot P_4} = \mathbb{E}\{ W_3 \cdot W_4 \}
\]

(2.21)

\[
= \mathbb{E}_{W_4} \{ W_4 \cdot \mathbb{E}_{W_3}\{ W_3 \mid W_4 \} \} .
\]  

(2.22)

Note that \( \mathbb{E}_{W_3}\{ W_3 \mid W_4 \} \) is a random variable that is a function of \( W_4 \). To make this explicit, define the function

\[
f(W_4) = \mathbb{E}_{W_3}\{ W_3 \mid W_4 \} .
\]  

(2.23)

Thus from (2.22),

\[
|\rho| \sqrt{P_3 \cdot P_4} = |\mathbb{E}_{W_4}\{ W_4 \cdot f(W_4) \}|
\]

(2.24)

\[
\leq \sqrt{\mathbb{E}_{W_4}\{ W_4^2 \} \cdot \mathbb{E}_{W_4}\{ f^2 (W_4) \}}
\]

(2.25)

\[
= \sqrt{P_4} \cdot \sqrt{\mathbb{E}_{W_4}\{ f^2 (W_4) \}},
\]  

(2.26)
where (2.25) follows from the Cauchy-Schwarz inequality, and (2.26) follows from (2.19). Rearranging (2.26) a bit,
\[ p^2 \cdot P_3 \leq \mathbb{E}_{W_4} \{ f^2 (W_4) \}. \] (2.27)

Also,
\[
P_3 = \mathbb{E} \{ W_3^2 \} = \mathbb{E}_{W_4} \{ \mathbb{E}_{W_3} \{ W_3^2 \mid W_4 \} \}. \] (2.28)

For each sample value of $W_4$, denoted by $w_4$,
\[
\mathbb{E}_{W_3} \{ W_3^2 \mid w_4 \} = (\mathbb{E}_{W_3} \{ W_3 \mid w_4 \})^2 + \text{VAR} (W_3 \mid w_4) = f^2 (w_4) + \text{VAR} (W_3 \mid w_4),
\] (2.30)

where we have used the notation $\text{VAR}(W_3 \mid w_4)$ to mean the variance of $W_3$ conditioned on the particular sample value of $W_4 = w_4$. Therefore, taking expectations in (2.31), we get
\[
P_3 = \mathbb{E}_{W_4} \{ f^2 (W_4) \} + \mathbb{E}_{W_4} \{ \text{VAR} (W_3 \mid w_4) \}. \] (2.32)

Combining (2.27) and (2.32),
\[
\mathbb{E}_{W_4} \{ \text{VAR} (W_3 \mid w_4) \} \leq (1 - p^2) \cdot P_3. \] (2.33)

Therefore, since the Gaussian variable maximizes entropy subject to a second moment constraint,
\[
H (Y \mid W_4) \leq \frac{1}{2} \log_2 \left( 2 \pi e \cdot ((1 - p^2) \cdot P_3 + N_z) \right), \] (2.34)
\[
H (Y) \leq \frac{1}{2} \log_2 \left( 2 \pi e \cdot (P_3 + P_4 + 2p \cdot \sqrt{P_3 \cdot P_4} + N_z) \right). \] (2.35)
2.2. **Converses for Gaussian Parallel Relay Networks**

On the other hand,

\[ H(Y \mid W_3, W_4) = \frac{1}{2} \log_2 \left( 2\pi e \cdot N_Z \right). \]

(2.36)

Therefore, combining (2.34)-(2.36),

\[ I(W_3; Y \mid W_4) \leq \frac{1}{2} \log_2 \left( 1 + \frac{(1 - \rho^2) \cdot P_3}{N_Z} \right), \]

(2.37)

\[ I(W_3, W_4; Y) \leq \frac{1}{2} \log_2 \left( 1 + \frac{P_3 + P_4 + 2\rho \sqrt{P_3 \cdot P_4}}{N_Z} \right). \]

(2.38)

The \( \beta \) in the theorem statement is analogous to \( \rho^2 \) in (2.37) and (2.38).

\[ \square \]

2.2 Converse for Gaussian Parallel Relay Networks

Some of the achievable coding methods are seen from a better perspective once we have established the converse results, and therefore we begin with the converse results. It is not surprising that, due to the processing at the two relays, tight upper bounds to network capacity are difficult to determine. In this section we present three upper bounds based on Fano’s inequality, the data processing inequality, and the same technique which yields the converse to the extended multiaccess capacity region just proved.

We derive all of the upper bounds by starting from the same point — Fano’s inequality. We then use various methods for bounding the mutual information between the source transmission, \( X \), and the receiver observation, \( Y \).

Assume we are attempting to transmit a source \( U \). Let \( U \) be defined on a finite alphabet of size \( M \). Assume that the source terminal encodes \( L \) source symbols, denoted \( U^L \), into \( n \) input channel symbols, denoted \( W_3^n \). At the decoder, let \( V(Y^n) \) be an estimator for \( U^L \) based only on the received observations, \( Y^n \). Denote the entropy of the source symbols, in bits per input channel use, by \( \frac{1}{n} H(U^L) \). Finally, denote by \( P_e \) the average probability of symbol error for the estimator \( V(Y^n) \). Then
Fano’s inequality and the data processing theorem imply [21, Ths. 4.3.2, 4.3.3]

\[
\frac{1}{n} H(U^L) \leq \frac{1}{n} I(X^n; Y^n) + \frac{n}{L} \left( P_e \cdot \log_2(M - 1) + H_{\text{bin}}(P_e) \right).
\] (2.39)

By reliable transmission of the source \( U \), we mean that for any \( \epsilon > 0 \), we can find a coding method with \( P_e \leq \epsilon \). Together with Fano’s inequality (2.39), this implies that there must be a series of codes with blocklength \( n \to \infty \) such that

\[
\lim_{n \to \infty} P_e = 0, \quad \lim_{n \to \infty} \frac{1}{n} H(U^{L(n)}) \leq \frac{1}{n} I(X^n; Y^n).
\] (2.41)

For a memoryless source \( U \),

\[
\frac{1}{n} H(U^{L(n)}) = \frac{L(n)}{n} H(U_1).
\] (2.42)

Substituting (2.42) into (2.41), we conclude that if a memoryless source \( U \) can be reliably transmitted through the network, then there must be a series of codes with blocklength \( n \to \infty \) such that

\[
\lim_{n \to \infty} P_e = 0, \quad \lim_{n \to \infty} \frac{L(n)}{n} H(U_1) \leq \frac{1}{n} I(X^n; Y^n).
\] (2.44)

As \( n \to \infty \), the rate of mapping source symbols to codewords, \( \frac{L(n)}{n} \), is bounded above by a constant. Based on inequality (2.44), the capacity of the parallel relay network, denoted by \( C_{\text{net}} \), is defined as the supremum of achievable \( \frac{1}{n} I(X^n; Y^n) \) (such that none of the average power constraints are violated). To see why we use this mathematical definition for \( C_{\text{net}} \), assume that there is some input and relay processing that yields a given average mutual information between input and output, \( \frac{1}{n} I(X^n; Y^n) \), without violating any of the average power constraints. We will later argue that this implies we can reliably achieve communication rates arbitrarily close to \( \frac{1}{n} I(X^n; Y^n) \) (see Theorem 3.4.1 and the discussion following the proof). That conclusion gives the mathematical definition of \( C_{\text{net}} \) operational significance. Equation (2.44) states that
2.2. *CONVERSES FOR GAUSSIAN PARALLEL RELAY NETWORKS*

if we can reliably transmit the source $U$ through the network, then the entropy rate of the source, measured in bits per input channel use, is less than or arbitrarily close to the average mutual information between the input transmission $X$ and the received observation $Y$. The discussion up to this point applies to both the Gaussian and the discrete parallel relay networks. We proceed by deriving various bounds on any achievable $\frac{1}{n}I(X^n; Y^n)$ in the Gaussian parallel relay network.

\[
\begin{align*}
\frac{1}{n}I(X^n; Y^n) &\leq \frac{1}{n}I(X^n; Y_1^n, Y_2^n) \\
&= \frac{1}{n} \left( H(Y_1^n, Y_2^n) - H(Y_1^n, Y_2^n | X^n) \right) \\
&= \frac{1}{n} \left( \sum_{i=1}^{n} H(Y_{1,i}, Y_{2,i} | Y_{1,i-1}, Y_{2,i-1}) \right) - \frac{1}{2} \log_2 \left( (2\pi e)^2 N_{Z_1}N_{Z_2} \right) \\
&\leq \frac{1}{n} \left( \sum_{i=1}^{n} H(Y_{1,i}, Y_{2,i}) \right) - \frac{1}{2} \log_2 \left( (2\pi e)^2 N_{Z_1}N_{Z_2} \right) \\
&\leq \frac{1}{2} \log_2 \left( 1 + S_1 + S_2 \right).
\end{align*}
\]

Here (2.45) follows from the data processing inequality applied to the broadcast side of the network, (2.48) follows since conditioning reduces entropy, and (2.49) follows since the jointly Gaussian random vector maximizes entropy subject to a second moment constraint. Inequality (2.49) holds for all input codes which satisfy the input power
CHAPTER 2. GAUSSIAN PARALLEL RELAY NETWORKS

constraint, regardless of the processing at the relays. Therefore

\[
C_{\text{net}} \leq \frac{1}{2} \log_2 (1 + S_1 + S_2).
\]

(2.50)

This upper bound is reminiscent of a cut-set inequality, applied to the broadcast side of the network, for a traditional network flow problem — as pictured in Figure 2-5. We refer to (2.50) as the broadcast cut-set bound for the Gaussian parallel relay network.

The second bound is a result of the data processing inequality applied to the multiaccess side of the network.

\[
\frac{1}{n} I(X^n; Y^n) \leq \frac{1}{n} I(W_3^n, W_4^n; Y^n) \leq \frac{1}{n} (H(Y^n) - H(Y^n | W_3^n, W_4^n))
\]

(2.52)

\[
\leq \frac{1}{n} \left( \sum_{i=1}^{n} H(Y_i | Y^{i-1}) \right) - \frac{1}{2} \log_2 (2\pi eN_Z)
\]

(2.53)

\[
\leq \frac{1}{n} \left( \sum_{i=1}^{n} H(Y_i) \right) - \frac{1}{2} \log_2 (2\pi eN_Z)
\]

(2.54)

\[
\leq \frac{1}{n} \left( \sum_{i=1}^{n} \frac{1}{2} \log_2 (2\pi e \cdot \mathbb{E} \{ Y_i^2 \}) \right) - \frac{1}{2} \log_2 (2\pi eN_Z)
\]

(2.55)

\[
\leq \frac{1}{2} \log_2 \left( 2\pi e \cdot \mathbb{E} \{ Y_i^2 \} \right) - \frac{1}{2} \log_2 (2\pi eN_Z)
\]

(2.56)

\[
\leq \frac{1}{2} \log_2 \left( 2\pi e \cdot \left( \sqrt{P_{W_3}} + \sqrt{P_{W_4}} \right)^2 + N_Z \right) - \frac{1}{2} \log_2 (2\pi eN_Z)
\]

(2.57)

\[
= \frac{1}{2} \log_2 \left( 1 + \left( \sqrt{S_3} + \sqrt{S_4} \right)^2 \right).
\]

(2.58)

Once again, (2.51) follows from the data processing inequality, this time applied to the multiaccess side of the network, (2.54) follows since conditioning reduces entropy, (2.55) follows since the Gaussian random variable maximizes entropy subject to a second moment constraint, (2.56) follows from the concavity of the logarithm,
and (2.57) follows from the relay power constraints on $W_3$ and $W_4$. Inequality (2.58) holds for all input codes and arbitrary relay processing provided all three power constraints are satisfied. Therefore

$$C_{\text{net}} \leq \frac{1}{2} \log_2 \left( 1 + \left( \sqrt{S_3} + \sqrt{S_4} \right)^2 \right).$$

This upper bound on capacity, (2.59), is a combination of the data processing inequality applied in the chain $X \rightarrow (W_3, W_4) \rightarrow Y$ and an upper bound on the mutual information $I(W_3^n, W_4^n; Y^n)$. The upper bound on $I(W_3^n, W_4^n; Y^n)$ is achievable when there is perfect cooperation between the relays, that is, when the relay signals $W_3$ and $W_4$ are perfectly correlated. The upper bound (2.59) is reminiscent of a cut-set inequality, applied to the multiaccess side of the network, in a traditional network flow problem — as pictured in Figure 2-6. We refer to (2.59) as the multiaccess cut-set bound for the Gaussian parallel relay network.

![Multiaccess cut-set bound](image)

Figure 2-6: Multiaccess cut-set bound

We next derive a third bound reminiscent of a diagonal cut-set. For the derivation, we consider the equivalent network pictured in Figure 2-7. We have defined an auxiliary variable $\hat{Y} = W_3 + Z$ such that $Y = \hat{Y} + W_4$. 
\begin{equation}
I(X^n; Y^n) \leq I(X^n; Y_2^n, Y^n) \tag{2.60}
= I(X^n; Y_2^n) + I(X^n; \tilde{Y}^n \mid Y_2^n) \tag{2.61}
= I(X^n; Y_2^n) + H(\tilde{Y}^n \mid Y_2^n) - H(\tilde{Y}^n \mid Y_2^n, X^n) \tag{2.62}
\leq I(X^n; Y_2^n) + H(\tilde{Y}^n) - H(\tilde{Y}^n \mid Y_2^n, X^n) \tag{2.63}
= I(X^n; Y_2^n) + H(\tilde{Y}^n) - H(\tilde{Y}^n \mid X^n) \tag{2.64}
= I(X^n; Y_2^n) + I(X^n; \tilde{Y}^n) \tag{2.65}
\leq I(X^n; Y_2^n) + I(W_3^n; \tilde{Y}^n) \tag{2.66}
\leq n \cdot \frac{1}{2} \log_2 (1 + S_2) + n \cdot \frac{1}{2} \log_2 (1 + S_3). \tag{2.67}
\end{equation}

Inequality (2.60) follows from the data processing inequality. Inequality (2.63) follows since conditioning reduces entropy. Equality (2.64) follows since $Y_2^n$ and $\tilde{Y}^n$ are conditionally independent given $X^n$. Inequality (2.66) follows from the data processing inequality since $X^n$ and $\tilde{Y}^n$ are conditionally independent given $W_3^n$. Inequality (2.67) follows by substituting the known capacity of an additive white Gaussian noise channel as a function of the SNR. Then since inequality (2.67) holds for all input codes and arbitrary relay processing provided all three power constraints are satisfied,

\begin{equation}
C_{\text{net}} \leq \frac{1}{2} \log_2 (1 + S_2) + \frac{1}{2} \log_2 (1 + S_3). \tag{2.68}
\end{equation}
We will refer to (2.68) as the pure cross-cut bound for the Gaussian parallel relay network, reflecting Figure 2-7.

We can tighten this result by simultaneously considering the multiaccess cut-set and the diagonal cut-set. We will derive this tightened bound in a way similar to the derivation of the converse for the extended Gaussian multiaccess capacity; for the tightened bound, we will exploit the fact that the diagonal bound is maximized when $W_3$ and $W_4$ are independent, while the multiaccess bound is maximized when $W_3$ and $W_4$ are scaled versions of each other (and are therefore highly dependent). We will need a fair amount of manipulation to move from blockwise to symbolwise mutual information and then to enforce the power constraints. For this derivation, we will explicitly use the same manipulations that we will use in the appendix to derive the converse to the extended Gaussian multiaccess capacity region. We begin by rederiving the diagonal cut-set result in a slightly different way, more reflective of Figure 2-8 than Figure 2-7.

\[
I(X^n; Y^n) \leq I(X^n; Y^n, W_4^n) \\
= I(X^n; W_4^n) + I(X^n; Y^n \mid W_4^n) \\
\leq I(X^n; Y_2^n) + I(X^n; Y^n \mid W_4^n) \\
\leq n \cdot \frac{1}{2} \log_2 (1 + S_2) + I(X^n; Y^n \mid W_4^n) \\
= n \cdot \frac{1}{2} \log_2 (1 + S_2) + \sum_{i=1}^{n} I(X^n; Y_i \mid W_4^n, Y^{i-1})
\]
Inequality (2.71) follows from the data processing inequality since $X^n$ and $W_4^n$ are conditionally independent given $Y_2^n$. Inequality (2.72) follows from the average power constraint at the input and the capacity of the additive white Gaussian noise channel. We proceed by upper bounding the second term of (2.73).

\[
\sum_{i=1}^{n} I(X^n; Y_i | W_4^n, Y^{i-1}) = \sum_{i=1}^{n} H(Y_i | W_4^n, Y^{i-1}) - H(Y_i | W_4^n, Y^{i-1}, X^n) \tag{2.74}
\]

\[
\leq \sum_{i=1}^{n} H(Y_i | W_4^n, Y^{i-1}) - H(Y_i | W_4^n, Y^{i-1}, X^n, W_{3,i}) \tag{2.75}
\]

\[
= \sum_{i=1}^{n} H(Y_i | W_4^n, Y^{i-1}) - H(Y_i | W_{4,i}, W_{3,i}) \tag{2.76}
\]

\[
\leq \sum_{i=1}^{n} H(Y_i | W_{4,i}) - H(Y_i | W_{4,i}, W_{3,i}) \tag{2.77}
\]

\[
= \sum_{i=1}^{n} I(W_{3,i}; Y_i | W_{4,i}) \tag{2.78}
\]

Inequality (2.75) follows since conditioning reduces entropy, and we have added conditioning in the last term. Equality (2.76) follows by the memorylessness of the multiaccess channel. Inequality (2.77) follows since conditioning reduces entropy, and we have dropped conditioning from the middle term. Combining (2.73) with (2.78), we have determined

\[
I(X^n; Y^n) \leq n \cdot \frac{1}{2} \log_2 (1 + S_2) + \sum_{i=1}^{n} I(W_{3,i}; Y_i | W_{4,i}) \tag{2.79}
\]

If we stopped here, we would end up with the same bound that we derived in (2.68). Specifically, from (2.79) and the power constraint at Relay 1,

\[
\frac{1}{n} I(X^n; Y^n) \leq \frac{1}{2} \log_2 (1 + S_2) + \frac{1}{2} \log_2 (1 + S_3). \tag{2.80}
\]

Now we tighten this result by simultaneously considering a second bound. We have
already derived the second bound in (2.54), which we repeat here.

\[
\frac{1}{n} I(X^n; Y^n) \leq \frac{1}{n} \left( \sum_{i=1}^{n} H(Y_i) \right) - \frac{1}{2} \log_2 (2\pi e N) \\
= \frac{1}{n} \sum_{i=1}^{n} I(W_{3,i}, W_{4,i}; Y_i). 
\]  

(2.81)

(2.82)

Now (2.79) and (2.82) are essentially the two inequalities (A.22) and (A.27) we will use in the appendix to derive the converse for the extended Gaussian multiaccess capacity region. The only difference here is that (2.79) contains an added constant \(\frac{1}{2} \log_2 (1 + S_2)\). We can apply the exact same series of mathematical manipulations on correlation functions that we will apply in the appendix. Applying that series of manipulations here, we arrive at the fourth upper bound to network capacity,

\[
C_{\text{net}} \leq \max_{\gamma \in [0,1]} \min \left[ \frac{1}{2} \log_2 ((1 + S_2) (1 + (1 - \gamma)S_3)); \frac{1}{2} \log_2 \left( 1 + S_3 + S_4 + 2\sqrt{\gamma S_3 S_4} \right) \right]. 
\]

(2.83)

We can switch the roles of the relays in the derivation of this bound to get an analogous fifth bound,

\[
C_{\text{net}} \leq \max_{\delta \in [0,1]} \min \left[ \frac{1}{2} \log_2 ((1 + S_1) (1 + (1 - \delta)S_4)); \frac{1}{2} \log_2 \left( 1 + S_3 + S_4 + 2\sqrt{\delta S_3 S_4} \right) \right]. 
\]

(2.84)

Consider the maximization over \(\gamma \in [0, 1]\) that defines the first correlation bound, (2.83). The first term in brackets is decreasing in \(\gamma\), while the second is increasing in \(\gamma\). This makes the maximization a simple function of the SNR’s. The same structure exists in bound (2.84). We refer to these bounds as the correlation bounds for the Gaussian parallel relay network.

When \(S_1 \geq S_2\) and \(S_4 \geq S_3\), then the first correlation bound (2.83) is as small or smaller than the second (2.84). This makes intuitive sense because in this regime, the diagonal cut displayed in Figure 2-8, which is used in deriving (2.83), cuts across a weaker pair of links than the opposing diagonal cut. Similarly, when \(S_1 \leq S_2\) and
$S_4 \leq S_3$, then the second correlation bound (2.84) is as small or smaller than the first (2.83). When the set of four SNR’s is not in one of these two regimes, then it is not as easy to express when one correlation bound dominates the other.

We can interpret these bounds intuitively based on the derivation itself. Essentially, in (2.83), we have upper bounded $I(X; Y)$ by both $I(X; Y_2) + I(W_3; Y \mid W_4)$ and $I(W_3, W_4; Y)$. Assume both relays transmit using all of their allowed power. Think in terms of parametrizing the correlation between signals $W_3$ and $W_4$, and assume the correlation is positive. The correlation coefficient is analogous to $\sqrt{\gamma}$ in (2.83). When the two signals are highly correlated ($\gamma \to 1$), then $I(W_3, W_4; Y)$ can be near its maximum, but $I(W_3; Y \mid W_4)$ cannot. Conversely, when the two signals are highly uncorrelated ($\gamma \to 0$), then $I(W_3, W_4; Y)$ cannot be near its maximum, while $I(W_3; Y \mid W_4)$ can. In the end, for any given coding scheme, there must be some second moment relationship between the relay signals’ transmissions. To get the bound, we must allow the signals to be correlated in the most advantageous way. This corresponds to the maximization over $\gamma \in [0, 1]$. Unfortunately, since we do not know whether these bounds are tight, we do not know whether they have a more operational significance.

Note that each of these correlation bounds, (2.83) and (2.84), are less than or equal to the pure multiaccess cut-set bound, (2.59), which states

$$C_{\text{net}} \leq \frac{1}{2} \log_2 \left( 1 + \left( \sqrt{S_3} + \sqrt{S_4} \right)^2 \right)$$

Specifically, (2.86) is greater than or equal to the second term of (2.83), with equality if and only if $S_2 \geq (\sqrt{S_3} + \sqrt{S_4})^2$. This condition, $S_2 \geq (\sqrt{S_3} + \sqrt{S_4})^2$, implies that the maximization in (2.83) is achieved by $\gamma^* = 1$, and that the second term is no larger than the first. Otherwise, $\gamma^* < 1$. Similarly, (2.86) is greater than or equal to the second term of (2.84), with equality if and only if $S_1 \geq (\sqrt{S_3} + \sqrt{S_4})^2$. We presented the second bound, the multiaccess cut-set bound of (2.59), only because it is a natural bound to consider. Collecting our converse results, then, our tightest upper bound equals the minimum of three terms: the broadcast data processing bound (2.50),
which states
\[ C_{\text{net}} \leq \frac{1}{2} \log_2 (1 + S_1 + S_2), \quad (2.87) \]

and the two correlation bounds, (2.83) and (2.84).

For completeness, we determine when the first correlation bound, (2.83), equals either the pure multiaccess cut-set bound of (2.59) or the pure cross-cut bound of (2.68). Let \( \gamma^* \in [0, 1] \) uniquely achieve the maximum in (2.83). Then \( \gamma^* = 0 \) if and only if
\[ S_2 \leq \frac{S_4}{S_3 + 1}. \quad (2.88) \]

When \( \gamma^* = 0 \), the first correlation bound states
\[ C_{\text{net}} \leq \frac{1}{2} \log_2 ((1 + S_2)(1 + S_3)). \quad (2.89) \]

This corresponds to the pure cross-cut bound, (2.68). Also, as already mentioned, \( \gamma^* = 1 \) if and only if
\[ S_2 \geq S_3 + S_4 + 2\sqrt{S_3 S_4}. \quad (2.90) \]

When \( \gamma^* = 1 \), the first correlation bound states
\[ C_{\text{net}} \leq \frac{1}{2} \log_2 \left( 1 + S_3 + S_4 + 2\sqrt{S_3 S_4} \right). \quad (2.91) \]

This corresponds to a pure multiaccess cut-set bound. Otherwise, \( \gamma^* \in (0, 1) \) uniquely satisfies
\[ (1 + S_2)(1 + (1 - \gamma^*)S_3) = 1 + S_3 + S_4 + 2\sqrt{\gamma^* S_3 S_4}. \quad (2.92) \]
This is a quadratic function of $\gamma^*$ with the uniquely valid solution

$$
\gamma^* = \left( \frac{\sqrt{S_2} \cdot \sqrt{(1 + S_2)(1 + S_3) - S_4 - \sqrt{S_4}}}{\sqrt{S_3} \cdot (1 + S_2)} \right)^2.
$$

(2.93)

When $\gamma^* \in (0, 1)$, the first correlation bound states

$$
C_{\text{net}} \leq \frac{1}{2} \log_2 \left( (1 + S_2)(1 + (1 - \gamma^*)S_3) \right).
$$

(2.94)

In this case, the first correlation bound is strictly smaller than both the pure cross-cut bound, (2.89), and the pure multiaccess cut-set bound, (2.91). Analogous statements hold when addressing the maximization over $\delta \in [0, 1]$ defining the second correlation bound (2.84).

When we specialize to the symmetric Gaussian parallel relay network, where $S_1 = S_2$ and $S_3 = S_4$, the two correlation bounds, (2.83) and (2.84), are identical. In this symmetric case, $\gamma^* = 0$, implying that the correlation bounds equal the pure cross-cut bound (2.68), if and only if $S_1 \leq \frac{S_4}{S_3 + 1}$. Incidentally, this implies $S_1 \leq \min[S_3; 1]$, and therefore this is the low broadcast SNR regime. At the other extreme, $\gamma^* = 1$, implying that the correlation bounds equal the pure multiaccess cut-set bound (2.59), if and only if $S_1 \geq 4S_3$.

We present these converse results graphically in Figure 2-9 for a representative symmetric Gaussian parallel relay network. We fixed the broadcast SNR, $S_1 = S_2 = 2$. We increased the multiaccess SNR, $S_3 = S_4$, from 0. The multiaccess SNR corresponds to the horizontal axis. The vertical axis corresponds to network communication rate. We have sketched two curves in the figure. One curve corresponds to the minimum of the broadcast and the multiaccess cut-set bounds, (2.50) and (2.59). Incidentally, when $S_1 = S_2 = 2$, the pure cross-cut bound, (2.68), is strictly larger than the minimum of the broadcast and multiaccess cut-set bounds. The second curve in the figure corresponds to the minimum of the broadcast cut-set and the correlation bound, (2.50) and (2.83). We do this in order to demonstrate the impact of the correlation bound. As $S_3$ increases from zero, the two curves diverge when $4S_3 = S_1$, which is when the maximizing $\gamma^*$ defining the correlation bound (2.83) becomes less than one. As $S_3$ increases further, the two curves reconverge when the correlation bound (2.83) equals
the broadcast cut-set bound (2.50). For moderate broadcast SNR’s, the correlation bound has little numerical effect. However, it does show that the minimum of these pure cut-set bounds is not tight. In fact, we will later show that there are asymmetric Gaussian networks where this correlation bound is asymptotically tight (in particular, tight with $\gamma^* \neq \{0, 1\}$). This shows that there is fundamental significance to the correlation bounds, (2.83) and (2.84).

![Graph](image)

Figure 2-9: Evaluation of converses for a symmetric Gaussian network, $S_1 = S_2 = 2$
CHAPTER 2. GAUSSIAN PARALLEL RELAY NETWORKS

2.3 Coding for Gaussian Parallel Relay Networks

2.3.1 Coding to the Relays

Assume without loss of generality that $S_1 \geq S_2$, and thus Relay 1 has a better quality observation than Relay 2. We design the first achievable scheme by reliably transmitting independent discrete messages from the sender to the two relays using broadcast coding techniques. Let $R_{top}$ be the rate of information sent to Relay 1. Let $R_{bot}$ be the rate of independent information sent to Relay 2. The relays reliably decode the relevant messages, remove the broadcast channel noise, and re-encode the information for transmission on the multiaccess channel. This is a natural idea for communication through the network. Though conventional design methodologies would generally lead to this, they would also lead to something very different from what we will do in theory on the multiaccess side. Specifically, a conventional design methodology would have the two relays send independent pieces of the overall message to the receiver. Generally, Relay 1 would send the information corresponding to $R_{top}$, while Relay 2 would send the information corresponding to $R_{bot}$. This would involve a very simple form of cooperation wherein the two relays avoid sending redundant information. Instead, we will use the cooperative technique that we exploited to extend the multiaccess capacity. Specifically, Relay 1 will send a scaled version of Relay 2’s signal plus an independent signal.

If the rate pair $(R_{top}, R_{bot})$ is in both the capacity region for the Gaussian broadcast channel, (2.2) and (2.3), and the extended capacity region for the Gaussian multiaccess channel, (2.7), then the rate $R_{ach} = R_{top} + R_{bot}$ is reliably achievable for the Gaussian parallel relay network. Note that in this scheme, when the relays correctly decode, they completely remove the noise induced by the broadcast channel before transmitting on the multiaccess channel.

For Gaussian parallel relay networks with a symmetric broadcast channel, where $S_1 = S_2$, this coding scheme simplifies significantly because the broadcast channel coding scheme degenerates. Specifically, since the relay observations are stochastically equivalent, both relays can decode both $R_{top}$ and $R_{bot}$. Since each relay can
reliably decode both messages, there is really only one message which is broadcast to both relays. The broadcast code, then, degenerates to a point-to-point channel code. Therefore, if we insist upon the relays decoding the input message stream reliably, the total rate is limited to, at most, the capacity of one broadcast channel link, i.e., to \( \frac{1}{2} \log_2 (1 + S_1) \) bits per channel use. On the multiaccess side, the relays can cooperate perfectly by sending the same codeword, scaled to meet their power constraints, which coherently combines at the receiver. The multiaccess side is then effectively a single-user Gaussian channel with input power constraint \( (\sqrt{P_{W_3}} + \sqrt{P_{W_4}})^2 \). Therefore, we can reliably transmit up to \( \frac{1}{2} \log_2 \left( 1 + \left( \sqrt{S_3} + \sqrt{S_4} \right)^2 \right) \) bits per channel use from the relays to the receiver. For a completely symmetric Gaussian parallel relay network, where \( S_3 = S_4 \) in addition to \( S_1 = S_2 \), this means we can reliably achieve

\[
R_{\text{ach}} = \frac{1}{2} \log_2 \left( 1 + \min \{ S_1; 4S_3 \} \right). 
\] (2.95)

When \( S_1 \geq 4S_3 \), (2.95) meets the multiaccess cut-set bound, (2.59), and thereby defines the capacity of the network. When \( S_1 < 4S_3 \), on the other hand, there is a gap between (2.95) and our tightest upper bound on capacity. The tightest upper bound equals the minimum of the broadcast cut-set, (2.50), and the correlation bound, (2.83). The broadcast cut-set bound states

\[
C_{\text{net}} \leq \frac{1}{2} \log_2 (1 + 2S_1). 
\] (2.96)

The correlation bound states

\[
C_{\text{net}} \leq \frac{1}{2} \log_2 \left( (1 + 2S_3(1 + \sqrt{\gamma^*})) \right) 
\] (2.97)

\[
= \frac{1}{2} \log_2 \left( (1 + S_1) \cdot (1 + (1 - \gamma^*)S_3) \right), 
\] (2.98)

and \( \gamma^* < 1 \) when \( S_1 < 4S_3 \) (c.f. (2.90)). Therefore when \( S_1 < 4S_3 \), the tightest upper bound is strictly larger than the rate achieved by coding to the relays.

Now consider the general asymmetric Gaussian network, still assuming \( S_1 \geq S_2 \) without loss of generality. The broadcast capacity region is given by (2.2) and (2.3), while the extended multiaccess capacity region is given by Theorem 2.1.1. We look
at the straightforward combination for coding to the relays. We want to determine the maximum achievable rate, $R_{\text{ach}} = R_{\text{top}} + R_{\text{bot}}$, such that $(R_{\text{top}}, R_{\text{bot}})$ is in the intersection of the broadcast and extended multiaccess capacity regions. The solution depends on the relative values of the four SNR’s. When the multiaccess side has much larger SNR’s than the broadcast side, then the extended multiaccess capacity region contains the broadcast capacity region. Conversely, when the broadcast side has much larger SNR’s than the multiaccess side, then the broadcast capacity region contains the extended multiaccess capacity region. In general, however, the two regions have a non-trivial intersection. For completeness we will explicitly present a closed-form solution to handle all cases where $S_1 \geq S_2$, but this will be rather uninformative.

Consider the broadcast and the extended multiaccess capacity regions. From (2.2) and (2.3), the capacity region of the broadcast channel is the set of $(R_{\text{top}}, R_{\text{bot}})$ such that, for some $\alpha \in [0, 1],

\begin{align}
R_{\text{top}} &\leq \frac{1}{2} \log_2 (1 + \alpha \cdot S_1), \\ R_{\text{bot}} &\leq \frac{1}{2} \log_2 \left( 1 + \frac{(1 - \alpha)S_2}{\alpha S_2 + 1} \right). 
\end{align}

For this broadcast channel, denote the maximum achievable $R_{\text{top}}$ by

$$R_{t,\text{BCST}}^* = \frac{1}{2} \log_2 (1 + S_1).$$

From (2.7), the extended capacity region of the multiaccess channel is the set of rate pairs $(R_{\text{top}}, R_{\text{bot}})$ such that, for some $\beta \in [0, 1],

\begin{align}
R_{\text{top}} &\leq \frac{1}{2} \log_2 (1 + (1 - \beta) \cdot S_3), \\ R_{\text{top}} + R_{\text{bot}} &\leq \frac{1}{2} \log_2 \left( 1 + S_3 + S_4 + 2 \sqrt{\beta S_3 S_4} \right).
\end{align}

For this multiaccess channel, define the maximum achievable $R_{\text{top}}$ by

$$R_{t,\text{MAC}}^* = \frac{1}{2} \log_2 (1 + S_3).$$

These two regions are roughly sketched in Figure 2-10.
We want to maximize $(R_{\text{top}} + R_{\text{bot}})$ over the set of pairs $(R_{\text{top}}, R_{\text{bot}})$ that lie in both of these capacity regions. We solve this explicitly. If, in addition to $S_1 \geq S_2$, we have $S_1 \leq S_3$, then $R^{*}_{t,BCST} \leq R^{*}_{t,MAC}$. We can therefore achieve any value of $R_{\text{top}}$ in the broadcast capacity region. We can choose $0 \leq \alpha \leq 1$ in (2.99) arbitrarily and set $\beta$ in (2.102) accordingly, implying $(1 - \beta) = \alpha \cdot \frac{S_3}{S_1}$. Then $R_{\text{bot}}$ equals the minimum of (2.100) and (2.103) for this choice of $\alpha$ and $\beta$. Therefore using this coding technique (and a little algebra), the best we can achieve by coding to the relays is given by

$$R_{\text{ach}} = \max_{\alpha \in [0,1]} \left[ \frac{1}{2} \log_2 (1 + \alpha S_1) + \frac{1}{2} \log_2 \left( 1 + \min \left[ \frac{(1-\alpha)S_2}{\alpha S_2 + 1} ; \frac{S_3 - \alpha S_1 + S_4 + 2 \sqrt{(S_3 - \alpha S_1)S_4}}{\alpha S_1 + 1} \right] \right) \right]. \quad (2.105)$$

On the other hand, if, in addition to $S_1 \geq S_2$, we have $S_1 \leq S_3$, then $R^{*}_{t,MAC} \leq R^{*}_{t,BCST}$. We can therefore achieve any value of $R_{\text{top}}$ in the extended multiaccess capacity region. We can choose $0 \leq \beta \leq 1$ in (2.102) arbitrarily and set $\alpha$ in (2.99) accordingly, implying $\alpha = (1 - \beta) \cdot \frac{S_3}{S_1}$. Then $R_{\text{bot}}$ equals the minimum of (2.100) and (2.103) for this choice of $\alpha$ and $\beta$. Therefore using this coding technique (and a little algebra), the best we can achieve by coding to the relays is given by

$$R_{\text{ach}} = \max_{\beta \in [0,1]} \left[ \frac{1}{2} \log_2 (1 + (1 - \beta)S_3) + \frac{1}{2} \log_2 \left( 1 + \min \left[ \frac{S_1 - (1-\beta)S_3S_2}{S_1 + (1-\beta)S_3S_2} ; \frac{\beta S_3 + S_4 + 2 \sqrt{\beta S_3 S_4}}{(1-\beta)S_3 + 1} \right] \right) \right]. \quad (2.106)$$

We can manipulate these algebraic expressions further or describe them geometrically,
but we do not find either approach particularly insightful. Instead, in what follows, we will describe specific coding choices which work well in various regimes. We can partition the full parametrization into three regions: \( S_2 \geq S_3 + S_4 + 2\sqrt{S_3 S_4} \), \( S_2 \leq \frac{S_4}{S_3+1} \), and \( \frac{S_4}{S_3+1} < S_2 < S_3 + S_4 + 2\sqrt{S_3 S_4} \). This partition is inspired by the converse results described in Section 2.2. From (2.90), \( S_2 \geq S_3 + S_4 + 2\sqrt{S_3 S_4} \) implies \( \gamma^* = 1 \) in the first correlation bound, (2.83). In this case, the first correlation bound equals the pure multiaccess cut-set bound, (2.59),

\[
C_{\text{net}} \leq \frac{1}{2} \log_2 \left( 1 + \left( \sqrt{S_3} + \sqrt{S_4} \right)^2 \right). \tag{2.107}
\]

From (2.88), \( S_2 \leq \frac{S_4}{S_3+1} \) implies \( \gamma^* = 0 \). In this case, the first correlation bound equals the pure cross-cut bound, (2.68),

\[
C_{\text{net}} \leq \frac{1}{2} \log_2 (1 + S_2) + \frac{1}{2} \log_2 (1 + S_3). \tag{2.108}
\]

Finally, from (2.92), \( \frac{S_4}{S_3+1} < S_2 < S_3 + S_4 + 2\sqrt{S_3 S_4} \) implies \( \gamma^* \in (0, 1) \). In this case, from (2.83) and (2.92), the first correlation bound states

\[
C_{\text{net}} \leq \frac{1}{2} \log_2 \left( (1 + S_2)(1 + (1 - \gamma^*)S_3) \right), \tag{2.109}
\]

where \( \gamma^* \) satisfies

\[
(1 + S_2)(1 + (1 - \gamma^*)S_3) = 1 + S_3 + S_4 + 2\sqrt{\gamma^*S_3 S_4}. \tag{2.110}
\]

**Case 1:** \( S_2 \geq S_3 + S_4 + 2\sqrt{S_3 S_4} \)

In this case, the broadcast side is very strong relative to the multiaccess side. The largest effective SNR we could achieve to the receiver, via perfect cooperation between the relays, is \( S_3 + S_4 + 2\sqrt{S_3 S_4} \). Because the broadcast side is so strong, we can get enough common information to both relays to allow for perfect cooperation at the maximum rate. We are thus limited only by the multiaccess side of the network. In this case, we send one message from the receiver which is decoded reliably by both relays. We can achieve \( R_{\text{ach}} = \frac{1}{2} \log_2 \left( 1 + S_3 + S_4 + 2\sqrt{S_3 S_4} \right) \) (choosing \( \beta = 1 \)
in (2.106)). This meets the multiaccess cut-set bound, (2.59), and thereby defines the
capacity of the network.

Case 2: \( S_2 \leq \frac{S_4}{S_3+1} \)

We consider a reasonable approach for a particular, highly asymmetric situation
lying in this regime. Consider the situation sketched pictorially in Figure 2-11. Here
the upper left and lower right channels are relatively strong, while the upper right
and lower left are relatively weak. Note that it is still possible to have \( S_1 \ll 1 \) or
\( S_2 \gg 1 \). Here we keep in mind the relevant pure cross-cut bound, (2.68), which states
\( C_{\text{net}} \leq \frac{1}{2} \log_2 (1 + S_2) + \frac{1}{2} \log_2 (1 + S_3) \). As mentioned above, in this case, the first
correlation bound equals the pure cross-cut bound. We will try to fill up the two
weak links with independent information. Let us start by choosing to completely fill
the top right link, i.e., choose \( R_{\text{top}} = \frac{1}{2} \log_2 (1 + S_3) \). We will now give an intuitive
explanation for why this is a reasonable choice in this regime.

![Diagram of a highly asymmetric network regime — weak cross-cut](image_url)

Consider the Gaussian broadcast capacity region, (2.99) and (2.100). In this
regime. \( S_1 \) is large relative to \( S_2 \). Therefore, by increasing \( \alpha \), we can significantly
increase \( R_{\text{top}} \), decodable only by Relay 1, while only slightly decreasing \( R_{\text{bot}} \), decodable by both relays. Now consider the Gaussian multiaccess capacity region
in this regime. Since \( S_3 \) is small relative to \( S_1 \), Relay 1 does not have enough
power to transmit a relatively large \( R_{\text{top}} \) to the decoder. Keep in mind that Relay
2 cannot decode \( R_{\text{top}} \), and therefore Relay 1 must transmit all of \( R_{\text{top}} \) to the
decoder on its own. Therefore, in (2.99) and (2.100), \( \alpha \) must be small. Furthermore,
\( R_{\text{bot}} \leq \frac{1}{2} \log_2 \left( 1 + \frac{(1-\alpha)S_2}{\alpha S_2 + 1} \right) \leq \frac{1}{2} \log_2 (1 + S_2) \).

Next, recall our intuitive explanation for achieving the Gaussian multiaccess channel
capacity region, (2.4). Since \( S_2 \leq \frac{S_4}{S_3+1} \), Relay 2 can successfully transmit all of
\( R_{\text{bot}} \) to the receiver without help from Relay 1. Indeed, the decoder can decode and strip out \( R_{\text{bot}} \) before \( R_{\text{top}} \), even if Relay 1 is transmitting \( R_{\text{top}} \) using all of its available power. We are thus free to make \( R_{\text{top}} \) as large as \( \frac{1}{2} \log_2 (1 + S_3) \).

To reiterate, we choose to fill the top right link, i.e., we choose \( R_{\text{top}} = \frac{1}{2} \log_2 (1 + S_3) \). Then \( R_{\text{bot}} \) is determined by the capacity region of the broadcast channel, (2.99) and (2.100). To connect this with our development above, assuming \( S_1 \geq S_3 \) as implied in Figure 2-11, we choose \( \beta = 0 \) in (2.106). This corresponds to no cooperation between relays. Then with \( \beta = 0 \), (2.106) becomes

\[
R_{\text{ach}} = \frac{1}{2} \log_2 (1 + S_3) + \frac{1}{2} \log_2 \left( 1 + \frac{(S_1 - S_3)S_2}{S_1 + S_3 S_2} \right).
\]  

(2.111)

In the limit \( \frac{S_3}{S_1} \to 0 \) and \( \frac{S_3 S_4}{S_1} \to 0 \), (2.111) becomes

\[
R_{\text{ach}} = \frac{1}{2} \log_2 (1 + S_3) + \frac{1}{2} \log_2 (1 + S_2).
\]  

(2.112)

The weak link SNR’s, \( S_2 \) and \( S_3 \), need not go to zero in this limit. For example, it is sufficient to fix \( S_2, S_3, \) and \( S_4 \) and to let \( S_1 \to \infty \). In this limit, from (2.112), \( R_{\text{ach}} \) meets the pure cross-cut bound, (2.68), and thereby defines the capacity of the network. In this case, the capacity equals the sum of the capacities of the two weak links.

As an alternative to filling the top right link, we could also choose to fill the bottom left link. This corresponds to choosing \( \beta = 1 \) — full cooperation between relays — in (2.106). With this choice we have \( R_{\text{top}} = 0 \), and we can achieve \( R_{\text{ach}} = \frac{1}{2} \log_2 (1 + S_2) \). However, using only the assumption \( S_1 \geq S_2 \), it can be shown algebraically that this is always inferior to (2.111).

\textit{Case 3:} \( \frac{S_4}{S_3 + 1} < S_2 < S_3 + S_4 + 2\sqrt{S_3 S_4} \)

We can again consider an asymmetric situation such as in Figure 2-11, but now the bottom right link may be weak as well. Intuitive reasoning similar to that of Case 2 applies here to Case 3. However, considering the broadcast capacity region, (2.99) and (2.100), we could now choose \( R_{\text{bot}} \) to be larger than \( \frac{1}{2} \log_2 \left( 1 + \frac{S_4}{S_3 + 1} \right) \). If we were to make such a choice for \( R_{\text{bot}} \), and if Relay 1 were to devote all its available power to transmitting \( R_{\text{top}} \), then the decoder could not successfully decode and strip out
2.3. CODING FOR GAUSSIAN PARALLEL RELAY NETWORKS

$R_{\text{bot}}$ in the presence of $R_{\text{top}}$. On the other hand, if $R_{\text{top}} < \frac{1}{2} \log_2 (1 + S_3)$, then Relay 1 can cooperate with Relay 2 to help send $R_{\text{bot}}$ to the receiver.

In this regime, the first correlation bound, (2.83), asserts

$$C_{\text{net}} \leq \frac{1}{2} \log_2 ((1 + S_2) (1 + (1 - \gamma^*) S_3)),$$

(2.113)

where $\gamma^* \in (0, 1)$ uniquely satisfies

$$(1 + S_2)(1 + (1 - \gamma^*)S_3) = 1 + S_3 + S_4 + 2\sqrt{\gamma^* S_3 S_4}.$$

(2.114)

Assuming $S_1 \geq S_3$, we can choose $\beta = \gamma^*$ in (2.106) to achieve

$$R_{\text{ach}} = \frac{1}{2} \log_2 (1 + (1 - \gamma^*) S_3) + \frac{1}{2} \log_2 \left( 1 + \frac{(S_1 - (1 - \gamma^*) S_3) S_2}{S_1 + (1 - \gamma^*) S_3 S_2} \right).$$

(2.115)

Herein we have chosen $R_{\text{top}} = \frac{1}{2} \log_2 (1 + (1 - \gamma^*) S_3)$, and both relays cooperate in sending $R_{\text{bot}}$.

