AN INFORMATION THEORETIC APPROACH TO NON-CENTRALIZED MULTI-USER COMMUNICATION SYSTEMS

CHONG HON FAH

NATIONAL UNIVERSITY OF SINGAPORE

2008

AN INFORMATION THEORETIC APPROACH TO NON-CENTRALIZED MULTI-USER COMMUNICATION SYSTEMS

CHONG HON FAH

(*B. Eng., (Hons.), M. Eng, NUS*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2008

Summary

In this thesis, we take a fundamental information-theoretic look at three non-centralized multi-user communication systems, namely, the relay channel, the interference channel (IFC), and the "Z"-channel (ZC). Such multi-user configurations may occur for example in wireless ad-hoc networks such as a wireless sensor network.

For the general relay channel, the best known lower bound is a generalized strategy of Cover & El Gamal, where the relay superimposes both cooperation and facilitation. We introduce and study three new generalized strategies: The first strategy makes use of sequential backward (SeqBack) decoding, the second strategy makes use of simultaneous backward (SimBack) decoding, and the third strategy makes use of sliding window decoding. We also establish the equivalence of the rates achievable by both SeqBack and SimBack decoding. For the Gaussian relay channel, assuming zero-mean, jointly Gaussian random variables, all three strategies give higher achievable rates than Cover & El Gamal's generalized strategy. Finally, we extend the rate achievable for SeqBack decoding to the relay channel with standard alphabets.

For the general IFC, a simplified description of the Han-Kobayashi rate region, the best known rate region to date for the IFC, is established. Using this result, we prove the equivalence between the Han-Kobayashi rate region and the recently discovered Chong-Motani-Garg rate region. Moreover, a tighter bound for the cardinality of the time-sharing auxiliary random variable emerges from

our simplified description. We then make use of our simplified description to establish the capacity region of a class of discrete memoryless IFCs. Finally, we extend the result to prove the capacity region of the same class of IFCs, where both transmitters now have a common message to transmit.

For the two-user ZC, we study both the discrete memoryless ZC and the Gaussian ZC. We first establish achievable rate regions for the general discrete memoryless ZC. We then specialize the rate regions obtained to two different types of degraded discrete memoryless ZCs and also derive respective outer bounds to their capacity regions. We show that as long as a certain condition is satisfied, the achievable rate region is the capacity region for one type of degraded discrete memoryless ZC. The results are then extended to the two-user Gaussian ZC with different crossover link gains. We determine an outer bound to the capacity region of the Gaussian ZC with strong crossover link gain and establish the capacity region for moderately strong crossover link gain.

Acknowledgments

I would like to express my heart-felt thanks to both of my supervisors, Prof. Hari Krishna Garg and Dr Mehul Motani, for their invaluable guidance, continuing support and constructive suggestions throughout my research in NUS. Their deep insight and wide knowledge have helped me out at the various phase of my research. It has been an enjoyable and cultivating experience working with them.

Next, I would like to thank my colleagues at ECE-I2R lab for all their help and for making my research life so wonderful.

Last but not least, I would like to thank my family members who have always been the best supporters of my life.

Contents

Nomenclature

Roman Symbols

$\Pr(E)$ Probability of an event E taking place.

$A_{\epsilon}^{(N)}(X_1, X_2, ...X_k)$ The set of ϵ-typical N-sequences $\left(x_1^N, x_2^N, ..., x_k^N\right)$.

f_X The density (Radom-Nikodym derivative) of the random variable X.

$H(X)$ The entropy of a discrete random variable X.

$h(X)$ The differential entropy of a continuous random variable X.

$I(X;Y)$ Mutual information between random variables X and Y.

p A probability distribution function.

$P_e^{(N)}$ Average probability of error for a block of size N.

P_i Power Constraint of node-i.

R Achievable rate.

Script Symbols

$(\mathcal{X}, \mathcal{F}_X)$ A measurable space consisting of a sample space $\mathcal{X}$ together with a σ-field $\mathcal{F}_X$ of subsets of $\mathcal{X}$.

$\mathbb{E}(X)$ Expectation of the random variable X.

$\mathcal{N}$ Normal distribution.

$\mathcal{P}$ A set of probability distributions.

$\mathcal{R}$ A rate region.

$\mathcal{X}$ A set.

Greek Symbols

Φ_n Relay encoding function.

Φ_{N+1} Relay decoding function.

Ψ An encoding function.

σ^2 Variance of a random variable.

Mathematical Symbols

$\mathcal{A} \subseteq \mathcal{B}$ $\mathcal{A}$ is a subset of $\mathcal{B}$.

$\mathcal{A} \subsetneq \mathcal{B}$ $\mathcal{A}$ is a proper subset of $\mathcal{B}$.

$M_X >> P_X$ The probability distribution M_X dominates the probability distribution P_X.

Abbreviations

RX_i Receiver i.

TX_i Transmitter i.

IFC The Interference Channel.

ZC The "Z"-channel.

ZIFC The "Z"-Interference Channel.

List of Figures

List of Tables

Chapter 1

Introduction

In 1948, Claude E. Shannon developed the mathematical theory of communication with the publication of his landmark paper "A mathematical theory of communication" [1]. In this paper, Shannon showed that reliable communication between a transmitter and a receiver is possible if and only if the rate of transmission is below the channel capacity. He gave a single letter characterization of the channel capacity, which is a function of the channel statistics. Shannon's work provided a crucial "knowledge base" for the discipline of communication engineering. The communication model is general enough so that the fundamental limits and general intuition provided by Shannon theory provide an extremely useful "road map" to designers of communication and information storage systems.

In his original paper, Shannon focused solely on communication between a single transmitter and receiver. However, almost all modern communication systems involve multiple transmitters and receivers attempting to communicate on the same channel. Shannon himself studied the two-way channel [2], and derived simple upper and lower bounds for the capacity region.

