Coding Schemes for Relay Networks

MAJID NASIRI KHORMUJI

Doctoral Thesis in Telecommunications
Stockholm, Sweden 2011

© Majid Nasiri Khormuji, September 2011

Tryck: Universitetsservice US AB
Cooperative communications by pooling available resources—for example, power and bandwidth—across the network, is a distributed solution for providing robust wireless transmission. Motivated by contemporary applications in multi-hop transmission and ad hoc networks, the classical three-node relay channel (RC) consisting of a source-destination pair and a relay node has received a renewed attention. One of the crucial aspects of the communication over relay networks (RNs) is the design of proper relaying protocols; that is, how the relay should take part in the transmission to meet a certain quality of service.

In this dissertation, we address the design of reliable transmission strategies and quantification of the associated transmission rates over RNs. We consider three canonical examples of RNs: the classical RC, the multiple-access RC (MARC) and the two-way RC. We also investigate the three-node RC and MARC with state. The capacity of the aforementioned RNs is an open problem in general except for some special cases. In the thesis, we derive various capacity bounds, through which we also identify the capacity of some new classes of RNs. In particular, we introduce the class of state-decoupled RNs and prove that the noisy network coding is capacity achieving under certain conditions.

In the thesis, we also study the effect of the memory length on the capacity of RNs. The investigated relaying protocols in the thesis can be categorized into two groups: protocols with a finite relay memory and those with infinite relay memory requirement. In particular, we consider the design of instantaneous relaying (also referred to as memoryless relaying) in which the output of the relay depends solely on the presently received signal at the relay. For optimizing the relay function, we present several algorithms constructed based on grid search and variational methods. Among other things, we surprisingly identify some classes of semi-deterministic RNs for which a properly constructed instantaneous relaying strategy achieves the capacity. We also show that the capacity of RNs can be increased by allowing the output of the relay to depend on the past received signals as well the current received signal at the relay. As an example, we propose a hybrid digital–analog scheme that outperforms the cutset upper bound for strictly causal relaying.

**Keywords:** Relay networks, capacity, decode-and-forward, amplify-and-forward, compress-and-forward, nonlinear relaying, noisy network coding, hybrid relaying.
Acknowledgments

It has been a long but immensely rewarding journey to reach the point to write my doctoral thesis. This would have been impossible without the support and inspiration of many individuals that I met during my life. In particular, I take the opportunity to acknowledge all of my teachers who inspired me and gave me more strength to pursue my studies further. In addition to the inspiring teachers, I also had the privilege to do my higher education in best schools with creative environments; Sharif University of Technology, Stanford University and KTH (Royal Institute of Technology).

First and foremost, I would like to sincerely thank my advisor Prof. Mikael Skoglund. I am grateful to his generous support to overcome difficulties and willingness for sharing experiences. I highly value his humble attitude. He gave me the freedom to explore different ideas and always supported my decisions.

I would like to express my gratitude to my former advisor Prof. Erik G. Larsson who gave me the opportunity to begin my doctoral studies and exposed me to the topic of communications over relay networks. His close supervision and systematic approach to the problems always inspired me.

I would like to thank Prof. Abbas El Gamal for giving me the opportunity to visit his group at Stanford University. I certainly benefitted from his outstanding knowledge of the field, clarity of thoughts and constructive criticism. I would like to thank John och Karin Engbloms Stipendiefond for the financial support of my visit. In Stanford, I met several individuals who made my stay more enjoyable. In particular, I would like to thank Samer Medawar, Davood Shamsi, Yeow-Khiang Chia, and Han-I Su.

I would like to thank Dr. Sha Yao and Ali Zaidi for joint work which resulted to some publications, and for proofreading some parts of the thesis. I would also like to thank Assistant Prof. Ming Xiao, Dr. Johannes Kron, and Hieu Do for proofreading some other parts of the thesis.

I would like to thank Prof. Gerhard Kramer for taking the time to act as faculty opponent, and Prof. Rolf Johannesson, Associate Prof. Petar Popovski and Associate Prof. Viktoria Fodor for acting as grading committee. I would also like
to thank Prof. Lars K. Rasmussen for the quality review of the thesis.

I would also like to thank Associate Prof. Mats Bengtsson for his help with WINNER project, Prof. Eduard A. Jorswieck for several discussions that we had when he was doing his postdoctoral studies at KTH, Dr. Per Zetterberg for discussions on practical implementation of relaying algorithms and Associate Prof. Magnus Jansson for considering my situations when assigning the teaching responsibilities.

I would like to express my appreciations to all my current and former colleagues at communication theory and signal processing groups—it has been an enjoyable experience to work on “Plan 4.” I would also like to thank Cecilia Forssman, Marie Androv Broms, Karin Demin, Annika Augustsson, Tove Schwartz, and Irene Kindblom for helping me with various administrative issues.

I would like to express my sincere gratitude to my parents and parents-in-law for their tremendous support and love. Finally I would like to deeply thank Farzaneh for the happiness she brings to my life.

Majid Nasiri Khormuji
Stockholm, September 2011
Abstract

Acknowledgments

Contents

1 Introduction

1.1 Background

1.2 Potential Applications of Relaying

1.2.1 Relay-Based Cellular Network

1.2.2 Wireless Sensor Networks

1.2.3 Relay-Assisted Cognitive Radios

1.3 Outline and Contributions

1.4 Contributions Outside the Thesis

1.5 Notation

1.6 Acronyms

2 The Classical Relay Channel

2.1 Definitions

2.2 Infinite-Letter Characterization of the Capacity

2.3 Single-Letter Bounds on the Capacity

2.3.1 Cutset Upper Bound

2.3.2 Direct-Transmission Lower Bound

2.3.3 Two-Hop Lower Bound

2.3.4 Decode-and-Forward Lower Bound

2.3.5 Partial Decode-and-Forward Lower Bound

2.3.6 Compress-and-Forward Lower Bound

2.4 Instantaneous Relaying

2.4.1 A Binary Symmetric Relay Channel

2.4.2 On the Optimality of Instantaneous Relaying

2.4.3 Innovative Relaying

2.A Typicality and Joint Typicality

2.B Mrs. Gerber’s Lemma
### 3 Binary Symmetric Relay Channels

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>37</td>
</tr>
<tr>
<td>3.2 Channel Models</td>
<td>37</td>
</tr>
<tr>
<td>3.3 Capacity Bounds: Infinite Memory Relay Case</td>
<td>39</td>
</tr>
<tr>
<td>3.4 Capacity Bounds: Finite Memory Relay Case</td>
<td>43</td>
</tr>
<tr>
<td>3.4.1 Achievable Rate</td>
<td>43</td>
</tr>
<tr>
<td>3.4.2 Mapping Optimization for an Arbitrary $k$</td>
<td>44</td>
</tr>
<tr>
<td>3.4.3 Fourier Spectrum of the Optimized Mappings</td>
<td>45</td>
</tr>
<tr>
<td>3.4.4 Effect of Channel Parameters on the Structure of the Optimized Mappings</td>
<td>47</td>
</tr>
<tr>
<td>3.5 Numerical Examples</td>
<td>51</td>
</tr>
<tr>
<td>3.6 Summary and Concluding Remarks</td>
<td>51</td>
</tr>
<tr>
<td>3.A Proof of Proposition 3.3.1</td>
<td>53</td>
</tr>
<tr>
<td>3.B Proof of Proposition 3.3.2</td>
<td>54</td>
</tr>
<tr>
<td>3.C Proof of Proposition 3.3.3</td>
<td>54</td>
</tr>
<tr>
<td>3.D Proof of Proposition 3.3.5</td>
<td>56</td>
</tr>
<tr>
<td>3.E Proof of Proposition 3.4.1</td>
<td>58</td>
</tr>
<tr>
<td>3.F Proof of Lemma 3.4.2</td>
<td>58</td>
</tr>
</tbody>
</table>

### 4 Orthogonal Gaussian Relay Channels

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>61</td>
</tr>
<tr>
<td>4.1.1 Scope and Motivation</td>
<td>61</td>
</tr>
<tr>
<td>4.1.2 Related Work</td>
<td>61</td>
</tr>
<tr>
<td>4.1.3 Structure of the Chapter and Contributions</td>
<td>62</td>
</tr>
<tr>
<td>4.2 Orthogonal Instantaneous Relay Channel</td>
<td>63</td>
</tr>
<tr>
<td>4.2.1 Upper Bound on the Capacity</td>
<td>64</td>
</tr>
<tr>
<td>4.3 A Semi-Deterministic Gaussian Relay Channel</td>
<td>66</td>
</tr>
<tr>
<td>4.3.1 Constructing the Optimal Instantaneous Mapping</td>
<td>66</td>
</tr>
<tr>
<td>4.3.2 Connection to High-Rate Uniform Quantization</td>
<td>71</td>
</tr>
<tr>
<td>4.3.3 Suboptimal Approaches</td>
<td>71</td>
</tr>
<tr>
<td>4.3.4 Numerical Examples</td>
<td>73</td>
</tr>
<tr>
<td>4.4 A Gaussian Relay Channel with Side Information</td>
<td>78</td>
</tr>
<tr>
<td>4.4.1 Instantaneous Relaying</td>
<td>79</td>
</tr>
<tr>
<td>4.4.2 CF Relaying</td>
<td>83</td>
</tr>
<tr>
<td>4.4.3 DF Relaying</td>
<td>84</td>
</tr>
<tr>
<td>4.5 The General Orthogonal Gaussian Relay Channel</td>
<td>85</td>
</tr>
<tr>
<td>4.5.1 Basic Bounds</td>
<td>86</td>
</tr>
<tr>
<td>4.5.2 Proposed Scheme: Piecewise Linear Relaying</td>
<td>87</td>
</tr>
<tr>
<td>4.5.3 Numerical Examples and Discussion</td>
<td>90</td>
</tr>
<tr>
<td>4.6 Summary and Concluding Remarks</td>
<td>91</td>
</tr>
<tr>
<td>4.A Capacity Bound for Noncausal Relaying</td>
<td>98</td>
</tr>
</tbody>
</table>

### 5 Interference Management Using Relays

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>101</td>
</tr>
</tbody>
</table>
5.2 System Model .................................................. 102
5.3 Cutset Upper Bound ........................................... 104
5.4 Achievable Rates Using a Memoryless Relay ................. 105
   5.4.1 Benchmark: Linear Relaying .......................... 106
   5.4.2 Optimized Nonlinear Relaying ......................... 107
5.5 Relaying via Compress-and-Forward ........................ 110
5.6 Capacity of the Interference-Limited Channel ............... 112
5.7 Numerical Examples ........................................ 113
5.8 Summary and Concluding Remarks ............................ 114

6 Capacity Achieving Instantaneous Relaying .................... 119
   6.1 Introduction ................................................ 119
   6.2 Main Result ................................................ 121
      6.2.1 Proof of Theorem 6.2.1 ............................. 122
   6.3 Application to Gaussian Channels ......................... 126
   6.4 Summary and Concluding Remarks ......................... 128

7 State-Decoupled Relay Channels ............................... 129
   7.1 Introduction ................................................ 129
   7.2 A Capacity Theorem ....................................... 130
   7.3 A Fading Relay Channel .................................... 131
      7.3.1 Channel Model ...................................... 132
      7.3.2 Direct-Link Transmission ............................ 132
      7.3.3 Capacity of Strictly Causal Relaying ............... 132
      7.3.4 Capacity of Causal and Noncausal Relaying ........ 133
   7.4 A Relay Channel with Additive Interference ............... 134
      7.4.1 Channel Model ...................................... 134
      7.4.2 Capacity of Strictly Causal Relaying ............... 134
      7.4.3 Causal Relaying ..................................... 135
      7.4.4 Receiver Noise ....................................... 137
      7.4.5 Noncausal Relaying ................................... 138
      7.4.6 Numerical Examples ................................... 140
   7.5 Summary and Concluding Remarks .......................... 141

8 The Relay Channel with Noncausal State Information ........... 147
   8.1 Introduction ............................................... 147
   8.2 Definitions ............................................... 147
   8.3 Main Results ............................................. 149
      8.3.1 A Cutset Interpretation of $R$ ....................... 150
      8.3.2 Connection to Compress-and-Forward ................ 150
   8.4 Proof of Proposition 8.3.1 ................................ 151
      8.4.1 Codebook Generation ................................. 151
      8.4.2 Encoding ............................................. 151
      8.4.3 Decoding ............................................. 152
12 Conclusions and Future Research

12.1 Concluding Remarks ........................................ 197
12.2 Some Possible Extensions .................................... 200
12.3 Conjectures ....................................................... 201
   12.3.1 Capacity of Strictly Causal State-Decoupled Relay Channels 201
   12.3.2 Capacity of a Gaussian State-Decoupled Relay Channel . 202
   12.3.3 Capacity of Causal State-Decoupled Relay Channels ... 203

Bibliography ......................................................... 205
1.1 Background

Consider a wireless network with some active nodes including a source and a destination. Due to the broadcast nature of wireless transmission, some nodes in the network might overhear the transmitted signal from the source. If the direct communication of the source–destination pair fails (for example, due to the channel variations), those nodes that have a copy of the transmitted signal can help to re-establish or enhance the communication between the intended source–destination pair. The nodes that take part in the transmission are called relays. Generally speaking, when the relays (i.e., the nodes take part in the transmission in order to help the source–destination pair) cooperate, multiple paths for conveying the transmitter’s message becomes available. In other words, a virtual antenna array is created. Transmission via multiple paths in wireless communication often translates into diversity gain. From this setup one can envision various applications of having the relays.

The classical three-node relay channel [vdM71, CG79] models the communication between the source–destination pair and the relay. See the blue nodes in Figure 1.1 for an illustration. The relay channel has received a renewed attention in recent years due to its potential in wireless applications and ad hoc networks [LTL+06, SGL06]. Similar concepts such as user cooperation [SEA03a, SEA03b] and cooperative diversity [NHH04, HN06], has been introduced as a means to enhance the reliability of communication by an additional transmission of an intended message through another node. An immediate question which emerges from the setup in the relay channel is that how the relay should utilize its available resources—for example, power and bandwidth—to enhance the reliability of communication.

\footnote{In the thesis, we refer to a node that has an own message to transmit as the source. In other words, the message originates from the source and the source hence knows the message exactly. In the literature, the source is also called the sender or the transmitter. We also refer to a node which is interested in the message transmitted from the source as the destination. The destination is also known as the sink or the receiver in the literature. We, in the thesis, assume that no message originates from the relay, and that the relay is not the intended destination of the transmitted message.}
main theme of this thesis is to investigate this question. In the thesis, we explore coding schemes for three canonical examples of the relay networks: the three-node relay channel, the multiple-access relay channel and the two-way relay channel. Figure 1.1 illustrates the aforementioned relay channels.
1.2 Potential Applications of Relaying

In what follows, we briefly describe some potential applications of relaying.

1.2.1 Relay-Based Cellular Network

One immediate application is the deployment of fixed relays in cellular networks [PWS+04]. In the uplink (i.e., mobile to base station) transmission, practical constraints on the size of the handset and the battery lifetime preclude the use of multiple-input multiple-output (MIMO) antenna techniques. MIMO systems are well-known to offer multiplexing gain (i.e., an increase in data rate) and diversity gain (i.e., an increase in the slope at which the symbol-error-probability decreases with the average signal-to-noise ratio (SNR)) [TV05]. Thus one way to obtain the benefits of MIMO links is to use distributed MIMO by the deployment of the relays. Additionally, relay-based networks are more robust to shadow fading (i.e., attenuation of the transmitted signal due to large obstacles) since the placement of relays is not as restricted as that of collocated antennas in the conventional MIMO systems.

1.2.2 Wireless Sensor Networks

Wireless sensor networks [ASSC02] are composed of many nodes that are densely deployed in the environment. The main purpose of such networks usually is detection or estimation of a physical phenomenon. The sensors are supposed to work with a low power and have a sort of intelligence to establish the communication and to maintain the network. The collaboration among nodes is of crucial importance in such networks to achieve

- reliable estimation and detection; and
- reasonably long lifetime of the network.

In other words, some nodes in the network should take the role of being relays to maintain reliable interconnections between the nodes and to save energy.

1.2.3 Relay-Assisted Cognitive Radios

Cognitive radio [MM99] is motivated by the apparent geographical and temporal under-utilization of the spectrum. The idea is to sense a spectrum hole and try to utilize it in such a way not to harm primary users (i.e., licensed users). Clearly, one of main challenges of the cognitive radio is to sense the spectrum. This detection problem is intimately related to the cooperative communication in which the cognitive users help one another by relaying relevant information to sense the spectrum [GL07].
1.3 Outline and Contributions

As mentioned earlier, we study three examples of relay networks: three-node relay channels, multiple-access relay channels and two-way relay channels. We also investigate relay networks with state. In particular, we consider three-node relay channels with state and multiple-access relay channels with state. Figure 1.2 shows the dependency graph among the chapters, which can be used as a guideline for reading the thesis. Chapters 2–6 consider the three-node relay channels, Chapters 7 and 8 investigate the three-node relay channels with state, Chapter 9 studies multiple-access relay channels, and Chapters 10 and 11 discuss two-way relay channels.

We next give an outline of the thesis and describe the contributions within each chapter.

Chapter 2

In this chapter, we formally define the classical three-node relay channel and the performance metric that we use throughout the thesis. The main goal of this chapter is to review existing capacity bounds. We begin with a multi-letter characterization of the capacity and then discuss known single-letter bounds on the capacity: cutset upper bound, decode-and-forward (DF), partial decode-and-forward (PDF), compress-and-forward (CF) and instantaneous relaying. We also discuss classes of the relay channels for which the capacity is known. We also report some novel results in this chapter.

Chapter 3

We investigate capacity bounds for two instances of binary symmetric relay channels in this chapter. For one of these, we establish the capacity, and for the other one we present various capacity bounds. We first evaluate the cutset bound and the rates achievable using DF, PDF and CF relaying. We then focus on relaying strategies with finite memory-length, and present an efficient algorithm for optimizing the relay functions. We use the boolean Fourier transform to unveil the structure of the optimized mappings. Interestingly, the optimized relay functions exhibit a simple structure. The numerical results show that the rates achieved using the optimized low-dimensional functions are either comparable or superior to those achieved by DF and CF relaying. In particular, the optimized low-dimensional relaying scheme can improve on DF relaying when the quality of the source–relay link is worse than or comparable to that of other links.

The material in this chapter is partly published in [KS10c] and partly presented in [KSG11] for possible publication.
Figure 1.2: Dependency graph.
Chapter 4

This chapter investigates the three-node Gaussian relay channel with orthogonal receive components (i.e., the transmitted signals from the source and the relay do not interfere with each other). For such channels linear relaying (also known as amplify-and-forward (AF)) is a suboptimal strategy in general. This is because a linear scheme merely repeats the received noisy signal and does not utilize the available degrees of freedom efficiently. At this background, we hence develop nonlinear, symbol-wise (as opposed to block-wise) relaying strategies to compensate for the shortcomings of the linear strategy.

We present optimal strategies for two special cases of the general scenario and show that memoryless relaying can achieve the capacity. Furthermore, for the general Gaussian relay channel, a parametric piecewise linear (PL) mapping is proposed and analyzed. The achievable rates obtained by the PL mapping are computed numerically and optimized for a certain number of design parameters. It is concluded that the optimized PL relaying always outperforms the conventional instantaneous linear relaying (i.e. AF). We also illustrate that the proposed PL relaying scheme can improve on sophisticated DF when the source–relay link is ill-conditioned (relative to other links). Furthermore, PL relaying can work at rates close to those achieved by CF, but at a much lower complexity.

The material in this chapter is partly published in the papers [KS09c, KS10b].

Chapter 5

This chapter studies the three-node Gaussian relay channel where the relay observes a noisy version of the interference present in the source–destination link. Motivated by the results in Chapter 4, we design nonlinear strategies for this scenario. In this chapter, we apply a variational method to derive a necessary condition for the optimality and then use the fixed point iteration method to solve the necessary condition. We additionally investigate linear relaying and CF. For interference-limited cases, we illustrate that optimized memoryless nonlinear relaying almost achieves the capacity.

The material in this chapter is partly published in the paper [KZS10].

Chapter 6

We, in this chapter, introduce a class of semi-deterministic relay channels and show that for channels in the class, capacity can be achieved using an instantaneous relaying scheme. We then discuss applications to Gaussian channels. This class of relay channels subsumes some special cases considered in Chapter 4 and Chapter 5. In particular, we prove that instantaneous relaying is enough to achieve the capacity for the interference limited relay channel in Chapter 5.

The material in this chapter will be partly presented in the paper [KS11a].
Chapter 7
In this chapter, we introduce the class of state-decoupled relay channels and establish the capacity of some semi-deterministic cases of such channels. We then compute the capacity of two examples of relay channels with antipodal multiplicative fading and Gaussian additive interference, respectively. In this chapter, we also discuss the effect of delay on the capacity of the relay channels. In particular, we show that one can increase the capacity of state-decoupled relay channels when the relay incorporates the current received signal in addition to the past ones.

The material in this chapter is partly presented in the paper [KGS11] for possible publication.

Chapter 8
We study the three-node relay channel with state in this chapter. We assume that the channel state is determined by a random parameter, and that noncausal knowledge of the state is partially available at the nodes. In this chapter, we first derive an achievable rate inspired by the noisy network coding (NNC) approach in [LKGC11] and the random binning scheme in [GP80]. We then derive an equivalent representation of the rate similar to that in [GMZ06] and show that the proposed NNC-based scheme achieves the same rate as that achieved by CF for the relay channel with state as proposed by [AMA10].

The material in this chapter is partly published in the paper [KS11e].

Chapter 9
This chapter considers multiple-access relay channels (MARCs). We find the capacity of two semi-deterministic classes of MARCs: MARCs with orthogonal receive components and state-decoupled MARCs.

The material in this chapter is partly presented in the paper [KGS11] for possible publication.

Chapter 10
In this chapter, we investigate a Gaussian two-way relay channel (TWRC) and develop nonlinear analog noisy network coding (ANNC) for this channel. When the received signal at the relay is noiseless, we present an optimal nonlinear mapping. For the general case when the received signal at the relay is noisy, we provide an optimization algorithm based on the fixed point iteration method to find optimized mappings. Our results show that the proposed scheme with optimized nonlinear ANNC can significantly outperform CF as proposed in [RW06] and linear ANNC [RW06, PY07, KMG+07]. Additionally, the proposed scheme achieves the rates close to those achieved by layered NNC proposed in [LKG10].

The material in this chapter is partly published in the paper [KS11d].
Chapter 11

We investigate TWRCs and propose three schemes based on hybrid digital–analog relaying, in this chapter. In the proposed schemes, the relay first employs digital compression to generate a digital signal. The relay then component-wise combines the encoded compression index with the analog noisy received signal. We refer to our proposed schemes as hybrid digital–analog NNC protocols, since the digital signal used for encoding of the compression index is combined with the analog noisy signal in the network. Our main conclusion is that the proposed hybrid NNC outperforms both purely-digital and purely-analog NNC.

The material in this chapter is partly published in the paper [KS11b].

Chapter 12

In this chapter, we finally conclude the thesis and suggest some directions for future research.

1.4 Contributions Outside the Thesis

The contributions in the published papers [KL08, KL07a, KS11c, KL07c, KL07d, ZKYS09a, ZKYS09b, ZKYS09c, KL09b, KL07b, KL09a, KS09d, KS09a, KS09b, KS09c] are not included in the thesis. These papers can be categorized as follows:

- Constellation Constrained Relaying
  In [KL08, KL07a], we propose a constellation rearrangement technique in which the relay uses a bit-symbol mapping that is different from the one used by the source. We find the optimal bit-symbol mappings for both the source and the relay and the associated optimal detectors, and show that the improvement over conventional relaying with Gray mapping at the source and the relay can amount to a power gain of several dBs. This technique is extended to MIMO link in [KS11c]. In [KL07c], we study the receiver design for relay networks with antipodal transmission. High-SNR and low-SNR approximations of the detectors are presented as well. A relaying scheme based on the constellation rotation for multi-dimensional signal constellations is proposed in [KL07d]. Additionally, in the papers [ZKYS09a, ZKYS09b, ZKYS09c], we propose a general method for optimizing memoryless relay mappings including the constellation constrained relay channels.

- Finite-SNR Optimization of DF Relaying
  In the papers [KL09b, KL07b, KL09a], we investigate the finite-SNR optimization of relaying schemes based on DF relaying. In [KL09a], we propose a novel half-duplex DF relaying scheme based on partial repetition coding at the relay. In the proposed scheme, if the relay decodes the received message successfully, it re-encodes the message using the same channel code as the one
used at the source, but retransmits only a fraction of the codeword. We analyze the proposed scheme and optimize the cooperation level (i.e., the fraction of the message that the relay should transmit).

- **Relay Channels with State**

In [KS09d], we investigate the relay channel with a random parameter whose causal knowledge is available partially at the source, the relay and the destination. We define four different classes of this scenario and establish lower and upper bounds on the capacity. The proposed schemes are combinations of different block Markov encoding schemes in [CG79] and Shannon strategy [Sha58] for point-to-point links. We also define a class of degraded relay channel for which the bounds are tight.

In [KS09a, KS09b, KS09e], we study a three-node Gaussian relay channel with interference which is noncausally known at the source. It is assumed that the interference affects only the relay–destination link. This model is motivated by a downlink scenario, where the source (base station), communicates with two destinations. We present several transmission strategies for this class of relay channels. Our proposed relaying schemes can be divided into two main categories: instantaneous relaying and strictly causal relaying. In the former, the relay functionality is restricted to a memoryless, symbol-by-symbol mapping (linear as well as nonlinear). While in the latter, the relay has an infinite memory and utilizes past received blocks to cooperate in the present block. For causal relaying, we investigate DF, CF, combined DF and CF. To utilize the knowledge of the interference at the source, superposition coding [CT06] at the source and Costa encoding [Cos83] at the relay are employed. One interesting finding is that instantaneous relaying can achieve higher rates than those achieved by strictly causal relaying.
1.5 Notation

\( M \) : A set of messages.
\( M \) : A message to be transmitted.
\( \hat{M} \) : A detected message when \( M \) is transmitted.
\( p(x) \) : probability mass function (pmf) or probability density function (pdf) of the random variable \( X \).
\( p(x, y) \) : joint pmf or pdf of the random variables \( X \) and \( Y \).
\( p(x|y) \) : conditional pmf or pdf of the random variable \( X \) given \( Y = y \).

\( X \sim p(x) \) : The random variable \( X \) is distributed according to \( p(x) \).
\( X \rightarrow Y \rightarrow Z \) : A Markov chain; i.e. \( p(z|y, x) = p(z|y) \).
\((X, p(y|x), Y)\) : The channel with input alphabet \( X \), output alphabet \( Y \), and the pmf \( p(y|x) \) where \( x \in X \) and \( y \in Y \).

\( \mathbb{E}[X] \) : Expectation of the random variable \( X \).
\( \mathbb{E}[X|Y] \) : Expectation of the random variable \( X \) given \( Y \).

\( H(X) \) : Entropy of the discrete random variable \( X \).
\( H(X, Y) \) : Joint entropy of the discrete random variables \( X \) and \( Y \).
\( H(X|Y) \) : Conditional entropy of the discrete random variable \( X \) given \( Y \).
\( h(X) \) : Differential entropy of the continuous random variable \( X \).
\( h(X, Y) \) : Joint differential entropy of the continuous random variables \( X \) and \( Y \).
\( h(X|Y) \) : Conditional differential entropy of the continuous random variable \( X \) given \( Y \).
\( h_\Delta(X) \) : Differential entropy of the random variable \( X \) when the associated pdf depends on \( \Delta \).
\( I(X; Y) \) : Mutual information between the random variables \( X \) and \( Y \).
\( I(X; Y|Z) \) : Conditional mutual information between the random variables \( X \) and \( Y \) given \( Z \).
\( \text{Ber}(\epsilon) \) : Bernoulli pmf with parameter \( \epsilon \).
\( \mathcal{N}(\mu, \sigma^2) \) : Gaussian pdf with the mean \( \mu \) and the variance \( \sigma^2 \).
1.5. Notation

\( GF(2) \): Denotes the binary Galois filed, i.e. \( \{0, 1\} \).

\( GF^k(2) \): Denotes the \( k \)-dimensional binary Galois filed, i.e. \( \{0, 1\}^k \).

\( x^n \): \((x_1, x_2, \ldots, x_n)\).

\( x_{ij}^k \): \((x_{i,j}, x_{i,j+1}, \ldots, x_{i,k})\).

\([1 : n]\): Defines the set \( \{1, 2, \ldots, n\} \).

\( \mathbb{R} \): The set of real numbers.

\( \{f_i\}_{i=1}^k \): The set of functions \( \{f_1, f_2, \ldots, f_k\} \).

\( T^{(n)}_\epsilon \): \( \epsilon \)-typical sequences of length \( n \).

\( \mathcal{X} \times \mathcal{Y} \): Cartesian product of \( \mathcal{X} \) and \( \mathcal{Y} \).

\( A \subseteq B \): \( B \) includes \( A \).

\(|A|\): Cardinality of the set \( A \).

\( A^c \): Complement of the set \( A \).

\( := \): Equality by definition.

\( \in \): Belongs to.

\( C(x) \): \( \frac{1}{2} \log_2 (1 + x) \).

\( Q(x) \): \( \frac{1}{\sqrt{2\pi}} \int_1^\infty \exp \left( -\frac{t^2}{2} \right) dt \).

\( H_\epsilon(\epsilon) \): \(-\epsilon \log_2(\epsilon) - (1 - \epsilon) \log_2(1 - \epsilon)\).

\( \epsilon_1 * \epsilon_2 \): \( \epsilon_1 + \epsilon_2 - 2 \epsilon_1 \epsilon_2 \).

\( D_H(x^n, y^n) \): The Hamming distance between the binary sequences \( x^n \) and \( y^n \).

\( \delta(t) \): Dirac delta function.

\( \delta_b(t) \): Binary Dirac delta function.

\( I_n \): The identity matrix of dimension \( n \times n \).

\(|K|\): Determinant of the matrix \( K \).

\( \|x\| \): The Euclidean norm of the vector \( x \).

\( \lfloor x \rfloor \): The largest integer less than or equal to \( x \).

\( \min \): Minimum.

\( \max \): Maximum.

\( \inf \): Infimum.

\( \sup \): Supremum.

\( \text{Supp}(p) \): Support of \( p \).
1.6 Acronyms

AF Amplify-and-Forward
ANNC Analog Noisy Network Coding
BPSK Binary-Phase Shift Keying
BSC Binary Symmetric Channel
BSRC Binary Symmetric Relay Channel
bpcu bits per channel use
CF Compress-and-Forward
CR Constellation Rearrangement
cdf cumulative density function
DF Decode-and-Forward
DeF Demodulate-and-Forward
dB decibel
EF Estimate-and-Forward
HDA Hybrid Digital–Analog
HF Hash-and-Forward
i.i.d. independent and identically distributed
MAC Multiple-Access Channel
MARCs Multiple-Access Relay Channel
MIMO Multiple-Input Multiple-Output
MMSE Minimum Mean Squared Error
MSE Mean Squared Error
NNC Noisy Network Coding
PDF Partial Decode-and-Forward
PL Piecewise Linear
pdf probability density function
pmf probability mass function
RC Relay Channel
SIMO Single-Input Multiple-Output
SNR Signal-to-Noise Ratio
s.t. Subject to
TWRC Two-Way Relay Channel
QF Quantize-and-Forward
Chapter 2

The Classical Relay Channel

This chapter introduces the classical three-node relay channel and discusses capacity bounds. We first give a multi-letter characterization of the capacity of the channel and then present known single-letter bounds. We discuss an upper bound and several lower bounds. For each lower bound, we give a class or an example of relay channels for which the lower bounds are tight. However the single-letter characterization of the capacity remains open for the general case.

2.1 Definitions

The discrete memoryless relay channel was introduced by van der Meulen in 1971 [vdM71]. The channel consists of three nodes; a source-destination pair and a relay node. We assume that the relay has no private message and its sole purpose is to help the communication between the source and the destination. Figure 2.1 shows a block diagram of the relay channel. The channel is defined using two finite input sets $\mathcal{X}$ and $\mathcal{X}_r$, two finite output sets $\mathcal{Y}$ and $\mathcal{Y}_r$, and a probability mass function (pmf) $p(y, y_r | x, x_r)$ which governs the stochastic input-output relationship over the channel.\footnote{For brevity, we denote $p_{Y|X}(y|x)$ as $p(y|x)$ when there is no risk of confusion.} Therefore, in Figure 2.1 we have $X^n \in \mathcal{X}^n$, $Y^n \in \mathcal{Y}^n$, $X^n_r \in \mathcal{X}^n_r$ and $Y^n_r \in \mathcal{Y}^n_r$.\footnote{We adopt the notation $X^i := (X_1, X_2, \cdots, X_i)$.} Moreover, the channel is assumed to be memoryless:

$$p \left( y^n, y^n_r | x^n, x^n_r \right) = \prod_{i=1}^{n} p \left( y_i, y_{ri} | x_i, x_{ri} \right).$$

That is, the received signals at the relay and the destination at time instant $i$ depend only on the transmitted signals from the source and the relay at time instant $i$ (i.e., $x_i, x_{ri}$). The memoryless relay channel is generally denoted as

$$(\mathcal{X} \times \mathcal{X}_r, p(y, y_r | x, x_r), \mathcal{Y} \times \mathcal{Y}_r).$$

The goal of the communication is to reliably convey the message $M$ uniformly drawn from the set $\mathcal{M}^{(n)}$ to the destination in $n$ channel uses. To accomplish this task, one needs to design three main components:
• an encoder to map the message $M$ to $X^n$;
• a set of relay functions: $\{f_i\}_{i=1}^n$ such that

$$x_{ri} = f_i(y_{r1}, y_{r2}, \ldots, y_{ri-1}), \quad \forall i \in [1:n];$$

and

• a decoder to map the received signal $Y^n$ to an estimate of the transmitted message $\hat{M}$.

Having designed a transmission strategy according to the above, the joint pmf of the transmitted message, the transmitted and received sequences, and the detected message is given by

$$p(m, x^n, x^n_r, y^n, y^n_r, \hat{m}) = p(m)p(x^n|m)p(\hat{m}|y^n) \prod_{i=1}^n p(x_{ri}|y_{ri}^{i-1})p(y_{ri}|y_{ri}^{i-1}, x_{ri}, x_{ri}).$$

(2.1)

The rate $R$ is achievable, if there exists a communication strategy, such that the average probability of error

$$P_e^{(n)} := \Pr \{ \hat{M} \neq M \},$$

at the decoder tends to zero as $n \to \infty$ and at the same time

$$\liminf_{n \to \infty} \frac{1}{n} \log |M(n)| \geq R.$$

The capacity of the channel is defined as the supremum of all achievable rates $R$.

### 2.2 Infinite-Letter Characterization of the Capacity

The capacity of the relay channel is not known in general. However, an infinite-letter characterization of the capacity is possible and it follows from a standard random coding argument.

**Theorem 2.2.1** ([vdM71, Zah05]). The capacity of the relay channel is given by

$$C = \lim_{k \to \infty} \sup_{\{f_i\}_{i=1}^k, s(x^k)} \frac{1}{k} I(X^k, Y^k).$$

(2.2)

---

3Here we assume **strictly causal** relaying, that is the transmitted signal at instant $i$ only depends on the received signals up to time instant $i - 1$. We define and investigate causal and noncausal relaying in Chapter 7.

4Unless otherwise specified, the logarithms are taken to the base two and the rates are hence measured in bits per channel use [bpcu].
2.2. Infinite-Letter Characterization of the Capacity

\[ \prod_{i=1}^{n} p(y_i, y_{ri}|x_i, x_{ri}) \]

\[ \text{Encoder} \]
\[ \text{Source} \]
\[ \text{Relay} \]
\[ \text{Decoder} \]
\[ X^n \]
\[ Y^n \]
\[ X^n_r \]
\[ n \]
\[ \hat{M} \]

Figure 2.1: Block diagram of the relay channel [CG79].

**Proof.** We first prove the achievability. Let \( \mathcal{M}^{(n)} := [1 : 2^{nR}] \). Then fix the relay functions \( f_1, f_2, \ldots, f_k \) and the pmf \( p_{X^k}(x^k) \). (Note that the relay channel can be seen as a point-to-point link parameterized by the pmf \( p(x^k|y^k) \).) Now generate \( 2^{nR} \) independent and identically distributed (i.i.d.) \( x^{nk}(m) \) sequences of length \( nk \) each distributed according \( \prod_{j=1}^{n} p_{X^k}(x_{(j-1)k+1}^k) \), for all \( m \in \mathcal{M}_n \). In order to send the message \( m \), the source transmits \( x^{nk}(m) \). The relay operates strictly causally using the functions \( \{f_i\}_{i=1}^k \). The destination upon receiving \( y^{nk} \), looks for a unique index \( \hat{m} \) such that

\[ (x^{nk}(\hat{m}), y^{nk}) \in T^{nk}_\epsilon; \]
otherwise it declares an error. Here \( T^{nk}_\epsilon \) denotes the set of \( \epsilon \)-typical sequences of length \( nk \) and is defined in Appendix 2.A. The probability of error at the destination is negligible if

\[ R < \sup_{\{f_i\}_{i=1}^k} R_k(f_1, \ldots, f_k) = \sup_{\{f_i\}_{i=1}^k, p(x^k)} \frac{1}{k} I(X^k; Y^k) =: C_k. \quad (2.3) \]

This follows from the packing lemma in Appendix 2.A. We note that the sequence \( \{C_k\} \) is non-decreasing and is bounded from above. Thus \( \lim_{k \to \infty} C_k =: C \) exists.

We next prove the converse; i.e., no higher rate than \( C \) is achievable when the average message error probability is negligible. Consider

\[ nR \overset{(a)}{=} H(M) \]
\[ = I(M; Y^n) + H(M|Y^n) \]
\[ \overset{(b)}{\leq} I(M; Y^n) + H(M|\hat{M}(Y^n)) \]
\[ \overset{(c)}{\leq} I(X^n; Y^n) + H(M|\hat{M}(Y^n)) \]
\[ \overset{(d)}{\leq} I(X^n; Y^n) + n\epsilon_n \]
\[ \overset{(e)}{\leq} n \left( \sup_n C_n + \epsilon_n \right) \]
The Classical Relay Channel

\[(f) \leq n (C + \epsilon_n), \quad (2.4)\]

where

\((a)\) holds because the messages are independent and are uniformly drawn from the set \(M^{(n)};\)

\((b)\) follows by the data processing inequality [CT06];

\((c)\) holds because \(M \rightarrow X^n \rightarrow Y^n\) form a Markov chain;

\((d)\) follows by Fano’s inequality [CT06]

\[H(M|M(Y^n)) \leq 1 + \mathbb{P}^{(n)}_n R =: n\epsilon_n, \quad (2.5)\]

where \(\epsilon_n \rightarrow 0\) as \(n \rightarrow \infty\) for a negligible probability of error;

\((e)\) follows by the definition of \(C_n;\) and

\((f)\) holds because \(C_n\) is non-decreasing in \(n.\)

This completes the proof. \(\square\)

Unfortunately, the above expression is not computable in general. We next provide an overview of some known computable bounds on the capacity.

2.3 Single-Letter Bounds on the Capacity

2.3.1 Cutset Upper Bound

In the following, we present an upper bound on the capacity. This upper bound is known as \textit{max-flow min-cut} or \textit{cutset} bound [CG79] and it consists of two terms:

- Broadcast bound: \(C \leq I(X;Y,Y_r|X_r);\)
- Multiple-access bound: \(C \leq I(X,X_r;Y).\)

Figure 2.2 schematically illustrates these two bounds. One can easily see the analogy between the broadcast bound and receive diversity in which the relay plays the role of another receiver. On the contrary in the multiple-access bound, the relay acts as another transmitter as it knows the true message. Combining these two bounds, we obtain

\[C \leq R_{ub} = \max_{p(x,x_r)} \min \{I(X,X_r;Y), I(X;Y,Y_r|X_r)\}, \quad (2.6)\]

where the maximum is taken over all joint pmfs of the form \(p(x,x_r).\)
2.3. Single-Letter Bounds on the Capacity

2.3.2 Direct-Transmission Lower Bound

Let the relay choose to transmit a fixed symbol over a block. That is $X_{ri} = x_r$ for all $i \in [1 : n]$. Then apply the standard random coding argument for the equivalent point-to-point channel. Thus the rate

$$R(X_r = x_r) = \max_{p(x)} I(X; Y|X_r = x_r),$$  \hfill (2.7)

is achievable. Now by optimizing over the relay’s symbols, we obtain that

$$R_D = \max_{p(x), x_r} I(X; Y|X_r = x_r),$$  \hfill (2.8)

is achievable.

This simple strategy achieves the capacity if the channel pmf can be decomposed as

$$p(y, y_r|x, x_r) = p(y|x, x_r)p(y_r|y, x_r).$$  \hfill (2.9)

That is $X \rightarrow (X_r, Y) \rightarrow Y_r$ form a Markov chain or equivalently $X \rightarrow Y \rightarrow Y_r$ form a Markov chain conditioned on $X_r$. This class of the relay channel is referred to as *reversely degraded* relay channel [CG79].

2.3.3 Two-Hop Lower Bound

Assume that the relay decodes the received signal from the sender at the end of a block, and then it re-encodes the message and transmits the encoded message. It is easy to verify that [GK10]

$$R_{2-\text{Hop}} = \max_{p(x)p(x_r)} \min \{I(X_r; Y), I(X; Y_r|X_r)\},$$  \hfill (2.10)

is achievable. This strategy achieves the capacity if the channel pmf can be decomposed as

$$p(y, y_r|x, x_r) = p(y_r|x)p(y|x_r).$$  \hfill (2.11)
That is the relay channel is a cascade of two point-to-point links. Note that under the condition given in (2.11), (2.10) simplifies to

\[ C = \max_{p(x)p(x_r)} \min \{ I(X_r;Y), I(X;Y_r) \} \]

\[ = \min \left\{ \max_{p(x_r)} I(X_r;Y), \max_{p(x)} I(X;Y_r) \right\}. \tag{2.12} \]

The converse follows from the cutset bound in (2.6).

2.3.4 Decode-and-Forward Lower Bound

With block Markov encoding [CG79] (also known as decode-and-forward (DF) [LWT08, KGG05, HMZ06]), the transmission occurs in \( B \) consecutive blocks. The relay transmits a fresh codeword by decoding the previous block received at the relay. This scheme relies on the successful decoding at the relay and hence its performance is limited by the quality of the source–relay link. Using DF, the following rate is achievable [CG79]

\[ R_{DF} = \max_{p(x,x_r)} \min \{ I(X,X_r;Y), I(X;Y_r|X_r) \}, \tag{2.13} \]

where

- \( I(X,X_r;Y) \) is the multiple-access bound on the capacity; and
- \( I(X;Y_r|X_r) \) reflects the capacity of the source–relay link.

The proposed DF scheme in [CG79] uses binning. However one can implement DF without binning using either backward decoding [WvdM85] or sliding window decoding as proposed in [XK05] and yet achieve the same rate as that given in (2.13). We next briefly explain how DF works with backward decoding. Consider transmission of \( B \) blocks, where each block consumes \( n \) channel uses. In this scheme a sequence of \( B - 1 \) i.i.d. messages \( m_j \), uniformly chosen from \( \mathcal{M}^{(n)} := [1:2^{nR}] \), for all \( j \in [1: B - 1] \) is transmitted to the destination. The transmission rate \( \frac{R(B-1)R}{nB} \) approaches \( R \) as \( B \to \infty \).

- Codebook generation: Fix \( p(x,x_r) \). For each \( j \in [1:B] \), randomly and independently generate \( 2^{nR} \) sequences \( x^n(m_j) \) for all \( m_j \in \mathcal{M}^{(n)} \), each according to \( \prod_{i=1}^{n} p_{X_i}(x_i) \). For each \( m_{j-1} \in \mathcal{M}^{(n)} \), randomly and conditionally independently generate \( 2^{nR} \) sequences \( x^n(m_j|m_{j-1}) \) for all \( m_j \in \mathcal{M}^{(n)} \), each according to \( \prod_{i=1}^{n} p_{X_i|X_{r_i}}(x_i|x_{r_i}(m_{j-1})) \). The codebook is then given by

\[ \mathcal{C}_j = \left\{ (x^n(m_j|m_{j-1}), x^n(m_{j-1}^r)) : m_{j-1}, m_j \in \mathcal{M}^{(n)} \right\}, \forall j \in [1:B]. \]
• **Encoding:** To send $m_j$ in block $j$, the encoder transmits $x^n(m_j|m_{j-1})$. At the end of block $j$, the relay has an estimate $\hat{m}_j$ of message $m_j$ and transmits $x^n_r(\hat{m}_j)$ in block $j + 1$. Additionally set $m_0 = 1$ and $m_B = 1$.

• **Decoding:** The relay at the end of block $j$ receives $y^n\left(j\right)$ and knows $m_{j-1}$. It then looks for a unique index $\tilde{m}_j \in \mathcal{M}_n(\epsilon)$ such that

$$(x^n(m_j|\hat{m}_{j-1}), x^n_r(\hat{m}_{j-1}), y^n\left(j\right)) \in T(\epsilon);$$

otherwise it declares an error. By the packing lemma in Appendix 2.A, this is done with a negligible probability of error if

$$R < I(X; Y_r|X_r) - \delta(\epsilon). \quad (2.14)$$

The destination waits until it receives all blocks and performs the decoding backward at the end of block $B$. The destination at step $j$ knows $m_j$ and looks for a unique index $\hat{m}_{j-1} \in \mathcal{M}_n(\epsilon)$ such that

$$(x^n(m_j|\hat{m}_{j-1}), x^n_r(\hat{m}_{j-1}), y^n\left(j\right)) \in T(\epsilon);$$

otherwise it declares an error. By the packing lemma in Appendix 2.A, this is done with a negligible probability of error if

$$R < I(X, X_r; Y) - \delta(\epsilon). \quad (2.15)$$

Combining (2.14) and (2.15) yields (2.13).

As it can be understood for the above code construction, DF with backward decoding suffers from long decoding delay at the destination. However the code construction using binning in [CG79] or sliding window decoding in [XK05] has much smaller decoding delay.

One can readily deduce that DF operates close to the capacity upper bound when the quality of the source–relay link is “reasonably” good. In particular, DF relaying achieves the capacity when the relay channel is physically degraded [CG79]; that is the channel pmf can be decomposed as

$$p(y, y_r|x, x_r) = p(y_r|x, x_r)p(y|y_r, x_r),$$

or equivalently $X \rightarrow (Y_r, X_r) \rightarrow Y$ form a Markov chain. This, loosely speaking, means that the relay receives a better copy of the transmitted message than the destination. Note that by this definition the broadcast bound in (2.6) simplifies to $I(X; Y_r|X_r)$ and the rate obtained by DF hence coincides with the cutset bound.
2.3.5 Partial Decode-and-Forward Lower Bound

In some cases, it is possible to improve on DF by using partial decode-and-forward (PDF) at the relay. That is, the relay only decodes a part of the transmitted message. The achievable rate of PDF is given by [CG79, GA82]

$$R_{PDF} = \max_{p(u, x, x_r)} \min \{ I(X, X_r; Y), I(U; Y_r | X_r) + I(X; Y | X_r, U) \}, \quad (2.17)$$

where $U \rightarrow (X, X_r) \rightarrow (Y_r, Y)$ form a Markov chain. The random variable $U$ denotes a part of the transmitted message that the relay decodes. Note that by choosing $U = X$, (2.17) simplifies to (2.13).

The PDF protocol achieves the capacity of the relay channel if $y_r = g(x, x_r)$ [GA82]. That is, the received signal at the relay is a deterministic function of transmitted symbols from the source and the relay. Since the source knows both $X$ and $X_r$, it then knows $Y_r = g(X, X_r)$. Now let $U = Y_r$. Then (2.17) simplifies to

$$C = \max_{p(x, x_r)} \min \{ I(X, X_r; Y), H(Y_r | X_r) + I(X; Y | X_r, Y_r) \}, \quad (2.18)$$

which coincides with the cutset bound.

The PDF approach also achieves the capacity of the relay channel with orthogonal transmit components. See Figure 2.3 for an illustration. Let $X = (X_1, X_2)$ where $X_1$ is transmitted via the direct link to the destination and $X_2$ is only transmitted to the relay. If the channel pmf is decomposed as

$$p(y, y_r | x, x_r) = p(y_r | x_2, x_r)p(y | x_1, x_r). \quad (2.19)$$

then the capacity is given by

$$C = \max_{p(x_r) p(x_1 | x_r) p(x_2 | x_r)} \min \{ I(X_1, X_r; Y), I(X_2; Y_r | X_r) + I(X_1; Y | X_r) \}. \quad (2.20)$$

The achievability follows by choosing $U = X_2$ in (2.17) and the converse follows by the cutset bound.

The relay channel with orthogonal transmit components is closely related to the multiple-access channel (MAC) with partially cooperative encoders studied by
Willems in [Wil83]. Figure 2.4 illustrates the MAC with partially cooperate links where $C_{12}$ and $C_{21}$ denote the capacity of noiseless links between the encoders. The capacity region of this channel is given by

$$\mathcal{R} = \left\{ (R_1, R_2) : R_1 \leq I(X_1; Y|X_2; U) + C_{12} \right. \\
R_2 \leq I(X_2; Y|X_1; U) + C_{21} \\
\left. R_1 + R_2 \leq \min\{I(X_1, X_2; Y), I(X_1, X_2; Y|U) + C_{12} + C_{21}\} \right\}, \quad (2.21)$$

for some $p(u)p(x_1|u)p(x_2|u)$.