In the limit $\frac{S_3}{S_1} \to 0$ and $\frac{S_3 S_2}{S_1} \to 0$, (2.115) becomes

$$R_{\text{ach}} = \frac{1}{2} \log_2 (1 + (1 - \gamma^*) S_3) + \frac{1}{2} \log_2 (1 + S_2).$$

(2.116)

Once again, the weak link SNR’s, $S_2$, $S_3$, and $S_4$, need not go to zero in this limit. For example, it is again sufficient to fix $S_2$, $S_3$, and $S_4$, and to let $S_1 \to \infty$. In this limit, $R_{\text{ach}}$ meets the first correlation bound, (2.83), and thereby defines the capacity of the network. In contrast to taking this same limit in the regime of Case 2 (c.f. (2.112)), the capacity here is strictly smaller than the sum of the capacities of the two weak links. Furthermore, the first correlation bound, (2.83), is tight with $\gamma^* \in (0, 1)$. This shows that there is fundamental significance to the correlation bounds, (2.83) and (2.84).
2.3.2 Forwarding Observations Without Decoding at the Relays

The coding technique we used in the last section cannot achieve rates greater than the capacity of a single broadcast link. Specifically, from the broadcast channel capacity region, \((2.99)\) and \((2.100)\), we are restricted to \(R_{\text{top}} + R_{\text{bot}} \leq \frac{1}{2} \log_2 (1 + \max \{S_1; S_2\})\).

Alternately, without relying on the defining equations, observe that the relay with less noise in its observation can reliably decode both \(R_{\text{top}}\) and \(R_{\text{bot}}\), and thus \((R_{\text{top}} + R_{\text{bot}})\) is limited to the capacity of the link to that relay. Instead of reliably sending discrete messages to each relay, we can take a very different approach. Between the two relays, we have two independent observations of the input signal, \(X\). This statistical diversity in the network potentially allows us to exceed the capacity of a single broadcast link.

![Logically equivalent network when both relays access both observations](image)

If the two observations were at the same point, or equivalently, if each relay had access to both observations, then the communication problem becomes easy. The two relays become a single combined relay with two observations, \(Y_1\) and \(Y_2\), and two transmission signals, \(W_3\) and \(W_4\). The logically equivalent network is drawn in Figure 2-12. In terms of reliable communication from the input terminal to the decoder, the network degenerates into a series of two point-to-point channels, the first with capacity \(\frac{1}{2} \log_2 (1 + S_1 + S_2)\) and the second with capacity \(\frac{1}{2} \log_2 (1 + (\sqrt{S_3} + \sqrt{S_4})^2)\).

The capacity of the network, \(C_{\text{net}}'\), is then equal to the minimum of the capacities of the two point-to-point links: \(C_{\text{net}}' = \frac{1}{2} \log_2 \left(1 + \min \left[S_1 + S_2; (\sqrt{S_3} + \sqrt{S_4})^2\right]\right)\). In this case, then, the capacity equals the minimum of the broadcast cut-set bound, \((2.50)\), and the pure multiaccess cut-set bound, \((2.59)\). To achieve this rate, the sender reliably transmits the discrete message to the combined relay. The combined relay re-encodes the message, transmitting the same codeword on \(W_3\) and \(W_4\) (scaled to meet the power constraints). Incidentally, when the capacities of the two effective
point-to-point links are unequal, that is, when \((S_1 + S_2) \neq (\sqrt{S_3} + \sqrt{S_4})^2\), one of the effective links will have unutilizable capacity. Continuing with the combined relay network, and assuming \((\sqrt{S_3} + \sqrt{S_4})^2 > \max[S_1; S_2]\), the network capacity in this case is greater than either broadcast link capacity alone. The point of discussing this combined relay network is to show that if both intermediate relays had access to both observations \(Y_1\) and \(Y_2\), then we would know exactly how to exploit the statistical diversity available in \(Y_1\) and \(Y_2\).

Return now to the original parallel relay network that we are studying, where these two relay observations are at distributed points. We now construct a coding technique for the parallel relay network that exploits the statistical diversity in the two relay observations. Consider a random code construction where we generate codewords using an i.i.d. Gaussian input ensemble, \(X \sim \mathcal{N}(0, P_X)\). Think first in terms of the symmetric Gaussian network. We can simply amplify \(Y_1\) at Relay 1 and \(Y_2\) at Relay 2 (each with one unit delay) subject to the relay power constraints. That is, we can set \(W_{3,k} = c \cdot Y_{1,k-1}\) and \(W_{4,k} = c \cdot Y_{2,k-1}\), where \(c = \sqrt{\frac{P_{w_3}}{P_X + N_{z_1}}} = \sqrt{\frac{P_{w_4}}{P_X + N_{z_2}}}\). The relays make no attempt to decode or even estimate the input codeword. This makes sense only if (and because) we are trying to communicate at rates above the capacity of a single broadcast link. We can view this procedure from an estimation perspective, where we want to minimize the mean square error in an estimate of \(X\) at the receiver. If \(X\) is i.i.d. Gaussian, this combines the relay observations optimally (and the core signal component \(X\) constructively) at the receiver before the multiaccess receiver noise, \(N_z\), is added. We would suspect that this is a good approach when the final multiaccess noise power, \(N_z\), is small.

This scheme results in an effective single-user Gaussian channel from \(X\) to \(Y\) with effective SNR

\[
\frac{4c^2 P_X}{c(N_{z_1} + N_{z_2}) + N_Z} = \frac{4S_3 S_1}{2S_3 + 1 + S_1} \\
= 4S_3 \left( \frac{S_1}{2S_3 + 1 + S_1} \right) \\
= 2S_1 \left( \frac{2S_3}{2S_3 + 1 + S_1} \right). \tag{2.17}
\]

Any codebook for this single-user channel that satisfies the average input power con-
CHAPTER 2. GAUSSIAN PARALLEL RELAY NETWORKS

constraint, $P_X$, will also satisfy the average power constraints at each relay. Therefore, from the single-user Gaussian channel capacity result, this approach yields an achievable rate

$$R_{ach} = \frac{1}{2} \log_2 \left( 1 + \frac{4S_3 S_1}{2S_3 + 1 + S_1} \right). \quad (2.120)$$

For comparison’s sake, we repeat the multiaccess cut-set bound, (2.59),

$$C_{net} \leq \frac{1}{2} \log_2 (1 + 4S_3), \quad (2.121)$$

and the broadcast cut-set bound, (2.50),

$$C_{net} \leq \frac{1}{2} \log_2 (1 + 2S_1). \quad (2.122)$$

If we fix the broadcast SNR, $S_1$, and we let the multiaccess SNR grow, $S_3 \to \infty$, then (2.120) becomes

$$R_{ach} = \frac{1}{2} \log_2 (1 + 2S_1). \quad (2.123)$$

In this limit, from (2.123), $R_{ach}$ meets the broadcast cut-set bound, (2.122), and thereby defines the asymptotic capacity of the network. We see that in this limit, we take full advantage of the two independent observations of $X$ in the network. It is worthwhile to note that in this limit as $S_3 \to \infty$, this approach asymptotically achieves network capacity even though both relays spend part of their allotted power amplifying the broadcast channel noise.
2.3. CODING FOR GAUSSIAN PARALLEL RELAY NETWORKS

Figure 2-13: Evaluation of converses and the two basic achievability schemes for a symmetric Gaussian network, $S_1 = S_2 = 2$

In Figure 2-13, we have presented the achievable coding results along with the converses for the same symmetric Gaussian parallel relay network we chose earlier. We will present a copy of this figure again in Figure E-1 on page 223 for easy reference. To reiterate, we fixed the broadcast SNR, $S_1 = S_2 = 2$. We increased the multiaccess SNR, $S_3 = S_4$, from 0. The multiaccess SNR corresponds to the horizontal axis. The vertical axis corresponds to network communication rate. We have again sketched the converses both with and without the correlation bound, (2.83). The point where the two versions of the converse diverge is the same point where the pure coding
achievability scheme diverges from the converses. Specifically, they all meet when 
$S_1 = 4S_3$. Note that we can time-share between the two achievability schemes. Time-
sharing between the two schemes with $S_1 = 2$ simply convexifies the maximum of the 
two achievable curves in Figure 2-13. We see from the figure that this time-sharing 
significantly improves the achievable rate in the “middle” region, where $4 \cdot S_3 > S_1$ 
but yet $S_3$ is not too large.

This begs the question whether a more complicated time-sharing scheme could 
 improve things further. Specifically, we use the coding approach with SNR’s $(S_1, S_3)_1$ 
a fraction $f$ of the time, and then we use the transponding approach with SNR’s 
$(S_1, S_3)_2$ for the remaining $(1 - f)$ fraction of time. The various parameters must 
be chosen so we do not violate either of the average power constraints on $S_1$ and $S_3$. 
For all reasonable SNR regimes, it turns out that, after numerically optimizing the 
various parameters, the best choice is the straightforward time-sharing scheme with 
constant $S_1$. After including the straightforward time-sharing scheme, Figure 2-13 
summarizes the state-of-the-analytical-art for the symmetric Gaussian network with 
$S_1 = 2.0$.

Let us now generalize the approach of forwarding the observations for the asym-
metric Gaussian network. We still assume, without loss of generality, that $S_1 \geq S_2$. 
For now, let us restrict each relay to simplify amplify its observation by a constant 
(again, with one unit delay). Let us assume that we use an i.i.d. Gaussian input, 
$X \sim \mathcal{N}(0, P_X)$, and that we wish to maximize the mutual information, $I(X^n; Y^n)$.

Contrary to initial intuition, it is not always best to use all available power at both 
relays. Returning to our estimation point of view, we would like to combine our ob-
servations in the correct ratio (known as optimal ratio combining). Initially, it seems 
intuitive that we should use all available power at Relay 1, since it has a higher quality 
observation. We may want to use all available power at both relays. However, con-
sidering that the relays are amplifying both signal and noise, this is not always best. 
Taken to an extreme, just to develop some intuition, we could end up swamping one 
of the observations when the opposite relay has too much power. On the other hand,

\footnote{Remember there is a coding theorem, which we mentioned but did not prove, that states that 
if we can achieve an end-to-end mutual information $\frac{1}{n} I(X^n; Y^n)$, then we can communicate reliably 
at rates arbitrarily close to $\frac{1}{n} I(X^n; Y^n)$.}
the decoder does not have access to the sum of the relay signals, $W_3 + W_4$. Rather, it receives $W_3 + W_4 + Z$, and thus we must fight against the multiaccess noise, $Z$.

Consider the following problem, pictured in Figure 2-14. We are given an input $X \sim \mathcal{N}(0, P_X)$, and we are restricted to using the relays as pure amplifiers with average power constraints. That is, $W_{3,k} = \alpha \cdot Y_{1,k-1}$ and $W_{4,k} = \beta \cdot Y_{2,k-1}$. The power constraint for Relay 1 implies $\alpha \in \left[ -\sqrt{\frac{P_{W3}}{P_X + N_{Z_1}}}, \sqrt{\frac{P_{W3}}{P_X + N_{Z_1}}} \right]$. The power constraint for Relay 2 implies $\beta \in \left[ -\sqrt{\frac{P_{W4}}{P_X + N_{Z_2}}}, \sqrt{\frac{P_{W4}}{P_X + N_{Z_2}}} \right]$. The goal is to choose $\alpha$ and $\beta$ to maximize the mutual information between sender and receiver, $I(X^n; Y^n)$. This is the same objective as choosing $\alpha$ and $\beta$ to maximize the effective SNR from $X$ to $Y$. This can be done by restricting $\alpha$ and $\beta$ to have the same sign, and therefore without loss of generality, we assume $\alpha \in [0, \alpha_{\text{max}}]$ and $\beta \in [0, \beta_{\text{max}}]$. The effective SNR is then equal to

$$\text{SNR}_{\text{eff}} = \frac{(\alpha + \beta)^2 P_X}{\alpha^2 N_{Z_1} + \beta^2 N_{Z_2} + N_Z}. \quad (2.124)$$

We will deal later with the case where there is no multiaccess noise, $N_Z = 0$. For now, assume that $N_Z > 0$ (and, of course, $P_X > 0$). Then we know that at the maximizing point, at least one of the two variables, $\alpha$ or $\beta$, must be equal to their upper bound. It is easy to see (and intuitive) that both $\alpha$ and $\beta$ must be non-zero at a maximizing point. If in addition they are both less than their upper bounds, then we can increase both $\alpha$ and $\beta$ while keeping $\frac{\beta}{\alpha}$ constant. This strictly increases the effective SNR.

Now we consider whether both variables must be equal to their maximum, and if not, which one must be. Evaluating the partial derivatives with respect to $\alpha$ and $\beta,$
we get

\[
\frac{\partial (\text{SNR}_{\text{eff}})}{\partial \alpha} = \left(\frac{2(\alpha + \beta)P_X}{(\alpha^2 N_{Z_1} + \beta^2 N_{Z_2} + N_Z)^2}\right) \cdot \left(\beta^2 N_{Z_2} + N_Z - \alpha \beta N_{Z_1}\right), \tag{2.125}
\]

\[
\frac{\partial (\text{SNR}_{\text{eff}})}{\partial \beta} = \left(\frac{2(\alpha + \beta)P_X}{(\alpha^2 N_{Z_1} + \beta^2 N_{Z_2} + N_Z)^2}\right) \cdot \left(\alpha^2 N_{Z_1} + N_Z - \alpha \beta N_{Z_2}\right). \tag{2.126}
\]

Assume we have found a maximizing pair \(\alpha^*\) and \(\beta^*\). It is clear from first principles that if \(\alpha^* \in (0, \alpha_{\text{max}})\), it must satisfy \(\frac{\partial (\text{SNR}_{\text{eff}})}{\partial \alpha}|_{\alpha^*} = 0\). If \(\alpha^* = \alpha_{\text{max}}\), it must satisfy \(\frac{\partial (\text{SNR}_{\text{eff}})}{\partial \alpha}|_{\alpha_{\text{max}}} \geq 0\) (e.g., see Bertsekas [10, Ch. 2] for justification of these statements.)

Considering (2.125), we see that \(\frac{\partial (\text{SNR}_{\text{eff}})}{\partial \alpha}|_{\alpha_{\text{max}}} \geq 0\) if and only if \(\alpha \leq \frac{\beta^2 N_{Z_2} + N_Z}{\beta N_{Z_1}}\). Similarly, \(\frac{\partial (\text{SNR}_{\text{eff}})}{\partial \beta} \geq 0\) if and only if \(\beta \leq \frac{\alpha^2 N_{Z_1} + N_Z}{\alpha N_{Z_2}}\).

When \(N_Z > 0\), these conditions imply that it is not always best to use all available power at both relays (for this restricted problem). In fact, somewhat surprisingly, these conditions imply that it is not always best to use all available power at Relay 1. This is somewhat surprising because Relay 1 has the better quality observation. But upon further thought, it is no longer surprising. Suppose we let the power constraint at Relay 1 go to \(\infty\) while holding the power constraint at Relay 2 fixed. If we let \(\alpha \to \infty\), then the effective SNR becomes (from (2.124))

\[
\text{SNR}_{\text{eff}} = \frac{P_X}{N_{Z_1}}. \tag{2.127}
\]

We see from (2.127) that we lose all diversity gain by letting \(\alpha \to \infty\) while keeping \(\beta\) fixed.

Now consider what happens when the multiaccess noise is absent. When \(N_Z = 0\), the effective SNR is maximized for any \(\alpha\) and \(\beta\) satisfying \(\frac{\beta}{\alpha} = \frac{N_{Z_1}}{N_{Z_2}}\). This corresponds to optimal ratio combining, in which case the effective SNR equals \((S_1 + S_2)\). This meets the broadcast cut-set bound on mutual information, (2.50):

\[
\frac{1}{n} I(X^n; Y^n) \leq \frac{1}{n} I(X^n; Y_1^n, Y_2^n) \tag{2.128}
\]

\[
\leq \frac{1}{2} \log_2 (1 + S_1 + S_2). \tag{2.129}
\]
2.3. **CODING FOR GAUSSIAN PARALLEL RELAY NETWORKS**

Therefore we can achieve network capacity using an arbitrarily small amount of power at each relay when we need not fight against multiaccess noise.

We next make explicit the optimality as the multiaccess noise power decreases to zero. We can set \( \frac{\beta}{\alpha} = \frac{N_2}{N_1} \) and increase their absolute values until the first of the two power constraints is met. Using this method of setting \( \alpha \) and \( \beta \), and using some simple algebra, we can see that in the limit as

\[
\min \left[ \frac{S_3}{S_1 + 1}, \frac{S_4}{S_2 + 1} \right] \rightarrow \infty,
\]

(2.130)

the effective SNR becomes (from (2.124))

\[
\text{SNR}_{\text{eff}} = S_1 + S_2.
\]

(2.131)

This implies that in this limit, we can achieve

\[
R_{\text{ach}} = \frac{1}{2} \log_2 (1 + S_1 + S_2).
\]

(2.132)

Therefore in this limit, from (2.132), \( R_{\text{ach}} \) meets the broadcast cut-set bound, (2.50), and thereby defines the capacity of the network.

As a final remark about this approach, we have restricted ourselves to holding \( \alpha \) and \( \beta \) constant for all symbols. Alternately, we could vary \( \alpha \) and \( \beta \) from symbol to symbol. We could also vary the power of the input symbol, \( P_X \). To be fully satisfied with the restriction we chose, we would need to determine whether there is a convex structure to the problem when we vary \( \alpha, \beta \), and the input symbol power on a symbol-by-symbol basis. This would then show that our restriction to constant values does not hurt. In fact, we will show in Section 2.5.1 that the problem does not have such a structure in general. However, based on the motivation for this approach, it makes sense to use this coding technique only when the multiaccess side has large SNR's. In the limit for this regime, our converse is equal to the broadcast cut-set bound, (2.50), which states

\[
C_{\text{net}} \leq \frac{1}{2} \log_2 (1 + S_1 + S_2).
\]

(2.133)
CHAPTER 2. GAUSSIAN PARALLEL RELAY NETWORKS

Based on (2.124) and the algebraic results that followed, it appears that this method can only approach the broadcast cut-set bound asymptotically as the multiaccess side gets strong in some sense. There is some subtlety in interpreting this statement. One way in which the multiaccess side can get strong is when both $S_3$ and $S_4$ get large. As we just found, when $S_3$ and $S_4$ get large, we already know how to asymptotically achieve the broadcast cut-set bound without varying $\alpha$, $\beta$, or the input symbol power, $P_X$. However, we may also think of a strong multiaccess side as simply having $S_3$ and $S_4$ large \textit{relative to} $S_1$ and $S_2$. We will explore this in a little more detail in Section 2.5.1 for the symmetric Gaussian network, where we will show that in the low input SNR limit, where $S_1 \to 0$, we must vary $\alpha$, $\beta$, and the power $P_X$ in a very simple way to get the analogous result.

2.4 On Degraded Channels

We now have an appropriate context within which to differentiate the concepts of physical and stochastic degradedness. In the introductory chapter, we defined a closely related communication problem called the relay channel. Most of the currently known analytical results on the relay channel were developed by Cover and El Gamal [13]. In that work, Cover and El Gamal found the capacity only of the physically degraded (and reversely physically degraded) relay channel. For simplicity, we will differentiate the two concepts of degradedness by the example of the Gaussian relay channel. The difference is the same as that outlined in [15, pg. 422] for the broadcast channel.

The Gaussian physically degraded relay channel is pictured in Figure 2-15. In this case, $Y_{1,k} = x_k + Z_{1,k}$, where $Z_{1,k}$ is a zero-mean Gaussian random variable with variance $N_{Z_1}$. Additionally, $Y_k = y_{1,k} + w_{3,k} + Z'_k$, where $Z'_k$ is a zero-mean Gaussian random variable with variance $N_{Z'}$.

For comparison, the Gaussian stochastically degraded relay channel is pictured in Figure 2-16. In this case, $Y_{1,k} = x_k + Z_{1,k}$ and $Y_k = x_k + Z_{2,k} + w_{3,k} + Z'_k$. Here $\{Z_{1,k}\}$, $\{Z_{2,k}\}$, and $\{Z'_k\}$, are independent Gaussian noise processes with i.i.d. components. $Z_{1,k}$, $Z_{2,k}$, and $Z'_k$ have variances $N_{Z_1}$, $N_{Z_2}$, and $N_{Z'}$, respectively. The stochastically degraded relay channel satisfies $(N_{Z_2} + N_{Z'}) \geq N_{Z_1}$ (conversely, according to [13], it
Figure 2-15: The physically degraded Gaussian relay channel

is called a reversely stochastically degraded relay channel when \((N_{Z_2} + N_{Z'}) < N_{Z_1}\). For the physically degraded relay channel of Figure 2-15, the analogous stochastically degraded relay channel would have \(N_{Z_1} = N_{Z_2}\).

Figure 2-16: The stochastically degraded Gaussian relay channel

To change the perspective somewhat, consider this latter picture, Figure 2-16. The physically degraded Gaussian relay channel is equivalent to the network pictured in Figure 2-16 when \(Z_1\) and \(Z_2\) are identical. That is, when \(Z_{1,k} = Z_{2,k}\) with probability one. This does not appear to be physically motivated, except for a particular case which we will mention below. For the stochastically degraded Gaussian relay channel, \(Z_{1,k}\) and \(Z_{2,k}\) are independent but may have the same power. At first, this difference may seem innocuous. Indeed, for the broadcast channel, the capacity region of the stochastically degraded channel is precisely the same as that of its physically degraded counterpart. However, for both the relay channel and for our parallel relay network, the difference is fundamentally important. The difference is that there is statistical diversity available in the stochastically degraded versions that is not available in the physically degraded versions.
CHAPTER 2. GAUSSIAN PARALLEL RELAY NETWORKS

For the physically degraded relay channel, if we have access to the relay observation \( Y_1 \), then the (degraded) receiver observation \( Y \) yields no useful information about the source input \( X \). Stated differently, the receiver observation \( Y \) is conditionally independent of the source input \( X \) when conditioned on the relay observation \( Y_1 \). The capacity of the physically degraded Gaussian relay channel is thus upper bounded by the capacity of the relay link,

\[
C_{\text{relay}} \leq \frac{1}{2} \log_2 \left( 1 + \frac{P_X}{N_{Z_1}} \right).
\]

(2.134)

In contrast, for the stochastically degraded relay channel, the direct path from source to receiver provides an additional noisy observation of the source input \( X \). In the stochastically degraded Gaussian relay channel, this additional noisy observation is provided via the direct path with independent additive noise \( Z_2 \). Speaking roughly and from an estimation point of view, we can reduce our uncertainty about \( X \) by somehow combining the information from the two relevant observations, \( X + Z_1 \) and \( X + Z_2 \). This is advantageous when \( Z_2 \neq Z_1 \) (i.e., when they are different sources of noise).

From this perspective, we can see the importance of recognizing the difference between a physically degraded and a stochastically degraded network. A physically degraded network has one relevant observation and an additional, degraded copy of the same observation. A stochastically degraded network has two relevant observations with possibly different qualities. Indeed, this is fundamentally why the converse to the physically degraded Gaussian relay channel does not apply to the stochastically degraded Gaussian relay channel. In fact, the capacity of the stochastically degraded Gaussian relay channel is at least as large as that of its physically degraded counterpart, and sometimes the capacity is strictly larger (specifically, when the input to relay observation SNR, \( \frac{P_X}{N_{Z_1}} \), is small).

There is an idealized situation where physical degradedness appears well motivated. One idealized model of feedback is when the relay observes both \( Y_{1,k} \) and \( Y_k \) at time \( k \). Starting with an arbitrary relay channel (possibly stochastically degraded, possibly not degraded at all), assume that the relay observes both its own private observation, \( Y_1 \), as well as the ultimate decoder’s observation \( Y \), as pictured in Fig-
2.4. ON DEGRADED CHANNELS

Figure 2-17: Idealized model for feedback for the Gaussian relay channel

In this model, there is neither delay nor noise in a feedback link from the decoder to the relay. This is indeed a physically degraded model since the decoder’s observation, $Y$, is a physically degraded version of the relay’s observation, $(Y_1, Y)$. Apart from this idealized feedback model, physically degraded networks appear to be of limited physical relevance.

Returning to our Gaussian parallel relay network, we have thus far been discussing the stochastically degraded version. If we still assume without loss of generality that $S_1 \geq S_2$, then the physically degraded version of our Gaussian parallel relay network would look exactly as drawn in Figure 2-1. However, the noise processes $\{Z_{1,k}\}$ and $\{Z_{2,k}\}$ would not be independent. Rather, we would have $Z_{2,k} = Z_{1,k} + Z'_{2,k}$ where $\{Z'_{2,k}\}$ is an independent Gaussian noise process with variance $N_{Z_2} - N_{Z_1}$. With the physically degraded Gaussian parallel relay network, our transponding approach no longer makes any sense because we no longer have two independent observations. In fact, the capacity of the symmetric, physically degraded parallel relay network can be determined by inspection (and is equal to $\sup_{p(x)p(w_3,w_4)} \min[I(X;Y_1), I(W_3,W_4;Y)]$). For the symmetric, physically degraded Gaussian parallel relay network,

$$C_{\text{net}} = \frac{1}{2} \log_2 (1 + \min [S_1; 4S_3]).$$

(2.135)

This is achieved using our pure coding technique, where the relays reliably decode the input message. As for the relay channel, the capacity of the stochastically degraded Gaussian parallel relay network is at least as large as that of its physically degraded counterpart, and we have shown that the capacity is strictly larger if and only if $S_1 \leq 4S_3$. 
2.5 Focusing on Symmetric Gaussian Networks

For the remainder of this chapter, we focus on symmetric Gaussian networks. Though asymmetric networks are important, and they are more generally representative of the physical situation we wish to understand, they are more difficult to think about. As we have seen, the symmetry allows us to describe our various approaches much more easily (since there are two SNR’s rather than four). It appears that the same fundamental questions arise even after we specialize to symmetric networks. In particular, the distributed estimation problem remains at the core of the communication problem. We have seen that when we are limited by the multiaccess side, we should code to the relays. Conversely, when we are limited by the broadcast side, we should try to take advantage of having two independent relay observations. When neither side dominates, we have yet to figure out exactly what to do.

2.5.1 Forwarding Observations Without Decoding at the Relays; Low SNR Effects

Consider the additive white Gaussian noise channel of Figure 2-18, and assume there

![Additive white Gaussian noise channel](image)

Figure 2-18: Additive white Gaussian noise channel

is an average power constraint \( P_X \) on the input to the channel, \( X \). We know that mutual information, \( I(X; Y) \), characterizes the maximum rate of reliable communication through the channel. For an additive white Gaussian noise channel with any SNR, an i.i.d. \( \sim \mathcal{N}(0, P_X) \) input process \( X \) maximizes the mutual information \( I(X; Y) \) subject to the average energy constraint \( P_X \) on \( X \). With this input process, the average
2.5. **FOCUSING ON SYMMETRIC GAUSSIAN NETWORKS**

mutual information equals

\[
\frac{1}{n} I(X^n; Y^n) = \frac{1}{2} \log_2 \left( 1 + \frac{P_X}{N_Z} \right) \\
= \frac{1}{2} \log_2 (1 + \text{SNR}) \\
= \frac{1}{2 \ln 2} \ln (1 + \text{SNR}).
\] (2.136)

In the limit as \( \text{SNR} \to 0 \), (2.138) becomes

\[
\frac{1}{n} I(X^n; Y^n) = \frac{1}{2 \ln 2} \cdot \text{SNR},
\] (2.139)

which is easy to see using the approximation (valid for small \( x \))

\[
\ln (1 + x) \approx x.
\] (2.140)

We can think of this low SNR regime as a very wideband communication situation, though for simplicity we continue to assume there is no intersymbol interference.

Asymptotically in the low SNR regime, there are other distributions which also maximize \( I(X; Y) \). One approach which works well in the low SNR regime is to use repetition codes. Equivalently, we can consider a random input process \( X \) where every \( k^{th} \) input symbol is generated i.i.d. \( \sim \mathcal{N}(0, P_X) \), and each random input symbol is repeated \( k \) times. It is straightforward to evaluate the mutual information. We evaluate the mutual information \( I(X; Y) \) per input symbol for one symbol of the repetition code, which is repeated \( k \) times. Since every \( k^{th} \) repetition code symbol is i.i.d., this equals the mutual information per input symbol evaluated over multiple repetition code symbols, each repeated \( k \) times.

\[
\frac{1}{nk} I(X^{nk}; Y^{nk}) = \frac{1}{k} I(X^k; Y^k) \]

\[
= \frac{1}{k} \left( H(Y^k) - H(Y^k \mid X^k) \right) \\
= \frac{1}{k} \left( H(Y^k) - \frac{1}{2} \log_2 \left( (2\pi eN_Z)^k \right) \right)
\] (2.143)
\[ \frac{1}{k} \left( \frac{1}{2} \log_2 \left( (2\pi e)^k |\Sigma_{Y^k}| \right) - \frac{1}{2} \log_2 \left( (2\pi eN_Z)^k \right) \right). \] (2.144)

In (2.144), \( \Sigma_{Y^k} \) is the covariance matrix of \( Y^k \), where \( Y_i = X + Z_i, i = 1, 2, \ldots k \), the \( Z_i \) are i.i.d. ~ \( \mathcal{N}(0,N_Z) \), and \( X \sim \mathcal{N}(0,P_X) \). Note that the input symbol is fixed over the block of \( k \) symbols. The \( k \times k \) covariance matrix \( \Sigma_{Y^k} \) has the following form:

\[
\Sigma_{Y^k} = \begin{bmatrix}
P_X + N_Z & P_X & P_X & \cdots & P_X \\
P_X & P_X + N_Z & P_X & \cdots & P_X \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
P_X & P_X & \cdots & P_X & P_X + N_Z
\end{bmatrix}. \quad (2.145)
\]

Because of the regular structure, it is straightforward to evaluate the determinant:

\[
|\Sigma_{Y^k}| = (N_Z)^{k-1} \cdot (k \cdot P_X + N_Z). \quad (2.146)
\]

Substituting (2.146) in (2.144), we conclude that the repetition code achieves the mutual information per input symbol

\[
\frac{1}{nk} I(X^{nk}; Y^{nk}) = \frac{1}{2k} \cdot \log_2 \left( 1 + \frac{k \cdot P_X}{N_Z} \right) \quad (2.147)
\]

\[
= \frac{1}{2k} \cdot \log_2 \left( 1 + k \cdot \text{SNR} \right). \quad (2.148)
\]

In the limit as SNR \( \rightarrow 0 \), (2.148) becomes

\[
\frac{1}{nk} I(X^{nk}; Y^{nk}) = \frac{1}{2 \ln 2} \cdot \text{SNR}. \quad (2.149)
\]

Again, this limit is easy to see using the approximation (2.140)

\[
\ln (1 + x) \approx x. \quad (2.150)
\]

Comparing (2.149) with (2.139), we see that a repetition code is as good as an i.i.d. ~ \( \mathcal{N}(0,P_X) \) input in the low SNR limit.

Another approach that works well in the low SNR regime is to remain silent for a
fraction of time, intermittently transmitting input symbols at higher power. We refer to this as the bursting approach. If we transmit one symbol with average energy $k \cdot P_X$ for every $k^{th}$ opportunity, and if these intermittent symbols are i.i.d. $\sim \mathcal{N}(0, k \cdot P_X)$, then the mutual information per input symbol equals

$$
\frac{1}{nk} I(X^{nk}; Y^{nk}) = \frac{1}{2k} \cdot \log_2 (1 + k \cdot \text{SNR}),
$$

(2.151)

which is precisely the same as that achieved with a repetition code, as we found in (2.148). Once again, in the limit as $\text{SNR} \to 0$, (2.151) becomes

$$
\frac{1}{nk} I(X^{nk}; Y^{nk}) = \frac{1}{2 \ln 2} \cdot \text{SNR}.
$$

(2.152)

For both the repetition code and the burst approach, the result is an effective SNR equal to

$$
\text{SNR}_{\text{eff}} = k \cdot \text{SNR}
$$

(2.153)

for every Gaussian i.i.d. variable generated in the input process. The repetition code achieves this by repeating a symbol $k$ times, where each symbol has average energy $P_X$. In this case, the receiver exploits statistical averaging of the $k$ independent noise variables $Z$ added to the input symbol. On the other hand, the bursting approach achieves the same effective SNR with a single channel use at $k$ times the average input power. In both cases, however, we use only a fraction $\frac{1}{k}$ of the degrees of freedom in the input process. With this perspective, it is easy to derive (i.e., to argue) that the equations we derived above, (2.148) and (2.151), are correct. Using the approximation $\ln(1 + x) \approx x$, valid for small $x$, we again see that for both the repetition and burst input processes,

$$
\frac{1}{k} \cdot \frac{1}{2} \log_2 (1 + k \cdot \text{SNR}) \approx \frac{1}{2 \ln 2} \cdot \frac{1}{k} \cdot k \cdot \text{SNR}
$$

(2.154)

$$
\approx \frac{1}{2 \ln 2} \cdot \text{SNR}
$$

(2.155)

$$
\approx \frac{1}{2 \ln 2} \cdot \frac{1}{2} \log_2 (1 + \text{SNR}).
$$

(2.156)
In other words, there is nothing lost, nothing gained with these variations of input structure for $X$ on an additive white Gaussian noise channel in the low SNR limit.

Returning to our Gaussian parallel relay network, it seems like a good idea to explore these variations of input structure in low SNR regimes. We will explicitly present some results for the bursting approach. Consider our approach of forwarding the observations, which from (2.120), achieves

$$R_{ach} = \frac{1}{2} \log_2 \left( 1 + \frac{4S_3S_1}{2S_3 + 1 + S_1} \right).$$

Since the relays simply amplify their observations when using this coding method, we will alternately refer to this as the transponding approach. We would like to think about this transponding approach in an appropriate low SNR regime. To this end, assume that we fix the ratio between $S_1$ and $S_3$; specifically, assume $S_3 = F \cdot S_1$ for some fixed factor $F$. Then (2.157) becomes

$$R_{ach} = \frac{1}{2} \log_2 \left( 1 + \frac{4F \cdot S_1^2}{(2F + 1) \cdot S_1 + 1} \right).$$

This approach has an effective SNR equal to

$$\text{SNR}_{eff} = \frac{4F \cdot S_1^2}{(2F + 1) \cdot S_1 + 1}.$$  

If we use the transponding approach in a burst mode, intermittently bursting both the input and the relays once every $k$ symbols, then both $S_1$ and $S_3$ increase linearly with $k$ for those symbols we use. In the small SNR regime where $S_1 \ll 1$ and $S_3 = F \cdot S_1 \ll 1$, from (2.159), this in turn increases the effective SNR quadratically. On the other hand, since we use the network only a fraction $\frac{1}{k}$ of the time, we take a linear penalty for not using all of the degrees of freedom available in the choice of inputs. Using the small $x$ approximation of (2.140), the net effect is an achievable rate that increases linearly with $k$. This approximation is reasonable provided our approximation

$$\ln(1 + x) \approx x$$
remains reasonable. As we increase \( k \), for those symbols we use, we transition from a low SNR regime into a higher SNR regime, where the concavity of the logarithm in (2.158) increasingly diminishes the quadratic gain promised by the low SNR approximation. Eventually it behooves us to stop increasing \( k \). We now proceed with presenting some explicit results for the burst transponding approach.

As we did above, we fix the ratio between the broadcast SNR \( S_1 \) and the multi-access SNR \( S_3 \). To make the analysis slightly easier, we will consider the behavior in the limit as the \( S_1 \) goes to zero. Specifically, assume \( S_3 = F \cdot S_1 \) for some constant factor \( F \). Assume the input transmits and both relays transpond a fraction \( \frac{1}{k} \) of time, implying \( k \in [1, \infty) \). We define \( k \) in this way to be consistent with the development above and because we like the form of the following equation better. Using (2.117) and some simple algebra, the achievable rate is given by

\[
R_{ach}(k) = \frac{1}{k} \cdot \frac{1}{2} \log_2 \left( 1 + \frac{4F(kS_1)^2}{(2F + 1)kS_1 + 1} \right). \tag{2.161}
\]

We should optimize over \( k \in [1, \infty) \) to maximize this achievable rate. When \( S_1 \) is not very small, then \( R_{ach}(k) \) is maximized with \( k = 1 \). This means that optimizing over the fraction of time we stay silent results in the previous, unoptimized approach (i.e., no bursting). To examine the low SNR effects, it is more illustrative to first take the limit as \( S_1 \to 0 \) and then to optimize over \( k \). There is no analytical problem in switching the order of the limit and the optimization — it just makes the solution to the problem easier to see. Optimized over the fraction of time we stay silent, we get

\[
R_{ach} = \frac{S_1}{2 \ln(2)} \cdot g(F), \tag{2.162}
\]

where

\[
g(F) = \sup_{z \geq 0} \frac{1}{z} \ln \left( 1 + \frac{4Fz^2}{(2F + 1)z + 1} \right). \tag{2.163}
\]

\( R_{ach} \) is still in bits per channel use — the \( \ln(2) \) term arises because we have changed the base of the logarithm from base 2 to base \( e \) in defining \( g(F) \). In the expression for \( g(F) \), (2.163), we have made the variable substitution \( z = kS_1 \). Since we are
examining the limit as $S_1 \to 0$, we optimize over $z \geq 0$. To examine the behavior for a fixed $S_1$, rather than examining the behavior in the limit as $S_1 \to 0$, we optimize over $z \geq S_1$ instead. We would like to point out that these equations, (2.162) and (2.163), are very good approximations when $S_1$ is small; we do not require $S_1 \to 0$ for this analysis to accurately predict low SNR behavior. One way to interpret (2.162) and (2.163) is to note that for any factor $F$, there is a fixed $S_{1,\text{min}}(F)$ such that, if $S_1 < S_{1,\text{min}}(F)$, bursting improves the simple transponding scheme.

![Graph](image)

Figure 2-19: Asymptotic behavior of burst-optimized approach.

The result of the maximization over $z \geq 0$ is plotted in Figure 2-19. It can be verified that $\lim_{F \to \infty} g(F) = 2$. To put this in a different perspective, we can think of the function $g(F)$ somewhat as a gain relative to the pure coding approach, where
we reliably decode the input message at the relays. From (2.95), the pure coding technique achieves

\[ R_{ach} = \frac{1}{2} \log_2 (1 + \min [S_1; 4S_3]). \]  \hfill (2.164)

In the low SNR regime, we can use the first order approximation \( \ln(1 + x) \approx x \) (reasonable for small \( x \)) implying that (2.164) becomes

\[ R_{ach} \approx \frac{S_1}{2 \ln(2)} \min [1; 4F]. \]  \hfill (2.165)

Therefore when \( F \geq 0.25 \), the pure coding approach achieves

\[ R_{ach} \approx \frac{S_1}{2 \ln(2)}. \]  \hfill (2.166)

Also note from (2.50) that the broadcast cut-set bound states

\[ C_{net} \leq \frac{1}{2} \log_2 (1 + 2S_1) \]  \hfill (2.167)

\[ \approx \frac{S_1}{2 \ln(2)} \cdot 2. \]  \hfill (2.168)

With this approximation, the broadcast cut-set bound is a factor of 2 larger than the rate of the pure coding approach. The first order approximation \( \ln(1 + x) \approx x \) is actually an upper bound,

\[ \ln(1 + x) \leq x \quad \forall \ x \geq 0. \]  \hfill (2.169)

Therefore the approximation (2.168) is actually an upper bound:

\[ C_{net} \leq \frac{1}{2} \log_2 (1 + 2S_1) \]  \hfill (2.170)

\[ \leq \frac{S_1}{2 \ln(2)} \cdot 2. \]  \hfill (2.171)

Comparing (2.162) with (2.171), we must have \( g(F) \leq 2 \) for all ratios \( F \).

We want to point out a behavioral difference in the low input SNR regime. We
saw earlier that when the multiaccess side gets strong in absolute terms, which corresponds to \( F \to \infty \) while holding \( S_1 \) fixed, then the unoptimized approach (equivalent to setting \( k = 1 \) in (2.161)) asymptotically achieves the broadcast cut-set bound (c.f. (2.123)). Moreover, when \( S_1 \) is not too small, then the maximization over \( k \geq 1 \) is achieved at \( k = 1 \), and thus the optimized approach is the same as the unoptimized approach. In other words, when \( S_1 \) is not too small, it is best to use the transponding approach without bursting.

Consider what happens in the low input SNR limit when we do not burst, which corresponds to setting \( k = 1 \) in (2.161). When the multiaccess side get strong in relative terms (\( F \to \infty \)), then we do not achieve the broadcast cut-set bound, (2.50). Specifically, using (2.120) and (2.50), and applying L’Hôpital’s rule to evaluate the limit,

\[
\lim_{s_3 = F \cdot s_1, s_1 \to 0} \frac{\frac{1}{2} \log_2 \left( 1 + \frac{4 s_3 s_1}{2 s_3 + 1 + s_1} \right)}{\frac{1}{2} \log_2 (1 + 2 s_1)} = 0 \quad \forall \ F.
\]  

(2.172)

The simple transponding scheme with no bursting thus becomes disastrous in the low SNR limit as \( S_3 \to 0 \), regardless of \( F = \frac{S_3}{S_1} \).

On the other hand, consider what happens in the low SNR limit when we optimize over \( k \). Now when the multiaccess side gets strong in relative terms (\( F \to \infty \)), once again we can asymptotically achieve the broadcast cut-set bound, (2.50). Mathematically, using (2.162), (2.163), and (2.50), we mean this in the following sense:

\[
\lim_{F \to \infty} \lim_{s_3 = F \cdot s_1, s_1 \to 0} \frac{\frac{S_1}{2 \log_2 (1 + 2 S_1)} \cdot g(F)}{\frac{1}{2} \log_2 (1 + 2 S_1)} = 1.
\]

(2.173)

To conclude this section, recall that in addition to bursting, repetition codes work equally well on the point-to-point additive white Gaussian noise channel of Figure 2-18. We developed several interesting results in this section concerning burst transponding for the Gaussian parallel relay network. It is natural to explore how repetition coding performs for the Gaussian parallel relay network. Based on the intuition developed for the point-to-point channel, we would suspect that the performance is identical. Indeed, the performance is identical for the straightforward
approaches we explored. Specifically, consider the following methods:

1. Burst (with factor $k$) each input symbol $X$, and simultaneously burst (with factor $k$) the associated received observation $Y_i$ at each relay.

2. Repeat each input symbol $X$ $k$ times, and simply amplify the received observation $Y_i$ at each relay.

3. Repeat each input symbol $X$ $k$ times, form the minimum mean square error (MMSE) estimate of the input symbol at each relay, then repeat the MMSE estimate $k$ times at each relay.

4. Repeat each input symbol $X$ $k$ times, form the MMSE estimate of the input symbol at each relay, then simultaneously burst (with factor $k$) the MMSE estimate at each relay.

5. Burst (with factor $k$) each input symbol $X$, then repeat the associated received observation $Y_i$ $k$ times at each relay.

The first approach is the burst transponding method we studied earlier in this section. Each of these approaches yields the same mutual information, $\frac{1}{nk}I(X^{nk}; Y^{nk})$, expressed in (2.161). In other words, they are all equivalent.

### 2.5.2 Combining the Two Main Approaches

We can combine the two basic approaches, coding to the relays and transponding the relay observations, on the Gaussian network. But the question is how best to combine them. We can first try to combine them simultaneously. We can structure the input much like we do when communicating independent information streams over a broadcast channel, splitting the power at the source $X$ for two independent data streams. The first stream is decoded and stripped out at the relays, while the second is amplified as before with the additional broadcast link noise. The relays also split their power between the two streams. To the final decoder, it looks like a two-user Gaussian multiaccess channel with independent messages. One “user” corresponds to the cooperating relays, which have decoded the first stream in common. The second “user” corresponds to the transponded stream.
CHAPTER 2. GAUSSIAN PARALLEL RELAY NETWORKS

We can show algebraically that decoding and stripping out the first “user” at the receiver is always better than decoding the second “user” first. Then for all choices of power splits and for every pair of SNR’s (S₁, S₃), we can show algebraically that either the pure coding scheme, where all information is reliably decoded at each relay, or the pure transponding scheme, where the observations are simply amplified, is at least as good as the simultaneously combined scheme. Therefore after optimizing over how the input power is split between information streams, the simultaneously combined approach degenerates to one of the two primary approaches for every pair of SNR’s S₁ and S₃.

However, from Figure 2-13, we see that time-sharing between the two approaches can significantly improve performance over either approach alone. Apparently, this is the “correct” way to combine the two approaches. After all, decoding at the relays works well when \( \frac{S₁}{S₃} \) is large, while transponding works well when \( \frac{S₃}{S₁} \) is large. We operate in these regimes when we time-share to convexify the achievable curve. Specifically, we time-share between coding at large \( \frac{S₁}{S₃} \) and transponding at large \( \frac{S₃}{S₁} \). But we can also operate in these regimes with the simultaneously combined approach by devoting a large fraction of power at the relays to the transponded stream. Specifically, in this scenario, the decoded stream is effectively operating with a large \( \frac{S₁}{S₃} \), while the transponded stream is effectively operating with a large \( \frac{S₃}{S₁} \). And yet this simultaneously combined scheme is always inferior to one of the pure schemes. We thus need some other intuition to explain this algebraic result. Our best explanation, which should be viewed with some skepticism, is as follows. When we time-share between approaches rather than combine them simultaneously, neither method must fight against the other as additional noise. With the simultaneously combined scheme, the first “user” must fight against the second when being decoded at the relays. Additionally, at the ultimate decoder, whichever “user” is decoded and stripped out first must fight against the other “user”.

Incidentally, with the straightforward time-sharing scheme we just described, we use the two approaches at the same S₁ but with different S₃. We could opt for a more general power allocation where both S₁ and S₃ are different. However, it turns out that it is best to keep S₁ the same for both approaches when time-sharing.
2.5.3 An Idea That Does Not Work — Using a Gauss-Markov Structure for the Input

We explore changing the input distribution from an i.i.d. Gaussian, \( X \sim \mathcal{N}(0, P_X) \), to a first-order Gauss-Markov process. This reduces the net information rate (entropy) sent into the broadcast channel, and thus we cannot hope to meet the broadcast data processing bound, (2.50), using non-degenerate Markov processes. The idea is to put some temporal structure in the source \( X \), thereby improving the ability of the relays to estimate the source. Specifically, we do not require the relays to fully decode the input source, but they can perform some rudimentary input sequence estimation. This is a sort of compromise between the two extremes of reliable decoding and transponding, and we would hope this approach would be useful in the middle SNR range where time-sharing is particularly beneficial.