Besides the two-way channel, Shannon's information theory has been applied to other multi-user communication networks. Fig. 1.1 shows a multiple access channel where there are m transmitters simultaneously transmitting to a common receiver. This is in fact one of the best understood multi-user communication

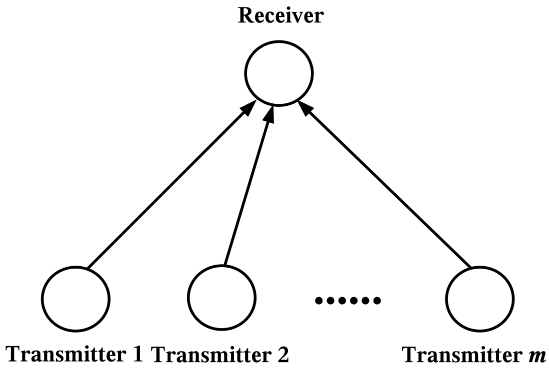

Figure 1.1: Multiple access channel

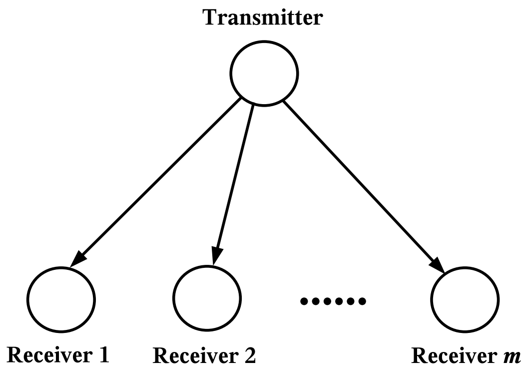

Figure 1.2: Broadcast Channel

network. The channel capacity for the multiple access channel was completely characterized by Ahlswede [3] and Liao [4].

On the other, we obtain the broadcast channel when the multiple access channel network is reversed. In the broadcast channel, one transmitter broadcasts information (common/independent) simultaneously to m receivers as shown in Fig. 1.2. Broadcast channels were first studied by Cover in 1972 [5]. The capacity for the degraded broadcast channels were determined by Gallager [6] for the discrete memoryless broadcast channel and Bergmans [7] for the Gaussian broadcast channel. The best known achievable rate region to date for the general broadcast channel is due to Marton [8]. Recently, the capacity region of the Gaussian MIMO broadcast channel, which is not a degraded broadcast channel, has been established [9], [10], [11].

1.1 Motivation

In the past, the study of multi-user information theory has largely been motivated by wireline and cellular systems. Hence, much emphasis has been placed upon multi-user channel configurations with a central node, such as the multiple access channel (cellular uplink, where the receiving base station is the central node) and the broadcast channel (cellular downlink, where the transmitting base station is the central node).

However, with recent advances and interests in wireless ad-hoc networks, there has been a growing interest in the study of other multi-user channels. A wireless ad-hoc network is a collection of two or more devices equipped with transmitting capabilities or receiving capabilities or a combination of both. Such devices can transmit to another device with the help of an available intermediate node. Recently, there has also been much focus on wireless sensor networks, which is a form of a wireless ad-hoc network. In a wireless sensor network, the sensors might be autonomously collecting information at different locations and attempting to communicate the information to one or more data-collection centers or sinks. The potential of wireless sensor networks cannot be overemphasized.

"*In the health care industry, sensors allow continuous monitoring of life-critical information. In the food industry, biosensor technology applied to quality control can help prevent rejected products from being shipped out, thus enhancing consumer satisfaction levels. In agriculture, sensors can help to determine the quality of soil and moisture level; they can also detect other bio-related compounds. Sensors are also widely used for environmental and weather information gathering. They enable us to make preparations in times of bad weather and natural disaster.*"—C. K. Toh [12, pp. 30]

Certain questions naturally arise when one attempts to study wireless ad-hoc networks. How should the nodes communicate with each other? What is the

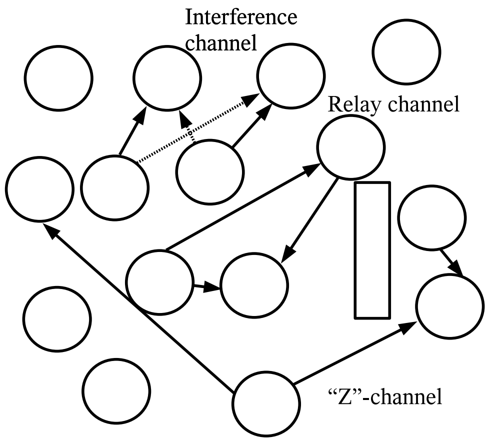

Figure 1.3: Simple multi-user configurations that may occur in a wireless ad-hoc network

rate achievable for such a network? The ideal would be to arrive at a general multi-terminal network information theory. However, attempting to solve the most general case for a wireless ad-hoc or sensor network even for a few number of nodes may be prohibitively difficult.

At the other extreme, in most strategies commonly implemented, a node simply attempts to communicate with a node within its radio range, and if it is out of its radio range, it attempts to relay the data via an intermediate node. If a nearby node is transmitting in the same bandwidth, the current node simply withholds itself from transmitting. Such a strategy however does not fully exploit cooperation and competition amongst the nodes close by.

Rather than attempting to arrive at a general multi-user information theory or study simple forwarding strategies, we take an intermediate stand. Our focus in this thesis is to study in-depth three non-centralized multi-user channel communication systems, namely, the relay channel, the interference channel (IFC), and the "Z"-channel (ZC) that often arise in a wireless ad-hoc network as shown in Fig. 1.3.