Now let $C_{21} = 0$, $C_{12} = R_0$ and think of the second user as a relay with no private message; i.e. $M_2 = \emptyset$. Thus, the channel in Figure 2.4 simplifies to that in Figure 2.5. The capacity of the relay channel in Figure 2.5 is given by

$$C = \max_{p(x_1, x_r)} \min\{I(X_1, X_r; Y), I(X_1; Y|X_r) + R_0\}. \quad (2.22)$$

This can be concluded either from (2.21) by choosing $R_2 = 0$, $C_{12} = R_0$, $C_{21} = 0$ and $U = \emptyset$ or from (2.20).
2.3.6 Compress-and-Forward Lower Bound

DF relaying requires that the relay decodes the received signal and it therefore results in a performance degradation when the source–relay link is not strong enough. In a situation when successful decoding is not possible, the relay can transmit an estimate of the received signal ($\hat{Y}_r$) and the destination can first recover the estimate $\hat{Y}_r$ using the side information $Y$ and then decode the transmitted message.

There are several ways to implement compression-based relaying. Figure 2.6 illustrates compress-and-forward (CF) relaying as proposed in [CG79]. In this scheme the relay combines quantization and binning together. This approach is also known as Wyner-Ziv based CF since the relay employs binning. In Figure 2.6, $R_0$ is the binning rate and $\hat{R}$ is the compression rate. In the CF scheme in [CG79], a block Markov encoding scheme is used to send $B-1$ i.i.d. messages in $B$ blocks. At the end of block $j$, the reproduction sequence $\hat{y}_n^j(j)$ which describes $y_n^j(j)$ conditioned on $x_n^j(j)$ is constructed at the relay using a test channel. The sequence $x_n^j(j)$ is known to both the relay and receiver at the end of block $j$. As illustrated in Figure 2.6 the relay transmits the encoded binning index $x_n^j(j+1)$ at the beginning of block $j+1$. The binning is used to reduce the transmission rate required to describe $y_n^j$ to the destination. The reduction in rate is possible since the destination has access to $y^n$ which is correlated to $y_n^j$. The destination at the end of block $j+1$ decodes the binning index $j+1$ and uses $y_n^j(j)$ and $x_n^j(j)$ to simultaneously decode $\hat{y}_n^j(j)$ and the message transmitted in block $j$. Using this scheme one can show that the rate [CG79]

$$ R_{CF} = \max_{p(x)p(x_r)p(\hat{y}_r|x_r,y)} I(X;Y,\hat{Y}_r|X_r), $$

s.t. $I(Y_r;\hat{Y}_r|X_r,Y) \leq I(X_r;Y)$ (2.23)

is achievable.

If the relay–destination link supports high data rate transmission (i.e., $I(X_r;Y)$ is very large), the destination can recover $Y_r$ with low distortion and the rate obtained by this scheme hence approaches the broadcast bound on the capacity. This scheme is generally referred to as CF [KGG05, HMZ06]. The representation in (2.23) is not generally suitable for optimization of the achievable rate. Therefore an equivalent representation is obtained in [GMZ06] where

$$ R_{CF} = \max_{p(x)p(x_r)p(\hat{y}_r|x_r,y_r)} \min \left\{ I(X,X_r;Y) - I(Y_r;\hat{Y}_r|X,X_r,Y), I(X;Y,\hat{Y}_r|X_r) \right\}. $$

(2.24)

In [LKGC11], it is shown that the above equivalent representation can be in fact obtained from a so-called noisy network coding (NNC) scheme. In NNC, the relay simply performs quantize-and-forward (QF) and does not use Wyner-Ziv compression as it is done in the original CF protocol described in above. Figure 2.7 illustrates QF relaying.
2.3. Single-Letter Bounds on the Capacity

Figure 2.6: Illustration of compress-and-forward.

Figure 2.7: Illustration of quantize-and-forward.
Because of practical limitations in some applications, the relay node is incapable of simultaneously receiving and transmitting in the same frequency band. Therefore, we also consider a relay with distinct frequency bands for reception and transmission. This assumption results in orthogonal reception at the destination. That is, the received signals from the source and the relay do not interfere with each other.

**Definition 2.3.1 (The Relay Channel with Orthogonal Receive Components).** The relay channel is said to have orthogonal receive components if the received signal at the destination is \( Y \equiv (Y_1, Y_2) \) where \( Y_1 \) and \( Y_2 \) respectively are signals received from the source and the relay, and the channel pmf is decomposed as

\[
p(y, y_r | x, x_r) = p(y_1, y_r | x)p(y_2 | x_r).
\] (2.25)

That is there is an orthogonal link from the relay to the destination. See Figure 2.8 for an illustration.

If the relay channel has orthogonal receive components and \( Y_r = f(X, Y_1) \), i.e. the received signal at the relay is a deterministic function of \( X \) and \( Y_1 \), then the capacity is given by

\[
C = \max_{p(x)} \min \{ I(X; Y_1) + R_0, I(X; Y_1, Y_r) \}.
\] (2.26)

where \( R_0 := \max_{p(x_r)} I(X_r; Y_2) \). The achievability follows by choosing \( \hat{Y}_r = Y_r \) in (2.24) and the converse follows from the cutset bound. In [Kim08], it is proved that one can achieve the same rate by simply hashing the received signals at the relay. Figure 2.9 illustrates hash-and-forward (HF) relaying. With HF, the relay uniformly partition all possible received sequences into \( 2^{nR_0} \) bins and associates to each bin a randomly and independently generated codeword. Upon receiving \( y_r^n \) the relay finds the corresponding bin index \( b = B(y_r^n) \) and transmits the associated codeword \( x_r^n(b) \) to that bin.

---

Binning is defined as a random and uniform partitioning of a set of sequences.
2.3. Single-Letter Bounds on the Capacity

| \(|Y| n\) sequences | \(2^{nR_0}\) bins | \(2^{nR_0}\) codewords |
|----------------------|-----------------|------------------|
| ![Diagram](image.png) | ![Diagram](image.png) | ![Diagram](image.png) |

Figure 2.9: Illustration of hash-and-forward.

The destination first constructs the list of codewords \(x^n(m)\) that are jointly typical with the received sequence \(y^n_1\):

\[
\mathcal{L}_1 = \left\{ m : (x^n(m), y^n_1) \in T_\varepsilon^{(n)} \right\}. \tag{2.27}
\]

The destination then decodes the bin index \(b\). This is done with a negligible probability of error if \(R_0 < I(X_r, Y_2)\). The destination then constructs the set

\[
\mathcal{L}_2 = \{ m : m \in \mathcal{L}_1, b = B(f(x^n(m), y^n_1)) \}. \tag{2.28}
\]

The decoder finally declares that \(\hat{m} \in \mathcal{L}_2\) as the transmitted message if \(|\mathcal{L}_2| = 1\); otherwise it declares an error. The decoding of the message is successful if

\[
R < \min\{I(X; Y_1) + R_0, I(X; Y_1, Y_r)\}.
\]

Note that the relay does not perform any quantization when employing HF protocol. See also [RY10] for another variant of compression-based relaying referred to as generalized HF which achieves the same rate as that achievable by CF for the general relay channel.

We next give another example which does not belong to the above semi-deterministic class and CF is still optimal.

**Example 1**
Consider a Gaussian relay channel with received signals \(Y_1 = X + Z_1\) and \(Y_2 = X_r + Z_2\) at the destination. The received signal at the relay is given by \(Y_r = X + Z_r\). Then assume that \(Z_r \sim \mathcal{N}(0, N_r)\) and it is independent of \(Z_1 \sim \mathcal{N}(0, N_1)\), and
$Z_2 = \rho Z_r$ where $\rho$ is a non-zero constant. Further assume that the sender and the relay operate under average power constraints; i.e. $\mathbb{E}[X^2] \leq P$ and $\mathbb{E}[X_r^2] \leq P_r$. See Figure 2.10 for an illustration.

The relay channel in Example 1 does not satisfy Definition 2.3.1 because

$$p(y_1, y_2, y_r | x, x_r) = p(y_2 | x_r) p(y_1, y_r | x, x_r, y_2)$$

$$= p(y_2 | x_r) p(y_1, y_r | x, x_r, y_2)$$

$$\neq p(y_2 | x_r) p(y_1, y_r | x).$$

The last inequality in (2.29) holds since $Y_r$ depends on $(Y_2 - X_r)$.

**Proposition 2.3.2.** The capacity of the Gaussian relay channel in Example 1 is given by

$$C = C \left( \frac{P}{N_1} \right) + C \left( \frac{P_r}{\rho^2 N_r} \right).$$

**Proof.** We first note that the received signal at the relay can be constructed from $X, X_r$ and $Y$. That is

$$Y_r = X + Z_r$$

$$= X + \frac{1}{\rho} (Y_2 - X_r)$$

$$=: f(X, X_r, Y_2).$$

Then let $\hat{Y}_r = Y_r$. Thus (2.24) simplifies to

$$R_{CF} = \max_{p(x)p(x_r)} \min \{ I(X, X_r; Y_1, Y_2), I(X; Y_1, Y_2, Y_r | X_r) \}$$

$$= \max_{p(x)p(x_r)} I(X, X_r; Y_1, Y_2)$$

---

6Throughout the thesis, capacity results proved for discrete memoryless channels can be extended to Gaussian channels by a discretization procedure as discussed in [McE77, GK10].
2.4. Instantaneous Relaying

\[ \begin{aligned} &\max_{p(x)} I(X; Y_1) + \max_{p(x_r)} I(X_r; Y_2) \\ &= C \left( \frac{P}{N_1} \right) + C \left( \frac{P_r}{P^2 N_r} \right). \end{aligned} \]  

(2.32)

The converse follows by the cutset bound which completes the proof. \[ \square \]

2.4 Instantaneous Relaying

The relaying schemes discussed in Section 2.3 except the direct-transmission lower bound, operates block-wise at the relay. In this section, we focus on relaying schemes where the relay uses memoryless encoding. That is the relay operates component-wise, in contrast to block-wise, on the received symbols at the relay.

Consider the relay channel with orthogonal receive components (see Definition 2.3.1). Specializing the achievable rate in (2.3) to \( r = 1 \), we obtain

\[ C_1 = \sup_{p(x), x_r = f(y_r)} I(X; Y_1, Y_2). \]  

(2.33)

where \( f(\cdot) \) is a deterministic, memoryless, one-dimensional function. \(^7\) That is, the present output of the relay depends only on the currently received sample at the input. Hence, we term this instantaneous relaying. \(^8\) This strategy is also known as memoryless relaying. Investigating instantaneous relaying is important for at least two reasons. Firstly, constraining the relay mapping to be one-dimensional significantly simplifies the relay functionality. Secondly, there is no processing delay at the relay. These facts make instantaneous relaying an attractive alternative in various applications, for example in wireless sensor networks where efficient mass deployment of relays is expected, and applications in which low-latency transmission can be crucial. In the thesis, we explore this strategy for different configuration of relay networks.

2.4.1 A Binary Symmetric Relay Channel

To motivate instantaneous relaying, we first give an example of the relay channels for which instantaneous relaying is optimal.

Example 2

Consider the following binary symmetric relay channel with orthogonal receive components. Let \( Y_r = Z \oplus Z_r \) denote the received signal at the relay where \( Z_r \sim \text{Ber}(\epsilon_r) \).

\(^7\) The direct-transmission lower bound in (2.8) is a special case of memoryless relaying for the relay channel with orthogonal receive components, in which the relay transmits a fixed letter regardless of the received signal.

\(^8\) Note that, \( C_1 \) can be also obtained by a strictly causal memoryless mapping at the relay, i.e., \( x_r = f(y_r, r-1) \). This is possible because the received signals from the source and the relay at the destination are orthogonal.
28 The Classical Relay Channel

Relay

Figure 2.11: A binary symmetric relay channel.

and \( Z \sim \text{Ber}(\frac{1}{2}) \). Additionally let \((Y_1, Y_2)\) be the received signals at the destination where \( Y_1 = X \oplus Z, \ Y_2 = X_r \oplus Z_2 \) and \( Z_2 \sim \text{Ber}(\epsilon_2) \). We assume that the Bernoulli noises \( Z, Z_r \) and \( Z_2 \) are independent of each other. In this example, all quantities are binary (i.e., \( X, X_r, Z, Z_2, Z_r, Y_1, Y_2, Y_r \in \{0, 1\} \)) and the addition is done in \( \mathbb{GF}(2) \) (i.e., XOR addition). See Figure 2.11 for an illustration of the channel.

This example is constructed by a slight modification of the channel studied in [ARY09]. Note that without the relay, no positive reliable rate can be achieved over the channel.

Proposition 2.4.1. The capacity of the relay channel in Example 2 is given by

\[
C = 1 - H_b(\epsilon_r * \epsilon_2),
\]

where \( H_b(\epsilon) := -\epsilon \log_2(\epsilon) - (1 - \epsilon) \log_2(1 - \epsilon) \) denotes the binary entropy function and \( \epsilon_1 * \epsilon_2 := \epsilon_1 + \epsilon_2 - 2\epsilon_1\epsilon_2 \).

Proof. We first prove the achievability. Let the memoryless relaying strategy be \( x_r = f(y_r) = y_r \). That is the relay transmits the received noisy bit to the destination. The achievable rate is then given by

\[
R = I(X; Y_1, Y_2) = H(Y_2, Y_1) - H(Y_1, Y_2|X) = H(Y_2, Y_1) - H(Z, Z_r \oplus Z_2) = H(Y_1) + H(Y_2|Y_1) - H(Z) - H(Z_r \oplus Z_2|Z) = H(Y_1) + H(Y_2|Y_1) - H(Z) - H(Z_r \oplus Z_2) = H(Y_1) + H(Y_2 \oplus Y_1|Y_1) - H(Z) - H(Z_r \oplus Z_2) = H(X \oplus Z) + H(Z \oplus Z_r \oplus Z_2|X \oplus Z) - H(Z) - H(Z_r \oplus Z_2) \leq 1 - H_b(\epsilon_r * \epsilon_2),
\]

where the upper bound is achieved by choosing \( X \sim \text{Ber}(\frac{1}{2}) \).
Now using a weak converse, we next show that no higher rate can be achieved. Consider

\[ nR = H(M) \]
\[ = I(M; Y_1^n, Y_2^n) + H(M|Y_1^n, Y_2^n) \]
\[ \leq I(M; Y_1^n, Y_2^n) + n\epsilon_n \]  
\[ \leq I(X^n; Y_1^n, Y_2^n) + n\epsilon_n \]
\[ = I(X^n; Y_1^n|Y_2^n) + I(X^n; Y_2^n) + n\epsilon_n \]
\[ \leq I(X^n; Y_1^n|Y_2^n) + n\epsilon_n \]
\[ = H(Y_1^n|Y_2^n) - H(Y_1^n|X^n, Y_2^n) + n\epsilon_n \]
\[ = H(Y_1^n|Y_2^n) - H(Z^n|Y_2^n) + n\epsilon_n \]
\[ \leq H(Y_1^n) - H(Z^n|Y_2^n) + n\epsilon_n \]
\[ \leq n - H(Z^n|Y_2^n) + n\epsilon_n \]
\[ \leq n - n\mathbb{H}_{b}(\epsilon_r + \mathbb{H}^{-1}_b(\frac{1}{n}H(Y^n_r|Y_2^n))) + n\epsilon_n \]
\[ \leq n (1 - \mathbb{H}_{b}(\epsilon_r + \epsilon_2) + \epsilon_n), \]

where

(a) follows by Fano’s inequality

\[ H(M|Y_1^n, Y_2^n) \leq H(M|\hat{M}(Y_1^n, Y_2^n)) \leq 1 + \mathbb{P}_e^{(n)} nR = n\epsilon_n, \]

where \( \epsilon_n \to 0 \) as \( n \to \infty \) for a negligible probability of error.

(b) holds since \( M \to X^n \to (Y_1^n, Y_2^n) \) form a Markov chain;

(c) follows because \( Y_2 \) is independent of \( X \);

(d) holds because conditioning reduces the entropy [CT06];

(e) follows by noting that \( Z_r = Y_r \oplus Z \) is independent of \( Y_r \) and applying Mrs. Gerber’s lemma [WZ73, SW90] given in Appendix 2.B; and

(f) holds because

\[ H(Y_1^n|Y_2^n) \]
\[ = H(Y_1^n) - I(Y_1^n; Y_2^n) \]
\[ = n - I(Y_1^n; Y_2^n) \]
\[ \geq n - I(X^n_p; Y_2^n) \]
\[ \geq n - n(1 - \mathbb{H}_{b}(\epsilon_2)) \]
\[ = n\mathbb{H}_{b}(\epsilon_2). \]
Thus the capacity of the above binary symmetric relay channel is given by
\[ C = 1 - H_b(\epsilon_r * \epsilon_2). \] (2.38)
This completes the proof.

A direct evaluation of the cutset bound in (2.6) gives that
\[ R_{ub} = 1 - \max \{ H_b(\epsilon_r), H_b(\epsilon_2) \} \geq 1 - H_b(\epsilon_r * \epsilon_2) = C, \] (2.39)
where the last inequality is strict if \( \epsilon_1, \epsilon_r \notin \{0, 0.5, 1\} \). Thus, from the above capacity result, we conclude that the cutset bound is not tight in general.

It is worth mentioning that the capacity of the relay channel in Example 2 can also be achieved using the CF strategy. One can prove this by direct computation of the bound in (2.23) and by choosing \( \hat{Y}_r = Y_r \oplus Z_q \) where \( Z_q \sim \text{Ber}(\epsilon_q) \) is independent of other random variables.

In summary, from the above example, we emphasize the following three points:

1. By employing the relay, one boosts the reliable transmission rate from zero to \( 1 - H_b(\epsilon_r * \epsilon_2) \).

2. Instantaneous relaying can be an optimal strategy. In the thesis, we explore this strategy in some detail and prove interesting results regarding its optimality and the associated optimal relaying function \( f \) for Gaussian relay channels.

3. The cutset bound is not tight in general.

### 2.4.2 On the Optimality of Instantaneous Relaying

We next generalize Example 2 to a class of relay channels and give some sufficient conditions for the optimality of the memoryless relaying.

**Proposition 2.4.2.** If the relay channel has orthogonal receive components according to Definition 2.3.1, then \( C_1 \leq R_{CF} \) where \( C_1 \) is given in (2.33).

**Proof.** CF achieves the rate \( R_{CF} = \max I(X; Y_1, \hat{Y}_r) \) subject to the constraint \( I(\hat{Y}_r; Y_r|Y_1) \leq I(X_r; Y_2) \) where the maximum is taken over \( p(x)p(x_r)p(\hat{y}_r|y_r) \). Now consider the pmf
\[ P_0 = p(x)p_f(\hat{y}_r)(x_r)p_{Y_2|X_r}(\hat{y}_r|f(y_r)). \] (2.40)
We next show that \( P_0 \) is a valid pmf subject to the constraint regardless of \( p(x) \). Consider
\[
I(\hat{Y}_r; Y_r|Y_1) = H(\hat{Y}_r|Y_1) - H(\hat{Y}_r|Y_r, Y_1)
= H(\hat{Y}_r|Y_1) - H(\hat{Y}_r|Y_r)
\leq H(\hat{Y}_r) - H(\hat{Y}_r|Y_r)
= H(Y_2) - H(Y_2|X_r = f(y_r))
= I(X_r; Y_2).
\]
Thus the rate $I^*(X; Y_1, \hat{Y}_r)$ evaluated using $P_0$ is achievable. This yields $R_{CF} \geq \max_{p(x)} I^*(X; Y_1, \hat{Y}_r) = \max_{p(x)} I(X; Y_1, Y_2|X_r = f(Y_r)) = C_1$. This completes the proof.

Thus for relay channels with orthogonal receive components, memoryless relaying cannot outperform CF. In other words, if memoryless relaying is optimal, so is CF. We next give sufficient conditions for optimality of memoryless relaying.

**Proposition 2.4.3.** The memoryless mapping $X_r = f(Y_r)$, where $f$ is a one-dimensional function, is an optimal relaying strategy if

i) CF achieves the capacity with the optimal pmf $P^* = p(x)p(x_r)p(\hat{y}_r|y_r)$;

ii) $H(\hat{Y}_r|Y_1) = H(\hat{Y}_r)$;

iii) $p(\hat{y}_r|y_r) = p_{Y_2|X_r}(\hat{y}_r|f(y_r))$; and

iv) $p(x_r) = p(f(Y_r)|x_r)$.

**Proof.** CF achieves the rate $R_{CF} = I(X; Y_1, \hat{Y}_r)$ for an optimal pmf $P^* = p(x)p(x_r)p(\hat{y}_r|y_r)$ subject to the constraint $I(\hat{Y}_r; Y_r|Y_1) \leq I(X_r; Y_2)$. We next show that the pmf $P = p(x)p(f(Y_r)|x_r)p_{Y_2|X_r}(\hat{y}_r|f(y_r))$ under conditions i)–iii) is also optimal. We first show that $P$ satisfy the constraint. Let at the optimal input distribution

$I_0 := I(\hat{Y}_r; Y_r|Y_1) \leq I(X_r; Y_2) =: I_1$.

By conditions ii) and iii) we have

$I_0 = I(\hat{Y}_r; Y_r|Y_1) = I(\hat{Y}_r; Y_r) = I(Y_2; Y_r)$,

and by condition iv) we have

$I_1 = I(X_r; Y_2) = I(Y_r; Y_2)$.

Hence $I_0 \leq I_1$ holds at $P$. This yields

$I(X; Y_1, \hat{Y}_r|P^*) = I(X; Y_1, \hat{Y}_r|P) = I(X; Y_1, Y_2|X_r = f(Y_r))$.

(2.41)

This completes the proof.

It is easy to check that the binary symmetric relay channel in Example 2 satisfies all conditions in Proposition 2.4.3.
2.4.3 Innovative Relaying

Note that Proposition 2.4.3 can be easily extended to the case with average cost constraints at the source and the relay. We next give two examples of Gaussian relay channels for which the simple linear mapping $x_r = \kappa y_r$ where $\kappa$ is a constant, is optimal. This scheme is known as amplify-and-forward (AF) [LWT08, GMZ06, NBK04, AT07] since the relay scales the received signal to meet the power constraint at the relay.

**Example 3**

Let $Y_1 = X + Z_1$ be the received signal at the destination where $Z_1 \sim \mathcal{N}(0, N_1)$ and $X$ denotes the transmitted signal from the source. Now let the received signal at the relay be $Y_r = g(X, Y_1) = X - \mathbb{E}[X|Y_1]$. That is the relay receives the error in estimating $X$ from $Y_1$. Finally let $Y_2 = X_r + Z_2$ be the received signal at the destination from the relay where $Z_2 \sim \mathcal{N}(0, N_2)$ is independent of $Z_1$, and $X_r$ denotes the transmitted signal from the relay. See Figure 2.12 for an illustration. We further assume that the source and the relay operate under average power constraints; i.e. $\mathbb{E}[X] \leq P$ and $\mathbb{E}[X_r] \leq P_r$.

**Proposition 2.4.4.** The capacity of the relay channel in Example 3 is given by

$$C = C \left( \frac{P}{N_1} \right) + C \left( \frac{P_r}{N_2} \right). \quad (2.42)$$

**Proof.** Using memoryless relaying, we can achieve the rate $R = I(X; Y_1, Y_2)$. Now let $X \sim \mathcal{N}(0, P)$ and $X_r = \kappa Y_r$ where $\kappa$ is chose to ensure the power constraint at the relay.

The received signal at the relay is given by

$$Y_r = X - \mathbb{E}[X|Y_1]$$

$$= X - \frac{P}{P + N_1} Y_1. \quad (2.43)$$

Figure 2.12: A semi-deterministic Gaussian relay channel.
2.4. Instantaneous Relaying

Hence the achievable rate is given by

\[ R = I(X; Y_1, Y_2) \]
\[ = h(Y_1, Y_2) - h(Y_1, Y_2|X) \]
\[ = h(Y_1) + h(Y_2|Y_1) - h(Y_1|X) - h(Y_2|X, Y_1) \]
\[ = h(Y_1) + h(Y_2|Y_1) - h(Y_1|X) - h(kg(X, Y_1) + Z_2|X, Y_1) \]
\[ = h(Y_1) + h(Y_2|Y_1) - h(Z_1) - h(Z_2) \]
\[ \overset{(a)}{=} h(Y_1) + h(Y_2) - h(Z_1) - h(Z_2) \]
\[ = C\left(\frac{P}{N_1}\right) + C\left(\frac{P_r}{N_2}\right), \quad (2.44) \]

where (a) follows by the fact that \( Y_r \) is independent of \( Y_1 \). The converse follows from the cutset bound, which completes the proof. One can also prove the optimality of linear relaying by checking that the conditions in Proposition 2.4.3 are met. \( \square \)

We next give another semi-deterministic example of relay channels for which AF is optimal.

**Example 4**

Let \( Y_1 = X + S \) be the received signal at the destination where \( S \sim \mathcal{N}(0, Q) \) denotes the interference on the direct link and \( X \) denotes the transmitted signal from the source. Now let the received signal at the relay be \( Y_r = g(S, Y_1) = S - E[S|Y_1] \). That is the relay receives the error in estimating the interference from \( Y_1 \). Finally let \( Y_2 = X_r + Z_2 \) be the received signal at the destination from the relay where \( Z_2 \sim \mathcal{N}(0, N_2) \) is independent of \( S \), and \( X_r \) denotes the transmitted signal from the relay. See Figure 2.13 for an illustration. Similar to Example 3, we assume that the source and the relay operate under average power constraint, i.e. \( E[X] \leq P \) and \( E[X_r] \leq P_r \).

**Proposition 2.4.5.** The capacity of the relay channel in Example 4 is given by

\[ C = C\left(\frac{P}{Q}\right) + C\left(\frac{P_r}{N_2}\right). \quad (2.45) \]
Proof. The proof is similar to that of Proposition 2.4.4 by choosing $X \sim \mathcal{N}(0, P)$ and $X_r = \kappa Y_r$ where $\kappa$ is chosen to ensure the power constraint at the relay.

From the proof of Proposition 2.4.5, we see that the optimal input distribution is Gaussian in spite of the fact that a Gaussian input creates the worst noise in estimating the interference. Two reasons for optimality of a Gaussian input for this example is that the estimation error matches the channel from the relay to the destination and additionally what the relay transmits is “new” compared to the signal that the destination has received from the source.
Appendix

2.A Typicality and Joint Typicality

In this section, we define $\epsilon$-typical and jointly $\epsilon$-typical sequences. The definitions and the results in this section can be found in [OR01, GK10].

Definition 2.A.1. ($\epsilon$-Typical Sequences) Let $X \sim p_X(x)$ be a random variable with finite alphabet $\mathcal{X}$. Let $x^n$ be a sequence with elements drawn from $\mathcal{X}$. Define the empirical pmf of $x^n$ as

$$\pi(x|x^n) := \frac{1}{n} |\{i : x_i = x\}|, \ \forall x \in \mathcal{X}. \quad (2.46)$$

Let $\epsilon > 0$. Then the set $T^n_\epsilon(X)$ of $\epsilon$-typical sequences of length $n$ is defined as

$$T^n_\epsilon(X) := \left\{ x^n : |\pi(x|x^n) - p(x)| \leq \epsilon \cdot p(x), \forall x \in \mathcal{X} \right\}. \quad (2.47)$$

Definition 2.A.2. (Jointly $\epsilon$-Typical Sequences) Let $(X, Y) \sim p_{X,Y}(x, y)$ be a pair of random variables with finite alphabets $\mathcal{X} \times \mathcal{Y}$. Let $(x^n, y^n)$ be a pair of sequences with elements drawn from $\mathcal{X} \times \mathcal{Y}$. Define the joint empirical pmf as

$$\pi(x, y|x^n, y^n) := \frac{1}{n} |\{i : (x_i, y_i) = (x, y)\}|, \ \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \quad (2.48)$$

Let $\epsilon > 0$. Then the set $T^n_\epsilon(X, Y)$ of jointly $\epsilon$-typical sequences of length $n$ is defined as

$$T^n_\epsilon(X, Y) := \left\{ (x^n, y^n) : |\pi(x, y|x^n, y^n) - p(x, y)| \leq \epsilon \cdot p(x, y), \forall (x, y) \in \mathcal{X} \times \mathcal{Y} \right\}. \quad (2.49)$$

We next mention some consequences of the above definitions.

Lemma 2.A.3. (Conditional Typicality Lemma) Let $X$ and $Y$ be finite random variables with the joint pmf $p(x, y)$ and $\epsilon > \epsilon' > 0$. If $x^n \in T^n_\epsilon(X)$ and $y^n$ drawn according to $\prod_{i=1}^n p_Y(y_i|x_i)$, then

$$\Pr\left\{ (x^n, y^n) \in T^n_\epsilon(X, Y) \right\} \rightarrow 1 \text{ as } n \rightarrow \infty. \quad (2.50)$$

Lemma 2.A.4. (Joint Typicality Lemma) Let $X$ and $Y$ be finite random variables with the joint pmf $p(x, y)$ and $\epsilon > 0$. If $\tilde{Y}^n$ is distributed according to an arbitrary pmf $p(\tilde{y}^n)$ over alphabet $\tilde{\mathcal{Y}}^n$, and is independent of $X^n \sim \prod_{i=1}^n p_X(x_i)$, then for $n$ sufficiently large, there exists a function $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ such that

$$\Pr\left\{ (X^n, \tilde{Y}^n) \in T^n_\epsilon(X, Y) \right\} \leq 2^{-n(I(X;Y) - \delta(\epsilon))}. \quad (2.51)$$

---

9 For brevity, we denote $T^n_\epsilon(X)$ as $T_\epsilon^n$ when it is understood from the context.

10 For brevity, we denote $T^n_\epsilon(X, Y)$ as $T_\epsilon^n$ when it is understood from the context.
Lemma 2.A.5. (Packing Lemma) Let $X$ and $Y$ be finite random variables with the joint pmf $p(x, y)$ and $\epsilon > 0$. Let $\tilde{Y}^n$ be distributed according to an arbitrary pmf $\tilde{p}(\tilde{y}^n)$ over alphabet $\tilde{Y}^n$. Let $X^n(m)$, $m \in \mathcal{M}^{(n)}$ where $|\mathcal{M}^{(n)}| \leq 2^nR$, be random sequences, each distributed according $\prod_{i=1}^n p_X(x_i)$. Assume that $X^n(m)$ is independent of $\tilde{Y}^n$. Then there exists a function $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that

$$\Pr\left\{ \left( X^n(m), \tilde{Y}^n \right) \notin T^{(n)}_{\epsilon}(X,Y), \text{ for some } m \in \mathcal{M}^{(n)} \right\} \to 0,$$

as $n \to \infty$, if $R < I(X; Y) - \delta(\epsilon)$.

Lemma 2.A.6. (Covering Lemma) Let $X$ and $\hat{X}$ be finite random variables with the joint pmf $p(x, \hat{x})$ and $\epsilon > 0$. Let $X^n \sim \prod_{i=1}^n p_X(x_i)$. Let $\hat{X}^n(m)$, $m \in \mathcal{M}^{(n)}$ where $|\mathcal{M}^{(n)}| \geq 2^nR$, be random sequences, each distributed according $\prod_{i=1}^n p_{\hat{X}}(\hat{x}_i)$ and independent of $X^n(m)$. Then there exists a function $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that

$$\Pr\left\{ \left( X^n, \hat{X}^n(m) \right) / \notin T^{(n)}_{\epsilon}(X,\hat{X}), \forall m \in \mathcal{M}^{(n)} \right\} \to 0,$$

as $n \to \infty$, if $R > I(X; \hat{X}) + \delta(\epsilon)$.

2.B Mrs. Gerber’s Lemma

We next present Mrs. Gerber’s Lemma [WZ73, SW90].

- Scalar version: Consider binary random variables $X, Z$ and $Y = X \oplus Z$, and an arbitrary random variable $U$. If $Z \sim \text{Ber}(p)$ is independent of $(X, U)$ then

$$H^{-1}_b(H(Y|U)) \geq H^{-1}_b(H(X|U) * p),$$

where $H^{-1}_b : [0, 1] \to [0, 0.5]$ denotes the inverse of the binary entropy function $H_b$.

- Vector version: Consider binary random vectors $X^n, Z^n$ and $Y^n = X^n \oplus Z^n$, and an arbitrary random variable $U$. If $Z_i \sim \text{Ber}(p)$ is i.i.d. and independent of $(X^n, U)$ then

$$H^{-1}_b \left( \frac{1}{n} H(Y^n|U) \right) \geq H^{-1}_b \left( \frac{1}{n} H(X^n|U) * p \right).$$

(2.55)
Chapter 3

Binary Symmetric Relay Channels

3.1 Introduction

In Chapter 2, we studied a binary symmetric relay channel (BSRC) in Example 2 and proved that the capacity of this relay channel can be achieved using memoryless relaying. In this chapter, we consider other examples of BSRCs and investigate relaying via decode-and-forward (DF), partial decode-and-forward (PDF), compress-and-forward (CF) and general finite-memory relay mappings. One of our main contributions is to propose a general structure for finite-dimensional relaying. We illustrate that one can get close to the capacity upper bound for some cases by using the proposed low-dimensional mappings.

The remaining part of the chapter is organized as follows. In Section 3.2, we present a three-node BSRC with orthogonal receive components at the destination. Section 3.3 derives capacity bounds for the BSRC when the relay is assumed to have infinite memory. Section 3.4 investigates capacity bounds for the BSRC when the relay is assumed to have a finite memory. We present an iterative algorithm to optimize the relay mapping in Section 3.4. In Section 3.5, we compare the achievable rates as a function of the memory length at the relay. Finally, Section 3.6 concludes the chapter.

3.2 Channel Models

We next introduce the channel models that we explore in this chapter. Figure 3.1 shows a BSRC consisting of three nodes: a source, a relay and a destination. In this model, we however assume that received signals at the destination are orthogonal. That is, the signals transmitted from the source and the relay do not interfere with each other. Note that this model falls into the class of relay channels defined in Definition 2.3.1. We further assume that all links are corrupted with modulo-sum noises distributed according to the Bernoulli distribution and all quantities are binary; i.e., \( X, X_r, Z_1, Z_r, Z_2, Y_1, Y_2, Y_r \in \{0, 1\} \).

The received signal at the relay \( Y_r \) is given by

\[
Y_r = X \oplus Z_r,
\]

(3.1)
where $X$ is the transmitted symbol from the source and $Z_r \sim \text{Ber}(\epsilon_r)$ is the additive Bernoulli noise. The received signal from the source at the destination $Y_1$ is given by

$$Y_1 = X \oplus Z_1,$$

where $Z_1 \sim \text{Ber}(\epsilon_1)$ is the additive Bernoulli noise. Similarly, the received signal from the relay at the destination $Y_2$ is given by

$$Y_2 = X_r \oplus Z_2,$$

where $X_r$ is the transmitted symbol from the relay and $Z_2 \sim \text{Ber}(\epsilon_2)$ is the additive Bernoulli noise. We assume that the random variables $Z_r$, $Z_1$, and $Z_2$ are mutually independent. Note that the addition in (3.1)–(3.3) is done in $\mathbb{GF}(2)$.

**Remark 3.2.1.** Figure 3.2 shows a BSRC with non-orthogonal reception at the destination. The received signal at the destination is given by

$$Y = X \oplus X_r \oplus Z,$$

where $X$ and $X_r$ respectively denote the symbols transmitted by the source and the relay. The random variable $Z \sim \text{Ber}(\epsilon)$ is the additive Bernoulli noise and is independent of $Z_r$. The capacity of this channel is

$$C_{no} = 1 - H_b(\epsilon).$$

By setting $X_r = 0$, we have $Y = X \oplus Z$. Then the achievability follows since $\max_{p(x)} I(X;Y|X_r = 0) = C_{no}$ (see also direct-transmission lower bound discussed in Section 2.3.2). The converse follows from the cutset bound in (2.6). Invoking the multiple-access bound, we have

$$C \leq I(X, X_r; Y)$$

$$= H(Y) - H(Y|X, X_r)$$

$$= H(Y) - H(Z)$$

$$\leq 1 - H_b(\epsilon).$$

Figure 3.1: Orthogonal Binary Symmetric Relay Channel.
3.3 Capacity Bounds: Infinite Memory Relay Case

In this section, we consider the cutset bound and three lower bounds on the capacity: DF, PDF and CF. These bounds are described in Section 2.3 for the general relay channel where the relay has infinite memory and unlimited processing capability.

**Proposition 3.3.1** (cutset bound). For the relay channel in Figure 3.1, the capacity is upper bounded by

$$R_{UB} = \min \left\{ 1 + \mathbb{H}_b(\epsilon_1 \ast \epsilon_r) - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_r), 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2) \right\}.$$  \hspace{1cm} (3.7)

**Proof.** See Appendix 3.A.

**Proposition 3.3.2** (Relaying via DF). For the relay channel in Figure 3.1, the capacity is lower bounded by

$$R_{DF} = \max \left\{ 1 - \mathbb{H}_b(\epsilon_1), \min \{1 - \mathbb{H}_b(\epsilon_r), 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2)\} \right\}.$$  \hspace{1cm} (3.8)

**Proof.** See Appendix 3.B.

The achievable rate of DF is limited by the quality of the source–relay link. Note that in (3.8), we take the maximum of the rates achieved using DF and the only-direct-link transmission. In some cases it is possible to improve on DF by using partial DF (PDF) at the relay. That is the relay only decodes a part of the transmitted message. Recall that the achievable rate of PDF is given by

$$R_{PDF} = \max_{p(u,x,x_r)} \min \{ I(X, X_r; Y_1, Y_2), I(U; Y_r | X_r) + I(X; Y_1, Y_2 | X_r, U) \},$$  \hspace{1cm} (3.9)

where $U$ denotes the part of the transmitted message that the relay decodes (see Section 2.3.5).
Proposition 3.3.3 (Relaying via PDF). For the relay channel in Figure 3.1, PDF attains the same rate as DF given in (3.8).

Proof. See Appendix 3.C. □

Corollary 3.3.4. DF is an optimal relaying strategy if
\[ H_b(\epsilon_1) + H_b(\epsilon_2) - H_b(\epsilon_r) \geq 1. \]  
(3.10)

Proof. The proof follows from Propositions 3.3.2 and 3.3.1. □

Proposition 3.3.5 (Relaying via CF). For the relay channel in Figure 3.1, the capacity is lower bounded by
\[
R_{CF} = \begin{cases} 
1 + H_b(\epsilon_1 * \epsilon_r * \epsilon_q) - H_b(\epsilon_r * \epsilon_q) - H_b(\epsilon_1), & \text{if } 1 < H_b(\epsilon_2) + H_b(\epsilon_1 * \epsilon_r) \\
1 + H_b(\epsilon_1 * \epsilon_r) - H_b(\epsilon_r) - H_b(\epsilon_1), & \text{if } 1 \geq H_b(\epsilon_2) + H_b(\epsilon_1 * \epsilon_r)
\end{cases}
\](3.11)

where \( \epsilon_q \) satisfies
\[ H_b(\epsilon_1 * \epsilon_r * \epsilon_q) - H_b(\epsilon_q) + H_b(\epsilon_2) = 1. \]  
(3.12)

Proof. See Appendix 3.D. □

Corollary 3.3.6. CF is an optimal relaying strategy if \( H_b(\epsilon_1 * \epsilon_r) + H_b(\epsilon_2) \leq 1 \).

Proof. The proof follows from Propositions 3.3.5 and 3.3.1. □

Figures 3.3 and 3.4 show optimal capacity regions as a function of \( \epsilon_r \) and \( \epsilon_1 \) in the plane \([0, 1] \times [0, 1]\) when \( \epsilon_2 \) is set to 0.05 and 0.2, respectively. One can see that for some parameters of \( \epsilon_r \) and \( \epsilon_1 \), CF or DF are optimal. Without loss of generality, we next focus on the section \([0, 0.5] \times [0, 0.5]\) of the plane.

From Figures 3.3 and 3.4, we see that DF is optimal when \( \epsilon_r \) is small and \( \epsilon_1 \) is close to 0.5. When \( \epsilon_1 = 0.5 \) the BSRC simplifies to a two-hop channel when DF is capacity achieving (see also Section 2.3.3). However, DF is also optimal when the relay is relatively strong compared to the direct link.

From Figures 3.3 and 3.4, we similarly see that CF is optimal at the corner of the plane, i.e. when \( \epsilon_1 \) and \( \epsilon_r \) are small. Small value of \( \epsilon_2 \) allows the relay to use higher rate to describe the received signal to destination and small value of \( \epsilon_1 \) means the destination has better side-information in order to decode the reproduction sequence generated at the relay.
Figure 3.3: Optimal regions for $\epsilon_2 = 0.05$. 
Figure 3.4: Optimal regions for $\epsilon_2 = 0.2$. 
We next consider the case when the relay has a finite memory length. If the relay has a storage memory of \( k \) bits, it can process the last \( k - 1 \) and the presently received symbol to generate \( k \) new symbols using \( k \) possibly different \( k \)-dimensional functions. This results in a low-complexity relaying protocols, which is suitable for delay-sensitive or inexpensive applications. In the following, we denote the relay functions by

\[
 f_i : \mathbb{GF}^k(2) \longrightarrow \mathbb{GF}(2)
\]

\[
x_{r,i} = f_i(y_{r,1}, \ldots, y_{r,k})
\]

for \( i \in \{1, 2, \ldots, k\} \). Note that we here consider the whole sub-block of \( k \) symbols to generate \( k \) new symbols to be transmitted to the destination. This is different from the classical definition with strictly causal relaying discussed in Section 2.1. We are allowed to do this, without any particular condition in signal reception at the relay, since the relay has an orthogonal link to the destination.

### 3.4.1 Achievable Rate

For a given set of relay functions \( \{f_i\}_{i=1}^{k} \), the channel is parameterized by the pmf \( p(y|x) \), by defining \( y := (y_{1,1}^k, y_{2,1}^k) \) and \( x := x_1^k \), where

\[
p(y|x) := p\left(y_{1,1}^k, y_{2,1}^k \mid x_1^k\right)
\]

\[
= \sum_{y_{1,1}^k, y_{2,1}^k \in \mathbb{GF}^k(2)} p\left(y_{1,1}^k, y_{2,1}^k, y_{r,1}^k, x_1^k\right) p\left(y_{r,1}^k \mid x_1^k\right)
\]

\[
= \sum_{y_{1,1}^k, y_{2,1}^k \in \mathbb{GF}^k(2)} p\left(y_1^k \mid x_1^k\right) p\left(y_2^k \mid y_1^k\right) p\left(y_{r,1}^k \mid x_1^k\right)
\]

\[
= \sum_{y_{1,1}^k, y_{2,1}^k \in \mathbb{GF}^k(2)} \prod_{i=1}^{k} p\left(y_{1,i} \mid x_1^k\right) p\left(y_{r,i} \mid x_1^k\right) p\left(y_{2,i} \mid f_i\left(y_{r,1}^k\right)\right)
\]

\[
= \sum_{y_{1,1}^k, y_{2,1}^k \in \mathbb{GF}^k(2)} \prod_{i=1}^{k} \left[ (1 - \epsilon_1) \delta_b\left(y_{1,i} \oplus x_i\right) + \epsilon_1 \delta_b\left(y_{1,i} \oplus x_i \oplus 1\right) \right] \times
\]

\[
\left[ (1 - \epsilon_r) \delta_b\left(y_{r,i} \oplus x_i\right) + \epsilon_r \delta_b\left(y_{r,i} \oplus x_i \oplus 1\right) \right] \times
\]

\[
\left[ (1 - \epsilon_2) \delta_b\left(y_{2,i} \oplus f_i\left(y_{r,1}^k\right)\right) + \epsilon_2 \delta_b\left(y_{2,i} \oplus f_i\left(y_{r,1}^k\right) \oplus 1\right) \right].
\]
44 Binary Symmetric Relay Channels

codeword is distributed according to $p(x_k^k)$ (i.e., $p(x_k^k) = \prod_{i=1}^n p(x_i^k)$). Thus, the achievable rate using the finite memory relay is given by

$$C_k = \sup_{\{f_i\}_{i=1}^k : p(x_k^k)} \frac{1}{k} I(X_1^k; Y_1^k, Y_2^k),$$  

(3.15)

where the supremum is taken over the set of boolean functions $\{f_i\}_{i=1}^k$ and the joint pmf $p(x_k^k)$ of $k$ symbols at the source. Since the channel is used $k$ times, the mutual information in (3.15) is divided by $k$ (see also Section 2.2).

Achievable Rate for $k = 1$: The simplest case is the memoryless relay in which the relay just transmits the received noisy bit to the destination without any further processing. That is $x_{r,i} = y_{r,i}$ for $1 \leq i \leq n$. For this relay function, the optimal input distribution is $X \sim \text{Ber}(\frac{1}{2})$.

Proposition 3.4.1. For the relay channel in Figure 3.1, the rate

$$C_1 = 1 + \mathbb{H}_b(\epsilon_1 \ast \epsilon_r \ast \epsilon_2) - \mathbb{H}_b(\epsilon_r \ast \epsilon_2) - \mathbb{H}_b(\epsilon_1),$$  

(3.16)

is achievable.

Proof. See Appendix 3.E.

Computation of $C_k$ for $k \geq 2$ is a cumbersome task. Nevertheless, there is no unique set of relay functions which is optimal for all channel parameters $\epsilon_r, \epsilon_1, \epsilon_2$. To see this, consider the case with $\epsilon_2 = 0$. For this case the simple strategy with $k = 1$ used in Proposition 3.4.1 is optimal. However, this relay function is not necessarily optimal for cases with $\epsilon_2 \neq 0$ since one can potentially provide error protection on the relay–destination link by utilizing functions with higher dimensions.

3.4.2 Mapping Optimization for an Arbitrary $k$

In the following, we confine the pmf at the source to be $p(x_k^k) = \prod_{i=1}^k p(x_i)$, and $p(x) = \frac{1}{2} \delta_b(x) + \frac{1}{2} \delta_b(x \oplus 1)$, i.e. $X \sim \text{Ber}(\frac{1}{2})$.

Lemma 3.4.2. The mutual information given in (3.15) can be written as

$$\frac{1}{k} I(X_1^k; Y_1^k, Y_2^k) = 1 - \mathbb{H}_b(\epsilon_1) - \frac{1}{k} \mathbb{E} \left[ \log(p(y_{2,1}^k \mid y_{1,1}^k)) \right] + \frac{1}{k} \mathbb{E} \left[ \log(p(y_{2,1}^k \mid x_1^k)) \right],$$  

(3.17)

where

$$p(y_{2,1}^k \mid x_1^k) = (1 - \epsilon_2)^k (1 - \epsilon_r)^k \times \sum_{y_{r,1}^k \in \mathbb{GF}_2^k} \left( \frac{\epsilon_2}{1 - \epsilon_2} \right)^{\mathbb{D}_n(y_{2,1}^k, f(y_{r,1}^k))} \left( \frac{\epsilon_r}{1 - \epsilon_r} \right)^{\mathbb{D}_n(y_{r,1}^k, x_1^k)},$$  

(3.18)
and

\[ p(y_{2,1}^k | y_{1,1}^k) = (1 - \epsilon_1)^k (1 - \epsilon_2)^k (1 - \epsilon_r)^k \times \]

\[ \sum_{y_{1,1}^k \in \text{GF}(2)^k} \left( \frac{\epsilon_2}{1 - \epsilon_2} \right)^{D_H(y_{2,1}^k, f(y_{1,1}^k))} \sum_{x_{1,1}^k \in \text{GF}(2)^k} \left( \frac{\epsilon_r}{1 - \epsilon_r} \right)^{D_H(y_{1,1}^k, x_{1,1}^k)} \left( \frac{\epsilon_1}{1 - \epsilon_1} \right)^{D_H(y_{1,1}^k, x_{1,1}^k)}. \]

(3.19)

**Proof.** See Appendix 3.F.

To compute the rate in (3.15), one needs to select the best functions among \(2^{2k^2}\) possible choices, which has an exponential complexity. (For \(k = 4\), there are approximately \(1.8 \times 10^{19}\) possible functions.) In order to cope with the complexity, we implement an efficient hill-climbing search algorithm as follows. For a given \(k\), we first initialize the relay functions with a random mapping and compute the rate using Lemma 3.4.2. Then we randomly select one function and one corresponding dimension, and flip the mapping and recompute the rate. If the new mapping provides a higher rate, we accept the change. Otherwise we repeat the process until the mapping converges. Since the algorithm by construction may terminate in a local optimum, we repeat the whole algorithm with different initializations and pick the mapping that attains the highest rate.

One example of the optimized mapping for \(k = 4\) when \(\epsilon_1 = \epsilon_2 = \epsilon_r = 0.01\) is

\[
F = \begin{bmatrix}
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1
\end{bmatrix}.
\]

Here \(F_{ij}\) denotes the output of the relay along the \(i\)th dimension for the \(j\)th input configuration. (We have \(1 \leq i \leq 4\) and \(1 \leq j \leq 2^4\).) By studying the matrix \(F\), we can get some insight into the underlying structure of the mapping. Finding an efficient structure at the relay simplifies the design and the implementation of relaying. We employ the Fourier transform to accomplish this task. Our use of the binary Fourier (or Hadamard) transform is related to how it was used in e.g. [Sko99, MZ00] to analyze the performance of quantizers over noisy channels.