We can do two reasonable things at the relays. First, we can run a Kalman estimator at each relay, based on the observation \( Y_1 \) or \( Y_2 \), and then amplify the current symbol estimate, \( \hat{X}(Y_1) \) or \( \hat{X}(Y_2) \). Second, observe that a Kalman estimator based on both observations, \( \hat{X}(Y_1, Y_2) \), consists of a linear function of \( Y_1 \) and \( Y_2 \) — there are no cross-terms. Therefore, such an estimator can be implemented in a distributed manner by the two relays acting on their observations alone. The two components are added automatically on the multiaccess channel when the relay transmissions are received at the decoder, albeit with additional noise, \( Z \). Intuitively and numerically, the second method is superior to the first. In both cases, there is a core source signal component in each relay transmission signal which combines constructively at the decoder.

We can analytically derive the mutual information rate, \( \lim_{n \to \infty} \frac{1}{n} I(X^n; Y^n) \), for each method. The derivation is cumbersome and nontrivial. Having derived the mutual information rate, we believe we could prove that this is a reliably achievable data rate for the Gaussian network. The unfortunate result is that for all reasonable broadcast SNR’s (specifically, when \( S_1 \) is not too small), the first-order Gauss-Markov process that maximizes this mutual information rate is a degenerate one — i.i.d. \( \mathcal{N}(0, P_X) \). The estimator at the relays then degenerates to symbol-by-symbol amplification by a constant, and we end up with the transponding scheme.

For very low broadcast SNR’s, a non-degenerate Gauss-Markov process maximizes
the mutual information rate. However, numerically, the mutual information rate is inferior to our improved time-sharing scheme outlined in the “low SNR” section, 2.5.1. We did not explore the possibility of staying silent for a fraction of time and then using a Gauss-Markov process for the remainder at higher power. The intuition for not exploring this is that optimizing over input Gauss-Markov processes at higher power resulted in a degenerate Markov process.

Because this is a negative result, and because the derivation is complicated, we do not provide it here. We also see no reason to pursue the random coding argument.

2.5.4 Applying Block Quantization and Independent Messages

We next borrow an idea that works well for discrete alphabet networks and attempt to apply it to the Gaussian network. Observe that in both the Gaussian and discrete networks, if the decoder had access to both relay observations $Y_1$ and $Y_2$, then communication is limited only by the ability to get information across the memoryless broadcast channel. Specifically, we would then be limited only by the broadcast cut-set bound, $\max_{p(x)} I(X; Y_1, Y_2)$. For the Gaussian average power limited network we have been studying in this chapter, an i.i.d. Gaussian input, $p(x) \sim \mathcal{N}(0, P_X)$, maximizes the mutual information across the broadcast channel. Further, when the input is i.i.d. Gaussian and the broadcast channel is symmetric, $p(x) \sim \mathcal{N}(0, P_X)$ and $S_1 = S_2$, then $I(X; Y_1, Y_2) = I(X; c \cdot (Y_1 + Y_2))$ for any constant $c$. This is consistent with our development of the transponding approach to communicating through the network, and it is consistent with the fact that $c \cdot (Y_1 + Y_2)$ is a sufficient statistic for the a posteriori distribution of $X$ based on the relay observations $Y_1$ and $Y_2$.

In the discrete networks, if the multiaccess side is sufficiently strong, then the decoder can recover the pair of observations, $(Y_1, Y_2)$, exactly. The relays can simply encode their observations using, say, Huffman codes, and then reliably transmit their observations to the decoder. The decoder can then decode the input $X$ based on the exact pair of relay observations. Together, the relay observations contain all relevant information about the source available anywhere in the network away from the input terminal.

If the multiaccess side is not quite strong enough to admit an error free reconstruc-
2.5. FOCUSING ON SYMMETRIC GAUSSIAN NETWORKS

...tion of \((Y_1, Y_2)\), then the relays can instead encode a reasonable representation of the observations. The decoder can reconstruct a distorted version of \((Y_1, Y_2)\) if the relays use rate-distortion codebooks (or, more precisely, block quantization codebooks) for their observations. For the Gaussian network, as long as the multiaccess noise \(Z\) has positive power, \(N_Z > 0\), it is impossible to reproduce the pair of observations at the decoder without distortion. However, we can quantize the observations and send the quantized representation. The stronger the multiaccess side, the finer the quantization achievable, and the lower the resulting distortion of the observations reproduced by the decoder.

We now sketch how one could apply this approach for the Gaussian network. Given a symmetric Gaussian parallel relay network, we begin by determining a communication rate for the relays. In other words, we effectively convert the multiaccess channel into a pair of independent, noiseless binary links to the receiver. We do so by considering the capacity region of the multiaccess side with independent messages, and we maximize the data rate achievable for the relays. According to (2.4), both relays can communicate reliably and simultaneously up to \(\frac{1}{4} \log_2 (1 + 2S_3)\) bits per channel use.

The difficult aspect of this approach is to determine the block mutual information achievable from the input to the decoder. In particular, after block quantizing at the relays, it is difficult to determine the joint relationship between the input and the pair of quantized observations. This joint statistical relationship is necessary both to compute the mutual information and to exploit random binning, which is a technique often used in the information-theoretic source coding literature. The argument will be carried out precisely for discrete networks, but the analysis relies on the notion of strong typicality, which in turn relies on the finiteness of the alphabets.

We made some reasonable approximations in a pseudo-analysis for the Gaussian network. We then pushed through the computations to see how well this approach could potentially perform (if a precise analysis were tractable). Unfortunately, this approach appears significantly inferior to the simple transponding scheme. Therefore, we see no reason to pursue a precise analysis.

We believe we understand why this approach is inferior. Both of our best approaches thus far, coding to the relays and transponding, exploit an effective power...
boost in the received signal at the decoder. This power boost arises from a common component in the relay transmission signals which combines constructively at the decoder. With this block quantization approach, on the other hand, we transform the multiaccess channel into a pair of noiseless channels by assuming independent messages from the relays. We thereby eliminate the possibility of cooperating through constructive combining on the multiaccess channel. Such a scheme intuitively appears doomed to failure on the Gaussian network. We have attempted to support this intuition with a reasonable coding approach and approximate analysis.

2.5.5 Difficulty in Applying the Entropy Power Inequality for Gaussian Parallel Relay Networks

If we want to derive a converse for the Gaussian network using only entropy and mutual information identities, data processing inequalities, and the entropy power inequality, then we have very few tools at our disposal. These tools are sufficient for proving tight converses for the Gaussian broadcast channel, the white Gaussian multiaccess channel with feedback, and the Gaussian CEO problem (in the limit of an infinite number of relays). Basically, we know the normal (i.e., Gaussian) distribution maximizes differential entropy under a second moment constraint. Denoting by $\text{VAR}(U)$ the variance of the random variable $U$,

$$\text{VAR}(U) \leq P_U \Rightarrow h(U) \leq \frac{1}{2} \log(2\pi e P_U). \quad (2.174)$$

We also have the entropy power inequality, which states that if $U^n$ and $V^n$ are independent and have probability densities, then

$$2^{\frac{1}{n} h(U^n + V^n)} \geq 2^{\frac{1}{n} h(U^n)} + 2^{\frac{1}{n} h(V^n)}. \quad (2.175)$$

In (2.175), $h(U)$ is the differential entropy of the random variable $U$. If one of the variables $U$ or $V$ is (conditionally) i.i.d. Gaussian, we can derive this same inequality in conditional probability space. For the multiaccess channel in our Gaussian parallel relay network, we have relay transmission signals $W_3^n$ and $W_4^n$ and independent additive white Gaussian noise $Z^n$. Then applying the entropy power inequality of (2.175)
when $X^n = x^n$,

$$
h(W_3^n + W_4^n + Z^n \mid X^n = x^n) \geq \frac{n}{2} \log \left( 2 \pi e \left( \frac{1}{2} h(W_3^n) + \frac{1}{2} h(W_4^n) + \frac{1}{2} h(Z^n) \right) \right)$$

(2.176)

$$
eq \frac{n}{2} \log \left( 2 \pi e \left( \frac{1}{2} h(W_3^n + W_4^n | X^n = x^n) + 2 \pi e N_Z \right) \right).$$

(2.177)

The right-hand side is convex in $h(W_3^n + W_4^n | X^n = x^n)$, and thus

$$
h(W_3^n + W_4^n + Z^n \mid X^n) \geq \frac{n}{2} \log \left( 2 \pi e \left( \frac{1}{2} h(W_3^n + W_4^n | X^n) + 2 \pi e N_Z \right) \right).$$

(2.178)

Recall that this inequality holds provided both $(W_3^n + W_4^n)$ and $Z^n$ (conditioned on $X^n$) have probability densities. If $(W_3^n + W_4^n)$ is discrete, as it is when the relays send discrete codewords, then this inequality still holds if we take the limit $h(W_3^n + W_4^n | X^n) \to -\infty$.

Applying (2.178) in the Gaussian parallel relay network, we get

$$
I(X^n; Y^n) = h(Y^n) - h(Y^n | X^n) \leq \frac{n}{2} \log \left( \frac{2 \pi e h(Y^n)}{2 \pi e (h(W_3^n + W_4^n | X^n) + 2 \pi e N_Z)} \right).$$

(2.179)

After thinking in depth about the Gaussian parallel relay network, it appears intuitive that if $Y^n$ can be used to reliably decode $X^n$, and if the relays cannot reliably decode $X^n$ on their own, then we should not be able to achieve the upper bound

$$
h(Y^n) \leq \frac{n}{2} \log (2 \pi e (4P_{W_3} + N_Z)).$$

(2.180)

It seems that this upper bound on $h(Y^n)$ can only be achieved when the relays essentially decode the input message $X^n$ and then re-encode the message using an identical codebook for transmission on the multiaccess channel. In turn, from (2.179), this would imply that $C_{net} < \frac{1}{2} \log_2 (1 + 4S_3)$ whenever $S_1 < 4S_3$. Of course, it would be better to bound $h(Y^n)$ away from the upper bound of (2.180) rather than simply stating that it must be smaller. After all, when $S_1 < 4S_3$, we already know that $C_{net} < \frac{1}{2} \log_2 (1 + 4S_3)$ since the correlation bounds (2.83) and (2.84) are strictly smaller here (c.f. (2.90)).
On the other hand, consider the implication of (2.179) if our intuition were wrong. If the upper bound of (2.180) could be achieved arbitrarily closely, then it could be achieved arbitrarily closely using discrete codebooks for $W_3^n$ and $W_4^n$. To see this, if $W_3^n$ and $W_4^n$ are not discrete codebooks to begin with, then simply discretize them finely — since $Y = W_3 + W_4 + Z$, and since $Z^n$ has a smooth probability density, $h(Y^n)$ will remain relatively unchanged. When $W_3^n$ and $W_4^n$ are discrete, $h(W_3^n + W_4^n \mid X^n) \to -\infty$. Then (2.179) merely states that

$$\frac{1}{n} I(X^n; Y^n) \leq \frac{1}{2} \log_2 (1 + 4S_3),$$

(2.181)

which is nothing more than the pure multiaccess cut-set bound, (2.59). As mentioned, when $S_1 > 4S_3$, this is strictly larger than our known converse.

Without proof, we cannot rule out the possibility that $W_3^n$ and $W_4^n$ are discrete codebooks which approximately meet the upper bound of (2.180). To see that this upper bound can sometimes be met, we can construct such a pair of random vectors as follows. Find a good communication codebook of rate $\frac{1}{2} \log_2 (1 + 4S_3)$ for the point-to-point AWGN channel with an average input power constraint yielding SNR equal to $4S_3$. For $X^n(i)$, the $i$th codeword from this codebook, let $W_3^n(i) = W_4^n(i) = \frac{1}{2} X^n(i)$. If we could manage to transmit the same corresponding codeword from each relay essentially all of the time, then indeed, the upper bound on $h(Y^n)$ would be met with approximate equality. In fact, this is exactly what we do with the pure coding method. The pure coding method works this way provided $S_1 \geq 4S_3$, which puts us back in the regime where each relay can reliably decode the input message.

We have tried many variations of manipulations based on the entropy power inequality, but invariably we encounter the same difficulty when considering discrete codewords for the relay transmissions. Apart from the implication of the correlation bounds (2.83) and (2.84), we cannot get a handle on bounding $h(Y^n)$ when the communication rate exceeds the capacity of the broadcast channel links.
Chapter 3

Discrete Memoryless Parallel Relay Networks

Converse results seem to be particularly difficult to derive for our parallel relay networks. We have thought extensively about achievability schemes as well as approaches for deriving upper bounds to network capacity. We believe that, in general, taking the minimum of various cut-set bounds does not provide a tight upper bound to network capacity. Indeed, for the Gaussian network that we considered in the previous section, we derived correlation bounds which tightened the cut-set bounds by considering more than one simultaneously, imposing the relay power constraints, and manipulating the second-order statistical relationship between the relay transmissions.

We hope to gain some insight into this complex communication problem by eliminating the power constraints altogether. A simpler system to consider is a discrete memoryless network. Discrete networks are somewhat simpler analytically than continuous alphabet networks with power constraints.

There are fundamental differences, however. In a discrete system, we can describe all relevant events with a finite number of bits.\(^1\) As a deeper example, in the discrete case, we can develop the notion of strong typicality. Sequences in the strongly typical set have a sort of uniformity not present in the weakly typical set. We will use strong typicality to develop an achievable coding scheme based on block quantization.

---

\(^1\)In the information theory literature, a discrete alphabet is usually considered finite. Mathematically, of course, a discrete set may have a countably infinite number of members.
at the relays. For our purposes, strong typicality is primarily useful in developing achievable coding techniques. We derive converse bounds using the usual arsenal of Fano’s inequality, data processing inequalities, and convexity. It appears that combinatorial approaches may be another useful weapon for our arsenal, but we have been unable to successfully exploit combinatorics thus far.

![Diagram](image)

Figure 3-1: The asymmetric PBSC network with noiseless relay channels

We primarily study discrete networks where the relays are followed by a pair of noiseless binary links with rates $R_3$ and $R_4$. We will most often assume that $R_3 = R_4$. In Figure 3-1, we redraw the discrete network on which we focus. This is called the parallel binary symmetric channel (PBSC) network. The received signal, $Y$, is now the pair of binary relay transmission signals, $(W_3, W_4)$. We normalize network capacity to bits per input symbol use. Recall from our definition in Section 1.1 that for $n$ input channel uses, denoted $X^n$, Relay 1 can noiselessly transmit $nR_3$ binary symbols, denoted $W_3^nR_3$, while Relay 2 can noiselessly transmit $nR_4$ binary symbols, denoted $W_4^nR_4$.

We define $C_{\text{net}}$ as the supremum of communication rates from input to decoder with arbitrarily low probability of error. We will prove in Theorem 3.4.1 that we can equivalently define network capacity as the maximum achievable mutual information rate between input and output. Specifically, if we define

$$C_{\text{net}}' = \sup \frac{1}{n} I(X^n; W_3^nR_3, W_4^nR_4),$$

(3.1)

then

$$C_{\text{net}} = C_{\text{net}}'.$$ 

(3.2)
That is, starting with the definition of $C_{\text{net}}'$, we will prove in Theorem 3.4.1 that we can reliably achieve communication rates arbitrarily close to $C_{\text{net}}'$, and conversely, that reliable communication is impossible for rates above $C_{\text{net}}'$. This gives the definition of $C_{\text{net}}'$ operational significance, and thus we will use these two equivalent definitions interchangeably until we prove Theorem 3.4.1 in Section 3.4.

Next, consider the class of parallel relay networks where the multiaccess channel is a pair of noiseless binary links. The PBSC network of Figure 3-1 is in this class. In a sense that we will explain shortly, this class of parallel relay networks is both the worst class we could use from a communication standpoint and the best class we could study when trying to develop converse theorems.

Let us first explain why, from a communication standpoint, this class of parallel relay networks is the worst class we could have. If we start with an arbitrary multiaccess channel rather than a pair of noiseless binary links, we can restrict the relays to using multiaccess codes that are designed for users with independent messages. If the rate pair $(R_3, R_4)$ is in the capacity region of the given multiaccess channel, then we can effectively convert the multiaccess channel into a pair of noiseless bitpipes with rates $R_3$ for Relay 1 and $R_4$ for Relay 2. This is one way of communicating information from the relays to the decoder over the multiaccess channel, but it is a restriction because the relays could potentially use other codes for communicating over the multiaccess channel. For example, for the Gaussian parallel relay network with a Gaussian multiaccess channel, we developed a technique where the relays transpond their observations rather than attempting to send their own reliably decodable message to the receiver. That technique actually relies on the fact that the multiaccess channel combines the relay transmissions at the receiver.

For essentially the same reason, from an analytical standpoint, this class of parallel relay networks is the best class we could study. Mathematically, the capacity of a parallel relay network with noiseless bitpipes of rates $R_3$ and $R_4$ will be upper bounded by the capacity of a more general parallel relay network where $(R_3, R_4)$ is in the capacity region of the multiaccess channel. Thus we cannot hope to understand what is limiting our ability to communicate through the general parallel relay network if we do not first understand this restricted class of parallel relay networks.
3.1 Discrete Broadcast and Multiaccess Channels

When studying discrete parallel relay networks with a symmetric broadcast channel, we do not need broadcast coding techniques for reliable coding to the relays. Since the relay observations are statistically identical, any message that can be reliably decoded by one relay can also be reliably decoded by the other.

However, we will also focus on a particular, highly asymmetric PBSC network. Specifically, we will consider the case where the BSC link from the sender to Relay 1 is a clean binary link. In other words, $Y_1 = X$. Referring to Figure 3-1, we set $\alpha_1 = 0$. A number of interesting phenomena occur in this case, at least in the non-trivial regime where $R_3 < 1$. To study this, we need the degraded broadcast channel capacity result.

The capacity region of a stochastically degraded broadcast channel is the same as that of its physically degraded counterpart (e.g., see [15, p.454, Ex. 10]). The input signal is $X$, and the output signals are $Y_1$ and $Y_2$. We assume that $Y_2$ is a physically degraded version of $Y_1$ (i.e., $X \rightarrow Y_1 \rightarrow Y_2$). For convenience, we refer to the terminal observing $Y_1$ as Relay 1, rather than Receiver 1. Similarly, we refer to the terminal observing $Y_2$ as Relay 2, rather than Receiver 2. Denote by $R_{\text{top}}$ the rate to Relay 1, which observes $Y_1$. Denote by $R_{\text{bot}}$ the rate to Relay 2, which observes $Y_2$. Then the capacity region, as shown by combining the results of Bergmans [8] and Gallager [22], is the convex hull of the closure of all rate pairs $(R_{\text{top}}, R_{\text{bot}})$ satisfying

$$R_{\text{bot}} \leq I(U; Y_2),$$
$$R_{\text{top}} \leq I(X; Y_1 | U),$$

for some joint distribution $p(u)p(x|u)p(y_1|x)p(y_2|y_1)$, where the auxiliary random variable $U$ has cardinality bounded by $|U| \leq \min\{|X|, |Y_1|, |Y_2|\}$. Additionally, Relay 1 can reliably decode $R_{\text{bot}}$ as well as $R_{\text{top}}$.

We will give a brief description of the coding structure which achieves such rate pairs. There are $2^{nR_{\text{bot}}}$ essentially non-intersecting groups of codewords. The distance between groups is sufficiently large to allow the poorer receiver, Relay 2, to distinguish between groups after observing the input signal through its noisy link. Each such group consists of a relatively tightly packed set of $2^{nR_{\text{top}}}$ actual codewords. The
3.2. CONVERSES FOR DISCRETE MEMORYLESS NETWORKS

The distance between codewords within each group is sufficiently large to allow the better receiver, Relay 1, to distinguish between codewords after observing the input signal through its noisy link. The overall structure is often described as cloud centers, corresponding to the groups, and fine points which define each cloud. The two information streams are used as follows. The first, $R_{\text{bot}}$, determines the cloud center. The second, $R_{\text{top}}$, determines the exact codeword within the cloud, and this is ultimately sent as the input signal $X^n$. In terms of the single-letter characterization, the auxiliary random variable $U$ is used to generate the cloud centers, and $X$ conditioned on $U$ is used to generate the fine points.

Now we turn to the multiaccess part of the parallel relay network. Again for convenience, we refer to the terminal with input $W_3$ as Relay 1, rather than User 1. Similarly, we refer to the terminal with input $W_4$ as Relay 2, rather than User 2. Since the multiaccess side consists of a pair of independent, noiseless binary links, the capacity region (with independent messages) is simply a rectangle. It equals the set of rate pairs $(R_{\text{top}}, R_{\text{bot}})$ such that $R_{\text{top}} \leq R_3$ and $R_{\text{bot}} \leq R_4$. These rates are given in terms of bits per input channel use.

We can easily extend the multiaccess capacity region to the case where Relay 1 knows the message of Relay 2, but not vice versa. In this case, the capacity region is the set of rate pairs $(R_{\text{top}}, R_{\text{bot}})$ such that $R_{\text{top}} \leq R_3$ and $(R_{\text{top}} + R_{\text{bot}}) \leq (R_3 + R_4)$. We refer to this as the extended multiaccess capacity region for the pair of noiseless binary links. To achieve these rate pairs, Relay 2 sends as much of $R_{\text{bot}}$ as it is able, up to $R_4$. Relay 1 then sends the remaining bits of $R_{\text{bot}}$, if there are any, as well as all bits of $R_{\text{top}}$. This form of cooperation differs significantly from the form we exploited in the Gaussian network. Here, rather than each relay sending a common signal, the relays partition $R_{\text{bot}}$ and send the independent parts.

3.2 Converse for Discrete Memoryless Networks

We derive converse results that are very similar in nature to those we derived for the Gaussian network. Of course, we cannot derive a converse based on second moment connections between signals, as we did when deriving the simultaneous bounds for the Gaussian network, (2.83) and (2.84). However, we can still derive the pure cut-set
bounds.

To derive the broadcast cut-set bound, we proceed in exactly the same way we did for the Gaussian network.

\[
\frac{1}{n} I(X^n ; W_3^n, W_4^n) \leq \frac{1}{n} I(X^n ; Y_1^n, Y_2^n)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} (H(Y_{1,i}, Y_{2,i} | Y_1^{i-1}, Y_2^{i-1})

- H(Y_{1,i}, Y_{2,i} | Y_1^{i-1}, Y_2^{i-1}, X^n))
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} (H(Y_{1,i}, Y_{2,i} | Y_1^{i-1}, Y_2^{i-1}) - H(Y_{1,i}, Y_{2,i} | X_i))
\]

\[
\leq \frac{1}{n} \sum_{i=1}^{n} (H(Y_{1,i}, Y_{2,i}) - H(Y_{1,i}, Y_{2,i} | X_i))
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} (I(X_i ; Y_{1,i}, Y_{2,i}))
\]

\[
\leq \max_{p(x)} I(X ; Y_1, Y_2).
\]

Inequality (3.4) follows from the data processing inequality. Equality (3.6) follows from the memorylessness of the broadcast channel. Inequality (3.7) follows because conditioning reduces entropy, and we have dropped conditioning variables from the first term. Equality (3.9) follows again from the memorylessness of the channel. We have abused notation a bit in (3.9), where we have written \(p(x)\) to indicate an arbitrary input distribution on the input symbol \(X\), and \(I(X; Y_1, Y_2)\) is computed with respect to the given broadcast channel \(p(y_1, y_2 | x)\). Inequality (3.9) holds for all input codes and all possible relay processing. Consequently,

\[
C_{\text{net}} \leq \max_{p(x)} I(X ; Y_1, Y_2).
\]
3.2. CONVERSES FOR DISCRETE MEMORYLESS NETWORKS

broadcast channel.

Before evaluating this bound for the PBSC network, we need to define some notation. We use $H_{\text{bin}}(x)$ to denote the binary entropy function evaluated at $x \in [0, 1]$. Recall that $H_{\text{bin}}(x) = H_{\text{bin}}(1 - x)$. By simply relabelling what we call a 1 or a 0, we can and will always restrict the argument of $H_{\text{bin}}(x)$ to $x \in [0, 0.5]$. Then $H_{\text{bin}}(x)$ is an increasing function of $x$ on this domain, $x \in [0, 0.5]$. For the cascade of two BSC’s with crossover probabilities $\alpha_1, \alpha_2 \in [0, 0.5]$, we define $\alpha_1 \otimes \alpha_2$ as the net crossover probability. That is,

$$\alpha_1 \otimes \alpha_2 = \alpha_1(1 - \alpha_2) + (1 - \alpha_1)\alpha_2.$$  \hfill (3.11)

Note that when $\alpha_1, \alpha_2 \leq 0.5$, this implies $\max[\alpha_1; \alpha_2] \leq \alpha_1 \otimes \alpha_2 \leq 0.5$, i.e., the cascade of two BSC’s only gets noisier. Also note that

$$\alpha \otimes \alpha = 2\alpha(1 - \alpha).$$  \hfill (3.12)

We will always assume, without loss of generality, that $\alpha_1, \alpha_2 \in [0, 0.5]$.

For the PBSC network of Figure 3-1, the general broadcast cut-set bound, (3.10), states

$$C_{\text{net}} \leq 1 + H_{\text{bin}}(\alpha_1 \otimes \alpha_2) - H_{\text{bin}}(\alpha_1) - H_{\text{bin}}(\alpha_2).$$  \hfill (3.13)

We refer to (3.13) as the broadcast cut-set bound for the general PBSC network. For any $\alpha_1, \alpha_2 \in [0, 0.5],

$$1 + H_{\text{bin}}(\alpha_1 \otimes \alpha_2) - H_{\text{bin}}(\alpha_1) - H_{\text{bin}}(\alpha_2) \geq \max[1 - H_{\text{bin}}(\alpha_1); 1 - H_{\text{bin}}(\alpha_2)],$$  \hfill (3.14)

with equality only when at least one of the links is degenerate ($\alpha_1$ or $\alpha_2$ equals 0 or 0.5). Therefore, when there is no degeneracy, the broadcast cut-set bound is strictly larger than the capacity of either broadcast link. When one of the links is noiseless, e.g., when $\alpha_1 = 0$, the broadcast cut-set bound (3.10) states that $C_{\text{net}} \leq 1$, as we expect. For the PBSC network with a symmetric broadcast channel, where
\( \alpha_1 = \alpha_2 = \alpha \), the broadcast cut-set bound, (3.10), states
\[
C_{\text{net}} \leq 1 + H_{\text{bin}} (2\alpha (1 - \alpha)) - 2H_{\text{bin}} (\alpha).
\] (3.15)

We refer to (3.15) as the broadcast cut-set bound for the symmetric PBSC network. We turn next to the multiaccess cut-set bound.

\[
\frac{1}{n} I(X^n ; W_3^{nR_3}, W_4^{nR_4}) = \frac{1}{n} H(W_3^{nR_3}, W_4^{nR_4}) - \frac{1}{n} H(W_3^{nR_3}, W_4^{nR_4} | X^n)
\] (3.16)
\[
\leq \frac{1}{n} H(W_3^{nR_3}, W_4^{nR_4})
\] (3.17)
\[
\leq R_3 + R_4.
\] (3.18)

Inequality (3.17) follows because \( H(W_3^{nR_3}, W_4^{nR_4} | X^n) \geq 0 \), and (3.18) follows because \( W_3 \) and \( W_4 \) are binary signals. Consequently,
\[
C_{\text{net}} \leq R_3 + R_4.
\] (3.19)

We refer to (3.19) as the multiaccess cut-set bound.

Finally, we derive the two pure cross-cut bounds. Define \( C_1 \) as the capacity of the link from sender to Relay 1. Similarly, define \( C_2 \) as the capacity of the link from sender to Relay 2. In other words,
\[
C_1 = \max_{p(X)} I(X; Y_1).
\] (3.20)
\[
C_2 = \max_{p(X)} I(X; Y_2).
\] (3.21)

Proceeding,
\[
\frac{1}{n} I(X^n ; W_3^{nR_3}, W_4^{nR_4}) = \frac{1}{n} \left( I(X^n ; W_4^{nR_4}) + I(X^n ; W_3^{nR_3} | W_4^{nR_4}) \right)
\] (3.22)
\[
\leq \frac{1}{n} \left( I(X^n ; Y_2^n) + I(X^n ; W_3^{nR_3} | W_4^{nR_4}) \right)
\] (3.23)
\[
\leq C_2 + \frac{1}{n} I(X^n ; W_3^{nR_3} | W_4^{nR_4})
\] (3.24)
\[ C_2 + \frac{1}{n} \left( H(W_3^n | W_4^n) - H(W_3^n | W_4^n, X^n) \right) \]
\[ \leq C_2 + \frac{1}{n} \left( H(W_3^n) \right) \]
\[ \leq C_2 + R_3. \] (3.25)

Inequality (3.23) follows from the data processing inequality since \( X^n \to Y_2^n \to W_4^n \), (3.24) follows by definition of \( C_2 \) and memorylessness of the broadcast channel, (3.26) follows since conditioning reduces entropy and since conditional entropy is non-negative, and (3.27) follows because \( W_3 \) is a binary signal. Consequently,

\[ C_{\text{net}} \leq C_2 + R_3. \] (3.28)

We refer to (3.28) as the first cross-cut bound. We derive the second cross-cut bound analogously, resulting in

\[ C_{\text{net}} \leq C_1 + R_4. \] (3.29)

For symmetric networks with noiseless binary links from the relays to the decoder, where \( C_1 = C_2 \) and \( R_3 = R_4 \), these two cross-cut bounds are identical. Furthermore, when the broadcast observations are conditionally independent given the input, which is often a natural assumption, then the two cross-cut bounds are superfluous given the broadcast and multiaccess cut-set bounds, (3.10) and (3.19). We demonstrate this by considering the two possible cases. For the first case, when \( C_1 \geq R_3 \), then \( C_1 + R_3 \geq 2R_3 \). Therefore in this case, the multiaccess cut-set bound, (3.19), is at least as small as the cross-cut bound, (3.29). For the second case, when \( C_1 < R_3 \), we begin with

\[ I(X; Y_1, Y_2) \leq I(X; Y_1) + I(X; Y_2) \] (3.30)

Inequality (3.30) holds because the relay observations are conditionally independent given the input (by assumption). Inequality (3.30) does not hold for all joint distributions \( p(x, y_1, y_2) \). As a counterexample with conditionally dependent observations,
consider $X$ equally likely 0 or 1. Let $Z$ be uniformly distributed on the integers from 1 to $N$. Let $Y_1 = X + Z$ and let $Y_2 = X - Z$. Then $I(X; Y_1, Y_2) = H(X) = 1$ but $I(X; Y_1) = I(X; Y_2) = \frac{1}{N}$. Continuing from (3.30) for the second case, when $C_1 < R_3$,

$$I(X; Y_1, Y_2) \leq I(X; Y_1) + I(X; Y_2)$$  
$$\leq 2C_1$$  
$$< C_1 + R_3.$$  

Therefore given symmetry and conditional independence of the relay observations, the minimum of the broadcast cut-set bound, (3.10), and the multiaccess cut-set bound, (3.19), is at least as small as the cross-cut bound, (3.29). On the other hand, for asymmetric networks with conditionally independent relay observations and noiseless binary links from the relays to the decoder, the cross-cut bounds are not necessarily superfluous.

![Diagram](attachment:Figure_3-2.png)

**Figure 3-2**: Evaluation of converse results for typical symmetric and asymmetric PBSC networks

We have plotted the two relevant converse results for a representative symmetric
3.3. **CODING FOR DISCRETE SYMMETRIC NETWORKS**

PBSC network on the left in Figure 3-2. We have fixed the broadcast channel such that $C_1 = C_2 = 0.5$. This corresponds to $\alpha_1 = \alpha_2 \approx 0.11$ for the two BSC’s. We increased the noiseless binary link rates, $R_3 = R_4$, from 0 to 1. The binary link rates correspond to the horizontal axis. The vertical axis corresponds to network communication rate.

We have plotted the three relevant converse results for a representative highly asymmetric PBSC network on the right in Figure 3-2. As usual, we have assumed for simplicity that $R_3 = R_4$. We have fixed the broadcast channel such that $C_1 = 1$, $C_2 = 0.5$. This corresponds to $\alpha_1 = 0$, $\alpha_2 \approx 0.11$ for the two BSC’s. In this case, the broadcast cut-set bound (3.10) states

$$C_{\text{net}} \leq 1. \quad (3.34)$$

The broadcast cut-set upper bound equals one purely because $X \equiv Y_1$ is binary. The relevant cross-cut bound (3.28), which states

$$C_{\text{net}} \leq (R_3 + C_2), \quad (3.35)$$

corresponds to the dashed line in the figure. The multiaccess cut-set bound has slope two, while the cross-cut bound has slope one. In light of the broadcast and multiaccess cut-set bounds of (3.10) and (3.19), the cross-cut bound of (3.28) is relevant for the highly asymmetric PBSC network if and only if $C_2 < 0.5$.

### 3.3 Coding for Discrete Symmetric Networks

We develop achievable coding techniques which are similar to those derived for the Gaussian network. In particular, we first develop the technique of reliably coding to the relays. We then develop the technique of forwarding the relay observations. This latter approach is similar in title but markedly different in character for discrete networks with noiseless binary relay links. The difference arises from the degeneracy of the multiaccess side of the network, not from the discreteness of the alphabets.

We first consider the symmetric PBSC network of Figure 3-1, where $\alpha_1 = \alpha_2 = \alpha$ and $R_3 = R_4$. The assumption that $\alpha_1 = \alpha_2$ greatly simplifies the analysis of
coding to the relays, but it is unimportant in the analysis of forwarding the relay observations. After considering the symmetric PBSC network, we then turn to the highly asymmetric PBSC network, where \( \alpha_1 = 0 \) but \( \alpha_2 \in (0,0.5) \). We will again apply the same approaches. For this highly asymmetric network, a new perspective on forwarding the relay observations arises because Relay 1 observes the input codeword without corruption. Rather than reproducing both observations at the receiver, we need only reproduce \( Y_1 \equiv X \). We will use Relay 2, with its correlated observation \( Y_2 \), to help the decoder resolve \( Y_1 \).

3.3.1 Coding to the Relays

For networks with statistically identical relay observations and noiseless binary relay links, the coding approach is simple. Since the relays have statistically identical observations, the same bits are decoded at each relay. The capacity of a link from the input to a single relay is \( C_1 \), and thus we can get up to \( C_1 \) bits per input symbol to each relay. The relays decode the same bits, and therefore they can partition them and send independent sets of bits to the decoder. Therefore, in total, the two relays can send up to \( R_3 + R_4 \) decoded bits per input symbol to the decoder. We can thus achieve

\[
R_{\text{ach}} = \min[C_1; R_3 + R_4].
\]

When \( C_1 \geq R_3 + R_4 \), this meets the multiaccess cut-set bound, (3.18), and thereby defines the capacity of the network. Thus, when the broadcast side is sufficiently strong, reliably coding to the relays can achieve network capacity. We found an analogous result when studying the Gaussian network (Section 2.3.1). Evaluating this scheme for the symmetric PBSC network, we can achieve

\[
R_{\text{ach}} = \min[1 - H_{\text{bin}}(\alpha); 2R_3].
\]
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

3.3.2 Binning to Communicate Observations Without Decoding at the Relays

For discrete networks, since we have noiseless binary links from the relays to the receiver, there is neither interference nor an opportunity to “statistically cooperate” on the multiaccess channel, as we did with both Gaussian achievability schemes. Since coding to the relays achieves network capacity whenever $C_{\text{net}} \leq C_1$, we only need worry about communicating at rates $R_{\text{ach}} > C_1$ from here on.

Our intuition when $R_{\text{ach}} > C_1$ is that the relays have no hope of decoding or even reasonably estimating the source codewords on their own. We therefore explore coding schemes where the relays attempt to get a reasonable representation of their observations to the decoder. Because the broadcast observation alphabets are finite (binary in the PBSC network), given large enough noiseless relay link rates, the decoder can recover an exact representation of the observations. We can achieve this reproduction at the decoder by having the relays encode blocks of observation symbols. Thus with a strong enough multiaccess side (i.e., with $R_3$ and $R_4$ large enough), we can achieve the broadcast cut-set bound (3.10), $\max_{p(x)} I(X; Y_1, Y_2)$. We do not require an asymptotically strong pair of noiseless relay links — we can achieve the broadcast cut-set bound with finite values of $R_3$ and $R_4$. An analogous result does not appear to hold for the Gaussian network.

We develop random coding arguments based on an i.i.d. input distribution $p(x)$, usually chosen to be equally likely 0 or 1. An i.i.d. input distribution induces an artificial i.i.d. “observation source”, $(Y_1, Y_2)$, distributed as $p(y_1, y_2) = \sum_x p(x)p(y_1, y_2 \mid x)$. Basically, we will design coding schemes where the decoder can accurately recover the artificial “observation source”, $(Y_1^n, Y_2^n)$, and then decode the input codeword based on the reconstructed observations.

We will use block codes throughout for the random coding arguments. The arguments, in essence, require two steps. In the first step, the decoder reproduces the pair of relay observations, $(Y_1^n, Y_2^n)$. When averaged over the input codeword ensemble, the relay observations are in some sense equivalent to the induced “observation source”. The intuition for this step is based on distributed source coding. Essentially, the first step is achievable when the noiseless link rates are large enough to enable the decoder to reproduce the full “observation source” (or, at least, to the designed
CHAPTER 3. DISCRETE MEMORYLESS PARALLEL RELAY NETWORKS

distortion with high probability). In the second step of the argument, the decoder recovers the input codeword, \(X^n\), based on the reproduced observations, \((Y_1^n, Y_2^n)\). The intuition for this step is based on traditional, point to point channel coding. Essentially, the second step is achievable with a channel code that is good for a channel with input \(X\) and output \((Y_1, Y_2)\).

**Essentially Perfect Representation of \((Y_1, Y_2)\)**

If the decoder had access to the relay observations \(Y_1\) and \(Y_2\), then we would effectively be trying to communicate over a point to point channel with input \(X\) and output \((Y_1, Y_2)\). If this were the case, then Shannon’s basic channel coding theorem proves we could achieve rates up to

\[
R_{ach} = I(X; Y_1, Y_2).
\]  

We explore how to communicate the relay observations to the decoder. First, we could opt to compress \(Y_1^n\) at Relay 1 and independently compress \(Y_2^n\) at Relay 2. We can describe \(Y_1^n\) at Relay 1 noiselessly if \(nR_3 > H(Y_1^n)\), and \(Y_2^n\) at Relay 2 noiselessly if \(nR_4 > H(Y_2^n)\), e.g., by using Huffman codes. In this case, the decoder can in fact recover all possible \((Y_1^n, Y_2^n)\) pairs without error. However, we can do better by using the Slepian-Wolf distributed compression (binning) scheme to describe only the jointly typical \((Y_1^n, Y_2^n)\) pairs. Here we do not describe all possible pairs \((Y_1^n, Y_2^n)\) — rather, we describe essentially all of the most likely pairs. From the Slepian-Wolf theorem on distributed source compression, this requires

\[
R_3 > H(Y_1 | Y_2), \\
R_4 > H(Y_2 | Y_1), \\
(R_3 + R_4) > H(Y_1, Y_2).
\]  

If (3.39) is satisfied then we can reproduce essentially all of the relevant \((Y_1^n, Y_2^n)\) pairs. We can then achieve rates up to

\[
R_{ach} = I(X; Y_1, Y_2).
\]
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

We delay the proof of this achievability result until the next section, where we will prove a slightly more general result. We proceed with evaluating the result.

For the PBSC network with a symmetric broadcast channel, choosing \( p(x) = (0.5, 0.5) \) to maximize \( I(X; Y_1, Y_2) \), we will claim in Theorem 3.3.1 that we can achieve

\[
R_{\text{ach}} = 1 + H_{\text{bin}} (\alpha \otimes \alpha) - 2H_{\text{bin}} (\alpha)
\]

provided the binary link rates satisfy (3.39). In symmetric networks, where \( R_3 = R_4 \), the third inequality of (3.39) dominates. Indeed, assuming the third inequality is satisfied,

\[
2R_3 = R_3 + R_4
\]

\[
> H(Y_1, Y_2)
\]

\[
= H(Y_1) + H(Y_2 | Y_1)
\]

\[
\geq H(Y_1 \mid Y_2) + H(Y_2 \mid Y_1)
\]

\[
= 2H(Y_1 \mid Y_2),
\]

where the last equality follows from the symmetry of the broadcast channel. For this i.i.d. input \( p(x) = (0.5, 0.5) \), then, we require

\[
R_3 > 0.5H(Y_1, Y_2)
\]

\[
= 0.5 \left( 1 + H_{\text{bin}} (\alpha \otimes \alpha) \right).
\]

For the symmetric PBSC network, we present the two achievability schemes graphically at fixed \( C_1 = C_2 = 0.5 \) as we increase the rate \( R_3 = R_4 \) from 0 to 1. This is done in Figure 3-3. We also plot the converse results for comparison. Note that there is no need to increase \( R_3 \) above \( 0.5H(Y_1, Y_2) = 0.5(1 + H_{\text{bin}} (\alpha \otimes \alpha)) \approx 0.857 \) since we can achieve the broadcast cut-set bound with that noiseless link rate. The dashed line in the figure corresponds to time-sharing between the two techniques, which convexifies the achievable curve.
Figure 3-3: Evaluation of the two basic achievability schemes for the symmetric PBSC network
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

Essentially Perfect Representation of \((Y_1, Y_2)\); Backing Off from \(I(X; Y_1, Y_2)\) by Throwing Away Input Codewords

If we do not quite meet the \(R_3\) and \(R_4\) link rate requirement of (3.39) for transmitting the artificial “observation source”, we can back \(R_{ach}\) off of the limit \(I(X; Y_1, Y_2)\) to loosen the requirement. The idea is that a typical codeword \(X^n\) has a potential “image” in \((Y_1^n, Y_2^n)\) of size \(2^{nH(Y_1, Y_2)}\) strings, while the induced observation source has an overall image of size \(2^{nH(Y_1, Y_2)}\) strings. For a good channel code, the images of different input codewords are essentially non-overlapping when \(R_{ach} \leq I(X; Y_1, Y_2)\). If we reduce the number of codewords used, then we need accurately describe fewer observation source strings, thereby reducing the rate requirement on \(R_3\) and \(R_4\).

Choose a single-letter distribution on the input, \(p(x)\). If we let

\[
R_{ach} = I(X; Y_1, Y_2) - \delta, \tag{3.49}
\]

we prove in Theorem 3.3.1 that we can linearly relax the rate constraints to

\[
R_3 > H(Y_1 \mid Y_2) - \delta, \tag{3.50}
\]

\[
R_4 > H(Y_2 \mid Y_1) - \delta, \tag{3.51}
\]

\[
(R_3 + R_4) > H(Y_1, Y_2) - \delta. \tag{3.52}
\]

This theorem holds for any discrete broadcast channel, not just those with conditionally independent observations, and for any pair of noiseless relay links, not necessarily of equal rate. Before stating the theorem, we evaluate the result for the symmetric PBSC network. Again, in a symmetric network with conditionally independent relay observations, the final inequality dominates (c.f. (3.46)), yielding the requirement

\[
R_3 = R_4 > 0.5H(Y_1, Y_2) - 0.5\delta. \tag{3.53}
\]

At the boundary, from (3.49) and (3.53), \(R_{ach}\) and \(R_3\) are related by

\[
R_{ach} = 2R_3 - 2H(Y_1 \mid X) \tag{3.54}
\]

\[
= 2(R_3 - H_{bin}(\alpha)) \tag{3.55}
\]
for any $R_3$ satisfying

$$H_{\text{bin}}(\alpha) \leq R_3 \leq 0.5(1 + H_{\text{bin}}(2\alpha(1 - \alpha))).$$

(3.56)

Figure 3-4: Evaluation of the backoff scheme for the symmetric PBSC network

We plot the performance of this backoff scheme in Figure 3-4. Unfortunately, for the symmetric PBSC network, the performance of this scheme for any $\delta > 0$ is inferior to time-sharing between the pure coding scheme and the full Slepian-Wolf / binning scheme (corresponding to $\delta = 0$). In fact, this backoff scheme can-
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

not perform better than time-sharing for any symmetric network with 
\( p(y_1, y_2 | x) = p(y_1 | x)p(y_2 | x) \) and \( R_3 = R_4 \). For this backoff scheme, from (3.54), the achievable curve approaches the point \( (R_3, R_{ach}) = (0.5H(Y_1, Y_2), I(X; Y_1, Y_2)) \) linearly with slope 2. On the other hand, the time-sharing curve approaches this same point, \( (R_3, R_{ach}) = (0.5H(Y_1, Y_2), I(X; Y_1, Y_2)) \), linearly with slope

\[
\frac{I(X; Y_1, Y_2) - I(X; Y_1)}{0.5(H(Y_1, Y_2) - I(X; Y_1))} \leq 2. \tag{3.57}
\]

Equality is achieved in (3.57) if and only if \( H(Y_1, Y_2 | X) = 0 \), i.e., if the broadcast channel outputs are deterministic functions of those input symbols for which \( p(x) \) has positive probability.

Observe that when the broadcast channel is deterministic, and when \( R_3 = R_4 \), then using the broadcast and multiaccess cut-set bounds, (3.10) and (3.19),

\[
C_{\text{net}} \leq \min \{2R_3; \sup_{p(x)} H(Y_1, Y_2) \}. \tag{3.58}
\]

This upper bound can be achieved by time-sharing between the origin and the point achieved with the basic approach of binning to communicate the observations. It does not appear worthwhile to further generalize this result for \( R_3 \neq R_4 \), for it does not seem relevant to the fundamental problems studied in this thesis.