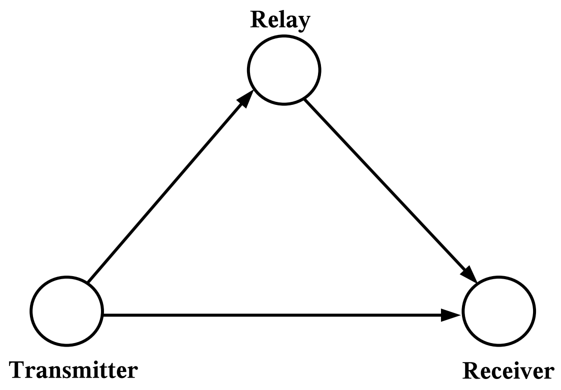

Figure 1.4: Relay Channel

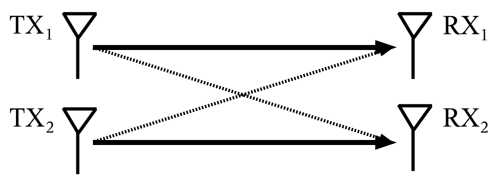

Figure 1.5: Interference Channel

1.1.1 Relay Channel

The relay channel is a channel in which there is one sender and one receiver with a number of intermediate nodes which acts as a relay network to help the communication from the sender to the receiver. The simplest relay channel has only one intermediate or relay node as shown in Fig. 1.4. Relay channels model situations where one or more relays help a pair of terminals communicate. This often occurs in a multi-hop wireless network, where nodes have limited power to transmit data. In fact, a node can help as a relay even when the receiving node is within the radio range of the transmitting node. This might also occur in a broadcast channel where the users are allowed to cooperate. Each of the users can then serve as a relay for the other user.

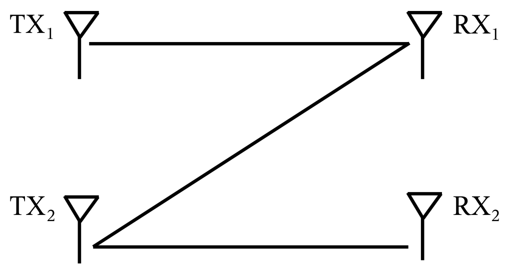

Figure 1.6: The configuration of the ZC

1.1.2 Interference channel

The simplest IFC consist of two transmitters and two receivers where there is no cooperation between the two transmitters or the two receivers as shown in Fig. 1.5. Each user is attempting to transmit information to its own intended receiver but interferes with the other non-intended receiver. This might occur when two nodes are attempting to communicate information to two different sinks in a wireless sensor network or in two overlapping wireless LAN where two users are attempting to communicate to their respective base stations. For the IFC with common information, both the senders transmit not only their own private information but also a common information to their corresponding receivers.

1.1.3 "Z"-channel

Recently, Vishwanath, Jindal, and Goldsmith [13] introduced the ZC shown in Fig. 1.6. The ZC consists of two senders and two receivers. The transmission of sender TX_1 can reach only receiver RX_1, while that of sender TX_2 can reach both receivers.

The Z-interference channel (ZIFC) has the same topology as the ZC shown in Fig. 5.1. In both the ZC and ZIFC, there is no cooperation between the two senders or between the two receivers. However, in the ZIFC, sender TX_2 has no information to transmit to receiver RX_1, while the ZC allows transmission of information from sender TX_2 to receiver RX_1. Hence, the ZC models a more

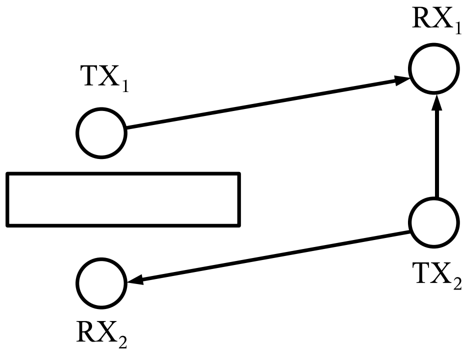

Figure 1.7: A ZC: transmission of sender TX_1 is unable to reach receiver RX_2 due to an obstacle

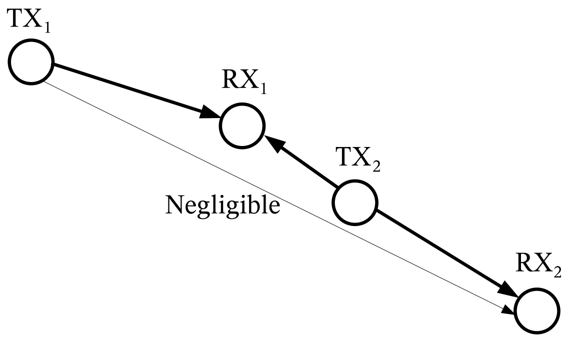

Figure 1.8: A ZC: transmission of sender TX_1 is unable to reach receiver RX_2 due to distance

general multi-user network compared to the ZIFC. The capacity region of the ZC includes the capacity region of the broadcast channel (sender TX_2 is transmitting information to both receivers), the capacity region of the multiple access channel (sender TX_1 and TX_2 are both transmitting information to receiver RX_1), and the capacity region of the ZIFC (both senders are transmitting information to their own intended receivers).

Such a multi-user configuration may correspond to a local scenario (with two users and two receivers) in a large sensor or wireless ad-hoc network. As shown in Fig. 1.7, sender TX_1 is unable to transmit to receiver RX_2 due to an obstacle, while sender TX_2 is able to transmit to both receivers. Another possible scenario is shown in Fig. 1.8, where sender TX_1 is so far away from receiver RX_2 that its transmission is negligible.

1.2 Thesis Outline and Contributions

In this thesis, we take an information-theoretic look at three non-centralized multi-user channels, the relay channel, the IFC, and the ZC, building from the information theoretic work of Claude Shannon and others.

The thesis is organized as follows.