### 3.4.3 Fourier Spectrum of the Optimized Mappings

In order to use the Fourier transform, we need an orthonormal basis [Rud90]. Consider the following set of functions

\[
X_S(x) : \{-1, +1\}^k \mapsto \{-1, +1\}
\]

\[
X_S(x) = \prod_{i \in S} x_i,
\]

(3.20)
where \( S \subseteq \{1, 2, \ldots, k\} \). Then, any function \( f : \{0, 1\}^k \to \{0, 1\} \) can uniquely be represented as

\[
f = \sum_{S \subseteq \{1, 2, \ldots, k\}} \hat{f}(S)X_S,
\]

(3.21)

where \( \hat{f}(S) \) is the Fourier coefficient of \( f \) and is given by

\[
\hat{f}(S) = \langle f, X_S \rangle = \frac{1}{2^k} \sum_{x \in \{+1, -1\}^k} f(x)X_S(x)
\]

\[
= \mathbb{E}[f \cdot X_S].
\]

(3.22)

The expectation in (3.22) is taken uniformly over \( x \in \{+1, -1\}^k \). Note that the Fourier expansion of \( f \) can potentially have up to \( 2^k \) terms. As an example consider the following randomly chosen function.

\[
x_3 \quad x_2 \quad x_1 \quad f(x_1, x_2, x_3)
\]

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
</tr>
</tbody>
</table>

This function can be expanded as

\[
f(x_1, x_2, x_3) = \frac{1}{4} - \frac{1}{4}x_1 + \frac{1}{4}x_2 - \frac{1}{4}x_3 - \frac{1}{4}x_1x_2 + \frac{1}{4}x_1x_3 + \frac{3}{4}x_2x_3 + \frac{1}{4}x_1x_2x_3.
\]

(3.23)

That is, the Fourier spectrum has 8 terms and the function is not linear. In general, we would like a sparse Fourier spectrum for efficient implementation.

Table 3.1 presents the Fourier expansion of optimized relay functions for \( k = 2, 3, 4, 6 \) when \( \epsilon_1 = \epsilon_2 = \epsilon_r = 0.01 \). Interestingly, we see that the Fourier expansion of the optimized functions is indeed sparse. Using the results in Table 3.1, we can rewrite the functions in the following form

\[
f : \{0, 1\}^k \mapsto \{0, 1\}^k
\]

\[
x_r = A_ky_r + b_k,
\]

(3.24)

\[\text{We use the one-to-one mapping } 0 \leftrightarrow +1 \text{ and } 1 \leftrightarrow -1.\]
Table 3.1: Fourier expansion of the optimized relay functions for the orthogonal BSRC with $\epsilon_1 = \epsilon_2 = \epsilon_r = 0.01$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$f_1(y_{r1})$</th>
<th>$f_2(y_{r1})$</th>
<th>$f_3(y_{r1})$</th>
<th>$f_4(y_{r1})$</th>
<th>$f_5(y_{r1})$</th>
<th>$f_6(y_{r1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$y_{r1}$</td>
<td>$y_{r2}y_{r1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$y_{r3}y_{r2}y_{r1}$</td>
<td>$-y_{r3}y_{r2}y_{r1}$</td>
<td>$y_{r3}y_{r2}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$-y_{r4}y_{r3}y_{r2}y_{r1}$</td>
<td>$-y_{r4}y_{r3}y_{r2}y_{r1}$</td>
<td>$-y_{r3}y_{r2}y_{r1}$</td>
<td>$y_{r4}y_{r3}y_{r2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>$y_{r5}y_{r4}y_{r3}y_{r2}$</td>
<td>$-y_{r6}y_{r5}y_{r4}y_{r3}y_{r1}$</td>
<td>$-y_{r5}y_{r4}y_{r3}y_{r2}$</td>
<td>$-y_{r4}y_{r3}y_{r2}y_{r1}$</td>
<td>$y_{r6}y_{r5}y_{r4}y_{r3}$</td>
<td>$-y_{r6}y_{r5}y_{r4}y_{r3}y_{r2}$</td>
</tr>
</tbody>
</table>

where $x_r = [x_{r1}, \ldots, x_{r,k}]^T$ and $y_r = [y_{r1}, \ldots, y_{r,k}]^T$, and $A_k \in \{0,1\}^{[k \times k]}$ and $b_k \in \{0,1\}^{[k \times 1]}$. For example for $k = 6$, we have

$$A_6 = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}, \quad b_6 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Note that the mapping given in (3.24) is not linear in the binary field, when $b_k \neq 0$. However, the linear mapping $x_r = A_ky_r$ gives the same performance as $x_r = A_ky_r + b_k$. In other words the bias term $b_k$ does not improve the rate. This essentially follows from the data processing inequality. Therefore, the underlying relay functions define a linear code of rate one on the noisy received bits at the relay. Additionally, the code used at the relay performs joint source–channel coding, it therefore should be good for both source and channel coding.

### 3.4.4 Effect of Channel Parameters on the Structure of the Optimized Mappings

In this section, we investigate the structure of the optimized mappings for different channel parameters. Our numerical search indicates that the linear mapping $x_r = A_ky_r$ is an efficient strategy among all classes of mappings for low-dimensional
relaying. That is the relay employs the binary matrix $A_k$ to generate the relay outputs using $k$ received bits.

Tables 3.2 and 3.3 show the optimized generator matrices $A_k$ for various values of $\epsilon_r$ when $k = 6$, $\epsilon_1 = \epsilon_2 = 0.05$. In particular, for $\epsilon_r = 0.25$ the optimized generator matrix is the identity matrix, i.e., $A_6 = I_6$. For this case the relay is better off transmitting the received noisy bits without any further processing. However, as $\epsilon_r$ decreases, the relay starts combining the received bits at the relay before transmitting. The number of ones in a row of the generator matrix indicates the number of inputs that the relay combines. The density of ones in the generator matrices $\rho := \frac{\# \text{ of ones}}{k^2}$ is also shown in Tables 3.2 and 3.3. We see that as $\epsilon_r$ decreases $\rho$ increases. That is, the relay starts to transmit combinations of more bits in one single channel use. This occurs because of two main reasons: firstly when $\epsilon_r$ decreases the relay receives less noisy bits on average and secondly the destination has some partial knowledge of individual bits via the received signal from the source.

Tables 3.4 and 3.5 show the optimized generator matrices for various values of $\epsilon_1$ when $k = 6$, $\epsilon_r = 0.01$ and $\epsilon_2 = 0.1$. We similarly see that as $\epsilon_1$ decreases $\rho$ increases. This is due to the fact that when $\epsilon_1$ decreases the destination receives better descriptions of the transmitted bits via the source–destination link. The relay then forwards combinations of several incoming bits when the destination has access to more reliable side information.

Tables 3.6 and 3.7 show the optimized generator matrices for various values of $\epsilon_2$ when $k = 6$, $\epsilon_r = 0.01$ and $\epsilon_1 = 0.1$. We can see that as $\epsilon_2$ decreases, the density of ones $\rho$ decreases as well. For $\epsilon_2 = 0.0001$, we have $A_6 = I_6$. This is due to the fact that for small value of $\epsilon_2$ the relay can reliably transmit the received bits to the destination. (Note that this strategy is optimal if $\epsilon_2 = 0$). However, for high values of $\epsilon_2$, the relay combines several bits prior to the transmission. This helps the destination to combat the noise on the relay–destination link.
3.4. Capacity Bounds: Finite Memory Relay Case

Table 3.2: Optimized generator matrices as a function of $\epsilon_r$ for $k = 6$ and $\epsilon_1 = \epsilon_2 = 0.05$.

<table>
<thead>
<tr>
<th>$\epsilon_r$</th>
<th>0.25</th>
<th>0.1</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 0 0 0 0 0</td>
<td>0 0 1 0 0 0</td>
<td>1 0 0 1 0 0</td>
<td></td>
</tr>
<tr>
<td>0 1 0 0 0 0</td>
<td>0 1 0 0 0 0</td>
<td>1 1 1 1 0 0</td>
<td></td>
</tr>
<tr>
<td>0 0 1 0 0 0</td>
<td>0 0 0 0 0 0</td>
<td>1 0 1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>0 0 0 1 0 0</td>
<td>0 0 0 1 1 0</td>
<td>1 0 0 1 1 1</td>
<td></td>
</tr>
<tr>
<td>0 0 0 0 1 0</td>
<td>1 0 0 0 1 0</td>
<td>0 1 0 0 1 0</td>
<td></td>
</tr>
<tr>
<td>0 0 0 0 0 1</td>
<td>1 1 1 1 1 1</td>
<td>0 0 0 0 1 1</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.1667</td>
<td>0.3611</td>
<td>0.4444</td>
</tr>
</tbody>
</table>

Table 3.3: Optimized generator matrices as a function of $\epsilon_r$ for $k = 6$ and $\epsilon_1 = \epsilon_2 = 0.05$.

<table>
<thead>
<tr>
<th>$\epsilon_r$</th>
<th>0.01</th>
<th>0.001</th>
<th>0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1 0 0 0 1</td>
<td>1 1 1 0 0 1</td>
<td>1 0 0 1 0 1</td>
<td></td>
</tr>
<tr>
<td>0 1 0 1 1 0</td>
<td>0 1 1 1 0 0</td>
<td>0 1 0 1 0 1</td>
<td></td>
</tr>
<tr>
<td>0 0 1 1 1 0</td>
<td>0 1 1 0 1 1</td>
<td>1 1 1 0 0 1</td>
<td></td>
</tr>
<tr>
<td>1 0 1 1 0 0</td>
<td>1 1 0 1 1 0</td>
<td>1 1 1 1 0 0</td>
<td></td>
</tr>
<tr>
<td>1 0 1 0 1 1</td>
<td>1 0 0 0 1 1</td>
<td>1 1 0 0 1 1</td>
<td></td>
</tr>
<tr>
<td>0 1 1 0 1 1</td>
<td>1 0 1 1 1 0</td>
<td>0 0 1 1 1 1</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.5556</td>
<td>0.6111</td>
<td>0.6389</td>
</tr>
</tbody>
</table>

Table 3.4: Optimized generator matrices as a function of $\epsilon_1$ for $k = 6$ and $\epsilon_r = 0.01$, $\epsilon_2 = 0.1$.

<table>
<thead>
<tr>
<th>$\epsilon_1$</th>
<th>0.4</th>
<th>0.25</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 0 0 0 0 0</td>
<td>0 1 1 0 0 0</td>
<td>1 1 0 0 1 1</td>
<td></td>
</tr>
<tr>
<td>0 1 0 0 0 0</td>
<td>1 0 0 1 0 0</td>
<td>0 1 0 1 0 1</td>
<td></td>
</tr>
<tr>
<td>0 0 1 0 0 0</td>
<td>0 0 1 0 0 1</td>
<td>0 1 1 0 1 0</td>
<td></td>
</tr>
<tr>
<td>0 0 0 1 0 0</td>
<td>1 0 0 0 0 1</td>
<td>1 0 0 1 1 1</td>
<td></td>
</tr>
<tr>
<td>0 0 0 0 1 0</td>
<td>0 1 0 0 1 0</td>
<td>1 0 1 0 1 0</td>
<td></td>
</tr>
<tr>
<td>0 0 0 0 0 1</td>
<td>0 0 0 1 1 0</td>
<td>1 0 1 1 0 0</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.1667</td>
<td>0.3333</td>
<td>0.5556</td>
</tr>
</tbody>
</table>
Table 3.5: Optimized generator matrices as a function of $\epsilon_1$ for $k = 6$ and $\epsilon_r = 0.01$, $\epsilon_2 = 0.1$.

<table>
<thead>
<tr>
<th>$\epsilon_1$</th>
<th>0.05</th>
<th>0.01</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1 1 1 0 1 0</td>
<td>1 0 1 0 1 1</td>
<td>1 1 1 1 0 0</td>
<td></td>
</tr>
<tr>
<td>1 0 1 1 0 0</td>
<td>1 1 0 1 0 1</td>
<td>1 1 1 1 1 1</td>
<td></td>
</tr>
<tr>
<td>1 0 1 0 1 1</td>
<td>0 1 0 1 1 1</td>
<td>1 1 1 0 1 0</td>
<td></td>
</tr>
<tr>
<td>0 1 1 1 1 1</td>
<td>1 0 1 1 1 0</td>
<td>1 0 1 1 1 1</td>
<td></td>
</tr>
<tr>
<td>1 0 0 1 1 0</td>
<td>1 1 1 0 1 0</td>
<td>1 1 1 0 0 1</td>
<td></td>
</tr>
<tr>
<td>1 1 0 1 0 1</td>
<td>0 1 1 1 0 1</td>
<td>1 1 0 1 1 1</td>
<td></td>
</tr>
</tbody>
</table>

$\rho$ | 0.6389 | 0.6667 | 0.7500 |

Table 3.6: Optimized generator matrices as a function of $\epsilon_2$ for $k = 6$ and $\epsilon_r = 0.01$, $\epsilon_1 = 0.1$.

<table>
<thead>
<tr>
<th>$\epsilon_2$</th>
<th>0.4</th>
<th>0.25</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 1 1 1 1 1</td>
<td>1 1 0 1 1 0</td>
<td>0 1 1 1 1 1</td>
<td></td>
</tr>
<tr>
<td>1 1 1 1 0 1</td>
<td>0 1 1 0 1 1</td>
<td>1 0 1 1 0 1</td>
<td></td>
</tr>
<tr>
<td>1 0 1 1 1 1</td>
<td>1 1 1 1 0 0</td>
<td>0 1 0 1 0 1</td>
<td></td>
</tr>
<tr>
<td>1 1 1 0 1 1</td>
<td>1 1 0 1 0 1</td>
<td>1 1 1 0 1 1</td>
<td></td>
</tr>
<tr>
<td>1 1 1 1 1 0</td>
<td>1 0 1 0 1 1</td>
<td>0 1 0 0 1 1</td>
<td></td>
</tr>
<tr>
<td>1 1 0 1 1 1</td>
<td>0 0 1 1 1 1</td>
<td>1 0 0 0 1 1</td>
<td></td>
</tr>
</tbody>
</table>

$\rho$ | 0.8333 | 0.6667 | 0.6389 |

Table 3.7: Optimized generator matrices as a function of $\epsilon_2$ for $k = 6$ and $\epsilon_r = 0.01$, $\epsilon_1 = 0.1$.

<table>
<thead>
<tr>
<th>$\epsilon_2$</th>
<th>0.01</th>
<th>0.001</th>
<th>0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 1 1 0 1 0</td>
<td>0 0 1 0 0 1</td>
<td>1 0 0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>0 1 0 1 1 0</td>
<td>0 1 0 1 0 1</td>
<td>0 1 0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>0 0 0 1 1 1</td>
<td>1 1 0 0 0 1</td>
<td>0 0 1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>1 1 1 1 1 0</td>
<td>0 1 1 0 1 0</td>
<td>0 0 0 1 0 0</td>
<td></td>
</tr>
<tr>
<td>1 0 1 0 1 1</td>
<td>1 0 0 0 0 0</td>
<td>0 0 0 0 1 0</td>
<td></td>
</tr>
<tr>
<td>1 0 0 1 0 1</td>
<td>0 1 0 1 1 0</td>
<td>0 0 0 0 0 1</td>
<td></td>
</tr>
</tbody>
</table>

$\rho$ | 0.5833 | 0.4167 | 0.1667 |
3.5 Numerical Examples

Figure 3.5 shows the capacity results for the orthogonal BSRC shown in Figure 3.1 as a function of $\epsilon_r$ when $\epsilon_1 = \epsilon_2 = 0.01$. In this figure, we have plotted the cutset upper bound (UB) (Equation (3.7)), rates achieved using DF (Equation (3.8)), CF (Equation (3.11)), and optimized finite memory relay (Equation (3.15)) for different memory size. The relay functions are optimized for the channel parameters $\epsilon_1 = \epsilon_2 = \epsilon_r = 0.01$ and are given in Table 3.1.

From Figure 3.5, we see that the achievable rate of DF decreases as $\epsilon_r$ increases to 0.01. This is due to the requirement of successful decoding at the relay. This constraint makes the relay useless when $\epsilon_r \geq 0.01$. On the other hand, the rates achieved by CF are very close to the upper bound. More interestingly, optimized low-dimensional relaying with $k = 6$ achieves rates close to those achieved by CF.

3.6 Summary and Concluding Remarks

We introduced a binary symmetric relay channel with orthogonal receive components. We studied three main relaying strategies: DF, CF, and optimized low-dimensional relaying. Using numerical search, we identified a general efficient structure exhibited by the optimized mappings. We also illustrated that one can obtain rates very close to the upper bound by using optimized low-dimensional relaying. It is worth noting that DF and CF require codebooks with infinite block length codewords at the relay. This stands in a sharp contrast to the proposed low-dimensional relaying scheme. Additionally, the suggested relaying protocol has low-delay processing and paves the way for implementation of inexpensive relaying protocols.
Figure 3.5: Capacity results for the binary symmetric relay channel in Figure 3.1 as a function of $\epsilon_r$ when $\epsilon_1 = \epsilon_2 = 0.01$. The dashed curves show the achievable rates using the low-dimensional relaying with memory length $k$. The relay functions are optimized for the channel parameters $\epsilon_1 = \epsilon_2 = \epsilon_r = 0.01$ and are given in Table 3.1.
Appendix

3.A Proof of Proposition 3.3.1

In order to proceed, we first present a lemma which we occasionally use in the sequel.

Lemma 3.A.1. Consider a binary symmetric channel with input X and outputs Y_1 and Y_2, where

\[ Y_1 = X \oplus Z_1, \]
\[ Y_2 = X \oplus Z_2. \]

The random variable \( Z_1 \sim \text{Ber}(\epsilon_1) \) and is independent of \( Z_2 \sim \text{Ber}(\epsilon_2) \). The capacity of this channel is given by

\[ C = 1 + H_b(\epsilon_1 \ast \epsilon_2) - H_b(\epsilon_1) - H_b(\epsilon_2). \] (3.25)

Proof. The channel is a standard 1 \( \times \) 2 single-input multiple-output (SIMO) link and its capacity is given by

\[ C = \max_{p(x)} I(X; Y_1, Y_2). \] (3.26)

Next consider

\[
C = H(Y_1, Y_2) - H(Y_1, Y_2 | X) \\
= H(Y_1, Y_2) - H(Z_1, Z_2 | X) \\
= H(Y_1, Y_2) - H(Z_1) - H(Z_2) \\
= H(Y_1) + H(Y_2 | Y_1) - H_b(\epsilon_1) - H_b(\epsilon_2) \\
= H(Y_1) + H(Y_2 \oplus Y_1 | Y_1) - H_b(\epsilon_1) - H_b(\epsilon_2) \\
\leq H(Y_1) + H(Y_2 \oplus Y_1 | Y_1) - H_b(\epsilon_1) - H_b(\epsilon_2) \\
= H(Y_1) + H(Z_2 \oplus Z_1) - H_b(\epsilon_1) - H_b(\epsilon_2) \\
\leq 1 + H_b(\epsilon_1 \ast \epsilon_2) - H_b(\epsilon_1) - H_b(\epsilon_2). \] (3.27)

We finally note that the upper bound can be achieved by choosing \( X \sim \text{Ber}(\frac{1}{2}) \). \( \square \)

Proof of Proposition 3.3.1: Using the cutset bound given in (2.6), we have

\[ C \leq \max_{p(x_x)} \min \{ I(X, X_r; Y_1, Y_2), I(X; Y_1, Y_2, Y_r | X_r) \}. \]

Now we bound each term under min. Consider

\[
I(X, X_r; Y_1, Y_2) = H(Y_1, Y_2) - H(Y_1, Y_2 | X, X_r) \\
= H(Y_1, Y_2) - H(Z_1, Z_2 | X, X_r) \\
= H(Y_1, Y_2) - H(Z_1) - H(Z_2) \\
\leq H(Y_1) + H(Y_2) - H(Z_1) - H(Z_2) \\
\leq 2 - H_b(\epsilon_1) - H_b(\epsilon_2). \] (3.28)
Similarly, we have

\[ I(X;Y_r,Y_1,Y_2|X_r) = I(X;Y_1,Y_r|X_r) + I(X;Y_2|Y_r,X_r) \]
\[ = I(X;Y_1,Y_r|X_r) + H(Y_2|Y_r,X_r) \]
\[ - H(Y_2|Y_1,Y_r,X_r,X) \]
\[ = I(X;Y_1,Y_r|X_r) + H(Z_2|Y_1,Y_r,X_r) \]
\[ - H(Z_2|Y_1,Y_r,X_r,X) \]
\[ = I(X;Y_1,Y_r|X_r) \]
\[ \leq 1 + \mathbb{H}_b(\epsilon_1 + \epsilon_r) - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_r), \quad (3.29) \]

where the last inequality follows from Lemma 3.A.1. Combining (3.28) and (3.29) and noting that (3.28) and (3.29) are optimized when the input distribution is chosen as

\[ p(x,x_r) = p(x)p(x_r) = [0.5\delta_b(x) + 0.5\delta_b(x \oplus 1)] [0.5\delta_b(x_r) + 0.5\delta_b(x_r \oplus 1)], \]

yield the result.

3.B Proof of Proposition 3.3.2

Using DF lower bound given in (2.13) the rate

\[ R_{\text{DF}} = \max_{p(x,x_r)} \min \{I(X,X_r;Y_1,Y_2), I(X;Y_r|X_r)\}, \]

is achievable. The first term is evaluated in Appendix 3.A and is maximized when \( X \) and \( X_r \) are independent and have uniform distribution. One can show the same distribution maximizes the second term. This yields

\[ R_{DF} = \min\{1 - \mathbb{H}_b(\epsilon_r), 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2)\}. \quad (3.30) \]

3.C Proof of Proposition 3.3.3

Using the partial DF lower bound given in (2.17) the rate

\[ R_{\text{PDF}} = \max_{p(u,x,x_r)} \min \{I(X,X_r;Y_1,Y_2), I(U;Y_r|X_r) + I(X;Y_1,Y_2|X_r,U)\}, \]
is achievable. The first term is evaluated in Appendix 3.A and is bounded as

$$I(X, X_r; Y_1, Y_2) \leq 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2).$$  \hspace{1cm} (3.31)

Next consider

$$I(U; Y_r | X_r) + I(X; Y_1, Y_2 | X_r, U)$$
$$= I(U; Y_r | X_r) + H(Y_1, Y_2 | X_r, U) - H(Y_1, Y_2 | X, X_r, U)$$
$$= I(U; Y_r | X_r) + H(Y_1, Y_2 | X_r, U) - H(Z_1, Z_2 | X, X_r, U)$$
$$= I(U; Y_r | X_r) + H(Y_1, Y_2 | X_r, U) - H(Z_1) - H(Z_2)$$
$$= I(U; Y_r | X_r) + H(Y_1 | X_r, U) + H(Y_2 | X_r, U, Y_1) - H(Z_1) - H(Z_2)$$
$$= I(U; Y_r | X_r) + H(Y_1 | X_r, U) + H(Z_2 | X_r, U, Y_1) - H(Z_1) - H(Z_2)$$
$$= H(Y_r | X_r) - H(Y_r | X_r, U) + H(Y_1 | X_r, U) - H(Z_1)$$
$$\leq H(Y_r) - H(Y_r | X_r, U) + H(Y_1 | X_r, U) - H(Z_1)$$
$$\leq 1 - \mathbb{H}_b(\epsilon_1) + H(Y_1 | X_r, U) - H(Y_r | X_r, U),$$  \hspace{1cm} (3.32)

Now we bound $H(Y_1 | X_r, U) - H(Y_r | X_r, U)$. First, define $V := (X_r, U)$ where $p(V = v_i) = p_i$ and $\sum_i p_i = 1$. Further assume that $p(x = 0 | V = v_i) = \delta_i$, $p(x = 1 | V = v_i) = 1 - \delta_i$. Next consider

$$H(Y_1 | X_r, U) - H(Y_r | X_r, U)$$
$$= H(X + Z_1 | V) - H(X + Z_r | V)$$
$$= \sum_i (H(X + Z_1 | V = v_i) - H(X + Z_r | V = v_i)) p_i$$
$$\leq \sum_i \max_{\delta_1} (H(X + Z_1 | V = v_i) - H(X + Z_r | V = v_i)) p_i$$
$$= \max_{\delta_1} (H(X + Z_1 | V = v_1) - H(X + Z_r | V = v_1)) \sum_i p_i$$
$$= \max_{\delta_1} \left( \mathbb{H}_b(\delta_1 * \epsilon_1) - \mathbb{H}_b(\delta_1 * \epsilon_r) \right).$$  \hspace{1cm} (3.33)

In the following, let $h(\delta) := \mathbb{H}_b(\delta * \epsilon_1) - \mathbb{H}_b(\delta * \epsilon_r)$ and $\max{\epsilon_1, \epsilon_r} \leq 0.5$. We then obtain\(^2\)

$$\frac{\partial h}{\partial \delta} = (1 - 2\epsilon_1) \log \left( \frac{1 - (\delta * \epsilon_1)}{\delta * \epsilon_1} \right) - (1 - 2\epsilon_r) \log \left( \frac{1 - (\delta * \epsilon_r)}{\delta * \epsilon_r} \right),$$

$$\frac{\partial^2 h}{\partial \delta^2} = \left( \frac{1 - 2\epsilon_r}{(\delta * \epsilon_r)(1 - (\delta * \epsilon_r))} - \frac{(1 - 2\epsilon_1)^2}{(\delta * \epsilon_1)(1 - (\delta * \epsilon_1))} \right).$$  \hspace{1cm} (3.34)

\(^2\)Here, we assume natural logarithms.
Now let $g(\epsilon) := \frac{(1-2\epsilon)^2}{(3+2(1-3\epsilon))}$. One can show that $\frac{\partial g}{\partial \epsilon} \leq 0$ if $\epsilon \leq 0.5$ and hence $g(\epsilon)$ is a non-increasing function. Thus we conclude that

\[
\begin{align*}
\epsilon_r \leq \epsilon_1 & \implies g(\epsilon_r) \geq g(\epsilon_1) \implies \frac{\partial^2 g}{\partial \epsilon^2} \geq 0 \implies h(\delta) \text{ is convex}
\epsilon_r \geq \epsilon_1 & \implies g(\epsilon_r) \leq g(\epsilon_1) \implies \frac{\partial^2 g}{\partial \epsilon^2} \leq 0 \implies h(\delta) \text{ is concave}
\end{align*}
\]

(3.35)

Therefore

\[
\begin{align*}
\left\{ \begin{array}{ll}
h(\delta) \leq \max\{h(0), h(1)\}, & \text{if } \epsilon_r \leq \epsilon_1 \\
h(\delta) \leq h(\delta^*), & \text{if } \epsilon_r \geq \epsilon_1
\end{array} \right.
\]

(3.36)

where $\delta^*$ is the solution of $\frac{\partial h}{\partial \epsilon} = 0$. Finally we obtain the following bound

\[
H(Y_1|X_r, U) - H(Y_r|X_r, U) \leq \left\{ \begin{array}{ll}
\mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_r), & \text{if } \epsilon_r \leq \epsilon_1 \\
0, & \text{otherwise}
\end{array} \right.
\]

(3.37)

Combining the equations (3.31), (3.32) and (3.37), proves that partial DF does not improve on DF for the BSRC.

3.D Proof of Proposition 3.3.5

We use the equivalent formulation of the original CF given in (2.24) with time-sharing. CF achieves the rate

\[
R_{CF} = \max_{p(q)p(x|q)p(y_r|z_r,y_r,q)} \min \left\{ I(X; Y_1, Y_2, \hat{Y}_r|X_r, Q), \right. \\
I(X, X_r; Y_1, Y_2|Q) - I(Y_r; \hat{Y}_r|X, X_r, Y_1, Y_2, Q) \left. \right\},
\]

(3.38)

where $Q$ denotes the time-sharing random variable. First consider

\[
I(X, X_r; Y_1, Y_2|Q) = I(X; Y_1, Y_2|Q) + I(X_r; Y_1, Y_2|X, Q)
= I(X; Y_1|Q) + I(X_r; Y_2|Y_1, Q) + I(X_r; Y_2|X, Q)
+ I(X_r; Y_1|Y_2, X, Q)
= I(X; Y_1|Q) + I(X_r; Y_2|X, Q)
= I(X; Y_1|Q) + I(X_r; Y_2|Q)
\leq 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2),
\]

(3.39)

where the upper bound can be achieved by choosing $X \sim \text{Ber}(\frac{1}{2})$ and $X_r \sim \text{Ber}(\frac{1}{2})$.

In order to proceed we choose the following binary test channel:

\[
\hat{Y}_r = Y_r \oplus Z_q,
\]

(3.40)

where $Z_q \sim \text{Ber}(\epsilon_q)$ and is independent of other random variables. This yields

\[
I(Y_r; \hat{Y}_r|X, X_r, Y_1, Y_2, Q) = H(Y_r|X, X_r, Y_1, Y_2, Q) - H(\hat{Y}_r|Y_r, X, X_r, Y_1, Y_2, Q)
= H(Z_r \oplus Z_q) - H(Z_q) = \mathbb{H}_b(\epsilon_r \ast \epsilon_q) - \mathbb{H}_b(\epsilon_q).
\]

(3.41)
and
\[ I(X; Y_1, Y_2, \hat{Y}_r | X_r, Q) = I(X; Y_1, \hat{Y}_r | X_r, Q) + I(X; Y_2 | Y_1, \hat{Y}_r, X_r, Q) \]
\[ = I(X; Y_1, \hat{Y}_r | Q) \]
\[ \leq 1 + H_b(\epsilon_1 * \epsilon_r * \epsilon_q) - H_b(\epsilon_r * \epsilon_q) - H_b(\epsilon_1), \tag{3.42} \]
where the last inequality follows from Lemma 3.A.1 and it is achieved by choosing \( X \sim \text{Ber}(\frac{1}{2}) \).

Putting all together, the following rate is achievable
\[ R_{CF} = \max_{\epsilon_q \in [0, 1]} \min \left\{ 2 - H_b(\epsilon_1) - H_b(\epsilon_2) - H_b(\epsilon_r * \epsilon_q) + H_b(\epsilon_q), 1 + H_b(\epsilon_1 * \epsilon_r * \epsilon_q) - H_b(\epsilon_r * \epsilon_q) - H_b(\epsilon_1) \right\}. \tag{3.43} \]

Now define
\[ R_1(\epsilon_q) := 2 - H_b(\epsilon_1) - H_b(\epsilon_2) - H_b(\epsilon_r * \epsilon_q) + H_b(\epsilon_q), \]
\[ R_2(\epsilon_q) := 1 + H_b(\epsilon_1 * \epsilon_r * \epsilon_q) - H_b(\epsilon_r * \epsilon_q) - H_b(\epsilon_1). \tag{3.44} \]
Using the fact that \( f(\epsilon) := H_b(\epsilon * \delta) - H_b(\epsilon) \) is convex \( \forall \delta \in [0, 1] \), we conclude that \( R_1(\epsilon_q) \) is concave and \( R_2(\epsilon_q) \) is convex in \( \epsilon_q \). We next note that
\[ \max_{\epsilon_q} R_1 = R_1(0.5) = 2 - H_b(\epsilon_1) - H_b(\epsilon_2), \]
\[ \min_{\epsilon_q} R_2 = R_2(0.5) = 1 - H_b(\epsilon_1). \tag{3.45} \]
Since \( R_2(0.5) \leq R_1(0.5) \) we only need to consider two following cases:

- **Case 1**: \( R_1(0) < R_2(0) \)
  If \( R_1(0) < R_2(0) \) we have \( 1 - H_b(\epsilon_2) < H_b(\epsilon_1 * \epsilon_r) \) and there exists \( \epsilon_q \) such that \( R_1(\epsilon_q) = R_2(\epsilon_q) \). Thus
  \[ R_{CF} = 1 + H_b(\epsilon_1 * \epsilon_r * \epsilon_q) - H_b(\epsilon_r * \epsilon_q) - H_b(\epsilon_1), \tag{3.46} \]
  is achievable where \( \epsilon_q \) satisfies
  \[ H_b(\epsilon_1 * \epsilon_r * \epsilon_q) - H_b(\epsilon_q) + H_b(\epsilon_2) = 1. \tag{3.47} \]

- **Case 2**: \( R_1(0) \geq R_2(0) \)
  If \( R_1(0) \geq R_2(0) \) we have \( 1 - H_b(\epsilon_2) \geq H_b(\epsilon_1 * \epsilon_r) \). Thus
  \[ R_{CF} = \max_{\epsilon_q} \min \{ R_1(\epsilon_q), R_2(\epsilon_q) \} = \max_{\epsilon_q} R_2(\epsilon_q) = R_2(0) \]
  \[ = 1 + H_b(\epsilon_1 * \epsilon_r) - H_b(\epsilon_r) - H_b(\epsilon_1), \tag{3.48} \]
  is achievable.
3.E Proof of Proposition 3.4.1

For the memoryless relay, we have \( x_r = y_r \) and hence

\[
Y_1 = X \oplus Z_1, \\
Y_2 = X_r \oplus Z_2 = X \oplus Z_r \oplus Z_2 = X \oplus Z_{eq},
\]

where \( Z_{eq} := Z_r \oplus Z_2 \sim \text{Ber}(\epsilon_r + \epsilon_2) \). Then the achievable rate is given by \( C_1 = \max_{p(x)} I(X; Y_1, Y_2) \). Now using Lemma 3.A.1, we obtain

\[
C_1 = 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r * \epsilon_2) - \mathbb{H}_b(\epsilon_r * \epsilon_2) - \mathbb{H}_b(\epsilon_1).
\]

3.F Proof of Lemma 3.4.2

Consider

\[
I(X_1^k; Y_{1,1}^k, Y_{2,1}^k) = I(X_1^k; Y_{1,1}^k) + I(X_1^k; Y_{2,1}^k | Y_{1,1}^k)
\]

\[
\stackrel{(a)}{=} k I(X_1; Y_1) + I(X_1^k; Y_{2,1}^k | X_1)
\]

\[
= k I(X_1; Y_1) + H(Y_{2,1}^k | Y_{1,1}^k) - H(Y_{2,1}^k | Y_{1,1}^k, X_1)
\]

\[
\stackrel{(b)}{=} k I(X_1; Y_1) + H(Y_{2,1}^k | Y_{1,1}^k) - H(Y_{2,1}^k | X_1^k)
\]

\[
= k(1 - \mathbb{H}_b(\epsilon_1)) - \mathbb{E} [\log_2(p(y_{2,1}^k | y_{1,1}^k))] + \mathbb{E} [\log_2(p(y_{2,1}^k | x_1^k))],
\]

where \((a)\) holds since \(\{X_i\}_{i=1}^k\) are i.i.d. and the channel is memoryless and \((b)\) holds since \(Y_{1,1}^k = X_1^k \oplus Z_{1,1}^k\) and \(Z_{1,1}^k\) is independent of other random sequences. The conditional probabilities can be computed as follows

\[
p(y_{2,1}^k | x_1^k) = \sum_{y_{1,1}^k} p(y_{2,1}^k | x_1^k, y_{1,1}^k) p(y_{1,1}^k | x_1^k)
\]

\[
= \sum_{y_{1,1}^k} p(y_{2,1}^k | y_{1,1}^k) p(y_{1,1}^k | x_1^k)
\]

\[
= \sum_{y_{1,1}^k} p(y_{2,1}^k | y_{1,1}^k) \prod_{i=1}^k p(y_{r,i} | x_i)
\]

\[
= \sum_{y_{1,1}^k} \prod_{i=1}^k p(y_{2,i} | f_i(y_{r,i}^k)) \prod_{i=1}^k p(y_{r,i} | x_i)
\]

\[
= \sum_{y_{1,1}^k} \delta_{y_{1,1}^k, f(y_{r,i}^k)} (1 - \epsilon_2)^k \delta_{y_{2,i}, f(y_{r,i}^k)}
\]

\[
\times \epsilon_r \delta_{y_{r,i}, x_i} (1 - \epsilon_r)^k \delta_{y_{r,i}, x_i}.
\]
3.F. Proof of Lemma 3.4.2

\[ (1 - \epsilon_2)^k (1 - \epsilon_r)^k \times \sum_{y_{r_1}^k} \left( \frac{\epsilon_2}{1 - \epsilon_2} \right)^{D_H(y_{r_1}^k, f(y_{r_1}^k))} \left( \frac{\epsilon_r}{1 - \epsilon_r} \right)^{D_H(y_{r_1}^k, x_{r_1}^k)}. \]  

(3.52)

We similarly obtain

\[ p(y_{2,1}^k | y_{1,1}^k) = \sum_{y_{r_1}^k} p(y_{2,1}^k | y_{1,1}^k, y_{r_1}^k) p(y_{r_1}^k | y_{1,1}^k) \]

\[ = \sum_{y_{r_1}^k} p(y_{2,1}^k | y_{r_1}^k) p(y_{r_1}^k | y_{1,1}^k) \]

\[ = \sum_{y_{r_1}^k} p(y_{2,1}^k | y_{r_1}^k) \sum_{x_{1}^k} p(y_{r_1}^k | x_{1}^k) p(x_{1}^k | y_{r_1}^k) \]

\[ = \sum_{y_{r_1}^k} p(y_{2,1}^k | y_{r_1}^k) \sum_{x_{1}^k} p(y_{r_1}^k | x_{1}^k) p(x_{1}^k | y_{r_1}^k) \]

\[ \prod_{i=1}^k p(y_{2,1}^k | f_i(y_{r_1}^k)) \sum_{x_{1}^k} \prod_{i=1}^k p(y_{r_1}^k | x_i) \prod_{i=1}^k p(y_{1,1} | x_i) \]

\[ = (1 - \epsilon_1)^k (1 - \epsilon_2)^k (1 - \epsilon_r)^k \sum_{y_{r_1}^k} \left( \frac{\epsilon_2}{1 - \epsilon_2} \right)^{D_H(y_{r_1}^k, f(y_{r_1}^k))} \times \]

\[ \sum_{x_{1}^k} \left( \frac{\epsilon_r}{1 - \epsilon_r} \right)^{D_H(y_{r_1}^k, x_{r_1}^k)} \left( \frac{\epsilon_1}{1 - \epsilon_1} \right)^{D_H(y_{1,1}^k, x_{1}^k)}, \]  

(3.53)

where (a) follows by Bayes’ rule and (b) holds since \( X_{1,i} \sim \text{Ber} \left( \frac{1}{2} \right) \) and \( Y_{1,i} \sim \text{Ber} \left( \frac{1}{2} \right) \).
Chapter 4

Orthogonal Gaussian Relay Channels

4.1 Introduction

4.1.1 Scope and Motivation

In this chapter, we consider some special cases of the general Gaussian three-node relay channel, with a source–destination pair and a relay node whose sole purpose is to support the communication between the source and the destination. We focus on Gaussian relay channels with orthogonal receive components in this chapter. That is, the received signals from the source and the relay do not interfere with each other (see also Definition 2.3.1). A similar channel model has been investigated in, e.g., [GMZ06] as well.

A natural way to implement instantaneous relaying for Gaussian relay channels is to retransmit the received signal at the relay. This strategy is known as linear relaying and is usually referred to as amplify-and-forward (AF) [LWT08, GMZ06, NBK04, AT07]. In Chapter 2, we showed that AF is optimal for some Gaussian channels. In particular, in the relay channels considered in Examples 3 and 4, AF is optimal because the signal that the relay transmits is new compared to that the destination has received via the direct link (i.e., source–destination link). However, AF is a suboptimal strategy in general. To see this, for simplicity, assume the link from the source to the destination and the relay as a broadcast channel with a noiseless source–relay link. Thus, when employing AF, the relay repeats the same signal as the one transmitted by the source to the destination. In other words, AF can be considered as a strategy that implements repetition coding which is in general highly suboptimal. On the other hand, by using nonlinear relaying one can potentially pack more codewords in the available signal space, subject to a given power constraint. This motivates us to investigate nonlinear instantaneous relaying schemes.

4.1.2 Related Work

In what follows we briefly review previous papers that investigate instantaneous relaying. These articles can be divided in two major groups: uncoded and coded...
transmission. The paper [FM04] finds the optimal relaying function in conjunction with uncoded BPSK signaling for two-hop transmission (i.e., the relay channel without the source–destination link). The work of [GJ07] formulates a design criterion based on the signal-to-noise ratio (SNR) under a modulation constraint with a focus on BPSK modulation, and investigates amplify-and-forward, demodulate-and-forward (DeF) and estimate-and-forward (EF) relaying. In [KL07a, KL08] mapping designs for higher-order modulation using DeF relaying are investigated. The paper [CHK08] considers instantaneous relaying for the two-way relay channel using uncoded binary antipodal signaling. In [GJ06], the relay mapping is optimized for maximum mutual information, and numerical results demonstrate that instantaneous EF is optimal in the case of BPSK signaling for two-hop transmission. The paper [GHM07] studied instantaneous relaying as a special case of relay channels with delay. However, [GHM07] assumes non-orthogonal receive signals at the destination in contrast to our work where we study the relay channel with orthogonal receive components. The work in [GMZ06] considers instantaneous relaying with orthogonal signals at the destination, however [GMZ06] only investigates conventional linear relaying (amplify-and-forward). Finally, the paper [YKS08] proposes a memoryless sawtooth mapping for the Gaussian relay channel with orthogonal receive signals. In this chapter, we show that this mapping is however not optimal in general.

4.1.3 Structure of the Chapter and Contributions

- Throughout we focus on continuous-amplitude channels, and in particular Gaussian channels. Section 4.2 presents the considered class of relay channels with orthogonal receive components. We present an achievable rate for the relay channel with instantaneous relaying and provide an upper bound on the capacity.

- Section 4.3 studies a special case of the Gaussian relay channel with a noise-free source–relay link. We show that memoryless relaying can achieve capacity in this scenario. The optimal mapping is a periodic function formed by a high-rate sawtooth mapping followed by nonlinear shaping. We emphasize that, for this class of relay channels, conventional linear relaying can have an unbounded gap to the capacity. We then discuss the connection between the optimal mapping and the theory of high-rate quantization. We finally verify our analytical results using Monte Carlo simulations.

- Section 4.4 establishes the capacity of a Gaussian relay channel where the destination knows the noise on the source–relay link. (Such noise in the source–relay link can model, e.g., interference that is generated by the destination when communicating with another user.) We show that compress-and-forward (CF) and memoryless relaying achieve the capacity. For this scenario, decode-and-forward (DF) and linear relaying are unable to attain the capacity bound and can have an unbounded gap to the capacity.
4.2 Orthogonal Instantaneous Relay Channel

Figure 4.1: Block diagram of the instantaneous relay channel with orthogonal receive components.

- Section 4.5 investigates the general Gaussian relay channel with orthogonal receive components. We propose a parametric piecewise linear (PL) relaying scheme, in which the relay employs a PL mapping to reshape the received signal at the relay. By construction, conventional linear relaying is a special case of the proposed PL mapping. Therefore, the performance of properly designed PL relaying is always at least as good as that of conventional linear relaying (AF). We then optimize the PL relay functionality, given a certain number of parameters, to assess the potential gain that this scheme can offer. We demonstrate that PL relaying can outperform both AF and DF relaying in some cases and it supports rates close to those achieved by CF. We stress that PL relaying is much less complex than DF and CF which require encoding of block codes at the relay.

- Finally, Section 4.6 concludes the chapter.

4.2 Orthogonal Instantaneous Relay Channel

Figure 4.1 illustrates the relay channel that we investigate in this chapter. This channel model falls into the class of relay channels with orthogonal receive components considered in Definition 2.3.1. The source intends to reliably communicate the message $m$ chosen uniformly from the set $\mathcal{M}^{(n)} := \left[ 1 : 2^n R \right]$ to the destination in $n$ channel uses. The symbols $x \in \mathcal{X}$ and $x_r \in \mathcal{X}_r$ respectively denote the signal transmitted from the source and the relay. The received symbols at the relay and the destination are denoted by $y_r \in \mathcal{Y}_r$ and $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$, respectively. We focus on continuous-amplitude and real-valued channels, so all alphabets are subsets of $\mathbb{R}$. We assume that the channel is memoryless and the interaction among the source, the relay, and the destination is governed by the following probability density function (pdf),

$$ p(y_1^n, y_2^n, y_r^n | x^n, x_r^n) = \prod_{i=1}^{n} p(y_1 | x_i) p(y_2 | x_r) p(y_r | x_i). \quad (4.1) $$
We additionally assume that the relay works *instantaneously* using a deterministic mapping defined as

\[ f : \mathcal{Y}_r \rightarrow \mathcal{X}_r \]

\[ x_{ri} = f(y_{ri}), \quad \forall i \in [1:n]. \]

For any fixed and given memoryless relay mapping \( f(\cdot) \), the instantaneous relay channel can be seen as a memoryless point-to-point communication link with input \( x \) and output \((y_1, y_2)\) parameterized by the pdf \( p(y_1, y_2|x) \) given as

\[
p(y_1, y_2|x) = \int_{y_r \in \mathcal{Y}_r} p(y_1, y_2|x, y_r)p(y_r|x)dy_r
\]

\[
= \int_{y_r \in \mathcal{Y}_r, x_r = f(y_r)} p(y_1|x)p(y_2|x_r)p(y_r|x)dy_r.
\]

(4.2)

For a given mapping \( f(\cdot) \), the achievable rate from \( x \) to \((y_1, y_2)\) is obtained as

\[ R_f = \max_{p(x)} I(X; Y_1, Y_2), \quad (4.3) \]

where the maximum is taken over the source pdf \( p(x) \in \mathcal{P} \), where \( \mathcal{P} \) denotes the class of permissible \( p(x) \) due to a cost constraint (power constraint). Then, maximizing over all possible mappings \( f(\cdot) \in \mathcal{F} \) at the relay, where \( \mathcal{F} \) denotes a cost constraint for the relay output, gives the best possible rate achievable from \( x \) to \((y_1, y_2)\). This yields the achievable rate

\[ C_1 = \sup_{f(\cdot) \in \mathcal{F}} C_f. \quad (4.4) \]

One can obtain lower bounds on \( C_1 \) by using any arbitrary choices of \( p(x) \in \mathcal{P} \) and \( f(\cdot) \in \mathcal{F} \). Thus, for given \( p(x) \) and \( f(\cdot) \), the number

\[ R = I(X; Y_1, Y_2) \quad (4.5) \]

provides a lower bound on \( C_1 \). That is \( R \leq C_1 \).

### 4.2.1 Upper Bound on the Capacity

Specializing the general cost-constraints to power constraints, that is \( \mathcal{P} \) and \( \mathcal{F} \) are specified by \( \mathbb{E}[X^2] \leq P_s \) and \( \mathbb{E}[X_r^2] \leq P_r \), for some \( P_s \) and \( P_r \), we get the following upper bound on the capacity.

**Proposition 4.2.1.** For the instantaneous relay channel with orthogonal receive components we have

\[
C_1 \leq \max_{p(x)p(x_r)} \min \{ I(X; Y_1) + I(X_r; Y_2), I(X; Y_1, Y_r) \}, \quad (4.6)
\]

where the maximum is subject to \( \mathbb{E}[X^2] \leq P_s \) and \( \mathbb{E}[X_r^2] \leq P_r \).
4.2. Orthogonal Instantaneous Relay Channel

Proof. Consider the following chain of inequalities

\[ I(X; Y_1, Y_2) = I(X; Y_1) + I(X; Y_2 | Y_1) \]
\[ = I(X; Y_1) + h(Y_2 | Y_1) - h(Y_2 | X, Y_1) \]
\[ \overset{(a)}{=} I(X; Y_1) + h(Y_2) - h(Y_2 | X, Y_1) \]
\[ \overset{(b)}{=} I(X; Y_1) + h(Y_2) - h(Y_2 | X, Y_1, X_r) \]
\[ \overset{(c)}{=} I(X; Y_1) + h(Y_2) - h(Y_2 | X_r) \]
\[ = I(X; Y_1) + I(X_r; Y_2), \quad (4.7) \]

where

(a) since conditioning reduces entropy;

(b) the same as (a); and

(c) follows from the fact that given \( X_r, Y_2 \) is independent of \( X \) and \( Y_1 \).

Now consider,

\[ I(X; Y_1, Y_2) \overset{(a)}{=} I(X; Y_1, Y_2) + I(X; Y_r | Y_1, Y_2) \]
\[ = I(X; Y_1, Y_2, Y_r) \]
\[ = I(X; Y_1, Y_r) + I(X; Y_2 | Y_1, Y_r) \]
\[ = I(X; Y_1, Y_r) + h(Y_2 | Y_1, Y_r) - h(Y_2 | X, Y_1, Y_r) \]
\[ \overset{(b)}{=} I(X; Y_1, Y_r) + h(Y_2 | Y_r) - h(Y_2 | Y_r) \]
\[ = I(X; Y_1, Y_r), \quad (4.8) \]

where

(a) follows from the fact that \( I(X; Y_r | Y_1, Y_2) \geq 0 \); and

(b) holds because given \( Y_r, Y_2 \) is independent of \( X \) and \( Y_1 \).

Combining (4.7) and (4.8) yields (4.6).

Remark 4.2.2. One can verify that the upper bound in (4.6) also holds for strictly causal relaying as defined in Chapter 2. In other words, the cutset bound in (2.6), for the relay channel in Figure 4.1, simplifies to (4.6). In Appendix 4.A, we show that this bound also holds for noncausal relaying in which the relay uses the entire received block to generate a symbol to be transmitted to the destination: \( x_{ri} = f(y^{n}_r), \forall i \in [1 : n] \). This essentially follows from two main observations:

i) \( X^n \rightarrow Y^n_r \rightarrow Y^n_2 \) form a Markov chain; and
Figure 4.2: Block diagram of the instantaneous relay channel with orthogonal receive components and perfect source–relay link.

ii) $Y^n_1$ and $Y^n_2$ are orthogonal.