![Parallel relay network with noiseless relay channels](image)

Figure 3-5: Parallel relay network with noiseless relay channels

We now prove the achievability theorem that leads to the results already developed in this section. The theorem may also be of independent interest. We first define the notion of joint (weak) typicality.
Definition 1. [15, pp. 384–5] Let \((X_1, X_2, \ldots, X_k)\) denote a finite collection of discrete random variables with some fixed joint distribution, \(p(x_1, x_2, \ldots, x_k)\), \((x_1, x_2, \ldots, x_k) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k\). The set of jointly \(\epsilon\)-typical sequences of blocklength \(n\), denoted \(\mathcal{A}_\epsilon^n(X_1, X_2, \ldots, X_k)\) or \(\mathcal{A}_\epsilon^n(X_1^n, X_2^n, \ldots, X_k^n)\), is the set of \(n\)-sequences \((x_1^n, x_2^n, \ldots, x_k^n)\) such that

\[
\left| -\frac{1}{n} \log p(s^n) - H(S) \right| < \epsilon, \quad \forall S \subseteq \{X_1, X_2, \ldots, X_k\}. \tag{3.59}
\]

In (3.59), \(s^n\) is the restriction of \((x_1^n, x_2^n, \ldots, x_k^n)\) to the subset of indices \(S\).

Before proceeding, we would like to make a few observations about this definition. For simplicity, consider just a pair of random variables \((X, Y)\). Assume that we are given a joint distribution \(p(x, y)\), and that we generate a sequence of pairs \((X^n, Y^n)\) i.i.d. \(\sim p(x, y)\). The set of jointly typical sequences is denoted \(\mathcal{A}_\epsilon^n(X, Y)\). Then for a particular pair of sequences \((x^n, y^n)\), from (3.59), \((x^n, y^n) \in \mathcal{A}_\epsilon^n(X, Y)\) if and only if

\[
\left| -\frac{1}{n} \log p(x^n) - H(X) \right| < \epsilon, \tag{3.60}
\]

\[
\left| -\frac{1}{n} \log p(y^n) - H(Y) \right| < \epsilon, \tag{3.61}
\]

\[
\left| -\frac{1}{n} \log p(x^n, y^n) - H(X, Y) \right| < \epsilon. \tag{3.62}
\]

Thus from (3.60), \(x^n\) looks typical. From (3.61), \(y^n\) looks typical. From (3.62), \((x^n, y^n)\) together look typical. Another way of thinking about the set \(\mathcal{A}_\epsilon^n(X, Y)\) is to think in terms of conditional typicality. In particular, \((x^n, y^n) \in \mathcal{A}_\epsilon^n(X, Y)\) means that \(x^n\) looks typical and that \(y^n\) looks conditionally typical conditioned on \(X^n = x^n\). Specifically, from (3.60),

\[
\left| -\frac{1}{n} \log p(x^n) - H(X) \right| < \epsilon. \tag{3.63}
\]
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

From (3.60), (3.62), the expansion \( H(X, Y) = H(X) + H(Y \mid X) \), and Bayes’ rule,

\[
-\frac{1}{n} \log p(y^n \mid x^n) - H(Y \mid X) < 2\epsilon. \tag{3.64}
\]

Notice that in (3.64), the measure of typicality has to be changed from \( \epsilon \) to \( 2\epsilon \). This is essentially unimportant but is a consequence of the mathematical definition of joint typicality. We proceed with the theorem of interest.

**Theorem 3.3.1.** We are given the parallel relay network of Figure 3-5, with discrete memoryless broadcast channel \( p(y_1, y_2 \mid x) \) and noiseless binary link rates \( R_3 \) and \( R_4 \). The relay observations need not be conditionally independent given the input. Choose any single-letter input distribution \( p(x) \) and any \( \delta \in [0, I(X; Y_1, Y_2)] \). We can reliably achieve rate \( R_{ach} \) provided

\[
R_{ach} < I(X; Y_1, Y_2) - \delta, \tag{3.65}
\]

\[
R_3 > H(Y_1 \mid Y_2^T) - \delta, \tag{3.66}
\]

\[
R_4 > H(Y_2 \mid Y_1) - \delta, \tag{3.67}
\]

\[
(R_3 + R_4) > H(Y_1, Y_2) - \delta. \tag{3.68}
\]

**Proof.** Assume the conditions of the theorem are satisfied. For now, we choose the typicality measure, \( \epsilon \), and the integer blocklength, \( n \), arbitrarily. We will later set \( \epsilon \) sufficiently small and \( n \) sufficiently large to make the probability of message error as small as we desire. We will ignore the integrality constraints by assuming \( 2^{nR_{ach}} \), \( 2^{nR_3} \), and \( 2^{nR_4} \) are integer valued for every integer \( n \) — the integrality constraints can be handled trivially at the expense of notational inconvenience.

Randomly generate \( 2^{nR_{ach}} \) input codewords of blocklength \( n \). Generate each symbol of each codeword independently according to \( p(x) \). Denote the input codewords by \( X^n(m), m = 1, 2, \ldots, 2^{nR_{ach}} \). Denote the input codebook by \( \mathcal{C} = \{\cup_{m=1}^{2^{nR_{ach}}} X^n(m)\} \). At the decoder, we will use (weakly) typical set decoding for the distribution \( p(x, y_1, y_2) = p(x)p(y_1, y_2 \mid x) \). Denote the set of jointly \( \epsilon \)-typical triples \( (x^n, y_1^n, y_2^n) \) by \( \mathcal{A}_\epsilon^n \), or equivalently, by \( \mathcal{A}_\epsilon^n(X^n, Y_1^n, Y_2^n) \). Similarly, denote the set of pairs \( (x^n, y_1^n) \) that are jointly \( \epsilon \)-typical with a particular Relay 2 observation, \( Y_2^n = y_2^n \), by \( \mathcal{A}_\epsilon^n(X^n, Y_1^n \mid y_2^n) \).

Randomly and uniformly assign every possible Relay 1 observation, \( y_1^n \in \mathcal{Y}_1^n \),
to one of $2^{nR_3}$ bins. Denote this randomly chosen bin assignment by the function $f_1(y_1^n)$. Similarly, randomly and uniformly assign every possible Relay 2 observation, $y_2^n \in \mathcal{Y}_2^n$, to one of $2^{nR_4}$ bins. Denote this randomly chosen bin assignment by the function $f_2(y_2^n)$.

Reveal the input codebook, $\mathcal{C}$, and the bin assignments, $f_1(\cdot)$ and $f_2(\cdot)$, to the decoder.

**Encoding:** The source transmits the input codeword $X^n(m)$ corresponding to message $m$. After receiving $Y_1^n$, Relay 1 sends the bin index $f_1(Y_1^n)$ to the receiver. Simultaneously, after receiving $Y_2^n$, Relay 2 sends the bin index $f_2(Y_2^n)$ to the receiver.

**Decoding:** The decoder receives the pair of bin numbers $f_1(Y_1^n)$ and $f_2(Y_2^n)$. The decoder declares message $m$ was sent if it is the unique message such that there is a pair of relay observations, each assigned to the given bin number, (weakly) jointly $\epsilon$-typical with the input codeword. I.e., the decoder declares message $m$ was sent if it is the unique message such that $X^n(m) \in \mathcal{C}$, and there is a pair $(y_1^n, y_2^n) \in \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ satisfying $f_1(y_1^n) = f_1(Y_1^n)$, $f_2(y_2^n) = f_2(Y_2^n)$, and $(X^n(m), y_1^n, y_2^n) \in \mathcal{A}_n$. Otherwise, the decoder declares an error.

**Average Probability of Decoding Error:** We compute the average probability of message error by averaging over the choice of input codebooks and relay observation bin assignments. We choose the codewords for the input codebook by randomly generating them independently from each other. Furthermore, we use the same probability distribution to choose each input codeword. Therefore, from the symmetry of these random choices, the average probability of message error for the randomly chosen input codebook equals the average probability of message error for the randomly chosen input codebook conditioned on the source transmitting the first input codeword. We will therefore upper bound the average probability of message error conditioned on the source transmitting the first input codeword. We denote this by $\mathbb{E}\{\Pr_{\text{error}}\}$.

The input codeword corresponding to the first message is $X^n(1)$. During operation, the relays receive the pair of observations $(Y_1^n, Y_2^n)$. For a particular triple of strings $(x^n, y_1^n, y_2^n)$, the probability that $X^n(1) = x^n$ and $(Y_1^n, Y_2^n) = (y_1^n, y_2^n)$ is denoted by $p_{X^n,Y_1^n,Y_2^n}(x^n, y_1^n, y_2^n)$. Before proceeding, recall our notation for the input codebook, $\mathcal{C} = \{\cup_{m=1}^{\mathcal{P}_{\text{ach}}} X^n(m)\}$. When computing the average probability of mes-
3.3. **Coding for Discrete Symmetric Networks**

sage error, we will consider the input codebook without the first codeword. Define \( \mathcal{C}_{-1} = \{ \cup_{m=2}^{n} X^n (m) \} \). Then \( \mathcal{C}_{-1} \subseteq \mathcal{C} \) with equality if and only if the first codeword, \( X^n(1) \), happens to be identical to the codeword for another message.

We group all error events into the union of five events. They are:

\[
E_0 = \{ (X^n(1), Y_1^n, Y_2^n) \notin \mathcal{A}_e^n \} ;
\]

(3.69)

\( E_0 \) is the event that the codeword and actual relay observations are not jointly typical.

\[
E_1 = \left\{ (X^n(1), Y_1^n, Y_2^n) \in \mathcal{A}_e^n, \ \exists x' \in \mathcal{C}_{-1} \text{ such that:} \right. \\
\left. (x', Y_1^n, Y_2^n) \in \mathcal{A}_e^n. \right\} ;
\]

(3.70)

\( E_1 \) is the event that \( (X^n(1), Y_1^n, Y_2^n) \) is jointly typical, and there is an incorrect codeword, \( x' \), such that the triple \( (x', Y_1^n, Y_2^n) \) is jointly typical.

\[
E_2 = \left\{ (X^n(1), Y_1^n, Y_2^n) \in \mathcal{A}_e^n, \ \exists x' \in \mathcal{C}_{-1} \text{ and } y'_1 \neq Y_1^n \text{ such that:} \right. \\
\left. f_1(y'_1) = f_1(Y_1^n) \text{ and } (x', y'_1, Y_2^n) \in \mathcal{A}_e^n. \right\} ;
\]

(3.71)

\( E_2 \) is the event that \( (X^n(1), Y_1^n, Y_2^n) \) is jointly typical, and there is an incorrect codeword, \( x' \), and a different Relay 1 observation, \( y'_1 \), assigned to the same bin as the actual Relay 1 observation, such that the triple \( (x', y'_1, Y_2^n) \) is jointly typical.

\[
E_3 = \left\{ (X^n(1), Y_1^n, Y_2^n) \in \mathcal{A}_e^n, \ \exists x' \in \mathcal{C}_{-1} \text{ and } y'_2 \neq Y_2^n \text{ such that:} \right. \\
\left. f_2(y'_2) = f_2(Y_2^n) \text{ and } (x', Y_1^n, y'_2) \in \mathcal{A}_e^n. \right\} ;
\]

(3.72)

\( E_3 \) is the event that \( (X^n(1), Y_1^n, Y_2^n) \) is jointly typical, and there is an incorrect codeword, \( x' \), and a different Relay 2 observation, \( y'_2 \), assigned to the same bin as the actual Relay 2 observation, such that the triple \( (x', Y_1^n, y'_2) \) is jointly typical.
\[ E_4 = \left\{ (X^n(1), Y_1^n, Y_2^n) \in \mathcal{A}_x^n, \exists x' \in C_{-1}, y'_1 \neq Y_1^n, \text{ and } y'_2 \neq Y_2^n \text{ such that:} \\
\begin{align*}
&f_1(y'_1) = f_1(Y_1^n), f_2(y'_2) = f_2(Y_2^n), \text{ and } (x', y'_1, y'_2) \in \mathcal{A}_x^n.
\end{align*}
\right\}; \]

(3.73)

\( E_4 \) is the event that \((X^n(1), Y_1^n, Y_2^n)\) is jointly typical, and there is an incorrect codeword, \( x' \), and two different relay observations, \( y'_1 \) and \( y'_2 \), each assigned to the same bins as the actual relay observations, such that the triple is jointly typical.

Using basic set theory and the union bound,

\[ \mathbb{E}\{\Pr_{\text{error}}\} = \Pr \left( \bigcup_{i=0}^{4} E_i \right) \leq \sum_{i=0}^{4} \Pr \left( E_i \right). \]

(3.74)

We will show that, for sufficiently small \( \epsilon \), each of these five probabilities can be made arbitrarily small by increasing the blocklength \( n \). The same measure of typicality, \( \epsilon \), will be used consistently in the argument that follows.

Note that ensuring error events \( E_0 \) and \( E_1 \) have vanishing probability is equivalent to proving the existence of a good point to point channel code for a channel with input \( X \) and output \( (Y_1, Y_2) \). This requires only \( R_{ach} < I(X; Y_1, Y_2) \). We provide the proof for error events \( E_0 \) and \( E_1 \) for completeness.

It follows immediately from the weak law of large numbers that \( \Pr(E_0) \to 0 \) as \( n \to \infty \) [15, Th. 14.2.1].

Next we bound \( \Pr(E_1) \). For a particular triple of strings \((x^n, y_1^n, y_2^n)\), the probabili-
Inequality (3.76) follows from the union bound. Inequality (3.77) would be an equality if we had not dropped the minus one from the middle term.

We next upper bound the inner summation. For every \( x' \) such that \( (x', y_1^n, y_2^n) \in \mathcal{A}_\epsilon^n \), we know from the definition of (weakly) typical sequences, (3.59), that

\[
\Pr \{ X^n(2) = x' \} \leq 2^{-n(H(X) - \epsilon)}. \tag{3.78}
\]

Additionally, by definition, the number of \( x' \) sequences such that \( (x', y_1^n, y_2^n) \in \mathcal{A}_\epsilon^n \) is denoted \( |\mathcal{A}_\epsilon^n (X^n \mid y_1^n, y_2^n)| \). For sufficiently large blocklength \( n \), and for \( (y_1^n, y_2^n) \in \mathcal{A}_\epsilon^n (Y_1, Y_2) \), it follows as a straightforward consequence of the definition of weak typicality that [15, Th. 14.2.2]

\[
|\mathcal{A}_\epsilon^n (X^n \mid y_1^n, y_2^n)| \leq 2^{n(H(X)Y_1Y_2) + 2\epsilon}. \tag{3.79}
\]

Therefore, combining (3.78) and (3.79),

\[
\sum_{x' \text{ s.t.} \quad (x', y_1^n, y_2^n) \in \mathcal{A}_\epsilon^n} \Pr \{ X^n(2) = x' \} \leq 2^{-n(I(X; Y_1, Y_2) - 3\epsilon)}. \tag{3.80}
\]

Substituting (3.80) into (3.77), we have that for sufficiently large \( n \),

\[
\Pr (E_1) \leq 2^{-n(I(X; Y_1, Y_2) - R_{ach} - 3\epsilon)}. \tag{3.81}
\]
Consequently, from (3.81), \( \Pr(E_1) \to 0 \) as \( n \to \infty \) provided \( R_{\text{ach}} < I(X;Y_1,Y_2) - 3\epsilon \). From (3.65), this is indeed satisfied when \( \epsilon \) is small enough.

To prove that the probability of error events \( E_2, E_3, \) and \( E_4 \) vanish asymptotically with \( n \), we integrate the above proof with Cover and Thomas’ achievability proof of the Slepian-Wolf theorem [15, Th. 14.4.1]. The feature that differentiates this problem from the more familiar distributed source coding problem of Slepian and Wolf is that in this context, the “observation source” outcomes are generated by sending one of \( 2^{nR_{\text{ach}}} \) input codewords \( \{X^n(i)\} \) into the broadcast channel. We will see that, relative to the results of Slepian and Wolf for reproducing the full “observation source,” we can relax the constraints on \( R_3 \) and \( R_4 \) when \( R_{\text{ach}} < I(X;Y_1,Y_2) \). We proceed with bounding the probability of error event \( E_2 \).

\[
\Pr(E_2) = \sum_{(x^n,y_1^n,y_2^n) \in \mathcal{A}_n^g} p_{X^n,Y_1^n,Y_2^n}(x^n,y_1^n,y_2^n) \\
\cdot \Pr\{\exists x' \in \mathcal{C}_1, y'_1 \neq y_1^n \text{ s.t. } f_1(y'_1) = f_1(y_1^n), (x',y'_1,y_2^n) \in \mathcal{A}_e^n\}
\]

\[
\leq \sum_{(x^n,y_1^n,y_2^n) \in \mathcal{A}_n^g} p_{X^n,Y_1^n,Y_2^n}(x^n,y_1^n,y_2^n) \cdot (2^{nR_{\text{ach}}} - 1) \\
\cdot \Pr\{\exists y'_1 \neq y_1^n \text{ s.t. } f_1(y'_1) = f_1(y_1^n), (X^n(2),y'_1,y_2^n) \in \mathcal{A}_e^n\}
\]

\[
\leq \sum_{(x^n,y_1^n,y_2^n) \in \mathcal{A}_n^g} p_{X^n,Y_1^n,Y_2^n}(x^n,y_1^n,y_2^n) \cdot 2^{nR_{\text{ach}}} \\
\cdot \sum_{x',y'_1 \neq y_1^n \text{ s.t.}} \Pr\{X^n(2) = x'\} \cdot \Pr\{f_1(y'_1) = f_1(y_1^n)\}.
\]

Inequality (3.83) follows from the union bound. Inequality (3.84) follows because strings in \( \mathcal{Y}_1^n \) are randomly and uniformly assigned to the \( 2^{nR_3} \) bins, independent from the choice of input codewords. It would be an equality if we had not dropped the minus one from the middle term.

We next upper bound the inner summation in much the same way we did when upper bounding the probability of \( E_1 \). For every \( x' \) such that \( (x',y'_1,y_2^n) \in \mathcal{A}_n^g \), we
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

know from the definition of (weakly) typical sequences, (3.59), that

\[
\Pr \{ X^n(2) = x' \} \leq 2^{-n(H(X) - \epsilon)}. \tag{3.85}
\]

Additionally, for any triplet \((x', y_1', y_2'')\) where \(y_1' \neq y_1''\), the probability that \(y_1'\) and \(y_1''\) are assigned to the same bin equals

\[
\Pr \{ f_1(y_1') = f_1(y_1'') \} = 2^{-nR_3}. \tag{3.86}
\]

Finally, by definition, the number of \((x', y_1')\) sequences such that \((x', y_1', y_2'') \in \mathcal{A}_c^n\) is denoted \(|\mathcal{A}_c^n(X^n, Y_1 \mid y_2'')|\). For sufficiently large blocklength \(n\), and for \(y_2'' \in \mathcal{A}_c^n(Y_2)\), it follows as a straightforward consequence of the definition of weak typicality that [15, Th. 14.2.2]

\[
|\mathcal{A}_c^n(X^n, Y_1^n \mid y_2'')| \leq 2^{n(H(X, Y_1|Y_2) + 2\epsilon)}. \tag{3.87}
\]

Therefore, combining (3.85)–(3.87),

\[
\sum_{x', y_1' \neq y_1'' \text{ s.t.} \ (x', y_1', y_2'') \in \mathcal{A}_c^n} \Pr \{ X^n(2) = x' \} \cdot \Pr \{ f_1(y_1') = f_1(y_1'') \} \leq 2^{-n(R_3 + H(X) - H(X, Y_1|Y_2) - 3\epsilon)}. \tag{3.88}
\]

Substituting (3.88) into (3.84), we have that for sufficiently large \(n\),

\[
\Pr(E_2) \leq 2^{-n(R_3 + H(X) - R_{ach} - H(X, Y_1|Y_2) - 3\epsilon)}. \tag{3.89}
\]

Consequently, \(\Pr(E_2) \to 0\) as \(n \to \infty\) provided

\[
(R_3 + H(X) - R_{ach} - H(X, Y_1|Y_2) - 3\epsilon)
= (I(X; Y_1, Y_2) - R_{ach}) + H(X \mid Y_1, Y_2) + R_3 - H(X, Y_1 \mid Y_2) - 3\epsilon \tag{3.90}
\]

\[
= (I(X; Y_1, Y_2) - R_{ach}) + (R_3 - H(Y_1 \mid Y_2)) - 3\epsilon \tag{3.91}
\]

\[
> 0. \tag{3.92}
\]
CHAPTER 3. DISCRETE MEMORYLESS PARALLEL RELAY NETWORKS

From (3.65) and (3.66), these conditions are indeed satisfied when $\epsilon$ is small enough.

Analogously we can derive

$$\Pr (E_3) \leq 2^{-n(R_4 + H(X) - R_{ach} - H(X | Y_1) - 3\epsilon)}. \quad (3.93)$$

Consequently, $\Pr (E_3) \to 0$ as $n \to \infty$ provided

$$(I(X; Y_1, Y_2) - R_{ach}) + (R_4 - H(Y_2 | Y_1)) - 3\epsilon > 0. \quad (3.94)$$

From (3.65) and (3.67), these conditions are indeed satisfied when $\epsilon$ is small enough.

We also upper bound the probability of event $E_4$ analogously.

$$\Pr (E_4) = \sum_{[x^n, y_1^n, y_2^n] \in A_n^a} p_{X^n, Y_1^n, Y_2^n} (x^n, y_1^n, y_2^n)$$

$$\cdot \Pr \left\{ \exists x' \in C_{-1}, \ y'_1 \neq y_1^n, \ y'_2 \neq y_2^n \text{ such that:} \right.$$ 

$$f_1(y'_1) = f_1(y_1^n), \ f_2(y'_2) = f_2(y_2^n), \ \text{and} \ (x', y'_1, y'_2) \in A^n_c \right\}$$

$$\leq \sum_{[x^n, y_1^n, y_2^n] \in A_n^a} p_{X^n, Y_1^n, Y_2^n} (x^n, y_1^n, y_2^n) \cdot (2^{nR_{ach}} - 1)$$

$$\cdot \Pr \left\{ \exists y'_1 \neq y_1^n, \ y'_2 \neq y_2^n \text{ such that:} \right.$$ 

$$f_1(y'_1) = f_1(y_1^n), \ f_2(y'_2) = f_2(y_2^n), \ \text{and} \ (X^n(2), y'_1, y'_2) \in A^n_c \right\}$$

$$\leq \sum_{[x^n, y_1^n, y_2^n] \in A_n^a} p_{X^n, Y_1^n, Y_2^n} (x^n, y_1^n, y_2^n) \cdot 2^{nR_{ach}}$$

$$\cdot \sum_{x', y'_1 \neq y_1^n, y'_2 \neq y_2^n \text{ s.t.} \ (x', y'_1, y'_2) \in A^n_c} \Pr \{x' = X^n(2)\} \cdot \Pr \{f_1(y'_1) = f_1(y_1^n)\} \cdot \Pr \{f_2(y'_2) = f_2(y_2^n)\}. \quad (3.97)$$
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

The upper bound on the inner summation equals, for sufficiently large $n$,

\[
\sum_{x', y'_1 \neq y''_1, y'_2 \neq y''_2 \text{ s.t. } (x', y'_1, y'_2) \in \mathcal{A}^n} \Pr\{x' = X^n(2)\} \cdot \Pr\{f_1(y'_1) = f_1(y''_1)\} \cdot \Pr\{f_2(y'_2) = f_2(y''_2)\} \\
\leq 2^{-n[R_3 + R_4 + H(X) - H(X, Y_1, Y_2) - 3\varepsilon]}.
\] (3.98)

Substituting (3.98) into (3.97) we have that for sufficiently large $n$,

\[
\Pr(E_4) \leq 2^{-n[R_3 + R_4 + H(X) - R_{ach} - H(X, Y_1, Y_2) - 3\varepsilon]}.
\] (3.99)

Consequently, $\Pr(E_4) \to 0$ as $n \to \infty$ provided

\[(I(X; Y_1, Y_2) - R_{ach}) + (R_3 + R_4 - H(Y_1, Y_2)) - 3\varepsilon > 0. \] (3.100)

From (3.65) and (3.68), these conditions are indeed satisfied when $\varepsilon$ is small enough.

Therefore, averaged over the choice of input codebooks and relay observation bin assignments, and for sufficiently small $\varepsilon > 0$,

\[
\mathbb{E}\{\Pr_{\text{error}}\} \leq \sum_{i=1}^{4} \Pr(E_i)
\] (3.101)

\[
\to 0 \text{ as } n \to \infty. \] (3.102)

The result now follows from a standard argument. Pick any positive probability of error, $p_e > 0$. From (3.102), there is a sufficiently small $\varepsilon$ and sufficiently large $n$ such that

\[
\mathbb{E}\{\Pr_{\text{error}}\} < p_e.
\] (3.103)

$\mathbb{E}\{\Pr_{\text{error}}\}$ is computed by averaging over the random choice of input codebooks and relay observation bin assignments. Since the average probability of error is less than $p_e$, there must be at least one deterministic input codebook and a deterministic pair of relay observation bin assignments with average probability of message error less than $p_e$. The theorem follows since this holds for any $p_e > 0$. \qed
Lower Entropy of Observations with a Different Input Distribution

We found that when using any input distribution \( p(x) \), we can time-share between reliable coding to the relays and binning to communicate essentially all pairs of relay observations. We also found that, when using the latter approach, we can back \( R_{ach} \) off of the limit \( I(X;Y_1,Y_2) \) to linearly relax the link rate requirement. However, we also found that this is always a bad idea for symmetric networks.

We can also relax the link rate requirement by changing the input distribution, \( p(x) \). Based on our intuition from distributed source coding, this should make the artificial “observation source” less entropic and therefore easier to compress. Unfortunately, for our symmetric PBSC network, this has exactly the same numerical implication as the linear backoff scheme. Specifically,

\[
H(Y_1,Y_2 \mid X) = 2H(Y_1 \mid X), \tag{3.104}
\]

\[
H(Y_1 \mid X) = H(Y_1 \mid X = 0) \tag{3.105}
\]

\[
= H(Y_1 \mid X = 1). \tag{3.106}
\]

This shows that \( H(Y_1,Y_2 \mid X) \) does not depend on \( p(x) \), and therefore changing \( p(x) \) to reduce \( H(Y_1,Y_2) \) also reduces \( I(X;Y_1,Y_2) \) by exactly the same amount. Using this scheme, then, \( R_{ach} \) and \( R_3 \) are related at the boundary by

\[
R_{ach} = 2R_3 - H(Y_1,Y_2 \mid X) \tag{3.107}
\]

\[
= 2(R_3 - H_{\text{bin}}(\alpha)). \tag{3.108}
\]

From (3.55), this is exactly what we found with the linear backoff scheme. This result is plotted in Figure 3-6.
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

Figure 3-6: Evaluation of the reduced $H(Y_1, Y_2)$ scheme for the symmetric PBSC network
Simultaneously Combine Coding and Communicating Observations Using Broadcast Codes, Binning the Conditional Observations

In this section we describe a network coding technique that works for any symmetric parallel relay network with noiseless binary links to the decoder, as in Figure 3-5.

Our first main approach, coding to the relays, allows each relay to reliably decode all of the input information. Our second main approach, binning to communicate the relay observations, uses input codes designed for the channel with input $X$ and output $(Y_1, Y_2)$. With this second approach, the relays encode their observations, which are reliably reproduced by the decoder. We can simultaneously combine these two approaches using broadcast channel input codes with two streams of input information. The first stream, corresponding to the coarse information, will be reliably decoded at each relay. The second stream, corresponding to the fine information, will be reliably decoded only at the decoder. For the decoder to decode the fine information, the relays will communicate the relevant portion of their observations to the decoder. We will apply the binning technique we used previously, but we will adjust the technique somewhat. Both relays decode the coarse information, and they transmit this coarse information noiselessly to the decoder. Therefore, the relays can effectively strip out that part of their observations corresponding to the coarse information. They need only communicate their observations conditioned on the coarse information. We will clarify this momentarily.

![Diagram of broadcast channel](image)

**Figure 3-7:** Artificial broadcast channel used in the simultaneously combined approach

For any broadcast channel, a single relay observation, $Y_1$ or $Y_2$ alone, is a degraded version of the pair of observations $(Y_1, Y_2)$. Additionally, for symmetric broadcast channels, $Y_1$ and $Y_2$ are stochastically equivalent. Therefore we can use broadcast channel input codes designed for the degraded broadcast channel with input $X$, “bet-
3.3. **CODING FOR DISCRETE SYMMETRIC NETWORKS**

... output \((Y_1, Y_2)\), and degraded output \(Y_1 = Y_2\), as pictured in Figure 3-7.

Using (3.3), we explicitly describe the broadcast channel capacity region for this artificial broadcast channel. Pick a single letter distribution, \(p(u)\), to generate the cloud centers (coarse information). Pick a conditional single letter distribution, \(p(x \mid u)\), to generate the cloud points (fine information). Then we use a broadcast channel code at the input with rates

\[
R_{\text{coarse}} = I(U; Y_1), \\
R_{\text{fine}} = I(X; Y_1, Y_2 \mid U),
\]

where \(R_{\text{coarse}}\) is the rate of the coarse information in bits per input symbol, and \(R_{\text{fine}}\) is the rate of the fine information.

The overall coding scheme works as follows. The relays both decode the coarse information. They each send half of the corresponding bits to the decoder, using

\[
R_{\text{decoded}} = 0.5 \cdot R_{\text{coarse}}.
\]

The relays then use binning codes to transmit their observations conditioned on the coarse information; there are \(2^{R_{\text{coarse}}}\) different binning codes, one for each coarse information possibility. For such binning codes to work, with a symmetric broadcast channel, each relay requires

\[
R_{\text{bin} \text{ codes}} = 0.5 \cdot H(Y_1, Y_2 \mid U).
\]

bits per input symbol. We have described how the relays operate; focus now on the decoder. Having received the coarse information directly from the relays, the decoder reconstructs the conditional relay observations using the conditional binning codes, and then it decodes the fine information. From (3.111) and (3.112), the total rate requirement on the relay links is

\[
R_3 = R_4 > 0.5 \cdot (R_{\text{coarse}} + H(Y_1, Y_2 \mid U)).
\]
The total information rate from input to the decoder is

\[
R_{\text{ach}} = R_{\text{coarse}} + R_{\text{fine}}
\]

\[
= R_{\text{coarse}} + H(Y_1, Y_2 \mid U) - H(Y_1, Y_2 \mid X, U)
\]

\[
= R_{\text{coarse}} + H(Y_1, Y_2 \mid U) - H(Y_1, Y_2 \mid X)
\]

\[
= 2 \cdot R_3 - H(Y_1, Y_2 \mid X).
\]

Equality (3.116) follows because \( U \to X \to (Y_1, Y_2) \) by choice of single-letter distributions for constructing broadcast channel codes. Equality (3.117) follows by substituting (3.113) in (3.116).

For any symmetric broadcast channel with conditionally independent relay observations, such as in the symmetric PBSC network,

\[
H(Y_1, Y_2 \mid X) = 2 \cdot H(Y_1 \mid X).
\]

To evaluate this scheme for the symmetric PBSC network, we substitute (3.118) into (3.117) to find

\[
R_{\text{ach}} = 2 \cdot (R_3 - H(Y_1 \mid X))
\]

\[
= 2 \cdot (R_3 - H_{\text{bin}} (\alpha)).
\]

Comparing (3.120) with (3.55) and (3.108), we find that for the symmetric PBSC network, this new approach achieves exactly the same performance as the linear backoff approach of Theorem (3.3.1) and the reduced entropy approach of the previous section. This is pictured graphically in Figure 3-8. We will repeat this and add the converses in Figure E-2 on page 224 for easy reference.
Figure 3-8: Evaluation of using broadcast codes to combine coding to the relays and communicating observations for the symmetric PBSC network
Remap Relay Observations Symbol by Symbol; Essentially Perfect Representation of Remapped Observations

We can remap the relay observation letters \( Y_{1,2} \) symbol by symbol to new letters \( V_{1,2} \) and then consider \( (V_1, V_2) \) as the new observation source. We do so by choosing two, possibly random, mappings \( p(v_1 \mid y_1) \) and \( p(v_2 \mid y_2) \). We apply these mappings memorylessly to the observations at Relay 1 and Relay 2, respectively. Unfortunately, we cannot simply choose a joint mapping \( p(v_1, v_2 \mid y_1, y_2) \) and then apply it memorylessly to the observations. This would require access to both observations at each relay. If we could apply a joint remapping, this would yield better communication performance. It may be possible to approximate this joint remapping using complicated blockwise remappings at the relays, but we have not directly explored this. That is, we have not picked a joint mapping \( p(v_1, v_2 \mid y_1, y_2) \) and then tried to figure out how close we can come to mimicking that mapping. Rather, we will take a different approach based on block quantization that, in essence, accomplishes this indirectly. We will present the block quantization method later in Theorem 3.3.2.

For now, let us return to this very simple approach of remapping the observations symbol by symbol. Consider an arbitrary broadcast channel — not necessarily a parallel binary symmetric broadcast channel. The idea is to compress the observation alphabet to something less entropic, thereby making it easier to describe. Usually, this entails compressing the cardinality of the observation alphabets. As a simple extension to the previous theorem, having chosen and applied the two mappings, we can reliably achieve \( R_{\text{ach}} \) provided

\[
R_{\text{ach}} < I(X; V_1, V_2), \tag{3.121}
\]
\[
R_3 > H(V_1 \mid V_2), \tag{3.122}
\]
\[
R_4 > H(V_2 \mid V_1), \tag{3.123}
\]
\[
R_3 + R_4 > H(V_1, V_2). \tag{3.124}
\]

These terms are computed using the joint distribution

\[
p(x, v_1, v_2) = \sum_{y_1, y_2} p(x, y_1, y_2)p(v_1 \mid y_1)p(v_2 \mid y_2). \tag{3.125}
\]
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

We can also linearly back off from this limit on $R_{ach}$ in exactly the same way we did in Theorem 3.3.1.

For the symmetric PBSC network, this method does not improve our time-sharing scheme at all — this can be shown numerically. Intuitively, since the observation alphabets are binary, we do not have much room for improvement when choosing $p(v_1 \mid y_1)$ and $p(v_2 \mid y_2)$. However, note that there are dumb examples where this method is quite useful. For example, think of broadcast channels with artificially expanded observation alphabets (which can be compressed without losing much information about the input letter). As an extreme example, construct a binary input, 10-output channel as follows. Start with a core BSC and then cascade it with something which expands each output letter to, say, one of 5 different stochastically equivalent letters, as in Figure 3-9. Construct a broadcast channel by using two such independent links to the relays. The “observation source” would then be comparatively difficult to compress as is. However, we could first deterministically compress the artificially expanded alphabet back to its core binary alphabet. In this case, we would lose no information about the input letter from the compressed source.

\[ \begin{array}{c}
X \\
\ \ \ Y_1
\end{array} \]

Figure 3-9: Artificially expanded output alphabet

The point of this example is to show that there can be much improvement over the two main achievability schemes we have described thus far. Further, a general converse may require more insight than we could gather from the PBSC network alone — a converse derived only for the PBSC network may hinge upon the observation alphabets being binary.

We will develop an improved communication technique using block quantization at the relays rather than remapping symbol-by-symbol. Therefore, we will not explore the current approach further.
3.3.3 Block Quantization and Binning of the Relay Observations

Rather than remapping the relay observations symbol-by-symbol at the relays, we now adopt a more natural approach to communication. If the binary link rates are insufficient to allow the decoder to accurately reproduce the relay observations, we can instead block quantize the observations at each relay. The decoder can then reproduce a reasonable (quantized) representation of the relay observations. We will combine block quantization with random binning, which we used in earlier achievability schemes. We provide a helpful sketch in Figure 3-10 for the asymmetric PBSC network.

![Block quantization and binning approach for a PBSC network](image)

Figure 3-10: Block quantization and binning approach for a PBSC network

We will prove the following theorem. Choose a single-letter distribution on the input, $p(x)$. Choose two more single-letter distributions for the relay quantizers, $p(v_1 \mid y_1)$ and $p(v_2 \mid y_2)$. Together with the given broadcast channel, $p(y_1, y_2 \mid x)$, this determines a joint distribution $p(x, y_1, y_2, v_1, v_2) = p(x)p(y_1, y_2 \mid x)p(v_1 \mid y_1)p(v_2 \mid y_2)$. Then we can reliably achieve rates up to

$$R_{ach} = I(X; V_1, V_2),$$

(3.126)

provided

$$R_3 > I(Y_1; V_1) - I(V_1; V_2),$$

(3.127)

$$R_4 > I(Y_2; V_2) - I(V_1; V_2),$$

(3.128)

$$R_3 + R_4 > I(Y_1; V_1) + I(Y_2; V_2) - I(V_1; V_2).$$

(3.129)
3.3. \textit{Coding for Discrete Symmetric Networks}

We delay proving this theorem until the end of this section. First note from the data processing theorem that $I(X;V_1,V_2) \leq I(X;Y_1,Y_2)$, and therefore the performance is indeed dominated by the broadcast cut-set bound (as it must be). In particular, non-degenerate choices for the relay quantizers will result in $R_{\text{ach}}$ strictly less than the broadcast cut-set bound of (3.10). This makes sense because the decoder will be attempting to reproduce a quantized, and therefore noisy, version of the relay observations.

The basic idea behind this scheme is as follows. We generate $2^{nR_{\text{ach}}}$ input codewords using the input distribution $p(x)$. We generate approximately $2^{nI(Y_1;Y_2)}$ quantization codewords for Relay 1 using the marginal distribution $p(v_1)$. We randomly assign these quantization codewords to $2^{nR_3}$ bins. Similarly, we generate approximately $2^{nI(Y_2;Y_1)}$ quantization codewords for Relay 2 using the marginal distribution $p(v_2)$. We randomly assign these quantization codewords to $2^{nR_4}$ bins.

We would like to think of the quantization codebooks as independent rate-distortion codes (with Hamming distortion) for the dependent observation sources $Y_1$ and $Y_2$, though there is no explicit distortion measure. Upon receiving $Y_1$, Relay 1 looks for any quantization codeword that is jointly typical with $Y_1$. If one is found, Relay 1 sends the appropriate index to the decoder. From the rate-distortion theorem, such a quantization codeword exists with high probability provided we generate more than $2^{nI(Y_1;Y_2)}$ quantization codewords. Thinking in terms of rate-distortion, we can choose the Relay 1 quantization map $p(v_1 \mid y_1)$ that achieves the rate-distortion function for the observation source $Y_1$. For intuition, compare using two different Relay 1 maps, $p(v_1 \mid y_1)$, each achieving the rate-distortion function for different values of distortion. The choice that achieves the larger distortion has fewer codewords. This reduces the binary link rate requirement, $R_3$, at the expense of communication performance, $I(X;V_1,V_2)$.

The positive terms in the link rate requirements (3.127)-(3.129) thus come from the need to generate enough quantization codewords to reliably perform the quantization step. The negative contributions, which relax the link rate requirements, come from binning the quantization codewords. Since $Y_1^n$ and $Y_2^n$ are dependent, the quantization codewords chosen by the relays during operation are also dependent. The decoder can therefore eliminate all pairs of quantization codewords in the
indicated bins that are not appropriately consistent. Specifically, the relays each indicate the bin to which their chosen quantization codeword belongs. But there are other quantization codewords assigned to those same bins. For those other quantization codewords, if we consider one from each bin, the probability that any two are consistent with each other is approximately $2^{-nI(V_1;V_2)}$.

Let us compare this block quantization result with the symbol-by-symbol remapping result. For any input distribution, $p(x)$, and pair of single-letter relay maps, $p(v_1 | y_1)$ and $p(v_2 | y_1)$, the achievable communication rate is $I(X;V_1,V_2)$ for both schemes. However, the binary link rate requirements are larger with symbol remapping. From (3.122), for symbol remapping we require

$$R_3 > H(V_1 | V_2).$$  \hspace{1cm} (3.130)

With block quantization, on the other hand, from (3.127), we require

$$R_3 > I(Y_1;V_1) - I(V_1;V_2)$$  \hspace{1cm} (3.131)

$$= H(V_1 | V_2) - H(V_1 | Y_1).$$  \hspace{1cm} (3.132)

Similarly, with symbol remapping, from (3.124), we require

$$(R_3 + R_4) > H(V_1,V_2).$$  \hspace{1cm} (3.133)

With block quantization, from (3.129), we require

$$(R_3 + R_4) > I(Y_1;V_1) + I(Y_2;V_2) - I(V_1;V_2)$$  \hspace{1cm} (3.134)

$$= H(V_1,V_2) - H(V_1 | Y_1) - H(V_2 | Y_2).$$  \hspace{1cm} (3.135)

Block quantization thus typically improves the performance over symbol remapping by reducing the link rate requirements. Note that if we choose deterministic single-letter quantization maps (i.e., when $p(v_1 | y_1)$ and $p(v_2 | y_2)$ equal 0 or 1 for all $y_1, y_2, v_1$, and $v_2$), then block quantization requires exactly the same link rates as symbol remapping. In particular, if we choose the mappings so that $V_1 \equiv Y_1$ and $V_2 \equiv Y_2$, then the achievable rate as well as the link rate requirements for both sym-
bol remapping and block quantization are precisely the same as when reproducing the full relay observations, \((Y_1, Y_2)\). Thus when the relay maps \(p(v_1 | y_1)\) and \(p(v_2 | y_2)\) are degenerate, both the symbol remapping and the block quantization approaches degenerate to the approach of binning to communicate the observations, Theorem 3.3.1 (with \(\delta = 0\)).

We now evaluate this theorem for our symmetric PBSC network. Given a symmetric network, if we choose identical single letter relay maps \(p(v_1 | y_1)\) and \(p(v_2 | y_2)\), then the third link rate requirement, (3.129), dominates the other two, (3.127) and (3.128). We do indeed now choose \(p(v_1 | y_1) \equiv p(v_2 | y_2)\) for evaluating this block quantization approach. We also need to choose an input distribution, \(p(x)\). The only input distribution that makes sense is \(p(x) = (0.5, 0.5)\), i.e., Bernoulli(0.5). So chosen, the marginal distributions of the artificial i.i.d. “observation sources”, \(p(y_1)\) and \(p(y_2)\), are also Bernoulli(0.5). Returning to the single letter relay maps, we need to choose \(p(v_1 | y_1)\). We now choose the conditional distribution \(p(v_1 | y_1)\) as Bernoulli(D), which achieves the rate-distortion function for the artificial Bernoulli(0.5) “source” \(Y_1\) at distortion \(D\). Then for any such chosen target distortion \(D \in [0, 0.5]\), and with \(\beta\) defined as

\[
\beta = \alpha \otimes D, \tag{3.136}
\]

we evaluate the achievable rate and link rate requirements; from (3.126)–(3.129),

\[
R_{ach} < I(X; V_1, V_2) \tag{3.137}
\]

\[
= 1 + H_{\text{bin}}(\beta \otimes \beta) - 2H_{\text{bin}}(\beta), \tag{3.138}
\]

\[
R_3 = R_4 > 0.5 \left( 1 + H_{\text{bin}}(\beta \otimes \beta) - 2H_{\text{bin}}(D) \right). \tag{3.139}
\]

For any pair \(\alpha\) and \(D\), from (3.136),

\[
\max[\alpha; D] \leq \beta \leq 0.5. \tag{3.140}
\]

Equality on one side or the other in (3.140) can only be achieved when \(\alpha\) or \(D\) are
CHAPTER 3. DISCRETE MEMORYLESS PARALLEL RELAY NETWORKS

degenerate — that is, when $\alpha$ or $D$ equal 0 or 0.5. Assuming

$$0 < \alpha < 0.5 \quad \text{and} \quad 0 < D < 0.5,$$

equations (3.138) and (3.139) imply that

$$R_{\text{ach}} < \min[I(X; Y_1, Y_2); 2R_3]. \tag{3.142}$$

In other words, this approach does not meet either the broadcast cut-set bound (3.10) or the multiaccess cut-set bound (3.19). When $D = 0$, we get $V_1 \equiv Y_1$ and $V_2 \equiv Y_2$, and this results in essentially perfect reproduction of $(Y_1, Y_2)$. On the other extreme, when $D = 0.5$, we get zero communication rate, $R_{\text{ach}} = 0$, with zero link rate requirements.

For the symmetric PBSC network, we again present the result graphically at fixed $C_1 = C_2 = 0.5$ as we increase the link rate $R_3 = R_4$ from 0 to 1. This is done in Figure 3-11.\footnote{The performance curve in Figure 3-11 corresponding to quantization and binning, (3.136)–(3.139), is reasonably accurate but was not mathematically generated.} We will repeat this figure in Figure E-3 on page 225 for easy reference. The dotted line in the figure corresponds to time-sharing between pure coding to the relays and binning to communicate the relay observations. As implied earlier, the communication rate achievable with this quantization approach decreases as the distortion $D$ increases. Note from the figure that the performance drops off rapidly as $D$ increases from zero. However, it is possible to show that the performance approaches the point $(0.5H(Y_1, Y_2), I(X; Y_1, Y_2))$ from the left with zero slope. Therefore, this is a very slight but still a strict improvement for our symmetric PBSC network. We can time-share between pure coding to the relays (again at the point $(0.5C_1, C_1)$) and this new approach, achieving very slightly improved performance relative to our previous time-sharing scheme.
3.3. **CODING FOR DISCRETE SYMMETRIC NETWORKS**

![Graph](image)

**Figure 3-11:** Evaluation of the block quantization and binning scheme for the symmetric PBSC network, $C_1 = C_2 = 0.5$
CHAPTER 3. DISCRETE MEMORYLESS PARALLEL RELAY NETWORKS

We now prove the relevant theorem. To do so, we need the concept of strong typicality rather than weak typicality. Strong typicality was introduced by Berger [6] and developed extensively by Csiszár and J. Körner [16]. For our purposes, the primary difference between the two concepts is embodied in the lower bound of Lemma 5 (to follow). It is this which technically allows us to combine channel coding on the broadcast side with block quantization (rate-distortion) at the relays. With rate-distortion based on weak typicality, we prove that on average, we can block quantize the “observation source” to the chosen distortion. With strong typicality, the rate-distortion theorem works as follows. For every typical “observation source” outcome, with high probability, there is a quantization codeword jointly strongly typical with that outcome. Thus there is a sort of uniformity in the strongly typical set. If we were interested in distortion, as we are when deriving the rate-distortion theorem, then we note that joint strong typicality implies low distortion. It is worthwhile to mention that the definition of strong typicality has nothing to do with the strong law of large numbers.