- In Chapter 2, we take a look at the three-node relay channel. We come up with new coding strategies for the discrete memoryless relay channel and then apply the results to the Gaussian relay channel. We also compare the performance of these strategies with respect to the best known lower bound for the general relay channel.
  - The main contributions of the chapter are Thm. 2.1, Thm. 2.2, and Thm. 2.3.
  - Thm. 2.1 establishes a potentially better lower bound for the achievable rate of the relay channel. This rate can be achieved by either a sequential backward (*SeqBack*) decoding strategy or a sliding window decoding strategy.
  - Thm. 2.2 establishes a new lower bound for the achievable rate of the relay channel using a simultaneous backward (*SimBack*) decoding strategy. All three strategies combine the decode-and-forward strategy [14, Thm. 1] and the compress-and-forward strategy [14, Thm. 6].
  - Thm. 2.3 establishes the equivalence of the rates achieved by Thm. 2.1 and Thm. 2.2.
  - Finally, we show that the rate achievable by SeqBack decoding, SimBack decoding or the sliding window decoding strategy includes the best known lower bound of Cover & El Gamal [14, Thm. 7]. When

applied to the Gaussian relay channel, assuming zero-mean Gaussian random variables, the new rate is shown to be strictly greater than the generalized strategy of Cover & El Gamal.

- Strictly, speaking Thm. 2.1 and Thm. 2.2 hold only for discrete random variables. In Chapter 3, we extend Thm. 2.2 to relay channels with more general alphabets.

  - The main contribution of the chapter is Thm. 3.2 which extends Thm. 2.1 to relay channels with more general alphabets, i.e., to the class of probability distributions with well-defined probability densities.
  - Thm. 3.2 allows us to obtain achievable rates for the Gaussian relay channel with well-defined continuous input probability density functions. We may also obtain achievable rates for mixed input distributions by setting the dominating measure to be the Lebesgue measure plus the counting measure.

- In Chapter 4, we take a look at the IFC. We establish a simplified description of the best known achievable rate region to date for the IFC. We then make use of our simplified description to establish the capacity of a new class of IFCs. We also extend this result to the case of the IFC with common information.

  - The main contributions of the chapter are Thm. 4.2, Thm. 4.7, and Thm. 4.8.
  - Thm. 4.2 gives a simplified description of the Han-Kobayashi rate region [15, Thm. 3.1] for the IFC. Using this result, we establish the equivalence between the Han-Kobayashi rate region and the recently discovered Chong-Motani-Garg representation [16, Thm. 3] of the Han-Kobayashi rate region. Moreover, a tighter bound for the cardinality of the time-sharing auxiliary random variable emerges from

our simplified description.

- Thm. 4.7 establishes the capacity region of a new class of IFCs and Thm. 4.8 extends the result to the IFC with common information. The setup is similar to the class of deterministic IFCs studied by El Gamal & Costa [17], which was later extended to the class of deterministic IFCs with common information by Jiang, Xin and Garg [18]. We relax certain deterministic constraints (see (4.101) and (4.102)) that were originally imposed by El Gamal & Costa. We show by a specific example that this class of IFC is strictly larger than the class of deterministic IFCs of El Gamal & Costa.

- In Chapter 5, we take a look at the ZC. We first establish achievable rates for the general discrete memoryless ZC. We then specialize the rates obtained to two different types of degraded, discrete memoryless ZCs (DMZC) and also derive respective outer bounds to their capacity regions. We show that as long as a certain condition (see Thm. 5.9) is satisfied, the achievable rate region is the capacity region for one type of degraded discrete memoryless ZC. The results are then extended to the two-user Gaussian ZC with different crossover link gains (see Section 5.4). We determine an outer bound to the capacity region of the Gaussian ZC with strong crossover link gain and establish the capacity region for moderately strong crossover link gain.

  - The main contributions of the chapter are Thm. 5.3, Thm. 5.9 and Thm. 5.13.
  - Thm. 5.3 establishes an achievable rate for the general ZC making use of rate-splitting and joint decoding.
  - Next, we specialize the result for the general setting to one type of degraded DMZC. We also determine an outer bound to the capacity region. The result is extended directly to the two-user Gaussian ZC with weak crossover link gain.

  - We then specialize the result for the general setting to another type of degraded DMZC. The result is extended directly to the Gaussian ZC with strong crossover link gain. We also determine respective outer bounds to their capacity regions. We establish the capacity region of the Gaussian ZC with moderately strong crossover link gain in Thm. 5.13. For the discrete case, we show in Thm. 5.9 that the achievable rate region is the capacity region if a certain condition is satisfied.

- In Chapter 6, we conclude the thesis with some directions for future work.

Thm. 2.1 and Thm. 2.2 are based on [19], while Thm. 2.3 is based on [20] presented at the Information Theory and Applications Workshop, 2007. Chapter 2 is based on [21]. Thm. 4.2 is based on [22] while Thm. 4.7 and Thm. 4.8 are based on [23] presented at the International Symposium of Information Theory, 2007. Finally, Chapter 5 is based on [24].

1.3 Notations and preliminaries

We denote a random variable with capital letter X and its realization with lower case letter x. The associated measurable space $(\mathcal{X}, \mathcal{F}_X)$ is a pair consisting of a sample space $\mathcal{X}$ together with a σ-field $\mathcal{F}_X$ of subsets of $\mathcal{X}$. We denote vectors with a superscript, e.g., X^N denotes a random vector of length N and x^N denotes a realization of the random vector. The associated measurable space is given by $(\mathcal{X}_1 \times \mathcal{X}_2 ... \times \mathcal{X}_N, \mathcal{F}_{X_1} \times \mathcal{F}_{X_2} ... \times \mathcal{F}_{X_N})$ or its short form $\left(\mathcal{X}_1^N, \mathcal{F}_{X_1^N}\right)$.

The usual notation for entropy and mutual information is used. $H\left(X\right)$ denotes the entropy of a discrete random variable and $h\left(X\right)$ denotes the differential entropy of a continuous random variable. $H\left(X|Y\right)$ is the conditional entropy of the random variable X given Y and $h\left(X|Y\right)$ is the conditional differential entropy of the random variable X given Y. $I\left(X;Y\right)$ is the mutual information between the random variable X and Y and $I\left(X;Y|Z\right)$ is the mutual information between the random variable X and Y conditioned on the random variable Z.