If one removes the orthogonality condition, the bound is no longer valid (see [GHM07]).

4.3 A Semi-Deterministic Gaussian Relay Channel

In this section, we consider the relay channel shown in Figure 4.2, assuming Gaussian channels and with a noise-free source–relay link. This scenario is, of course, not practical. However, it will turn out that studying it will give useful insight into the general problem. Furthermore, the results we obtain here can be expected to approximately predict the general structure of the optimal system for very high SNR in the source–relay link.

The received signals at the destination are given by

$$y_1 = x + z_1$$

$$y_2 = x_r + z_2 = f(x) + z_2,$$

where $x$ and $x_r = f(x)$ denote the transmitted symbols from the source and the relay, respectively, $f(\cdot)$ denotes the one-dimensional symbol-by-symbol mapping used at the relay, and $Z_1 \sim N(0, 1)$ and $Z_2 \sim N(0, 1)$ are assumed to be independent zero-mean white Gaussian noise terms with unit variance. We assume average power constraints at the source and the relay; i.e., $\mathbb{E}[X^2] \leq P_s$ and $\mathbb{E}[f^2(X)] \leq P_r$.

4.3.1 Constructing the Optimal Instantaneous Mapping

Lemma 4.3.1. Using a Gaussian source codebook (i.e., $X \sim N(0, P_s)$) and with a noise-free source–relay link, the following are sufficient conditions to achieve capacity.

i) $I(X_r; Y_1) = 0$; and
ii) \(X_r \sim \mathcal{N}(0, P_r)\).

**Proof.** For a given relay mapping the following rate is achievable

\[
I(X; Y_1, Y_2) = I(X; Y_1) + I(X; Y_2|Y_1) \\
= I(X; Y_1) + h(Y_2|Y_1) - h(Y_2|Y_1, X) \\
\overset{(a)}{=} I(X; Y_1) + h(Y_2|Y_1) - h(Z_2|Y_1, X) \\
\overset{(b)}{=} I(X; Y_1) + h(Y_2|Y_1) - h(Z_2),
\]

where (a) holds since \(Y_2 = f(X) + Z_2\), and (b) follows because \(Z_2\) is independent of \(X\) and \(Y_1\). Next we bound \(h(Y_2|Y_1)\). Consider

\[
h(Y_2|Y_1) = h(X_r + Z_2|Y_1) \\
\leq \frac{1}{2} \log \left( 2^{2h(X_r|Y_1)} + 2^{2h(Z_2|X+Z_1)} \right) \\
= \frac{1}{2} \log \left( 2^{2h(X_r|Y_1)} + 2^{2h(Z_2)} \right) \\
= \frac{1}{2} \log \left( 2^{2h(X_r)-I(X_r;Y_1)} + 2^{2h(Z_2)} \right),
\]

where we used the entropy power inequality [CT06]. Substituting the conditions that \(X \sim \mathcal{N}(0, P_s)\), \(X_r \sim \mathcal{N}(0, P_r)\) and \(I(X_r;Y_1) = 0\) in (4.11) and (4.12), we obtain

\[
I(X; Y_1, Y_2) \geq C(P_s) + C(P_r),
\]

where \(C(x) := \frac{1}{2} \log(1 + x)\). Now using the upper bound given in (4.6), we have

\[
I(X; Y_1, Y_2) \leq C(P_s) + C(P_r).
\]

We thus conclude that \(X \sim \mathcal{N}(0, P_s)\), \(X_r \sim \mathcal{N}(0, P_r)\) and \(I(X_r;Y_1) = 0\) are sufficient to achieve the capacity. \(\square\)

Note that to make \(I(X_r;Y_1) = I(X_r;X+Z_1) = 0\), the random variable \(X + Z_1\) should not contain any information regarding \(X_r\) and vice versa. That
is \( h(X_r | X + Z_1) = h(X_r) \). This seems to stand in stark contrast to using a relay in which \( X_r = f(X) \) and \( f \) is a deterministic mapping. However, in the following we establish that, under certain conditions, it is possible to construct a mapping that asymptotically makes the conditions in Lemma 4.3.1 fulfilled, i.e., 
\[ h(X_r | X + Z_1) \rightarrow h(X_r) \] and \( p(x_r) \rightarrow N(0, P_r) \). The main idea is to use a high-rate sawtooth mapping to generate a uniform output which is uncorrelated with the input, and then construct a Gaussian output using an appropriate memoryless mapping. See Figure 4.3 for an illustration.

**Proposition 4.3.2.** Let \( X \) be a continuous random variable with the probability density function \( p(x) \), and let \( X_r = g_\Delta(X) \) where \( g_\Delta \) is a deterministic and periodic function with period \( \Delta \) which is invertible within each period. If the densities \( p(x) \) and \( p(y_1 | x) \) are Riemann integrable, then

\[
\lim_{\Delta \rightarrow 0} I_\Delta(X_r; Y_1) = 0.
\]  

**Proof.** For a fixed \( \Delta \), let \( \Delta_i \) denote the \( i \)th period (an interval of length \( \Delta \)) of \( g_\Delta \). We have

\[
p(y_1 | x_r) = \int p(y_1 | x, x_r)p(x | x_r)dx
\]

\[
\stackrel{(a)}{=} \int p(y_1 | x)p(x)\frac{p(x_r | x)}{p(x_r)}dx
\]

\[
= \int p(y_1 | x)p(x)\frac{p(x_r | x)}{\int p(x)p(x_r | x)dx}dx
\]

\[
\stackrel{(b)}{=} \sum \int_{x \in \Delta_i} p(y_1 | x)p(x)\frac{p(x_r | x)}{\sum_{j} \int_{x \in \Delta_j} p(x)p(x_r | x)dx}dx
\]

\[
\stackrel{(c)}{=} \sum \frac{\int_{x \in \Delta_i} p(y_1 | x)p(x)\delta(x_r - g_\Delta(x))dx}{\sum_{j} \int_{x \in \Delta_j} p(x)\delta(x_r - g_\Delta(x))dx}
\]

\[
\stackrel{(d)}{=} \sum \frac{p(y_1 | x_i)p(x_i)}{\sum_i p(x_i)}, \quad (4.16)
\]

where \( \delta(\cdot) \) denotes the Dirac \( \delta \)-function, and

(a) follows by applying Bayes’ rule and using the fact that \( p(y_1 | x, x_r) = p(y_1 | x) \);

(b) follows by taking the integration over each period and summing up the contribution of each term;

(c) follows by using the fact that \( x_r = g_\Delta(x) \) and \( g_\Delta \) is a deterministic mapping; and

(d) follows because for each period there is a unique \( x \) such that \( g_\Delta(x) = x_r \). Here \( x_i \) denotes the root of \( g_\Delta(x) - x_r = 0 \) in the \( i \)th period.
This yields
\[
\lim_{\Delta \to 0} p(y_1|x_r) = \lim_{\Delta \to 0} \frac{\sum_i p(y_1|x_i)p(x_i)}{\sum_i p(x_i)}
\]
\[
= \lim_{\Delta \to 0} \frac{\sum_i p(y_1|x_i)p(x_i)\Delta}{\sum_i p(x_i)\Delta}
\]
\[
= \left( a \right) \frac{\int p(y_1|x)p(x)dx}{\int p(x)dx}
\]
\[
= p(y_1),
\]
(4.17)
where (a) follows since the densities \( p(x) \) and \( p(y_1|x) \) are Riemann integrable. Thus
\[
h(Y_1|X_r) = -\int p(x_r) \left\{ \int p(y_1|x_r) \log p(y_1|x_r)dy_1 \right\} dx_r \rightarrow h(Y_1),
\]
(4.18)
since the inner integral approaches \( h(Y_1) \) as \( \Delta \to 0 \) (where changing the order of the limit and the integral is permitted due to the dominated convergence theorem). Consequently, as \( \Delta \to 0 \) the variable \( Y_1 \) does not depend on \( X_r \) and
\[
\lim_{\Delta \to 0} I_{\Delta}(X_r;Y_1) = 0.
\]
This completes the proof.  

Heuristically, knowing \( X_r \) gives less and less information about \( Y_1 = X + Z_1 \) for smaller \( \Delta \) because the possible \( X \) that fulfill \( X_r = g_\Delta(X) \) lie denser and denser, and as \( \Delta \to 0 \), “all \( X \)” are possible solutions to \( X_r = g_\Delta(X) \).

Lemma 4.3.3. Let \( X \) be a random variable with the density \( p(x) \) and \( T = s(\frac{X}{\Delta}) \) where \( s(x) \) is a periodic function with period one such that \( s(x) = 2x \) for \(-\frac{1}{2} \leq x < \frac{1}{2}\). If the density \( p(x) \) is Riemann integrable, the distribution of \( T \) approaches the uniform distribution as \( \Delta \) goes to zero. That is
\[
\lim_{\Delta \to 0} p(t) = \begin{cases} 
\frac{1}{2}, & -1 \leq t < 1 \\
0, & \text{otherwise}
\end{cases}
\]
(4.19)

Proof. Since \(-1 \leq T < 1\), we have \( p(t) = 0 \) if \( T \geq 1 \) or \( T < -1 \). Thus, consider the case when \(-1 \leq T < 1\). For \( t \in [-1,1) \) we have
\[
p(t) = \sum_i \frac{p(x_i)}{|s'(x_i)|}
\]
\[
= \frac{1}{2} \sum_i p(x_i)\Delta,
\]
(4.20)
where \( s'(x_i) \) denotes the derivative of \( s(\cdot) \) evaluated at \( x_i \), and \( x_i \) denotes the unique root of \( s(x) = t \) in the ith period. Since the density \( p(x) \) is Riemann integrable, we obtain
\[
\lim_{\Delta \to 0} p(t) = \frac{1}{2} \int p(x)dx = \frac{1}{2},
\]
(4.21)
for \( t \in [-1,1) \). This completes the proof.
Theorem 4.3.1. In the case of a noise-less source–relay link, memoryless relaying can achieve the capacity
\[ C = C(P_r) + C(P_s). \] (4.22)

Proof. Let \( X \sim N(0, P_s) \) and the relay mapping be
\[ f_\Delta(x) = \sqrt{P_r} F^{-1} \left( \frac{1}{2} s\left( \frac{x}{\Delta} \right) + \frac{1}{2} \right), \] (4.23)
where
- \( \Delta \) is a positive number;
- \( s(x) \) is a periodic function with period one (i.e., \( s(x+1) = s(x) \)) and \( s(x) = 2x \) for \( -\frac{1}{2} \leq x < \frac{1}{2} \); and
- \( F^{-1}(x) \) is the inverse of \( F(x) = 1 - Q(x) \) where \( Q(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt. \)

Using Lemma 4.3.3, the distribution of \( T = s\left( \frac{X}{\Delta} \right) \) approaches a uniform distribution over \(-1 \leq t < 1\) as \( \Delta \) goes to zero. Hence \( \frac{1}{2} s\left( \frac{X}{\Delta} \right) + \frac{1}{2} \) is uniformly distributed over \([0, 1)\). Now since \( F^{-1}(x) \) is the inverse of the cumulative distribution function (cdf) of the Gaussian distribution with variance one, the distribution of \( X_r \) tends to a Gaussian distribution with variance \( P_r \) as \( \Delta \) goes to zero. Using Proposition 4.3.2, as \( \Delta \to 0 \), we additionally have \( I_{\Delta}(X_r; Y_1) \to 0. \)

Thus using the mapping \( f_\Delta(x) \) at the relay, we can satisfy both conditions in Lemma 4.3.1 as \( \Delta \to 0 \). This proves that memoryless relaying achieves the capacity.

Remark 4.3.4. Using the mapping \( f_\Delta \) at the relay, we can achieve the rate
\[ R = I(X; Y_1) + C(P_r) \]
as long as the density of \( X \) is Riemann integrable. Thus the relay works at its best, for example, when the input distribution at the source is uniform.

Remark 4.3.5. The optimal mapping is not unique. To see this, let for example the relay function be
\[ g_\Delta(x) = \sqrt{P_r} F^{-1} \left( \frac{1}{2} r\left( \frac{x}{\Delta} \right) + \frac{1}{2} \right), \] (4.24)
where \( F^{-1} \) is defined in Theorem 4.3.1, \( \Delta \) is a positive number, and \( r(x) \) is a continuous triangular function with period one (i.e., \( r(x+1) = r(x) \)) and
\[ r(x) = \begin{cases} \frac{1}{2} + 2x, & \text{if } -\frac{1}{2} \leq x < 0 \\ -2x + \frac{1}{2}, & \text{if } 0 \leq x < \frac{1}{2} \end{cases} \] (4.25)

With the same analysis as in Theorem 4.3.1, one can verify that the conditions in Lemma 4.3.1 are fulfilled as \( \Delta \to 0 \).
4.3. A Semi-Deterministic Gaussian Relay Channel

4.3.2 Connection to High-Rate Uniform Quantization

It is well known that the error term in uniform scalar quantization is obtained by feeding the source sample through a sawtooth mapping. It is also well-known that the distribution of the error term gets more and more uniform as the quantization rate increases to infinity (see e.g. [WKL96]). Thus one can realize the optimal mapping $f_\Delta$ using a uniform quantizer followed by nonlinear shaping, as illustrated in Figure 4.4. That is, the optimal mapping $f_\Delta$ can be interpreted as analog coding of the quantization error at the relay. Hence the relay performs “quantize-and-forward,” however it is not the received and quantized samples that are being forwarded but a shaped version of the quantization errors (cf. innovative relaying in Section 2.4.3). This stands in a sharp contrast to linear relaying where the relay transmits a scaled version of the received signal.

4.3.3 Suboptimal Approaches

Linear Relaying: With linear relaying, the relay transmits $x_r = \sqrt{\frac{P_r}{P_s}} x$. The received signals at the destination are thus given by

$$y_1 = x + z_1$$
$$y_2 = \sqrt{\frac{P_r}{P_s}} x + z_2.$$

The achievable rate can be computed to be

$$R_{AF} = \max_{p(x)} I(X; Y_1, Y_2) = C (P_s + P_r),$$  \hspace{1cm} (4.26)

where the maximum is attained using a Gaussian input; i.e. $X \sim \mathcal{N}(0, P_s)$. Linear relaying can have an unbounded gap to the capacity. To see this, let $P_s = P_r = P$, then

$$C - R_{AF} = C \left( \frac{P^2}{2P + 1} \right) \longrightarrow +\infty \text{ as } P \longrightarrow +\infty.$$
Sawtooth relaying: Sawtooth relaying is introduced in [YKS08]. Following the same approach as that in Theorem 4.3.1, the rate achieved by a sawtooth mapping with a Gaussian codebook at the source is given by

\[ R_{\text{sawtooth}} = I(X; Y_1) + h(Y_2|Y_1) - h(Z_2) \rightarrow C(P_s) + h(X_r + Z_2) - h(Z_2). \]  

(4.27)

For a high-rate sawtooth mapping \( X_r \) is uniformly distributed with average power \( P_r \) (see also Lemma 4.3.3). The gap to the capacity then is

\[ \text{Gap} = C - R_{\text{sawtooth}} \]

\[ = \frac{1}{2} \log(2\pi e(P_r + 1)) - h(X_r + Z_2). \]  

(4.28)

This gap is numerically computed and is illustrated in Figure 4.5. One can also bound (4.28) using the entropy power inequality as follows.

\[ \text{Gap} \leq \frac{1}{2} \log(2\pi e(P_r + 1)) - \frac{1}{2} \log \left( 2^{2h(X_r)} + 2^{2h(Z_2)} \right) \]

\[ = \frac{1}{2} \log \left( \frac{1 + P_r}{1 + \frac{2}{\pi} P_r} \right). \]  

(4.29)

This upper bound is also plotted in Figure 4.5. The gap increases with \( P_r \) and is less than \( \frac{1}{2} \log \left( \frac{2}{\pi} \right) \approx 0.2546 \). We have also evaluated the achievable rate for the sawtooth mapping with \( P_r = 10 \) and 20 dB using Monte Carlo simulation. The gap to the cutset bound is illustrated in Figure 4.5 using bold circles. Sawtooth relaying is a sub-optimal approach since it generates a uniformly distributed output at the relay. However, in contrast to linear relaying, the gap to the optimal performance is bounded.

Laplace relaying: An alternative approach is to use the logarithmic mapping

\[ x_r = -\sqrt{\frac{P_r}{2}} \text{sign}(u) \log_e(1 - 2|u|), \]

where \( -\frac{1}{2} \leq u \leq \frac{1}{2} \) and \( u \) is generated using a sawtooth mapping. Thus \( X_r \) is distributed according to a Laplace distribution. Using the entropy power inequality, the gap between the achievable rate of this strategy and the capacity is bounded as

\[ \text{Gap} = C - R_{\text{Laplace}} \leq \frac{1}{2} \log \left( \frac{1 + P_r}{1 + \frac{2}{\pi} P_r} \right). \]  

(4.30)

In this case, the gap increases with \( P_r \) and is less than \( \frac{1}{2} \log \left( \frac{2}{\pi} \right) \approx 0.1044 \). We see that this approach is superior to triangular sawtooth relaying at moderate to high SNRs.
4.3. A Semi-Deterministic Gaussian Relay Channel

4.3.4 Numerical Examples

In the following we assume that the source uses a Gaussian codebook. A plot of $f_{x_1}^*(x)$ for $\Delta = 1$ and $P_r = 1$ is shown in Figure 4.6. Figure 4.7 shows the achievable rates for memoryless relaying as a function of $\Delta^{-1}$ when $P_s = P_r = 10$. The rates are computed based on Monte Carlo simulation using $N = 5 \times 10^6$ samples. We see that as $\Delta$ decreases the rate increases and reaches the upper bound. Figure 4.8 illustrates the resulting empirical distribution of $X_r$ for $\Delta = 1$. We see that the empirical distribution matches very well with a Gaussian distribution. For this particular sample set we have $E[X X_r] \approx \frac{1}{N} \sum_{i=1}^{N} x_i x_{ri} = 0.00038$. Figure 4.9 illustrates the contours of the joint empirical distribution of $(Y_1, Y_r)$ for $\Delta = 1$. We see that the contours are circles and they match very well with a bivariate i.i.d Gaussian distribution.
Figure 4.6: The optimal relay function for $\Delta = 1$ and $P_r = 1$.

Figure 4.10 shows the achievable rate of memoryless relaying using $f_\Delta^*$ with $\Delta = 1$ when $P_s = P_r = P$. We see that the rates computed using Monte Carlo simulation match very well with the upper bound as predicted in Theorem 4.3.1. The rate achieved by linear relaying is also plotted. However, in this scenario we see that the linear strategy performs poorly.

We finally remark that in addition to memoryless relaying, both CF and DF can achieve the capacity in the case of a perfect source–relay link [GMZ06]. However, CF and DF require block-wise encoding and decoding at the relay in contrast to the proposed instantaneous relaying scheme.
Figure 4.7: Achievable rate of memoryless relaying using $f_\Delta^*$ as a function of $\Delta$ when $X \sim \mathcal{N}(0, P_s)$ and $P_s = P_r = 10$. 
Figure 4.8: The empirical distribution of $X_r$ when $P_r = 10$ and $\Delta = 1$. The plot with solid line shows the zero-mean Gaussian distribution with variance one.

Figure 4.9: Contours of the empirical distribution of $(Y_1, Y_2)$ when $X \sim \mathcal{N}(0, P_s)$, $P_r = P_r = 10$ and $\Delta = 1$. The solid lines illustrate the contours of an uncorrelated bivariate Gaussian distribution with variance one.
Figure 4.10: Achievable rates of instantaneous relaying with $X \sim \mathcal{N}(0, P)$ using a linear mapping and optimal mapping $f^{*}_\Delta$ with $\Delta = 1$. 
4.4 A Gaussian Relay Channel with Side Information

Figure 4.11 shows a block diagram illustrating the Gaussian relay channel that we consider in this section. The received signal at the relay is given by

\[ y_r = ax + z_r \]  \hspace{1cm} (4.31)

where \( x \) is the transmitted symbol from the source, \( a \) is the channel gain between the source and the relay, and \( Z_r \sim \mathcal{N}(0,1) \) is zero-mean additive white Gaussian noise with unit variance. We assume that the destination knows \( z_r \). The proposed scenario can be used to model, e.g., a situation in which the destination simultaneously communicates with another user in the same frequency band as the one used by the source. Consequently, the destination creates a known interference term on the source–relay link.

The received signal from the source at the destination is given by

\[ y_1 = x + z_1, \]  \hspace{1cm} (4.32)

where \( Z_1 \sim \mathcal{N}(0,1) \). Note that the channel gain of the source–destination link is normalized to one. The received signal from the relay at the destination is given by

\[ y_2 = bx_r + z_2, \]  \hspace{1cm} (4.33)

where \( x_r \) is the symbol transmitted from the relay, \( b \) is the channel gain between the relay and the destination, and \( Z_2 \sim \mathcal{N}(0,1) \). We assume that \( Z_r, Z_1, \) and \( Z_2 \) are mutually independent. We further assume average power constraints at the source and the relay, i.e., \( E[X^2] \leq P_s \) and \( E[X_r^2] \leq P_r \).
4.4.1 Instantaneous Relaying

Here we consider, as in Section 4.2, that upon receiving $y_r$, the relay transmits the signal $x_r$ which is given by

$$x_r = f(y_r), \quad (4.34)$$

where $f(\cdot)$ is a one-dimensional, symbol-by-symbol, mapping used at the relay. Thus the received signal from the relay at the destination is given by

$$y_2 = bx_r + z_2 = bf(ax + z_r) + z_2. \quad (4.35)$$

**Proposition 4.4.1.** Memoryless relaying can achieve the capacity of the Gaussian relay channel shown in Figure 4.11 when the destination knows $Z_r$ and the capacity is given by

$$C = C(b^2 P_r) + C(P_s). \quad (4.36)$$

**Proof.** The proof closely follows the approach presented in Section 4.3. For a fixed relay mapping the following rate is achievable

$$I(X; Y_1, Y_2|Z_r) = I(X; Y_1|Z_r) + I(X; Y_2|Y_1, Z_r)$$

$$= I(X; Y_1|Z_r) + h(Y_2|Y_1, Z_r) - h(Y_2|X, Y_1, Z_r)$$

$$\overset{(a)}{=} I(X; Y_1|Z_r) + h(Y_2|Y_1, Z_r) - h(Z_2|X, Y_1, Z_r)$$

$$\overset{(b)}{=} I(X; Y_1) + h(Y_2|Y_1, Z_r) - h(Z_2)$$

$$= I(X; Y_1) + h(bX_r + Z_2|Y_1, Z_r) - h(Z_2)$$

$$\overset{(c)}{\geq} \frac{1}{2} \log \left(2^{2h(X_r|Y_1, Z_r)} + 2^{2h(Z_2|Y_1, Z_r)} \right) + I(X; Y_1) - h(Z_2)$$

$$= \frac{1}{2} \log \left(2^{2h(X_r|Y_1, Z_r) + \log(b)} + 2^{2h(Z_2)} \right) +$$

$$I(X; Y_1) - h(Z_2)$$

$$= \frac{1}{2} \log \left(2^{2h(X_r) + \log(b)} + I(X_r; Y_1, Z_r) + 2^{2h(Z_2)} \right) +$$

$$I(X; Y_1) - h(Z_2) \quad (4.37)$$

where

(a) holds since $Y_2 = bf(ax + z_r) + Z_2$;

(b) holds because $Z_r$ is independent of $(X, Y_1)$, and $Z_2$ is independent of $(X, Y_1, Z_r)$; and

(c) follows by the entropy power inequality.
Now let \( X \sim \mathcal{N}(0, P_x) \) and the relay mapping be \( f_\Delta^* \) as given in Theorem 4.3.1. Next consider

\[
p(y_1, z_r | x_r) = \int p(y_1, z_r | y_r, x_r)p(y_r | x_r)dy_r
\]

\[
= \int p(y_1, z_r | y_r)p(y_r)p(x_r | y_r)dy_r
\]

\[
= \int p(y_1, z_r | y_r)p(y_r)\frac{p(x_r | y_r)}{\int p(y_r)p(x_r | y_r)dy_r}dy_r
\]

\[
= \sum_i \int_{x_i} p(y_1, z_r | y_r)p(y_r)\sum_j f_{\Delta}(y_r)dy_r
\]

\[
= \sum_i p(y_1, z_r | y_r)p(y_r)
\]

\[
\sum_i p(y_r)
\]

where

(a) follows by applying Bayes’ rule and using the fact that \( p(y_1, z_r | y_r, x_r) = p(y_1, z_r | y_r) \);

(b) follows by taking the integration over each period and summing up the contribution of each term;

(c) follows by using the fact that \( x_r = f_\Delta^*(y_r) \) and \( f_\Delta^* \) is a deterministic mapping and using the Dirac delta function; and

(d) follows because for each period there is a unique \( y_r \) such that \( x_r = f_\Delta^*(y_r) \).

Here \( y_{ri} \) denotes the root of \( f_\Delta^*(y_r) - x_r = 0 \) in the \( i \)th period.

This yields

\[
\lim_{\Delta \to 0} p(y_1, z_r | x_r) = \lim_{\Delta \to 0} \frac{\sum_i p(y_1, z_r | y_r)p(y_r)}{\sum_i p(y_r)}
\]

\[
= \lim_{\Delta \to 0} \frac{\sum_i p(y_1, z_r | y_r)p(y_r)\Delta}{\sum_i p(y_r)\Delta}
\]

\[
= \int p(y_1, z_r | y_r)p(y_r)dy_r
\]

\[
\int p(y_r)dy_r
\]

\[
p(y_1, z_r)
\]

\[
(4.39)
\]

where (a) holds since \( p(y_r) \) and \( p(y_1, z_r | y_r) \) are Riemann integrable. Thus \( h(Y_1, Z_r | X_r) \to h(Y_1, Z_r) \) as \( \Delta \to 0 \) (invoking the dominated convergence theorem). Now using Lemma 4.3.3, the distribution of \( X_r \) approaches a Gaussian distribution.
as $\Delta$ goes to zero. Hence, by combining the above results with (4.37), we conclude that $R = C(b^2P_r) + C(P_s)$ is achievable. This completes the proof since $R$ coincides with the upper bound.

### Encoding and Decoding

Knowing the noise-term $Z_r$ at the destination is crucial to achieve the capacity using memoryless relaying. To clarify this, we review the encoding and the decoding procedures that achieve the capacity.

**Codebook Generation:**

- Generate i.i.d sequences of length $n$ for each $m \in [1 : 2^nR]$ where each sequence is distributed according to $\prod_{i=1}^{n} p(x)$ and $X \sim \mathcal{N}(0, P_s)$. Label them $X^n(m).

**Encoding:**

- The source for a given message $m$ transmits the associated codeword $X^n(m)$.
- The relay using the symbol-by-symbol mapping $f^*_\Delta$ transmits $X^n_r(m) = (f^*_\Delta(aX_1(m)+Z_r,1), f^*_\Delta(aX_2(m)+Z_r,2), \cdots, f^*_\Delta(aX_n(m)+Z_r,n))$ in $n$ channel uses.

**Decoding:**

- Because the destination knows i) the codebook at the source, ii) the relay mapping $f^*_\Delta$, and iii) the noise sequence on the source–relay link, it consequently knows the “codebook” used at the relay. By the codebook at the relay, we mean the collection of possible $\{X_r,i\}_{i=1}^{n}$ in each block, which in fact varies with the noise on the source–relay link. Thus, upon receiving $Y^n_1$ and $Y^n_2$, the destination declares that the message $\hat{m} \in [1 : 2^nR]$ is transmitted if it is a unique index such that $(X^n(\hat{m}), X^n_r(\hat{m}))$ is jointly typical with $(Y^n_1, Y^n_2)$. Otherwise, it declares an error. In the following, $I_\Delta$ and $h_\Delta$ respectively denote mutual information and entropy of random variables when the relay employs $f^*_\Delta$. Thus, the decoding is successful if

$$R < I_\Delta(X, X_r; Y_1, Y_2 | Z_r)$$

$$\overset{(a)}{=} I_\Delta(X; Y_1, Y_2 | Z_r) + I_\Delta(X_r; Y_1, Y_2 | X, Z_r)$$

$$\overset{(b)}{=} I_\Delta(X; Y_1, Y_2 | Z_r)$$

$$\overset{(c)}{=} I_\Delta(X; Y_1 | Z_r) + I_\Delta(X; Y_2 | Y_1, Z_r)$$

$$\overset{(d)}{=} I(X; Y_1) + I_\Delta(X; Y_2 | Y_1, Z_r)$$

$$= I(X; Y_1) + h_\Delta(Y_2 | Y_1, Z_r) - h_\Delta(Y_2 | Y_1, Z_r, X)$$

$$\overset{(e)}{=} I(X; Y_1) + h_\Delta(Y_2 | Y_1, Z_r) - h_\Delta(Y_2 | Y_1, Z_r, X, X_r)$$
Orthogonal Gaussian Relay Channels

\[ I(X; Y_1) + h_\Delta(Y_2|Y_1, Z_r) - h(Z_2) \]
\[ C(P_s) + h_\Delta(Y_2|Y_1, Z_r) - \frac{1}{2} \log(2\pi e) \]
\[ C(P_s) + h_\Delta(Y_2) - \frac{1}{2} \log(2\pi e) \text{ as } \Delta \to 0 \]
\[ C(P_s) + C(b^2 P_r) \text{ as } \Delta \to 0, \]  \hspace{1cm} (4.40)

where

(a) follows by applying the chain rule;
(b) follows by the fact that \( X_r = f^*_\Delta(aX + Z_r) \);
(c) the same as (a);
(d) holds since \( X \) and \( Y_1 \) are independent of \( Z_r \) and \( f^*_\Delta \);
(e) the same as (b);
(f) follows because \( Y_2 = X_r + Z_2 \) and \( Z_2 \) is independent of \( (Y_1, Z_r, X, f^*_\Delta) \);
(g) follows since the \( X \sim \mathcal{N}(0, P_s) \) and \( Z_2 \sim \mathcal{N}(0, 1) \);
(h) holds since the relay employs \( f^*_\Delta \) with an arbitrarily small \( \Delta \); and
(i) holds since \( Y_2 \sim \mathcal{N}(0, P_r + 1) \) for an arbitrarily small \( \Delta \).

To summarize, the destination sees an augmented codeword \((x^n(m), x^n_r(m))\) where the second half of the codeword is generated using the first half and knowing \( f^*_\Delta \) and \( z^n_r \) at the destination. This code construction is reminiscent of the idea of the parity bits in a systematic linear code. Here, the relay actually generates analog parity constraints.

**Linear Relaying**

With linear relaying, the relay uses an instantaneous linear mapping, i.e.,

\[ f(y_r) = \sqrt{\frac{P_r}{\mathbb{E}[Y_r^2]}} y_r. \]

Thus, the received signals from the source and the relay at the destination are given by

\[ y_1 = x + z_1 \]
\[ y_2 = \beta ax + \beta z_r + z_2, \]
where $\beta = \sqrt{\frac{b^2 P_s}{a^2 P_s + 1}}$ is chosen to ensure the power constraint at the relay. The achievable rate of AF can be computed as follows

\[
R_{AF} = I(X; Y_1, Y_2 | Z_r) = h(Y_1, Y_2 | Z_r) - h(Y_1, Y_2, X, Z_r) \\
= (a) h(Y_1, \tilde{Y}_2) - h(Z_1, Z_2) \\
\leq \frac{1}{2} \log \left( \frac{(2\pi e)^2}{{\mathbb{E}}[Y_1^2]} \right) - h(Z_1, Z_2) \\
= \frac{1}{2} \log \left( \frac{(2\pi e)^2}{{\mathbb{E}}[Y_2 Y_1]} \right) - h(Z_1, Z_2) \\
= \frac{1}{2} \log \left( \frac{(2\pi e)^2}{1} \right) - h(Z_1, Z_2) \\
= C \left( P_s + \frac{a^2 b^2 P_s P_r}{1 + a^2 P_s} \right),
\]

(4.41)

where $(a)$ follows by defining $\tilde{Y}_2 := Y_2 - \beta Z_r = a \beta X + Z_2$. The bound is achieved if $(Y_1, \tilde{Y}_2)$ are jointly Gaussian, which results from choosing $X \sim \mathcal{N}(0, P_s)$. That is, using a Gaussian codebook at the source maximizes the achievable rate when the relay employs a linear strategy.

Also in the scenario considered here, linear relaying can have an unbounded gap to the capacity. To see this, let $P_s = P_r = P$, and $a = b = 1$, then

\[
C - R_{AF} = C \left( \frac{P^3 + P^2 + P}{2P^2 + 2P + 1} \right) \to +\infty \quad \text{as} \quad P \to +\infty.
\]

The main drawback of the linear relay is that it forwards the noise term to the destination. Even though the forwarded noise can be cancelled at the destination, this drains power at the relay. Consequently, a linear strategy is not optimal even for the cases when $|b| \to +\infty$ or $N_2 \to 0$.

**Remark 4.4.2.** For a Gaussian codebook at the source, the capacity of instantaneous estimate-and-forward (EF) coincides with that of AF, since then

\[
x_r = \lambda {\mathbb{E}}[X | y_r] = \lambda \frac{a P_s}{a^2 P_s + 1} y_r,
\]

where $\lambda$ is a power normalization factor.

### 4.4.2 CF Relaying

**Proposition 4.4.3.** CF relaying achieves the capacity of the Gaussian relay channel in Figure 4.11 when the destination knows $Z_r$ and the capacity is given by (4.36).
Proof. Note that the received signal at the relay $y_r$ is a deterministic function of the channel input $x$ and the output at the destination (i.e., $z_r$). Hence from the result in Section 2.3.6, CF is capacity achieving.

For completeness, we also compute the rate using the achievable rate in (2.23) in Section 2.3.6. Recall that the rate
\[
\sup_{p(y_r|x_r)p(x_r)p(x)} I(X; Y_1, \hat{Y}_r | Z_r),
\]
subject to $I(Y_r; \hat{Y}_r | X_r, Y_1, Z_r) \leq I(X_r; Y_2)$ is achievable. Now let
\[
\begin{bmatrix}
X \\
X_r
\end{bmatrix} \sim \mathcal{N}
\begin{bmatrix}
P_s & 0 \\
0 & P_r
\end{bmatrix},
\]
and define $\hat{Y}_r := Y_r + Z_q$ where $Z_q \sim \mathcal{N}(0, N_q)$ and is independent of other random variables. Using these choices of random variables, we obtain
\[
I(Y_r; \hat{Y}_r | X_r, Y_1, Z_r) = h(Y_r | Y_1, Z_r) - h(Y_r | Y_r, Y_1, Z_r) = h(aX + Z_q | X + Z_1) - h(Z_q) = C \left( \frac{a^2 P_s}{N_q(P_s + 1)} \right);
\]
\[
I(X; Y_1, \hat{Y}_r | Z_r) = h(X + Z_1, aX + Z_q) - h(Z_1, Z_q) = C \left( P_s + \frac{a^2 P_s}{N_q} \right);
\]
\[
I(X_r; Y_2) = C(b^2 P_r).
\]
Choosing the optimal $N_q^*$, we find $N_q^* = \frac{a^2 P_s}{b^2 P_r (P_s + 1)}$ and hence obtain
\[
R_{\text{CF}} = I(X; Y_1, \hat{Y}_r | Z_r, N_q = N_q^*) = C(P_s + b^2 P_r P_s + b^2 P_r) = C(P_s) + C(b^2 P_r). \tag{4.42}
\]
This completes the proof. \hfill \Box

### 4.4.3 DF Relaying

The DF strategy does not achieve the capacity in this case, since the relay has no access to the side information. Under the constraint of successful decoding at the relay, knowing the side information at the destination does not increase the capacity. Thus the rate achieved by DF [CG79, GMZ06] is given as
\[
R_{\text{DF}} = \max \left\{ C(P_s), \min \left\{ C(a^2 P_s), C(P_s) + C(b^2 P_r) \right\} \right\}, \tag{4.43}
\]
which is achieved by choosing
\[
\begin{bmatrix}
X \\
X_r
\end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix}
P_s & 0 \\
0 & P_r
\end{bmatrix}\right).
\]

For given \((P_s, P_r, a, b)\), DF is optimal if \(|a| \geq \sqrt{1 + b^2 P_r (1 + P_s^{-1})}\). However, DF can have an unbounded gap to the capacity. To see this, let \(|a| \leq 1\) and \(b \neq 0\), then
\[
C - R_{DF} = C (b^2 P_r) \longrightarrow +\infty \quad \text{as} \quad P_r \longrightarrow +\infty.
\]

The main shortcoming of DF is the requirement that the relay has to recover the transmitted message. However, unlike DF, CF as well as instantaneous relaying can efficiently utilize the side information in a distributed manner and leave the detection of the transmitted message to the destination.

### 4.5 The General Orthogonal Gaussian Relay Channel

In this section we consider a Gaussian relay channel model similar to that in Section 4.4. However, we assume that the destination does not know \(z_r\), i.e. the noise component on the source–relay link. See Figure 4.12 for an illustration.

From Figure 4.12, we can see that as the magnitude of \(a\) increases, the channel model in Figure 4.12 can be approximated as that in Section 4.3. Or equivalently as the variance of the noise on the source–relay link \((N_r)\) vanishes, the channel model in Figure 4.12 approaches those in Sections 4.3 and 4.4. Earlier in the chapter, we have established that linear relaying is a suboptimal strategy for the asymptotic cases considered in Sections 4.3 and 4.4. One can expect this to hold true also for finite values of \(a\) and \(N_r\).

We first briefly discuss some basic bounds on the capacity and then introduce our proposed instantaneous relaying scheme. We finally end this section by some numerical examples.

![Figure 4.12: Block diagram of the general Gaussian relay channel with orthogonal receive components.](image)
4.5.1 Basic Bounds

Upper Bound

**Proposition 4.5.1.** For the relay channel with instantaneous relaying and orthogonal receive components we have

\[ C \leq C \left( \min \{ P_s + b^2 P_r + b^2 P_s P_r, (1 + a^2)P_s \} \right). \]  

(4.44)

**Proof.** The proof follows from Proposition 4.2.1. \qed

Lower Bounds

For the sake of comparison, in this section, we briefly review well-known relaying protocols studied in previous work.

**Instantaneous Linear Relaying (AF):** With AF, as before, the relay uses an instantaneous linear mapping, i.e.,

\[ f(y_r) = \sqrt{\frac{P_r}{E[Y_r^2]}} y_r. \]

Thus, the received signals from the source and the relay at the destination are given by

\[ y_1 = x + z_1 \]
\[ y_2 = \beta ax + \beta z_r + z_2, \]

where \( \beta = \sqrt{\frac{b^2 P_r}{a^2 P_s + 1}}. \) The achievable rate of AF computed to be

\[ R_{AF} = \max_{p(x)} I(X;Y_1,Y_2) = C \left( P_s + \frac{a^2 b^2 P_s P_r}{1 + \frac{a^2}{a^2 P_s + b^2 P_r}} \right), \]

(4.45)

where the maximum is achieved by choosing a Gaussian input at the source. (See Section 4.4.1 and also [GMZ06, NBK04].)

**Remark 4.5.2.** When \( b^2 P_r \gg \max\{a^2 P_s, 1\}, \) the rate given in (4.45) simplifies to

\[ R_{AF} \approx C \left( (1 + a^2)P_s \right), \]

which coincides with the capacity upper bound given in (4.44). Therefore, AF operates close to the capacity when the relay–destination link is very strong relative to other links (e.g., when the relay is located in the proximity of the destination).

**Decode-and-forward (DF):** The rate achieved by DF is given in (4.43). This rate is limited by the SNR of the source–relay link. However, DF is optimal when \( |a| \geq \sqrt{1 + b^2 P_r(1 + P_s^{-1})}. \)
4.5. The General Orthogonal Gaussian Relay Channel

Compress-and-forward (CF): Employing a Gaussian codebook at the source and the relay and Gaussian Wyner–Ziv quantization\(^1\) similar to that in Proposition 4.4.3 at the relay results in the achievable rate [CG79, GMZ06, KGG05, HMZ06]

\[
R_{CF} = C\left(P_s + \frac{a^2b^2P_rP_s(P_s + 1)}{1 + (1 + a^2)P_s + b^2P_r(1 + P_s)}\right). \tag{4.46}
\]

Comparing (4.45) with (4.46), we see that \(R_{CF} \geq R_{AF}\) and \(R_{AF} \rightarrow R_{CF}\) as \(|b| \rightarrow \infty\). Note that CF asymptotically achieves the capacity upper bound given by (4.44) when \(|a| \rightarrow \infty\) or \(|b| \rightarrow \infty\).

4.5.2 Proposed Scheme: Piecewise Linear Relaying

We next introduce a novel relaying protocol based on a piecewise linear (PL) mapping. In the following, we assume that the input of the relay channel is Gaussian with variance \(P_s\), i.e. \(X \sim \mathcal{N}(0, P_s)\).\(^2\) Note that the rate obtained by the proposed new scheme serves as a lower bound on the overall capacity of instantaneous relaying.

Fully linear (AF) relaying is an efficient solution to the problem of instantaneous relaying when the relay–destination link is stronger than the other links, e.g., when the relay is located close to the destination (see also Remark 4.5.2). However, AF relaying results in a significant loss when the source–relay link is strong (see Section 4.3). Therefore, we modify linear relaying such that the relay is adapted to take into consideration the quality of the source–relay link. The proposed piecewise linear mapping is defined according to

\[
f_{PL}(y_r) := \begin{cases} 
    a_1y_r + b_1 & \text{if } y_r \in T_1 = (t_1, t_2] \\
    a_2y_r + b_2 & \text{if } y_r \in T_2 = (t_2, t_3] \\
    \vdots & \\
    a_iy_r + b_i & \text{if } y_r \in T_i = (t_i, t_{i+1}] \\
    \vdots & \\
    a_Ny_r + b_N & \text{if } y_r \in T_N = (t_N, t_{N+1})
\end{cases} \tag{4.47}
\]

where \(t_1 < t_2 < \cdots < t_{N+1}\). Thus, the proposed PL relaying mapping has two sets of parameters:

- \(T = \{T_i\}_{i=1}^N\) where \(\bigcup_i T_i = \mathbb{R}\) and \(T_i \cap T_j = \emptyset\) for \(i \neq j\);
- \(A = \{a_i\}_{i=1}^N\) and \(B = \{b_i\}_{i=1}^N\) such that \(f_{PL}(y_r) = a_iy_r + b_i\) when \(y_r \in T_i\).

The parametric PL mapping therefore provides flexibility to be adapted to the quality of the channel links by optimization of the parameters. It is worth noting

\(^1\)Gaussian quantization is not necessarily optimal. For example, see [DS08].
\(^2\)The proposed scheme is not limited to any particular choice of the channel input distribution.
that conventional instantaneous linear relaying (i.e., AF) is a special case of the proposed PL mapping by choosing \(a_1 = a_2 = \cdots = a_N\) and \(b_1 = b_2 = \cdots = b_N = 0\). Therefore, the achievable rate of PL relaying is lower bounded by that of AF, provided that \(\{a_i\}_{i=1}^N\) and \(\{b_i\}_{i=1}^N\) are optimally chosen. Additionally, the proposed PL mapping can approximate any nonlinear mapping by choosing \(N, T, A\) and \(B\) appropriately. Note that it is crucial to include the bias terms \(b_i\) in (4.47). This is because when \(b_i = 0\), the relay mapping is forced to be piecewise linear with segments that pass through the origin. In general, however, it is not efficient to approximate an arbitrary function using the mapping in (4.47) with \(b_i = 0\). For example, the mapping in Figure 4.6, can be more efficiently approximated when \(b_i \neq 0\).

**Mapping Optimization**

Fixing the number of intervals \(N\) (a design parameter), the PL relaying function given in (4.47) can be optimized according to

\[
\arg \max_{\{t_i\}_{i=1}^{N+1}, \{a_i\}_{i=1}^N, \{b_i\}_{i=1}^N} I(X; Y_1, Y_2),
\]

subject to the power constraint

\[
E[f_{PL}(Y_r)] = \sum_{i=1}^N \int_{T_i} p(y_r)(a_i y_r + b_i)^2 dy_r = P_r.
\]

Note that the optimized mapping will in general depend on the “design point” defined by \((a, b, P_s, P_r)\). In solving the problem in (4.48), consider that

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

\[
(a) = I(X; Y_1) + I(X; Y_2|Y_1)
\]

\[
= I(X; Y_1) + h(Y_2|Y_1) - h(Y_2|X, Y_1)
\]

\[
(b) = I(X; Y_1) + h(Y_2|Y_1) - h(Y_2|X, Z_1)
\]

\[
(c) = I(X; Y_1) + h(Y_2|Y_1) - h(Y_2|X)
\]

\[
(d) = C(P_s) - E[\log p(y_2|y_1)] + E[\log p(y_2|x)],
\]

where

(a) follows from the chain rule;

(b) since \(Y_1 = X + Z_1\);

(c) follows from the fact that \(Z_1\) is independent of \((X, Y_2)\); and

(d) holds by the assumption that \(X \sim \mathcal{N}(0, P_s)\).
It therefore suffices to find the conditional densities \( p(y_2|x) \) and \( p(y_2|y_1) \) to numerically compute \( I(X; Y_1, Y_2) \). The following lemma gives closed form expressions for \( p(y_2|x) \) and \( p(y_2|y_1) \) in the case of the general mapping defined in (4.47). The proof (omitted) is a straightforward exercise.

**Lemma 4.5.3.** Let the source variable be Gaussian (i.e., \( X \sim N(0, P_s) \)), and let the relay employ the instantaneous mapping given in (4.47). Then,

- the conditional pdf of \( y_2 \) given \( x \) is
  \[
p(y_2|x) = \sum_{i=1}^{N} \frac{\gamma_i}{\sqrt{2\pi}} \exp \left( -\frac{a^2x^2 + (y_2 - bb_i)^2 - \gamma_i^2 \beta_i^2}{2} \right) \times \left[ Q \left( \frac{t_i - \gamma_i^2 \beta_i}{\gamma_i} \right) - Q \left( \frac{t_{i+1} - \gamma_i^2 \beta_i}{\gamma_i} \right) \right],
  \]
  where
  \[
  \gamma_i^2 := \frac{1}{1 + a_i^2 b^2},
  \]
  \[
  \beta_i := ax + a_i b (y_2 - bb_i),
  \]
  \[
  Q(\alpha) := \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} \exp \left( -\frac{1}{2} s^2 \right) ds;
  \]

- the conditional pdf of \( y_2 \) given \( y_1 \) is
  \[
p(y_2|y_1) = \sum_{i=1}^{N} \frac{\gamma_i}{\sqrt{2\pi}\xi^2} \exp \left( -\frac{k^2 y_1^2 + (y_2 - bb_i)^2 - \gamma_i^2 \beta_i^2}{2} \right) \times \left[ Q \left( \frac{t_i - \gamma_i^2 \beta_i}{\gamma_i} \right) - Q \left( \frac{t_{i+1} - \gamma_i^2 \beta_i}{\gamma_i} \right) \right],
  \]
  where
  \[
  \mu := \mathbb{E}[Y_1|y_1] = \frac{a P_s}{P_s + 1} y_1 := k y_1,
  \]
  \[
  \xi^2 := \mathbb{E} \left[ (Y_r - \mathbb{E}[Y_r|Y_1])^2 \right] = \frac{1 + (1 + a^2) P_s}{1 + P_s},
  \]
  \[
  \gamma_i^2 := \frac{1}{a_i^2 b^2 + \xi^2},
  \]
  \[
  \beta_i := \frac{k}{\xi^2} y_1 + a_i b (y_2 - bb_i).
  \]

Using Lemma 4.5.3, we can numerically compute \( I(X; Y_1, Y_2) \) and then optimize the mapping for a certain number of parameters. This optimization problem is in general non-convex. We therefore conduct the optimization based on an iterative grid search method as follows.
• **Initialization:** Fix the number of intervals \( N = N_{\text{int}} \) (a prescribed value) and the sets \( A = \{a_1, \ldots, a_M\} \) and \( B = \{b_1, \ldots, b_M\} \) defining a two dimensional grid with \( M^2 \) points. Next initialize the relay function with a linear mapping, i.e., \( a_i = a \) and \( b_i = 0 \) where \( a \) is chosen to ensure the power constraint at the relay.

• **Step 1:**

  1.1 *Iteration over intervals:* For each interval, using (4.50) and Lemma 4.5.3, compute the achievable rate \( R_{\text{PL}} \) for all \( b_i \in B \) and \( a_i \in A \) subject to the power constraint given in (4.49). Then find the pair \((a_i, b_i)\) that gives the highest rate and update the mapping.

  1.2 *Termination:* Repeat Step 1.1 until the number of iterations exceeds \( K \) (a given number).

• **Step 2:**

  2.1 *Increasing the number of intervals:* Divide each interior region into two equal regions and repeat Step 1.

  2.2 *Termination:* Repeat Step 2.1 until there is no significant change in the rate.

4.5.3 Numerical Examples and Discussion

In this section we present numerical results to compare the performance of different forwarding strategies.

In order to run the design algorithm we choose the following parameters. We divide the interval \((-3.5\sqrt{P_r}, 3.5\sqrt{P_r})\) uniformly into 50 points in order to obtain the sets \( A \) and \( B \). We further choose \( N_{\text{int}} = 5 \) and \( K = 10 \). We additionally fix the two outermost regions \( T_1 = (-\infty, -3.5\sqrt{P_r}) \) and \( T_N = (3.5\sqrt{P_r}, +\infty) \), and we terminate the design if the change in the resulting achieved rate is less than 1%. We have observed that in most cases the algorithm terminates with \( N < 100 \).