Now consider our communication problem. We can use the strong typicality approach by first asserting that the input codeword $x^n$ and the relay observations $(y_1^n, y_2^n)$ are jointly typical with high probability. Then, for these relay observations $y_1^n$ and $y_2^n$, with high probability, there are a pair of relay quantization codewords, $v_1^n$ and $v_2^n$, such that $(y_1^n, v_1^n)$ and $(y_2^n, v_2^n)$ are jointly typical. Finally, we use a few lemmas to show that this implies $(x^n, v_1^n, v_2^n)$ are jointly typical with high probability. All of these probabilities are computed by averaging over an ensemble of input codebooks, an ensemble of relay quantization codebooks, and the broadcast channel outcomes. We were unable to prove this theorem based on weak typicality.

We need several definitions and relatively simple lemmas to prove the theorem. For the reader familiar with the conventions, we use the strong typicality definition of Csiszár and J. Körner [16], rather than that of Berger [6] and of Cover and Thomas [15], because we do not want to implicitly normalize the typicality measure, $\epsilon$, by the discrete alphabet sizes.

For any sequence $x^n$ with alphabet $\mathcal{X}^n$, define the function $N(a \mid x^n)$ to be the number of occurrences of the letter $a \in \mathcal{X}$ in the string $x^n$. Then we have the following two definitions.
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

**Definition 2.** [16, Def. 2.8, p. 33] For an arbitrary distribution \( p_X(x) \) on \( X \), a sequence \( x^n \in X^n \) is \( \epsilon \)-strongly typical if

\[
\frac{1}{n} N(a \mid x^n) - p_X(a) \leq \epsilon \quad \forall \ a \in X,
\]

and further, \( N(a \mid x^n) = 0 \) whenever \( p_X(a) = 0 \). In this case, we write \( x^n \in A^*_\epsilon(X) \) when the distribution, \( p_X(x) \), and the blocklength, \( n \), are understood.

This definition is extended in the obvious way to joint distributions \( p_{X,Y}(x,y) \) with the function \( N(a,b \mid x^n,y^n) \).

**Definition 3.** [16, Def. 2.9, p. 34] For an arbitrary transition matrix \( p_{Y\mid X}(y \mid x) \) and any sequence \( x^n \in X^n \), a sequence \( y^n \in Y^n \) is \( \epsilon \)-strongly conditionally typical if

\[
\frac{1}{n} N(a,b \mid x^n,y^n) - \frac{1}{n} N(a \mid x^n) \cdot p_{Y\mid X}(b \mid a) \leq \epsilon \quad \forall \ (a,b) \in X \times Y,
\]

and further, \( N(a,b \mid x^n,y^n) = 0 \) whenever \( p_{Y\mid X}(b \mid a) = 0 \). In this case, we write \( y^n \in A^*_\epsilon(Y \mid x^n) \) when the conditional distribution, \( p_{Y\mid X}(y \mid x) \), is understood. Here, the blocklength, \( n \), is either understood or implied directly from the notation.

We will need several lemmas concerning strongly typical sequences. The first follows directly from the weak law of large numbers.

**Lemma 1.** [15, Lemma 13.6.1, p. 359] Let \( X_i, i = 1,2,\ldots,n \) be drawn i.i.d. \( \sim p_X(x) \). Then for any \( \epsilon > 0 \), \( \Pr(X^n \in A^*_\epsilon(X)) \to 1 \) as \( n \to \infty \).

The next lemma states that joint typicality implies marginal typicality. That is, if \( (x^n,y^n) \) are jointly typical, then \( x^n \) and \( y^n \) are each marginally typical. This follows immediately from the definition of strong typicality, but the measure of typicality, \( \epsilon \), must be loosened to make the statement. This is probably the reason for the difference in the two conventional definitions of strong typicality.

**Lemma 2.** [15, p. 359] If \( (x^n,y^n) \in A^*_\epsilon(X,Y) \), then \( x^n \in A^*_\epsilon(Y \mid X) \).

The next lemma states that if \( (x^n,y^n) \) are jointly typical, then \( y^n \) is conditionally typical. Again, the measure of typicality, \( \epsilon \), must be loosened to make the statement.
Lemma 3. If \((x^n, y^n) \in \mathcal{A}_\varepsilon^*(X, Y)\), then \(y^n \in \mathcal{A}_\varepsilon^*(Y \mid x^n)\), where

\[
\varepsilon' = \varepsilon \cdot \left(1 + |\mathcal{Y}| \cdot \max_{a,b} p_{Y \mid X}(b \mid a)\right) \leq \varepsilon \cdot (1 + |\mathcal{Y}|).
\]

Remark: Before proving the lemma, consider first a back of the envelope calculation. Since \((x^n, y^n)\) are typical, \(x^n\) is typical. The fraction of locations where \(x_i = a\) is approximately \(p_X(a)\). The fraction of locations where \(x_i = a\) and \(y_i = b\) is approximately \(p_{X,Y}(a,b)\). Therefore, the fraction of locations where \(y_i = b\) given that \(x_i = a\) is approximately \(\frac{p_{X,Y}(a,b)}{p_X(a)} = p_{Y \mid X}(b \mid a)\). Therefore \(y^n\) looks conditionally typical. Now for the details.

Proof. First note that \(p_{Y \mid X}(b \mid a) = 0 \Rightarrow p_{X,Y}(a,b) = 0\). Therefore, since \((x^n, y^n)\) are jointly typical, \(N(a,b \mid x^n, y^n) = 0\) whenever \(p_{Y \mid X}(b \mid a) = 0\). Also,

\[
\frac{1}{n} N(a \mid x^n) = \frac{1}{n} \cdot \sum_{b \in \mathcal{Y}} N(a,b \mid x^n, y^n) = \frac{1}{n} \cdot \sum_{b \in \mathcal{Y}} \left(p_{X,Y}(a,b) \pm \varepsilon\right)
\]

\[
\in p_X(a) \pm \varepsilon \cdot |\mathcal{Y}|.
\]

Inclusion (3.144) follows by definition of joint typicality for \((x^n, y^n)\), and inclusion (3.145) is obvious. Continuing,

\[
\frac{1}{n} N(a,b \mid x^n, y^n) - \frac{1}{n} N(a \mid x^n) \cdot p_{Y \mid X}(b \mid a)
\]

\[
\in \left[\frac{1}{n} N(a,b \mid x^n, y^n) - p_{X,Y}(a,b)\right] \pm \varepsilon \cdot |\mathcal{Y}| \cdot p_{Y \mid X}(b \mid a)
\]

\[
\in \left[\frac{1}{n} N(a,b \mid x^n, y^n) - p_{X,Y}(a,b)\right] \pm \varepsilon \cdot |\mathcal{Y}| \cdot \max_{a,b} p_{Y \mid X}(b \mid a)
\]

\[
\in \pm \varepsilon \cdot \left(1 + |\mathcal{Y}| \cdot \max_{a,b} p_{Y \mid X}(b \mid a)\right)
\]

\[
\in \pm \varepsilon \cdot (1 + |\mathcal{Y}|) .
\]
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

Inclusion (3.146) follows by substituting (3.145), and inclusion (3.148) follows since 
\((x^n, y^n)\) are jointly typical by assumption. The lemma follows immediately. 

\(\square\)

The next lemma states that if \(x^n\) is typical and \(y^n\) is conditionally typical, then the 
pair \((x^n, y^n)\) is jointly typical. Once again, the measure of typicality, \(\epsilon\), must be 
loosened when making this statement.

**Lemma 4.** [16, Lemma 2.10, p. 34] If \(x^n \in \mathcal{A}^*_{\epsilon_1}(X)\) and \(y^n \in \mathcal{A}^*_{\epsilon_2}(Y \mid x^n)\), then 
\((x^n, y^n) \in \mathcal{A}^*_{\epsilon_1 + \epsilon_2}(X, Y)\).

The next lemma concerns the probability that independently generated sequences 
are jointly typical. This is an extension and strengthening (due to the lower bound) 
of the same result for weak typicality.

**Lemma 5.** [15, Th. 13.6.2, p. 359] Let \(Y_1, Y_2, \ldots, Y_n\) be generated i.i.d. \(\sim p_Y(y)\) 
and independent from \(X^n\). For any \(x^n \in \mathcal{A}^*_{\epsilon}(X)\), the probability that 
\((x^n, Y^n) \in \mathcal{A}^*_{n(\epsilon)}(X, Y)\) is bounded by

\[
2^{-n(I(X;Y) + \epsilon)} \leq \Pr \left( (x^n, Y^n) \in \mathcal{A}^*_{n(\epsilon)}(X, Y) \right) \leq 2^{-n(I(X;Y) - \epsilon)},
\]

where \(\epsilon \to 0\) as \(\epsilon \to 0\) and \(n \to \infty\).

In addition to these basic lemmas concerning strong typicality, we will need the 
following lemma to prove our theorem.

**Lemma 6.** Fix any \(0 \leq f, k \leq 1\) and choose any \(\epsilon > 0\). We have \(S\) slots. Mark any 
\(F = \lfloor f \cdot S \rfloor\) of these slots as "special" slots. Throw \(K = \lfloor k \cdot S \rfloor\) balls randomly and 
uniformly into the \(S\) slots, allowing at most one ball per slot. Let \(V = v \cdot S\) be the 
number of balls that end up in "special" slots. Then for all \(S\), \(\max[0; K + F - S] \leq V \leq \min[K; F]\) and

\[
\Pr \left( |v - f \cdot k| > \epsilon \right) \leq \left( S^2 + S \right) \cdot \left( 2^{-S \cdot g_{f,k}(f \cdot k - \epsilon)} + 2^{-S \cdot g_{f,k}(f \cdot k + \epsilon)} \right),
\]

where

\[
g_{f,k}(v) = H_{\text{bin}}(k) - f H_{\text{bin}} \left( \frac{v}{f} \right) - (1 - f) H_{\text{bin}} \left( \frac{k - v}{1 - f} \right).
\]
In particular, the function \( g_{f,k}(v) \) is strictly convex on \( \max[0;k + f - 1] \leq v \leq \min[k; f] \) and \( g_{f,k}(f \cdot k) = 0 \). This implies that for any \( \epsilon > 0 \),

\[
\lim_{S \to \infty} \Pr (|v - f \cdot k| > \epsilon) = 0.
\]  

(3.152)

Proof. See the appendix. \( \square \)

We are now ready to prove the following theorem.

**Theorem 3.3.2.** We are given the parallel relay network of Figure 3-5 with discrete memoryless broadcast channel \( p(y_1, y_2 \mid x) \) and noiseless binary link rates \( R_3 \) and \( R_4 \). The relay observations need not be conditionally independent given the input. Choose any single-letter input distribution \( p(x) \) and any pair of transition matrices \( p(v_1 \mid y_1) \) and \( p(v_2 \mid y_2) \). We can reliably achieve rate \( R_{ach} \) provided

\[
R_{ach} < I(X; V_1, V_2),
\]

(3.153)

\[
R_3 > I(Y_1; V_1) - I(V_1; V_2),
\]

(3.154)

\[
R_4 > I(Y_2; V_2) - I(V_1; V_2),
\]

(3.155)

\[
R_3 + R_4 > I(Y_1; V_1) + I(Y_2; V_2) - I(V_1; V_2).
\]

(3.156)

These values are computed with respect to the distribution

\[
p(x, y_1, y_2, v_1, v_2) = p(x)p(y_1, y_2 \mid x)p(v_1 \mid y_1)p(v_2 \mid y_2).
\]

(3.157)

Proof. Assume the conditions of the theorem are satisfied. For now, we choose the decoder’s typicality measure, \( \epsilon_{\text{dec}} \), and the integer blocklength, \( n \), arbitrarily. We will later set \( \epsilon_{\text{dec}} \) sufficiently small and \( n \) sufficiently large to make the probability of message error as small as we desire. We will ignore the integrality constraints by assuming \( 2^{nR_{ach}} \), \( 2^{nR_3} \), and \( 2^{nR_4} \) are integer valued for every integer \( n \) — the integrality constraints can be handled trivially at the expense of notational inconvenience.

Randomly generate \( 2^{nR_{ach}} \) input codewords of blocklength \( n \). Generate each symbol of each codeword independently according to \( p(x) \). Denote these input codewords by \( X^n(m), m = 1, 2, \ldots, 2^{nR_{ach}} \), and denote the randomly chosen input codebook by \( C = \{ U_{m=1}^{2^{nR_{ach}}} X^n(m) \} \). For a particular codebook choice, denote the in-
put codewords by \( x^n(m), m = 1, 2, \ldots, 2^{nR_{\text{ach}}} \), and denote the input codebook by \( c = \{ \bigcup_{m=1}^{2^{nR_{\text{ach}}}} x^n(m) \} \).

Set \( \epsilon_{\text{quant}} > 0 \) and \( \delta > 0 \) as arbitrary parameters for now. We will later show how to choose them. Randomly generate \( 2^{n(I(Y_1;V_1)+\delta)} \) Relay 1 quantization codewords of blocklength \( n \). Generate each symbol of each codeword independently according to \( p(v_1) \). Denote these Relay 1 quantization codewords by \( V_1^n(j), j = 1, 2, \ldots, 2^{n(I(Y_1;V_1)+\delta)} \). For a particular quantization codebook choice, denote the quantization codebook by \( c_{q1} = \{ \bigcup_j v_1^n(j) \} \). Next, having randomly generated the Relay 1 quantization codebook, randomly and uniformly assign each Relay 1 quantization codeword, \( v_1^n \), to one of \( 2^{nR_3} \) bins. Denote the randomly chosen bin assignments by the function \( f_1(v_1^n(j)) \).

Similarly, randomly generate \( 2^{n(I(Y_2;V_2)+\delta)} \) Relay 2 quantization codewords of blocklength \( n \). Generate each symbol of each codeword independently according to \( p(v_2) \). Denote these Relay 2 quantization codewords by \( V_2^n(j), j = 1, 2, \ldots, 2^{n(I(Y_2;V_2)+\delta)} \). For a particular quantization codebook choice, denote the quantization codebook by \( c_{q2} = \{ \bigcup_j v_2^n(j) \} \). Next, having randomly generated the Relay 2 quantization codebook, randomly and uniformly assign each Relay 2 quantization codeword, \( v_2^n \), to one of \( 2^{nR_4} \) bins. Denote the randomly chosen bin assignments by the function \( f_2(v_2^n(j)) \).

Reveal the input codebook, \( c \), the quantization codebooks, \( c_{q1} \) and \( c_{q2} \), and the bin assignments, \( f_1(\cdot) \) and \( f_2(\cdot) \), to the decoder.

**Encoding:** The sender transmits the codeword \( x^n(m) \) corresponding to message \( m \). After receiving \( Y_1^n \), Relay 1 searches for any quantization codeword \( v_1^n(j_1) \in c_{q1} \) such that \( (Y_1^n, v_1^n(j_1)) \in \mathcal{A}_{\epsilon_{\text{quant}}}^*(Y_1, V_1) \). If one or more such quantization codewords exist, Relay 1 chooses one of them randomly (uniformly amongst the candidates) and sends \( w_3^{nR_3} = f_1(v_1^n(j_1)) \). Otherwise, Relay 1 declares an error (e.g., sends \( w_3^{nR_3} \equiv 0 \)). In this case, for notational convenience, we define the quantization codeword \( v_1^n = \emptyset \). Similarly, after receiving \( Y_2^n \), Relay 2 searches for any quantization codeword \( v_2^n(j_2) \in c_{q2} \) such that \( (Y_2^n, v_2^n(j_2)) \in \mathcal{A}_{\epsilon_{\text{quant}}}^*(Y_2, V_2) \). If one or more such quantization codewords exist, Relay 2 chooses one of them randomly (uniformly amongst the candidates) and sends \( w_4^{nR_4} = f_2(v_2^n(j_2)) \). Otherwise, Relay 2 declares an error (e.g., sends \( w_4^{nR_4} \equiv 0 \)). Again, in this case, we define the quantization codeword \( v_2^n = \emptyset \).

**Decoding:** At the decoder, we will use strongly typical set decoding. The de-
Chapter 3. Discrete Memoryless Parallel Relay Networks

coder receives the pair of bin numbers \( w_3^{nR_3} \) and \( w_4^{nR_4} \). If either of these bin numbers equal the relay error message 0, the decoder declares an error. This indicates that one or both of the relays failed the quantization step. Otherwise, the decoder declares message \( m \) was sent if it is the unique message with a pair of quantization codewords, \( v_1^n \) and \( v_2^n \), such that \( f_1(v_1^n) = w_3^{nR_3} \), \( f_2(v_2^n) = w_4^{nR_4} \), and \( (x^n(m), v_1^n, v_2^n) \in \mathcal{A}_{\text{dec}}^*(X, V_1, V_2) \). If no message or if more than one message satisfies this criterion, the decoder declares an error.

**Average Probability of Decoding Error:** We compute the average probability of message error by averaging over the choice of input codewords, the choice of relay quantization codebooks, the quantization codeword bin assignments, and the broadcast channel outcomes. We choose the codewords for the input codebook by randomly generating them independently from each other. Furthermore, we use the same probability distribution to choose each input codeword. Therefore, from the symmetry of these random choices, the average probability of message error for the randomly chosen input codebook equals the average probability of message error for the randomly chosen input codebook conditioned on the source transmitting the first message. We will therefore upper bound the average probability of message error conditioned on the source transmitting the first message. We denote this by \( \mathbb{E}\{\Pr_{\text{error}}\} \).

Before averaging over the input codebook ensemble, the input codeword corresponding to the first message is denoted \( X^n(1) \). During operation, the relays receive the pair of observations \((Y_1^n, Y_2^n)\). For a particular triple of strings \((x^n, y_1^n, y_2^n)\), the probability that \( X^n(1) = x^n \) and \((Y_1^n, Y_2^n) = (y_1^n, y_2^n)\) is denoted \( p_{X^n,Y_1^n,Y_2^n}(x^n, y_1^n, y_2^n) \). Before proceeding, recall our notation for the input codebook, \( \mathcal{C} = \{\cup_{m=1}^{2^{nR_{\text{enc}}}} X^n(m)\} \). When computing the average probability of message error, we will consider the input codebook without the first codeword. Define \( \mathcal{C}_{-1} = \{\cup_{m=2}^{2^{nR_{\text{enc}}}} X^n(m)\} \). Then \( \mathcal{C}_{-1} \subseteq \mathcal{C} \) with equality if and only if the first codeword, \( X^n(1) \), happens to be identical to the codeword for another message. Finally, we adopt the conventional set notation \( \mathcal{C}_{q_1} \setminus V_1^n = C_{q_1} - V_1^n \) and \( \mathcal{C}_{q_2} \setminus V_2^n = C_{q_2} - V_2^n \). Note that it is possible to have \( \mathcal{C}_{-1} \neq \mathcal{C} \setminus X^n(1) \).
3.3. Coding for Discrete Symmetric Networks

We group all error events into the union of five events. They are:

\[ E_0 = \left\{ (X^n(1), V_1^n, V_2^n) \notin \mathcal{A}_{\delta_{\text{dec}}}^* (X, V_1, V_2) \right\}; \tag{3.158} \]

\[ E_0 \] is the event that the input codeword and actual relay quantization codewords are not jointly typical.

\[
E_1 = \left\{ (X^n(1), V_1^n, V_2^n) \in \mathcal{A}_{\delta_{\text{dec}}}^* (X, V_1, V_2), \quad \exists x' \in \mathcal{C}_{-1} \text{ such that: } \right. \\
\left. (x', V_1^n, V_2^n) \in \mathcal{A}_{\delta_{\text{dec}}}^* (X, V_1, V_2). \right\}; \tag{3.159} \]

\[ E_1 \] is the event that \( (X^n(1), V_1^n, V_2^n) \) is jointly typical, and there is an incorrect input codeword, \( x' \), such that the triple \( (x', V_1^n, V_2^n) \) is jointly typical.

\[
E_2 = \left\{ (X^n(1), V_1^n, V_2^n) \in \mathcal{A}_{\delta_{\text{dec}}}^* (X, V_1, V_2), \quad \right. \\
\exists x' \in \mathcal{C}_{-1} \text{ and } v'_1 \in \mathcal{C}_{q_1} \backslash V_1^n \text{ such that: } \\
f_1(v'_1) = W_3^{-nR_3} \quad \text{and} \quad (x', v'_1, V_2^n) \in \mathcal{A}_{\delta_{\text{dec}}}^* (X, V_1, V_2). \right\}; \tag{3.160} \]

\[ E_2 \] is the event that \( (X^n(1), V_1^n, V_2^n) \) is jointly typical, and there is an incorrect input codeword, \( x' \), and a different Relay 1 quantization codeword, \( v'_1 \), assigned to the same bin as the chosen Relay 1 quantization codeword, such that the triple \( (x', v'_1, V_2^n) \) is jointly typical.

\[
E_3 = \left\{ (X^n(1), V_1^n, V_2^n) \in \mathcal{A}_{\delta_{\text{dec}}}^* (X, V_1, V_2), \quad \right. \\
\exists x' \in \mathcal{C}_{-1} \text{ and } v'_2 \in \mathcal{C}_{q_2} \backslash V_2^n \text{ such that: } \\
f_2(v'_2) = W_4^{-nR_4} \quad \text{and} \quad (x', V_1^n, v'_2) \in \mathcal{A}_{\delta_{\text{dec}}}^* (X, V_1, V_2). \right\}; \tag{3.161} \]

\[ E_3 \] is the event that \( (X^n(1), V_1^n, V_2^n) \) is jointly typical, and there is an incorrect input codeword, \( x' \), and a different Relay 2 quantization codeword, \( v'_2 \), assigned to the same bin as the chosen Relay 2 quantization codeword, such that the triple \( (x', V_1^n, v'_2) \) is jointly typical.
\[ E_4 = \left\{ \begin{array}{l} (X^n(1), V_1^n, V_2^n) \in \mathcal{A}_{\text{dec}}^*(X, V_1, V_2), \\
\exists x' \in C_{-1}, \ v'_1 \in C_{q_1} \setminus V_1^n, \text{ and } v'_2 \in C_{q_2} \setminus V_2^n \text{ such that:} \\
f_1(v'_1) = W_3^n R_3, \ f_2(v'_2) = W_4^n R_4, \text{ and } (x', V_1^n, v'_2) \in \mathcal{A}_{\text{dec}}^*(X, V_1, V_2). \end{array} \right\}; \] 

\[ (3.162) \]

\( E_4 \) is the event that \((X^n(1), V_1^n, V_2^n)\) is jointly typical, and there is an incorrect input codeword, \( x' \), and two different relay quantization codewords, \( v'_1 \) and \( v'_2 \), each assigned to the same bin as the chosen relay quantization codewords, such that the triple \((x', v'_1, v'_2)\) is jointly typical.

Using basic set theory and the union bound,

\[ \mathbb{E}\{\Pr_{\text{error}}\} = \Pr\left(\bigcup_{i=0}^{4} E_i\right) \leq \sum_{i=0}^{4} \Pr\left(E_i\right). \] 

\[ (3.163) \]

We will show that, for sufficiently small \( \epsilon_{\text{dec}} \), each of these five probabilities can be made arbitrarily small by appropriately choosing \( \epsilon_{\text{ch}}, \epsilon_{\text{quant}} \) and \( \delta \) and by increasing the blocklength \( n \).

We begin by bounding \( \Pr(E_0) \). From Lemma 1, we can choose \( \epsilon_{\text{ch}}(n) > 0 \) such that \( \epsilon_{\text{ch}}(n) \to 0 \) as \( n \to \infty \) and such that

\[ \Pr\left(\left(X^n(1), V_1^n, V_2^n\right) \in \mathcal{A}_{\text{ch}}^* (X, V_1, V_2)\right) \to 1 \text{ as } n \to \infty. \] 

\[ (3.164) \]

Note that \( \epsilon_{\text{ch}}(n) \) is a function of the blocklength \( n \). However, to simplify the notation, we will drop the dependence on \( n \). We thus write \( \epsilon_{\text{ch}} \to 0 \) as \( n \to \infty \).

We now show that if we choose our typicality measures appropriately, then with high probability as \( n \to \infty \), the relays will produce a pair of quantization codewords, \( V_1^n \) and \( V_2^n \), such that \((X^n, V_1^n, V_2^n)\) are jointly strongly typical (with typicality measure \( \epsilon_{\text{dec}} \)).

Let \((x^n, y_1^n, y_2^n)\) denote any particular jointly \( \epsilon_{\text{ch}} \)-strongly typical triple of input codeword and received relay observations. From Lemma 2, \( y_1^n \in \mathcal{A}_{\text{ch}}^* |_{X^n} (Y_1) \) and \( y_2^n \in \mathcal{A}_{\text{ch}}^* |_{X^n} (Y_2) \). We now use the achievability proof of the rate-distortion theorem
based on strong typicality, [15, pp. 358–62]. This, in turn, is based on the lower bound of Lemma 5. We have generated $2^{n(I(Y_1; V_1)+\delta)}$ quantization codewords for Relay 1, and similarly, we have generated $2^{n(I(Y_2; V_2)+\delta)}$ quantization codewords for Relay 2. Now choose $\epsilon_{\text{quant}}$ as a function of $n$ such that $\epsilon_{\text{ch}} \cdot |X| \cdot |Y_1| \leq \epsilon_{\text{quant}}$, $\epsilon_{\text{ch}} \cdot |X| \cdot |Y_2| \leq \epsilon_{\text{quant}}$, and $\epsilon_{\text{quant}} \to 0$ as $n \to \infty$. We can choose such an $\epsilon_{\text{quant}}$ because $\epsilon_{\text{ch}} \to 0$ as $n \to \infty$.

It follows directly from the achievability proof that as $\epsilon_{\text{quant}} \to 0$ and $n \to \infty$,

$$\Pr \left( \exists v_1^n \in \mathcal{C}_{q_1} \text{ s.t. } (y_1^n, v_1^n) \in \mathcal{A}_{\epsilon_{\text{quant}}}^*(Y_1, V_1) \right) \to 1, \quad (3.165)$$

$$\Pr \left( \exists v_2^n \in \mathcal{C}_{q_2} \text{ s.t. } (y_2^n, v_2^n) \in \mathcal{A}_{\epsilon_{\text{quant}}}^*(Y_2, V_2) \right) \to 1. \quad (3.166)$$

Next assume we have found such a pair of quantization codewords, $V_1^n$ and $V_2^n$. That is, for a particular jointly typical triple $(x^n, y_1^n, y_2^n)$, condition on the following event:

$$(X^n(1), Y_1^n, Y_2^n) = (x^n, y_1^n, y_2^n), \quad (3.167)$$

where

$$(x^n, y_1^n, y_2^n) \in \mathcal{A}_{\epsilon_{\text{ch}}}^*(X, Y_1, Y_2), \quad (3.168)$$

and quantization codewords $V_1^n$ and $V_2^n$ exist satisfying

$$(y_1^n, V_1^n) \in \mathcal{A}_{\epsilon_{\text{quant}}}^*(Y_1, V_1), \quad (3.169)$$

$$(y_2^n, V_2^n) \in \mathcal{A}_{\epsilon_{\text{quant}}}^*(Y_2, V_2). \quad (3.170)$$

We now show that when $\epsilon_{\text{quant}}$ is sufficiently small, then with high probability as $n \to \infty$, $(x^n, V_1^n, V_2^n) \in \mathcal{A}_{\epsilon_{\text{dec}}}^*(X, V_1, V_2)$. This is not true with probability 1.

Recall that in our notation, the function $N(a, b, c \mid x^n, y_1^n, y_2^n)$ equals the number of positions (out of $n$) where $x_i = a$, $y_{1,i} = b$, and $y_{2,i} = c$. By definition of strong typicality and from (3.168), the set of $|X| \cdot |Y_1| \cdot |Y_2|$ empirical counts, $N(a, b, c \mid}$
$x^n, y_1^n, y_2^n$, ) satisfy

$$\frac{1}{n} N(a, b, c \mid x^n, y_1^n, y_2^n) \in p_{X,Y_1,Y_2}(a, b, c) \pm \epsilon_{ch} \quad \forall \; (a, b, c) \in \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2,$$

(3.171)

$$N(a, b, c \mid x^n, y_1^n, y_2^n) = 0 \quad \text{when} \quad p_{X,Y_1,Y_2}(a, b, c) = 0.$$

(3.172)

From (3.169) and (3.170), quantization codewords $V_1^n$ and $V_2^n$ exist satisfying $(y_1^n, V_1^n) \in \mathcal{A}^{\epsilon_{\text{quant}}}_{\epsilon_{\text{dev}}}(Y_1, V_1)$ and $(y_2^n, V_2^n) \in \mathcal{A}^{\epsilon_{\text{quant}}}_{\epsilon_{\text{dev}}}(Y_2, V_2)$. We want to find the probability that $(x^n, V_1^n, V_2^n) \in \mathcal{A}^{\epsilon_{\text{dev}}}_{\epsilon_{\text{dev}}}(X, V_1, V_2)$. We start by showing that there is an $\epsilon_{01} > 0$ such that $\epsilon_{01} \to 0$ as $\epsilon_{\text{quant}} \to 0$, and such that with high probability as $n \to \infty$, $(x^n, y_1^n, y_2^n, V_1^n) \in \mathcal{A}^{\epsilon_{\text{dev}}}_{\epsilon_{\text{dev}}}(X, Y_1, Y_2, V_1)$. We show this by averaging over the set of possible $v_1^n$ jointly typical with the particular $y_1^n$. By definition, $(y_1^n, V_1^n) \in \mathcal{A}^{\epsilon_{\text{quant}}}_{\epsilon_{\text{quant}}}(Y_1, V_1)$ if there is a set of $|\mathcal{Y}_1| \cdot |\mathcal{V}_1|$ empirical counts, $N(b, d)$, satisfying

$$\frac{1}{n} N(b, d) \in p_{Y_1,V_1}(b, d) \pm \epsilon_{\text{quant}} \quad \forall \; (b, d) \in \mathcal{Y}_1 \times \mathcal{V}_1,$$

(3.173)

$$N(b, d) = 0 \quad \text{when} \quad p_{Y_1,V_1}(b, d) = 0.$$

(3.174)

We have already fixed the triplet $(x^n, y_1^n, y_2^n)$. We now further condition on the Relay 1 quantization codeword, $V_1^n$, satisfying a particular set of empirical counts $N(b, d)$. By the random choice of Relay 1 quantization codebook and the relay encoding step, for the given string $y_1^n$, every string $v_1^n$ with the same set of empirical counts $N(b, d \mid y_1^n, v_1^n)$ is equally likely to be the quantization codeword chosen by Relay 1. We now apply Lemma 6, assigning $S = N(b \mid y_1^n)$, $F = N(a, b, c \mid x^n, y_1^n, y_2^n)$, and $K = N(b, d)$. First, from (3.174),

$$N(a, b, c, d \mid x^n, y_1^n, y_2^n, V_1^n) = 0 \quad \text{when} \quad p_{X,Y_1,Y_2,V_1}(a, b, c, d) = 0,$$

(3.175)

since $p_{X,Y_1,Y_2,V_1}(a, b, c, d) = 0$ implies that $p_{Y_1,V_1}(b, d) = 0$. 


3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

Also, having conditioned on the particular set of empirical counts \( N(b, d) \), Lemma 6 states that for any \( \epsilon > 0 \),

\[
\lim_{N(b | y_n^1) \to \infty} \Pr \left( \left| \frac{N(a, b, c, d \mid x^n, y_1^n, y_2^n, V_{1}^n)}{N(b \mid y_1^n)} - \frac{N(b, d) \cdot N(a, b, c \mid x^n, y_1^n, y_2^n)}{N(b \mid y_1^n)} \right| > \epsilon \right) = 0.
\]

(3.176)

Now as \( \epsilon_{\text{quant}} \to 0 \) and \( n \to \infty \), we have

\[
\frac{1}{n} N(b \mid y_1^n) \to p_{Y_1}(b),
\]

(3.177)

\[
\frac{1}{n} N(a, b, c \mid x^n, y_1^n, y_2^n) \to p_{X,Y_1,Y_2}(a, b, c).
\]

(3.178)

For any set of empirical counts \( N(b, d) \) satisfying (3.173) and (3.174),

\[
\frac{1}{n} N(b, d) \to p_{Y_1,V_1}(b, d).
\]

(3.179)

Therefore from (3.176) and the discussion immediately preceding it,

\[
\frac{1}{n} N(a, b, c, d \mid x^n, y_1^n, y_2^n, V_{1}^n) \to p_{X,Y_1,Y_2,V_1}(a, b, c, d) \text{ in probability}
\]

\[
\forall (a, b, c, d) \in \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{V}_1,
\]

(3.180)

\[
\frac{1}{n} N(a, b, c, d \mid x^n, y_1^n, y_2^n, V_{1}^n) = 0 \text{ when } p_{X,Y_1,Y_2,V_1}(a, b, c, d) = 0.
\]

(3.181)

(3.182)

In other words,

\[
(x^n, y_1^n, y_2^n, V_{1}^n) \in \mathcal{A}_{\epsilon_{\text{quant}}}^{*}(X, Y_1, Y_2, V_1)
\]

(3.183)

with high probability as \( \epsilon_{\text{quant}} \to 0 \) and \( n \to \infty \). We technically justify (3.183) by applying continuity of the function \( g_{f,k}(v) \) of Lemma 6 in \( f, k, \) and \( v \), and by applying
the strict positivity of $g_{f,k}(v)$ at the points:

\[ f = p_{X,Y_1,Y_2}(a,b,c), \]
\[ k = p_{Y_1,V_1}(b,d), \]
\[ v \neq f \cdot k, \]
\[ \forall (a, b, c, d) \in \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{V}_1. \]

This technical argument justifies

\[ (X^n, Y_1^n, Y_2^n, V_1^n) \in \mathcal{A}_{\text{eq}}^n(X, Y_1, Y_2, V_1) \] (3.188)

with high probability as $\epsilon_{\text{quant}} \to 0$ and $n \to \infty$. The difference between (3.183) and (3.188) is that the latter, (3.188), is averaged over the input codebook ensemble and broadcast channel outcomes, whereas (3.183) is conditioned on a particular triple $(x^n, y_1^n, y_2^n)$.

An essentially identical argument shows that there is an $\epsilon_{02} > 0$ such that $\epsilon_{02} \to 0$ as $\epsilon_{\text{quant}} \to 0$, $\epsilon_{01} \to 0$, and $n \to \infty$, and such that

\[ \frac{1}{n} N(a, b, c, d, e \mid x^n, y_1^n, y_2^n, V_1^n, V_2^n) \to p_{X,Y_1,Y_2,V_1,V_2}(a,b,c,d,e) \] in probability

\[ \forall (a, b, c, d, e) \in \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{V}_1 \times \mathcal{V}_2, \] (3.189)

\[ \frac{1}{n} N(a, b, c, d, e \mid x^n, y_1^n, y_2^n, V_1^n, V_2^n) = 0 \text{ when } p_{X,Y_1,Y_2,V_1,V_2}(a,b,c,d,e) = 0. \] (3.190)

In other words,

\[ (x^n, y_1^n, y_2^n, V_1^n, V_2^n) \in \mathcal{A}_{\text{eq2}}^n(X, Y_1, Y_2, V_1, V_2) \] (3.192)

with high probability as $\epsilon_{\text{quant}} \to 0$, $\epsilon_{01} \to 0$, and $n \to \infty$. Again, we apply the same technique to technically justify (3.192), resulting in

\[ (X^n, Y_1^n, Y_2^n, V_1^n, V_2^n) \in \mathcal{A}_{\text{eq2}}^n(X, Y_1, Y_2, V_1, V_2) \] (3.193)
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

with high probability as $\epsilon_{\text{quant}} \to 0, \epsilon_{\text{ch}} \to 0$, and $n \to \infty$.

We conclude from (3.193) and Lemma 2, for any $\epsilon_{\text{dec}} > 0$, we can choose $\epsilon_{\text{ch}}$ and $\epsilon_{\text{quant}}$ such that with high probability as $n \to \infty$,

$$ (X^n, V_1^n, V_2^n) \in \mathcal{A}_{\epsilon_{\text{dec}}}^*(X, V_1, V_2). $$

(3.194)

We have shown thus far that $\Pr(E_0) \to 0$ as $\epsilon_{\text{dec}} \to 0$ and $n \to \infty$.3

The rest of the proof proceeds in a more straightforward manner. For each of the events $E_1$ through $E_4$, $(X^n(1), V_1^n, V_2^n) \in \mathcal{A}_{\epsilon_{\text{dec}}}^*(X, V_1, V_2)$ implies $(V_1^n, V_2^n)$ are jointly typical (specifically, from Lemma 2, with measure $\epsilon_{\text{dec}} \cdot |X|$). The remainder of the proof only deals with this jointly typical $(V_1^n, V_2^n)$ and the uniform binning of relay quantization codewords. Fortunately, we do not need to track the measure of typicality so carefully — we only need recognize that joint typicality implies marginal typicality (Lemma 2).

Consider event $E_1$. By the union bound and Lemma 5,

$$ \Pr(E_1) \leq (2^{nR_{\text{ach}}} - 1) \cdot \Pr\left( (X^n(2), V_1^n, V_2^n) \in \mathcal{A}_{\epsilon_{\text{dec}}}^*(X, V_1, V_2) \right) $$

(3.195)

$$ \leq 2^{-n(I(X; V_1, V_2) - R_{\text{ach}} - \epsilon_1)} $$

(3.196)

for some $\epsilon_1 > 0$, where $\epsilon_1 \to 0$ as $\epsilon_{\text{dec}} \to 0$ and $n \to \infty$. From (3.153), $\Pr(E_1) \to 0$ as $\epsilon_{\text{dec}} \to 0$ and $n \to \infty$.

We address the remaining three events only by considering the relay quantization codewords. Consider event $E_2$. If event $E_2$ occurs, then $(v'_1, V_2^n)$ are jointly typical (Lemma 2). But such another $v'_1$ with the same bin number is unlikely. For any given

---

3For the reader familiar with the Markov Lemma of Berger [6, p. 202], we could conclude that the quadruple $(X^n, Y_1^n, Y_2^n, V_1^n)$ will be jointly typical with high probability (cf. (3.180)-(3.182)). However, the Markov Lemma is not sufficient to prove that the quintuplet $(X^n, Y_1^n, Y_2^n, V_1^n, V_2^n)$ will be jointly typical with high probability (cf. (3.189)-(3.191)). Hence the need for Lemma 6. Upon a literature search, we found a variant of the Markov Lemma, stated but not proved in [7, p. 890], which is similar in nature to Lemma 6.
chapter 3. discrete memoryless parallel relay networks

typical \( V_2^n \), by the union bound and Lemma 5,

\[
\Pr \left( \exists v'_1 \in C_{v_1} \setminus V_1^n \text{ s.t. } f_1(v'_1) = f_1(V_1^n), (v'_1, V_2^n) \in A(R_1, V_2) \right)
\leq (2^{n(I(Y_1;V_1)+\delta)} - 1) \cdot 2^{-nR_3} \cdot 2^{-n(I(V_1;V_2)-\epsilon_2)}
\leq 2^{-n(R_3+I(Y_1;V_2)-I(Y_1;V_1)-\epsilon_2-\delta)}
\]

(3.197)

(3.198)

for some \( \epsilon_2 > 0 \), where \( \epsilon_2 \to 0 \) as \( \epsilon_{de} \to 0 \) and \( n \to \infty \). This is true for every possible \( V_2^n \). Therefore, from the first link rate requirement, (3.154), and since we can let \( \delta \to 0 \), \( \Pr(E_2) \to 0 \) as \( \epsilon_{de} \to 0 \) and \( n \to \infty \).

Similarly, consider event \( E_3 \). By the same reasoning,

\[
\Pr(E_3) \leq 2^{-n(R_3+I(Y_1;V_2)-I(Y_2;V_2)-\epsilon_3-\delta)}
\]

(3.199)

for some \( \epsilon_3 > 0 \), where \( \epsilon_3 \to 0 \) as \( \epsilon_{de} \to 0 \) and \( n \to \infty \). From the second link rate requirement, (3.155), and since we can let \( \delta \to 0 \), \( \Pr(E_3) \to 0 \) as \( \epsilon_{de} \to 0 \) and \( n \to \infty \).

Finally, consider event \( E_4 \). From Lemma 1 and Lemma 5, we see that the probability that independently generated relay quantization codewords are jointly typical is again approximately \( 2^{-n(I(Y_1;V_2))} \). We can thus use the same reasoning here.

\[
\Pr(E_4) \leq 2^{-n(R_3+R_4+I(Y_1;V_2)-I(Y_1;V_1)-I(Y_2;V_2)-\epsilon_4-2\delta)}
\]

(3.200)

for some \( \epsilon_4 > 0 \), where \( \epsilon_4 \to 0 \) as \( \epsilon_{de} \to 0 \) and \( n \to \infty \). From the third link rate requirement, (3.156), and since we can let \( \delta \to 0 \), \( \Pr(E_4) \to 0 \) as \( \epsilon_{de} \to 0 \) and \( n \to \infty \).

The theorem now follows from the same standard final argument we used in the proof of Theorem 3.3.1. See (3.102) and the short discussion that followed. \( \square \)

3.3.4 Reproducing a Sufficient Statistic for Estimating \( X^n \) Based on \( (Y_1^n, Y_2^n) \)

Consider an arbitrary parallel relay network, where the alphabets can be continuous or discrete, and where the multiaccess channel is general rather than consisting of a pair of noiseless binary links (see Figure 1-2). Continue to assume that the multiaccess channel is independent of the broadcast channel. If an input codebook
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

\{\phi_{k2^n}^{n_{\text{ach}}}\} can be decoded with small probability of error at the decoder, then, since the decoder observation is conditionally independent of the input given the pair of relay observations, the input codebook must be decodable based on the pair of relay observations \((Y_1^n, Y_2^n)\). Rather than attempting to faithfully reproduce the observations at the decoder, as we did in Theorem 3.3.1, or attempting to reproduce a slightly distorted version of the observations, as we did in Theorem 3.3.2, we can instead attempt to reproduce a sufficient statistic for the input \(\{x_{\delta}^n\}^{2^{n_{\text{ach}}} n_{\text{ach}}}\) based on the observations \((Y_1^n, Y_2^n)\). That is, if we can efficiently reproduce \(p(X^n | Y_1^n, Y_2^n)\) at the decoder without first reproducing \((Y_1^n, Y_2^n)\), then we would have all of the information we need to efficiently decode an arbitrary input codebook. The approach we will describe momentarily is highly tuned to the symmetric PBSC network, and a related result can be derived for a symmetric, parallel binary erasure channel (PBEC) network. We will mention the result for the PBEC network after we present that for the PBSC network.

Consider transmitting a single input symbol \(X \in \{0, 1\}\) on a symmetric PBSC broadcast channel. Then if we know the probability that \(X = 0\), and if we know whether \(Y_1 = Y_2 = 0\), \(Y_1 = Y_2 = 1\), or \(Y_1 \neq Y_2\), then we know \(p(X | Y_1, Y_2)\). The important point here is that the broadcast channel is symmetric with conditionally independent outputs \(Y_1\) and \(Y_2\). Therefore, regardless of the probability that \(X = 0\), the two events \((Y_1, Y_2) = (0, 1)\) and \((Y_1, Y_2) = (1, 0)\) are equally likely and yield the same a posteriori probability distribution on \(X\). Since we are only concerned about \(X\), we need not differentiate between \((Y_1, Y_2) = (0, 1)\) and \((Y_1, Y_2) = (1, 0)\). In this single symbol case, we can define a simple sufficient statistic for the a posteriori distribution of \(X\) given \(Y_1\) and \(Y_2\). Let \(Z_+ = Y_1 + Y_2\), where the addition is real addition, and thus \(Z_+ \in \{0, 1, 2\}\). Then \(Z_+\) contains all of the information in \(Y_1\) and \(Y_2\) relevant to \(X\).

This approach works identically for blocks as well as single symbols. That is, define \(Z_+^n = Y_1^n + Y_2^n\), where the addition is componentwise real addition, and thus \(Z_+^n \in \{0, 1, 2\}^n\). Then if we have \(Z_+^n\), we can exactly reproduce \(p(X^n | Y_1^n, Y_2^n)\). That is, for an arbitrary input codebook where the codewords are used with an

---

4It is important to point out that \(Z_+ = Y_1 + Y_2\) is sufficient for estimating \(X\) only when the PBSC broadcast channel is symmetric or degenerate (i.e., both crossover probabilities are 0 or 1).
arbitrary probability distribution,

\[ p(X^n | Z_+^n) = p(X^n | Y_1^n, Y_2^n), \tag{3.201} \]

where \( Z_+^n = Y_1^n + Y_2^n. \tag{3.202} \)

If we consider averaging over an input codebook ensemble where the codewords are generated i.i.d. using a single-letter input distribution \( p(X) \), then we induce an artificial i.i.d. correlated “observation source” \((Y_1, Y_2)\). We can make a random coding argument that proves we can achieve \( R_{ach} = I(X; Y_1, Y_2) \) provided \( R_3 \) and \( R_4 \) are sufficient to reproduce \( Z_+ \) reliably at the decoder (for an i.i.d. input \( p(X) \)).

We are now in the domain of distributed source coding, where we try to reproduce a symbolwise function of a pair of correlated variables. This problem has received some attention in the information theory literature, though with limited success (e.g., see [29], [16, p. 398–401], [3], [26]). We will consider this particular distributed source coding problem in detail, where \( X \) is i.i.d. \( p(X) = 0.5 \) and we want to reliably reproduce \( Z_+ \) at the decoder.