Except for Chapter 3, we denote the set of both ϵ-weakly typical and ϵ-strongly typical sequences w.r.t. the discrete probability distribution $p_X(x)$ by $A_\epsilon^{(N)}(X)$. In Chapter 3, we denote the set of ϵ-typical sequences w.r.t. the probability density $f_X(x)$ by $A_\epsilon^{(N)}(X)$ and the set of ϵ-strongly typical sequences w.r.t. the discrete probability distribution $p_X(x)$ by $A_\epsilon^{*(N)}(X)$. When the context is clear, we will ignore the subscript X for $f_X(x)$ and $p_X(x)$.

Most of the fundamental theory about entropy and mutual information used throughout the thesis can be found in [25]. In Chapter 3, we extend the results of Chapter 2 to relay channels with standard alphabets. We define relative entropy, conditional relative entropy, mutual information and conditional mutual information for random variables, with well defined probability densities, taking values in standard spaces. Most of this theory can be found in [26].

In Chapter 4, we make heavy use of the Fourier-Motzkin elimination method for eliminating variables and removing redundant inequalities. More information can be found at [27].

Chapter 2

On the Relay Channel

2.1 Introduction

The three-node relay channel was introduced by Van der Meulen [28], [29]. In [28], a time sharing strategy was used to establish a lower bound for the capacity of the relay channel. Outer bounds for the capacity of the relay channel were found in [28], [30]. Two important coding theorems for the single relay channel were established in a fundamental paper by Cover & El Gamal [14].

In the *cooperation* strategy via decode-and-forward [14, Thm. 1], the relay decodes the source message and forwards it to the destination. Cover & El Gamal made use of block Markov superposition encoding, random binning, and successive list decoding to achieve the rate for the decode-and-forward strategy. Two other techniques that have been proposed are commonly known as regular encoding/sliding window decoding ([31], [32]) and regular encoding/backward decoding ([33], [34]). These are summarized in [35]. The decode-and-forward strategy was shown in [14] to achieve the capacity of the degraded relay channel, the reversely degraded relay channel, and the relay channel with causal noiseless receiver-relay feedback. However, this strategy does not achieve the capacity of the general discrete memoryless relay channel or the Gaussian relay channel.

In the *facilitation* strategy via compress-and-forward, Cover & El Gamal made use of random binning and Wyner-Ziv source coding [36] to exploit side information at the destination. In this strategy, the relay transmits a compressed version of its channel outputs to the destination. This was also recently shown to be capacity achieving for a class of deterministic relay channels [37].

The decode-and-forward strategy and the compress-and-forward strategy were combined to give a generalized strategy for the relay channel in [14, Thm. 7]. The generalized strategy combines ideas such as block Markov superposition encoding, random binning, successive list decoding coupled with Wyner-Ziv source coding to exploit the side information at the destination. The purpose of this chapter is to investigate other generalizations of the two basic coding strategies for the three-node relay channel. We discuss and derive achievable rates for three alternative strategies that superimpose cooperation and facilitation. These alternate strategies are modifications of the decoder and hence, changes the error analysis at the decoder.

The first strategy performs sequential backward (*SeqBack*) decoding at the receiver. Backward decoding was introduced by Willems [33] for the multiple-access channel with feedback. Zeng, Kuhlmann, and Buzo [34] showed that many of the proofs for multi-user channel coding theorems could be simplified using backward decoding.

In [15], it was shown that simultaneous decoding results in superior performance compared to sequential decoding for the interference channel (IFC). Hence, our second strategy, (*SimBack*) decoding, investigates the performance of backward decoding coupled with simultaneous decoding. Our last strategy is a sliding window decoding strategy that achieves the same rate as SeqBack decoding. In fact, sliding window decoding rather than backward decoding is the preferred method used for multi-hopping [35], as the delay introduced by backward decoding strategies makes it impractical for implementation.

We then compute the achievable rates for these strategies in a Gaussian relay channel. As it may be formidable to compute the maximum achievable rate over all input distributions, we impose the customary restriction to the class of jointly Gaussian input distributions. It is shown that for certain parameters of the Gaussian relay channel, our generalized strategies outperform the generalized strategy of Cover & El Gamal.

The achievable rates for the different generalized strategies are expressed in different forms making it hard for comparison. Finally, we compare the various generalized strategies by casting the achievable rates into appropriate forms. We show that in fact all our strategies achieve the same rate; we also conjecture that in general our strategies outperform that of the generalized strategy of Cover & El Gamal as suggested by our numerical computation for the Gaussian relay channel.

2.1.1 Outline

This chapter is organized as follows:

- In Section 2.2, we define the mathematical model for the discrete memoryless relay channel and the Gaussian relay channel.
- In Section 2.3, we review some results for the general relay channel. We also derive the achievable rates for three new generalized strategies and apply the results to the Gaussian relay channel.
- In Section 2.4, we compute and compare the achievable rates for certain parameters of the Gaussian relay channel.
- In Section 2.5, we compare the performance of the various generalized strategies.

2.2 Mathematical Model

We closely follow the formulation and notation of [14]. A discrete memoryless relay channel consists of four finite sets $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}_2$, $\mathcal{Y}_3$, and a collection of probability distributions $p\left(.,.|x_1,x_2\right)$ on $\mathcal{Y}_2$, $\mathcal{Y}_3$. The quantity x_1 is the source input, x_2 is the relay input, y_2 is the relay output, and y_3 is the destination output.