In the following, we set \( P_s = P_r = 10 \) and present numerical results for three different cases of the channel gains, i.e. \( a \) and \( b \).

*Case 1:* Figure 4.13 shows the computed achievable rates of different forwarding strategies when \( b = 1 \), as a function of the source–relay link gain \( a \). Figure 4.14 shows the associated optimized relaying functions. In Figure 4.14 the input and output are normalized by the corresponding standard deviations (the probability that an input sample is in the illustrated region is higher than 0.999). From Figure 4.13, we see that AF is incapable of utilizing the increasing quality of the source–relay link and the achievable rate saturates very quickly as \( a \) increases. However, PL relaying compensates for this and performs very close to CF relaying. For large values of \( a \), DF coincides with the upper bound, however for small \( a \), PL linear relaying improves on both AF and DF.
4.6 Summary and Concluding Remarks

Case 2: Figure 4.15 illustrates the achievable rates of different forwarding strategies when \( a = 1 \), now instead as a function of the relay–destination link gain \( b \), and Figure 4.16 shows the associated optimized functions. From Figure 4.15, we see that DF relaying provides a constant rate and performs worst among all relaying strategies. This is due to the weak quality of the source–relay link in this example. On the other hand, PL relaying improves on AF and achieves rates close to those of CF.

Case 3: Finally, Figure 4.17 shows the achievable rates of different forwarding strategies \( b = 0.1 \) as a function of the source–relay link gain \( a \). Figure 4.18 illustrates the associated optimized functions. From Figure 4.15, we see that AF almost provides a constant rate and performs worst among all relaying strategies. The poor performance of AF is due to the weak quality of the relay–destination link. In this case, DF achieves the capacity for large \( a \). PL relaying outperforms AF and performs close to CF.

From the presented numerical examples, we stress the following points:

- Optimized PL relaying always outperforms AF and improves on DF when the quality of the source–relay link is low. Additionally, PL relaying provides rates close to those of CF but at a much lower complexity and delay, since its operation is symbol-wise and it does not require any encoding at the relay.

- Optimized PL functions are not one-to-one mappings, i.e., the same output value is used for different received values. This in general results in ambiguity, however in our case the destination has access to the side information \( Y_1 \) (i.e., received signal from the source at the destination) and decoding is performed by joint utilization of \( Y_1 \) and \( Y_2 \). Generally, these mappings therefore enjoy better power efficiency than AF relaying. The proposed PL relaying can be hence considered as a one-dimensional analogue of CF, where the side information in the source–destination link is efficiently utilized to save power in the relay transmission.

4.6 Summary and Concluding Remarks

We studied the design of instantaneous relaying strategies for Gaussian relay channels with orthogonal receive components. We showed that when either the source–relay link is perfect or the destination knows the noise on the source–relay link, instantaneous relaying can achieve the capacity. In these cases, the rates achieved by conventional AF can have a unbounded gap to the capacity. We also proposed a novel instantaneous forwarding strategy: piecewise linear (PL) relaying for the general Gaussian relay channel. We presented some interesting examples and the associated performance of PL relaying. We demonstrated that the proposed scheme can improve on DF and AF relaying. The achievable rate of PL relaying however maintains a gap to that of CF. This can be explained by the fact the proposed
scheme is a memoryless symbol-by-symbol mapping, while CF utilizes all past received signals. Nevertheless, PL relaying is much simpler than DF and CF since both require block-wise encoding and decoding at the relay.
Figure 4.14: Optimized relay functions when $P_s = P_r = 10$ and $b = 1$. 
Figure 4.15: Performance comparison of instantaneous relaying with causal relaying as a function of gain of the relay–destination link (i.e., $b$) when $a = 1$, $P_s = P_r = 10$. 
Figure 4.16: Optimized relay functions when $P_s = P_r = 10$ and $a = 1$. 
Figure 4.17: Performance comparison of instantaneous relaying with causal relaying as a function of gain of the source–relay link (i.e., $\alpha$) when $b = 0.1$, $P_s = P_r = 10$. 
Figure 4.18: Optimized relay functions when $P_s = P_r = 10$ and $b = 0.1$. 
Appendix

4.A Capacity Bound for Noncausal Relaying

In this appendix, we derive an upper bound on the capacity of the relay channel with orthogonal receive components and noncausal relaying (i.e., $X_{ri} = f_i(Y^n_r)$). Using Fano’s inequality, we have

$$nR = H(M)$$
$$= I(M; Y^n_1, Y^n_2) + H(M|Y^n_1, Y^n_2)$$
$$\leq I(M; Y^n_1, Y^n_2) + n\epsilon_n,$$

(4.60)

where $\epsilon_n \to 0$ as $n \to \infty$ if $R$ is achievable. Now consider the following series of inequalities:

$$I(M; Y^n_1, Y^n_2) = H(Y^n_1, Y^n_2) - H(Y^n_1, Y^n_2|M)$$
$$\overset{(a)}{=} \sum_{i}^{n} H(Y_{1i}, Y_{2i}|Y_{1i}^{i-1}, Y_{2i}^{i-1}) - H(Y_{1i}, Y_{2i}|Y_{1i}^{i-1}, Y_{2i}^{i-1})$$
$$\overset{(b)}{\leq} \sum_{i}^{n} H(Y_{1i}, Y_{2i}) - H(Y_{1i}, Y_{2i}|X_i, X_{ri}, M, Y_{1i}^{i-1}, Y_{2i}^{i-1})$$
$$\overset{(c)}{=} \sum_{i}^{n} H(Y_{1i}, Y_{2i}) - H(Y_{1i}, Y_{2i}|X_i, X_{ri})$$
$$\overset{(d)}{=} \sum_{i}^{n} H(Y_{1i}, Y_{2i}) - H(Y_{1i}|X_i) - H(Y_{2i}|X_{ri})$$
$$\leq \sum_{i}^{n} H(Y_{1i}) + H(Y_{2i}) - H(Y_{1i}|X_i) - H(Y_{2i}|X_{ri})$$
$$= \sum_{i}^{n} I(X_i; Y_{1i}) + I(X_{ri}; Y_{2i})$$
$$\overset{(e)}{\leq} n (I(X; Y_1) + I(X; Y_2)),$$

(4.61)

where

(a) follows from the chain rule;

(b) holds because conditioning reduces entropy;

(c) follows since the channel is memoryless;

(d) holds because $p(y_1, y_2|x, x_r) = p(y_1|x)p(y_2|x_r)$; and

(e) follows from the time sharing argument.
In a similar fashion, consider

\[ I(M; Y^n_1, Y^n_2) \leq I(M; Y^n_1, Y^n_2) + I(M; Y^n_1 | Y^n_1, Y^n_2) \]

= \( I(M; Y^n_1, Y^n_2) \)

= \( I(M; Y^n_1, Y^n_2) + I(M; Y^n_1 | Y^n_1, Y^n_2) \)

= \( I(M; Y^n_1, Y^n_2) + H(Y^n_2 | Y^n_1, Y^n_2) - H(Y^n_2 | Y^n_1, Y^n_2, M) \)

\leq I(M; Y^n_1, Y^n_2) + H(Y^n_2 | Y^n_2) - H(Y^n_2 | Y^n_2)

= I(M; Y^n_1, Y^n_2)

\[ = \sum^n_i H(Y^n_{1i}, Y^n_{ri} | Y^n_{1i-1}, Y^n_{ri-1}) - H(Y^n_{1i}, Y^n_{ri} | M, Y^n_{1i-1}, Y^n_{ri-1}) \]

\leq \sum^n_i H(Y^n_{1i}, Y^n_{ri}) - H(Y^n_{1i}, Y^n_{ri} | X^n_i, M, Y^n_{1i-1}, Y^n_{ri-1})

= \sum^n_i I(X^n_i; Y^n_{1i}, Y^n_{ri})

\[ \leq n I(X^n; Y^n_1, Y^n_r), \]  

(4.62)

where

(a) follows from non-negativity of mutual information;
(b) follows from the fact that \((M, Y^n_1) \rightarrow Y^n_r \rightarrow Y^n_2 \) form a Markov chain;
(c) follows from the chain rule;
(d) holds because conditioning reduces entropy;
(e) follows since the channel is memoryless; and
(f) follows from the time sharing argument.

Putting (4.60)–(4.62) together we obtain

\[ C \leq \max_{p(x)p(x, x, y)} \min \{ I(X; Y_1) + I(X_r; Y_2), I(X; Y_1, Y_r) \} \]

= \[ \max_{p(x)p(x_r)} \min \{ I(X; Y_1) + I(X_r; Y_2), I(X; Y_1, Y_r) \} \]  

(4.63)
Chapter 5

Interference Management Using Relays

5.1 Introduction

An important issue that might arise in wireless networks (in particular ad-hoc networks), is that different source–destination pairs may interfere with one another when operating in the same frequency band. As an example, consider the ad-hoc network shown in Figure 5.1. The network consists of five nodes with two source–destination pairs. Nodes 1 and 4 wish to communicate with nodes 2 and 5, respectively. We assume that nodes 1 and 4 use the same frequency band. This consequently creates interference between the links 1–2 and 4–5. Additionally, node 3 in the network is available to assist nodes 1 and 2. However, nodes 1 and 3 cannot cooperate with each other since the link 1–3 is blocked. The only way to assist node 1 is to let the relay (i.e., node 3) overhear the signal transmitted from node 4 and then inform the destination regarding the interference. That is, the relay is used in the network to manage the interference among the links. This aspect of relaying is also considered in [MDG08, DMG08, SSE09]. However, our channel model and proposed relaying schemes are different from those in [MDG08, DMG08, SSE09].

It is reasonable to assume that the relay (node 3) in Figure 5.1 operates obliviously. That is the relay does not know the codebook of node 4 and is not hence able to perform decoding. Therefore, decode-and-forward type of relaying is not implementable in such networks. On the other hand, one can envision other types of relaying including memoryless linear and nonlinear relaying, and compress-and-forward (CF) relaying. The deployment of the oblivious relay also provides more flexibility for choosing a source–destination pair since the relay can help as long as it overhears the transmitted signal from the source.

We, in this chapter, develop a simple yet rich enough Gaussian model for this scenario and study different types of relaying schemes. We quantify the achievable rates of linear relaying (i.e. amplify-and-forward (AF)) and CF as in Chapter 2 and provide an algorithm for optimization of the memoryless nonlinear relay mappings. In this chapter, we present a new algorithm for optimizing the nonlinear relaying as compared to that in Chapter 4. Our results unveil the shape of optimal nonlinear
relaying. The optimized mapping resembles a periodic mapping such as sinusoid, in some cases. Thus the mapping can be considered as an analog and memoryless implementation of CF. We also show that optimized low-complexity nonlinear relaying performs superior to AF and operates close to CF. In particular, for the interference-limited cases, our numerical results indicate that optimized nonlinear memoryless relaying almost achieves the capacity, while linear relaying may have an unbounded gap to the capacity.

5.2 System Model

Figure 5.2 shows a three-node Gaussian relay channel where the source–destination link suffers from an additive interference (this interference originates from another source–destination pair which uses the same frequency band, c.f., Figure 5.1). The relay node partially observes the interference as illustrated in Figure 5.2. In our model, we do not impose any particular structure on the interference. That is the
Figure 5.2: The three-node Gaussian relay channel with an additive Gaussian interference. The interference is partially observed by the relay. The relay has access to an orthogonal link to convey its knowledge about the interference to the destination.

relay and the destination do not know the modulation and the codebook that the interferer transmitter is using.

The communication task is to reproduce the transmitted message $M$, uniformly chosen from the set $\mathcal{M}(n) = [1: 2^{nR}]$, at the destination while maintaining arbitrarily small probability of error. The transmission of a message consumes $n$ channel uses. The channel in Figure 5.2 suffers from additive Gaussian noises $Z^n_r \sim \mathcal{N}(0, N_r I_n)$, $Z^n_1 \sim \mathcal{N}(0, N_1 I_n)$, $Z^n_2 \sim \mathcal{N}(0, N_2 I_n)$ and an additive Gaussian interference $S^n \sim \mathcal{N}(0, QI_n)$ on the source–destination link. The random sequences $Z^n_r$, $Z^n_1$, $Z^n_2$, and $S^n$ are assumed to be mutually independent.

The received signal vector at the relay $Y^n_r$ is given by

$$Y^n_r = S^n + Z^n_r,$$

where $S^n$ is the additive interference. That is, the relay observes a noisy version of the true interference. The channel outputs $(Y^n_1, Y^n_2)$ are given by

$$Y^n_1 = X^n + S^n + Z^n_1,$$

$$Y^n_2 = X^n_r + Z^n_2,$$

where $X^n$ and $X^n_r$ are the signal transmitted by the source and the relay, respectively. We assume an average power constraint on $X^n$ and $X^n_r$ such that $E[\|X^n\|^2] \leq n P_s$ and $E[\|X^n_r\|^2] \leq n P_r$. The channel model in Figure 5.2 falls into the class of relay channels considered in Definition 2.3.1.

We would like to quantify the supremum of the set of $R$ values for which the average message error probability at the destination can be made to approach zero as the number of channel uses $n$ goes to infinity. This number as defined in Chapter 2 is the capacity, $C$, in the communication between the source and the destination.
5.3 Cutset Upper Bound

The following proposition gives an upper bound on the capacity.

**Proposition 5.3.1.** For the relay channel in Figure 5.2, the capacity is upper bounded by

\[
C \leq \min \left\{ C \left( \frac{P_s}{N_1 + Q} \right) + C \left( \frac{P_r}{N_2} \right) , C \left( \frac{P_s}{N_1 + Q \left( 1 - \frac{Q}{Q + N_r} \right)} \right) \right\},
\]

(5.3)

where \( C(x) := \frac{1}{2} \log_2 (1 + x) \).

**Proof.** Using the cutset bound in (2.6), we have

\[
C \leq \max_{p(x,x_r)} \min \{ I(X;X_r|Y_1,Y_2), I(X;Y_1,Y_2,Y_r|X) \}.
\]

Now we bound each term. First consider

\[
I(X;X_r;Y_1,Y_2) = h(Y_1,Y_2) - h(Y_1,Y_2|X,X_r)
\]

\[= h(Y_1,Y_2) - h(\tilde{Y}_1,Z_2|X,X_r) \tag{a}\]

\[= h(Y_1,Y_2) - h(\tilde{Y}_1) - h(Z_2) \tag{b}\]

\[= h(Y_1) + h(Y_2|Y_1) - h(\tilde{Y}_1) - h(Z_2) \tag{c}\]

\[\leq h(Y_2|Y_1) + \frac{1}{2} \log_2 (2\pi e (P_s + Q + N_1)) \]

\[- \frac{1}{2} \log_2 (2\pi e (Q + N_1)) - \frac{1}{2} \log_2 (2\pi e N_2) \tag{d}\]

\[\leq \frac{1}{2} \log_2 \left( 2\pi e \left[ \mathbb{E} \left( Y_2 - \mathbb{E} [Y_2|Y_1] \right)^2 \right] \right) + \frac{1}{2} \log_2 \left( \frac{P_s + Q + N_1}{2\pi e (Q + N_1) N_2} \right) \tag{e}\]

\[= \frac{1}{2} \log_2 \left( 2\pi e \left( P_s + N_2 - \rho \sqrt{P_s P_r} \right) \right) \]

\[+ \frac{1}{2} \log_2 \left( \frac{P_s + Q + N_1}{2\pi e (Q + N_1) N_2} \right) \]

\[= \frac{1}{2} \log_2 \left( \frac{(P_r + N_2)(P_s + Q + N_1) - \rho \sqrt{P_s P_r}}{(Q + N_1) N_2} \right) \]

\[\leq C \left( \frac{P_s}{N_1 + Q} \right) + C \left( \frac{P_r}{N_2} \right), \tag{5.4}\]

where
5.4 Achievable Rates Using a Memoryless Relay

(a) follows by defining $\tilde{Y}_1 := Y_1 - X = S + Z_1$;

(b) holds since $\tilde{Y}_1$ and $Z_2$ are independent of $X$ and $X_r$;

(c) follows by entropy maximization under an average power constraint;

(d) same as (c); and

(e) follows by defining $\rho := \frac{E[X X_r]}{\sqrt{E[X^2|X]}} = \frac{E[X X_r]}{\sqrt{P_s P_r}}$.

Next consider

\[
I(X; Y_r, Y_1, Y_2|X_r) = I(X; Y_1, Y_r|X_r) + I(X; Y_2|Y_1, Y_r, X_r)
\]
\[
= I(X; Y_1, Y_r|X_r) + h(Y_2|Y_1, Y_r, X_r)
\]
\[
- h(Y_2|Y_1, Y_r, X_r, X)
\]
\[
= I(X; Y_1, Y_r|X_r) + h(Z_2|Y_1, Y_r, X_r)
\]
\[
- h(Z_2|Y_1, Y_r, X_r, X)
\]
\[
= I(X; Y_1, Y_r|X_r)
\]
\[
h(Y_1, Y_r|X_r) - h(Y_1, Y_r|X_r)
\]
\[
h(Y_1, Y_r|X_r) - h(\tilde{Y}_1, Y_r)
\]
\[
\leq h(Y_1, Y_r) - h(\tilde{Y}_1, Y_r)
\]
\[
\leq \frac{1}{2} \log_2 \left( \frac{P_s + Q + N_1}{Q + N_r} \right)
\]
\[
- \frac{1}{2} \log_2 \left( \frac{Q + N_1}{Q + N_r} \right)
\]
\[
= C \left( \frac{P_s}{N_1 + Q \left( 1 - \frac{Q}{Q + N_r} \right)} \right). \tag{5.5}
\]

Combining (5.4) and (5.5) and noting that both terms are maximized by choosing $\rho = 0$ yield (5.3).

5.4 Achievable Rates Using a Memoryless Relay

In this section, we restrict the relay to employ a memoryless, symbol-by-symbol, deterministic mapping (see also Section 2.4). That is, the present output at the relay $x_{ri}$ is a memoryless time-invariant function of the present input $y_{ri}$, such that $x_{ri} = f(y_{ri})$, for all $i \in [1 : n]$. 

\[\]
For the relay channel with known interference shown in Figure 5.2, the following rate is achievable

\[ R_{\text{ins}} = \sup_{p(x), x_r = f(y_r), \mathbb{E}[X^2] \leq P_s, \mathbb{E}[X_r^2] \leq P_r} I(X; Y_1, Y_2), \]

where \( f(y_r) \) denotes a memoryless, symbol-by-symbol, deterministic mapping at the relay.

### 5.4.1 Benchmark: Linear Relaying

With linear relaying, the relay transmits a scaled version of the analog received signal to the destination. That is

\[ x_r = f(y_r) = \beta y_r, \]

where \( \beta = \sqrt{P_r Q + N_r} \) is chosen to meet the power constraint at the relay. The achievable rate of linear relaying is given by

\[
R_{AF} = I(X; Y_1, Y_2) = h(Y_1, Y_2) - h(Y_1, Y_2|X) \\
= (a) h(Y_1, Y_2) - h(\tilde{Y}_1, Y_2) \\
= \left( b \right) \frac{1}{2} \log_2 \left( 1 + \frac{P_s + Q + N_1}{\beta Q P_r + N_2} \left( \frac{\beta Q}{P_r + N_2} \right) \right) \\
= C \left( \frac{P_s (1 - \frac{\beta^2 Q}{P_r + N_2})}{N_1 + Q \left( 1 - \frac{P_s Q}{P_r + N_2} \right)} \right), \tag{5.7}
\]

where (a) follows by defining \( \tilde{Y}_1 := Y_1 - X = S + Z_1 \) and (b) follows by choosing

\[ X \sim \mathcal{N}(0, P_s). \]

One natural way to utilize the received signal \( Y_2 \) at the destination is first to estimate \( S \) from \( Y_2 \) and then subtract the estimated interference from \( Y_1 \). The MMSE estimate of \( S \) given \( Y_2 \) can be computed as \( \hat{S} = \mathbb{E}[S|Y_2] = \frac{\beta Q}{P_r + N_2} Y_2 \). By subtracting \( \hat{S} \) from \( Y_1 \), we obtain \( \tilde{Y}_1 = Y_1 - \hat{S} = X + S + Z_1 - \frac{\beta Q}{P_r + N_2} Y_2 \). Now, by using only \( \tilde{Y}_1 \) for decoding at the destination, we can achieve the rate

\[
R_{AF} = I(X; \tilde{Y}_1) = C \left( \frac{P_s}{N_1 + \xi^2} \right)
\]
5.4. Achievable Rates Using a Memoryless Relay

\[ C \left( \frac{P_s}{N_1 + Q \left( 1 - \frac{P_r Q}{(P_r + N_2)(P_r + N_r)} \right)} \right), \]  

(5.8)

where \( \xi^2 := \mathbb{E} [(S - \hat{S})^2] = Q \left( 1 - \frac{P_r Q}{(P_r + N_2)(P_r + N_r)} \right) \). Note that the rate in (5.8) is the same as the one in (5.7). Hence this procedure of incorporating \( Y_2 \) in the decoding is optimal.

5.4.2 Optimized Nonlinear Relaying

As we demonstrated in Chapter 4 linear relaying is a suboptimal strategy in general. We therefore consider the design of optimal memoryless relay mappings. Here we confine our investigation to using a Gaussian codebook at the source (not necessarily optimal). In order to take into account the power constraint at the relay, we construct the Lagrangian

\[ J = I(X; Y_1, Y_2) + \lambda (\mathbb{E} [f^2(y_r)] - P_r), \]  

(5.9)

where \( \lambda \geq 0 \) and

\[
I(X; Y_1, Y_2) = h(Y_1, Y_2) - h(\tilde{Y}_1, Y_2) \\
= -E [\log(p(y_1, y_2))] + E [\log(p(\tilde{y}_1, y_2))] \\
= - \iint \log(p(y_1, y_2))p(y_1, y_2)dy_1dy_2 \\
+ \iint \log(p(\tilde{y}_1, y_2))p(\tilde{y}_1, y_2)d\tilde{y}_1dy_2.
\]  

(5.10)

The necessary condition for optimality is then given by \( \frac{\partial J}{\partial f} = 0 \). The final result after some manipulation is given in the following lemma.

**Lemma 5.4.1.** Let \( X \sim \mathcal{N}(0, P_s) \) and \( \tilde{Y}_1 := Y_1 - X = S + Z_1 \). If the joint densities \( p(\tilde{y}_1, y_2) \) and \( p(y_1, y_2) \) are differentiable, then the optimal mapping \( f^* \) satisfies

\[
f^*(y_r) = \frac{1}{2\lambda p(y_r)} \left( \iint (1 + \log(p(y_1, y_2))) \frac{\partial p(y_1, y_2)}{\partial f^*(y_r)} dy_1dy_2 \\
- \iint (1 + \log(p(y_1, y_2))) \frac{\partial p(y_1, y_2)}{\partial f^*(y_r)} dy_1dy_2 \right),
\]  

(5.11)

where

\[
\frac{\partial p(y_1, y_2)}{\partial f^*(y_r)} = \frac{y_2 - f^*(y_r)}{2N_2} p(y_2|y_r)p(y_1|y_r)p(y_r), \\
\frac{\partial p(\tilde{y}_1, y_2)}{\partial f^*(y_r)} = \frac{y_2 - f^*(y_r)}{2N_2} p(y_2|y_r)p(\tilde{y}_1|y_r)p(y_r),
\]
\begin{align*}
p(y_1, y_2) &= \int p(y_2|y_r)p(y_1|y_r)p(y_r)dy_r, \\
p(\tilde{y}_1, y_2) &= \int p(y_2|y_r)p(\tilde{y}_1|y_r)p(y_r)dy_r, \\
p(y_1|y_r) &= \mathcal{N}\left(\frac{Q}{Q+N_r}y_r, P_s + Q + N_1 - \frac{Q^2}{Q+N_r}\right), \\
p(\tilde{y}_1|y_r) &= \mathcal{N}\left(\frac{Q}{Q+N_r}y_r, Q + N_1 - \frac{Q^2}{Q+N_r}\right), \\
p(y_2|y_r) &= \mathcal{N}(f(y_r), N_2), \text{ and} \\
p(y_r) &= \mathcal{N}(0, Q+N_r).
\end{align*}

From (5.11), we see that the optimality condition can be written as \( f^* = \mathcal{F}(f^*) \). Thus the optimal mapping is a fixed point of \( f = \mathcal{F}(f) \). We use the fixed-point iteration method to solve the equation in (5.11). That is we compute the sequence of mappings \( \{f_i\} \) subject to the power constraint using \( f_{i+1} = \mathcal{F}(f_i) \) with \( f_1 = f_{\text{init}} \) as an initial solution. In our computer implementation we restrict the mappings to be \( f_i : [-M, M] \rightarrow \mathbb{R} \), where \( M := 3\sqrt{Q+N_r} \). The probability that a received sample at the relay is within the interval \([-M, M] \) is

\[
\text{Pr}\{y_r \in [-M, M]\} = 1 - 2Q(3) \approx 0.999, \tag{5.12}
\]

where

\[
Q(x) := \frac{1}{\sqrt{2\pi}} \int_{x}^{+\infty} e^{-\frac{t^2}{2}} dt.
\]

We have observed that the algorithm converges for a wide range of channel parameters \((P_s, P_r, Q, N_1, N_2, N_r)\). Additionally, the final solution depends on the initial mapping \( f_1 \), in general. Therefore, this procedure does not guarantee reaching a global optimum, however it in general reaches a reasonable local stationary point, as the numerical results indicate. Figure 5.3 shows an example of optimized mappings for \( P_s = 0, P_r = 0, Q = 5, N_1 = -30, N_r = -30, \) and \( N_2 = -5 \) dB. The mapping sequence \( \{f_i\} \) is shown for \( i = 1, 8, 20, 100 \) in Figure 5.3. We see that the algorithm converges to a periodic function with sharp peaks as \( i \) increases.

**Shape of Optimized Functions**

Figures 5.3 and 5.4 show two typical examples of optimized mappings obtained using (5.11) and linear mapping as an initial guess. We have observed similar shapes for various configurations of the channel parameters. These mappings possess some interesting features as follows.

- The mappings are non-invertible. That is the same output is used for different input values. This results in ambiguity, however the destination will resolve this by using the side-information received via the source–destination link.
5.4. Achievable Rates Using a Memoryless Relay

- The optimized mappings resemble periodic functions whose period depends on the quality of links and the strength of the interference. The received signals at the relay and at the destination are correlated through the interference. Hence, when the correlation is strong the relay utilizes the side information and employs a function with smaller periods (i.e., more zero crossings for a given input range). C.f. Figures 5.3 and 5.4.

- The mappings are power efficient, since the relay avoids large output values when employing optimized mappings. This stands in a sharp contrast to the case of linear relaying.

The proposed analog mappings include the scalar quantize-and-forward protocol as a special case. The optimal scalar quantizer for the underlying channel is nested in general because of the availability of side information at the destination. With this analogy, each period of the optimized mappings can be considered as a quantization bin in the analog domain. This further explains the periodic behavior of the optimized mappings. Motivated by the above features, the optimized mappings can be considered to perform one-dimensional analog compression at the relay. The
amount of compression depends on the quality of the links and in particular the side information available via the direct link.

5.5 Relaying via Compress-and-Forward

**Proposition 5.5.1.** For the relay channel in Figure 5.2, the following rate is achievable

\[
R_{CF} = C \left( \frac{P_s}{N_1 + Q \left( 1 - \frac{Q}{Q + N_r + N_q} \right)} \right),
\]

where

\[
N_q := \frac{N_2 [(P_s + N_1)(Q + N_r) + Q N_r]}{P_r (P_s + Q + N_1)}.
\]

**Proof.** Using the lower bound in (2.23), the rate \( R_{CF} = \sup I(X; Y_1, \hat{Y}_r | X_r) \) subject
5.5. Relaying via Compress-and-Forward

$I(\hat{Y}_r; Y_r | X_r, Y_1) \leq I(X_2; Y_2)$ is achievable, where the supremum is taken over

$$p(x)p(x_r)p(\hat{y}_r | y_r, x_r).$$

Now let $X \sim \mathcal{N}(0, P_s)$ and be independent of $X_r \sim \mathcal{N}(0, P_r)$ and define $\hat{Y}_r := Y_r + Z_q$ where $Z_q \sim \mathcal{N}(0, N_q)$ and is independent of other random variables in the channel. Next consider

$$I(\hat{Y}_r; Y_r | Y_1, X_r) = I(\hat{Y}_r; Y_r | Y_1) = h(\hat{Y}_r | Y_1) - h(\hat{Y}_r | Y_1)$$

$$= h(\hat{Y}_r | Y_1) - h(\hat{Y}_r | Y_1)$$

$$= h(\hat{Y}_r, Y_1) - h(Y_1) - h(Z_q)$$

$$= \frac{1}{2} \log_2 \left( \frac{Q + N_r + N_q}{Q} \right)$$

$$= C \left( \frac{(P_s + Q + N_1)(Q + N_r)}{N_q(P_s + Q + N_1)} - Q^2 \right).$$

Now by solving $I(\hat{Y}_r; Y_r | Y_1) = I(X_r; Y_2) = C \left( \frac{P_s}{N_q} \right)$, we obtain

$$N_q = \frac{N_2}{P_r(N_q)} \left[ (P_s + N_1)(Q + N_r) + QN_q \right].$$

Thus, by this choice of $N_q$, we find the following rate

$$R_{CF} = I(X; Y_1, \hat{Y}_r)$$

$$= h(Y_1, \hat{Y}_r) - h(\hat{Y}_1, \hat{Y}_r)$$

$$= \frac{1}{2} \log_2 \left( \frac{P_s + Q + N_1}{Q} \right)$$

$$= C \left( \frac{P_s}{N_1 + Q \left( 1 - \frac{Q}{Q + N_r + N_q} \right)} \right). \quad (5.14)$$

This completes the proof.

In the following, we briefly explain how the CF strategy works in this setup. The relay upon receiving the sequence $Y^n_r = S^n + Z^n_r$, quantizes this sequence using vector quantization with $Y^n_1 = X^n + S^n + Z^n_1$ as the side information (i.e., the Wyner–Ziv approach). Then the quantized index is transmitted via the relay–destination link using a Gaussian codebook. The destination then uses $Y^n_2 = X^n + Z^n_2$ to decode.
the transmitted codeword from the relay. Having successfully decoded the relay’s codeword, the destination uses \( X^n \) and \( Y^n_1 \) as side information to retrieve \( \hat{Y}_n \). The destination estimates the interference using \( \hat{Y} = Y^n_1 - \hat{S}^n \). Then the estimated interference sequence \( \hat{S}^n \) is subtracted from \( Y^n_1 \) to obtain \( \bar{Y} = Y^n_1 - \hat{S}^n \). Finally, the destination decodes \( X^n \) using \( \bar{Y} \). By doing so, the rate \( I(X; \bar{Y}) \) is achievable, which is given by (5.13).

**Remark 5.5.2.** A binary counterpart of the channel in Figure 5.2 with \( Z_1 \equiv 0 \) and modulo-sum noises is investigated in [ARY09]. It is shown that the capacity is achieved using a quantize-and-forward scheme without Wyner–Ziv coding (see Theorem 1 in [ARY09]). Now let \( X \sim N(0, P_s) \) and be independent of \( X_r \sim N(0, P_r) \), and define \( U := Y_r + Z_q \) where \( Z_q \sim N(0, N_q) \) and is independent of other random variables in the channel. Then invoking Theorem 1 in [ARY09] by optimizing the choice of \( N_q \) yields the same rate as given in (5.7), which is achieved by linear relaying. Thus, for our Gaussian channel model joint source-channel coding is essential to obtain rates above those achieved by linear relaying.

### 5.6 Capacity of the Interference-Limited Channel

Assume that the noise components on the source–relay and the source–destination links are negligible compared to the additive interference. That is \( Y_1 = X + S + Z_1 \approx X + S \) and \( Y_r = S + Z_r \approx S \). This is reasonable when the destination and the relay are located close to the interferer node.

**Corollary 5.6.1.** The capacity of the interference limited case is

\[
C = C\left(\frac{P_s}{Q}\right) + C\left(\frac{P_r}{N_2}\right).
\]

**Proof.** From Proposition 5.3.1, by substituting \( N_1 = N_r = 0 \) in (5.3), we find

\[
C \leq C\left(\frac{P_s}{Q}\right) + C\left(\frac{P_r}{N_2}\right).
\]

This bound coincides with the achievable rate of CF in (5.13). This completes the proof.

**Remark 5.6.2.** Note that in the interference limited case, we have \( Y_r \approx Y_1 - X \). That is the received signal at the relay is a deterministic function of the received signal at the destination \( (Y_1) \) and the transmitted symbol from the source \( (X) \). Thus the hash-and-forward strategy in [Kim08, CK07] followed by point-to-point channel coding on the relay–destination link, also achieves the capacity (see also Section 2.3.6).
Remark 5.6.3. Let $\gamma_1 := \frac{P_s}{Q}$, and $\gamma_2 := \frac{P_r}{N_2}$. Then fix $\gamma_1 \neq 0$ and define

$$\Delta(\gamma_2) := C - R_{AF} = C\left(\frac{\gamma_2}{1 + \gamma_1 + \gamma_1 \gamma_2}\right).$$ (5.17)

One can verify that $\Delta(\gamma_2)$ is an increasing function and $\Delta(\gamma_2) \leq \lim_{\gamma_2 \to \infty} \Delta(\gamma_2) = C\left(\frac{1}{\gamma_1}\right)$. That is linear relaying can have an unbounded gap to the capacity if the source has a very little power to consume. However, we show that for such cases the optimized nonlinear strategy significantly closes the gap. See Section 5.7 and Figure 5.7.

5.7 Numerical Examples

We next present numerical examples of transmission rates achieved with linear relaying, the proposed nonlinear relaying, and CF relaying and compare them against the cutset bound.

In Figures 5.5 and 5.6, we analyze the effect of varying the strength of the source–destination link, on the achievable rate of different relaying schemes under the discussion. In Figure 5.5, we fix source power $P_s$ to 0 dB, and plot achievable rate for several values of the source–destination noise power $N_1$. In Figure 5.5, the interferer power $Q$ is set to 10 dB higher than the source power, and the remaining channel parameters are chosen such that $P_r = 0$, $N_r = -30$, $N_2 = -5$ dB. The achievable rate of the linear relay slightly increases in the beginning and then stays almost constant. This saturation in achievable rate is due to the high interference power and the linear relay’s incapability to provide sufficient information about the interference to the destination. This incapability makes the achievable rate limited by the presence of the interference, and the rate cannot be improved further by reducing the source–destination noise. On the other hand, the achievable rate with CF relaying (shown in Figure 5.5) increases monotonically with decreasing source–destination noise, and finally achieves the upper bound at very low noise values.

There is a significant performance gap between the optimized memoryless nonlinear relay and the memoryless linear relay. Additionally, a significant part of the performance gap between CF with infinite memory and memoryless linear relaying is compensated for with our proposed nonlinear relaying as illustrated in Figure 5.5. Most notably, the achievable rate with the nonlinear relay increases monotonically (without saturation) with decreasing source–destination noise and gets very close to the CF relaying and the upper bound at very low source–destination noise power.

In Figure 5.6, we fix the source–destination noise power $N_1$ to $-35$ dB and plot achievable rates for several values of source power $P_s$. The interferer power $Q$ is set to 15 dB, and the remaining channel parameters are the same as those in Figure 5.5 (i.e., $P_r = 0$, $N_r = -30$, and $N_2 = -5$ dB). In Figure 5.6 (similar to Figure 5.5), the nonlinear relay provides far better achievable rates as compared to the linear relay and the rates get close to those of CF relaying and the upper bound when the source–destination link quality is good. For example at $P_s = -15$ dB (when
source power is 30 dB lower than interferer power), the rate achieved with a linear relay is almost zero whereas with a nonlinear relay we can achieve almost 0.65 bits. In other words, with linear relaying to achieve the rate 0.65 bits, one needs 26 dB higher source power as shown in Figure 5.6. The performance gap between linear and nonlinear relaying reduces at higher $P_s$. This is because of the fact that the relay has a significant role only when the received signal is dominated by the interference.

We finally investigate the interference-limited scenario where the additive noise at the relay and the noise component on the source–destination link are negligible. We consider a source power 10 dB lower than the interferer power ($P_s = 0$, $Q = 10$ dB) and assume noise-free source–destination and interferer–relay links. (See also Section 5.6.) Figure 5.7 shows the achievable rate for different schemes versus the relay–destination SNR $P_r/N_2$. The achievable rate of the nonlinear relay very closely follows that of CF relaying for all noise power values shown in Figure 5.7. (We recall that in the given situation CF and hash-and-forward attain the capacity for all $N_2$). Thus, our simple memoryless nonlinear relay cleverly manages interference to provide rates very close to those achieved by infinite memory CF relaying (the upper bound) in the interference-limited scenario.

5.8 Summary and Concluding Remarks

We presented a class of relay channels where there is interference on the source–destination link, and the relay observes this interference partially. We presented three main relaying strategies: memoryless linear relaying (AF), memoryless nonlinear relaying and compress-and-forward (CF) relaying. The numerical results indicate a significant gain of nonlinear relaying over linear relaying. In particular, we showed that for the interference-limited scenario CF is optimal and, more interestingly, we demonstrated that the optimized memoryless nonlinear relay performs very close to the capacity bound. This result is of crucial importance, since memoryless relaying is far more practical than CF. Additionally, memoryless nonlinear relaying has low delay which is suitable for delay sensitive applications, as for example in closed-loop control applications. Finally, we stress that the deployment of inexpensive relays, operating with memoryless nonlinear mappings, in ad-hoc wireless networks in order to manage the interference adds a new dimension for designing future communication systems.
Figure 5.5: Capacity bounds when $P_s = 0$ dB, $P_r = 0$ dB, $Q = 10$ dB, $N_r = -30$ dB, $N_2 = -5$ dB.
Figure 5.6: Capacity bounds when $P_r = 0$ dB, $Q = 15$ dB, $N_1 = -35$ dB, $N_r = -30$ dB, $N_2 = -5$ dB.
Figure 5.7: Capacity bounds when $P_s = 0$ dB, $P_r = 0$ dB, $Q = 10$ dB, $N_1 = N_r = -\infty$ dB.
Chapter 6

Capacity Achieving Instantaneous Relaying

6.1 Introduction

We investigated memoryless relaying (i.e., instantaneous relaying) in particular in Chapters 4 and 5 and demonstrated that instantaneous relaying, in which the output of the relay depends only on the presently received input, can be optimal for some special cases of a general Gaussian model. In this chapter, we unify the ideas in Chapters 4 and 5, and introduce a more general class of semi-deterministic relay channels for which instantaneous relaying is optimal. We then discuss some applications of our new result to the Gaussian case. In particular, our result also proves the optimality of instantaneous relaying for a Gaussian relay channel with interference studied in Section 5.6.

Figure 6.1 illustrates the three-node relay channel that we investigate in this chapter. The channel is constituted of two sub-channels; the broadcast sub-channel from the source to the relay and the destination \((X, p(y_1, y_r|x), \mathcal{Y}_1 \times \mathcal{Y}_r)\), and the point-to-point sub-channel from the relay to the destination \((X_r, p(y_2|x_r), \mathcal{Y}_2)\). Here \(\mathcal{Y}_1, \mathcal{Y}_2\) and \(\mathcal{Y}_2\) are subsets of \(\mathbb{R}\), and \(p(y_1, y_r|x)\) and \(p(y_2|x_r)\) denote conditional probability density functions (pdf’s) on \(\mathcal{Y}_1 \times \mathcal{Y}_r\) and \(\mathcal{Y}_2\), respectively. The goal is to reliably transmit a message index \(M\) uniformly chosen from the set \(M^\binom{n}{n}\) in \(n\) channel uses. Given \(M = m\), an encoder \(\alpha_n\) produces a codeword \(x_1^n(m)\), with letters \(x_i(m)\) such that the channel input is \(x = x_i(m)\) at the \(i\)th channel use, \(i \in [1:n]\). The received signal from the source at the relay and the destination are denoted by \(y_r \in \mathcal{Y}_r\) and \(y_1 \in \mathcal{Y}_1\), respectively. The relay employs the function \(f\) to generate a symbol \(x_r \in \mathcal{X}_r \subseteq \mathbb{R}\) based only on the presently received signal \(y_r\). The received signal from the relay at the destination is denoted by \(y_2 \in \mathcal{Y}_2\). Note that \(y_1\) and \(y_2\) do not interfere with each other. Finally, the destination employs the decoder \(\beta_n\) to decode the transmitted message. We also assume that the channel is memoryless, i.e.

\[
p(y_{11}^n, y_{1}^n|x_1^n) = \prod_{i=1}^{n} p(y_{1i}, y_{ri}|x_i),
\]

119
The rate $R$ is achievable, if there exists a triple $(\alpha_n, f, \beta_n)$, such that the probability of error $P_e^{(n)} = \Pr\{\hat{M} \neq M\}$ at the decoder tends to zero as $n \to \infty$ and at the same time
\[
\liminf_{n \to \infty} \frac{1}{n} \log |\mathcal{M}^{(n)}| \geq R.
\]
As before in the thesis, the capacity is then defined as the supremum of all achievable rates subject to a cost constraint. We consider the following three possible cases for a constraint:

- no cost constraint, i.e. $\{(x_i \times x_{ri}) \in (\mathcal{X} \times \mathcal{X}_r)\}$;
- letter cost constraint, i.e. $\{(x_i \times x_{ri}) \in (\mathcal{X}_c \times \mathcal{X}_{rc}) \subseteq (\mathcal{X} \times \mathcal{X}_r)\}$ (for example peak power constraints at the source and the relay); and
- average cost constraint, i.e.
\[
\left\{(x^n \times x^n_r) \in (\mathcal{X}^n \times \mathcal{X}_r^n) : \frac{1}{n} \sum_{i=1}^{n} c_s(x_i(m)) \leq \Gamma_s, \right. \\
\left. \frac{1}{n} \sum_{i=1}^{n} E[c_r(X_i)|X = x_i(m)] \leq \Gamma_r, \ m \in \mathcal{M}^{(n)} \right\},
\] (6.1)
where $c_s : \mathcal{X} \mapsto \mathbb{R}$ and $c_r : \mathcal{X}_r \mapsto \mathbb{R}$ denote per-letter cost functions at the source and the relay respectively, and $\Gamma_s$ and $\Gamma_r$ are two arbitrary constants. (C.f., average power constraints at the source and the relay.)
6.2 Main Result

Theorem 6.2.1. Assume that

- \( y_r = g(x, y_1) \), i.e., the received signal at the relay is a deterministic function of \( x \) and \( y_1 \); and
- \( p(y_r) \) and \( p(y_1 | y_r) \) are Riemann integrable.

Then the capacity of the channel in Figure 1, with instantaneous relaying, is given by

\[
C = \max_{p(x) \in \mathcal{P}} I(X; Y_1) + \max_{p(x_r) \in \mathcal{P}_r} I(X_r; Y_2),
\]

where \( \mathcal{P} \) and \( \mathcal{P}_r \) describe the sets of permissible \( p(x) \) and \( p(x_r) \) such that the cost constraints are met. That is

- no cost constraint: \( \mathcal{P} := \{ p(x) : \text{Supp}(p(x)) \subseteq X \} \);
- letter cost constraint: \( \mathcal{P} := \{ p(x) : \text{Supp}(p(x)) \subseteq X_c \} \); and
- average cost constraint: \( \mathcal{P} := \{ p(x) : E[c_s(X)] \leq \Gamma_s, \text{Supp}(p(x)) \subseteq X \} \).

The first condition in Theorem 6.2.1 indicates that the broadcast sub-channel is deterministic in the sense stated. This constraint helps the destination to be able to recover the transmitted symbols from the relay if the destination knows \( X \) (i.e., a transmitted component of a codeword), and \( Y_1 \). The second condition indicates that the broadcast sub-channel should be continuous in the sense that \( p(y_r) \) and \( p(y_1 | y_r) \) are Riemann integrable. This property is needed in order to construct an instantaneous operation at the relay such that the generated symbols at the relay possess capacity achieving properties.

Remark 6.2.1. Recall from Chapter 2, for this class of relay channels, both hash-and-forward [Kim08, CK07] and compress-and-forward [CG79] achieve the capacity as well. Additionally, if \( y_r = g(x) \) the capacity can be achieved using partial decode-and-forward [GA82]. The contribution in the chapter, however, is that we propose an instantaneous strategy that does not require block encoding or decoding at the relay.
6.2.1 Proof of Theorem 6.2.1

We first prove the converse. By the upper bound for the relay-without-delay [GHM07, Theorem 2] we have

\[
C \leq I(X; Y_1, Y_2) \\
= h(Y_1, Y_2) - h(Y_1, Y_2|X, V) \\
= h(Y_1) - h(Y_1|X, V) + h(Y_2|Y_1) - h(Y_2|X, V, Y_1) \\
\leq (a) \leq h(Y_1) - h(Y_1|X, V) + h(Y_2) - h(Y_2|X_r, X, V, Y_1) \\
\leq (b) \leq h(Y_1) - h(Y_1|X, V) + h(Y_2) - h(Y_2|X_r) \\
\leq (c) \leq h(Y_1) - h(Y_1|X) + h(Y_2) - h(Y_2|X_r) \\
= I(X; Y_1) + I(X_r; Y_2), \tag{6.3}
\]

where

(a) follows since conditioning reduces entropy;

(b) follows because \((X, V, Y_1) \rightarrow X_r \rightarrow Y_2\) form a Markov chain; and

(c) holds since \(V \rightarrow X \rightarrow Y_1\) form a Markov chain because

\[
I(V; Y_1|X) \leq I(V; Y_1, Y_r|X) \\
= h(Y_1, Y_r|X) - h(Y_1, Y_r|X, V) \\
= 0. \tag{6.4}
\]

We next prove the positive part, that is, we establish that the right hand side of (6.2) is an achievable rate, in a few steps as follows.

- **Random codebook at the source:** Generate \(|\mathcal{M}^{(n)}|\) i.i.d. codewords of length \(n\) with pdf \(p(x^n) = \prod_{i=1}^{n} p^*(x)\) where \(p^*(x) \in \mathcal{P}\) maximizes \(I(X; Y_1)\). Note that in the case of an average cost constraint, the codewords can be made to satisfy the constraint with probability \(1 - \epsilon\) for any \(\epsilon > 0\) by choosing \(n\) large enough.

- **Parameterized relay mapping:** Let the relay mapping be parameterized by \(k \geq 0\) as

\[
f_k(y_r) = F^{-1} \left( \frac{1}{2} + \left( \frac{y_r}{\Delta_k} \right) \right), \tag{6.5}
\]

where

- \(\Delta_k := \Delta_0 2^{-k}\) and \(\Delta_0 > 0\) is a constant;
6.2. Main Result

- $s(x)$ is a periodic function with period one (i.e., $s(x + 1) = s(x)$) and $s(x) = 2x$ for $-rac{1}{2} \leq x < \frac{1}{2}$, and

- $F^{-1}(t) := \inf\{x : F(x) > t\}$, denotes the inverse of $F(t)$ and $F(t)$ is the cumulative distribution function (cdf) of $X_r \sim p^*(x_r)$ where $p^*(x_r) \in \mathcal{P}_r$ maximizes $I(X_r, Y_2)$.

Apply the same relay mapping $f_i(y_i) = f_k(y_r)$ for all channel uses $i \in [1 : n]$.

- **Achievable rate:** For the given relay function $f_k$, the channel is a memoryless point-to-point link parameterized by

$$p_k(y_1, y_2|x) = p(y_1|x)p_k(y_2|x, y_1) = p(y_1|x)p_k(y_2|f_k(g(x, y_1))),$$

where the second equality holds since $x_r = f_k(y_r) = f_k(g(x, y_1))$ and the fact that $(X, Y_1) \rightarrow f_k(g(X, Y_1)) \rightarrow Y_2$ form a Markov chain. Thus, denoting mutual information and differential entropy, when the corresponding random variables are affected by $f_k$, as $I_k$ and $h_k$, respectively, the following rate is achievable

$$R_k = I_k(X; Y_1, Y_2) = I(X; Y_1) + I_k(X; Y_2|Y_1) = I(X; Y_1) + h_k(Y_2|Y_1) - h_k(Y_2|Y_1, X)$$

$$= I(X; Y_1) + h_k(Y_2|Y_1) - h_k(Y_2|Y_1, X, X_r) \quad (a)$$

$$= I(X; Y_1) + h_k(Y_2|Y_1) - h_k(Y_2|X_r), \quad (6.6)$$

where $(a)$ holds since $X_r = f_k(Y_r) = f_k(g(X, Y_1))$ and $(b)$ follows because given $X_r, Y_2$ is independent of $(X, Y_1)$.