It follows from the principles of Slepian-Wolf compression that in order to reliably reproduce \( Z_+ \), we necessarily require

\[ R_3 \geq H(Z_+ | Y_2), \tag{3.203} \]
\[ = H(Y_1 | Y_2), \tag{3.204} \]
\[ R_4 \geq H(Z_+ | Y_1), \tag{3.205} \]
\[ = H(Y_2 | Y_1), \tag{3.206} \]
\[ R_3 + R_4 \geq H(Z_+). \tag{3.207} \]

Equalities (3.204) and (3.206) follow simply by definition of \( Z_+ \). We next evaluate these entropy terms for the symmetric PBSC broadcast channel with crossover probability \( \alpha \) and input distribution \( p(X = 0) = 0.5 \). Defining

\[ \beta = \alpha \otimes \alpha = 2 \cdot \alpha \cdot (1 - \alpha), \tag{3.208} \]
it follows that

\[
H(Z_+ | Y_2) = H(Y_1 | Y_2) = H_{\text{bin}}(\beta), \quad (3.209)
\]

\[
H(Z_+ | Y_1) = H(Y_2 | Y_1) = H_{\text{bin}}(\beta), \quad (3.210)
\]

\[
H(Z_+) = 1 + H_{\text{bin}}(\beta) - \beta, \quad (3.211)
\]

\[
I(X; Z_+) = I(X; Y_1, Y_2) = 1 + H_{\text{bin}}(\beta) - 2 \cdot H_{\text{bin}}(\alpha). \quad (3.212)
\]

We include (3.212) only to verify that our symbolwise function $Z_+$ indeed has exactly the same information about $X$ as that contained in $(Y_1, Y_2)$. Recall from (3.46) that in the symmetric PBSC network,

\[
\frac{1}{2} H(Y_1, Y_2) \geq H(Y_1 | Y_2). \quad (3.213)
\]

Comparing (3.211) with (3.48), we verify that

\[
H(Z_+) \leq H(Y_1, Y_2) \quad (3.214)
\]

with equality if and only if $\alpha = 0$ or $\alpha = 1$. That is, as long as the broadcast channel is nondegenerate, the sufficient statistic $Z_+$ has strictly smaller entropy than the pair of observations $(Y_1, Y_2)$. Therefore in principle, from (3.203)–(3.207) and (3.209)–(3.211), we could potentially reduce the rate requirement on $R_3$ and $R_4$ by reproducing $Z_+$ rather than $(Y_1, Y_2)$ at the decoder.

Unfortunately, for a PBSC network with a symmetric broadcast channel and with $p(X = 0) = 0.5$, it turns out that reproducing $Z_+$ reliably at the decoder has the same requirements on $R_3$ and $R_4$ as when reproducing $(Y_1, Y_2)$. In light of (3.204) and (3.206), to prove this we need only prove that reliably reproducing $Z_+$ at the decoder also requires

\[
R_3 + R_4 \geq H(Y_1, Y_2). \quad (3.215)
\]

From Slepian and Wolf, (3.204), (3.206), and (3.215) are the same requirements as when reproducing $(Y_1, Y_2)$. In other words, reproducing $Z_+$ at the decoder is no easier than reproducing $(Y_1, Y_2)$. We will prove (3.215) using a technique developed
by Körner [16, p. 400].

The first step in proving (3.215) is to prove that the sufficient statistic $Z_+$ is “closer” (in an appropriate sense) to $Y_1$ than the other observation $Y_2$ is to $Y_1$. Specifically, let $W$ be any random variable satisfying

$$W \rightarrow Y_1^n \rightarrow (Z_+^n, Y_2^n). \tag{3.216}$$

That is, $W$ is any random variable conditionally independent of $(Z_+^n, Y_2^n)$ when conditioned on $Y_1^n$. Then the first step is to prove

$$I(Y_1^n; Z_+^n \mid W) \geq I(Y_1^n; Y_2^n \mid W). \tag{3.217}$$

In particular, after proving (3.217), we will set $W = W_3^{nR_3}$, where $W_3^{nR_3}$ is the transmitted binary signal of Relay 1.

To prove (3.217) we use a mathematical trick. We have assumed that $X$ is generated i.i.d. $p(X = 0) = 0.5$, and this induces an i.i.d. joint distribution on $(Y_1, Y_2, Z_+)$. Consider the two memoryless channels, $A : Y_1 \rightarrow Z_+$ and $B : Y_1 \rightarrow Y_2$. Channel $A$ is equivalent to the conditional probability matrix $p(z_+ \mid y_1)$. Similarly, channel $B$ is equivalent to the conditional probability matrix $p(y_2 \mid y_1)$. We will show that channel $B$ is a stochastically degraded version of channel $A$. Consider a new random variable $Y_2'$, which is related to $Y_1$ and $Z_+$ as shown in Figure 3-12. Specifically, when defining the new random variable over a block of $n$ symbols, $Y_2'^n$ is connected to $Y_1^n$ by the memoryless channel $B' : Y_1 \rightarrow Y_2'$.

![Figure 3-12: Mathematical trick to prove (3.217)]
3.3. **CODING FOR DISCRETE SYMMETRIC NETWORKS**

By definition of channel $A$, referring to Figure 3-12, we must have

$$\beta = \alpha \otimes \alpha. \quad (3.218)$$

As indicated in the figure,

$$Y_1 \rightarrow Z_+ \rightarrow Y'_2, \quad (3.219)$$

and thus channel $B' : Y_1 \rightarrow Y'_2$ is a physically degraded version of channel $A : Y_1 \rightarrow Z_+$. The trick is to show that we can pick $\gamma$ in Figure 3-12 such that channel $B'$ is equivalent to the original channel $B$. It can be verified that choosing

$$\gamma = \frac{1}{2} \cdot \frac{\beta}{1 - \beta} \quad (3.220)$$

works (recall that $0 \leq \alpha \leq 0.5$, implying $0 \leq \beta \leq 0.5$). That is, by construction,

$$p(Y'_2 \mid Y_1) = p(Y_2 \mid Y_1). \quad (3.221)$$

We have shown that channel $B$ is a stochastically degraded version of channel $A$ (by definition of stochastic degradedness). Since the original channel $B$ and the new channel $B'$ are memoryless and equivalent, for any a priori distribution on $Y_1^n$,

$$I(Y_1^n; Y_2^n) = I(Y_1^n; Y'_2^n). \quad (3.222)$$

Also, since channel $B'$ is a physically degraded version of channel $A$, for any a priori distribution on $Y_1^n$, we have from the data processing inequality,

$$I(Y_1^n; Z_+^n) \geq I(Y_1^n; Y'_2^n). \quad (3.223)$$

Finally, consider any random variable $W$ satisfying (3.216). In terms of the joint distribution on $(W, Y_1^n, Y_2^n, Z_+^n)$, conditioning on any sample value $W = w$ simply
places an a priori distribution on $Y_1^n$. That is, with a slight abuse of notation:

$$p(Y_1^n, Y_2^n, Z_+^n \mid W = w) = p(Y_1^n \mid W = w) \cdot p(Y_2^n, Z_+^n \mid Y_1^n, W = w)$$

$$= p(Y_1^n \mid W = w) \cdot p(Y_2^n, Z_+^n \mid Y_1^n).$$

(3.224)  

(3.225)

Therefore, combining this observation with (3.222) and (3.223) yields the desired result (3.217):

$$I(Y_1^n; Z_+^n \mid W) \geq I(Y_1^n; Y_2^n \mid W).$$

(3.226)

In particular, for $W = W_3^{nR_3}$, where $W_3^{nR_3}$ is the transmitted binary signal of Relay 1,

$$I(Y_1^n; Z_+^n \mid W_3^{nR_3}) \geq I(Y_1^n; Y_2^n \mid W_3^{nR_3}).$$

(3.227)

We will now convert (3.227) into a more useful form. First note that by definition of $Z_+$,

$$H(Z_+^n \mid Y_1^n) = H(Y_2^n \mid Y_1^n).$$

(3.228)

Additionally, since $W_3^{nR_3} \rightarrow Y_1^n \rightarrow (Y_2^n, Z_+^n)$, the definition of $Z_+$ implies

$$H(Z_+^n \mid Y_1^n, W_3^{nR_3}) = H(Y_2^n \mid Y_1^n, W_3^{nR_3}).$$

(3.229)

Substituting (3.228) and (3.229) into (3.227) yields

$$H(Z_+^n \mid W_3^{nR_3}) \geq H(Y_2^n \mid W_3^{nR_3}).$$

(3.230)

This completes the first step in proving (3.215). For the second step, we will incorporate (3.230) into marginal requirements on $R_3$ and $R_4$, assuming that $Z_+^n$ is reliably
3.3. CODING FOR DISCRETE SYMMETRIC NETWORKS

 decodable at the decoder. We start with a lower bound on $R_4$.

$$nR_4 \geq H(W_4^{nR_4})$$

$$\geq H(W_4^{nR_4} | W_3^{nR_3})$$

$$= I(Z_+^n; W_4^{nR_4} | W_3^{nR_3}) + H(W_3^{nR_3}, Z_+^n)$$

$$\geq I(Z_+^n; W_4^{nR_4} | W_3^{nR_3})$$

$$= H(Z_+^n | W_3^{nR_3}) - H(Z_+^n | W_3^{nR_3}, W_4^{nR_4})$$

$$\geq H(Z_+^n | W_3^{nR_3}) - n \cdot \delta_1$$

$$\geq H(Y_2^n | W_3^{nR_3}) - n \cdot \delta_1,$$  (3.231)

where $\delta_1$ is non-negative and $\delta_1 \to 0$ as the probability of error in reproducing $Z_+^n$ at the decoder goes to 0. Inequality (3.231) follows since $W_4$ is a binary signal. Inequality (3.232) follows since conditioning reduces entropy. Inequality (3.234) follows since discrete entropy is non-negative. Inequality (3.236) follows from Fano’s inequality since $Z_+^n$ is decodable from $(W_3^{nR_3}, W_4^{nR_4})$ by assumption. Finally, inequality (3.237) follows from (3.30).

We next derive a lower bound on $R_3$.

$$nR_3 \geq H(W_3^{nR_3})$$

$$\geq H(W_3^{nR_3}) + H(Z_+^n | W_3^{nR_3}, Y_2^n) - n \cdot \delta_2$$

$$= H(W_3^{nR_3} | Z_+^n, Y_2^n) - H(Y_2^n | W_3^{nR_3}) - n \cdot \delta_2$$

$$= H(Z_+^n, Y_2^n) + H(W_3^{nR_3} | Z_+^n, Y_2^n) - H(Y_2^n | W_3^{nR_3}) - n \cdot \delta_2$$

$$\geq H(Z_+^n, Y_2^n) - H(Y_2^n | W_3^{nR_3}) - n \cdot \delta_2$$

$$= H(Y_1^n, Y_2^n) - H(Y_2^n | W_3^{nR_3}) - n \cdot \delta_2,$$  (3.38)

where $\delta_2$ is non-negative and $\delta_2 \to 0$ as the probability of error in reproducing $Z_+^n$ at the decoder goes to 0. Inequality (3.238) follows since $W_3$ is a binary signal. Inequality (3.239) follows from Fano’s inequality since $Z_+^n$ is decodable from $(W_3^{nR_3}, W_4^{nR_4})$ by assumption, and since $W_4^{nR_4}$ is a (possibly stochastic) function of $Y_2^n$. Inequality (3.243) follows since discrete entropy is non-negative. Finally, inequality (3.244)
follows from the symmetric counterpart of (3.228),
\[ H(Z_+^n \mid Y_2^n) = H(Y_1^n \mid Y_2^n). \] (3.245)

Combining (3.237) with (3.244) we find that
\[ n(R_3 + R_4) \geq H(Y_1^n, Y_2^n) - n \cdot (\delta_1 + \delta_2), \] (3.246)

where both \( \delta_1 \) and \( \delta_2 \) go to zero as the probability of error at the decoder goes to 0. We have thus proved the desired converse statement (3.215). We conclude that when \( X \) is i.i.d. \( p(X = 0) = 0.5 \), reliably reproducing \( Z_+ \) places the same constraints on \( R_3 \) and \( R_4 \) as reliably reproducing the full set of observations \( (Y_1, Y_2) \).

We can analyze the symmetric PBEC network similarly, but the converse statement, which would be analogous to our conclusion (3.246) for the PBSC network, is not quite as strong. That is, making the same argument for the PBEC network, we will not be able to conclude that reproducing the straightforward sufficient statistic (which we will define shortly) is as difficult as reproducing the full set of observations \( (Y_1, Y_2) \). We now present the analogous result for the symmetric PBEC network.

The input \( X \) is still binary, \( X \in \{0, 1\} \), while both relay observations \( Y_1 \) and \( Y_2 \) are ternary. In particular, let the ternary alphabet be \( Y_1, Y_1 \in \{0, 1, 2\} \). We identify an erasure with the outcome 1; thus \( Y_1 = 1 \) means that the Relay 1 observation was erased. In this case, we could still use \( Z_+ = Y_1 + Y_2 \) as a sufficient statistic, but there is a more natural, smaller sufficient statistic. Specifically, let \( Z_u \) be a ternary symbol, \( Z_u \in \{0, 1, 2\} \), defined as follows.

\[
Z_u = \begin{cases} 
0 & \text{if } Y_1 = 0 \text{ or } Y_2 = 0 \text{ (or both).} \\
1 & \text{if } Y_1 = 1 \text{ and } Y_2 = 1. \\
2 & \text{if } Y_1 = 2 \text{ or } Y_2 = 2 \text{ (or both).}
\end{cases} \] (3.247)

Then \( Z_u \) is a sufficient statistic, containing all of the information relevant to \( X \) contained in \( Y_1 \) and \( Y_2 \).

Proceeding analogously to the PBSC case, assume the input \( X \) is generated i.i.d. \( p(X = 0) = 0.5 \). This input induces an i.i.d. joint distribution on \( (Y_1, Y_2, Z_u) \). Consider the two channels \( A : Y_1 \to Z_u \) and \( B : Y_1 \to Y_2 \). Using the same trick, we can
show that channel $B$ is a stochastically degraded version of channel $A$. Specifically, referring to Figure 3-13, set

$$\gamma = \epsilon,$$  \hfill (3.248)

$$\delta = \frac{1}{2} (1 - \epsilon),$$  \hfill (3.249)

where $\epsilon$ is the erasure probability of the BEC’s — i.e., $\text{Prob}(Y_1 = 1) = \epsilon$. This leads to the important result that for the transmission signal from Relay 1, $W_3^n R_3$,

$$I(Y_1^n; Z^n U | W_3^n R_3) \geq I(Y_1^n; Y_2^n | W_3^n R_3).$$  \hfill (3.250)

This is analogous to (3.227) for the PBSC network. However, for a nondegenerate PBEC network and this sufficient statistic $Z_U$,

$$H(Z_U^n | Y_1^n) < H(Y_2^n | Y_2^n).$$  \hfill (3.251)

(compare with (3.228)). Consequently, the bounding approach we took for the PBSC network does not yield quite as strong a converse statement when applied to the PBEC network. Following the derivation for the PBSC network up to (3.236), we get

$$n R_4 \geq H(Z_U^n | W_3^n R_3) - n \cdot \delta_1,$$  \hfill (3.252)

where $\delta_1 \to 0$ as the probability of error at the decoder goes to 0. Following the
derivation for the PBSC network up to (3.243), we get

\[ nR_3 \geq H(Z_{1,n}^n, Y_2^n) - H(Y_2^n \mid W_3^n R_3) - n \cdot \delta_2, \tag{3.253} \]

where \( \delta_2 \to 0 \) as the probability of error at the decoder goes to 0. Putting together (3.250), (3.252), and (3.253), and ignoring \( \delta_1 \) and \( \delta_2 \), we find

\[
\begin{align*}
  n(R_3 + R_4) &\geq H(Z_{1,n}^n, Y_2^n) + H(Z_{1,n}^n \mid Y_1^n) - H(Y_2^n \mid Y_1^n) \\
  &= H(Y_1^n, Y_2^n) + H(Z_{1,n}^n \mid Y_2^n) - H(Y_1^n \mid Y_2^n) \\
  &\quad + H(Z_{1,n}^n \mid Y_1^n) - H(Y_2^n \mid Y_1^n) \\
  &= H(Y_1^n, Y_2^n) - 2 \cdot (H(Y_2^n \mid Y_1^n) - H(Z_{1,n}^n \mid Y_1^n)),
\end{align*}
\]

where (3.256) follows by symmetry since

\[
\begin{align*}
  H(Y_1^n \mid Y_2^n) &= H(Y_2^n \mid Y_1^n), \text{ and} \\
  H(Z_{1,n}^n \mid Y_1^n) &= H(Z_{1,n}^n \mid Y_2^n).
\end{align*}
\]

For the symmetric PBEC network with this sufficient statistic \( Z_{1,n} \),

\[
H(Z_{1,n}^n \mid Y_1^n) < H(Y_2^n \mid Y_1^n) \tag{3.259}
\]

whenever \( 0 < \epsilon < 1 \). \tag{3.260}

Therefore, the lower bound (3.256) is strictly smaller than \( H(Y_1^n, Y_2^n) \). With the PBEC network, then, and with this sufficient statistic \( Z_{1,n} \), we cannot conclude that it is as difficult to reproduce \( Z_{1,n} \) at the decoder as it is to reproduce the pair of relay observations \((Y_1, Y_2)\). Unfortunately, we have not been able to tighten this lower bound, nor have we succeeded in proving we can reproduce \( Z_{1,n} \) using less than \( R_3 + R_4 \approx H(Y_1, Y_2) \).

We make one final observation about the symmetric PBEC network. Any symbolwise sufficient statistic, such as \( Z_+ = Y_1 + Y_2 \) where the addition is real (and thus \( Z_+ \in \{0, 1, 2, 3, 4\} \)), can be used to construct \( Z_{1,n} \). Therefore, no other symbolwise sufficient statistic can be easier to distributively compress.

This concludes the detailed technical results of the section. Consider what we
have done from a higher level perspective. It is important to observe that for both the PBSC and the PBEC networks, our goal was to reproduce a particular sufficient statistic for estimating $X^n$ based on $(Y_1^n, Y_2^n)$. This goal is inherently lossless in the sense that the relays communicate all of the relevant information contained in $(Y_1^n, Y_2^n)$ to the decoder. Our intuition, based on applying block quantization at the relays to the symmetric PBSC network (Theorem 3.3.2), leads us to believe that lossless compression of all relevant information in $(Y_1^n, Y_2^n)$ may not be the most efficient way to communicate through the parallel relay network.

Additionally, the sufficient statistics we chose in this section were sufficient for an arbitrary input codebook $\{x^n\}_{i=1}^{n^{R_{\text{avg}}}}$. We chose them based on the specific symmetric structure of the broadcast channels. It is natural to ask how to take the particular input codebook into account. If we could manage it, we would like to process the observations at the relays so that the decoder could differentiate between the maximum likelihood decoding regions in $\mathcal{Y}_1^n \times \mathcal{Y}_2^n$ space. It may be possible to efficiently compress these maximum likelihood decoding regions for a highly structured input codebook. This is promising for a constructive algebraic approach, but it makes proving converses based on this perspective difficult. On the other hand, for achievable coding techniques, we often average over an ensemble of relatively unstructured input codebooks (as we have done throughout this thesis).\footnote{We may be able to restrict the ensemble of input codebooks to linear codes, which is sufficient to achieve capacity on a point to point BSC [20]. We did not aggressively explore this restriction — our preliminary work in this regard did not appear promising.}

We suspect that the average input codebook will not have sufficient structure to permit efficient compression of these maximum likelihood decoding regions, and this makes proving achievability results based on this perspective difficult.

\section{3.4 Capacity Equals End-to-End Mutual Information}

For conventional point to point channels, the noisy channel coding theorem shows that the maximum mutual information rate between channel input and output is
equal to the maximum rate at which data can be transmitted with arbitrarily small probability of error. This coding theorem is the fundamental result of information theory for point to point channels, since it both gives operational significance to the definition of mutual information and also separates the problem of finding the maximum mutual information from the usually much harder problem of finding codes that achieve low error probability at rates close to capacity.

In this section we will show that this same fundamental noisy channel coding theorem applies to the parallel relay network with noiseless binary relay links. We have defined capacity, \( C_{\text{net}} \), as the supremum of input rates that can be transmitted through the network with arbitrarily low error probability. Thus, this fundamental coding theorem, for these parallel relay networks, states that \( C_{\text{net}} \) is equal to the maximum mutual information rate from input to output. In the case of parallel relay networks, it appears difficult to determine the maximum mutual information rate. This theorem, however, says that if we could determine the maximum mutual information rate, then we would also know the supremum communication rate achievable with arbitrarily low error probability.

We will now prove the relevant theorem for the general parallel relay network with a pair of noiseless binary relay links to the decoder, as pictured in Figure 3-14. We will use a proof technique developed by Shannon for the discrete two-way communication channel [37, Sec. 15].

![Parallel relay network with noiseless relay channels](image)

**Figure 3-14:** Parallel relay network with noiseless relay channels

**Theorem 3.4.1.** For the parallel relay network of Figure 3-14, where the relay observations need not be conditionally independent given the input,

\[
C_{\text{net}} = \sup \frac{1}{n} I(X^n; W_3^{nR_3}, W_4^{nR_4}).
\]

(3.261)
3.4. CAPACITY EQUALS END-TO-END MUTUAL INFORMATION

The supremum in (3.261) is taken with respect to all blocklengths \( n \), all length \( n \) input codes \( \{ \tilde{x} \} \) used with an arbitrary probability distribution \( p(\tilde{x}) \), and all (possibly random) relay mappings \( M_1 : Y_1^n \rightarrow W_3^{nR_3} \) and \( M_2 : Y_2^n \rightarrow W_4^{nR_4} \), subject to the constraints

\[
(X^n, Y_2^n, W_4^{nR_4}) \rightarrow Y_1^n \rightarrow W_3^{nR_3}
\]

(3.262)

\[
(X^n, Y_1^n, W_3^{nR_3}) \rightarrow Y_2^n \rightarrow W_4^{nR_4}.
\]

(3.263)

Proof.

Remark: The two constraints (3.262) and (3.263) guarantee that if the relay transmissions are random functions of their observations, then the randomness is independent. It is not sufficient to assume

\[
(X^n, Y_2^n) \rightarrow Y_1^n \rightarrow W_3^{nR_3}
\]

(3.264)

\[
(X^n, Y_1^n) \rightarrow Y_2^n \rightarrow W_4^{nR_4}.
\]

(3.265)

As a counterexample, assume \( R_3 = R_4 = 1 \), let \( n = 1 \) and let \( X, Y_1, Y_2, \) and \( W_3 \) each be independent Bernoulli(0.5). Let \( W_4 \) be a parity check bit enforcing, say, odd parity amongst the five variables. Then (3.264) and (3.265) hold but (3.263) does not.

Converse: We begin by using Fano’s inequality to relate reliable communication rate to end-to-end mutual information. All communication must begin and terminate with a finite number \( n \) of channel uses, though \( n \) may be arbitrarily large. Let \( nR_{in} \) equal the number of information bits, denoted \( B_{in}^{nR_{in}} \), transmitted by the input terminal using \( n \) input symbols, denoted \( X^n \). The input terminal can map the \( nR_{in} \) information bits to input strings in an arbitrary way. Therefore there is an arbitrary probability distribution, \( p(\tilde{x}) \), on the length \( n \) input strings — specifically, \( p(\tilde{x}) \) is not necessarily uniform over \( 2^{nR_{in}} \) length \( n \) input strings.\(^6\) Relay 1 maps its received observation \( Y_1^n \) into a transmitted signal \( W_3^{nR_3} \), using a (possibly random) mapping \( M_1 : Y_1^n \rightarrow W_3^{nR_3} \). Similarly, Relay 2 maps its received observation \( Y_2^n \) into a transmitted signal \( W_4^{nR_4} \), using a (possibly random) mapping \( M_2 : Y_2^n \rightarrow W_4^{nR_4} \). As

\(^6\)Variable length input codes and relay mappings can be accommodated easily in this proof by using various network model definitions and by constraining the expected length of \( W_3 \) and \( W_4 \). However, variable length codes do not make much sense in the context of our discussion.
Chapter 3. **Discrete Memoryless Parallel Relay Networks**

(possibly random) functions of the observations, these mappings must satisfy

\[
(X^n, Y_2^n, W_4^n) \rightarrow Y_1^n \rightarrow W_3^n \]  
(3.266)

\[
(X^n, Y_1^n, W_3^n) \rightarrow Y_2^n \rightarrow W_4^n .
\]  
(3.267)

The decoder estimates the information bits based on its received signal \((W_3^n, W_4^n)\). Denote by \(\hat{B}_{in}^{nR_m}\) the estimated input bits. We thus have the conditional independence structure

\[
B_{in}^{nR_m} \rightarrow X^n \rightarrow (W_3^n, W_4^n) \rightarrow \hat{B}_{in}^{nR_m}.
\]  
(3.268)

With this structure, the data processing inequality [21, Th. 4.3.3] states

\[
I(B_{in}^{nR_m} ; \hat{B}_{in}^{nR_m}) \leq I(X^n, W_3^n, W_4^n).
\]  
(3.269)

Denote by \(\overline{P_{e_0}}\) the average bit error rate at the decoder. Combining (3.269) with Fano’s inequality [21, Th. 4.3.2] yields

\[
R_{in} \leq \frac{1}{n} I(X^n; W_3^n, W_4^n) + H_{bin}(\overline{P_{e_0}}).
\]  
(3.270)

For reliable communication we require \(\overline{P_{e_0}} \rightarrow 0\), and thus from (3.270) we find

\[
C_{net} \leq \sup \frac{1}{n} I(X^n; W_3^n, W_4^n).
\]  
(3.271)

The supremum in (3.271) is taken with respect to the same set of blocklength \(n\) input codes and relay mappings defined in the theorem statement.

**Achievability:** Assume there is a set of length \(n\) input strings, \(\{x_i^n\}\), with an associated probability distribution \(p(x_i^n)\). Additionally, assume there are two relay mappings \(M_1 : Y_1^n \rightarrow W_3^n\) and \(M_2 : Y_2^n \rightarrow W_4^n\) satisfying

\[
(X^n, Y_2^n, W_4^n) \rightarrow Y_1^n \rightarrow W_3^n\]  
(3.272)

\[
(X^n, Y_1^n, W_3^n) \rightarrow Y_2^n \rightarrow W_4^n .
\]  
(3.273)

The conclusion will follow when we prove that we can reliably achieve any communi-
3.4. CAPACITY EQUALS END-TO-END MUTUAL INFORMATION

In this section, we consider the problem of achieving direct communication between two points over a noisy channel. We seek to find the maximum rate at which information can be reliably transmitted between the two points. The set-up for our problem is as follows. We have two points, A and B, that wish to communicate in the presence of a noisy channel. Let $X$ be the input to the channel, $Y$ be the output, and $Z$ be the noise. The channel is given by $p(y | x, z)$.

**Proof.** We use the result of Theorem 3.41 to prove Theorem 3.42. Let $R_{ach}$ be the achievable rate. Then

$$R_{ach} < \frac{1}{n} I(X^n; W_3^{nR_3}, W_4^{nR_4}).$$

(3.274)

Pick any such rate $R_{ach}$. We will use the length $n$ input strings, $\{x_i^n\}$, as input superletters. At Relay 1, we will apply the given mapping $M_1 : Y_1^n \rightarrow W_3^{nR_3}$ with no memory between blocks of $n$ received letters. At Relay 2, we will apply the given mapping $M_2 : Y_2^n \rightarrow W_4^{nR_4}$ with no memory between blocks of $n$ received letters. We will use a random coding argument using $p(x^n)$ as a single-superletter input distribution. To make the notation easier to follow, define $\tilde{X} = X^n$, $\tilde{Y}_2 = Y_2^n$, $\tilde{W}_3 = W_3^{nR_3}$, and $\tilde{W}_4 = W_4^{nR_4}$. The single-superletter input distribution is then denoted $p(\tilde{X})$ and the relay mappings are denoted $M_1 : \tilde{Y}_1 \rightarrow \tilde{W}_3$ and $M_2 : \tilde{Y}_2 \rightarrow \tilde{W}_4$.

![Diagram](image)

Figure 3-15: Creating a point to point channel based on superletters

With this construction, the network can be used as a straightforward point to point discrete memoryless channel with input $\tilde{X}$ and decoder observation $(\tilde{W}_3, \tilde{W}_4)$, as pictured in Figure 3-15. The discrete memoryless channel transition probabilities equal

$$p(\tilde{w}_3, \tilde{w}_4 | \tilde{x}) = \sum_{\tilde{y}_1, \tilde{y}_2} p(\tilde{y}_1, \tilde{y}_2 | \tilde{x}) \cdot p(\tilde{w}_3 | \tilde{y}_1) \cdot p(\tilde{w}_4 | \tilde{y}_2),$$

where $p(\tilde{w}_3 | \tilde{y}_1)$ and $p(\tilde{w}_4 | \tilde{y}_2)$ are given by the relay mappings $M_1$ and $M_2$. It follows from Shannon’s original achievability theorem for discrete memoryless channels that $R_{ach}$ is reliably achievable.

It should be clear from the proof that Theorem 3.41 can be restated for more general networks. In particular, an analogous theorem holds when the pair of noiseless relay links is replaced by a discrete memoryless multiaccess channel $p(y | w_3, w_4)$. With such
a network, the capacity equals the supremum of the end-to-end mutual information rate \( \frac{1}{n} I(X^n; Y^n) \), where we maximize over the same set of input codes and relay mappings. The symbol rate of this multiaccess channel could also be different from the symbol rate of the broadcast channel. Additionally, Theorem 3.4.1 can be restated for the Gaussian parallel relay network of Chapter 2, where the input code and the relay mappings are constrained to satisfy the average power constraints (at the input and at each relay). A general principle thus follows that the capacity of a reasonably general set of parallel relay networks is determined by the maximum achievable end-to-end mutual information rate, a principle consistent with the fundamental coding theorem for point to point channels.

Unfortunately, determining the maximum end-to-end mutual information in parallel relay networks, either analytically or computationally, appears exceptionally difficult. In essence, Theorem 3.4.1 states that we should consider all possible input codes and all possible ways of mapping relay observations to transmission signals. It thus gives no hint of what are good and bad ideas for communicating through the network. Even if we could take a brute force approach to evaluate the maximum end-to-end mutual information for a fixed blocklength \( n \), we have found no way of quantifying or bounding the difference between the best strategy of length \( n \) and the best strategy in the limit as \( n \to \infty \). In the information theory literature, results such as Theorem 3.4.1 are usually called incomputable capacity results.

When thinking about why computing end-to-end mutual information appears difficult, two things stand out. First, the intermediate processing in the network must be done in parallel branches. As we mentioned in the introductory chapter, many such interesting distributed processing problems have proven to be notoriously difficult. Second, unlike traditional network flow problems, where information is modeled as fluid flowing through interconnected network pipes, there is no “continuity of flow” for mutual information. To make this explicit, consider three variables \( X, Y, \) and \( Z \), and assume for simplicity that

\[
X \rightarrow Y \rightarrow Z. \tag{3.276}
\]

This situation is analogous to a series connection of two links in an acyclic network, as pictured in Figure 3-16.
3.5. CODING FOR HIGHLY ASYMMETRIC NETWORKS

For a traditional network flow problem, assuming the links in Figure 3-16 do not contain any internal “sinks”, the flow on each link must be identical, and the end-to-end flow is equal to the flow on each of the links. In contrast, mutual information does not satisfy these natural, physical constraints. From the data processing inequality applied twice,

\[ I(X; Z) \leq \min [I(X; Y) ; I(Y; Z)]. \quad (3.277) \]

Even with this simple series of two links, it is possible that

\[ I(X; Y) \neq I(Y; Z), \quad (3.278) \]

and the inequality in (3.277) can be strict. For a concrete example, each link could be an additive white Gaussian noise channel (with no relay terminal between the two links). When there are parallel branches, as there are in parallel relay networks, this problem becomes more pronounced.

We will return to this issue in Section 3.6. There we will claim in Theorem 3.6.1 that to determine the capacity of some degenerate networks, we need to solve an apparently simpler end-to-end mutual information problem without parallel branches.

3.5 Coding for Highly Asymmetric Networks

We will develop several achievable coding techniques specific to highly asymmetric parallel relay networks, where Relay 1 observes the input \( X \) perfectly. We will collect all of our results for the highly asymmetric PBSC network and present them in Figure 3-18 on page 172 and again in Figure E-4 on page 226 for easy reference.
3.5.1 Coding to the Relays

Consider now the highly asymmetric PBSC network of Figure 3-17, where \( Y_1 \equiv X \).

\[
\begin{align*}
C_1 &= \max_{p(X)} I(X; Y_1) \\
&= 1 - H_{\text{bin}}(\alpha_1) \\
&= 1. \\
C_2 &= \max_{p(X)} I(X; Y_2) \\
&= 1 - H_{\text{bin}}(\alpha_2). 
\end{align*}
\]

We assume here and throughout that \( R_3 = R_4 < 1 \). For if \( R_3 = R_4 \geq 1 \), then Relay 1 can forward \( X \) directly to the receiver, and then communication is trivial. Recall our definition of the single link capacities \( C_1 \) and \( C_2 \), from (3.20) and (3.21):

\[
R_{\text{top}} \leq I(X; Y_2 \mid U) = H(X \mid U) = H_{\text{bin}}(\beta), \\
R_{\text{bot}} \leq I(U; Y_1) = 1 - H_{\text{bin}}(\alpha_2 \otimes \beta). 
\]

Recall that broadcast channel codes consist of two independent pieces of information. We encode the coarse information at rate \( R_{\text{bot}} \), which both Relay 1 and Relay 2 can reliably decode. We encode the fine information at rate \( R_{\text{top}} \), which only Relay 1 can reliably decode.

When coding to the relays for the highly asymmetric PBSC network, we use a broadcast channel code at the input. After decoding the broadcast code, Relay 1 uses
3.5. CODING FOR HIGHLY ASYMMETRIC NETWORKS

$R_{\text{bot}}$ bits per input symbol to transmit the fine information to the decoder, leaving $R_3 - R_{\text{bot}}$ bits per input symbol for Relay 1 to help send the coarse information. To maximize the total rate from input to the decoder, we want to maximize $(R_{\text{top}} + R_{\text{bot}})$ such that $(R_{\text{top}}, R_{\text{bot}})$ is in the broadcast capacity region, (3.284) and (3.285), and such that $R_{\text{top}} \leq R_3$ and $R_{\text{bot}} \leq (R_4 + R_3 - R_{\text{top}})$. Since the link to Relay 1 is perfect (and $R_3 = R_4$), we would intuitively expect that the maximizing rate pair satisfies $R_{\text{top}} = R_3$. Indeed, we prove in the appendix that there is always a maximizing rate pair with $R_{\text{top}} = R_3$. The maximizing rate pair is not necessarily unique. For example, we may have a weak multiaccess side with $2R_3 < C_2$, in which case we can choose $R_{\text{top}} = 0$ and still achieve the maximum (equal to the multiaccess cut-set bound, $2R_3$).

For this coding scheme, the resulting achievable rate is given by

$$R_{\text{ach}} = \min[2R_3; R_3 + 1 - H_{\text{bin}}^{-1}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_3))]. \quad (3.286)$$

The function $H_{\text{bin}}^{-1}(x) \in [0, 0.5]$ is the inverse of the binary entropy function, and we will always assume $x \in [0, 1]$. Recall we derived the multiaccess cut-set bound, (3.19):

$$C_{\text{net}} \leq 2R_3. \quad (3.287)$$

Then coding to the relays, with this highly asymmetric PBSC network, achieves network capacity provided $R_3 \leq R_3^*$, where

$$R_3^* = 1 - H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_3^*)). \quad (3.288)$$

We must solve this implicit equation numerically to evaluate it. We evaluate this for the same numerical example we chose before, where $C_2 = 0.5$. In this case, $\alpha_2 \approx 0.11$ and $R_3^* \approx 0.36$.

3.5.2 Binning to Communicate Observations Without Decoding at the Relays

We can again apply the technique of binning to communicate the pair of observations $(Y_1, Y_2)$ for the decoder. For the highly asymmetric PBSC network, $Y_1 \equiv X$. We
again assume $R_3 = R_4$ for simplicity. Then applying Theorem 3.3.1 and evaluating
this basic result with input distribution $p(x) = (0.5, 0.5)$, we can achieve rates up to

$$R_{ach} = I(X; Y_1, Y_2)$$
$$= H(X)$$
$$= 1,$$

provided

$$R_3 > H(X \mid Y_2)$$
$$= H(Y_2 \mid X)$$
$$= H_{bin}(\alpha_2)$$

and also provided

$$2R_3 > H(X, Y_2)$$
$$= 1 + H_{bin}(\alpha_2)$$
$$= 2 - C_2.$$
We need not do anything to turn this directly into an achievable channel coding theorem for our highly asymmetric PBSC network. For a channel code, we use all $2^n$ input codewords $X^n \in \{0,1\}^n$. The source coding with side information result states that the decoder can recover the entire input codeword with arbitrarily small probability of error. Using (3.298) and (3.299), for a fixed value of $R_4 \leq 1$, we can solve for the $R_3$ necessary to achieve $R_{\text{ach}} = 1$:

$$R_3 > H_{\text{bin}} (\alpha_2 \otimes d),$$

$$R_4 > 1 - H_{\text{bin}} (d).$$

We need not do anything to turn this directly into an achievable channel coding theorem for our highly asymmetric PBSC network. For a channel code, we use all $2^n$ input codewords $X^n \in \{0,1\}^n$. The source coding with side information result states that the decoder can recover the entire input codeword with arbitrarily small probability of error. Using (3.298) and (3.299), for a fixed value of $R_4 \leq 1$, we can solve for the $R_3$ necessary to achieve $R_{\text{ach}} = 1$:

$$R_3 > H_{\text{bin}} (\alpha_2 \otimes H_{\text{bin}}^{-1} (1 - R_4)).$$

Clearly if $R_4 > 1$, we use 1 in place of $R_4$ in (3.300).

To evaluate this result when $R_3 = R_4$, we must solve these equations implicitly for the unique $d^*$ satisfying

$$H_{\text{bin}} (\alpha_2 \otimes d^*) = 1 - H_{\text{bin}} (d^*).$$

This $d^*$ is related to the critical value $R_3^*$ we found in the previous section. Specifically, in (3.288), we determined the maximum link rate $R_3^*$ such that coding to the relays achieves network capacity. When $R_3 = R_4 \leq R_3^*$, network capacity equals $2R_3$. The relationship between that $R_3^*$ and this $d^*$ is

$$R_3^* = H_{\text{bin}} (d^*).$$
CHAPTER 3. DISCRETE MEMORYLESS PARALLEL RELAY NETWORKS

For our numerical example $C_2 = 0.5$, we can achieve rates up to

$$R_{ach} = H(X)$$

$$= 1$$

provided

$$R_3 > 1 - H_{bin} (d^*)$$

$$\approx 0.64.$$  

This approach significantly reduces the binary link rate required to achieve $R_{ach} = 1$. When reproducing both $X$ and $Y_2$, we required $R_3 > 0.75$. 

Consider next our coding approach based on quantization and binning at the relays, Theorem 3.3.2. It turns out that this theorem also says we can achieve this same point, $(R_3, R_{ach}) = (1 - H_{bin} (d^*), 1)$. We choose the input distribution $p(x) = (0.5, 0.5)$. We choose $V_1 \equiv Y_1 \equiv X$, and thus we are attempting to get a perfect representation of $Y_1 \equiv X$ to the decoder. Finally, we choose $p(v_2 | y_2)$ to be a BSC with crossover probability $d^*$. Thus we are also attempting to get $Y_2$ to the decoder with distortion $d^*$. We evaluate the relevant mutual information quantities for the theorem (equations (3.153)-(3.156)):

$$I(X; V_1, V_2) = I(X; X, V_2) = 1,$$

$$I(Y_1; V_1) = I(X; X) = 1,$$

$$I(V_1; V_2) = 1 - H_{bin} (\alpha_2 \otimes d^*),$$

$$I(Y_2; V_2) = 1 - H_{bin} (d^*).$$

Therefore, Theorem 3.3.2 shows we can achieve rates up to

$$R_{ach} = 1$$
3.5. **CODING FOR HIGHLY ASYMMETRIC NETWORKS**

provided

\[
R_3 > H_{\text{bin}}(\alpha_2 \otimes d^*),
\]

(3.312)

\[
= 1 - H_{\text{bin}}(d^*),
\]

(3.313)

\[
R_4 > H_{\text{bin}}(\alpha_2 \otimes d^* - H_{\text{bin}}(d^*),
\]

(3.314)

\[
= 1 - 2H_{\text{bin}}(d^*),
\]

(3.315)

\[
R_3 + R_4 > 1 + H_{\text{bin}}(\alpha_2 \otimes d^* - H_{\text{bin}}(d^*)
\]

(3.316)

\[
= 2(1 - H_{\text{bin}}(d^*)).
\]

(3.317)

The first and the third inequalities become tight as \( R_3 = R_4 \rightarrow 1 - H_{\text{bin}}(d^*) \approx 0.64 \). Apparently, with \( p(x) = (0.5, 0.5) \) and \( R_3 = R_4 \), it is equally difficult to describe \( X \) essentially perfectly or to describe \( X \) essentially perfectly and \( Y_2 \) with distortion \( d^* \). This begs the question (or, rather, the sanity check) whether our block quantization theorem states we can reduce the link rate requirements further by describing \( X \) essentially perfectly and \( Y_2 \) with distortion greater than \( d^* \). Let us see if this is possible. To meet \( \rho_{\text{ach}} = 1 \), we are forced to make the choices \( p(x) = (0.5, 0.5) \) and \( V_1 \equiv Y_1 \equiv X \). When we make these choices, the first inequality, \( R_3 > I(Y_1; V_1) - I(V_1; V_2) = H(X \mid V_2) \), is strictly increasing in the distortion of \( Y_2 \). Further, this inequality is becomes tight as the distortion of \( Y_2 \) approaches \( d^* \). Therefore, we cannot reduce the link rate requirements by increasing the distortion of \( Y_2 \).

Let us now consider a third perspective. We can argue from a very different point of view that this same point, \( (R_3, \rho_{\text{ach}}) = (1 - H_{\text{bin}}(d^*), 1) \), should be achievable. However, this time we make an intuitive argument rather than providing a formal proof. The approach is similar to that embodied in the block quantization and binning technique of Theorem 3.3.2. Here, however, we use a more natural and regular code structure based on coset codes. We will leave out the epsilons in this intuitive discussion. Consider applying a rate-distortion code at Relay 2 for the “observation source” \( Y_2 \). If we use an input codebook at \( X \) which looks essentially like an i.i.d. Bernoulli(0.5) stream, then the “observation source” also looks like an i.i.d. Bernoulli(0.5) stream. We have rate \( R_4 \) available at Relay 2 for a rate-distortion code. The distortion-rate function for an i.i.d. Bernoulli(0.5) source with rate \( R_4 \) is
\( D(R_4) = H_{\text{bin}}^{-1}(1 - R_4) \). Let us model the net effect of the channel and rate-distortion code at Relay 2 as the cascade of two memoryless channels. The first channel, from \( X \) to \( Y_2 \), is the given BSC(\( \alpha_2 \)). The second channel, from \( Y_2 \) to the rate-distortion reproduction vector \( W_4 \), is a BSC(\( H_{\text{bin}}^{-1}(1 - R_4) \)). Then the net effect from input \( X \) to rate-distortion reproduction vector \( W_4 \) is a BSC with crossover probability \( \alpha_2 \otimes D(R_4) = \alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4) \). We build upon this intuitive model, leaving out the formal proof.

It was proved by Elias [20] that parity check codes are sufficient to achieve capacity on a BSC. Take any good parity check code of rate \( 1 - R_3 \). Keep in mind that a parity check code that is good for one BSC can have only smaller probability of decoding error when used on a BSC with smaller crossover probability. This parity check code has exactly \( 2^{nR_3} \) cosets (translations). For the input channel code, we will use all \( 2^n \) binary input strings of length \( n \). Hence the attempted communication rate equals one bit per input symbol. We use this parity check code and its \( 2^{nR_3} \) translations to partition the set of input strings. For transmission, the input simply transmits \( n \) bits (uncoded). Relay 2 applies the rate-distortion code of rate \( R_4 \) to \( Y_2^n \). We have modeled the net effect from input \( X \) to reconstruction \( W_4 \) as a BSC with crossover probability \( \alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4) \). Relay 1, having observed the input \( X^n \) exactly, transmits the coset containing \( X^n \) to the receiver. Using the known coset, the decoder can determine \( X^n \) reliably provided the effective BSC has capacity exceeding the rate of the parity check code, \( 1 - R_3 \). Thus, this approach is successful provided

\[
1 - R_3 < 1 - H_{\text{bin}} \left( \alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4) \right), \tag{3.318}
\]

in other words, provided

\[
H_{\text{bin}} \left( \alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4) \right) < R_3. \tag{3.319}
\]

When the link rates satisfy \( R_3 = R_4 = 1 - H_{\text{bin}}(d^*) = H_{\text{bin}}(\alpha_2 \otimes d^*) \), the requirement
of (3.319) is equivalent to

\[ H_{\text{bin}} (\alpha_2 \otimes H_{\text{bin}}^{-1} (1 - R_4)) < R_3, \]  
\[ H_{\text{bin}} (\alpha_2 \otimes d^*) < R_3. \]

Therefore this construction indeed works at the point

\[ (R_3, R_{\text{ach}}) = (H_{\text{bin}} (\alpha_2 \otimes d^*), 1) \]

provided the effect of the rate-distortion code at Relay 2 can be modeled as we have described. This approach is very similar to the quantization and binning approach, which achieves the same point (cf. (3.311)-(3.317)). With this rate-distortion and parity check approach, however, we do not bin the rate-distortion representation points at Relay 2.

In Figure 3-18 we present all of our results for the highly asymmetric PBSC network with \( C_2 = 0.5 \). As we have done previously when presenting graphical results, we assume \( R_3 = R_4 \) and we increase \( R_3 \) from 0 to 1 on the horizontal axis.
Figure 3-18: Evaluation of results for the highly asymmetric PBSC network, $C_2 = 0.5$
3.5.3 Capacity of the Highly Asymmetric PBSC Network with Feedback to Relay 2

It is informative to change the network structure somewhat and continue with the same sort of coset code structure. Specifically, assume now that there is a single feedback link from the decoder to the poorer relay, Relay 2, and assume this link has capacity at least $R_3$. We can equivalently assume that Relay 2 can observe the output of Relay 1, $W_3^n$. This new network is pictured in Figure 3-19.