An $\left(2^{NR},N\right)$ code for the relay channel is composed of a set of integers $\mathcal{M}=\left\{1,2,...,2^{NR}\right\}$, an encoding function

$$\Psi:\left\{1,2,...,2^{NR}\right\}\rightarrow\mathcal{X}_1^N$$

a set of relay functions $\{\Phi_n\}_{n=1}^{n=N}$ such that

$$\Phi_n:\mathcal{Y}_2^n\rightarrow\mathcal{X}_2,\ \ 1\leq n\leq N,$$

and a decoding function

$$\Phi_{N+1}:\mathcal{Y}_3^N\rightarrow\left\{1,2,...,2^{NR}\right\}.$$

The relay is causal in nature. Hence, the relay transmission is allowed to depend only on the past observations $y_{21},y_{22},...,y_{2n-1}$. On the other hand, for an acausal relay, the relay transmission is allowed to depend also on the current observation y_{2n}. The channel is also assumed to be memoryless in the sense that the channel outputs (y_{2n},y_{3n}) depends on the past only through the current transmitted symbols (x_{1n},x_{2n}). Hence, for any choice $p\left(m\right)$, $m\in\mathcal{M}$, any code choice $\Psi:\left\{1,2,...,2^{NR}\right\}\rightarrow\mathcal{X}_1^N$, and relay functions $\{\Phi_n\}_{n=1}^{N}$, the joint probability distribution function on $\mathcal{M}\times\mathcal{X}_1^N\times\mathcal{X}_2^N\times\mathcal{Y}_2^N\times\mathcal{Y}_3^N$ is given by

$$p\left(m,x_1^N,x_2^N,y_2^N,y_3^N\right)=p\left(m\right)\prod_{n=1}^{N}p\left(x_{1n}|m\right)p\left(x_{2n}|y_2^{n-1}\right)\cdot p\left(y_{2n},y_{3n}|x_{1n},x_{2n}\right).$$

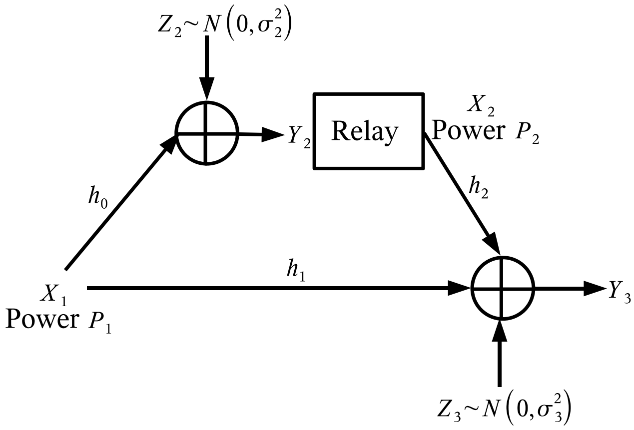

Figure 2.1: Gaussian Relay Channel

If the message $m \in \mathcal{M}$ is sent, let

$$\lambda\left(m\right) = \Pr\left\{\Phi_{N+1}\left(Y_3^N\right) \neq m | m \text{ sent}\right\}$$

be the conditional probability of error. The average probability of error is defined by

$$P_e^{(N)} = \frac{1}{2^{NR}} \sum_m \lambda\left(m\right).$$

The probability of error is calculated under the uniform distribution over the codewords $m \in \mathcal{M}$. The rate R is said to be achievable by the relay channel if there exists a sequence of $\left(2^{NR}, N\right)$ codes with $P_e^{(N)} \to 0$ as $N \to \infty$. The capacity C_{R} is the supremum of the set of achievable rates.

2.2.1 Model for the Gaussian Relay Channel

Consider the Gaussian relay channel of Fig. 2.1, in which the source node intends to transmit information to the destination node by using the direct link between source and destination as well as with the help of another relay node.

The dependency of the outputs on the inputs are as follows. The relay output is given by

$$Y_2 = h_0 X_1 + Z_2 \tag{2.1}$$

and the destination output is given by

$$Y_3 = h_1 X_1 + h_2 X_2 + Z_3. \tag{2.2}$$

The constants h_0, h_1, and h_2 are channel losses and are assumed to be constant. $Z_2 \sim \mathcal{N}(0, \sigma_2^2)$ and $Z_3 \sim \mathcal{N}(0, \sigma_3^2)$ are independent Gaussian noises. The input power constraints are given by $\mathbb{E}[X_1^2] \leq P_1$ and $\mathbb{E}[X_2^2] \leq P_2$.

Remark 2.1. Throughout this chapter, we make use of strong typicality in order to invoke Berger's Markov lemma [25, Lem. 14.8.1]. Even though strong typicality does not apply to continuous random variables, we will still make use of the coding theorems to compute achievable rates for the Gaussian relay channel. This is because the coding theorems can also be proven using weak typicality by making modifications along the lines of Oohama [38]. We will leave the derivation of the coding theorems using weak typicality for the next chapter. Our focus, in this chapter, is to look at new generalized strategies for the relay channel.

2.3 Coding Strategies for the Relay Channel

In this section, we review the cut-set upper bound on the capacity of the relay channel. We also review some achievable coding strategies of [14] and then derive two new generalized backward decoding strategies, and a generalized sliding window decoding strategy. For all the strategies, we compute rates for the Gaussian relay channel shown in Fig. 2.1.

2.3.1 Capacity Upper Bound

The capacity of the relay channel satisfies

$$R_{\mathrm{U}} \leq \sup_{p(x_1, x_2)} \min \left\{ I(X_1 X_2; Y_3), I(X_1; Y_2 Y_3 | X_2) \right\}. \tag{2.3}$$

This capacity upper bound follows directly from the cut-set upper bound [25, Thm. 14.10.1] and can be achieved under certain conditions. The source and the relay could transmit to the destination with rate $I(X_1X_2;Y_3)$ if the relay had complete knowledge of the source message. The rate $I(X_1;Y_2Y_3|X_2)$ could be achieved if the destination had knowledge of X_2 and Y_2.