- **Sufficient Conditions for Achieving Capacity:** The first term in the last row of $(6.6)$ is maximized by letting $X \sim p^*(x)$. Hence, one can see that it is sufficient to show that

$$\lim_{k \rightarrow \infty} h_k(Y_2|Y_1) = h(Y_2^*)$$

(6.7)

and that

$$\lim_{k \rightarrow \infty} h_k(Y_2|X_r) = h(Y_2^*|X_r^*).$$

(6.8)

Here $h(Y_2^*)$ and $h(Y_2^*|X_r^*)$ denote the values for $h(Y_2)$ and $h(Y_2|X_r)$ obtained when $X_r \sim p^*(x_r)$. Since $p^*(x_r) \in \mathcal{P}_r$ the cost constraint at the output of the relay will be fulfilled (with high probability for large enough $n$). Also, by evoking the dominated convergence theorem it is straightforward to prove
and we therefore focus on (6.7). Since \( Y_1 \rightarrow X_r \rightarrow Y_2 \) form a Markov chain, we have
\[
h_k(Y_1 | X_r) \leq h_k(Y_1 | Y_2) \leq h(Y_1).
\]
Because \( h_k(Y_2 | Y_1) = h_k(Y_1 | Y_2) + h_k(Y_2) - h(Y_1) \), it therefore suffices to show that

\[\text{i) } \lim_{k \to \infty} p_k(x_r) = p^*(x_r); \text{ and} \]

\[\text{ii) } \lim_{k \to \infty} h_k(Y_1 | X_r) = h(Y_1),\]

in order to prove (6.7).

- **Asymptotic behavior**: We next consider the behavior of \( p_k(x_r) \) and \( h_k(Y_1 | X_r) \) as \( k \to \infty \). We first recall the following two lemmas from Chapter 4.

**Lemma 6.2.2 (Generating uniform symbols at the relay)**. Let \( Y_r \) be a random variable with the density \( p(y_r) \) and \( T = s \left( \frac{Y_r}{\Delta} \right) \) where \( s(x) \) is a periodic function with period one such that \( s(x) = 2x \) for \( -\frac{1}{2} \leq x < \frac{1}{2} \). If the density \( p(y_r) \) is Riemann integrable, the distribution of \( T \) approaches the uniform distribution as \( \Delta \) goes to zero. That is
\[
\lim_{\Delta \to 0^+} p(t) = \begin{cases} 
\frac{1}{2}, & -1 \leq t < 1 \\
0, & \text{otherwise}
\end{cases}
\]  
(6.9)

**Lemma 6.2.3 (Generating independent symbols at the relay)**. Let \( U_r = g_{\Delta}(Y_r) \) where \( g_{\Delta} \) is a deterministic and periodic function with period \( \Delta \) which is invertible within each period. If the densities \( p(y_r) \) and \( p(y_1 | y_r) \) are Riemann integrable, then
\[
\lim_{\Delta \to 0^+} p(q_1 | u_r) = p(y_1).
\]  
(6.10)

**Fulfilling Condition i)**: Since \( F \) is the cdf of a random variable with probability density function \( p^*(x_r) \) that maximizes \( I(X_r, Y_2) \). Invoking Lemma 6.2.2, the distribution of \( F^{-1} \left( \frac{1}{2} s \left( \frac{Y_r}{\Delta} \right) + \frac{1}{2} \right) \) tends to \( p^*(x_r) \) as \( k \to \infty \).

**Fulfilling Condition ii)**: Define \( s_k(y_r) := \frac{1}{2} s \left( \frac{y_r}{\Delta_k} \right) + \frac{1}{2} \). We then observe that \( Y_1 \rightarrow S_k \rightarrow S_{k+1} \) form a Markov chain because \( S_{k+1} = 2S_k - \lfloor 2S_k \rfloor \) where \( \lfloor t \rfloor \) denotes the greatest integer less than or equal to \( t \). This yields
\[
h_k(Y_1 | S) \leq h_{k+1}(Y_1 | S) \leq h(Y_1) < \infty.
\]  
(6.11)

where \( h_k(Y_1 | S) := h(Y_1 | S_k) \). Thus \( \{h_k(Y_1 | S)\} \) is a bounded nondecreasing sequence. We therefore conclude that \( \lim_{k \to \infty} h_k(Y_1 | S) \) exists and equals to \( \sup_k \{h_k(Y_1 | S)\} \).
Finally, we note that $Y$ is arbitrarily close to $h$.

Next consider

$$\lim_{k \to \infty} h_k(Y_1|S) = \lim_{k \to \infty} \int p(y_r)h_k(Y_1|s(y_r))dy_r$$

\[= \int \lim_{k \to \infty} p(y_r)h_k(Y_1|s(y_r))dy_r \]

\[= \int p(y_r) \lim_{k \to \infty} \left\{ -p_k(y_1|s) \log p_k(y_1|s)dy_1 \right\} dy_r \]

\[\geq \int p(y_r) \lim_{k \to \infty} \left\{ \int_{-M}^M -p_k(y_1|s) \log p_k(y_1|s)dy_1 \right\} dy_r \]

\[= \int p(y_r) \left\{ \int_{-M}^M -p(y_1) \log p(y_1)dy_1 \right\} dy_r \]

\[= \int_{-M}^M -p(y_1) \log p(y_1)dy_1 \]  \hspace{1cm} (6.12)

where

(a) follows by the dominated convergence theorem since $0 \leq h_k(Y_1|s) \leq h(Y_1) < \infty$;

(b) holds for any $M > 0$, because $-x \log(x) \geq 0$ when $x \in (0,1]$;

(c) follows by the dominated convergence theorem since $-x \log(x)$ is bounded for $x \in (0,1]$;

(d) follows by Lemma 6.2.3.

For any $M > 0$, we thus have

$$h_M(Y_1) \leq \lim_{k \to \infty} h_k(Y_1|S) \leq h(Y_1),$$  \hspace{1cm} (6.13)

where $h_M(y_1) := -\int_{-M}^M p(y_1) \log p(y_1)dy_1$. If $Y_1$ has a finite range we can then choose $M$ such that $[-M,M]$ contains the range of $Y_1$ and hence $h_M(Y_1) = h(Y_1)$. Otherwise, since $h(Y_1)$ is bounded, we can choose $M$ such that $h_M(Y_1)$ is arbitrarily close to $h(Y_1)$. This essentially follows from

$$\lim_{m \to \infty} h_m(Y_1) = h(Y_1) \iff \forall \epsilon > 0 \exists M : |h_m(Y_1) - h(Y_1)| < \epsilon \text{ if } m \geq M.$$

Finally, we note that $Y_1 \to S_k \to X_r$ form a Markov chain and hence $h_k(Y_1|S) \leq h_k(Y_1|X_r) \leq h(Y_1)$. Thus

$$\lim_{k \to \infty} h_k(Y_1|X_r) = h(Y_1).$$
Hence, putting everything together, we obtain

$$
\lim_{k \to \infty} R_k = \max_{p(x) \in \mathcal{P}} I(X; Y_1) + \max_{p(x_r) \in \mathcal{P}_r} I(X_r; Y_2). \tag{6.14}
$$

This completes the proof.

**Remark 6.2.4.** The proposed relaying strategy can be interpreted as a one-dimensional modulo-lattice operation followed by nonlinear shaping. The nonlinear shaping is used in order to match the output of the modulo operation to the relay–destination link.

**Remark 6.2.5.** Note that Theorem 6.2.1 does not impose any condition on the structure of \( p(y_2|x_r) \). That is, the result is applicable to discrete as well as continuous relay–destination links.

### 6.3 Application to Gaussian Channels

We next present four examples of Gaussian relay channels where memoryless relaying is optimal. In the following, we assume that \( Y_2 = X_r + Z_2 \) where \( Z_2 \sim \mathcal{N}(0, N_2) \) and is independent of other random variables. We additionally assume that the source and the relay are operating under average power constraints, i.e., \( \mathbb{E}[X^2] \leq P_s \) and \( \mathbb{E}[X_r^2] \leq P_r \). The broadcast sub-channels \( p(y_1, y_r|x) \) are given as follows.

$$
Z_2 \sim \mathcal{N}(0, N_2)
$$

![Figure 6.2: Relay–destination link.](image)

**Example 1** (Relay channel with noiseless source–relay link): \( Y_1 = X + Z_1, Y_r = X \) where \( Z_1 \sim \mathcal{N}(0, N_1) \).

![Figure 6.3: Channel with noiseless source–relay link.](image)

**Example 2** (Relay channel with correlated noise [Kim08]): \( Y_1 = X + Z_1 \) and \( Y_r = X + Z_r \) where \( Z_1 \sim \mathcal{N}(0, N_1) \), \( Z_1 = kZ_r \), and \( k \neq 0 \) is a known constant.
6.3. Application to Gaussian Channels

\[ Z_r = kZ_1 \]
\[ Y = Z_1 \sim \mathcal{N}(0, N_1) \]
\[ Y_r = kZ_1 \]

Figure 6.4: Channel with correlated noise.

\[ S \sim \mathcal{N}(0, N_1) \]
\[ Y_1 = X + S \]
\[ Y_r = S \]

Figure 6.5: Channel with known interference at the relay.

Example 3 (Relay channel with interference): \( Y_1 = X + S \) and \( Y_r = S \) where \( S \sim \mathcal{N}(0, N_1) \).

Example 4 (Relay channel with interference and informed destination): \( Y_1 = X + Z_1 \) and \( Y_r = X + S \) where \( S \sim \mathcal{N}(0, Q) \) and is independent of \( Z_1 \sim \mathcal{N}(0, Z_1) \) and the destination knows \( S \).

\[ S \sim \mathcal{N}(0, Q) \]
\[ Y_1 = X + Z_1 \]
\[ Y_r = X + S \]

Figure 6.6: Channel with known interference at the destination.

Recall that we investigated the channel in Example 1 and Example 4 in Sections 4.3 and 4.4, respectively. We also considered the design of optimal instantaneous relaying schemes for the channel in Example 3 in Chapter 5 using a fixed point iteration method. In Chapter 5, we concluded that instantaneous relaying can perform very close to the capacity. In the present chapter, we in fact establish the optimality of instantaneous relaying.

Corollary 6.3.1. Instantaneous relaying achieves the capacity

\[ C = \frac{1}{2} \log \left( 1 + \frac{P_s}{N_1} \right) + \frac{1}{2} \log \left( 1 + \frac{P_r}{N_2} \right) \]

for the special cases in Examples 1–4.
Proof. Let $X \sim \mathcal{N}(0, P_s)$ and the relay mapping be

$$f_k^r(y_r) = \sqrt{P_r} F^{-1} \left( \frac{1}{2} s \left( \frac{y_r}{\Delta_k} \right) + \frac{1}{2} \right),$$  \hspace{1cm} (6.15)$$

where

- $\Delta_k := \Delta_0 2^{-k}$ and $\Delta_0 > 0$ is a constant;
- $s(x)$ is a periodic function with period one (i.e., $s(x+1) = s(x)$) and $s(x) = 2x$ for $-\frac{1}{2} \leq x < \frac{1}{2}$; and
- $F^{-1}(x)$ is the inverse of $F(x) = 1 - Q(x)$ where $Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \exp \left( -\frac{1}{2} t^2 \right) dt$.

By this choice of the relay mapping the density of $X_r$ approaches $\mathcal{N}(0, P_r)$. Now applying Theorem 6.2.1 gives the result as $k \to \infty$.

Linear relaying with $x_r = \kappa y_r$ where $\kappa$ is chosen to ensure the power constraint at the relay, is a suboptimal approach for the above examples. In fact one can verify the following result.

**Proposition 6.3.2.** The achievable rate of linear relaying for the channels in Examples 1–4 can have an unbounded gap to the capacity.

**Remark 6.3.3.** Note that it is essential to have an input distribution at the source or other random components in the channel such that $p(y_r)$ and $p(y_1 | y_r)$ are Riemann integrable, as assumed. As an example consider the channel given in Example 1, but with a channel input alphabet constrained to be finite. Then we cannot apply Theorem 6.2.1. For this case one approach that can yield good performance is constellation rearrangement, as proposed in [KL08]. However, instantaneous relaying is not necessarily optimal for a discrete source alphabet.

### 6.4 Summary and Concluding Remarks

In this chapter, we introduced a semi-deterministic class of relay channels and showed that the capacity of the channels in the class can be obtained using a relaying strategy with memoryless mapping.
7.1 Introduction

In this chapter, we introduce the class of state-decoupled relay channels and establish the capacity of some semi-deterministic cases of these channels. Our results generalize the capacity result reported in [CK07, Section VIII]. We also compute the capacity for two examples of relay channels with multiplicative fading and additive interference, respectively. Our examples do not fall in the classes studied in [GA82, CK07, Kim08].

As pointed out earlier in Chapter 2, the relay operates using strictly causal functions, i.e. $x(r_i(y_{i-1}))$, in the classical relay channel [vdM71, CG79]. The work of El Gamal, Hassanpour, and Mammen [GHM07] generalizes this notion to two other main classes: causal relaying where $x(r_i(y_i))$, i.e. the signal transmitted from the relay can also depend on the presently received signal in addition to the past ones, and noncausal relaying where $x(r_i(y^n))$, i.e. the signal transmitted from the relay at time instant $i$ can depend on the entire received signal at the relay. Table 7.1 summarizes these three types of relaying. As we discuss in the remainder of the chapter, we highlight that this notion is very insightful for the state-decoupled relay channels.

We, in this chapter, show that one can increase the capacity by relaxing the strict causality requirement at the relay. In particular, for a relay channel with Gaussian additive interference, we propose a hybrid digital–analog (HDA) relaying scheme that outperforms CF with strictly causal relaying. HDA relaying was proposed in [YS09] for the three-node Gaussian relay channel in slow fading and with orthogonal receive components. Using HDA coding in [YS09] was mainly motivated by achieving robustness toward the lack of channel state information at the relay, and new protocols were suggested for either dimension expansion or compression relaying. In contrast, in this chapter, we propose HDA relaying for the non-fading Gaussian relay channel. In the proposed scheme, the relay first employs digital compression to...
<table>
<thead>
<tr>
<th>Type of Relaying</th>
<th>Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strictly Causal</td>
<td>$x_r(y_r^{i-1})$</td>
</tr>
<tr>
<td>Causal</td>
<td>$x_r(y_r^i)$</td>
</tr>
<tr>
<td>Noncausal</td>
<td>$x_r(y_r^n)$</td>
</tr>
</tbody>
</table>

Table 7.1: Relay mappings.

generate a digital signal. The relay then linearly and component-wise combines the encoded compression index with the analog noisy received signal. Power splitting between the digital and analog parts is employed in order to optimize the scheme.

### 7.2 A Capacity Theorem

In this section, we define state-decoupled relay channels and present a capacity result for a semi-deterministic class of such channels.

**Definition 7.2.1.** The discrete memoryless relay channel $(X \times X_r \times S, p(y, y_r|x, x_r, s), Y \times Y_r)$ with a random state $s$ is said to be state-decoupled if the channel pmf is given by

$$p(y^n, y_r^n|x^n, x_r^n, s^n) = \prod_{i=1}^{n} p(y_i|x_i, x_r^i, s_i)p(y_r^i|s_i).$$

That is $(X_i, X_{ri}, Y_i) \rightarrow S_i \rightarrow Y_{ri}$ form a Markov chain. The channel state sequence $(s_1, s_2, \ldots, s_n)$ is assumed to be i.i.d. and it is determined by nature. See Figure 7.1 for an illustration of the channel.

In this chapter, we assume that the source has no knowledge of the channel state and the relay and the destination are informed about the channel state only through $y_r$ and $y$, respectively.

**Theorem 7.2.1.** If the relay channel is state-decoupled and $y_r = f(x, x_r, y)$, then the capacity with strictly causal relaying is given by

$$C = \max_{p(x)p(x_r)} \min \{I(X, X_r; Y), I(X; Y|X_r, Y_r)\}. \quad (7.2)$$

**Proof.** We first prove the positive part. Using the CF lower bound discussed in Section 2.3.6, the capacity is bounded as

$$C \geq R = \max_{p(x)p(x_r)p(y_r|y, x_r)} \min \{R_1, R_2\}, \quad (7.3)$$
7.3 A Fading Relay Channel

We next present an example of a semi-deterministic relay channel with a multiplicative fading. One of the main features of our example is that the fading coefficient is observed by the relay but it is transmitted to the destination over an interfering channel. This stands in contrast to previous examples in the literature where the knowledge of the channel state is conveyed to the destination over an orthogonal noisy, noiseless or rate-limited link.

7.3. A Fading Relay Channel

Figure 7.1: The state-decoupled three-node relay channel.

where

\[ R_1 = I(X, X_r; Y) - I(Y_r; \hat{Y}_r|X, X_r, Y), \] (7.4)
\[ R_2 = I(X; Y, \hat{Y}_r|X_r). \] (7.5)

Now let \( \hat{Y}_r = Y_r \). Since \( y_r = f(x, x_r, y) \), (7.3) simplifies to

\[ C \geq R = \max_{p(x)} \min \{ I(X, X_r; Y), I(X; Y|X_r, Y_r) \}. \] (7.6)

The converse follows by the cutset bound and noting that \( X_i \) is independent of \( X_{ir}(Y_{i-1}^r) \). This completes the proof.

Remark 7.2.2. Theorem 7.2.1 subsumes Theorem 3 in [CK07, Section VIII]. This essentially follows by the fact that the channel model in Theorem 7.2.1 includes the channel model in [CK07, Section VIII] as a special case. To see this, let \( y_r = s \), \( y = (y_1, y_2) \), and \( p(y|x_r, x, s) = p(y_1|x, s)p(y_2|x_r) \) in the general state-decoupled relay channel in Definition 7.2.1. Here \( y_1 \) is the signal received over the direct link from the source and \( y_2 \) is the signal received from the relay over an orthogonal link. Then without loss of generality we can replace the link from the relay to the destination with a noiseless link with the rate \( R_0 = \max_{p(x)} I(X_r; Y_2) \).
7.3.1 Channel Model

Consider the following relay channel with

\[ Y = SX + X_r, \]  
\[ Y_r = S, \]  

(7.7a) (7.7b)

where \( X, X_r, S \in \{\pm 1\} \) and

\[ \Pr\{S = +1\} = \Pr\{S = -1\} = \frac{1}{2}. \]

This example models an antipodal signaling with uniform phase fading at high signal-to-noise ratio. We assume that the source has no knowledge of \( s \) and the relay knows \( s \) through \( y_r \). The knowledge of the channel state is transmitted from the relay to the destination over a common channel shared by the source. That is, there is no orthogonal channel devoted to convey the channel state information to the destination.

7.3.2 Direct-Link Transmission

We first note that the direct-link transmission discussed in Section 2.3.2, does not amount to any positive reliable rate since

\[ R = \max_{p(x)} I(X;Y|X_r = x_r) \] 
\[ = \max_{p(x), x_r} \{H(Y|X_r = x_r) - H(Y|X, X_r = x_r)\} \] 
\[ = \max_{p(x)} H(SX) - H(S) \] 
\[ = 0. \]  

(7.8)

While the direct-link transmission fails, we next show that one can reliably transmit 0.5 bits by appropriately incorporating the relay.

7.3.3 Capacity of Strictly Causal Relaying

Proposition 7.3.1. The capacity of the relay channel described by (7.7) with strictly causal relaying is \( C = 0.5 \) bits per channel use.

Proof. From (7.7), we have \( Y_r = (Y - X_r)/X \). Therefore the capacity of the channel is given by Theorem 7.2.1. In order to compute the capacity, let \( \Pr\{X_r = +1\} = p, \)


<table>
<thead>
<tr>
<th>$y$</th>
<th>0</th>
<th>+2</th>
<th>−2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Pr{Y = y}$</td>
<td>$\frac{1}{1}$</td>
<td>$\frac{1}{1}$</td>
<td>$\frac{1}{1}$</td>
</tr>
</tbody>
</table>

Table 7.2: The pmf of $Y$.

\[ \Pr\{X_r = -1\} = 1 - p, \Pr\{X = +1\} = t, \Pr\{X = -1\} = 1 - t. \]

Then consider

\[
I(X;Y|X_r, Y_r) = H(Y|X_r, S) - H(Y|X, X_r, S) \\
= H(SX + X_r|X_r, S) - H(SX + X_r|X, X_r, S) \\
= H(X) \\
= \mathbb{H}_b(t), \tag{7.9}
\]

where $\mathbb{H}_b(t) := -t \log_2(t) - (1 - t) \log_2(1 - t)$ denotes the binary entropy function. Similarly consider

\[
I(X, X_r; Y) = H(Y) - H(Y|X, X_r) \\
= H(Y) - H(SX + X_r|X, X_r) \\
= H(Y) - H(S) \\
= \frac{1}{2} \mathbb{H}_b(p), \tag{7.10}
\]

where the last equality follows because of the pmf of $Y$ given in Table 7.2.

Combining (7.9) and (7.10) yields

\[
C = \max_{t,p} \min \left\{ \mathbb{H}_b(t), \frac{1}{2} \mathbb{H}_b(p) \right\} = 0.5, \tag{7.11}
\]

which completes the proof.

7.3.4 Capacity of Causal and Noncausal Relaying

In this section, we assume that the signal transmitted from the relay at time instant $i$ can also depend on the presently received signal at the relay, i.e. $x_{ri}(y^l_r)$. We next give a simple deterministic code that achieves the capacity of 1 bit per channel use using instantaneous relaying. Therefore instantaneous relaying outperforms strictly causal relaying for this example.

**Proposition 7.3.2.** The capacity of the relay channel given by (7.7) where $x_{ri}(y^l_r)$, $\forall l \in [i, n]$ is $C = 1$ bit per channel use.

**Proof.** The converse is immediate since $H(X) \leq 1$. We next prove the achievability. Let the source uniformly choose its symbol from the set $\{\pm 1\}$ and let relay use the
X | S | Xr = S | Y = S(X + 1)
---|---|---|---
+1 | +1 | +1 | +2
+1 | -1 | -1 | -2
-1 | +1 | +1 | 0
-1 | -1 | -1 | 0

Table 7.3: Instantaneous Relaying.

instantaneous mapping \( X_{ri} = S_i \). Then the inputs and outputs of the channel are given in Table 7.3.

Now it is easy to observe that the destination can recover \( X \) from \( Y \) without error. Thus 1 bit can be transmitted. Note that considering future received signals at the relay, i.e. \( y_{r,i+1} \) does not buy us any gain for this particular example.

7.4 A Relay Channel with Additive Interference

We next present an example of a semi-deterministic relay channel with additive interference and compute the capacity with strictly causal relaying. We also discuss causal and noncausal relaying schemes and show that one can achieve higher rates by causal relaying where the relay employs an HDA relaying scheme.

7.4.1 Channel Model

Consider the relay channel with

\[
Y = X + S + X_r, \tag{7.12a}
\]
\[
Y_r = S, \tag{7.12b}
\]

where \( S \sim \mathcal{N}(0, Q) \). Further assume that the source and the relay operate under average power constraints: \( \mathbb{E}[X^2] \leq P \) and \( \mathbb{E}[X_r^2] \leq P_r \). See Figure 7.2 for an illustration. This example models an interference-limited scenario in which additive noises at the relay and the destination are neglected. We assume that the source does not know the interference. In this example the relay perfectly observes the interference, but what it transmits also creates an interference at the destination. Recall that we investigated a similar model but with an orthogonal link from the relay to the destination in Section 5.6.

7.4.2 Capacity of Strictly Causal Relaying

**Proposition 7.4.1.** The capacity of the relay channel described by (7.12) with \( S \sim \mathcal{N}(0, Q) \) and strictly causal relaying is given by

\[
C = C \left( \frac{P + P_r}{Q} \right), \tag{7.13}
\]
where $C(x) := \frac{1}{2} \log_2(1 + x)$. 

Proof. We note that the received signal at the relay can be constructed from $X, X_r, \text{ and } Y$. That is $S = Y - X - X_r$. Thus we can apply Theorem 7.2.1. The proof then follows by noting that (7.2) is optimized by choosing $X \sim \mathcal{N}(0, P)$ and $X_r \sim \mathcal{N}(0, P_r)$.

Remark 7.4.2. The relay channel shown in Figure 7.2 can be generalized as follows. Let

\begin{align*}
Y &= f_1(X, X_r) + f_2(S), \quad \text{(7.14a)} \\
Y_r &= f_3(S), \quad \text{(7.14b)}
\end{align*}

where $f_1$ and $f_3$ are arbitrary functions, $f_2$ is an invertible function and $S$ denotes the channel state with an arbitrary distribution. Note that the relay channel given in (7.14) is state-decoupled and

$$
Y_r = f_3 \left( f_2^{-1}(Y - f_1(X, X_r)) \right) =: f(X, X_r, Y). \quad \text{(7.15)}
$$

Thus the capacity of the channel is achieved by CF and is given in (7.2).

7.4.3 Causal Relaying

The capacity of the relay channel can be increased when the relay’s output can also depend on the presently received signal at the relay.

Proposition 7.4.3. The rate

$$
R_L = \begin{cases} 
C \left( \frac{P}{\sqrt{Q - P_r}} \right) & \text{if } Q > P_r \\
\infty & \text{if } Q \leq P_r
\end{cases} \quad \text{(7.16)}
$$

is achievable.
Proof. If $Q \leq P_r$ the relay can perfectly cancel the interference and hence arbitrarily large rate is achievable. Now assume that $Q > P_r$ and let $X_{ri} = -\sqrt{\kappa}S_i$ where $\kappa := \frac{P_r}{Q}$ is chosen to ensure the power constraint at the relay. Thus $R_L$ is achievable where

\[ R_L = I(X; Y) \]
\[ = h(X + (1 - \sqrt{\kappa})S) - h((1 - \sqrt{\kappa})S) \]
\[ \leq C \left( \frac{P}{(\sqrt{Q} - \sqrt{P})^2} \right), \tag{7.17} \]

where the upper bound is achieved by choosing $X \sim \mathcal{N}(0, P)$.

As we next show, instantaneous relaying can be improved by utilizing HDA relaying scheme where the digital part is used to quantize the relay’s observation.

**Proposition 7.4.4.** The rate

\[ R_H = \begin{cases} 
\max_{\gamma \in [0, 1]} C \left( \frac{P + \gamma P_r}{\sqrt{Q} - \sqrt{P}} \right) & \text{if } Q > P_r \\
\infty & \text{if } Q \leq P_r 
\end{cases} \tag{7.18} \]

is achievable where $\bar{\gamma} := 1 - \gamma$.

Proof. Similar to Proposition 7.4.3, if $Q \leq P_r$ arbitrarily large rate is achievable. Now assume that $Q > P_r$. Using the hybrid CF lower bound in [GHM07], the capacity is bounded as

\[ C \geq R_H = \max_{p(x)p(v)p(y_r, v), x_r = f(y_r, v)} \min \{ R_1, R_2 \}, \tag{7.19} \]

where

\[ R_1 = I(X, V; Y) - I(Y_r; \hat{Y}_r|X, X), \tag{7.20} \]
\[ R_2 = I(X; Y, \hat{Y}_r|V). \tag{7.21} \]

The proof then follows by choosing $X \sim \mathcal{N}(0, P)$, $\hat{Y}_r = Y_r + Z_q$ where $Z_q \sim \mathcal{N}(0, N_q)$ is independent of other random variables, and

\[ f(v, y_r) = \sqrt{\bar{\gamma}}v - \sqrt{\kappa}y_r, \tag{7.22} \]

where $V \sim \mathcal{N}(0, P_r)$, $\gamma \in [0, 1]$ is a power splitting factor and $\kappa := \frac{P_r}{Q}$ is a power normalization constant.

\[^2\text{The rate in [GHM07, Eq. (18)] is mistakenly optimized over } p(x, v).\]
7.4.4 Receiver Noise

Now assume that the received signal at the destination is noisy and it is given by 
\[ Y = X + S + X_r + Z, \]
where \( Z \sim \mathcal{N}(0, N) \) is independent of other random variables. Similar to the noiseless case, we can compute the capacity bounds.

- **Lower Bounds:**
  
  1. CF with Strictly Causal Relaying:
     \[
     R_{\text{CF}} = C \left( \frac{P(P + P_r + N)}{(P + N)(N + Q) + P_r N} \right)
     \]  
     (7.23)
  
  2. Instantaneous Linear Relaying:
     \[
     R_{\text{IL}} = \begin{cases} 
     C \left( \frac{P}{(\sqrt{Q} - \sqrt{P_r})^2 + N} \right) & \text{if } Q > P_r \\
     C \left( \frac{P}{N} \right) & \text{if } Q \leq P_r 
     \end{cases}
     \]  
     (7.24)
  
  3. HDA Relaying:
     \[
     R_{\text{HDA}} = \begin{cases} 
     \max_{\gamma \in [0,1]} R & \text{if } Q > P_r \\
     C \left( \frac{P_r}{\gamma N} \right) & \text{if } Q \leq P_r 
     \end{cases}
     \]  
     (7.25)

  where
  \[
  R = C \left( \frac{P(P + \gamma P_r + N)}{(P + N) \left( (\sqrt{Q} - \sqrt{\gamma P_r})^2 + N \right) + \gamma P_r N} \right).
  \]  
  (7.26)

- **Upper Bounds:**
  
  1. Cutset Upper Bound:
     
     The evaluation of the cutset bound in (2.6) for strictly causal relaying gives
     \[
     C \leq \min \left\{ C \left( \frac{P}{N} \right), C \left( \frac{P + P_r}{Q + N} \right) \right\}.
     \]  
     (7.27)

  2. Interference-Free Upper Bound:
     
     The capacity of the channel with noncausal relaying is less than that of the counterpart channel when there is no interference. That is
     \[
     C_{\text{NC}} \leq C \left( \frac{P}{N} \right).
     \]  
     (7.28)

\[^3\text{Note that there is a misprint in [GHM07, Eq. (7)].} \]
7.4.5 Noncausal Relaying

Wang and Naghshvar in [WN11] have recently proposed an extension of CF by employing Gelfand-Pinsker type of encoding at the relay for the noncausal case. In [WN11] it is shown that the rate

\[ R_{HNC} = \max_{p(x)p(u|y_r)} \min \{ R_1, R_2 \}, \]

is achievable where

\[ R_1 = I(X, U; Y) - I(Y_r; U|X), \]
\[ R_2 = I(X; U, Y). \]

We next show that the achievable rate \( R_{HNC} \) does not improve the causal achievable rate \( R_H \) given in (7.19).

**Proposition 7.4.5.** \( R_{HNC} \leq R_H \).

**Proof.** Without loss of generality let \( U = (V, \hat{Y}_r) \) where \( V \sim p(v) \) is independent of other random variables and \( \hat{Y}_r \sim \sum_{y_r} p(\hat{y}_r|v,y_r)p(y_r)p(v) \). This yields

\[ R_1 = I(X, U; Y) - I(Y_r; U|X) \]
\[ = I(X, V, \hat{Y}_r; Y) - I(Y_r; V, \hat{Y}_r|X) \]
\[ = I(X, V, \hat{Y}_r; Y) - I(Y_r; V|X) - I(Y_r; \hat{Y}_r|X, V) \]
\[ = I(X, V; Y) + I(\hat{Y}_r; Y|X, V) - I(Y_r; \hat{Y}_r|X, V) \]
\[ = I(X, V; Y) - H(\hat{Y}_r|Y, X, V) + H(\hat{Y}_r|Y_r, X, V) \]
\[ = I(X, V; Y) - I(\hat{Y}_r, Y_r|Y, X, V), \]

and

\[ R_2 = I(X; U, Y) = I(X; V, \hat{Y}_r, Y) = I(X; Y, \hat{Y}_r|V). \]

Thus the rate

\[ R_{HNC} = \max_{p(x)p(v)p(\hat{y}_r|v,y_r), x_r = f(v,y_r, \hat{y}_r)} \min \{ R_1, R_2 \}, \]

is achievable. Note that since \( U = (V, \hat{Y}_r) \), it suffices to restrict \( \hat{Y}_r \) to be a deterministic function of \( V \) and \( Y_r \), i.e. \( \hat{Y}_r = g(V, Y_r) \) in order to get all possible mappings.
7.4. A Relay Channel with Additive Interference

\[ p(u|y_r). \] This yields

\[
R_{\text{HNC}} = \max_{p(x)p(v)g(\hat{y}_r|v,y_r), x_r = f(v,y_r)} \min \{R_1, R_2\} \\
= \max_{p(x)p(v)g(\hat{y}_r|v,y_r), x_r = f(v,y_r)} \min \{R_1, R_2\} \\
\leq \max_{p(x)p(v)\hat{y}_r|v,y_r), x_r = f(v,y_r)} \min \{R_1, R_2\} \\
= R_H. \quad (7.35)
\]

This completes the proof.

\[ \square \]

**Remark 7.4.6.** Comparing (7.19) with (7.34), we observe that both rates are the same except the generation of \( x_r \). In noncausal relaying the relay can also incorporate the reproduction symbol \( \hat{y}_r \) to generate \( x_r \). This is not however feasible in strictly causal or causal relaying. See Figures 7.3–7.5 for schematic descriptions of strictly causal, causal and noncausal relaying based on CF.

We note that \( R_{\text{HNC}} \leq R_H \) is valid if the rates are optimally evaluated. For example for Gaussian relay channels, the optimal evaluation of CF is not known. We next compute an achievable rate for the channel in Section 7.4.4 using the lower
bound in (7.34). Let $X \sim \mathcal{N}(0, P)$, $V \sim \mathcal{N}(0, 1)$, $\hat{Y}_r = Y_r + Z_q$ where $Z_q \sim \mathcal{N}(0, N_q)$ is independent of other random variables, and

$$X_r = f(V, Y_r, \hat{Y}_r) = -\alpha Y_r + \beta Z_q + \gamma V,$$

(7.36)

where $\alpha := \sqrt{\frac{aP_r}{Q}}$, $\beta := \sqrt{\frac{bP_r}{N_q}}$, $\gamma := \sqrt{(1 - a - b)P_r}$. The parameters $a$ and $b$ are power splitting factors such that $a \in [0, 1]$, $b \in [0, 1]$, and $a + b \leq 1$. This yields

$$R_{HNC} = \begin{cases} \max_{a,b \in [0,1], a+b \leq 1} \min \{R_1, R_2\} & \text{if } Q > P_r \\ C\left(\frac{P}{Q}\right) & \text{if } Q \leq P_r \end{cases}$$

(7.37)

where

$$R_1 = C\left(\frac{P(Q + N_q)}{(1 - \alpha)^2 QN_q + (\beta^2 N_q + N)(Q + N_q)}\right),$$

(7.38)

and

$$R_2 = C\left(\frac{N_q \left( P + (1 - \alpha)^2 Q + \beta^2 N_q + N + \gamma^2 \right)}{(1 - \alpha)^2 + \beta^2} QN_q + N(Q + N_q) - 1\right).$$

(7.39)

**7.4.6 Numerical Examples**

Figure 7.6 illustrates the capacity bounds for the semi-deterministic relay channel in Figure 7.2. We plot the capacity of strictly causal relaying (Eq. (7.13)), the achievable rate of linear instantaneous relaying (Eq. (7.16)) and HDA scheme (Eq. (7.18)) as a function of the source power $P$. The channel parameters are set to $Q = 20$ and $P_r = 2$. We observe that the achievable rate of the hybrid scheme is strictly larger than the capacity of strictly causal relaying. When the source has a low power, it is beneficial to use hybrid relaying. In fact when the source has a negligible power, linear instantaneous relaying performs poor.

Figures 7.7–7.9 illustrate the capacity bounds for the relay channel in Figure 7.2 but with receiver noise. We plot the achievable rate of strictly causal relaying (Eq. (7.23)), the achievable rate of linear instantaneous relaying (Eq. (7.24)), HDA scheme (Eq. (7.25)) and cutset bound for strictly causal relaying (Eq. (7.27)) as a function of the source power $P$. We set $Q = 100$, $P_r = 10$, and $N = 1$ in Figure 7.7. The same channel parameters but with $Q = 20$ are set for Figure 7.8. From Figures 7.7 and 7.8, we see that the relay channel with the causal operation can outperform the cutset bound. We also observe that for strong interference, the proposed hybrid coding is beneficial. Figure 7.9 plots the capacity bounds as function of $P_r$ when the channel parameters are set to $P = Q = N = 10$. We see that for this case the linear relaying performs as good as hybrid relaying. We finally note that for this channel (including noisy and noiseless), the achievable rate in (7.37) obtained by noncausal relaying does not give any gain over the causal counterpart.
We introduced the class of state-decoupled relay channels and established the capacity for some semi-deterministic cases with strictly causal relaying. We proved that capacity is achieved using CF. By constructing some examples, we showed that one can enlarge the capacity by causal relaying. In particular we proposed a HDA relaying protocol where the digital signal is linearly combined with the analog received signal at the relay for a Gaussian relay channel. Table 7.4 summarizes the achievable rates of CF relaying with strictly causal, causal and noncausal relaying. The random variable $Q$ denotes the time-sharing variable. The time-sharing is used since the objective function is not convex in general [GMZ06].

<table>
<thead>
<tr>
<th>Type of Relaying</th>
<th>Achievable Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strictly Causal $x_{r1}(y_r^{i-1})$</td>
<td>$R_{CF} = \max_{p(q)p(x</td>
</tr>
<tr>
<td>Causal $x_{ri}(y_r^i)$</td>
<td>$R_H = \max_{p(q)p(x</td>
</tr>
<tr>
<td>Noncausal $x_{ri}(y_r^n)$</td>
<td>$R_{HNC} = \max_{p(q)p(x</td>
</tr>
</tbody>
</table>

Table 7.4: Achievable rates using CF for strictly causal, causal and noncausal relaying.

**7.5 Summary and Concluding Remarks**

We introduced the class of state-decoupled relay channels and established the capacity for some semi-deterministic cases with strictly causal relaying. We proved that capacity is achieved using CF. By constructing some examples, we showed that one can enlarge the capacity by causal relaying. In particular we proposed a HDA relaying protocol where the digital signal is linearly combined with the analog received signal at the relay for a Gaussian relay channel. Table 7.4 summarizes the achievable rates of CF relaying with strictly causal, causal and noncausal relaying. The random variable $Q$ denotes the time-sharing variable. The time-sharing is used since the objective function is not convex in general [GMZ06].
Figure 7.6: Comparison of the capacity of strictly causal relaying and the achievable rates of instantaneous and hybrid relaying for the semi-deterministic relay channel in Figure 7.2. The channel parameters are set to $Q = 20$ and $P_r = 2$. We have also plotted the rate of only-direct-link transmission when the relay is off, i.e. $x_r = 0$. 
Figure 7.7: Comparison of the capacity of strictly causal relaying and the achievable rates of instantaneous and hybrid relaying for relay channel in Figure 7.2 but with the receiver noise. The channel parameters are set to $Q = 100$, $P_r = 10$, and $N = 1$. We have also plotted the rate of only-direct-link transmission when the relay is off, i.e. $x_r = 0$. 
Figure 7.8: Same as Figure 7.7 but for $Q = 20$. 
Figure 7.9: Same as Figure 7.7 but for $P = Q = N = 10$. 
Chapter 8

The Relay Channel with Noncausal State Information

8.1 Introduction

We study the three-node relay channel [vdM71, CG79] consisting of a source, a relay and a destination with state in this chapter. We assume that the channel state is determined by a random parameter, and that noncausal knowledge of the state is partially available at the nodes. We studied a similar model but with causal knowledge of the state in [KS09d]. In this chapter, we first derive an achievable rate inspired by the noisy network coding (NNC) approach in [LKGC11] and the random binning scheme in [GP80]. We then derive an equivalent representation of the rate similar to that in [GMZ06] and show that the proposed NNC-based scheme achieves the same rate that can be achieved by compress-and-forward for the relay channel with a random state as proposed by [AMA10].

The remainder of the chapter is organized as follows. Section 8.2 presents the system model and the definitions. Section 8.3 presents main results of the chapter and comments on them. Section 8.4 and 8.5 give the proofs of the results presented in Section 8.3. Section 8.6 concludes the chapter.

8.2 Definitions

We next present basic definitions and assumptions on the channel model.

Definition 8.2.1. The discrete memoryless, three-node relay channel with a random parameter is denoted by

\[(X_1 \times X_2, p(y_2, y_3|x_1, x_2, s), Y_2 \times Y_3) \cdot (S, p(s_1, s_2, s_3|s) p(s), S_1 \times S_2 \times S_3)\]  

where
- \(x_1 \in X_1\) denotes the source’s input;
- \(x_2 \in X_2\) denotes the relay’s input;
The Relay Channel with Noncausal State Information

- \( y_2 \in \mathcal{Y}_2 \) denotes the output of the channel at the relay
- \( y_3 \in \mathcal{Y}_3 \) denotes the output of the channel at the destination;
- \( s \in \mathcal{S} \) denotes the random state of the channel;
- \( s_1 \in \mathcal{S}_1 \) denotes the knowledge of the state of the channel at the source;
- \( s_2 \in \mathcal{S}_2 \) denotes the knowledge of the state of the channel at the relay;
- \( s_3 \in \mathcal{S}_3 \) denotes the knowledge of the state of the channel at the destination;
- \( p(y_2, y_3|x_1, x_2, s) \) denotes the pmf modelling the interaction between the variables \( (x_1, x_2, s, y_2, y_3) \), whose \( n \)-extension is given by

\[
p(y_{21}^n, y_{31}^n|x_{11}^n, x_{21}^n, s^n) = \prod_{i=1}^{n} p(y_{2i}, y_{3i}|x_{1i}, x_{2i}, s_i),
\]

- \( p(s_1, s_2, s_3|s)p(s) \) denotes the pmf modelling the interaction between the random state of the channel and the knowledge of the state at the different nodes. The \( n \)-extension is given by

\[
p(s^n, s_{11}^n, s_{21}^n, s_{31}^n) = \prod_{i=1}^{n} p(s_{1i}, s_{2i}, s_{3i}|s_i)p(s_i).
\]

See Figure 8.1 for an illustration of the relay channel with the random state \( s \).
Definition 8.2.2. A \((2^nR, n)\) code for the relay channel with a random parameter consists of the following three main components:

- an encoder that maps the message \(M\) uniformly drawn from the set \(\mathcal{M}^{(n)} := [1 : 2^nR]\) to \(X^n_i\) according to \(\alpha: \mathcal{M}^{(n)} \times S^n_{11} \to X^n\) such that \(X_{1i} = \alpha(M, S^n_{11})\) \(\forall i \in [1 : n]\). That is, the encoder incorporates the noncausal knowledge of the channel state.

- a set of relay functions: \(\{f_i\}_{i=1}^n\) such that \(\forall i \in [1 : n], X_{2i} = f_i(Y_{21}^{i-1}, S^n_{21})\). That is, the relay acts strictly causally on the received signals but it incorporates noncausal knowledge of the channel state.

- a decoder that maps the received signal \(Y^n_{31}\) to an estimate of the transmitted message \(\hat{M}\) according to \(\beta: Y^n_{31} \times S^n_{31} \to \mathcal{M}^{(n)}\) where \(\hat{M} = \beta(Y^n_{31}, S^n_{31})\).

Remark 8.2.3. Since we assume block decoding at the destination, assuming causal knowledge of the channel state at the destination does not affect the results.

Definition 8.2.4. The rate \(R\) is said to be achievable if there exists a sequence of communication strategies

\[ \mathcal{S}_n := (\alpha^{(n)}, \{f_i\}_{i=1}^n, \beta^{(n)}) \]

such that the average message error probability \(p_{e}^{(n)} := \Pr\{\hat{M} \neq M\}\) goes to zero as \(n \to \infty\).

8.3 Main Results

We next present the main results of the chapter.

Proposition 8.3.1. The rate

\[ R = \sup \min \{R_1, R_2\}, \]  

is achievable using NNC, where

\[ R_1 = I(U_1; \hat{Y}_2, Y_3, S_3|U_2) - I(U_1; S_1|U_2), \]  

\[ R_2 = I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2|U_1, U_2, Y_3, S_3) - I(U_1, U_2; S_1, S_2), \]

and the supremum in (8.2) is taken over the pmf

\[ p(x_1, u_1|s_1)p(x_2, u_2|s_2)p(\hat{y}_2|y_2, u_2, s_2). \]

Remark 8.3.2. Compression of both the received signal \(Y_2\) and the knowledge of the channel state \(S_2\) at the relay is considered in obtaining the rate in (8.2).
Remark 8.3.3. If the source and the relay causally know the channel state, then the rate in (8.2) using NNC is achievable where

\[ R_1 = I(U_1; \hat{Y}_2, Y_3, S_3 | U_2), \]  
\[ R_2 = I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2 | U_1, U_2, Y_3, S_3), \]

and the supremum is taken over

\[ p(u_1)p(x_1 | u_1, s_1)p(u_2)p(x_2 | u_2, s_2)p(\hat{y}_2 | y_2, u_2, s_2). \]

8.3.1 A Cutset Interpretation of \( R \)

- **Broadcast cut**: The relay acts as a virtual receiver and the rate \( R_1 \) is achievable. The relay only provides \( \hat{Y}_2 \), i.e. a compressed version of what it observes. The source pays a penalty due to the binning, which is captured in \( I(U_1; S_1 | U_2) \).

- **Multiple-access cut**: The relay acts as a virtual transmitter which is reflected in \( R_2 = I(U_1, U_2; Y_3, S_3) \). However, the rate is penalized by the amount of signal compression at the relay which is captured in \( I(\hat{Y}_2; Y_2, S_2 | U_1, U_2, Y_3, S_3) \) and by the amount of binning at the source and the relay captured in \( I(U_1, U_2; S_1, S_2) \).

8.3.2 Connection to Compress-and-Forward

**Proposition 8.3.4.** The achievable rate in (8.2) is equivalent to

\[ R \leq \sup I(U_1; \hat{Y}_2, Y_3, S_3 | U_2) - I(U_1; S_1 | U_2), \]  
\[ \text{s.t. } I(\hat{Y}_2; Y_2, S_2 | U_1, U_2, Y_3, S_3) \leq I(U_2; Y_3, S_3) - I(U_2; S_2), \]

where the supremum is taken over the pmf given in (8.5).

From Proposition 8.3.4, we conclude that both NNC and compress-and-forward (CF) give the same achievable rate for the relay channel with a random state.\(^1\) We note that, in contrast to CF, NNC does not utilize the Wyner-Ziv coding at the relay [LKGC11].

The remainder of the chapter is devoted to the proof of Propositions 8.3.1 and 8.3.4.

\(^1\)CF actually can obtain higher rates than the one given in [AMA10]. This is because Eq. (30) in [AMA10] should be \( I(U_1; Y_3, \hat{Y}_2, U_2) \), since \( U_1 \) and \( U_2 \) are correlated. This is because of the fact that \( S_1 \) and \( S_2 \) are correlated through \( S \). Thus, Eq. (1) in [AMA10] changes to (8.8) in this chapter.
8.4 Proof of Proposition 8.3.1

8.4.1 Codebook Generation

Fix the pmf
\[ p(u_1|s_1)p(x_1|u_1, s_1)p(u_2|s_2)p(x_2|u_2, s_2)p(\hat{y}_2|y_2, u_2, s). \]

Then for each block \( j \in [1 : b] \), randomly and independently generate \( 2^{nR} \times 2^{nR_1} \) sequences

\[ u_{1j}(m, \tilde{l}_{1j}) \sim \prod_{i=1}^{n} p_{U_1}(u_{1,(j-1)n+i}), \ \forall m \in [1 : 2^{nR}], \ \forall \tilde{l}_{1j} \in [1 : 2^{nR_1}] \quad (8.9) \]

Similarly, randomly and independently generate \( 2^{nR_2} \times 2^{nR_2} \) sequences

\[ u_{2j}(l_{2,j-1}, \tilde{l}_{2j}) \sim \prod_{i=1}^{n} p_{U_2}(u_{2,(j-1)n+i}), \ \forall l_{2,j-1} \in [1 : 2^{nR_2}], \ \forall \tilde{l}_{2j} \in [1 : 2^{nR_2}] \]

\[ (8.10) \]

For each \( u_{2j}(l_{2,j-1}, \tilde{l}_{2j}) \), \( l_{2,j-1} \in [1 : 2^{nR_2}], \tilde{l}_{2j} \in [1 : 2^{nR_2}] \) randomly and conditionally independent generate \( 2^{nR_2} \) sequences

\[ \hat{y}_{2j}(l_{2,j-1}, \tilde{l}_{2j}) \sim \prod_{i=1}^{n} p_{\hat{Y}_2}(\hat{y}_{2,(j-1)n+i}|u_{2,(j-1)n+i}(l_{2,j-1}, \tilde{l}_{2j})). \]

\[ (8.11) \]

8.4.2 Encoding

We next explain the encoding at the beginning of block \( j \in [1 : b] \). Let \( m \in [1 : 2^{nR}] \) be the message to be sent. The source node looks for the smallest index \( l_{1j} \in [1 : 2^{nR_1}] \) such that

\[ (u_{1j}(m, \tilde{l}_{1j}), s_{1j}) \in T_{e_1}^{(n)}. \]

\[ (8.12) \]

If there is no such index, it picks one at random.