![Network Diagram](image)

Figure 3-19: The highly asymmetric PBSC network with feedback to Relay 2

We consider the highly asymmetric PBSC network with feedback from the decoder to Relay 2, and we continue to assume $R_3 = R_4$. We can easily prove (rather than just argue) that the following coding technique is reliably achievable. We use $2^{nR_3}$ cosets of a good parity check code, where the parity check code has rate $\min[C_2; R_4; 1 - R_3]$. Relay 1 transmits the coset which contains the input codeword $X^n$. Both the decoder and Relay 2 then know that coset. Relay 2 then decodes the input codeword and transmits the index of the codeword within the coset. In this minimum, the first term guarantees that Relay 2 can decode the parity check code. The second term guarantees that Relay 2, having decoded the proper codeword, can enumerate the members of the coset for the receiver. The third term arises because there are exactly $2^n$ binary strings of length $n$, and therefore each coset cannot have more than $2^{n(1-R_3)}$ members. With this construction, therefore, we can reliably achieve rate

$$R_{ach} = \min[1; R_3 + R_4; R_3 + C_2].$$

(3.323)

This is precisely equal to the minimum of the four converse cut-set bounds, which remain valid with this feedback link. Therefore, we have completely determined
network capacity with this feedback link. Furthermore, the code structure which can achieve network capacity is particularly appealing.

This highlights the difficulty we face without feedback. If we are sending at rate $R_{\text{ach}} = 1$, we would like to take full advantage of the fact that Relay 1 observes $X$ exactly. The best idea we have found is to use Relay 1 to partition the possible codewords into groups where the members are reasonably well separated. The larger the rate $R_3$, the fewer members per group, and therefore, the larger the separation achievable within each group. We would like to combine this action of Relay 1 with a channel decoder at Relay 2. Ideally, this decoder would be a maximum likelihood decoder for the group that Relay 1 specifies. In the scheme we used above, we did exactly this. We exploited the structure of a parity check code and its translations. We used a single channel decoder (which is a minimum Hamming distance quantizer for $Y_2^n$) and translated it appropriately based on the coset of $X^n$. When we do not have feedback, we are restricted to using a single quantizer at Relay 2, regardless of which input codeword $X^n$ is sent. Therefore, without feedback, we cannot tailor the quantizer at Relay 2 to the group specified by Relay 1. If we imagine applying the same channel decoder at Relay 2 with the wrong coset translation, then we end up with a bad channel decoder.

### 3.6 Capacity Problem for Highly Asymmetric Networks

We proved in Section 3.4 that the capacity of the general discrete parallel relay network is determined by the maximum achievable mutual information rate from input to decoder observation, maximized over all block input codes and block mappings at the relays (with arbitrary blocklength). We also claimed that computing the maximum achievable end-to-end mutual information appears quite difficult. Part of the difficulty appears to come from having to process the relay observations in parallel branches of the network. In this section, we will prove that the capacity of the highly asymmetric PBSC network can also be characterized by the solution to a similar end-to-end mutual information problem, but the new problem involves only a single
branch of the network. We first define the new problem, and we will later prove in
Theorem 3.6.1 that its solution determines (but is not equal to) the capacity of the
highly asymmetric PBSC network.

\[ X \xrightarrow{p(y_2 \mid x)} Y_2 \xrightarrow{\text{Relay}} W_4 \xrightarrow{\text{Rate } R_4} \]

Figure 3-20: Series of two channels — an end-to-end mutual information problem

Consider the simple series network of Figure 3-20, corresponding to the lower
branch of the highly asymmetric PBSC network of Figure 3-17. There is an input \( X \)
connected to a discrete memoryless input channel, \( p(y_2 \mid x) \). For the highly asymmet-
cric PBSC network, the discrete memoryless input channel is the BSC with crossover
probability \( \alpha_2 \). For simplicity we will restrict our attention to the BSC, though the
discussion can be generalized easily to other memoryless channels. The output of
this channel, \( Y_2 \), is observed by a relay – the relay is the same type of arbitrarily
complex processor that we have considered throughout this thesis. The relay is in
turn connected to a noiseless binary link that can be used at rate \( R_4 \) bits per input
symbol; the signal transmitted on the noiseless binary link is labeled \( W_4 \). Since \( Y_2 \)
is a binary signal, to avoid trivial details, we assume throughout the remainder of
this section that \( R_4 \leq 1 \). The problem is to determine the maximum achievable
end-to-end mutual information, \( \frac{1}{n} I(X^n; W_4^{nR_4}) \), as a function of the input entropy
\( \frac{1}{n} H(X^n) \). Specifically, for the BSC channel, we want to determine the function \( f(v) \)
for all \( v \in [0, 1] \), where

\[ f(v) = \sup \frac{1}{n} I(X^n; W_4^{nR_4}). \] (3.324)

The supremum in (3.324) is taken with respect to all blocklengths \( n \), all length \( n \) input
codes \( \{x^n_i\} \) used with an arbitrary probability distribution \( p(x^n_i) \), and all (possibly
random) relay mappings \( M_2 : Y^n_2 \to W_4^{nR_4} \), subject to the constraint

\[ \frac{1}{n} H(X^n) = v. \] (3.325)
As implied by Figure 3-20, every (possibly random) relay mapping must satisfy

\[ X^n \rightarrow Y_2^n \rightarrow W_4^{nR_4}. \]  \hspace{1cm} (3.326)

Denote by \( C_{\text{DMC}} \) the capacity of the discrete memoryless channel. For the BSC,

\[ C_{\text{DMC}} = 1 - H_{\text{bin}} (\alpha_2). \]  \hspace{1cm} (3.327)

It follows immediately from the data processing inequality applied on both sides of the relay that the solution to (3.324) must satisfy, for all \( 0 \leq v \leq 1 \),

\[ f(v) \leq \min [C_{\text{DMC}} ; R_4] \]
\[ = \min [1 - H_{\text{bin}} (\alpha_2) ; R_4]. \]  \hspace{1cm} (3.328)

The upper bound, (3.329), is achievable for all \( v \) less than or equal to \( C_{\text{DMC}} \), the capacity of the BSC. To achieve the upper bound arbitrarily closely, we can use input codes of rate \( v \) that are arbitrarily good for the BSC. The relay can then decode the input code. When \( v \leq R_4 \), the relay maps each decoded message into a unique index \( W_4^{nR_4} \). When \( v > R_4 \), the relay partitions the decoded messages uniformly into the \( 2^{nR_4} \) possible binary signals. Therefore,

\[ f(v) = \min [C_{\text{DMC}} ; R_4] \hspace{1cm} \forall \, v \leq C_{\text{DMC}} \]
\[ = \min [1 - H_{\text{bin}} (\alpha_2) ; R_4] \hspace{1cm} \forall \, v \leq (1 - H_{\text{bin}} (\alpha_2)). \]  \hspace{1cm} (3.330)

The solution when \( v > C_{\text{DMC}} \) remains open, which is surprising considering the simplicity of the series network in Figure 3-20. This problem forces us to ask what happens when we attempt to communicate above the capacity of the discrete memoryless channel, and more specifically, what happens to the end-to-end mutual information when processing must be done without decoding at the relay. We would intuitively expect that we must lose end-to-end mutual information as the input entropy increases beyond the capacity of the discrete memoryless channel. Unfortunately, we have been unable to quantify or even bound the loss (with one exception; we will present the solution to \( f(1) \) after the following proof). This seems to be one of the simplest and most fundamental issues we need to understand before we can understand more
general network information theory problems.

Before discussing explicit solutions to this problem further, we will now prove that \( f(v) \) determines (but is not equal to) the capacity of the highly asymmetric PBSC network. An analogous result holds when we replace the highly asymmetric broadcast channel with any broadcast channel where Relay 1 observes the input \( X \) perfectly. In Theorem 3.4.1, we characterized the capacity as the supremum mutual information rate from input to decoder. In light of that characterization, we could approach this next theorem by upper and lower bounding the end-to-end mutual information in the particular case of the highly asymmetric network. For the converse statement, we will in fact upper bound the end-to-end mutual information. For the achievability statement, however, it will be more natural to prove a coding theorem than to lower bound the maximum end-to-end mutual information.

**Theorem 3.6.1.** For the highly asymmetric PBSC network, define \( f(v) \):

\[
f(v) = \sup \frac{1}{n} I(X^n; W_4^{nR_4}).
\]

The supremum in (3.331) is taken with respect to all blocklengths \( n \), all length \( n \) input codes \( \{x^n\} \) used with an arbitrary probability distribution \( p(x^n) \), and all (possibly random) Relay 2 mappings \( M_2 : Y_2^n \rightarrow W_4^{nR_4} \), subject to the constraints

\[
\frac{1}{n} H(X^n) = v,
\]

\[
X^n \rightarrow Y_2^n \rightarrow W_4^{nR_4}.
\]

Network capacity equals

\[
C_{\text{net}} = \sup_{0 \leq v \leq 1} \min \{ v + f(v) + R_3 \}.
\]

**Proof.**

**Converse:** We proved in Theorem 3.4.1 that

\[
C_{\text{net}} \leq \sup \frac{1}{n} I(X^n; W_3^{nR_3}, W_4^{nR_4}).
\]
The supremum in (3.335) is taken with respect to the same set of blocklength $n$ input codes and relay mappings that define the function $f(v)$ in the theorem statement. We thus need only show that for a particular choice of blocklength $n$, input code $\{x_i^n\}$, and relay mapping $M_2 : Y_2^n \rightarrow W_4^{nR_4}$,

$$\frac{1}{n} I(X^n; W_3^{nR_3}, W_4^{nR_4}) \leq \min [v ; f(v) + R_3],$$  \hspace{1cm} (3.336)

where

$$v = \frac{1}{n} H(X^n).$$ \hspace{1cm} (3.337)

To show the first term in (3.336) holds,

$$I(X^n; W_3^{nR_3}, W_4^{nR_4}) = H(X^n) - H(X^n | W_3^{nR_3}, W_4^{nR_4})$$ \hspace{1cm} (3.338)

$$\leq H(X^n)$$ \hspace{1cm} (3.339)

$$= n \cdot v.$$ \hspace{1cm} (3.340)

Inequality (3.339) follows because discrete entropy is non-negative. Equality (3.340) follows by assumption (3.337). To show the second term in (3.336) holds,

$$I(X^n; W_3^{nR_3}, W_4^{nR_4}) = I(X^n; W_4^{nR_4}) + I(X^n; W_3^{nR_3} | W_4^{nR_4})$$ \hspace{1cm} (3.341)

$$\leq n \cdot f(v) + I(X^n; W_3^{nR_3} | W_4^{nR_4})$$ \hspace{1cm} (3.342)

$$= n \cdot f(v) + H(W_3^{nR_3} | W_4^{nR_4}) - H(W_3^{nR_3} | W_4^{nR_4}, X^n)$$ \hspace{1cm} (3.343)

$$\leq n \cdot f(v) + H(W_3^{nR_3} | W_4^{nR_4})$$ \hspace{1cm} (3.344)

$$\leq n \cdot f(v) + H(W_3^{nR_3})$$ \hspace{1cm} (3.345)

$$\leq n \cdot f(v) + nR_3.$$ \hspace{1cm} (3.346)

Inequality (3.342) follows by definition of $f(v)$. Inequality (3.344) follows because discrete entropy is non-negative. Inequality (3.345) follows because conditioning reduces entropy, and we have dropped the conditioning on $W_4^{nR_4}$ from the second term. Inequality (3.346) follows since $W_3$ is a binary signal.
3.6. CAPACITY PROBLEM FOR HIGHLY ASYMMETRIC NETWORKS

Dividing (3.340) and (3.346) by \( n \) verifies (3.336).

Achievability: Assume there is a set of length \( n \) input strings, \( \{x^n_i\} \), with an associated probability distribution \( p(x^n_i) \). Additionally, assume there is a relay mapping
\[
M_2 : Y_2^n \rightarrow W_4^{nR4}
\]
satisfying
\[
X^n \rightarrow Y_2^n \rightarrow W_4^{nR4}.
\]  
(3.347)

The conclusion will follow when we prove that we can reliably achieve any communication rate \( R_{ach} \) satisfying
\[
R_{ach} < \min \left[ \frac{1}{n} H(X^n) ; \frac{1}{n} I \left( X^n ; W_4^{nR4} \right) + R_3 \right].
\]  
(3.348)

Pick any such rate \( R_{ach} \). We will use the length \( n \) input strings, \( \{x^n_i\} \), as input superletters. At Relay 2, we will apply the given mapping \( M_2 : Y_2^n \rightarrow W_4^{nR4} \) with no memory between blocks of \( n \) received letters. We will develop a random coding argument using \( p(x^n_i) \) as a single-superletter input distribution. To make the notation easier to follow, define \( \tilde{X} = X^n, \tilde{Y}_2 = Y_2^n, \tilde{W}_3 = W_3^{nR3}, \) and \( \tilde{W}_4 = W_4^{nR4} \). The single-superletter input distribution is then denoted \( p(\tilde{X}) \) and the relay mapping is denoted
\[
M_2 : \tilde{Y}_2 \rightarrow \tilde{W}_4.
\]

For a randomly chosen input codebook, generate \( 2^{knR_{ach}} \) supercodewords of blocklength \( k \). Generate each letter of each supercodeword independently according to \( p(\tilde{X}) \). For Relay 1, randomly and uniformly assign each of the \( 2^{kn} \) possible input superstrings to one of \( 2^{knR_3} \) bins — denote the bin assignment function by \( f_1(\tilde{X}^k) \).

During operation, Relay 1 observes the input \( \tilde{X}^k \) exactly and transmits \( \tilde{W}_3^{knR3} = f_1(\tilde{X}^k) \), indicating the appropriate bin assignment to the decoder. Relay 2 observes \( \tilde{Y}_2^k \) and transmits \( \tilde{W}_4^k \) by applying the given mapping \( M_2 : \tilde{Y}_2 \rightarrow \tilde{W}_4 \) independently to each of the \( k \) observed superletters.

The decoder looks for the unique supercodeword that is jointly (weakly) typical with \( \tilde{W}_4^k \) and that is assigned to the bin indicated by \( \tilde{W}_3^{knR3} \).

Average Probability of Decoding Error: We compute the average probability of

\[\text{footnote text} \]
message error by averaging over the choice of input supercodebooks and Relay 1 bin assignments. We choose the supercodewords for the input supercodebook by randomly generating them independently from each other. Furthermore, we use the same probability distribution to choose each input supercodeword. Therefore, from the symmetry of these random choices, the average probability of message error for the randomly chosen input supercodebook equals the average probability of message error for the randomly chosen input supercodebook conditioned on the source transmitting the first input supercodeword. We will therefore upper bound the average probability of message error conditioned on the source transmitting the first input supercodeword. Denote by \( \mathbb{E}\{\Pr_{\text{error}}\} \) the average probability of message error.

Denote by \( \tilde{X}^k(1) \) the first randomly generated input supercodeword. Define

\[
C_{-1} = \{\bigcup_{m=2}^{2^{kn_{\text{rel}}}} \tilde{X}^k(m)\}. \tag{3.349}
\]

Next define the following two error events:

\[
E_0 = \left\{ \left( \tilde{X}^k(1), \tilde{W}_4^k \right) \notin \mathcal{A}_c^k \right\}; \tag{3.350}
\]

\( E_0 \) is the event that the correct supercodeword, \( \tilde{X}^k(1) \), and the Relay 2 transmission, \( \tilde{W}_4^k \), are not jointly typical.

\[
E_1 = \left\{ \left( \tilde{X}^k(1), \tilde{W}_4^k \right) \in \mathcal{A}_c^k, \exists \tilde{x}' \in C_{-1} \text{ such that:} \right. \\
\{ f_1(\tilde{x}') = f_1(\tilde{X}^k(1)) \text{ and } (\tilde{x}', \tilde{W}_4^k) \in \mathcal{A}_c^k. \right\} \tag{3.351}
\]

\( E_1 \) is the event that the correct supercodeword and the Relay 2 transmission are jointly typical, and that there is an incorrect supercodeword, \( \tilde{x}'^k \), assigned to the same Relay 1 bin as the correct supercodeword, that is jointly typical with \( \tilde{W}_4^k \).
3.6. Capacity Problem for Highly Asymmetric Networks

Using basic set theory and the union bound,

$$\mathbb{E}\{\text{Pr}_{\text{error}}\} = \text{Pr}(E_0 \cup E_1) \leq \text{Pr}(E_0) + \text{Pr}(E_1).$$

We will show that, for sufficiently small typicality measure $\epsilon$, both of these probabilities can be made arbitrarily small by increasing the superblocklength $k$. The same measure of typicality, $\epsilon$, will be used consistently in the argument that follows.

It follows immediately from the weak law of large numbers [15, Th. 14.2.1]

$$\text{Pr}(E_0) \to 0 \text{ as } k \to \infty.$$  (3.353)

Next we bound $\text{Pr}(E_1)$.

$$\text{Pr}(E_1) = \sum_{(\tilde{x}, \tilde{w}_4^k) \in \mathcal{A}_k^k} p_{\tilde{x}, \tilde{w}_4^k} \left( \tilde{x}^k, \tilde{w}_4^k \right) \cdot \text{Pr}\left\{ \exists \tilde{x}' \in \mathcal{C}_1, \text{ s.t. } f_1(\tilde{x}') = f_1(\tilde{x}^k), (x', \tilde{w}_4^k) \in \mathcal{A}_\epsilon^k \right\}$$  (3.354)

$$\leq \sum_{(\tilde{x}, \tilde{w}_4^k) \in \mathcal{A}_k^k} p_{\tilde{x}, \tilde{w}_4^k} \left( \tilde{x}^k, \tilde{w}_4^k \right) \cdot (2^{knR_{ach}} - 1) \cdot \text{Pr}\left\{ f_1(\tilde{X}^k(2)) = f_1(\tilde{x}^k), (\tilde{X}^k(2), \tilde{w}_4^k) \in \mathcal{A}_\epsilon^k \right\}$$  (3.355)

$$\leq \sum_{(\tilde{x}, \tilde{w}_4^k) \in \mathcal{A}_k^k} p_{\tilde{x}, \tilde{w}_4^k} \left( \tilde{x}^k, \tilde{w}_4^k \right) \cdot 2^{knR_{ach}} \cdot \sum_{\tilde{x}' \text{ s.t. } (x', \tilde{w}_4^k) \in \mathcal{A}_\epsilon^k} \text{Pr}\left\{ \tilde{X}^k(2) = x' \right\} \cdot \text{Pr}\left\{ f_1(x') = f_1(\tilde{x}^k) \right\}.$$  (3.356)

Inequality (3.355) follows from the union bound. Inequality (3.356) follows because the superstrings $\tilde{X}^k$ are randomly and independently assigned to the $2^{knR_{ach}}$ Relay 1 bins, independent from the choice of input supercodewords. It would be an equality if we had not dropped the minus one from the middle term.

Consider the terms in the inner summation of (3.356). For every $\tilde{x}'$ such that $(\tilde{x}', \tilde{w}_4^k) \in \mathcal{A}_\epsilon^k$, we know from the definition of (weakly) typical sequences, (3.59),
that

\[
\Pr \left\{ \tilde{X}^k(2) = \tilde{x}' \right\} \leq 2^{-k(H(\tilde{X})-\epsilon)}. \tag{3.357}
\]

Additionally, by definition, the number of \( \tilde{x}' \) sequences such that \((\tilde{x}', \tilde{w}_4^k) \in \mathcal{A}_e^k \) is denoted \( |\mathcal{A}_e^k(\tilde{X}^k | \tilde{w}_4^k)| \). For sufficiently large superblocklength \( k \), and for \( \tilde{w}_4^k \in \mathcal{A}_e^k(\tilde{W}_4^k) \), it follows as a straightforward consequence of the definition of weak typicality that [15, Th. 14.2.2]

\[
|\mathcal{A}_e^k(\tilde{X}^k | \tilde{w}_4^k)| \leq 2^{k(H(\tilde{X}|\tilde{W}_4)+2\epsilon)}. \tag{3.358}
\]

Finally, for all \( \tilde{x}' \neq \tilde{x}^k \),

\[
\Pr \left\{ f_1(\tilde{x}') = f_1(\tilde{x}^k) \right\} = 2^{-knR_3}. \tag{3.359}
\]

However, one of the terms of the inner summation in (3.356) is \( \tilde{x}' = \tilde{x}^k \), corresponding to the second supercodecword being precisely the same as the first supercodecword. In this case they two supercodecwords, being identical, are assigned to the same Relay 1 bin. Therefore, putting together (3.356)-(3.359) and making note of the case \( \tilde{x}' = \tilde{x}^k \),

\[
\Pr (E_1) \leq \sum_{(\tilde{x}^k, \tilde{w}_4^k) \in \mathcal{A}_e^k} p_{\tilde{X}^k, \tilde{W}_4^k} \left( \tilde{x}^k, \tilde{w}_4^k \right) \cdot 2^{knR_{ach}} \cdot 2^{-k(H(\tilde{X})-\epsilon)}
\cdot \left( \left| \mathcal{A}_e^k(\tilde{X}^k | \tilde{w}_4^k) \right| - 1 \right) \cdot 2^{-knR_3} + 1
\]

\[
\leq \sum_{(\tilde{x}^k, \tilde{w}_4^k) \in \mathcal{A}_e^k} p_{\tilde{X}^k, \tilde{W}_4^k} \left( \tilde{x}^k, \tilde{w}_4^k \right) \cdot 2^{knR_{ach}} \cdot 2^{-k(H(\tilde{X})-\epsilon)}
\cdot \left( 2^{k(H(\tilde{X}|\tilde{W}_4)-nR_3+2\epsilon)} + 1 \right)
\]

\[
= 2^{-k(U(\tilde{X};\tilde{W}_4)+nR_3-nR_{ach}-3\epsilon)} + 2^{-k(H(\tilde{X})-nR_{ach}-\epsilon)} \tag{3.362}
\]

Consequently, from (3.362),

\[
\Pr(E_1) \rightarrow 0 \text{ as } k \rightarrow \infty \tag{3.363}
\]
provided

$$R_{ach} < \frac{1}{n} I(\tilde{X}; \tilde{W}_4) + R_3 - \frac{3\epsilon}{n}$$  \hspace{1cm} (3.364)$$

$$= \frac{1}{n} I(X^n; W_4^n) + R_3 - \frac{3\epsilon}{n}$$  \hspace{1cm} (3.365)$$

and

$$R_{ach} < \frac{1}{n} H(\tilde{X}) - \frac{\epsilon}{n}$$  \hspace{1cm} (3.366)$$

$$= \frac{1}{n} H(X^n) - \frac{\epsilon}{n}$$  \hspace{1cm} (3.367)$$

The result now follows from a standard argument. Pick any positive probability of error, $p_e > 0$. From (3.352), (3.353), (3.363), (3.365), and (3.367), there is a sufficiently small $\epsilon$ and sufficiently large $k$ such that

$$\mathbb{E}\{Pr_{\text{error}}\} < p_e.$$  \hspace{1cm} (3.368)$$

$\mathbb{E}\{Pr_{\text{error}}\}$ is computed by averaging over the random choice of input supercodebooks and Relay 1 bin assignments. Since the average probability of error is less than $p_e$, there must be at least one deterministic input supercodebook and a deterministic Relay 1 bin assignment with average probability of message error less than $p_e$. The theorem follows since this holds for any $p_e > 0$. \hfill $\Box$

Theorem 3.6.1 can be restated analogously when we replace the noiseless binary links of rates $R_3$ and $R_4$ with independent discrete memoryless channels of capacity $R_3$ and $R_4$, respectively. Achievability follows immediately if Relay 1 uses reliable codes of rate (arbitrarily close to) $R_3$. The converse follows with a simple adjustment after (3.341):

$$I(X^n; W_3^n \mid W_4^n) = H(W_3^n \mid W_4^n) - H(W_3^n \mid X^n, W_4^n)$$  \hspace{1cm} (3.369)$$

$$\leq H(W_3^n) - H(W_3^n \mid X^n, W_4^n)$$  \hspace{1cm} (3.370)$$

$$= H(W_3^n) - H(W_3^n \mid X^n)$$  \hspace{1cm} (3.371)$$
where (3.371) follows because \( W_3 \) is a (possibly random) function of \( X^n \), conditionally independent of \( Y_2^n \) and \( W_4^n \). It is not obvious, however, what happens when we replace the pair of noiseless binary links with a discrete multiaccess channel.

Comparing Theorem 3.4.1 with Theorem 3.6.1, we have just shown that, optimized over the same set of input codes and relay mappings,

\[
\sup_n \frac{1}{n} I(X^n; W_3^{nR_3}, W_4^{nR_4}) = \sup_{0 \leq v \leq 1} \min \left[ v ; f(v) + R_3 \right].
\]  

(3.374)

It is peculiar that to show (3.374), it is easier to construct a random coding theorem than to characterize end-to-end mutual information directly.

Return now to the problem we formulated in the beginning of the section, determining \( f(v) \) for the highly asymmetric PBSC network. Recall the definition of the function \( f(v) \):

\[
f(v) = \sup_n \frac{1}{n} I(X^n; W_4^{nR_4}).
\]  

(3.375)

To repeat the conditions of the problem, the supremum in (3.375) is taken with respect to all blocklengths \( n \), all length \( n \) input codes \( \{x_i^n\} \) used with an arbitrary probability distribution \( p(x_i^n) \), and all (possibly random) relay mappings \( M_2 : Y_2^n \rightarrow W_4^{nR_4} \), subject to the constraints

\[
\frac{1}{n} H(X^n) = v,
\]  

(3.376)

\[
X^n \rightarrow Y_2^n \rightarrow W_4^{nR_4}.
\]  

(3.377)

For a fixed input entropy \( \frac{1}{n} H(X^n) = v \), and since

\[
I(X^n; W_4^{nR_4}) = H(X^n) - H(X^n \mid W_4^{nR_4}),
\]  

(3.378)
the problem is equivalent to minimizing the conditional entropy

$$\inf \frac{1}{n} H(X^n \mid W_4^{nR_4})$$

(3.379)

subject to the same set of constraints. As we mentioned earlier in this section, this problem is open for all $v > C_{\text{DMC}}$, with one exception; we will now determine $f(1)$. Specifically, we will show that

$$f(1) = 1 - H_{\text{bin}} \left( \alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4) \right). \quad (3.380)$$

In order to show (3.380), we need to determine

$$\inf \frac{1}{n} H(X^n \mid W_4^{nR_4})$$

(3.381)

when

$$\frac{1}{n} H(X^n) = v = 1. \quad (3.382)$$

We now lower bound (3.381), which in turn upper bounds $f(\cdot)$. When $v = 1$, for every blocklength $n$, the input $X^n$ must be i.i.d. Bernoulli(0.5). This is because the i.i.d. Bernoulli(0.5) distribution uniquely maximizes the input entropy $\frac{1}{n} H(X^n)$ at 1. Assuming $X^n$ is i.i.d. Bernoulli(0.5) implies that $Y^n$ is also i.i.d. Bernoulli(0.5),

$$H(Y^n) = n. \quad (3.383)$$

Additionally, since $W_4$ is a binary signal of rate $R_4$ bits per input symbol,

$$I(Y^n; W_4^{nR_4}) \leq n \cdot R_4. \quad (3.384)$$

Combining (3.383) and (3.384) implies

$$H(Y_2^n \mid W_4^{nR_4}) \geq n \cdot (1 - R_4). \quad (3.385)$$

Instead of solving (3.381) subject to the given constraints, we now define a new problem by loosening the constraints on $W_4$. We will thus lower bound the answer to
the original problem by solving the new problem. We will then prove that the lower bound is tight for the original problem. To define the new problem, we incorporate the fact from (3.384) that \( I(Y^n; W_4^n) \leq n \cdot R_4 \), and then we throw away the cardinality constraint on \( W_4^n \) by replacing it with a new variable \( \widehat{W}_4 \). For the new problem, we maintain

\[
\frac{1}{n} H(X^n) = 1.
\]  

(3.386)

We now solve for

\[
\inf H(X^n \mid \widehat{W}_4),
\]  

(3.387)

minimizing over all blocklengths \( n \) and all random variables \( \widehat{W}_4 \) satisfying

\[
X^n \rightarrow Y_2^n \rightarrow \widehat{W}_4
\]  

(3.388)

and

\[
H(Y_2^n \mid \widehat{W}_4) \geq n \cdot (1 - R_4).
\]  

(3.389)

The solution to the new problem (3.387) in our special case where \( H(X^n) = n \) follows from a result of Wyner and Ziv [50, Cor. 4], the solution being\(^8\)

\[
H(X^n \mid \widehat{W}_4) \geq n \cdot H(\text{bin}) \left( \alpha \otimes H(\text{bin})^{-1} (1 - R_4) \right).
\]  

(3.390)

To be explicit, the lower bound in (3.390) is achievable for a particular \( \widehat{W}_4 \). Putting things together, since the new problem minimizes \( H(X \mid W_4) \) over a larger (more

---

\(^8\)Wyner and Ziv considered what happens to the conditional entropy when an input variable is passed through a channel (also see [51], [46], and [5]). When \( H(X^n) = n \) we can use their approach only because we can “flip” the BSC. We can consider \( Y^n \) as the input and \( X^n \) as the output of an equivalent BSC — the joint statistics of \( (X^n, Y^n) \) are identical in both cases, and this singularity occurs only when \( X^n \) is i.i.d. Bernoulli(0.5) or when the BSC is noiseless.
general) constraint set,

$$\inf H(X^n | W_4^n) \geq \inf H(X^n | \hat{W}_4)$$

$$\geq n \cdot H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4)). \quad (3.392)$$

Substituting (3.392) into the definition of $f(v)$ at $v = 1$, from (3.375),

$$f(1) \leq 1 - H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4)). \quad (3.393)$$

We next prove the complementary lower bound,

$$f(1) \geq 1 - H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4)). \quad (3.394)$$

To do this, we must show that the upper bound of (3.393) is achievable (arbitrarily closely) for a particular $W_4^n$, as opposed to being achievable for the more general $\hat{W}_4$. We can achieve this upper bound by using (at the relay) a rate-distortion code of rate $R_4$ designed for $Y^n$. The distortion-rate function $D_{\text{B}(0.5)}(R)$ and the rate-distortion function $R_{\text{B}(0.5)}(D)$ for an i.i.d. Bernoulli(0.5) source are [15, Th. 13.3.1]

$$D_{\text{B}(0.5)}(R) = H_{\text{bin}}^{-1}(1 - R) \quad (3.395)$$

$$R_{\text{B}(0.5)}(D) = 1 - H_{\text{bin}}(D). \quad (3.396)$$

In particular, since $Y^n$ is i.i.d. Bernoulli(0.5), we can achieve a bitwise error probability (i.e., Hamming distortion) on $Y^n$ arbitrarily close to

$$D_Y = H_{\text{bin}}^{-1}(1 - R_4). \quad (3.397)$$

Using the best estimate of $Y^n$ as the best estimate of $X^n$, we achieve a bitwise error probability (i.e., Hamming distortion) on $X^n$ arbitrarily close to

$$D_X = \alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4). \quad (3.398)$$

One piece of the traditional rate-distortion theorem states that if we can achieve an average distortion of $D_X$, then the end-to-end mutual information is lower bounded
by the rate-distortion function evaluated at distortion $D_X$ [21, Th. 9.2.1]. Therefore it follows from (3.396) and the rate-distortion theorem that

$$\sup \frac{1}{n} I(X^n; W_4^{nR_4}) \geq R_{B(0.5)}(D_X) \geq 1 - H_{\text{bin}} \left( \alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4) \right). \quad (3.399)$$

Therefore by definition of $f(v)$ in (3.375),

$$f(1) \geq 1 - H_{\text{bin}} \left( \alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4) \right). \quad (3.400)$$

Combining the upper bound of (3.393) with the lower bound of (3.401), we conclude

$$f(1) = 1 - H_{\text{bin}} \left( \alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4) \right), \quad (3.402)$$

as we had claimed. Incidentally, if we were interested in minimizing the end-to-end Hamming distortion of $X$ when $X$ is i.i.d. Bernoulli(0.5), then applying (at the relay) rate-distortion codes designed for $Y$, as we did above, minimizes the end-to-end achievable distortion of $X$ (see [36], [47]).

After some thought about the BSC channel, we suspect that $f(v)$ is continuous at $v = 1$, though we have been unable to prove it.

**Conjecture 1.** $\lim_{v \to 1} f(v) = f(1)$.

If our conjecture is correct, then for the highly asymmetric PBSC network,

$$C_{\text{net}} = 1 \quad (3.403)$$

if and only if

$$R_3 \geq 1 - H_{\text{bin}} \left( \alpha_2 \otimes H_{\text{bin}}^{-1}(1 - R_4) \right). \quad (3.404)$$

Before we show this implication, refer to (3.403)-(3.404) in the context of our graphical results in Figure 3-18. For the highly asymmetric network, the broadcast cut-set
bound (3.13) states

\[ C_{\text{net}} \leq 1. \] \hspace{1cm} (3.405)

We demonstrated, using three different approaches, that we can achieve this upper bound to capacity provided

\[ R_3 \geq 1 - R_3^*, \] \hspace{1cm} (3.406)

where \( R_3^* \) is the fixed point solution to (3.404) (that is, when \( R_3 = R_4 \) and equality holds in (3.404)). Therefore Conjecture 1 would imply that we have found the smallest rate \( R_3 \) such that \( C_{\text{net}} = 1 \). In other words, the capacity curve must diverge from the cut-set bound as we decrease \( R_3 \) below \( (1 - R_3^*) \).

We proceed to show that Conjecture 1 implies (3.403)–(3.404). From Theorem 3.6.1, network capacity equals

\[ C_{\text{net}} = \sup_{0 \leq v \leq 1} \min [v ; f(v) + R_3]. \] \hspace{1cm} (3.407)

Therefore, considering the first term in (3.407), we can achieve

\[ C_{\text{net}} \to 1 \] \hspace{1cm} (3.408)

only by insisting

\[ \frac{1}{n} H(X^n) = v \to 1. \] \hspace{1cm} (3.409)

If our continuity assumption holds, that is, if

\[ \lim_{v \to 1} f(v) = f(1), \] \hspace{1cm} (3.410)
then for any sequence $v \to 1$,

$$\limsup_{v \to 1} f(v) \leq f(1) \quad (3.411)$$

$$= 1 - \text{H}_{\text{bin}} \left( \alpha_2 \otimes \text{H}_{\text{bin}}^{-1} (1 - R_4) \right). \quad (3.412)$$

Then as $v \to 1$, from (3.407) and (3.412), we would have

$$C_{\text{net}} \leq R_3 + 1 - \text{H}_{\text{bin}} \left( \alpha_2 \otimes \text{H}_{\text{bin}}^{-1} (1 - R_4) \right), \quad (3.413)$$

and, in particular,

$$C_{\text{net}} < 1 \quad (3.414)$$

whenever

$$R_3 < \text{H}_{\text{bin}} \left( \alpha_2 \otimes \text{H}_{\text{bin}}^{-1} (1 - R_4) \right). \quad (3.415)$$

On the other hand, we have already shown

$$C_{\text{net}} = 1 \quad (3.416)$$

when

$$R_3 \geq \text{H}_{\text{bin}} \left( \alpha_2 \otimes \text{H}_{\text{bin}}^{-1} (1 - R_4) \right). \quad (3.417)$$

We showed this in (3.300), where we developed a communication code based on source coding with side information results.
Appendix A

Extended Gaussian multiaccess converse

Theorem A.0.2 (Converse for Theorem 2.1.1). Let $R_{\text{top}}$ be the rate of Message 1 in bits per channel symbol, and let $R_{\text{bot}}$ be the rate of Message 2. Message 1 and Message 2 are independent. The capacity region of the two-user Gaussian multiaccess channel, where User 1 knows both Message 1 and Message 2, while User 2 knows only Message 2, is $\mathcal{B}\{S_3, S_4\}$, where

$$\mathcal{B}\{S_3, S_4\} = \left\{ (R_{\text{top}}, R_{\text{bot}}) \, \mid \, \exists \beta \in [0, 1]: \begin{align*} R_{\text{top}} &\leq \frac{1}{2} \log_2 (1 + (1 - \beta)S_3) \\ R_{\text{top}} + R_{\text{bot}} &\leq \frac{1}{2} \log_2 \left( 1 + S_3 + S_4 + 2\sqrt{\beta S_3 S_4} \right) \end{align*} \right\}. \quad (A.1)$$

Proof.

Achievability: Already shown in Section 2.1.

Converse: Assume there are two sources, $U_t$ and $U_b$, to be transmitted reliably to the receiver. Let $U_t$ be defined on a finite alphabet of size $M_t$. Let $U_b$ be defined on a finite alphabet of size $M_b$. Assume that Relay 1 observes both $U_t$ and $U_b$, while Relay 2 observes only $U_b$. Relay 1 encodes $L_t$ symbols from $U_t$, denoted $U_t^{L_t}$, and $L_b$ symbols from $U_b$, denoted $U_b^{L_b}$, into its input codeword, denoted $W_3^n$. Therefore $W_3^n$ is a function of both $U_t^{L_t}$ and $U_b^{L_b}$. Relay 2 encodes $U_b^{L_b}$ into its input codeword, $W_4^n$. Therefore $W_4^n$ is a function of $U_b^{L_b}$ only. The two input codewords, $W_3^n$ and
$W_4^n$, are sent simultaneously using $n$ input channel uses. At the receiver, let $V_{t}^{L_t}(Y^n)$ be an estimator for $U_t^{L_t}$ based only on the received observations, $Y^n$. Similarly, let $V_{b}^{L_b}(Y^n)$ be an estimator for $U_b^{L_b}$ based only on the received observations, $Y^n$. Denote the entropy of the source symbols, in bits per input channel use, by $\frac{1}{n}H(U_t^{L_t})$ and $\frac{1}{n}H(U_b^{L_b})$. Then

\begin{align*}
R_{\text{top}} &= \frac{1}{n}H(U_t^{L_t}), \\
R_{\text{bot}} &= \frac{1}{n}H(U_b^{L_b}).
\end{align*}

(A.2)

We will prove that the point $(\frac{1}{n}H(U_t^{L_t}), \frac{1}{n}H(U_b^{L_b}))$ must be within or must get arbitrarily close to $B\{R_{\text{top}}, R_{\text{bot}}\}$ as the average symbolwise probability of estimation error for both source estimators gets arbitrarily small.

To prove this, we apply Fano’s inequality to derive two upper bounds on the entropy rates. Then we proceed with manipulation of mutual information and second order statistics. In the first of the two applications, we use Fano’s inequality in a slightly unconventional way. Before we proceed, we must define notation for various probabilities of error.

Denote by $P_{e,t,w_3}$ the average probability of symbol error for the estimator $V_t^{L_t}$ conditioned on Relay 2 transmitting the particular codeword $W_4^n = w_4^n$. Denote by $P_{e,t}$ the average probability of symbol error for $V_t^{L_t}$ (not conditioned on $W_4$). Denote by $P_{e,b}$ the average probability of symbol error for the estimator $V_b^{L_b}$.

Conditioned on Relay 2 transmitting $W_4^n = w_4^n$, Fano’s inequality states [21, Th. 4.3.2]

\begin{equation}
\frac{1}{L_t} H(U_t^{L_t} | V_t^{L_t}, W_4^n = w_4^n) \leq P_{e,t,w_3} \cdot \log_2(M_t - 1) + H_{\text{bin}}(P_{e,t,w_3}). 
\end{equation}

(A.4)

Averaging this over $W_4^n$, and using Jensen’s inequality and the convexity of $H_{\text{bin}}(\cdot)$, we get

\begin{equation}
\frac{1}{L_t} H(U_t^{L_t} | V_t^{L_t}, W_4^n) \leq P_{e,t} \cdot \log_2(M_t - 1) + H_{\text{bin}}(P_{e,t}).
\end{equation}

(A.5)
But we also have

\[
H \left(U_t^{L_t} \mid V_t^{L_t}, W_4^n\right) = H \left(U_t^{L_t} \mid W_4^n\right) - I \left(U_t^{L_t}; V_t^{L_t} \mid W_4^n\right) \tag{A.6}
\]

\[
= H \left(U_t^{L_t}\right) - I \left(U_t^{L_t}; V_t^{L_t} \mid W_4^n\right) \tag{A.7}
\]

\[
\geq H \left(U_t^{L_t}\right) - I \left(U_t^{L_t}; Y^n \mid W_4^n\right) \tag{A.8}
\]

\[
\geq H \left(U_t^{L_t}\right) - I \left(W_3^n; Y^n \mid W_4^n\right). \tag{A.9}
\]

Equality (A.7) follows since $U_t^{L_t}$ and $W_4^n$ are independent, (A.8) follows by the data processing inequality since $U_t^{L_t}$ and $V_t^{L_t}$ are conditionally independent given $Y^n$, and (A.9) follows since $H(Y^n \mid W_4^n, U_t^{L_t}) \geq H(Y^n \mid W_4^n, W_3^n)$. Combining (A.5) and (A.9) we find

\[
\frac{1}{n} H \left(U_t^{L_t}\right) \leq \frac{1}{n} I \left(W_3^n; Y^n \mid W_4^n\right) + \frac{n}{L_t} \left(P_{e,t} \cdot \log_2 (M_t - 1) + H_{\text{bin}} \left(P_{e,t}\right)\right). \tag{A.10}
\]

Next, we upper bound the sum of the entropy rates with a similar inequality. We use Fano’s inequality twice, once for each source estimator.

\[
\frac{1}{n} \left(H \left(U_t^{L_t} \mid V_t^{L_t}\right) + H \left(U_b^{L_b} \mid V_b^{L_b}\right)\right) \leq \frac{n}{L_t} \left(P_{e,t} \cdot \log_2 (M_t - 1) + H_{\text{bin}} \left(P_{e,t}\right)\right) + \frac{n}{L_b} \left(P_{e,b} \cdot \log_2 (M_b - 1) + H_{\text{bin}} \left(P_{e,b}\right)\right). \tag{A.11}
\]

But we also have

\[
H \left(U_t^{L_t} \mid V_t^{L_t}\right) + H \left(U_b^{L_b} \mid V_b^{L_b}\right) \geq H \left(U_t^{L_t} \mid Y^n\right) + H \left(U_b^{L_b} \mid Y^n\right) \tag{A.12}
\]

\[
\geq H \left(U_t^{L_t}, U_b^{L_b} \mid Y^n\right) \tag{A.13}
\]

\[
= H \left(U_t^{L_t}, U_b^{L_b}\right) - I \left(U_t^{L_t}, U_b^{L_b}; Y^n\right) \tag{A.14}
\]

\[
\geq H \left(U_t^{L_t}, U_b^{L_b}\right) - I \left(W_3^n, W_4^n; Y^n\right). \tag{A.15}
\]

\[
= H \left(U_t^{L_t}\right) + H \left(U_b^{L_b}\right) - I \left(W_3^n, W_4^n; Y^n\right). \tag{A.16}
\]

Inequality (A.12) follows since $V_t^{L_t}$ and $V_b^{L_b}$ are functions of $Y^n$, (A.13) follows from the chain rule and since conditioning reduces entropy, (A.15) follows from the data
processing inequality, and (A.16) follows since $U_i$ and $U_b$ are independent. Combining (A.11) and (A.16) we get

$$\frac{1}{n} (H(U_i^{L_i}) + H(U_b^{L_b})) \leq \frac{1}{n} I(W_3^n, W_4^n; Y^n) + \frac{n}{L_t} (P_{e,i} \cdot \log_2 (M_t - 1) + H_{\text{bin}}(P_{e,i}))$$

$$+ \frac{n}{L_b} (P_{e,b} \cdot \log_2 (M_b - 1) + H_{\text{bin}}(P_{e,b})).$$  \hfill (A.17)

Next we upper bound the block mutual information terms in (A.10) and (A.17) with their symbol-by-symbol counterparts to enforce the average power constraints. Referring to (A.10),

$$I(W_3^n; Y^n | W_4^n) = \sum_{i=1}^n I(W_3^n; Y_i | W_4^n, Y^{i-1})$$  \hfill (A.18)

$$= \sum_{i=1}^n H(Y_i | W_4^n, Y^{i-1}) - H(Y_i | W_3^n, W_4^n, Y^{i-1})$$  \hfill (A.19)

$$= \sum_{i=1}^n H(Y_i | W_4^n, Y^{i-1}) - H(Y_i | W_3,i, W_4,i)$$  \hfill (A.20)

$$\leq \sum_{i=1}^n H(Y_i | W_4,i) - H(Y_i | W_3,i, W_4,i)$$  \hfill (A.21)

$$= \sum_{i=1}^n I(W_3,i; Y_i | W_4,i).$$  \hfill (A.22)

Similarly, referring to (A.17),

$$I(W_3^n, W_4^n; Y^n) = \sum_{i=1}^n I(W_3^n, W_4^n; Y_i | Y^{i-1})$$  \hfill (A.23)

$$= \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | W_3^n, W_4^n, Y^{i-1})$$  \hfill (A.24)

$$= \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | W_3,i, W_4,i).$$  \hfill (A.25)
\[ \leq \sum_{i=1}^{n} H(Y_i) - H(Y_i \mid W_{3,i}, W_{4,i}) \tag{A.26} \]
\[ = \sum_{i=1}^{n} I(W_{3,i}, W_{4,i}; Y_i) . \tag{A.27} \]

Equalities (A.20) and (A.25) follow since \( Y_i \) is conditionally independent of previous inputs and outputs conditioned on \( (W_{3,i}, W_{4,i}) \), while (A.21) and (A.26) follow since conditioning reduces entropy.