A conditional maximum entropy theorem of [39] ensures that the capacity upper bound for the Gaussian relay channel can be maximized by making $p(x_1,x_2)$ zero-mean Gaussian. Hence, for the Gaussian relay channel, let $X_1 = aX_2 + W$, where a is a constant. In Appendix A.1, we compute the cut-set bound to be

$$R_{\mathrm{U}} \leq \sup_{0\leq\alpha\leq 1} \min \left\{ \begin{array}{c} \frac{1}{2}\log_2\left(1+\frac{h_1^2P_1+h_2^2P_2+2h_1h_2\sqrt{(1-\alpha)P_1P_2}}{\sigma_3^2}\right), \\ \frac{1}{2}\log_2\left(1+\alpha P_1\left(\frac{h_0^2}{\sigma_2^2}+\frac{h_1^2}{\sigma_3^2}\right)\right) \end{array} \right\}. \tag{2.4}$$

2.3.2 Cooperation via Decode-And-Forward

For the first strategy of Cover & El Gamal [14, Thm. 1], the relay decodes all the information transmitted to the receiver. Hence, the authors in [35] interpret this as a decode-and-forward strategy. This strategy can achieve any rate up to

$$R_1 = \sup_{p(x_1,x_2)} \left\{\min\left\{I(X_1X_2;Y_3), I(X_1;Y_2|X_2)\right\}\right\}. \tag{2.5}$$

In the literature, several different strategies have been suggested to achieve rate R_1. In [14], Cover & El Gamal use irregular block Markov superposition encoding and successive decoding. In [33], Willems suggests regular block Markov superposition encoding and backward decoding. In [32], Carleial uses regular block Markov superposition encoding and sliding window decoding. The advantage of the third strategy by Carleial is that both the source and the relay employ an equal number of codewords. Moreover, a delay of only one block length is necessary for the receiver to perform decoding.

For the Gaussian relay channel, the conditional maximum entropy theorem of [39] again ensures that R_1 is maximized by choosing X_1 and X_2 to be zero-mean Gaussian. Similar to the computation of the cut-set upper bound, we let $X_1 = aX_2 + W$, where a is a constant. Rate R_1 is then given by

$$R_1 = \sup_{0\leq\alpha\leq 1} \min \left\{ \begin{array}{c} \frac{1}{2}\log_2\left(1 + \frac{h_1^2 P_1 + h_2^2 P_2 + 2h_1 h_2 \sqrt{(1-\alpha)P_1 P_2}}{\sigma_3^2}\right), \\ \frac{1}{2}\log_2\left(1 + \frac{\alpha h_0^2 P_1}{\sigma_2^2}\right) \end{array} \right\}. \tag{2.6}$$

2.3.3 Facilitation via Compress-and-Forward

For the strategy of [14, Thm. 6], the relay forwards a compressed version of Y_2 to the destination. For any relay channel, the following rate is achievable:

$$R_2 = \sup_{p(x_1)p(x_2)} I\left(X_1; \hat{Y}_2 Y_3 | X_2\right) \tag{2.7}$$

where the supremum is taken over all joint probability density functions of the form

$$p\left(x_1, x_2, \hat{y}_2, y_2, y_3\right) = p\left(x_1\right) p\left(x_2\right) p\left(y_2, y_3 | x_1, x_2\right) p\left(\hat{y}_2 | y_2, x_2\right) \tag{2.8}$$

subject to the constraint

$$I\left(X_2; Y_3\right) \geq I\left(Y_2; \hat{Y}_2 | X_2 Y_3\right). \tag{2.9}$$

Remark 2.2. The optimal distribution on $\left(X_1, X_2, \hat{Y}_2\right)$ is currently still unknown. However, restricting $\left(X_1, X_2, \hat{Y}_2\right)$ to the Gaussian distribution allows one to compute an achievable rate for the Gaussian relay channel using compress-and-forward strategy. Throughout the rest of the chapter, we restrict our attention to the class of Gaussian input distributions, which may not necessarily be optimal.

To compute an achievable rate for the Gaussian relay channel, let $\hat{Y}_2 = Y_2 + Z_W$, where $Z_W \sim \mathcal{N}\left(0, \sigma_W^2\right)$. We also assume that X_1 and X_2 are independent, zero-mean Gaussian random variables. (see [35, (55) and (56)] for the same analysis). We compute the rate in Appendix D.4 and obtain

$$R_2 = \frac{1}{2}\log_2\left(1 + P_1\left(\frac{h_1^2}{\sigma_3^2} + \frac{h_0^2}{\sigma_2^2 + \sigma_W^2}\right)\right) \tag{2.10}$$

subject to the constraint

$$\sigma_W^2 \geq \frac{h_1^2 P_1 \sigma_2^2 + h_0^2 P_1 \sigma_3^2 + \sigma_2^2 \sigma_3^2}{h_2^2 P_2}. \tag{2.11}$$

2.3.4 Generalized Lower Bound of Cover & El Gamal

The strategy of [14, Thm. 7] is a combination of the decode-and-forward strategy with the compress-and-forward strategy. In (2.12) below, we have also included a discrete time-sharing random variable Q as El Gamal, Mohseni, and Zahedi [40] showed that the compress-and-forward strategy can be improved upon with time-sharing. By including a time-sharing parameter Q, the generalized strategy of Cover & El Gamal achieves any rate up to

$$R_3 = \sup\left\{\min\left\{\begin{array}{l} I\left(X_1; \hat{Y}_2 Y_3 | U X_2 Q\right) + I\left(U; Y_2 | V X_2 Q\right), \\ I\left(X_1 X_2; Y_3 | Q\right) - I\left(Y_2; \hat{Y}_2 | U X_1 X_2 Y_3 Q\right) \end{array}\right\}\right\} \tag{2.12}$$

where the supremum is taken over all joint probability density functions of the form

$$\begin{aligned} p\left(q, u, v, x_1, x_2, y_2, \hat{y}_2, y_3\right) &= p\left(q\right) p\left(v|q\right) p\left(u|v, q\right) \\ &\cdot p\left(x_1|u, q\right) p\left(x_2|v, q\right) p\left(y_2, y_3|x_1, x_2\right) p\left(\hat{y}_2|x_2, y_2, u, q\right) \end{aligned} \tag{2.13}$$

subject to the constraint

$$I\left(X_2; Y_3 | V Q\right) \geq I\left(\hat{Y}_2; Y_2 | U X_2 Y_3 Q\right). \tag{2.14}$$