At the end of block \( j - 1 \), the relay node knows \( s_{2,j-1}, s_{2j}, l_{2,j-2}, \) and \( \tilde{l}_{2,j-1} \). The relay node looks for the smallest index such that

\[ (\hat{y}_{2,j-1}(l_{2,j-1}|l_{2,j-2}, \tilde{l}_{2,j-1}), y_{2,j-1}, s_{2,j-1}, u_{2,j-1}(l_{2,j-2}, \tilde{l}_{2,j-1})) \in T_{e_2}^{(n)}. \]

\[ (8.13) \]

where \( l_{2,0} = 1 \) by convention. If there is no such index, it picks one at random.
Let $\epsilon > \max \{\epsilon_1, \epsilon_2, \epsilon_3\}$. The destination performs the decoding at the end of block $b$. The decoder looks for a unique index $\hat{m} \in [1 : 2^{bR_1}]$ such that
\[
\left( \bar{u}_{1j}(\hat{m}, \hat{l}_{1j}), \bar{y}_{1j}(\hat{m}, \hat{l}_{1j}, \hat{l}_{2j}), \bar{u}_{2j}(\hat{m}, \hat{l}_{1j}, \hat{l}_{2j}), \bar{y}_{2j}(\hat{m}, \hat{l}_{1j}, \hat{l}_{2j}) \right) \in T_{\epsilon}^{(n)},
\]
for some $\hat{l}_{1j}, \hat{l}_{2j}$, and for all $j \in [1 : b]$. (Table 8.1 summarize the encoding and decoding over $b$ blocks)

### 8.4.4 Probability of Error

Let $M = 1$ denote the message sent from the source node and $L_{2j}$ denote indices chosen at the relay in the block $j \in [1 : b]$. Now define the following events
\[
\mathcal{E}_1 := \bigcup_{j=1}^{b} \left\{ \left( \bar{u}_{1j}(1, l_{1j}), \bar{s}_{1j} \right) \notin T_{\epsilon}^{(n)} \right\},
\]
\[
\mathcal{E}_2 := \bigcup_{j=1}^{b} \left\{ \left( \bar{u}_{2j}(L_{2j} - 1, \hat{l}_{2j}), \bar{s}_{2j} \right) \notin T_{\epsilon}^{(n)} \right\},
\]
\[ \mathcal{E}_3 := \bigcup_{j=1}^b \left\{ (\tilde{y}_{2j}|l_{2,j-1}, \tilde{l}_{2j}), \tilde{y}_{2j}, s_{2,j}, u_{2,j}(L_{2,j-1}, \tilde{l}_{2j}) \not\in \mathcal{T}_e(n) \right\}, \quad (8.20) \]

\[ \mathcal{E}_{4m} := \left\{ (u_{1j}(m, \hat{l}_{1j}), \tilde{y}_{2j}|l_{2,j-1}, \hat{l}_{2j}), u_{2j}(l_{2,j-1}, \hat{l}_{2j}), \tilde{y}_{3j}, s_{3j}) \in \mathcal{T}_e(n) \right\} \quad \text{for some } \hat{l}_{1j}, \hat{l}_{2j}, \hat{l}_{2j}. \quad (8.21) \]

Then, the probability of error can be bounded as

\[ P(n)(\mathcal{E}) \leq P(n)(\mathcal{E}_1) + P(n)(\mathcal{E}_2) + P(n)(\mathcal{E}_3) + P(n)(\mathcal{E}_{41} \cap \mathcal{E}_{41} \cap \mathcal{E}_{42} \cap \mathcal{E}_{43}) + \]

\[ P(n)(\cup_{m \neq 1} \mathcal{E}_{4m}). \quad (8.22) \]

Using the covering lemma given in Appendix 2.A, we have

\[ P(n)(\mathcal{E}_1) \to 0 \text{ as } n \to \infty, \text{ if } \tilde{R}_1 > I(U_1; S_1) + \delta(\epsilon_1), \quad (8.23) \]

\[ P(n)(\mathcal{E}_2) \to 0 \text{ as } n \to \infty, \text{ if } \tilde{R}_2 > I(U_2; S_2) + \delta(\epsilon_2), \quad (8.24) \]

\[ P(n)(\mathcal{E}_3) \to 0 \text{ as } n \to \infty, \text{ if } \tilde{R}_2 > I(Y_2; S_2|U_2) + \delta(\epsilon_3). \quad (8.25) \]

By the conditional typicality lemma given in Appendix 2.A,

\[ P(n)(\mathcal{E}_{41} \cap \mathcal{E}_{42} \cap \mathcal{E}_{43}) \to 0 \text{ as } n \to \infty. \quad (8.26) \]

We next bound \( P(n)(\cup_{m \neq 1} \mathcal{E}_{4m}) \). Define

\[ \mathcal{A}_j(m, \hat{l}_{1j}, l_{2,j-1}, l_{2j}) := \left\{ (U_{1j}(m, \hat{l}_{1j}), \tilde{Y}_{2j}|l_{2,j-1}, \hat{l}_{2j}), U_{2j}(l_{2,j-1}, \hat{l}_{2j}), Y_{3j}, S_{3j}) \in \mathcal{T}_e(n) \right\}. \quad (8.27) \]

Then consider,

\[ P(n)(\mathcal{E}_{4m}) = P(n) \left( \bigcup_{\tilde{p}_1, \tilde{p}_2} \bigcup_{\tilde{l}_2} \bigcap_{j=1}^b \mathcal{A}_j(m, \hat{l}_{1j}, l_{2,j-1}, l_{2j}, \tilde{l}_{2j}) \right) \]

\[ \leq \sum_{\tilde{p}_1, \tilde{p}_2} \sum_{\tilde{l}_2} \prod_{j=1}^b P(n) \left( \mathcal{A}_j(m, \hat{l}_{1j}, l_{2,j-1}, l_{2j}, \tilde{l}_{2j}) \right) \]

\[ = \sum_{\tilde{p}_1, \tilde{p}_2} \sum_{\tilde{l}_2} \prod_{j=1}^b P(n) \left( \mathcal{A}_j(m, \hat{l}_{1j}, l_{2,j-1}, l_{2j}, \tilde{l}_{2j}) \right) \]

\[ \leq \sum_{\tilde{p}_1, \tilde{p}_2} \sum_{\tilde{l}_2} \prod_{j=2}^b P(n) \left( \mathcal{A}_j(m, \hat{l}_{1j}, l_{2,j-1}, l_{2j}, \tilde{l}_{2j}) \right). \quad (8.28) \]

If \( l_{j-1} = L_{j-1}, \hat{l}_{2j} = \tilde{l}_{2j} \) but \( m \neq 1 \), then \( U_{ij}(m, \hat{l}_{ij}) \) is independent of \( \tilde{Y}_{2j}|l_{2,j-1}, \hat{l}_{2j}, U_{2j}|l_{2,j-1}, \hat{l}_{2j}, Y_{3j}, S_{3j} \). Therefore by a similar approach as
that in the joint typicality lemma given in Appendix 2.A, we have
\[
\mathbb{P}^{(n)} \left( A_j(m, \hat{l}_{1j}, l_{2,j-1}, l_{2j}, \hat{l}_{2j} \mid l_{j-1} = L_{j-1}, l_{2j} = \hat{L}_{2j}, m \neq 1 \right)
\]
\[
= \sum_{(U_{1j}, Y_{2j}, U_{2j}, Y_{3j}, S_{3j})} p(U_{1j}) p(Y_{2j}, U_{2j}, Y_{3j}, S_{3j})
\]
\[
\leq \sum_{(U_{1j}, Y_{2j}, U_{2j}, Y_{3j}, S_{3j})} 2^{-n(1-\epsilon)H(U_{1j})} 2^{-n(1-\epsilon)H(U_{2j}, Y_{2j}, Y_{3j}, S_{3j})}
\]
\[
= \mathcal{T}_k(n)^{2^{-n(1-\epsilon)H(U_{1j})} - n(1-\epsilon)H(U_{2j}, Y_{2j}, Y_{3j}, S_{3j})}
\]
\[
\leq 2^n(1-\epsilon)H(U_{1j}, U_{2j}, Y_{2j}, Y_{3j}, S_{3j}) 2^{-n(1-\epsilon)H(U_{2j}, Y_{2j}, Y_{3j}, S_{3j})}
\]
\[
= 2^n(I(U_{1j}, U_{2j}, Y_{2j}, Y_{3j}, S_{3j}) - \delta(\epsilon))
\]
\[
= 2^n(I(U_{1j}, U_{2j}, Y_{2j}, Y_{3j}, S_{3j}) - \delta(\epsilon))
\]
\[
= 2^{-n(I_{2j} - \delta(\epsilon))},
\]
(8.29)

where

\[
I_1 := I(U_{1j}, Y_{2j}, Y_{3j}|U_{2j}) + I(U_{1j} U_{2j}).
\]
(8.30)

Similarly,

\[
\mathbb{P}^{(n)} \left( A_j(m, \hat{l}_{1j}, l_{2,j-1}, l_{2j}, \hat{l}_{2j} \mid \{l_{j-1} \neq L_{j-1} or \hat{l}_{2j} \neq \hat{L}_{2j}, m \neq 1 \right)
\]
\[
= \sum_{(U_{1j}, Y_{2j}, U_{2j}, Y_{3j}, S_{3j})} p(U_{1j}) p(Y_{2j}, U_{2j}, Y_{3j}, S_{3j})
\]
\[
\leq \sum_{(U_{1j}, Y_{2j}, U_{2j}, Y_{3j}, S_{3j})} 2^{-n(1-\epsilon)H(U_{1j})} 2^{-n(1-\epsilon)H(U_{2j}, Y_{2j}, Y_{3j}, S_{3j})}
\]
\[
= \mathcal{T}_k(n)^{2^{-n(1-\epsilon)H(U_{1j})} - n(1-\epsilon)H(U_{2j}, Y_{2j}, Y_{3j}, S_{3j})}
\]
\[
\leq 2^n(1-\epsilon)H(U_{1j}, U_{2j}, Y_{2j}, Y_{3j}, S_{3j}) 2^{-n(1-\epsilon)H(U_{2j}, Y_{2j}, Y_{3j}, S_{3j})}
\]
\[
= 2^n(I(U_{1j}, U_{2j}, Y_{2j}, Y_{3j}, S_{3j}) + I(Y_{2j} U_{1j}, Y_{3j}|U_{2j}) - \delta(\epsilon))
\]
\[
= 2^n(I_2 - \delta(\epsilon)),
\]
(8.31)

where the last equality holds since

\[
H(U_1) + H(U_2, \hat{Y}_2) + H(Y_3, S_3) - H(U_1, U_2, \hat{Y}_2, Y_3, S_3)
\]
\[
= H(U_1) + H(U_2) + H(\hat{Y}_2|U_2) + H(Y_3, S_3)
\]
\[
= H(U_1, U_2) - H(Y_3, S_3|U_1, U_2) - H(\hat{Y}_2|U_1, U_2, Y_3, S_3)
\]
\[
= [H(Y_3, S_3) - H(Y_3, S_3|U_1, U_2)] + [H(\hat{Y}_2|U_2) - H(\hat{Y}_2|U_1, U_2, Y_3, S_3)]
\]
Similarly consider

\[ + [H(U_1) + H(U_1) - H(U_1; U_2)] \]

\[ = I(U_1, U_2; Y_3, S_3) + I(Y_2; U_1, Y_3, S_3|U_2) + I(U_1; U_2). \]

\[ =: I_2 \quad (8.32) \]

Thus

\[ \mathbb{P}(\mathcal{E}_{4m}) \leq \sum_{\ell_1, \ell_2, \ell_3, \ell_4 = 1}^b \mathbb{P} \left( A_j(m, \ell_1, \ell_2, \ell_3, \ell_4) \right) \]

\[ \leq \sum_{\ell_1, \ell_2, \ell_3, \ell_4 = 1}^b \sum_{k=0}^{b-1} \left( \frac{b-1}{k} \right) \left\{ \begin{array}{l} 2^{-n(b-1-k)(\ell_2+\ell_4)} \leq 2^{-n(b-1-k)(\ell_2+\ell_4) - \delta} \\ 2^{-n(b-1)(\ell_1, \ell_2 - \ell_2 - \ell_4)} - \delta \end{array} \right\} \]

\[ \leq \sum_{\ell_1, \ell_2, \ell_3, \ell_4 = 1}^b \sum_{k=0}^{b-1} \left( \frac{b-1}{k} \right) \left\{ \begin{array}{l} 2^{-n(b-1-k)(\ell_2+\ell_4)} \leq 2^{-n(b-1-k)(\ell_2+\ell_4) - \delta} \\ 2^{-n(b-1)(\ell_1, \ell_2 - \ell_2 - \ell_4)} - \delta \end{array} \right\} \]

\[ = \sum_{\ell_1, \ell_2, \ell_3, \ell_4 = 1}^b \sum_{k=0}^{b-1} \left( \frac{b-1}{k} \right) \left\{ \begin{array}{l} 2^{-n(b-1-k)(\ell_2+\ell_4)} \leq 2^{-n(b-1-k)(\ell_2+\ell_4) - \delta} \\ 2^{-n(b-1)(\ell_1, \ell_2 - \ell_2 - \ell_4)} - \delta \end{array} \right\} \]

Employing the union bound, we obtain

\[ \mathbb{P}(\bigcup_{m \neq 1} \mathcal{E}_{4m}) \leq \sum_{m \neq 1} \mathbb{P}(\mathcal{E}_{4m}) \]

\[ \leq 2^{n(b)(R - \frac{b-1}{b}(\min\{I_1 - \ell_1, I_2 - \ell_2 - \ell_1 - \ell_2\} - \delta) + \frac{b-1}{b}(\ell_2 + \ell_4)).} \quad (8.33) \]

Thus, as \( n \to \infty \), the probability of error goes to zeros if

\[ R < \frac{b-1}{b} \left[ \min\{I_1 - \ell_1, I_2 - \ell_2 - \ell_1 - \ell_2\} - \delta \right] - \frac{1}{b}\ell_2 + \ell_4. \]

We next simplify each term under the min. Consider

\[ I_1 - \ell_1 = I(U_1; \tilde{Y}_2, Y_3, S_3|U_2) + I(U_1; U_2) - I(U_1; S_1) - \delta \]

\[ = I(U_1; \tilde{Y}_2, Y_3, S_3|U_2) + H(U_1) - H(U_1|U_2) - H(U_1|S_1) - \delta \]

\[ = I(U_1; \tilde{Y}_2, Y_3, S_3|U_2) - H(U_1|U_2) + H(U_1|S_1, U_2) - \delta \]

\[ = I(U_1; \tilde{Y}_2, Y_3, S_3|U_2) - I(U_1; S_1|U_2) - \delta. \quad (8.34) \]

Similarly consider

\[ I_2 - \ell_2 - \ell_1 - \ell_2 = I(U_1, U_2; Y_3, S_3) + I(\tilde{Y}_2; U_1, Y_3, S_3|U_2) + I(U_1; U_2) \]

\[ - I(\tilde{Y}_2; Y_3, S_3|U_2) - I(U_1; S_1) - I(U_2; S_2) - \delta' \]

\[ = I(U_1, U_2; Y_3, S_3) + H(\tilde{Y}_2|Y_2, S_2, U_2) - H(\tilde{Y}_2|U_1, U_2, Y_3, S_3). \]
Proof. Consider the following series of equalities

\[ + [-H(U_1|U_2) + H(U_1|S_1) - H(U_2) + H(U_2|S_2)] = \delta'(\epsilon) \]
\[ = I(U_1, U_2; Y_3, S_3) \]
\[ + \left[ H(\hat{Y}_2|Y_2, S_2, U_2, U_1, Y_3, S_3) - H(\hat{Y}_2|U_1, U_2, Y_3, S_3) \right] + \]
\[ - [H(U_1|U_2) + H(U_1|S_1) - H(U_2) + H(U_2|S_2)] = \delta'(\epsilon) \]
\[ = I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2|U_1, U_2, Y_3, S_3) \]
\[ + [-H(U_1, U_2) + H(U_1|S_1) + H(U_2|S_2, U_1, S_1)] - \delta'(\epsilon) \]
\[ = I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2|U_1, U_2, Y_3, S_3) \]
\[ - [H(U_1, U_2) - H(U_1, U_2|S_1, S_2)] - \delta'(\epsilon) \]
\[ = I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2|U_1, U_2, Y_3, S_3) \]
\[ = I(U_1, U_2; S_1, S_2) - \delta'(\epsilon), \text{(8.35)} \]

where \( \delta'(\epsilon) := \delta(\epsilon_1) + \delta(\epsilon_2) + \delta(\epsilon_3) \).

Now let \( b \to \infty \) and \( \{\epsilon, \epsilon_1, \epsilon_2, \epsilon_3\} \to 0 \). Thus the rate

\[ R < \min I(U_1; \hat{Y}_2, Y_3, S_3|U_2) - I(U_1; S_1|U_2), \]
\[ I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2|U_1, U_2, Y_3, S_3) - I(U_1, U_2; S_1, S_2) \text{, } \text{(8.36)} \]

is achievable. This completes the proof.

8.5 Proof of Proposition 8.3.4

In order to proceed with the proof, we first present two lemmas.

Lemma 8.5.1.

\[ I(\hat{Y}_2; Y_2, S_2|U_2, Y_3, S_3) - [I(U_2; Y_3, S_3) - I(U_2; S_2)] \]
\[ = \left[ I(U_1; \hat{Y}_2, Y_3, S_3|U_2) - I(U_1; S_1|U_2) \right] \]
\[ - \left[ I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2|U_1, U_2, Y_3, S_3) - I(U_1, U_2; S_1, S_2) \right] \text{. } \text{(8.37)} \]

Proof. Consider the following series of equalities

\[ I(U_1; \hat{Y}_2, Y_3, S_3|U_2) - I(U_1; S_1|U_2) \]
\[ = H(\hat{Y}_2, Y_3, S_3|U_2) - H(\hat{Y}_2, Y_3, S_3|U_1, U_2) - I(U_1; S_1|U_2) \]
\[ = H(Y_3, S_3|U_2) + H(\hat{Y}_2|U_2, Y_3, S_3) - H(Y_3, S_3|U_1, U_2) \]
\[ - H(\hat{Y}_2|U_1, U_2, Y_3, S_3) - I(U_1; S_1|U_2) \]
\[ = I(U_1; Y_3, S_3|U_2) + H(\hat{Y}_2|U_2, Y_3, S_3) \]
This completes the proof of the lemma.

8.5. Proof of Proposition 8.3.4

Let

\[ I(U_1; U_2; Y_3, S_3) = I(U_1; U_2; Y_3, S_3) + H(\hat{Y}_2|U_2, Y_3, S_3) \]

where

\[ G := I(\hat{Y}_2; Y_2, S_2|U_2, Y_3, S_3) - I(U_2; Y_3, S_3) + I(U_2; S_2). \]  

This completes the proof of the lemma.

Lemma 8.5.2. Let

\[ \mathcal{P}^* = p(x_1, u_1|s_1)p(x_2, u_2|s_2)p(\hat{y}_2|y_2, u_2, s_2), \]

be the joint pmf that optimizes the rate in (8.2). Then for \( \mathcal{P}^*\),

\[ I(U_1; \hat{Y}_2, Y_3, S_3|U_2) - I(U_1; S_1|U_2) \leq I(U_1; U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2|U_1, U_2, Y_3, S_3) - I(U_1, U_2; S_1, S_2). \]  

Proof. We give the proof by contradiction. We show that if

\[ I(U_1; \hat{Y}_2, Y_3, S_3|U_2) - I(U_1; S_1|U_2) > I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2; Y_2, S_2|U_1, U_2, Y_3, S_3) - I(U_1, U_2; S_1, S_2), \]  

then a contradiction follows.
then there is another
\[ P' = p(x_1, u_1|s_1)p(x_2, u_2|s_2)p(\hat{y}_2'|y_2, u_2, s_2), \quad (8.43) \]
that attains a higher rate. Now let \( \hat{Y}_2' = \hat{Y}_2 \) with probability \( p \) and \( \hat{Y}_2' = \emptyset \) otherwise. We observe that both terms under min in (8.2) are continuous in \( p \) and the first term increases in \( p \) while the second term decreases in \( p \). Thus there exists \( p^* \) such that
\[
I(U_1; \hat{Y}_2', Y_3, S_3|U_2) - I(U_1; S_1|U_2) = I(U_1, U_2; Y_3, S_3) - I(\hat{Y}_2'; Y_2, S_2|U_1, U_2, Y_3, S_3)
- I(U_1, U_2; S_1, S_2),
\]
which contradicts (8.42). This completes the proof of the lemma. □

Now, the proof of Proposition 8.3.4 follows by combining Lemma 8.5.1 and 8.5.2.

### 8.6 Concluding Remarks

We, in this chapter, proposed an achievable rate for a fairly general relay channel with noncausal state at the nodes. The main conclusion of the chapter is that the proposed scheme with quantize-and-forward attains the same rate as that of CF with Wyner-Ziv coding as proposed in [AMA10].
9.1 Introduction

In this chapter, we study the multiple-access relay channel (MARC)

\[ (X_1 \times \cdots \times X_K \times X_r \times S, p(y, y_r | x_1, \ldots, x_K, x_r, s)p(s), Y \times Y_r) \]

with a random state \( s \), \( K \) users, a relay, and a destination. User \( k \in [1:K] \) wants to transmit a message \( M_k \) uniformly chosen from the set \( \mathcal{M}_k^{(n)} \) to the destination. At time instant \( i \in [1:n] \), the \( k \)th user transmits \( X_{ki}(M_k) \), the relay receives \( Y_{ri} \) and transmits \( X_{ri}(Y_{ri}^{i-1}) \), and the destination receives \( Y_i \). The messages at the users are assumed to be independent from each other.

The rate tuple \( (R_1, R_2, \ldots, R_K) \) is achievable, if there exists a communication strategy, such that the average probability of error

\[ P_e^{(n)} := \text{Pr}\{ (\hat{M}_1, \ldots, \hat{M}_K) \neq (M_1, \ldots, M_K) \} \]

at the decoder tends to zero as \( n \to \infty \) and at the same time

\[ \liminf_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_k^{(n)}| \geq R_k, \ \forall k \in [1:K]. \]

The capacity region of the channel is defined as the closure of the set of achievable rate tuples \( (R_1, R_2, \ldots, R_K) \).

In this chapter, we establish the capacity region of some classes of MARCs. Some related work on MARCs is reported in e.g., [KvW00, KGG05, SKM07].

9.2 State-Decoupled MARC

We next introduce state-decoupled MARCs and establish the capacity of a semi-deterministic class of the state-decoupled MARCs.
Definition 9.2.1. The MARC with a random state $s$ is said to be state-decoupled if the channel pmf is given by
\[
p(y^n, y^n_r | x^n_1, \ldots, x^n_K, x^n_r, s^n) = \prod_{i=1}^{n} p(y_i | x^n_1, \ldots, x^n_{K}, x^n_r, s_i) p(y^n_r | s^n).
\] (9.1)
That is $(X_1, \ldots, X_K, X_r, Y) \rightarrow S_i \rightarrow Y_i$ form a Markov chain. The channel state sequence $(s_1, s_2, \ldots, s_n)$ is assumed to be i.i.d. and it is determined by nature. See Figure 9.1 for an illustration of the channel.

We assume that the users have no knowledge of the channel state and the relay and the destination are informed about the channel state only through $y_r$ and $y$, respectively.

Theorem 9.2.1. If the $K$-user MARC is state-decoupled and $y_r = f(x^K_1, x_r, y)$, then the capacity region is given by the union of all $(R_1, R_2, \ldots, R_K)$ tuples that satisfy
\[
\sum_{j \in \mathcal{J}} R_j \leq I(X(\mathcal{J}); Y(\mathcal{J}^c), Y|X(\mathcal{J}^c), Q) \text{ for all } \mathcal{J} \subseteq K \cup r
\] (9.2)
and for some joint pmf of the form $p(q) p(x_r | q) \prod_{k=1}^{K} p(x_k | q)$ with $|Q| \leq K$. Here $R_r = 0$, $X(\mathcal{J}) := \{ X_j : j \in \mathcal{J} \}$, $K := [1 : K]$ and $\mathcal{J}^c$ denotes the complement of $\mathcal{J}$.

Proof. We first consider achievable rates. Using the noisy network coding (NNC) lower bound in [LKG11], the rate region
\[
\sum_{j \in \mathcal{J}} R_j \leq I(X(\mathcal{J}); Y(\mathcal{J}^c), Y|X(\mathcal{J}^c), Q) - I(Y_r(\mathcal{J}); \hat{Y}_r(\mathcal{J})|X^K_1, X_r, \hat{Y}(\mathcal{J}^c), Y, Q),
\] (9.3)
is achievable for some
\[
p(q) p(y_r, x_r, q) p(x_r | q) \prod_{k=1}^{K} p(x_k | q).
\]
Observe that the subtractive term in (9.3) is active if the relay is included in the set $\mathcal{J}$ and it takes the form $I(Y_r; Y_r | X_r^K, X_r, Y)$. Now let $Y_r = Y_r$. Since $Y_r = f(X_r^K, X_r, Y)$, we have $I(Y_r; Y_r | X_r^K, X_r, Y) = 0$. This completes the proof of achievability. The converse follows by the cutset bound and noting that the input distribution can be restricted to $p(x_k | q) \prod_{k=1}^{K} p(x_k | q)$. This follows by noting that i) the users are separated from each other, ii) the associated messages are independent from each other, and iii) the users do not know the channel state. The cardinality bound on $Q$ follows by the shape of the capacity region. This completes the proof of the theorem.

Theorem 9.2.1 includes the state-decoupled MARC studied by Tandon and Poor in [TP11, Section IV-A] as a special case. Firstly, note that the channel model in [TP11] is a special case of that in Theorem 9.2.1. To see this, let $K = 2$, $y_r = s$, $y = (y_1, y_2)$ and $p(y|x_1, x_2, x_r, s) = p(y_1|x_1, x_2, s)p(y_2|x_r)$. That is, there is an orthogonal link from the relay to the destination. Now without loss of generality replace the link from the relay to the destination with a noiseless link with rate $R_0 = \max_{p(x)} I(X_r; Y_2)$. Secondly, note that the deterministic relation in [TP11] is more restrictive than that in Theorem 9.2.1. The relation in [TP11] is specified as $s = f_1(x_1, y_1)$ and $s = f_2(x_2, y_1)$, which yields that $s = f(x_1, x_2, x_r, y)$. However the condition $s = f(x_1, x_2, x_r, y)$ does not in general imply that $s = f_1(x_1, y_1)$ and $s = f_2(x_2, y_1)$. As a counter-example consider the channel in Example 5 where $s \neq f_1(x_1, y_1)$ and $s \neq f_2(x_2, y_1)$.

**Example 5**

Consider the MARC with

\begin{align}
Y_1 &= X_1 + X_2 + S \\
Y_r &= S
\end{align}

and assume that the relay has a noiseless link with capacity $R_0$ to the destination. We assume that the users operate under the average power constraint, i.e. $\mathbb{E} [X_k^2] \leq P_k$ for $k \in \{1, 2\}$.

**Proposition 9.2.2.** The capacity region of the MARC in Example 5 is

\[\mathcal{R} = \left\{ (R_1, R_2) : R_1 \leq C \left( \frac{P_1}{Q} \right) + R_0, \right.\]

\[R_2 \leq C \left( \frac{P_2}{Q} \right) + R_0, \]

\[R_1 + R_2 \leq C \left( \frac{P_1 + P_2}{Q} \right) + R_0 \}, \]

(9.5)

where $C(x) := \frac{1}{2} \log_2 (1 + x)$.

**Proof.** The proof follows by applying Theorem 9.2.1 and noting that Gaussian inputs are optimal. \qed
9.3 MARC with Orthogonal Receive Components

We next prove that NNC is also capacity achieving for a class of MARCs with orthogonal receive components.

Proposition 9.3.1. Consider the discrete memoryless MARC with pmf
\[ p(y, y_r | x_1, x_2) \]
and an orthogonal noiseless link with rate \( R_0 \) from the relay to the destination as illustrated in Figure 9.2. If
\[ y_r = f(x_1, x_2, y) \]
then the capacity region is given by

\[
\mathcal{R} = \left\{ (R_1, R_2) : R_1 \leq I(X_1; Y | X_2, Q) + \min\{R_0, I(X_1; Y_r | Y, X_2, Q)\} \\
R_2 \leq I(X_2; Y | X_1, Q) + \min\{R_0, I(X_2; Y_r | Y, X_1, Q)\} \\
R_1 + R_2 \leq I(X_1, X_2; Y | Q) + \min\{R_0, I(X_1, X_2; Y_r | Y, Q)\} \right\}
\]

for some \( p(q)p(x_1|q)p(x_2|q) \) with \(|Q| \leq 2\).

Proof. The achievability follows by choosing \( \hat{Y}_r = Y_r \) in the noisy networking coding lower bound in [LKGC11] and the converse follows by the cut-set bound. \( \square \)

Proposition 9.3.1, for example can be applied to the channel with

\[
Y = g_1(X_1, X_2) + g_2(Z),
\]
\[
Y_r = g_3(X_1, X_2) + g_4(Z),
\]
where \( Z \) is an arbitrary random variable, \( g_1, g_3 \) and \( g_4 \) are arbitrary functions, and \( g_2 \) is an invertible function. This follows because

\[
Y_r = g_3(X_1, X_2) + g_4\left(g_2^{-1}(Y - g_1(X_1, X_2))\right)
=: f(X_1, X_2, Y).
\]

Note that this channel is not state-decoupled.
9.4 Application to a MARC with Additive Interference

9.4.1 Channel Model

Consider the strictly causal MARC consisting of two users, a relay, and a destination. The input–output relations are given by

\[
Y = X_1 + X_2 + S + X_r, \quad (9.9a)
\]

\[
Y_r = S, \quad (9.9b)
\]

where \( S \sim \mathcal{N}(0, Q) \). We assume that the users and the relay operate under average power constraints, i.e. \( \mathbb{E} \left[ X_k^2 \right] \leq P_r, \mathbb{E} \left[ X_k^2 \right] \leq P_k \) for \( k \in \{1, 2\} \). This models interference-limited MARC when additive noises at the destination and the relay are neglected. Note that the relay observes the interference and it transmits signal \( X_r(S_i - 1) \) which creates an interference at the destination. See Figure 9.3 for an illustration.

9.4.2 Capacity of Strictly Causal Relaying

**Proposition 9.4.1.** The capacity region of the MARC given by (9.9) with strictly causal relaying when \( S \sim \mathcal{N}(0, Q) \) is

\[
\mathcal{R} = \left\{ (R_1, R_2) : R_1 \leq C \left( \frac{P_1 + P_r}{Q} \right), \right. \\
\left. R_2 \leq C \left( \frac{P_2 + P_r}{Q} \right), \right. \\
\left. R_1 + R_2 \leq C \left( \frac{P_1 + P_2 + P_r}{Q} \right) \right\}. \quad (9.10)
\]
Proof. Observe that
\[ Y_r = Y - X_1 - X_2 - X_r =: f(X_1, X_2, X_r, Y). \] (9.11)
Therefore, we can apply Theorem 9.2.1. The proof then follows by noting that Gaussian inputs are optimal. See Figure 9.4 for an illustration of the capacity region.

Remark 9.4.2. Using similar techniques as in Section 7.4.3 for the relay channel, one can enlarge the capacity region of the state-decoupled MARC by considering causal relaying.

9.4.3 Generalization
One can apply Theorem 9.2.1 to the MARC with
\[
Y = f_1(X_1, X_2, X_r) + f_2(S), \quad \text{(9.12a)}
\]
\[
Y_r = f_3(S), \quad \text{(9.12b)}
\]
where \( f_1 \) and \( f_3 \) are arbitrary functions, \( f_2 \) is an invertible function and \( S \) denotes the state of the channel with an arbitrary distribution.

9.4.4 Extension to Multiple Users
Consider MARC consisting of \( K \) users, a relay, and a destination. The input–output relations are given by
\[
Y = \sum_{k=1}^{K} X_k + S + X_r, \quad \text{(9.13a)}
\]
\[
Y_r = S, \quad \text{(9.13b)}
\]
where \( S \sim \mathcal{N}(0, Q) \). We assume that the users and the relay operate under average power constraints, i.e. \( \mathbb{E}[X_k^2] \leq P_r \), \( \mathbb{E}[X_k^2] \leq P_k \) for \( \forall k \in [1:K] \).

Proposition 9.4.3. The capacity of the above MARC with strictly causal relaying is
\[
\sum_{j \in J} R_j \leq C \left( \frac{\sum_{j \in J} P_j + P_r}{Q} \right), \quad \forall J \subseteq [1:K]. \] (9.14)
Proof. The proof is similar to that of Proposition 9.4.1.

9.5 Conclusions
In this chapter, we studied MARCs and proved that the noisy network coding protocol as proposed in [LKGC11] achieves the capacity of two classes of the MARCs.
9.5. Conclusions

Figure 9.4: The capacity region of the semi-deterministic MARC with additive Gaussian interference and strictly causal relaying.
10.1 Introduction

Noisy network coding (NNC) [LKGC11] and compress-and-forward (CF) [CG79] strategies are constructed based on an analog-to-digital interface in which the relay compresses the received signal and transmits a compression index in a digital manner. As we remarked in Chapter 2, NNC achieves the same rate as that achieved by the original CF strategy constructed in [CG79] for the classical three-node relay channel introduced in [vdM71]. NNC can however outperform CF for larger networks since it does not employ the Wyner-Ziv approach as it is done in CF, and the destination is not restricted to find the correct compression index. Recall that we discussed NNC in Chapter 9 for multiple-access relay channels (MARCs) and proved that NNC is in fact optimal for some semi-deterministic classes of such channels. It is also shown in [LKGC11] that NNC can outperform extensions of CF [CG79, KGG05, HMZ06] and amplify-and-forward (AF) [RW06, GMZ06, PY07, LWT08] for the two-way relay channel (TWRC). NNC can be further improved using layered NNC in which the nodes may transmit the compression index in two layers [LKGC10]. Thus the users have the flexibility to decode either one layer or two layers based on the channel quality and the available side information.

When employing AF, the relay transmits the noisy received signal using a linear mapping. Therefore, AF can be considered as a scheme that is based on an analog-to-analog interface. AF can be envisioned as a special case of a more general strategy in which the relay performs analog mappings directly on noisy received signals at the relay. We refer to such a general scheme as analog noisy network coding (ANNC). This strategy is fundamentally different from CF and NNC since the relay does not generate any digital compression index. With this generalization, for example a scheme in which the relay employs analog nonlinear mappings on the received signals at the relay, is also an ANNC scheme. Memoryless relaying is also considered in [FM04, GJo6, KMG+07, GJo7, YKS08, CHK08] for different types of channels. In particular [CHK08] investigates nonlinear relaying for the TWRC in the case of uncoded transmission. For coded transmission over the TWRC, to
the best of our knowledge, only linear relaying has been studied before, e.g. in \cite{RW06, PY07, KMG+07}. In this chapter, we study nonlinear ANNC strategies for the TWRC.

In the sequel, we first present the Gaussian TWRC that we investigate. We then briefly discuss some recent relaying schemes proposed for the TWRC. Assuming that the relay employs a memoryless analog mapping, we next present an achievable rate region for the TWRC. For a semi-deterministic case where the received signal at the relay is noiseless, we present an optimal nonlinear mapping. For the noiseless case, we show that memoryless ANNC is capacity achieving. For the general case when the received signal at the relay is noisy, we provide an optimization algorithm based on the fixed point iteration method to find optimized mappings. Our results show that the proposed scheme with optimized nonlinear ANNC can significantly outperform CF and AF and can achieve the rates close to the layered NNC. We note that the proposed scheme has virtually the same complexity at the relay as that of AF.

10.2 The Gaussian Two-Way Relay Channel

Figure 10.1 illustrates the TWRC that we consider in this chapter. We assume that the two users are separated and therefore cannot overhear each other. The users wish to exchange two messages $M_1 \in \mathcal{M}_1^{(n)}$ and $M_2 \in \mathcal{M}_2^{(n)}$ in $n$ channel uses with a shared relay node. We assume that the messages are uniformly chosen from $\mathcal{M}_1^{(n)}$ and $\mathcal{M}_2^{(n)}$.

Let $i \in [1:n]$ denote the $i$th time instant. The users transmit symbols $X_{1i}$ and $X_{2i}$ using encoders $\alpha_1$ and $\alpha_2$, respectively. The relay receives

$$Y_{ri} = X_{1i} + X_{2i} + Z_{ri},$$

where $Z_{ri} \sim \mathcal{N}(0,N_r)$ is the additive Gaussian noise at the relay. The relay employs
the encoder $f$ to generate the symbol $X_{ri}$. The user $j \in \{1, 2\}$ then receives
\[ Y_{ji} = X_{ri} + Z_{ji}, \] (10.2)
where $Z_{ji} \sim \mathcal{N}(0, N_j)$ denotes additive noise at the $j$th user. We assume that all noise sequences are white and mutually independent from each other. We additionally assume that the users and the relay operate under average power constraints, i.e. $\sum_{i=1}^{n} \mathbb{E}[X_{ri}^2] \leq nP_r$, $\sum_{i=1}^{n} \mathbb{E}[X_{ji}^2] \leq nP_j$ for $j \in \{1, 2\}$. Upon receiving $Y_{1i}^n$, user 1 employs the decoder $\beta_1$ to find an estimate of the transmitted message, denoted by $\hat{M}_1$. Similarly, user 2 employs the decoder $\beta_2$ to find $\hat{M}_2$.

Similar to the definition in Chapter 9 for the MARC, the rate pair $(R_1, R_2)$ is achievable if there exists a communication strategy such that the average message error probability
\[ P_e^n = \Pr\{ (M_1, M_2) \neq (\hat{M}_1, \hat{M}_2) \}, \]
approaches zero as $n \to \infty$ and at the same time
\[ \lim inf_{n \to \infty} \frac{1}{n} \log |M_1^n| \geq R_j, \quad \forall j \in \{1, 2\}. \] (10.3)

### 10.3 Compression-Based Achievable Rate Regions

In this section, we briefly review achievable rate results for relaying schemes with an analog-to-digital interface including CF, NNC, and layered NNC. The achievable rate results are tailored to the TWRC.

#### 10.3.1 Relaying via CF

The achievable rate region of CF is given by [RW06]
\[ \mathcal{R}_{CF} = \{ (R_1, R_2) : R_1 \leq I(X_1; Y_2, \hat{Y}_r | X_2, X_r), \]
\[ R_2 \leq I(X_2; Y_1, \hat{Y}_r | X_1, X_r), \]
\[ \max \left( I(Y_r; \hat{Y}_r | X_1, X_r, Y_1), I(Y_r; \hat{Y}_r | X_2, X_r, Y_2) \right) \leq \min \left( I(X_r; Y_1 | X_1), I(X_r; Y_2 | X_2) \right) \}, \]
for some $p(x_1)p(x_2)p(x_r)p(\hat{y}_r|y_r,x_r)$. Now let the input alphabets at the users be Gaussian and $\hat{Y}_r = Y_r + Z_q$ where $Z_q \sim \mathcal{N}(0, N_q)$ is independent of other random variables. Then by optimizing over $N_q$, we obtain
\[ \mathcal{R}_{CF} = \left\{ (R_1, R_2) : R_1 \leq \mathcal{C} \left( \frac{P_1 P_r}{P_r N_r + \max(P_1, P_2) + N_r} \max(N_1, N_2) \right), \right. \]
\[ R_2 \leq \mathcal{C} \left( \frac{P_2 P_r}{P_r N_r + \max(P_1, P_2) + N_r} \max(N_1, N_2) \right) \}, \]
where $\mathcal{C}(x) := \frac{1}{2} \log_2(1 + x)$. 

10.3.2 Relaying via NNC

Using Theorem 2 in [LKGC11], NNC achieves the rate region

\[
\mathcal{R}_{\text{NNC}} = \bigg\{(R_1, R_2) : R_1 \leq I(X_1; Y_2, \hat{Y}_r | X_2, X_r), \bigg\}
\]

\[
R_1 \leq I(X_1, X_r; Y_2 | X_2) - I(Y_r; \hat{Y}_r | X_1, X_2, X_r, Y_2),
\]

\[
R_2 \leq I(X_2; Y_1, \hat{Y}_r | X_1, X_r),
\]

\[
R_2 \leq I(X_2, X_r; Y_1 | X_1) - I(Y_r; \hat{Y}_r | X_1, X_2, X_r, Y_1)
\bigg\},
\]

for some \(p(x_1)p(x_2)p(x_r)p(y_r | y_r, x_r, u)\). By choosing the same distributions as those in evaluating the rate region achievable by CF in Section 10.3.1, we obtain

\[
\mathcal{R}_{\text{NNC}} = \bigg\{(R_1, R_2) : R_1 \leq C \left( \frac{P_1}{P_1 + N_q} \right), \bigg\}
\]

\[
R_1 \leq C \left( \frac{P_r}{P_2} \right) - C \left( \frac{N_r}{N_q} \right),
\]

\[
R_2 \leq C \left( \frac{P_2}{P_1 + N_q} \right),
\]

\[
R_2 \leq C \left( \frac{P_r}{P_1} \right) - C \left( \frac{N_r}{N_q} \right), \quad N_q > 0
\bigg\}.
\]

10.3.3 Relaying via Layered NNC

NNC can be improved using a layered transmission of the compression index. The layered NNC in [LKGC10] achieves the rate region

\[
\mathcal{R}_{\text{LNNC}} = \bigg\{(R_1, R_2) : R_1 \leq I(X_1; Y_2, \hat{Y}_r | X_2, U), \bigg\}
\]

\[
R_1 \leq I(X_1, U; Y_2 | X_2) - I(Y_r; \hat{Y}_r | X_1, X_2, U, Y_2),
\]

\[
R_2 \leq I(X_2; Y_1, \hat{Y}_r, \tilde{Y}_r | X_1, X_r, U),
\]

\[
R_2 \leq I(X_2, X_r; Y_1 | X_1) - I(Y_r; \hat{Y}_r, \tilde{Y}_r | X_1, X_2, X_r, U, Y_1)
\bigg\},
\]

for some \(p(x_1)p(x_2)p(x_r)p(y_r | y_r, x_r, u)\). Note that by choosing \(\hat{Y}_r = Y_r\) and \(U = X_r\), LNNC simplifies to NNC and hence \(\mathcal{R}_{\text{NNC}} \subseteq \mathcal{R}_{\text{LNNC}}\). Now let input alphabets at the users be Gaussian and \(\hat{Y}_r = Y_r + Z_q, Y_r = \hat{Y}_r + \tilde{Z}_q\) where \(\tilde{Z}_q \sim \mathcal{N}(0, \tilde{N}_q), \tilde{Z}_q \sim \mathcal{N}(0, \bar{N}_q)\) are independent of each other. Additionally let \(X_r = U + V\) where \(U \sim \mathcal{N}(0, \alpha P_r)\) and \(V \sim \mathcal{N}(0, \bar{P}_r)\) is independent of other random
variables and \(\alpha = 1 - \alpha\). Then

\[
\mathcal{R}_{\text{LNNC}} = \left\{ (R_1, R_2) : R_1 \leq C \left( \frac{P_1}{N_r + N_q + \hat{N}_q} \right) , \right. \\
R_1 \leq C \left( \frac{\alpha P_r}{\hat{\alpha} P_r + N_2} \right) - C \left( \frac{N_r}{N_q + N_q} \right) , \\
R_2 \leq C \left( \frac{P_2}{N_r + N_q} \right) , \\
R_2 \leq C \left( \frac{P_r}{N_1} \right) - C \left( \frac{N_r}{N_q} \right) , \\
R_2 \leq C \left( \frac{\alpha P_r}{N_1} + \frac{P_2(\alpha P_r + N_1)}{N_1(N_r + N_q + \hat{N}_q)} \right) \\
- C \left( \frac{N_r\hat{N}_q}{N_q(N_r + N_q + \hat{N}_q)} \right) , \hat{N}_q > 0, \tilde{N}_q > 0, \alpha \in [0,1] \}.
\]

### 10.4 Instantaneous Analog Noisy Network Coding

We, in this section, focus on relaying schemes based on an analog-to-analog interface. In the sequel, we assume that the relay operates using the memoryless one-dimensional function \(f(y_r)\) that meets the power constraint at the relay. The relay functionality is symbol-wise similar to that in instantaneous relaying discussed in Section 2.4. When employing the function \(f(\cdot)\), the input–output relation over the channel can be hence written as

\[
\begin{align*}
Y_1 &= f(X_1 + X_2 + Z_r) + Z_1, \quad (10.4a) \\
Y_2 &= f(X_1 + X_2 + Z_r) + Z_2. \quad (10.4b)
\end{align*}
\]

Observe that the channel in (10.4) can be seen as a two-way channel parameterized by the pdf

\[
p(y_1, y_2|x_1, x_2) = \int p(y_1, y_2|x_1, x_2, z_r)p(z_r)dz_r = \int p(y_1|x_1, x_2, z_r)p(y_2|x_1, x_2, z_r)p(z_r)dz_r = \frac{1}{\sqrt{(2\pi)^3 N_1 N_2 N_r}} \int e^{-\frac{(y_1-f(x_1+x_2+z_r))^2}{2N_1} - \frac{(y_2-f(x_1+x_2+z_r))^2}{2N_2}} \frac{i^2}{\pi N_r} dz_r.
\]

(10.5)
Now using Shannon’s inner bound in [Sha61], for the equivalent two-way channel $p(y_1, y_2|x_1, x_2)$ given in (10.4), the rate region

$$R_{\text{IANNC}} = \left\{ (R_1, R_2) : R_1 \leq I(X_1; Y_2|X_2), \quad R_2 \leq I(X_2; Y_1|X_1) \right\},$$

(10.6)
is achievable for some $\{p(x_1) \in \mathcal{P}_1, p(x_2) \in \mathcal{P}_2, f(y_r) \in \mathcal{F}_r\}$ where $\mathcal{P}_1$ and $\mathcal{P}_2$ denote the sets of all pdf’s that fulfill the average power constraints at the users, and $\mathcal{F}_r$ denotes the set of real-valued one-dimensional functions that make the power constraint at the relay hold. We refer to this relaying scheme as instantaneous analog noisy network coding (IANNC), since the relay acts instantaneously on the received noisy analog symbol in order to generate a new analog symbol to be broadcasted to the users.

### 10.4.1 Linear Relaying

One instance of the rate region given in (10.4) can be obtained using a linear relaying scheme, also (well) known as amplify-and-forward (AF) [RW06]. For AF, the relay mapping is given by $f(y_r) = \kappa y_r$, where $\kappa := \sqrt{P_r} / (P_1 + P_2 + N_r)$ is chosen to ensure the power constraint at the relay. By choosing Gaussian input alphabets, i.e. $X_1 \sim \mathcal{N}(0, P_1)$, and $X_2 \sim \mathcal{N}(0, P_2)$, the achievable rate region given in (10.4) can be computed to be

$$R_{\text{L}} = \left\{ (R_1, R_2) : R_1 \leq C \left( \frac{P_1 P_r}{P_1 N_r + P_1 N_2 + P_2 N_2 + N_2 N_r} \right), \quad R_2 \leq C \left( \frac{P_2 P_r}{P_1 N_r + P_1 N_1 + P_2 N_1 + N_1 N_r} \right) \right\},$$

(10.7)

### 10.4.2 Computation of the Rate Region for Nonlinear Relaying

Linear relaying is optimal in the limit of infinite SNR in the links from the relay to the users. However, linear relaying is not optimal for our setup since the users have some side information which is correlated to the received signal at the relay. The availability of side information allows the relay to save its power by transmitting only a part of its knowledge (i.e., the noisy received signal) in the analog domain. In the following, we therefore allow that the relay mapping be nonlinear. We first sketch, how we can compute the achievable rate for given $p(x_1), p(x_2)$ and $f(y_r)$. Consider

$$R_1 = I(X_1; Y_2|X_2)$$
$$= h(Y_2|X_2) - h(Y_2|X_1, X_2)$$
$$= h(Y_2|X_2) - h(f(X_1 + X_2 + Z_r) + Z_2|X_1, X_2)$$
\[ h(Y_2|X_2) - h(f(T + Z_r) + Z_2|T) = -\mathbb{E}[\log p(y_2|x_2)] + \mathbb{E}[\log p(y_2|t)], \tag{10.8} \]

where \( T := X_1 + X_2 \) and
\[
\begin{align*}
p(y_2|x_2) &= \frac{1}{\sqrt{2\pi N_2}} \int \exp\left(-\frac{(y_2 - f(y_r))^2}{2N_2}\right) p(y_r|x_2) dy_r, \\
p(y_2|t) &= \frac{1}{2\sqrt{\pi N_2 N_r}} \int \exp\left(-\frac{(y_2 - f(y_r))^2}{2N_2} - \frac{(y_r - t)^2}{2N_r}\right) dy_r.
\end{align*}
\]

Similarly, we have
\[
\begin{align*}
R_2 &= I(X_2; Y_1|X_1) = -\mathbb{E}[\log p(y_1|x_1)] + \mathbb{E}[\log p(y_1|t)], \\
\end{align*}
\]
where \( T = X_1 + X_2 \) and
\[
\begin{align*}
p(y_1|x_1) &= \frac{1}{\sqrt{2\pi N_1}} \int \exp\left(-\frac{(y_1 - f(y_r))^2}{2N_1}\right) p(y_r|x_1) dy_r, \\
p(y_1|t) &= \frac{1}{2\sqrt{\pi N_1 N_r}} \int \exp\left(-\frac{(y_1 - f(y_r))^2}{2N_1} - \frac{(y_r - t)^2}{2N_r}\right) dy_r.
\end{align*}
\]

Using the above steps, we are able to compute the rate region for a given nonlinear function and input alphabets.

### 10.5 Optimized Analog Noisy Network Coding

#### 10.5.1 Noiseless Uplink

We first consider the noiseless uplink for which the received signal at the relay is given by
\[ Y_r = X_1 + X_2. \tag{10.9} \]

Next consider
\[
\begin{align*}
R_1 &= I(X_1; Y_2|X_2) \\
&= h(Y_2|X_2) - h(f(X_1 + X_2) + Z_2|X_1, X_2) \\
&= h(f(X_1 + X_2) + Z_2|X_2) - h(Z_2).
\end{align*}
\]

Thus, in order to optimize \( R_1 \), it suffices to optimize \( h(f(X_1 + X_2) + Z_2|X_2) \). This problem can be cast into the one addressed in Chapter 6. To end this, let \( X_1 \sim \mathcal{N}(0, P_1) \), \( X_2 \sim \mathcal{N}(0, P_2) \), and
\[
\begin{align*}
f^*_k(y_r) &= \sqrt{P_r} f^{-1} \left( \frac{1}{2} s \left( \frac{y_r}{\Delta_k} \right) + \frac{1}{2} \right), \tag{10.10} \end{align*}
\]

where
• $\Delta_k := \Delta_0 2^{-k}$ and $\Delta_0 > 0$ is a constant;
• $s(x)$ is a periodic function with period one (i.e., $s(x+1) = s(x)$) and $s(x) = 2x$ for $-\frac{1}{2} \leq x < \frac{1}{2}$; and
• $F^{-1}(x)$ is the inverse of $F(x) = 1 - Q(x)$ where $Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$.