Now observe that to maximize (A.22), we need \( W_{3,i} \) and \( W_{4,i} \) to be independent. Conversely, to maximize (A.27), we need \( W_{3,i} \) and \( W_{4,i} \) to be scaled versions of each other. Since we have a Gaussian channel, it turns out to be sufficient to consider only the second moment relationship between these two signals. This allows us to enforce the average power constraints. The following series of manipulations is borrowed from Cover and El Gamal [13]. Continuing with the derivation,

\[ I(W_{3,i} ; Y_i \mid W_{4,i}) = H(Y_i \mid W_{4,i}) - H(Y_i \mid W_{3,i}, W_{4,i}) \tag{A.28} \]
\[ \leq \mathbb{E}_{W_{4,i}} \left\{ \frac{1}{2} \log_2 \left( \frac{\text{var}(Y_i \mid W_{4,i})}{N_Z} \right) \right\} \tag{A.29} \]
\[ \leq \frac{1}{2} \log_2 \left( \frac{\mathbb{E}_{W_{4,i}} \{ \text{var}(Y_i \mid W_{4,i}) \}}{N_Z} \right) \tag{A.30} \]
\[ = \frac{1}{2} \log_2 \left( 1 + \frac{\mathbb{E}_{W_{4,i}} \{ \text{var}(W_{3,i} \mid W_{4,i}) \}}{N_Z} \right) , \tag{A.31} \]

and therefore,

\[ \frac{1}{n} \sum_{i=1}^{n} I(W_{3,i} ; Y_i \mid W_{4,i}) \leq \frac{1}{2} \log_2 \left( 1 + \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_{4,i}} \{ \text{var}(W_{3,i} \mid W_{4,i}) \} \right) \tag{A.32} \]

Inequality (A.29) follows because the Gaussian variable maximizes differential entropy subject to a variance constraint, (A.30) follows by the concavity of the logarithm and Jensen’s inequality, and (A.32) follows again by the concavity of the logarithm and
Jensen’s inequality. We can similarly derive

\[
\frac{1}{n} \sum_{i=1}^{n} I(W_{3,i}, W_{4,i}; Y_i) \leq \frac{1}{2} \log_2 \left( 1 + \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_{3,i}, W_{4,i}} \left\{ (W_{3,i} + W_{4,i})^2 \right\} \right). \quad (A.33)
\]

We now use the correlation between \( W_3 \) and \( W_4 \) to bound (A.32) and (A.33). Denote the average powers of \( W_3 \) and \( W_4 \) by \( P_1 \) and \( P_2 \), respectively. That is,

\[
P_1 = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_{3,i}} \left\{ W_{3,i}^2 \right\}, \quad (A.34)
\]

\[
P_2 = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_{4,i}} \left\{ W_{4,i}^2 \right\}. \quad (A.35)
\]

From the relay power constraints, we must have \( P_1 \leq P_{W_3} \) and \( P_2 \leq P_{W_4} \). Next define the (normalized) correlation coefficient, \( \rho \), by

\[
\rho \sqrt{P_1 \cdot P_2} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_{3,i}, W_{4,i}} \left\{ W_{3,i} \cdot W_{4,i} \right\}. \quad (A.36)
\]

Note that a good communication scheme would have \( \rho \geq 0 \), but in general, \( -1 \leq \rho \leq 1 \). Then from (A.36), (A.33), and (A.27) we have

\[
\frac{1}{n} I(W_3^n, W_4^n; Y^n) \leq \frac{1}{2} \log_2 \left( 1 + \frac{P_1 + P_2 + 2\rho \sqrt{P_1 \cdot P_2}}{N_z} \right) \quad (A.37)
\]

We proceed with bounding (A.32). For any pair of random variables \( W_{3,i} \) and \( W_{4,i} \),

\[
\mathbb{E}_{W_{3,i}} \left\{ \text{var} \left( W_{3,i} \mid W_{4,i} \right) \right\} = \mathbb{E}_{W_{4,i}} \left\{ \mathbb{E}_{W_{3,i}} \left\{ W_{3,i}^2 \mid W_{4,i} \right\} - \left( \mathbb{E}_{W_{3,i}} \left\{ W_{3,i} \mid W_{4,i} \right\} \right)^2 \right\} \quad (A.38)
\]

\[
= \mathbb{E}_{W_{3,i}} \left\{ W_{3,i}^2 \right\} - \mathbb{E}_{W_{4,i}} \left\{ \left( \mathbb{E}_{W_{3,i}} \left\{ W_{3,i} \mid W_{4,i} \right\} \right)^2 \right\}. \quad (A.39)
\]
Also,
\[
\mathbb{E}_{W_3,i,W_4,i} \{ W_{3,i} \cdot W_{4,i} \} = \mathbb{E}_{W_4,i} \{ W_{4,i} \cdot \mathbb{E}_{W_3,i} \{ W_{3,i} \mid W_{4,i} \} \} \quad (A.40)
\]
\[
\leq \sqrt{\mathbb{E}_{W_4,i} \{ W_{4,i}^2 \}} \cdot \sqrt{\mathbb{E}_{W_4,i} \left\{ \left( \mathbb{E}_{W_3,i} \{ W_{3,i} \mid W_{4,i} \} \right)^2 \right\}}. \quad (A.41)
\]

Inequality (A.41) follows from the Cauchy-Schwarz inequality applied in the Hilbert space of scalar random variables. Returning to the problem of interest, from (A.39) and (A.34),
\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_4,i} \{ \text{var} (W_{3,i} \mid W_{4,i}) \} = P_1 - \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_4,i} \left\{ \left( \mathbb{E}_{W_3,i} \{ W_{3,i} \mid W_{4,i} \} \right)^2 \right\}. \quad (A.42)
\]

Also, from (A.41)
\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_3,i,W_4,i} \{ W_{3,i} \cdot W_{4,i} \} \leq \frac{1}{n} \sum_{i=1}^{n} \sqrt{\mathbb{E}_{W_4,i} \{ W_{4,i}^2 \}}
\]
\[
\cdot \sqrt{\mathbb{E}_{W_4,i} \left\{ \left( \mathbb{E}_{W_3,i} \{ W_{3,i} \mid W_{4,i} \} \right)^2 \right\}} \quad (A.43)
\]
\[
\leq \sqrt{\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_4,i} \{ W_{4,i}^2 \}}
\]
\[
\cdot \sqrt{\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_4,i} \left\{ \left( \mathbb{E}_{W_3,i} \{ W_{3,i} \mid W_{4,i} \} \right)^2 \right\}} \quad (A.44)
\]
\[
= \sqrt{P_2} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_4,i} \left\{ \left( \mathbb{E}_{W_3,i} \{ W_{3,i} \mid W_{4,i} \} \right)^2 \right\}}. \quad (A.45)
\]

Inequality (A.44) follows from the Cauchy-Schwarz inequality applied in \( n \)-dimensional
Euclidean space, and (A.45) follows from (A.35). Using (A.36) and (A.45)

$$\rho \cdot \sqrt{P_1} \leq \sqrt{\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_{4,i}} \left\{ \left( \mathbb{E}_{W_{3,i}} \{ W_{3,i} \mid W_{4,i} \} \right)^2 \right\}}. \quad (A.46)$$

Substituting (A.46) into (A.42), we get

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{W_{4,i}} \{ \text{var} (W_{3,i} \mid W_{4,i}) \} \leq P_1 - \rho^2 P_1 \quad (A.47)$$

$$= (1 - \rho^2) P_1. \quad (A.48)$$

Now we put these results together to arrive at the conclusion. Repeating (A.37),

$$\frac{1}{n} I(W_3^n, W_4^n; Y^n) \leq \frac{1}{2} \log_2 \left( 1 + \frac{P_1 + P_2 + 2\rho\sqrt{P_1 \cdot P_2}}{N_z} \right). \quad (A.49)$$

From (A.48), (A.32), and (A.22),

$$\frac{1}{n} I(W_3^n; Y^n \mid W_4^n) \leq \frac{1}{2} \log_2 \left( 1 + \frac{(1 - \rho^2) P_1}{N_z} \right). \quad (A.50)$$

Define $\beta$ to satisfy

$$(1 - \beta) \cdot P_{W_3} = (1 - \rho^2) \cdot P_1. \quad (A.51)$$

The power constraint at Relay 1 implies $P_1 \leq P_{W_3}$, and thus $0 \leq \beta \leq 1$. Substituting (A.51) into (A.49) and (A.50), we get

$$\frac{1}{n} I(W_3^n; Y^n \mid W_4^n) \leq \frac{1}{2} \log_2 (1 + (1 - \beta)S_3). \quad (A.52)$$

$$\frac{1}{n} I(W_3^n, W_4^n; Y^n) \leq \frac{1}{2} \log_2 \left( 1 + \frac{P_1 + P_2 + 2\sqrt{P_2 \cdot (P_1 - (1 - \beta)P_{W_3})}}{N_z} \right) \quad (A.53)$$

$$= \frac{1}{2} \log_2 \left( 1 + \frac{P_1 + P_2 + 2\sqrt{P_2 \cdot (\beta P_{W_3} + (P_1 - P_{W_3}))}}{N_z} \right) \quad (A.54)$$
\[
\leq \frac{1}{2} \log_2 \left( 1 + \frac{P_1 + P_2 + 2\sqrt{P_2 \cdot \beta P_{W_3}}}{N_Z} \right) \tag{A.55}
\]
\[
\leq \frac{1}{2} \log_2 \left( 1 + \frac{P_{W_3} + P_{W_4} + 2\sqrt{\beta P_{W_3} P_{W_4}}}{N_Z} \right) \tag{A.56}
\]
\[
= \frac{1}{2} \log_2 \left( 1 + S_3 + S_4 + 2\sqrt{\beta S_3 S_4} \right). \tag{A.57}
\]

The conclusion follows immediately from (A.52), (A.57), (A.10), and (A.17). \qed
Appendix B

Coding to Relays for an Asymmetric Network

Consider the highly asymmetric PBSC network, where $Y_1 \equiv X$. Also assume $R_3 = R_4 \in [0, 1]$. We want to maximize $(R_{\text{top}} + R_{\text{bot}})$ such that the rate pair is in the intersection of the broadcast capacity region and the extended multiaccess capacity region. We will show that there is always a maximizing rate pair with $R_{\text{top}} = R_3$. As mentioned in Section 3.5.1, the maximizing rate pair is not necessarily unique. If $\alpha_2 = 0.5$, the bottom link of the broadcast channel is completely noisy, and therefore $R_{\text{bot}} = 0$. Since $Y_1 \equiv X$, we can immediately set $R_{\text{top}} = R_3$. Without loss of generality, then, assume $\alpha_2 \in [0, 0.5]$.

Choose $R_{\text{top}} \leq R_3$ arbitrarily. Then we have $(R_3 - R_{\text{top}})$ bits left at Relay 1 to help send $R_{\text{bot}}$. Therefore, $R_{\text{bot}} \leq (2R_3 - R_{\text{top}})$ since it must lie in the extended multiaccess capacity region. Also, from the broadcast capacity region (3.284) and (3.285), the constraint on $R_{\text{bot}}$ is tightened with increasing $\beta \in [0, 0.5]$, and $H_{\text{bin}}^{-1}(x) \in [0, 0.5]$ is a monotonically increasing function of $x$. Therefore, we must have $R_{\text{bot}} \leq 1 - H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_{\text{top}}))$. For an arbitrary $R_{\text{top}} \leq R_3$, our net rate is thus limited to

$$R_{\text{ach}} = (R_{\text{top}} + R_{\text{bot}}) \leq R_{\text{top}} + \min [2R_3 - R_{\text{top}}; 1 - H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_{\text{top}}))]$$

(B.1)

$$= \min [2R_3; R_{\text{top}} + 1 - H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_{\text{top}}))] .$$

(B.2)
The conclusion follows once we establish that

$$f(R_{\text{top}}) = R_{\text{top}} + 1 - H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_{\text{top}}))$$

(B.4)

is an increasing (and, in fact, concave) function of $R_{\text{top}}$.

First, it was shown in Wyner and Ziv [50] that $H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_{\text{top}}))$ is convex in $R_{\text{top}}$. Incidentally, this immediately implies $f(R_{\text{top}})$ is concave. Second, $H_{\text{bin}}^{-1}(R_{\text{top}})$ is increasing in $R_{\text{top}} \in [0, 1]$, and therefore $H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_{\text{top}}))$ is increasing in $R_{\text{top}}$. Together, these two facts imply that

$$\left| \frac{\partial}{\partial R_{\text{top}}} \left( H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_{\text{top}})) \right) \right| \leq \lim_{R_{\text{top}} \to 1} \left| \frac{\partial}{\partial R_{\text{top}}} \left( H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_{\text{top}})) \right) \right|. \quad (B.5)$$

This limit, and those that follow, are left-hand limits. We will evaluate this limit explicitly. First note that $\frac{\partial (H_{\text{bin}}(x))}{\partial x} = \log_2 \left( \frac{1 - x}{x} \right)$. Now let $\lambda = H_{\text{bin}}^{-1}(R_{\text{top}})$. Then by the chain rule for derivatives and since $\left( \frac{\partial (\lambda)}{\partial R_{\text{top}}} \right) = \left( \frac{\partial (\lambda)}{\partial \lambda} \right)^{-1}$ for $R_{\text{top}} < 1$,

$$\frac{\partial}{\partial R_{\text{top}}} \left( H_{\text{bin}}(\alpha_2 \otimes H_{\text{bin}}^{-1}(R_{\text{top}})) \right) = \frac{\partial (H_{\text{bin}}(\alpha_2 \otimes \lambda))}{\partial \lambda} \left( \frac{\partial (\lambda)}{\partial R_{\text{top}}} \right)$$

(B.6)

$$= \frac{(1 - 2\alpha_2) \cdot \log_2 \left( \frac{1 - \alpha_2}{\alpha_2} \right)}{\log_2 \left( \frac{1}{\lambda} \right)}$$

(B.7)

$$= \frac{(1 - 2\alpha_2) \cdot \ln \left( \frac{1 - \alpha_2}{\alpha_2} \right)}{\ln (1 - \lambda) - \ln (\lambda)}.$$

(B.8)
As $R_{\text{top}} \to 1$, we have $\lambda \to 0.5$. We use L'Hôpital’s rule to evaluate the limit.

\[
\lim_{{R_{\text{top}} \to 1}} \left| \frac{\partial \left( H_{\text{bin}} \left( \alpha_2 \otimes H_{\text{bin}}^{-1} \left( R_{\text{top}} \right) \right) \right)}{\partial R_{\text{top}}} \right|
\]

\[
= \lim_{{\lambda \to 0.5}} \left| (1 - 2\alpha_2) \cdot \frac{\ln (1 - \alpha_2 - (1 - 2\alpha_2)\lambda) - \ln (\alpha_2 + (1 - 2\alpha_2)\lambda)}{\ln (1 - \lambda) - \ln (\lambda)} \right|
\]

(B.9)

\[
= \lim_{{\lambda \to 0.5}} \left| (1 - 2\alpha_2) \cdot \frac{-\frac{(1 - 2\alpha_2)}{1 - \alpha_2 - (1 - 2\alpha_2)\lambda} - \frac{1 - 2\alpha_2}{\alpha_2 + (1 - 2\alpha_2)\lambda}}{\frac{-1}{1 - \lambda} - \frac{1}{\lambda}} \right|
\]

(B.10)

\[
= (1 - 2\alpha_2) \cdot \frac{2(1 - 2\alpha_2) + 2(1 - 2\alpha_2)}{2 + 2}
\]

(B.11)

\[
= (1 - 2\alpha_2)^2,
\]

(B.12)

and therefore,

\[
\frac{\partial \left( f(R_{\text{top}}) \right)}{\partial R_{\text{top}}} \geq 1 - (1 - 2\alpha_2)^2
\]

(B.13)

\[
\geq 0 \quad \forall \ R_{\text{top}},
\]

(B.14)

where (B.10) follows from L'Hôpital’s rule, and (B.14) follows since $\alpha_2 \leq 0.5$ by assumption.
Appendix C

A Lemma About Randomly Tossed Balls

**Lemma 6.** Fix any $0 \leq f, k \leq 1$ and choose any $\epsilon > 0$. We have $S$ slots. Mark any $F = \lfloor f \cdot S \rfloor$ of these slots as “special” slots. Throw $K = \lfloor k \cdot S \rfloor$ balls randomly and uniformly into the $S$ slots, allowing at most one ball per slot. Let $V = v \cdot S$ be the number of balls that end up in “special” slots. Then for all $S$, $\max[0; K + F - S] \leq V \leq \min[K; F]$ and

$$
\Pr (|v - f \cdot k| > \epsilon) \leq (S^2 + S) \cdot (2^{-S \cdot g_{f,k}(f \cdot k - \epsilon)} + 2^{-S \cdot g_{f,k}(f \cdot k + \epsilon)}),
$$

where

$$
g_{f,k}(v) = H_{\text{bin}}(k) - f H_{\text{bin}} \left( \frac{v}{f} \right) - (1 - f) H_{\text{bin}} \left( \frac{k - v}{1 - f} \right).
$$

In particular, the function $g_{f,k}(v)$ is strictly convex on $\max[0; k + f - 1] \leq v \leq \min[k; f]$ and $g_{f,k}(f \cdot k) = 0$. This implies that for any $\epsilon > 0$,

$$
\lim_{S \to \infty} \Pr (|v - f \cdot k| > \epsilon) = 0.
$$

**Proof.** All boundary cases, where $f$ or $k$ equal 0 or 1, follow immediately. In what follows, then, assume that $0 < f, k < 1$. We write the probability in (C.3) exactly
and then use exponential bounds to bound each term in the exact expression. We use the conventional notation
\[
\binom{n}{r} = \frac{n!}{r!(n-r)!}.
\] (C.4)

We also use the following well-known bounds [15, p. 284],
\[
\frac{1}{n+1} 2^{\text{H_bin}(\hat{\pi})} \leq \binom{n}{r} \leq 2^{\text{H_bin}(\hat{\pi})}.
\] (C.5)

We ignore the integrality constraints on \(F\) and \(K\) by assuming \(F = f \cdot S\) and \(K = k \cdot S\) rather than \(F = \lfloor f \cdot S \rfloor\) and \(K = \lfloor k \cdot S \rfloor\). Enforcing the integrality constraints can be handled trivially at the expense of notation.

Proceeding, first note that \(\mathbb{E}\{v\} = f \cdot k\). This follows because
\[
\mathbb{E}\{\text{\# balls in slot 1}\} = \Pr(\text{a ball ends up in slot 1})
= \binom{S-1}{K-1} \binom{S}{K}
= k.
\] (C.6)

Therefore, by linearity of the expectation,
\[
\mathbb{E}\{v\} = f \cdot k.
\] (C.9)

Unfortunately we cannot easily use Chebyshev’s inequality to bound the probability in (C.3) because we cannot easily get a bound on the variance of the random variable \(V = v \cdot S\). After all, \(V\) is not the sum of independent, identically distributed random variables since each slot can contain at most one ball. We take a different approach.
We begin by writing the probability in (C.3) exactly.

$$\Pr \left( |v - f \cdot k| > \epsilon \right) = \sum_{V=\max[0;K+F-S]}^{F \cdot k - \epsilon S} \frac{\binom{F}{V} \cdot \binom{S - F}{K - V}}{\binom{S}{K}} + \sum_{V=\max\left[\frac{F \cdot k}{S} + \epsilon S;K+F-S\right]}^{\min[K;F]} \frac{\binom{F}{V} \cdot \binom{S - F}{K - V}}{\binom{S}{K}}. \quad (C.10)$$

The first summation in (C.10) corresponds to $v < f \cdot k - \epsilon$, while the second corresponds to $v > f \cdot k + \epsilon$. Each term in both summations equals the probability that exactly $V = v \cdot S$ balls end up in "special" slots. This follows from the uniform distribution of where the balls land, making every possible combination of balls in slots equally likely. The total number of combinations of $K$ balls in $S$ slots is $\binom{S}{K}$. The number of combinations with $V$ balls in the $F$ "special" slots, and thus with the remaining $(K - V)$ balls in the remaining $(S - F)$ slots, is $\binom{F}{V} \cdot \binom{S - F}{K - V}$.

Some explanation of the limits of the summations in (C.10) is in order. Observe that $V$ must always satisfy $V \geq K + F - S$. Otherwise, we would need $(K - V)$ balls to fit in $(S - F)$ slots. However, when $V < K + F - S$,

$$(S - F) - (K - V) = V - (K + F - S) < 0. \quad (C.11)$$

This must be taken into account in the lower limits of the summations. The upper limit of the second summation, $V = \min[K;F]$, follows because there are only $K$ balls and only $F$ "special" slots.

Finally, we have assumed that $\epsilon$ is relatively small in order to make (C.10) strictly
correct. Specifically, we have assumed that
\[
\max[0; K + F - S] \leq \frac{F \cdot K}{S} - \epsilon \cdot S, \quad (C.12)
\]
\[
\max\left[\frac{F \cdot K}{S} + \epsilon \cdot S; K + F - S\right] \leq \min[K; F]. \quad (C.13)
\]
However, the necessary adjustment for larger \(\epsilon\) is simple and the result follows easily using the same argument. In particular, if inequality (C.12) does not hold, then the first summation in (C.10) does not appear. Similarly, if inequality (C.13) does not hold, then the second summation in (C.10) does not appear. If neither inequality holds, then the probability of the event in question is exactly 0.

Proceeding, consider the exponential behavior (in \(S\)) of the individual terms in (C.10). Using (C.5) for the individual terms in (C.10),
\[
\left( \frac{F}{V} \right) \cdot \left( \frac{S - F}{K - V} \right) \cdot \left( \frac{S}{K} \right) \leq (S + 1) \cdot 2^{\left[FH_{bin}\left(\frac{V}{f}\right) + (S - F)H_{bin}\left(\frac{k - V}{1 - f}\right) - SH_{bin}\left(\frac{k}{1 - f}\right)\right]} \quad (C.14)
\]
\[
= (S + 1) \cdot 2^{-S} \left[H_{bin}(k) - fH_{bin}\left(\frac{V}{f}\right) - (1 - f)H_{bin}\left(\frac{k - V}{1 - f}\right)\right] \quad (C.15)
\]
\[
= (S + 1) \cdot 2^{-S} g_{f,k}(v), \quad (C.16)
\]
where we have defined the function
\[
g_{f,k}(v) = H_{bin}(k) - fH_{bin}\left(\frac{v}{f}\right) - (1 - f)H_{bin}\left(\frac{k - v}{1 - f}\right); \quad (C.17)
\]
We will define the domain of the function \(g_{f,k}(v)\) shortly.

We will show that the function \(g_{f,k}(v)\) looks like that sketched in Figure C-1. Specifically, we will show that the function \(g_{f,k}(v)\) is strictly convex in \(v\) over its domain, and that \(g_{f,k}(f \cdot k) = 0\). As a result, we will have shown that
\[
g_{f,k}(v) \geq g_{f,k}(f \cdot k - \epsilon) \quad (C.18)
\]
\[
> 0 \quad (C.19)
\]
whenever $v \leq f \cdot k - \epsilon$. Similarly,

$$g_{f,k}(v) \geq g_{f,k}(f \cdot k + \epsilon)$$  \hspace{1cm} (C.20)

$$> 0$$  \hspace{1cm} (C.21)

whenever $v \geq f \cdot k + \epsilon$. The conclusion of the lemma follows from this, in turn, since, from (C.10) and (C.16),

$$\Pr(|v - f \cdot k| > \epsilon) = \sum_{V=\max[0;K+F-S]}^{\min[K;F]} \frac{\binom{F}{V} \cdot \binom{S-F}{K-V}}{\binom{S}{K}}$$

$$+ \sum_{V=\max[\frac{F-K}{S}+\epsilon S;K+F-S]}^{\min[K;F]} \frac{\binom{F}{V} \cdot \binom{S-F}{K-V}}{\binom{S}{K}}$$  \hspace{1cm} (C.22)

$$\leq (S + 1) \cdot \sum_{V=\max[0;K+F-S]}^{\min[K;F]} 2^{-S g_{f,k}(v)} + (S + 1) \cdot \sum_{V=\max[\frac{F-K}{S}+\epsilon S;K+F-S]}^{\min[K;F]} 2^{-S g_{f,k}(v)}$$  \hspace{1cm} (C.23)

$$\leq (S^2 + S) \cdot (2^{-S g_{f,k}(f \cdot k - \epsilon)} + 2^{-S g_{f,k}(f \cdot k + \epsilon)})$$  \hspace{1cm} (C.24)

$$\rightarrow 0 \text{ as } S \rightarrow \infty.$$  \hspace{1cm} (C.25)
From (C.23), we define the domain of $g_{f,k}(v)$ as

$$\max[0; k + f - 1] \leq v \leq \min[k; f].$$  \hfill (C.26)

Notice that the binary entropy function, $H_{\text{bin}}(x)$, is twice continuously differentiable everywhere except $x = 0$ and $x = 1$. Furthermore, $0 \leq H_{\text{bin}}(x) \leq 1$ for all $x \in [0,1]$. Additionally, $H_{\text{bin}}(x)$ is right- and left-continuous at $x = 0$ and $x = 1$, respectively. Evaluating the derivative,

$$\frac{d (H_{\text{bin}}(x))}{dx} = \log \left( \frac{1 - x}{x} \right).$$  \hfill (C.27)

From the definition of $g_{f,k}(v)$ in (C.17), then, $g_{f,k}(v)$ is a twice continuously differentiable function of $v$ on the interior of its domain, which was specified in (C.26). Furthermore, $g_{f,k}(v)$ is finite and continuous on its whole domain (specifically, it is right- and left-continuous at the boundaries of its domain). Additionally, direct evaluation of (C.17) shows $g_{f,k}(f \cdot k) = 0$. It remains to show that $g_{f,k}(v)$ is strictly convex with a minimum at $v = f \cdot k$. Using (C.27) and the chain rule for differentiation,

$$\frac{d (g_{f,k}(v))}{dv} = - \log \left( \frac{f - v}{v} \right) + \log \left( \frac{1 - f - k + v}{k - v} \right)$$

$$= \log \left( \frac{(1 - f - k + v) \cdot v}{(f - v) \cdot (k - v)} \right).$$  \hfill (C.28)

Evaluating at $v = f \cdot k$, we find

$$\left. \frac{d (g_{f,k}(v))}{dv} \right|_{v = f \cdot k} = \log(1) = 0.$$  \hfill (C.30)

For completeness, we verify that we have not violated any assumption in evaluating the derivative at $v = f \cdot k$ in (C.30). Notice that $f \cdot k \in \{0, f, k, k + f - 1\}$ is only possible when $f$ or $k$ equals 0 or 1. To check whether it would be possible to satisfy
\[ f \cdot k = k + f - 1, \text{ notice that} \]

\[
f \cdot k - (k + f - 1) = 1 - k - f + f \cdot k \]
\[= (1 - k) \cdot (1 - f) \quad (C.32)\]
\[> 0 \quad \text{whenever } 0 < f, k < 1. \quad (C.33)\]

We next show that the second derivative, \( \frac{d^2(g_{f,k}(v))}{dv^2} \), is strictly positive in its domain. From (C.29),

\[
\frac{d}{dv} \left( g_{f,k}(v) \right) = \log((1 - f - k + v) \cdot v) - \log(k - v) - \log(f - v). \quad (C.34)
\]

Taking the second derivative,

\[
\frac{1}{\ln 2} \frac{d^2 \left( g_{f,k}(v) \right)}{dv^2} = \frac{(1 - f - k) + 2v}{(1 - f - k) \cdot v + v^2} + \frac{1}{k - v} + \frac{1}{f - v} \quad (C.35)
\]
\[= \frac{(1 - f - k) + v + v}{v((1 - f - k) + v)} + \frac{1}{k - v} + \frac{1}{f - v} \quad (C.36)
\]
\[= \frac{1}{v} \left( 1 + \frac{v}{(1 - f - k) + v} \right) + \frac{1}{k - v} + \frac{1}{f - v}. \quad (C.37)
\]

In the domain specified by (C.26), at all points except the boundaries \( v \in \{0, f, k, k + f - 1\} \), all three terms in (C.37) are strictly positive. Therefore, \( g_{f,k}(v) \) is strictly convex on the interior of the region of interest specified by (C.26) [10, Prop. B.4, p. 562]. Since, as pointed out earlier, \( g_{f,k}(v) \) is right- and left-continuous at the boundaries of its domain, the lemma follows. \( \square \)
APPENDIX C. A LEMMA ABOUT RANDOMLY TOSSED BALLS
Appendix D

Several As Yet Fruitless Ideas for Converges

D.1 Conjecture for Symmetric Discrete Networks

For a discrete network with a symmetric broadcast channel and with noiseless relay links, the multiaccess cut-set bound, (3.19), states

\[ C_{\text{net}} \leq R_3 + R_4. \]  \hspace{1cm} (D.1)

We conjecture that this upper bound is loose whenever the relays cannot decode the input message. Specifically,

Conjecture 2.

\[ C_{\text{net}} < R_3 + R_4 \]  \hspace{1cm} (D.2)

whenever

\[ R_3 + R_4 > C_{\text{link}}. \]  \hspace{1cm} (D.3)

In the following discussion, assume for simplicity that we have \( R_3 = R_4 \) as well as a symmetric broadcast channel. We can make an intuitive argument by contradic-
tion based on an extreme interpretation of some mutual information “approximate equalities”. In particular, if we assume that 2R_3 is achievable, then we must have

\[
I(X^n; W_3^{nR_3}) \approx nR_3 \Rightarrow H(W_3^{nR_3}|X^n) \approx 0, \quad H(W_3^{nR_3}) \approx nR_3.
\]

\[
I(X^n; W_4^{nR_4}) \approx nR_4 \Rightarrow H(W_4^{nR_4}|X^n) \approx 0, \quad H(W_4^{nR_4}) \approx nR_4. \quad \dagger
\]

\[
I(X^n; W_3^{nR_3}, W_4^{nR_4}) \approx n(R_3 + R_4) \Rightarrow H(W_4^{nR_4}|W_3^{nR_3}) \approx H(W_4^{nR_4}).
\]

We have labeled these approximate equalities as (†) for future reference. Thus, from (†), W_3 and W_4 are approximately independent, W_3 and W_4 are approximately uniformly distributed over their entire range, and the conditional distribution of W_3 and W_4 given the transmitted input codeword has extremely small normalized entropy. Intuitively, we believe this implies that the relays relatively coarsely quantize the \(Y_{1,2}\) space. Thus assume that 2R_3 > C_{1\text{ link}} and that 2R_3 is reliably achievable. We try to derive a contradiction by showing that if 2R_3 > C_{1\text{ link}} and if the above approximate equalities (†) are satisfied, then, in fact, we must have 2R_3 ≤ C_{1\text{ link}}.

**Point of view # 1 — string counting**

If we interpret these approximate equalities as actual equalities, we can make an intuitive argument and then see if we can generalize rigorously. In particular, assume \(H(W_3^{nR_3}|X^n) = H(W_4^{nR_4}|X^n) = 0\). Then every input codeword \(X^n\) must be mapped by the relays to a unique index \(W_{\{3,4\}}(X^n)\). There are approximately \(2^{nH(Y_1|X)} Y_1^n\) strings associated with a given codeword, and all of these must get mapped en masse by Relay 1 to some \(W_3(X^n)\). Similarly, by symmetry of the broadcast channel, these same \(2^{nH(Y_1|X)} Y_2^n\) strings must get mapped en masse by Relay 2 to some \(W_4(X^n)\). Finally, the decoder must decode the pair \((W_3(X^n), W_4(X^n))\) to \(X^n\).

To prevent confusion between input codewords in the network, every set of \(2^{nH(Y_1|X)} Y^n\)-strings associated with a codeword must have no significant overlap with the union of strings associated with the other codewords (assuming we have expurgated a good code to yield a maximum probability of error close to zero). However, there are only \(2^n Y^n\)-strings in total, and therefore \(2^{nR_{ach}} \cdot 2^{nH(Y_1|X)} \leq 2^n\). Therefore \(R_{ach} \leq C_{1\text{ link}}\), and we have our contradiction.

There are several equivalent ways of making this same basic argument under the critical assumption \(H(W_3^{nR_3}|X^n) = H(W_4^{nR_4}|X^n) = 0\). Unfortunately, we have not
been able to generalize this to a valid argument under the proper assumptions (†). In particular, \( H(W_3^{nR_3} | X^n) \approx 0 \) really means \( H(W_3^{nR_3} | X^n) \leq n \cdot \delta \) for some small \( \delta > 0 \) — therefore the unnormalized conditional entropy can grow linearly with the blocklength \( n \). This does not then imply that the set of \( 2^{nH(Y_1|X)} \) strings associated with a particular input codeword get mapped to only one or two bins. Rather, they could be mapped to an almost exponential number of bins. This makes the argument difficult to generalize.

**Point of view # 2 — broadcast channel implication**

We would like to verify the same converse by contradiction for the discrete, symmetric network (again, the intuition follows with equal validity when \( R_3 \neq R_4 \)). Assuming \( 2R_3 > C_{1 \text{ link}} \) is achievable, we would like to argue that this implies \( 2R_3 \leq C_{1 \text{ link}} \). Proceeding from the same approximate equalities (†), this looks very much like we are sending \( R_3 \) bits to a user connected downstream from Relay 1, and we are sending \( R_4 \) independent bits to a user connected downstream from Relay 2. Why *doesn’t* this imply that we can approximately achieve the independent rate pair \( (R_3, R_4) \) across the underlying broadcast channel \( X \rightarrow (Y_1, Y_2) \)? If it did, then since \( Y_1 \) and \( Y_2 \) are stochastically identical, they can decode each other’s message streams. This would imply that the rate \( (R_3 + R_4) \) is achievable to a user connected to \( Y_1 \) only, i.e., that \( (R_3 + R_4) \leq C_{1 \text{ link}} \), and we would have our contradiction.

We have assumed there is a block code of length \( n \) input symbols per codeword \( X^n \) (and \( nR_3 \) binary \( W_3 \) symbols, \( nR_4 \) binary \( W_4 \) symbols). There are \( 2^{n2R_3} \) codewords, and if we place a uniform probability distribution on them, then the approximate equalities (†) are satisfied. Though we have assumed a uniform probability distribution on the input codewords, which is not unusual in the literature when proving converses, this simplifying assumption is not fundamentally important.

We can view the assumed block code as a set of superletters, \( X = X^n \). We use these superletters as input symbols for a new broadcast channel. We include the relay mappings, \( Y_1^n \rightarrow W_3^{nR_3} \) and \( Y_2^n \rightarrow W_4^{nR_4} \), in the definition of the new broadcast channel, and thus it has output superletters \( \tilde{W}_3 = W_3^{nR_3} \) and \( \tilde{W}_4 = W_4^{nR_4} \). Since the underlying broadcast channel has conditionally independent observations when conditioned on the input, the new broadcast channel satisfies \( p(\tilde{W}_3, \tilde{W}_4 | \tilde{X}) = p(\tilde{W}_3 | \tilde{X}) \cdot p(\tilde{W}_4 | \tilde{X}) \). If we place a uniform distribution \( p(\tilde{X}) \) on the \( 2^{n2R_3} \) input superletters,
then the approximate equalities (†) hold.

Now we would like to develop a random coding argument by generating a number of length $k$ supercodewords using the uniform input distribution $p(\widetilde{X})$ a large number of times, $k \rightarrow \infty$. Our conclusion and thus the contradiction follows if we can show that some rate pair $(R'_3, R'_4) \approx (R_3, R_4)$ is achievable.

Unfortunately, we now have a general broadcast channel $\widetilde{X} \rightarrow (\widetilde{W}_3, \widetilde{W}_4)$, no longer with stochastically identical outputs (much less degraded). We only have the conditional independence structure, $p(\bar{w}_1, \bar{w}_2 | \bar{x}) = p(\bar{w}_1 | \bar{x}) \cdot p(\bar{w}_2 | \bar{x})$. As far as we have found, almost no progress has been made in the literature in studying the general broadcast channel with this structure (and without additional structure). However, there is one general achievability theorem proved Marton in [32] (see also El Gamal and Van der Meulen [24]). Choose an arbitrary joint ensemble $p(U, V, \widetilde{X})$. Then $(R'_3, R'_4)$ is achievable if

$$
R'_3 \leq I(U; \widetilde{W}_3), \\
R'_4 \leq I(V; \widetilde{W}_4), \\
(R'_3 + R'_4) \leq I(U; \widetilde{W}_3) + I(V; \widetilde{W}_4) - I(U; V).
$$

Note that $\widetilde{X}$ enters the picture via the broadcast channel, $p(\bar{w}_1, \bar{w}_2 | u, v, \bar{x}) = p(\bar{w}_1, \bar{w}_2 | \bar{x})$.

Now the hope would be that we could choose $U$ to be $nR_3$ bits and $V$ to be $nR_4$ independent bits. Then $X$ is the collection of $U$ and $V$, while $I(U; V) = 0$. If we could show that we can partition the original length $n$ codewords in such a way that $H(\widetilde{W}_3 | U) \approx 0$ and $H(\widetilde{W}_4 | V) \approx 0$, then we’d be done — Marton’s theorem says we can achieve $(R'_3, R'_4) = (R_3, R_4)$. However, demonstrating that we can partition the input codewords in this way appears to be difficult. Similarly, if we assume the extreme interpretation of the approximate equalities (†), namely that $H(\widetilde{W}_3 | \bar{X}) = H(\widetilde{W}_4 | \bar{X}) = 0$, then we could let $p(U, V, \bar{X})$ be distributed as $U = W_3(X^n)$ and $V = W_4(X^n)$ and reach exactly the same conclusion. But again this result does not follow when we relax the extreme interpretation. As an aside, note that $U$ and $V$ need not be independent to apply this theorem achieving independent information rates. This yields some intuitive validity to our somewhat
loose interpretation of the approximate equalities (†).

We have thus far not been able to make a valid argument applying Marton’s theorem, but there seems to be something to this point of view. There are two possibilities. One, there is something to this indirect approach, but we have not been able to demonstrate an appropriate $U$ and $V$ (and it may be analytically difficult). Two, our intuition in interpreting the approximate equalities (†) as actual equalities is way off.

Point of view # 3 — what we definitely need to consider simultaneously to establish the conjecture

Consider what we are trying to prove with the approximate equalities (†). We are trying to prove that we cannot simultaneously have

\[
\begin{align*}
I(X; W_3) & \to R_3, \\
I(X; W_4) & \to R_4, \\
I(W_3; W_4) & \to 0
\end{align*}
\]

for any $(R_3 + R_4) \in (C_{\text{link}_1}, C_{\text{link}_2})$. We should check what is happening at the various achievable points we have discovered thus far. In particular, when we code to the relays we have

\[
\begin{align*}
I(X; W_3) &= R_3, \\
I(X; W_4) &= R_4, \\
I(W_3; W_4) &= 0.
\end{align*}
\]

However $(R_3 + R_4) \leq C_{\text{link}_1}$. Thus $(R_3 + R_4) \in (C_{\text{link}_1}, C_{\text{link}_2})$ is crucial in establishing our conjecture.

At the Slepian-Wolf corner point, where we distributively encode the observations,
we have
\[
H(W_3) = R_3, \\
H(W_4) = R_4, \\
I(W_3; W_4) = 0, \\
I(X; W_3, W_4) = C_{\text{links}}.
\]
However \(H(W_3 | X) \approx H(Y_1 | X)\). Thus \(H(W_3 | X) \approx 0\) and \(H(W_4 | X) \approx 0\) are also crucial in establishing our conjecture.

Now let us consider a third situation with a different network. This will yield some additional insight into the difficulty in proving a general converse statement. Instead of a symmetric broadcast channel, consider a deterministic broadcast channel with a 2-bit input \(X\) and a pair of 1-bit outputs \(Y_1\) and \(Y_2\). Let \(Y_1\) be the first bit of \(X\) (noiseless) and let \(Y_2\) be the second bit of \(X\) (noiseless). Then the links from \(X \rightarrow Y_1\) and \(X \rightarrow Y_2\) are conditionally independent given the input \(X\), and they are both essentially the same by many information-theoretic measures — \(H(Y_1 | X) = H(Y_2 | X) = 0\), and both have capacity equal to one bit (we can therefore define \(C_{\text{link}} = 1\)). Further, \(I(X; Y_1, Y_2)\) is maximized by a uniform input distribution on \(X\), in which case \(I(X; Y_1, Y_2) = 2C_{\text{link}} = 2\). Note that with a non-uniform input distribution \(p(X)\), we typically have \(H(Y_1) \neq H(Y_2)\). Thus the broadcast links are not the same from all information-theoretic points of view.

For this network, the capacity equals
\[
C_{\text{net}} = (\min[1; R_3] + \min[1; R_4]). \tag{D.4}
\]
Let us prove this to see if we can gain any insight for our symmetric PBSC network. For the converse, we have
\[
I(X; W_3, W_4) \leq I(X; W_3) + I(X; W_4), \tag{D.5}
\]
which follows since

$$H(W_4 | W_3) \leq H(W_4),$$  \hfill (D.6)

and, from the conditional independence of the broadcast channel outputs,

$$H(W_4 | W_3, X) = H(W_4 | X).$$  \hfill (D.7)

Applying data processing inequalities separately to each of these terms, we find

$$I(X; W_3) \leq \min[1; R_3],$$  \hfill (D.8)

$$I(X; W_4) \leq \min[1; R_4].$$  \hfill (D.9)

On the achievability side, note that the broadcast channel capacity region with independent messages is a box: any \((R'_3, R'_4) \leq (1,1)\) is achievable. Therefore

$$(R'_3, R'_4) = (\min[R_3; 1], \min[R_4; 1])$$  \hfill (D.10)

is in the intersection of the capacity region of the broadcast channel and the capacity region of the multiaccess channel, assuming independent message streams. In this network, then, we can achieve rates \(R_{\text{ach}} = (R_3 + R_4)\) even when \((R_3 + R_4) > C_{\text{link}},\)

e.g., when \(R_3 = R_4 = 1.\) We can thus conclude from this example that the stochastically identical link assumption, \(p(y_1 | x) = p(y_2 | x),\) is also crucial in establishing our conjecture.

Finally, with this last network example in mind, let us turn back to our symmetric network examples. On the converse side, it follows similarly that

$$C_{\text{net}} \leq (\min[R_3; C_{\text{link}}] + \min[R_4; C_{\text{link}}]).$$  \hfill (D.11)

However, the minimum of the broadcast cut-set, (3.10), the multiaccess cut-set, (3.19), and the cross-cut bounds, (3.28) and (3.29), is at least as small as \((\min[R_3; C_{\text{link}}] + \min[R_4; C_{\text{link}}]).\) To show this, we require only some simple algebra and the fact that \(C_{\text{link}} \leq 2C_{\text{link}}.\) On the achievability side, from (3.3) and since the broadcast
channel is symmetric, the broadcast channel capacity region (assuming independent messages) is a triangle: \((R_3' + R_4') \leq C_{1\text{link}}\) is achievable. Therefore, basing our approach on coding to the relays, we can never prove achievability of \((R_3 + R_4)\) when \((R_3 + R_4) > C_{1\text{link}}\), (coincidentally, this would disprove our conjecture). Thus considering this last network example only seems to teach us that stochastically identical relay observations are crucial to our conjecture for symmetric parallel relay networks.

\section*{D.2 Find a network where we can reasonably guess \(C_{\text{net}}\)}

It would be useful to construct a discrete parallel relay network with appropriate structure for which we can reasonably guess the capacity. In such a case, knowing that converses are difficult in general, we would have a somewhat comfortable goal in mind. Then we could work to prove it. However, we can find no such network because invariably we observe the same fundamental behavior we have seen thus far. Namely, when the broadcast side is sufficiently strong, then coding to the relays and cooperating on the multiaccess side is best. On the other hand, when the broadcast side is sufficiently weak, then we should try to take advantage of multiple observations within the distributed system.

Let us take an example to illustrate the point. Consider a new parallel relay network. On the broadcast side, consider a symmetric, parallel erasure broadcast channel (PBEC). This is similar to a PBSC except that the BSC’s are replaced with binary erasure channels. For these erasure channels, let \(p(y_1 = 0|x = 0) = 1 - \epsilon_1, p(y_1 = 2|x = 1) = 1 - \epsilon_1,\) and \(p(y_1 = 1|x = 0) = p(y_1 = 1|x = 1) = \epsilon_1.\) Thus an erasure is received at the relays as a 1, occurs with probability \(\epsilon_1,\) and occurs independently from the choice of input symbol. Next, consider a new multiaccess side. Replace the pair of noiseless binary links with a structured, \(3 \times 3\)-input and 6-output multiaccess channel. We let the relay transmission signals, \(W_3\) and \(W_4,\) have the same alphabet, both equal to \(\{0, 1, 2\}.\) Then define the multiaccess channel as follows. With probability \(\epsilon_2\) the output is erased: \(p(y = 6 | w_3, w_4) \equiv \epsilon_2.\) Otherwise, the output equals the sum of the two inputs with probability one: \(p(y = w_3 + w_4|w_3, w_4) = (1-\epsilon_2).\)

This network is constructed with a sort of sufficient statistic “matching” in mind.
D.2. FIND A NETWORK WHERE WE CAN REASONABLY GUESS $C_{\text{net}}$

In particular, assume that the relays simply feed through what they receive (i.e.,
y they simply transpond their observations: $W_{3,k} = Y_{1,k-1}$ and $W_{4,k} = Y_{2,k-1}$). Then if
$y = 6$, we get no information about the input letter $X$ because the output is erased.
If $y \in \{0,1\}$, we conclude $X = 0$ with probability one. If $y \in \{3,4\}$, we conclude
$X = 1$ with probability one. If $y = 2$ then we conclude that both observations $Y_1$ and
$Y_2$ were erased, and thus we would have no information about the input symbol even
having access to the pair of relay observations (which, in fact, we do since $w_3 + w_4 = 2$
is only possible when $y_1 = y_2 = 1$).

Noting that with probability $\epsilon_2$ we get no information through on a multiaccess
channel use, we may hypothesize (at first) that the most we could hope for is to
achieve rate $\left(1 - \epsilon_2\right) \cdot \max I(X; Y_1, Y_2) = \left(1 - \epsilon_2\right) \cdot (1 - \epsilon_1^2) \leq \left(1 - \epsilon_2\right)$. This is achieved
by the simple transponding scheme. This would be the goal to have in mind in trying
to prove a converse. Unfortunately, the processing at the relays renders this intuition
invalid. Specifically, this is not the best we can do in all parametric regimes. Consider
the multiaccess data processing bound, $\max I(W_3, W_4; Y) = \left(1 - \epsilon_2\right) \cdot \log_2(5) > \left(1 - \epsilon_2\right)$.\nOnce again, if the broadcast side of the channel is good enough (i.e., if $\epsilon_1$ is small
enough relative to $\epsilon_2$: specifically, if $\epsilon_1 \leq 1 - (1 - \epsilon_2) \cdot \log_2(5)$), then we can achieve the
multiaccess cut-set bound by coding to the relays. Thus we do not know the capacity
of this specially constructed network. We have yet to think of a useful network which
does not exhibit similar behavior.
Appendix E

Summary of Main Graphical Results

Figure E-1: Evaluation of results for a symmetric Gaussian network, $S_1 = S_2 = 2$
Figure E-2: Evaluation of results for the symmetric PBSC network, $C_1 = C_2 = 0.5$
Figure E-3: Evaluation of the block quantization and binning scheme for the symmetric PBSC network, $C_1 = C_2 = 0.5$
Figure E-4: Evaluation of results for the highly asymmetric PBSC network, $C_2 = 0.5$
Bibliography