By this choice of the relay mapping, the density of $X_r$ approaches $\mathcal{N}(0, P_r)$ as $k$ increases. Now by applying Theorem 6.2.1, we obtain
\[
\lim_{k \to \infty} R_{1k} = \lim_{k \to \infty} h\left(f_k(X_1 + X_2 + Z_2|X_2)\right) = \frac{1}{2} \log(2\pi e (P_r + N_2)).
\]
Similarly, we have $\lim_{k \to \infty} R_{2k} = \frac{1}{2} \log(2\pi e (P_r + N_1))$. Thus, using the above achievability scheme and the cutset bound, we have the following proposition.

**Proposition 10.5.1.** Instantaneous analog noisy network coding for the Gaussian TWRC with a noiseless uplink achieves the capacity
\[
R_{NL} = \left\{(R_1, R_2) : R_1 \leq C\left(\frac{P_r}{N_2}\right), R_2 \leq C\left(\frac{P_r}{N_1}\right)\right\}.
\]

**10.5.2 Noisy Uplink**

We next consider the general case. In order to optimize the analog network coding strategy, we develop a numerical iterative algorithm for optimizing the sum-rate similar to that in Chapter 5. We first construct the Lagrangian under the power constraint:
\[
J = I(X_1; Y_2|X_2) + I(X_2; Y_1|X_1) + \lambda(\mathbb{E}[f^2(y_r)] - P_r),
\]
where $\lambda$ is a positive real number. Computation of the necessary condition $\frac{\partial J}{\partial f} = 0$ for optimality, after some manipulation gives
\[
f(y_r) = \frac{1}{\lambda p(y_r)} \sum_{j \in \{1, 2\}} \left[ \int \int p(x_j)\left[1 + \log p(y_j|x_j)\right] \frac{\partial p(y_j|x_j)}{\partial f} dy_j dx_j - \int \int p(t)\left[1 + \log p(y_j|t)\right] \frac{\partial p(y_j|t)}{\partial f} dy_j dt \right],
\]
where $t = x_1 + x_2$ and
\[
\frac{\partial p(y_j|x_j)}{\partial f} = \frac{y_j - f(y_r)}{N_j} p(y_j|y_r)p(y_r|x_j), \quad \frac{\partial p(y_j|t)}{\partial f} = \frac{y_j - f(y_r)}{N_j} p(y_j|y_r)p(y_r|t).
\]
In order to solve (10.13), we use the fixed-point iteration method with an initial function $f_0$ to compute \{f_k\}. In the following, we choose Gaussian inputs: $X_1 \sim \mathcal{N}(0, P_1)$ and $X_2 \sim \mathcal{N}(0, P_2)$. Figure 10.2 shows the optimized mappings for $P_1 = 2$, $P_2 = 1$, $P_r = 2$, $N_1 = 0$, $N_2 = 5$ and $N_r = -N$ dB, where $N \in \{9, 11, 12, 16, 25\}$. We use two different initialization: $f_{0,1}(y_r) = \kappa_1 y_r$ and $f_{0,2} = \kappa_2 \sin(y_r)$ where $\kappa_1$ and $\kappa_2$ are chosen to ensure power constraint at the relay. From Figure 10.2, we see that the algorithm converges. We additionally observe that the convergence of the algorithm is faster when it is initialized with $f_{0,2}$. We also note that the optimized curves with different initializations are not the same in general. However, the associated optimized rates are very close. Unlike AF, the optimized mappings are not invertible and exhibit a periodic behavior. We see that as the received signal at the relay becomes less noisy, the optimized mapping gets more zero crossings. Finally, we see that the shape of optimized mapping when $N_r = -25$ dB is similar to that for the noiseless case, whose equation is given in (10.10).

Figure 10.3 compares the achievable sum-rate for $P_1 = 2$, $P_2 = 1$, $P_r = 2$, $N_1 = 0$, $N_2 = 5$, $N_r = -N$ dB as a function of $N$. We see that optimized nonlinear ANNC significantly improves on AF. Additionally, nonlinear ANNC achieves the rates close to those achieved by the layered NNC. It is also worth noting that ANNC outperforms NNC when $N_r$ is high.

10.6 Concluding Remarks

We presented an analog noisy network coding (ANNC) scheme for the TWRC and demonstrated that the achievable rates of optimized instantaneous ANNC are superior to those achieved by AF. In fact, AF can be considered as a particular instantaneous ANNC which is suboptimal in general. Optimized instantaneous ANNC, which is nonlinear and periodic, outperforms AF because it is more power efficient and utilizes the side information in a distributed manner. In other words, ANNC transmits a compressed version of the received signal. ANNC also outperforms CF and provides rates close to those achieved by the layered NNC. ANNC, unlike CF and NNC, enjoys a joint analog source–channel coding approach which is potentially more powerful than digital compression. The current work hence motivates further research to develop new noisy network coding schemes for which the relays employ a joint source–channel coding scheme. Another useful direction would be to combine the proposed ANNC protocol with DF [RW06, OSBB08] or lattice encoding [NCL10, WNPS10].
Figure 10.2: Optimized mappings and convergence of the fixed point method when the channel parameters are set to $P_1 = 2$, $P_2 = 1$, $P_r = 2$, $N_1 = 0$, $N_2 = 5$, $N_r = -N$ dB, $N = \{9, 11, 12, 16, 25\}$ from top to bottom. Dashed and solid curves respectively show the result when the algorithm is initialized with $f_{0,1}(y_r) = \kappa_1 y_r$ and $f_{0,2} = \kappa_2 \sin(y_r)$ where $\kappa_1$ and $\kappa_2$ are chosen to ensure power constraint at the relay.
Figure 10.3: Comparison of achievable sum-rate for the Gaussian TWRC where $P_1 = 2$, $P_2 = 1$, $P_r = 2$, $N_1 = 0$, $N_2 = 5$, $N_r = -N$ dB.
Chapter 11

Hybrid Noisy Network Coding

11.1 Introduction

In this chapter, we focus on noisy network coding (NNC) type of schemes for the two-way relay channel (TWRC). That is, the relay does not know the codebook used at the users. Methods based on NNC can be divided into two main categories: analog and digital. Analog NNC operates based on an analog-to-analog interface. One well-known example in this category is linear relaying where the relay transmits the scaled version of the noisy received signal [RW06, PY07, KMG+07]. Digital NNC is however constructed based on an analog-to-digital interface. The compress-and-forward (CF) type of scheme in [RW06], which is an extension of the original CF approach in [CG79], and the NNC scheme proposed in [LKGC11] belong to this category. We studied aforementioned analog and digital NNC schemes in Chapter 10 for a Gaussian TWRC.

In this chapter, we introduce three new protocols based on the combination of analog and digital NNC. We investigated hybrid digital–analog (HDA) relaying [YS09] in Chapter 7 for three-node relay channels. In the following, we however propose three schemes based on HDA relaying for TWRC. In the proposed schemes, the relay first employs digital compression to generate a digital signal. The relay then linearly and component-wise combines the encoded compression index with the analog noisy received signal. Power splitting between the digital and analog parts is employed in order to optimize the achievable rate region. We refer to our proposed schemes as hybrid digital–analog NNC protocols, since the digital signal used for encoding of the compression index is combined with the analog noisy signal in the network. Our main conclusion is that the new hybrid NNC approach outperforms both purely-digital and purely-analog NNC.

The remaining part of the chapter is organized as follows. Section 11.2 presents the discrete memoryless TWRC that we consider in this chapter. Section 11.3 introduces three hybrid digital–analog NNC schemes. Section 11.4 evaluates the achievable rate regions of the proposed hybrid NNC schemes for a Gaussian TWRC. Section 11.5 numerically compares the achievable sum-rate of the proposed hybrid schemes with that of already existing NNC schemes. In particular, in Section 11.5,
11.2 Channel Model

Figure 11.1 illustrates the discrete memoryless TWRC that we investigate in this chapter. The users wish to exchange two messages $M_1 \in \mathcal{M}_1^n$ and $M_2 \in \mathcal{M}_2^n$ in $n$ channel uses with a shared relay node. We assume that the messages are independent from each other and are uniformly chosen from $\mathcal{M}_1$ and $\mathcal{M}_2$.

Let $i \in [1:n]$ denotes the $i$th transmission instant. The users transmit symbols $X_{1i}$ and $X_{2i}$ using encoders $\alpha_1$ and $\alpha_2$, respectively. The relay receives $Y_{ri}$ and employs an encoder to generate the symbol $X_{ri}(Y_{ri})$. That is the relay operates causally (see also Table 7.1). The user $j \in \{1, 2\}$ then receives $Y_{ji}$. The users then employ decoders $\beta_1$ and $\beta_2$ respectively to decode the transmitted messages.

Similar to the previous chapter, the rate pair $(R_1, R_2)$ is achievable if there exists a communication strategy such that the average message error probability

$$P_e^{(n)} = \Pr\{(M_1, M_2) \neq (\hat{M}_1, \hat{M}_2)\},$$

approaches zero as $n \to \infty$ and at the same time

$$\liminf_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_j^{(n)}| \geq R_j, \quad \forall j \in \{1, 2\}. \quad (11.1)$$
In this chapter, we assume that the two users are separated and therefore cannot overhear each other such that
\[ p(y_1, y_2, y_r|x_1, x_2, x_r) = p(y_1, y_2|x_r)p(y_r|x_1, x_2). \] (11.2)
That is the network is constituted of an uplink channel with the pmf \( p(y_r|x_1, x_2) \) and a downlink channel with the pmf \( p(y_1, y_2|x_r) \). Note that for such configuration the causal operation at the relay is meaningful.

11.3 Hybrid NNC

We next propose three hybrid NNC schemes.

11.3.1 Scheme 1

In this scheme the relay generates a compression codebook using the Wyner-Ziv principle for one user and then it combines the encoded compression index with the noisy analog signal. The scheme is constructed in a way that one user only recovers the digital signal and the other user treats the digital signal as a noise.

**Theorem 11.3.1.** The rate region \( \mathcal{R}_{H_1} = \mathcal{R}_1 \cup \mathcal{R}_2 \) is achievable where
\[
\mathcal{R}_1 = \left\{ (R_1, R_2) : R_1 \leq I(X_1; Y_2, \hat{Y}_r|X_2, V), \right.
\]
\[
R_2 \leq I(X_2; Y_1|X_1),
\]
\[
I(Y_r; \hat{Y}_r|X_2, V, Y_2) \leq I(V; Y_2|X_2) \\right\}, \] (11.3)
a nd
\[
\mathcal{R}_2 = \left\{ (R_1, R_2) : R_1 \leq I(X_1; Y_2|X_2), \right.
\]
\[
R_2 \leq I(X_2; Y_1, \hat{Y}_r|X_1, V),
\]
\[
I(Y_r; \hat{Y}_r|X_1, V, Y_1) \leq I(V; Y_1|X_1) \\right\}, \] (11.4)
for some \( p(x_1)p(x_2)p(v)p(\hat{y}_r|y_r), x_r = f(y_r, v) \).

**Proof.** Fix \( p(x_1)p(x_2)p(\hat{x}_r)p(\hat{y}_r|y_r) \) and \( x_r = f(v, y_r) \), and let \( \epsilon > \epsilon_r \).

- Codebook generation:
  i) Generate \( 2^{nR_1} \) sequences \( x_1(m_1) \) for \( m_1 \in [1 : 2^{nR_1}] \) each according to \( \prod_{i=1}^n p_{X_1}(x_{1i}). \)
  ii) Generate \( 2^{nR_2} \) sequences \( x_2(m_2) \) for \( m_2 \in [1 : 2^{nR_2}] \) each according to \( \prod_{i=1}^n p_{X_2}(x_{2i}). \)
iii) Generate $2^{n(R+\hat{R})}$ sequences $\hat{y}_r(k,l)$ for $k \in [1 : 2^n\hat{R}]$ and $l \in [1 : 2^n\hat{R}]$ each according to $\prod_{i=1}^{n} p_{\hat{Y}_r}(\hat{y}_r_i)$.

iv) Generate $2^n\hat{R}$ sequences $v(l)$ for $l \in [1 : 2^n\hat{R}]$ each according to $\prod_{i=1}^{n} p_{V}(v_i)$.

- **Encoding:**
  i) User 1 transmits $x_1(m_1)$.
  ii) User 2 transmits $x_2(m_2)$.
  iii) The relay looks for an index pair $(k,l)$ such that $(\hat{Y}_r(k,l), Y_r) \in T^{(n)}$. If there is no such pair it declares an error. If there are multiple pairs, it picks one in random. The encoding is successful if
  $$\hat{R} + \tilde{R} > I(Y_r; \hat{Y}_r) + \delta(\epsilon_r) \quad (11.5)$$
  The relay then transmits $x_r$ where $x_{rs} = f(v(l), y_{rs})$.

- **Decoding:**
  i) User 1 looks for a unique index $\hat{m}_2$ such that
  $$(X_2(\hat{m}_2), Y_1, X_1) \in T_t^{(n)}.$$ 
  If there is no such index it declares an error. The probability of decoding error is negligible if
  $$R_2 < I(X_2; Y_1|X_1) - \delta(\epsilon_1). \quad (11.6)$$
  ii) User 2 first finds a unique index $\hat{l}$ such that
  $$(V(\hat{l}), Y_2, X_2) \in T_a^{(n)}.$$ 
  If there is no such index it declares an error. This is done with a negligible probability of error if
  $$\hat{R} < I(V; Y_2|X_2) - \delta(\epsilon_a). \quad (11.7)$$
  iii) Having found $l$, User 2 looks for a unique index $\hat{k}$ such that
  $$(\hat{Y}_r(\hat{k},l), V(l), Y_2, X_2) \in T_e^{(n)}.$$ 
  If there is no such index it declares an error. This is done with an arbitrarily small probability of error if
  $$\tilde{R} < I(\hat{Y}_r; Y_2, X_2, V) - \delta(\epsilon). \quad (11.8)$$
iv) Having found $k$ and $l$, User 2 finally looks for a unique index $\hat{m}_1$ such that $(X_1(\hat{m}_1), \hat{Y}_r(k, l), Y_2, V(l), X_2) \in T^{(n)}_{2}$. If there is no such index it declares an error. The probability of error is negligible if
\[
R_2 < I(X_1; Y_2, \hat{Y}_r | X_2, V) - \delta(\epsilon_2). \tag{11.9}
\]
Combining the rate constraints in (11.5)–(11.9), we conclude that the rate region
\[
\mathcal{R}_1 = \{(R_1, R_2) : R_1 \leq I(X_1; Y_2, \hat{Y}_r | X_2, V),
R_2 \leq I(X_2; Y_1 | X_1),
I(Y_r; \hat{Y}_r | X_2, V, Y_2) \leq I(V; Y_2 | X_2)\}, \tag{11.10}
\]
for some $p(x_1)p(x_2)p(v)p(\hat{y}_r | y_r)$ and $x_r = f(v, y_r)$ is achievable. By changing the role of User 1 and User 2, we can obtain the following rate region
\[
\mathcal{R}_2 = \{(R_1, R_2) : R_1 \leq I(X_1; Y_2 | X_2),
R_2 \leq I(X_2; Y_1, \hat{Y}_r | X_1, V),
I(Y_r; \hat{Y}_r | X_1, V, Y_1) \leq I(V; Y_1 | X_1)\}. \tag{11.11}
\]
This completes the proof.

**Remark 11.3.1.** One can also achieve the rate region in Theorem 11.3.1 without using Wyner-Ziv compression. See Appendix 11.A for an alternative relaying scheme using quantize-and-forward relaying.

### 11.3.2 Scheme 2

In a similar fashion, we can apply the hybrid strategy as discussed in Scheme 1, to the CF scheme in [RW06]. In this scheme the relay employs Wyner-Ziv compression and both users decode the digital signal. That is the compression and the binning rates are chosen such that both users can construct the reproduction sequence generated at the relay (see also Figure 2.6).

**Theorem 11.3.2.** The rate region
\[
\mathcal{R}_{CF} = \{(R_1, R_2) : R_1 \leq I(X_1; Y_2, \hat{Y}_r | X_2, V),
R_2 \leq I(X_2; Y_1, \hat{Y}_r | X_1, V),
\max \left(I(Y_r; \hat{Y}_r | X_1, V, Y_1), I(Y_r; \hat{Y}_r | X_2, V, Y_2)\right)
\leq \min (I(V; Y_1 | X_1), I(V; Y_2 | X_2))\}, \tag{11.12}
\]
is achievable for some $p(x_1)p(x_2)p(v)p(\hat{y}_r | y_r, v)$ and $x_r = f(v, y_r)$.
Proof. The proof follows by combining CF in [RW06] with the Shannon strategy [Sha58] at the relay.

11.3.3 Scheme 3

We next apply the HDA protocol to the purely-digital NNC scheme in [LKGC11]. This scheme does not employ Wyner-Ziv compression.

Theorem 11.3.3. The rate region

\[ R_{\text{NNC}} = \left\{ (R_1, R_2) : R_1 \leq I(X_1; Y_2, \hat{Y}_r | X_2, V), \right. \\
\left. R_1 \leq I(X_1, V; Y_2 | X_2) - I(Y_r; \hat{Y}_r | X_1, X_2, V, Y_2), \right. \\
\left. R_2 \leq I(X_2; Y_1 | X_1, V), \right. \\
\left. R_2 \leq I(X_2, V; Y_1 | X_1) - I(Y_r; \hat{Y}_r | X_1, X_2, V, Y_1) \right\} \tag{11.13} \]

is achievable for some \( p(x_1)p(x_2)p(v)p(\hat{y}_r | y_r, v) \) and \( x_r = f(v, y_r) \).

Proof. The proof follows by combining NNC in [LKGC11] with the Shannon strategy [Sha58] at the relay.

11.4 The Gaussian TWRC

We next evaluate the achievable hybrid rate regions for the Gaussian channel model in Chapter 10. We first recall the channel model.

11.4.1 Channel Model

Figure 11.2 illustrates the Gaussian TWRC that we consider in this Section. The users transmit symbols \( X_{1i} \) and \( X_{2i} \) using encoders \( \alpha_1 \) and \( \alpha_2 \), respectively. The relay receives

\[ Y_{ri} = X_{1i} + X_{2i} + Z_{ri}, \tag{11.14} \]

where \( Z_{ri} \sim \mathcal{N}(0, N_r) \) is the additive Gaussian noise at the relay. The relay employs the encoder \( f \) to generate the symbol \( X_{ri} \). The user \( j \in \{1, 2\} \) then receives

\[ Y_{ji} = X_{ri} + Z_{ji}, \tag{11.15} \]

where \( Z_{ji} \sim \mathcal{N}(0, N_j) \) denotes the additive Gaussian noise at the \( j \)th user. We assume that all noise sequences are white and mutually independent from each other. We additionally assume that the users and the relay operate under average power constraints, i.e. \( \sum_{i=1}^{n} E[X_{ri}^2] \leq nP_r, \sum_{i=1}^{n} E[X_{ji}^2] \leq nP_j \) for \( j \in \{1, 2\} \).
11.4. Achievable Rate Regions

Proposition 11.4.1. The rate region $\mathcal{R}_H = \mathcal{R}_1 \cup \mathcal{R}_2$ is achievable where

$$\mathcal{R}_1 = \bigcup_{\gamma \in [0, 1]} \left\{ (R_1, R_2) : R_1 \leq C \left( \frac{\gamma \kappa P_2}{\gamma P_r + \gamma \kappa N_r + N_1} \right), \right. \right.$$ 

$$R_2 \leq C \left( \frac{P_1 N_2 + \gamma \kappa P_1 N_q}{N_r N_2 + \gamma \kappa N_r N_q + N_2 N_q} \right),$$

$$N_q = \frac{(P_1 + N_r) N_2}{\gamma P_r} \right\}, \tag{11.16}$$

$$\mathcal{R}_2 = \bigcup_{\gamma \in [0, 1]} \left\{ (R_1, R_2) : R_1 \leq C \left( \frac{\tilde{\gamma} \kappa P_1}{\gamma P_r + \gamma \kappa N_r + N_2} \right), \right. \right.$$ 

$$R_2 \leq C \left( \frac{P_2 N_1 + \gamma \kappa P_2 N_q}{N_r N_1 + \gamma \kappa N_r N_q + N_1 N_q} \right),$$

$$N_q = \frac{(P_2 + N_r) N_1}{\gamma P_r} \right\}, \tag{11.17}$$

$\tilde{\gamma} := 1 - \gamma$ and $C(x) := \frac{1}{2} \log_2 (1 + x)$.

Proof. Let $X_1 \sim \mathcal{N}(0, P_1)$, $X_2 \sim \mathcal{N}(0, P_2)$, $V \sim \mathcal{N}(0, \gamma P_r)$. Then let $\hat{Y}_r = Y_r + Z_q$ where $Z_q \sim \mathcal{N}(0, N_q)$ is independent of other random variables and

$$X_r = X_{rd} + X_{ra} = \sqrt{\gamma} V + \sqrt{\kappa} \gamma Y_r \tag{11.18}$$

where $\gamma$ is a power splitting parameter and $\kappa = P_r (P_1 + P_2 + N_r)^{-1}$ is a power normalization factor to ensure the power constraint at the relay. See Figure 11.3 for an illustration of the hybrid relaying.
By these choices of random variables, using Theorem 11.3.1 one can verify that the rate region in Proposition 11.4.1 is achievable. (See Appendix 11.B for the computation of $R_1$.) This completes the proof.

By choosing $\gamma = 0$, we can recover the achievable rate region of linear relaying, which is given by

$$
\mathcal{R}_L = \left\{ (R_1, R_2) : R_1 \leq C \left( \frac{P_1 P_r}{P_r N_r + P_1 N_2 + P_2 N_2 + N_2 N_r} \right) \right\},
$$

$$
R_2 \leq C \left( \frac{P_2 P_r}{P_r N_r + P_1 N_1 + P_2 N_1 + N_1 N_r} \right) \right\}. \quad (11.19)
$$

Thus $\mathcal{R}_L \subseteq \mathcal{R}_{H_1}$.

Figure 11.4 shows an illustration of the receiver at User 1. User 1 first removes $\sqrt{\alpha_r}X_1$ from the received signal $Y_1$ to obtain $\tilde{Y}_1$ and it then decodes the transmitted message using $\beta_1$. Figure 11.5 shows an illustration of the receiver at User 2. User 2 first removes $\sqrt{\gamma}X_2$ from the received signal $Y_2$ to obtain $\tilde{Y}_2$. It then decodes the digital signal $X_{rd}$ using $\beta_{21}$. Using the side information and $X_{rd}$, User 2 then employs $\beta_{22}$ to recover the reproduction signal $\hat{Y}_r = Y_r + Z_q$ where $Z_q$ denotes the reproduction noise. Finally, User 2 finds $\tilde{Y}_2 = \hat{Y}_2 - X_{rd}$ and $\tilde{Y}_r = Y_r - X_2$, and decodes the transmitted message using $\beta_{23}$. 
11.4. The Gaussian TWRC

Figure 11.5: Illustration of the receiver at User 2.

Proposition 11.4.2. The rate region $\mathcal{R}_{H_2}$ is achievable where

$$\mathcal{R}_{H_2} = \bigcup_{\gamma \in [0,1]} \left\{ (R_1, R_2) : R_1 \leq C \left( \frac{P_1 N_2 + \gamma \kappa P_1 N_q}{N_r N_2 + \gamma \kappa N_r N_q + N_2 N_q} \right), \right. \right.$$  

$$R_2 \leq C \left( \frac{P_2 N_1 + \gamma \kappa P_2 N_q}{N_r N_1 + \gamma \kappa N_r N_q + N_1 N_q} \right),$$

$$N_q = \frac{1}{\gamma P_r} \max \left( \frac{(P_1 + N_r)N_2}{N_2 + \gamma \kappa(P_1 + N_r)}, \frac{(P_2 + N_r)N_1}{N_1 + \gamma \kappa(P_2 + N_r)} \right),$$

$$\times \max (N_1 + \gamma \kappa(P_2 + N_r), N_2 + \gamma \kappa(P_1 + N_r)) \left\} \right.$$

(11.20)

Proof. The proof follows by Theorem 11.3.2 and using the same random variables as those in Proposition 11.4.1.

By choosing $\gamma = 0$, $\mathcal{R}_{H_2}$ simplifies to $\mathcal{R}_L$ given in (11.19). By choosing $\gamma = 1$, $\mathcal{R}_{H_2}$ reduces to the achievable rate region of the CF scheme in [RW06], which is given by

$$\mathcal{R}_{CF} = \left\{ (R_1, R_2) : R_1 \leq C \left( \frac{P_1 P_r}{P_r N_r + [\max(P_1, P_2) + N_r] \max(N_1, N_2)} \right), \right.$$  

$$R_2 \leq C \left( \frac{P_2 P_r}{P_r N_r + [\max(P_1, P_2) + N_r] \max(N_1, N_2)} \right) \left\} \right.$$  

Thus $\mathcal{R}_L \subseteq \mathcal{R}_{H_2}$ and $\mathcal{R}_{CF} \subseteq \mathcal{R}_{H_2}$. 
Proposition 11.4.3. The rate region $R_{H_3}$ is achievable where

$$R_{H_3} = \bigcup_{\gamma \in [0,1]} \left\{ (R_1, R_2) : \begin{array}{l}
R_1 \leq C \left( \frac{\gamma P_r + \kappa \bar{\gamma} P_1}{N_\gamma (\gamma N_r + N_2)} \right) - C \left( \frac{N_r N_2}{N_q (\gamma N_r + N_2)} \right), \\
R_2 \leq C \left( \frac{\gamma P_r + \kappa \bar{\gamma} P_2}{N_\gamma (\gamma N_r + N_1)} \right) - C \left( \frac{N_r N_1}{N_q (\gamma N_r + N_1)} \right), \\
R_1 \leq C \left( \frac{P_1 N_2 + \kappa \bar{\gamma} P_1 N_q}{N_r N_2 + \gamma \kappa N_r N_q + N_2 N_q} \right), \\
R_2 \leq C \left( \frac{P_2 N_1 + \bar{\gamma} \kappa P_2 N_q}{N_r N_1 + \gamma N_r N_q + N_1 N_q} \right), N_q > 0 \right\}. 
\right. \right.$$

(11.21)

Proof. The proof follows by Theorem 11.3.3 and using the same random variables as those in Proposition 11.4.1.

By choosing $\gamma = 0$, $R_{H_3}$ simplifies to $R_L$ given in (11.19). By choosing $\gamma = 1$, $R_{H_3}$ reduces to the achievable rate region of the NNC scheme in [LKGC11], which is given by

$$R_{NNC} = \left\{ (R_1, R_2) : \begin{array}{l}
R_1 \leq C \left( \frac{P_1}{N_r} \right), \\
R_1 \leq C \left( \frac{P_2}{N_r} \right), \\
R_2 \leq C \left( \frac{N_r}{N_q} \right), N_q > 0 \right\}. 
\right. \right.$$

Thus $R_L \subseteq R_{H_3}$ and $R_{NNC} \subseteq R_{H_3}$.

11.5 Numerical Examples

We next numerically evaluate the achievable sum-rate of various noisy network coding schemes. We assume that User 1, User 2 and the relay are located on a straight line, where the distance of User 1 to the relay is $d \in [0,1]$ and the distance of the relay to User 2 is $1-d$. In order to take into account the geometry of the network, we let the channel parameters be $P_1 = P d^{-3}$, $P_2 = P(1-d)^{-3}$, $N_1 = d^4$, $N_2 = (1-d)^3$ and $N_r = 1$. To avoid crowded figures, we plot the maximum sum-rate achieved by the three proposed hybrid schemes.

Case 1: In the first case we set $P = 5$ and $P_r = 10$ dB. Figure 11.6 shows the sum-rate of different schemes. (The achievable rate of the hybrid scheme, denoted in
the figure, is the maximum of those achieved by the three proposed hybrid schemes.)
We see that hybrid NNC improves on CF, AF and NNC. For small $d$, hybrid Scheme 1 performs best, and for larger $d$ hybrid Scheme 2 outperforms the others. Hybrid Scheme 3 does not provide any gain within the numerical precision.

Case 2: In the second case we set $P = 10$ and $P_r = 5$ dB. Figure 11.7 shows the sum-rate of different schemes. We see that hybrid NNC achieves higher sum-rates than those achieved using CF, AF and NNC schemes. For small $d$ hybrid Scheme 1 performs best, and for larger $d$ hybrid Scheme 3 outperforms others. Hybrid Scheme 2 provides negligible gain but still less than the two other hybrid schemes.

Case 3: In the third case we set $P = 5$ and $P_r = -20$ dB. Figure 11.8 shows the sum-rate of different schemes. We again see that hybrid NNC outperforms CF, AF and NNC. For small $d$, hybrid Scheme 1 performs best, and for larger $d$ hybrid Scheme 3 outperforms the others.

From the above numerical examples, we observe that the performance of the hybrid NNC schemes actually coincides with the cutset bound when $d \to 0$. In fact we have the following result.

**Corollary 11.5.1.** For the above geometric network, when $d \to 0$ the capacity region is given by

$$
R = \left\{ (R_1, R_2) : R_1 \leq C \left( \frac{P_r}{N_2} \right), R_2 \leq C \left( \frac{P_2}{N_r} \right) \right\}.
$$

(11.22)

**Proof.** The achievability follows from the rate region in (11.17) and the converse follows from the cutset bound.

**11.6 Concluding Remarks**

We presented three hybrid noisy network coding schemes for the TWRC. We demonstrated that for a Gaussian TWRC, hybrid relaying outperforms AF (linear analog network coding), CF, and NNC for the Gaussian TWRC.
Figure 11.6: Comparison of achievable sum-rate for the Gaussian TWRC where $P = 5$ and $P_r = 10$ dB.
Figure 11.7: Same as Figure 11.6 but for $P = 10$ and $P_r = 5$ dB.
Figure 11.8: Same as Figure 11.6 but for $P = 5$ and $P_r = -20$ dB.
11.A Equivalent Derivation of $R_1$

Fix $p(x_1)p(x_2)p(v)p(y_r|y_r)$ and let $\epsilon_2 > \epsilon_r$.

- **Codebook generation:**
  
  i) Generate $2^{nR_1}$ sequences $x_1(m_1)$ for $m_1 \in [1 : 2^{nR_1}]$ each according to $\prod_{i=1}^{n} P_{X_1}(x_{1i})$.
  
  ii) Generate $2^{nR_2}$ sequences $x_2(m_2)$ for $m_2 \in [1 : 2^{nR_2}]$ each according to $\prod_{i=1}^{n} P_{X_2}(x_{2i})$.
  
  iii) Generate $2^{n\hat{R}}$ sequences $\hat{y}_r(l)$ for $l \in [1 : 2^{n\hat{R}}]$ each according to $\prod_{i=1}^{n} P_{\hat{Y}_r}(\hat{y}_{ri})$.
  
  iv) Generate $2^{n\hat{R}}$ sequences $\hat{v}(l)$ for $l \in [1 : 2^{n\hat{R}}]$ each according to $\prod_{i=1}^{n} P_{\hat{V}}(v_{i})$.

- **Encoding:**
  
  i) User 1 transmits $x_1(m_1)$.
  
  ii) User 2 transmits $x_2(m_2)$.
  
  iii) The relay looks for the smallest index $l$ such that $(\hat{Y}_r(l), Y_r) \in T_{\epsilon_r}^{(n)}$. If there is no such index it declares an error. The encoding is successful if

  \[ \hat{R} > I(Y_r; \hat{Y}_r) + \delta(\epsilon_r). \]  
  \[ (11.23) \]

  The relay then transmits $x_r$ where $x_r = f(v_1(l), y_r)$.

- **Decoding:**
  
  i) User 1 looks for a unique index $m_2$ such that

  \[ (X_2(m_2), Y_1, X_1) \in T_{\epsilon_1}^{(n)}. \]

  If there is no such index it declares an error. The probability of decoding error is negligible if

  \[ R_2 < I(X_2; Y_1 | X_1) - \delta(\epsilon_1). \]  
  \[ (11.24) \]

  ii) User 2 looks for a unique index $m_1$ such that

  \[ (X_1(m_1), \hat{Y}_r(l), Y_2, V(l), X_2) \in T_{\epsilon_2}^{(n)}, \]

  for some $l$. If there is no such index it declares an error. Define

  \[ A(m_1,l) := \{(X_1(m_1), \hat{Y}_r(l), Y_2, V(l), X_2) \in T_{\epsilon_2}^{(n)}\}, \]
and assume that the User 2 has transmitted \( m_2 = 1 \) and the relay has chosen the compression index \( L \). The probability of error under the condition that the encoding at the relay was successful, can be bounded as

\[
P_e(n) \leq P_e^n(A^c(1, L)) + P^n \left( \bigcup_{m_2 \neq 1} A(m_2, l) \right)
\]  

(11.25)

The first term \( P_e^n(A^c(1, L)) \) goes to zero as \( n \to \infty \). The second term can be bounded as follows

\[
P^n \left( \bigcup_{m_2 \neq 1} A(m_2, l) \right) \leq \sum_{m_2 \neq 1} P^n(A(m_2, l))
\]

\[
\leq \sum_{m_2 \neq 1, l = L} P^n(A(m_2, l)) + \sum_{m_2 \neq 1, l \neq L} P^n(A(m_2, l))
\]

\[
\leq 2^n(R_1 - I(X_1; Y_r | X_2) - \delta(\epsilon_2)) + 2^n(R_1 + R - I(X_1, V; Y_r | X_2) - I(X_1, X_2; Y_r | V) + \delta(\epsilon_2))
\]

(11.26)

Using (11.24) and (11.26) we conclude that the probability of error goes to zeros if the rate \( R_1 \) satisfies

\[
R_1 < I(X_1; Y_r | X_2) - \delta(\epsilon_2),
\]

(11.27)

\[
R_1 < I(X_1, V; Y_r | X_2) - I(Y_r; Y_r | V, X_1, X_2, Y_2) - \delta(\epsilon_2) - \delta(\epsilon_r).
\]

(11.28)

Combining the above rate constraints, we obtain the following achievable rate region

\[
\mathcal{R}_{11} = \{(R_1, R_2) : R_2 \leq I(X_2; Y_2 | X_1),
\ R_1 \leq I(X_1; Y_2, \hat{Y}_r | X_2, V),
\ R_1 \leq I(X_1, V; Y_r | X_2) - I(Y_r; \hat{Y}_r | V, X_1, X_2, Y_2)\},
\]

(11.29)

for some \( p(x_1)p(x_2)p(v)p(y_r | y_r) \) and \( x_r = f(v, y_r) \). One can verify that \( \mathcal{R}_{11} \) is equivalent to \( \mathcal{R}_1 \) in (11.3).
11.B Computation of $R_1$

In the following, we compute the rate region $R_1$ in (11.3). Consider

$$R_1 \leq I(X_1; Y_2, \hat{Y}_r|X_2, V)$$
$$= h(Y_2, \hat{Y}_r|X_2, V) - h(Y_2, \hat{Y}_r|X_1, X_2, V)$$
$$= h(\sqrt{\gamma} V + \sqrt{\kappa} Y_r + Z_2, Y_r + Z_q|X_2, V)$$
$$- h(\sqrt{\gamma} V + \sqrt{\kappa} Y_r + Z_2, Y_r + Z_q|X_1, X_2, V)$$
$$= h(\sqrt{\kappa} Y_1 + Z_r) + Z_2, X_1 + Z_r + Z_q) - h(\sqrt{\kappa} Z_r + Z_2, Z_r + Z_q)$$
$$= \frac{1}{2} \log \left( \frac{P_1 + N_r + N_q}{N_r + N_q} \right)$$
$$= \frac{1}{2} \log \left( \frac{(P_1 + N_r + N_q)(\sqrt{\kappa} P_1 + N_r) + N_2}{(N_r + N_q)(\sqrt{\kappa} N_r + N_2) - \kappa^2 N_2^2} \right)$$
$$= C \left( \frac{N_r N_2 + \kappa^2 P_1 N_q}{N_r N_2 + \kappa^2 N_r N_q + N_2 N_q} \right), \quad (11.30)$$

$$R_2 \leq I(X_2; Y_1|X_1) = h(Y_1|X_1) - h(Y_1|X_1, X_2)$$
$$= h(\sqrt{\gamma} V + \sqrt{\kappa} Y_r + Z_1|X_1) - h(\sqrt{\gamma} V + \sqrt{\kappa} Y_r + Z_1|X_1, X_2)$$
$$= h(\sqrt{\gamma} V + \sqrt{\kappa} Y(X_2 + Z_r) + Z_1) - h(\sqrt{\gamma} V + \sqrt{\kappa} Z_r + Z_1)$$
$$= C \left( \frac{\kappa^2 P_1}{P_1 + \kappa^2 N_r + N_1} \right), \quad (11.31)$$

$$\bar{R} \geq I(Y_r; \hat{Y}_r|V, X_2, Y_2)$$
$$= h(\hat{Y}_r|V, X_2, Y_2) - h(\hat{Y}_r|V, X_2, Y_2, Y_r)$$
$$= h(X_1 + X_2 + Z_2, Y_r|V, X_2, \sqrt{\gamma} V + \sqrt{\kappa^2} Y_r + Z_2) - h(Z_q)$$
$$= h(X_1 + X_2 + Z_2, Y_q|\sqrt{\kappa^2}(X_1 + Z_r) + Z_2) - h(Z_q)$$
$$= h(X_1 + Z_r + Z_q, \sqrt{\kappa^2}(X_1 + Z_r) + Z_2) - h(\sqrt{\kappa^2}(X_1 + Z_r) + Z_q)$$
$$- h(Z_q)$$
$$= \frac{1}{2} \log \left( \frac{P_1 + N_r + N_q}{N_q(\sqrt{\kappa^2} P_1 + N_r)} \right)$$
$$= C \left( \frac{(P_1 + N_r) N_2}{N_q(\sqrt{\kappa^2} P_1 + N_r) + N_2} \right), \quad (11.32)$$
and

$$\hat{R} \leq I(V; Y_2 | X_2) = h(Y_2 | X_2) - h(Y_2 | X_2, V)$$

$$= h(\sqrt{\gamma V} + \sqrt{\kappa \gamma}(X_1 + Z_r) + Z_2) - h(\sqrt{\kappa \gamma}(X_1 + Z_r) + Z_2)$$

$$= C \left( \frac{\gamma P_r}{\kappa \gamma(P_1 + N_r) + N_2} \right).$$

(11.33)

Combining (11.30)–(11.33) and optimizing over $N_q$ yields the rate region in (11.16).
12.1 Concluding Remarks

In the thesis, we discussed various fundamental aspects of reliable communication over relay networks. We investigated three-node relay channels, multiple-access relay channels and two-way relay channels. We studied several existing relaying protocols and proposed some novel protocols. The investigated relaying protocols in the thesis can be divided into two main categories based on the amount of the memory required for relay processing: protocols with finite memory and protocols with infinite memory. For the former case we studied both linear and nonlinear relaying. For the latter case we considered decode-and-forward (DF), partial decode-and-forward (PDF), compress-and-forward (CF), noisy network coding (NNC) and hybrid digital–analog (HDA) relaying. Figure 12.1 illustrates the protocols considered within each chapter. Via several examples, we observed that

- nonlinear relaying is superior to linear relaying;
- DF is an optimal strategy when the relay is located in the proximity of the source;
- CF and NNC outperform other protocols when the relay is located close to the destination;
- optimized low-dimensional relaying protocols perform encouragingly in comparison to those with infinite memory, and they are interestingly optimal for some cases; and
- HDA relaying outperforms purely-digital and purely-analog relaying schemes.

In the thesis, in order to optimize low-dimensional relaying, we proposed different algorithms as summarized in the following.
Figure 12.1: Relaying protocols investigated in the thesis.
12.1. Concluding Remarks

Iterative Bit-Switching

This method was adopted for binary symmetric relay channels studied in Chapter 3. Using this method the relay function is initialized with a random mapping. Then the mapping in a randomly chosen dimension is switched. The change is accepted if it increases the rate, otherwise it is rejected. This procedure is repeated until there is no significant improvement in the rate. This method by construction may terminate in a locally optimal solution. One remedy to this shortcoming is to initialize the whole procedure with a different random mapping and choose the best optimized mapping at the end.

Iterative Grid Search

This method was applied to Gaussian relay channels in Chapter 4. Using this method the range of the received signal at the relay is finely quantized and to each quantization interval an output assigned. Then the assigned output values are iteratively optimized until there is no significant improvement in the rate. This method is time-consuming but generally reaches a reasonably good stationary solution.

Fixed Point Iteration

We used this method in Chapters 5 and 10. In order to apply this method, we borrowed tools from the calculus of variations to derive a fixed point equation. We then initialized the equation and found an optimized mapping iteratively. The convergence of the fixed point iteration method is one of important characteristics that we only qualitatively discussed in the thesis. One future research direction would be to analytically investigate the convergence of this method. This algorithm is in general fast and provides good solutions when it converges.

Structured Design

With this method, we imposed a parameterized structure on the relay mapping and then optimized the design parameters. We generally desire a structure with a small number of parameters to facilitate the design. In the thesis, we proposed an interesting structured mapping in Chapter 6. We showed that the proposed structured mapping is in fact optimal in some cases. In order to design well-structured mappings, one should consider the properties of the problem in hand and incorporate the insights acquired from optimized unstructured mappings. In [KL08, KL07a], we proposed a memoryless relaying strategy using constellation rearrangement (CR) which is another example of a good structured solution. In [KS10a], it is demonstrated that relaying with CR performs very close to that with optimized unstructured mappings and it is also optimal in some cases.
200 Conclusions and Future Research

12.2 Some Possible Extensions

Simulated Annealing

One can implement the iterative bit-switching algorithm in Chapter 3 using simulated annealing [KGV83] in order to get possibly improved mappings. In simulated annealing, the changes for which the rate decreases are also accepted with a certain probability. This gives the possibility to escape the local optima. Simulated annealing is used for designing short length codes for point-to-point channels in [GHSW87], and for designing vector quantizers over noisy channels in [Far90].

Multi-Dimensional Mappings

Interesting extensions of memoryless relaying would include a generalization of the proposed nonlinear schemes in the thesis to multi-dimensional mappings, and also joint optimization of source distribution (the law for the generation of a random codebook) and the relay function. In particular, we propose the multi-dimensional mapping illustrated in Figure 12.2. This relaying strategy is the generalization of that in Figure 4.4.

Optimized Hybrid Mappings

We, in Chapters 7 and 11, discussed hybrid relaying in which the relay generates a digital compression index and linearly combines the encoded compression index with the noisy received signal. One extension to the current work is to find the optimal two-dimensional mapping for combining the digital and the analog signals.

Constant Gap to the Capacity

In the thesis, via serval examples, we demonstrated that optimized instantaneous relaying provides rates close to those achieved by CF and NNC. It is known that
NNC operates within a constant gap from the capacity of some Gaussian networks \[\text{[LKGC11]}\]. Therefore one interesting topic for future research would be to verify whether this is possible also with instantaneous relaying. If proved right, this would be an exciting result.

**Short-Message NNC**

We, in Chapter 8, proposed a transmission strategy based on the NNC scheme in \[\text{[LKGC11]}\] for the relay channel with noncausal knowledge of the state. An extension of the current work would be to develop a coding scheme based on the “short-message” NNC as discussed in \[\text{[KH11]}\]. The short-message NNC has a much smaller decoding delay at the destination.

### 12.3 Conjectures

We introduced the state-decoupled relay channel in Chapter 7 and found the capacity of a semi-deterministic class of these channels. However, we believe that one can extend some of the capacity results as discussed in the following conjectures.

#### 12.3.1 Capacity of Strictly Causal State-Decoupled Relay Channels

**Conjecture 12.3.1.** The capacity of the state-decoupled relay channel with strictly causal relaying is given by

\[
C = \max_{p(q)p(x|q)p(x_r|q)p(y_r|y,x_r,q)} \min \{R_1, R_2\},
\]

where

\[
R_1 = I(X, X_r; Y|Q) - I(Y_r; \hat{Y}_r|X, X_r, Y, Q),
\]

\[
R_2 = I(X; Y, \hat{Y}_r|X_r, Q),
\]

and \(|\hat{Y}_r| \leq |X_r| \cdot |Y_r| + 1, |Q| \leq 2.1\).

Conjecture 12.3.1 includes that of Han-Ahlswede in \[\text{[AH83, Section V]}\] as a special case. This follows by a similar discussion as that given in Remark 7.2.2.

Related to the above conjecture, Tandon and Ulukus in \[\text{[TU08]}\] have established a new upper bound on the capacity of the state-decoupled relay channels with a noiseless link from the relay to the destination, which is tighter than the cutset bound. We also remark that the channel studied by Aleksic, Razaghi, and Yu in \[\text{[ARY09]}\] is state-decoupled and its capacity is achieved by CF. For this channel the

---

1Here, \(Q\) denotes the time-sharing random variable. The time-sharing is used since the objective function is not convex in general.
upper bound in [TU08] is tight. This channel however does not fall in the semi-deterministic classes studied in [GA82, Kim08, CK07] and Theorem 7.2.1 and the capacity is yet achieved by CF.

12.3.2 Capacity of a Gaussian State-Decoupled Relay Channel

In Chapter 7, we showed that causal relaying improves on strictly causal relaying. In particular, we investigated the relay channel with additive Gaussian interference as illustrated in Figure 7.2. In Section 7.4, we proved that instantaneous linear relaying can outperform CF for the channel in Figure 7.2. However, we did not discuss the design of nonlinear strategies. We next briefly explain how one can implement a powerful nonlinear instantaneous strategy for the channel in Figure 7.2. The received signal at the destination is given by \( Y = X + S + X_r \), where \( S \sim N(0, Q) \) and \( X_r = f(S) \). Note that one can always choose \( f \) such that \( S + f(S) \in \mathcal{D}_s \), where \( \mathcal{D}_s \) is a finite set whose elements are chosen from the real line. The smaller the power at the relay, the denser the set \( \mathcal{D}_s \) should be chosen in order to meet the power constraint at the relay. Now let \( X \) be uniformly distributed over the interval \( (-\frac{d_{\min}}{2}, \frac{d_{\min}}{2}) \), where \( d_{\min} := \min_{i \neq j} |d_i - d_j| \) for all \( d_i, d_j \in \mathcal{D}_s \). Because the effective interference is discrete, the destination can exactly recover \( X \) from \( Y \) and hence arbitrarily high transmission rate is achievable. This scheme can be also interpreted as the relay transmits the quantization error such that \( \mathcal{D}_s \) is the set of the reconstruction points of the quantizer (see also Figure 4.4).² One interesting extension is to find optimized nonlinear strategies for the case with the receiver noise, as illustrated in Figure 12.3.

The channel shown in Figure 12.3 is intimately related to the point-to-point dirty tape channel illustrated in Figure 12.4. In the dirty tape channel, the received signal is given by \( Y = X + S + Z \), where \( X \) is the transmitted signal, \( S \) is the additive interference and \( Z \) is the additive noise at the receiver. The encoder is assumed to causally know \( S \) [ESZ05]. Using the terminology of Costa in [Cos83], the above suggested strategy for the noiseless case, in fact organizes the dirt such that the encoder can write on the remaining clean space. Additionally this strategy is similar to the ‘interference concentration’ scheme suggested by Willems in [Wil00, Wil88].

²Discussions with Dr. J. Kron are gratefully acknowledged.
12.3. Conjectures

The suggested strategy can also be considered as a sort of interference alignment invented by Maddah-Ali, Motahari, and Khandani [MAMK08]. In the language of interference alignment, the above scheme operates in a way that the effective interference is aligned on a countable subset of the real line and the remaining space is reserved for the transmission of the desired signal (cf., [MOGMAK09, MOGMAK11]). We conjecture that the optimized instantaneous relaying performs within a constant gap from the interference-free capacity bound when the relay has a non-zero power.

**Conjecture 12.3.2.** Consider the relay channel in Figure 12.3. Let $Z \sim \mathcal{N}(0, N)$ and $\mathbb{E}[X^2] + \mathbb{E}[X^2_r] \leq P$. If the interference has finite first and second moments, then the capacity with causal relaying is given by

$$C = C \left( \frac{P}{N} \right) - \kappa,$$

where $\kappa$ is a constant that does not depend on the interference.

We can show that the capacity $C$ in (12.4) is upper bounded by that of the dirty tape channel. We also remark that with strictly causal relaying, Conjecture 12.3.2 does not hold (see Proposition 7.4.1).

### 12.3.3 Capacity of Causal State-Decoupled Relay Channels

We end the thesis by the following conjecture.

**Conjecture 12.3.3.** The capacity of the state-decoupled relay channel with causal relaying is given by

$$C = \max_{p(q)p(x|q)p(v|q)p(\hat{y}_r|y_r,v,q), x_r = f(y_r,v,q)} \min \{R_1, R_2\},$$

where

$$R_1 = I(X,V;Y|Q) - I(Y_r;\hat{Y}_r|X,V,Y,Q),$$

$$R_2 = I(X;Y,\hat{Y}_r|V,Q).$$