Remark 2.3. V represents the information that the relay has decoded in the previous block while U represents the information that the relay can decode from the current block. With $Q \triangleq \emptyset$, the strategy is simply a combination of the decode-and-forward strategy with the compress-and-forward strategy. For example, cooperation via decode-and-forward strategy is attained by setting $Q \triangleq \emptyset$, $V \triangleq X_2$, $U \triangleq X_1$, and $\hat{Y}_2 \triangleq \emptyset$ and facilitation via compress-and-forward strategy is attained by setting $Q \triangleq \emptyset$, $V \triangleq \emptyset$, and $U \triangleq \emptyset$. The parameter Q allows the time-sharing of different combined strategies.

We set $Q \triangleq \emptyset$ for ease of computation of an achievable rate region for the Gaussian relay channel. We also assume $\left(U, V, X_1, X_2, \hat{Y}_2\right)$ to be jointly Gaussian, zero-mean random variables. Let U, X_1, and X_2 be zero-mean Gaussian random variables of the following form:

$$\begin{aligned} U &= aV + W_0, \\ X_1 &= bU + W_1 = abV + bW_0 + W_1, \\ X_2 &= cV + W_2. \end{aligned} \tag{2.15}$$

where a, b, and c are constants, and V, W_0, W_1, and W_2 are independent, zero-mean Gaussian random variables. For $\alpha \in [0, 1]$, $\beta \in [0, 1]$ and $\gamma \in [0, 1]$, define the following:

$$\begin{aligned} \alpha &= \frac{\mathbb{E}\left[W_1^2\right]}{P_1}, \\ \beta &= \frac{\mathbb{E}\left[b^2 W_0^2\right]}{\left(1-\alpha\right) P_1}, \\ \gamma &= \frac{\mathbb{E}\left[W_2^2\right]}{P_2}. \end{aligned} \tag{2.16}$$

We also define the following:

$$\overline{\alpha} = 1 - \alpha, \quad \overline{\beta} = 1 - \beta, \quad \text{and} \quad \overline{\gamma} = 1 - \gamma. \tag{2.17}$$

The random variables Y_2, Y_3, and $\hat{Y}_2$ can then be written as

$$\begin{aligned} Y_2 &= abh_0 V + bh_0 W_0 + h_0 W_1 + Z_2 \\ Y_3 &= \left(abh_1 + ch_2\right) V + bh_1 W_0 + h_1 W_1 + h_2 W_2 + Z_3 \\ \hat{Y}_2 &= Y_2 + Z_W \end{aligned} \tag{2.18}$$

where $Z_W \sim \mathcal{N}\left(0, \sigma_W^2\right)$. We derive the following in Appendix A.3:

$$I\left(X_1; \hat{Y}_2 Y_3 | U X_2\right) = \frac{1}{2}\log_2\left(1 + \alpha P_1\left(\frac{h_1^2}{\sigma_3^2} + \frac{h_0^2}{\sigma_2^2 + \sigma_W^2}\right)\right) \tag{2.19}$$

$$I\left(U; Y_2 | V X_2\right) = \frac{1}{2}\log_2\left(1 + \frac{h_0^2 \beta \overline{\alpha} P_1}{h_0^2 \alpha P_1 + \sigma_2^2}\right) \tag{2.20}$$

$$I\left(X_1 X_2; Y_3\right) = \frac{1}{2}\log_2\left(1 + \frac{h_1^2 P_1 + h_2^2 P_2 + 2h_1 h_2 \sqrt{\overline{\alpha}\overline{\beta}\overline{\gamma} P_1 P_2}}{\sigma_3^2}\right) \tag{2.21}$$

$$I\left(\hat{Y}_2; Y_2 | U X_1 X_2 Y_3\right) = \frac{1}{2}\log_2\left(1 + \frac{\sigma_2^2}{\sigma_W^2}\right) \tag{2.22}$$

where constraint (2.14) translates to

$$\sigma_W^2 \geq \frac{\left[\alpha h_1^2 P_1 \sigma_2^2 + \sigma_3^2\left(\alpha h_0^2 P_1 + \sigma_2^2\right)\right]\left[\left(\beta - \alpha\beta + \alpha\right) h_1^2 P_1 + \sigma_3^2\right]}{\gamma h_2^2 P_2 \left(\alpha h_1^2 P_1 + \sigma_3^2\right)}. \tag{2.23}$$

2.3.5 SeqBack Decoding Strategy

In this section, we derive a new achievable rate for the discrete memoryless relay channel. Similar to the derivation of [14, Thm. 7], we superimpose *cooperation* and the transmission of $\hat{Y}_2$. However, the encoding and decoding methods differ from those of Cover & El Gamal. For encoding, we make use of regular block Markov superposition encoding and for decoding, we make use of backward decoding [33].

The regular encoding scheme is depicted in Fig. 2.2. The regular encoding scheme is depicted in Fig. 2.2. A sequence of B messages $w_{1i} \times w_{2i} \in \left\{1, 2, ..., 2^{NR'}\right\} \times \left\{1, 2, ..., 2^{NR''}\right\}$, $i = 1, 2, ..., B$ will be sent over the channel in $N\left(B + b' + 1\right)$ transmissions. The last b' blocks serve to transmit z_{B+1} from the

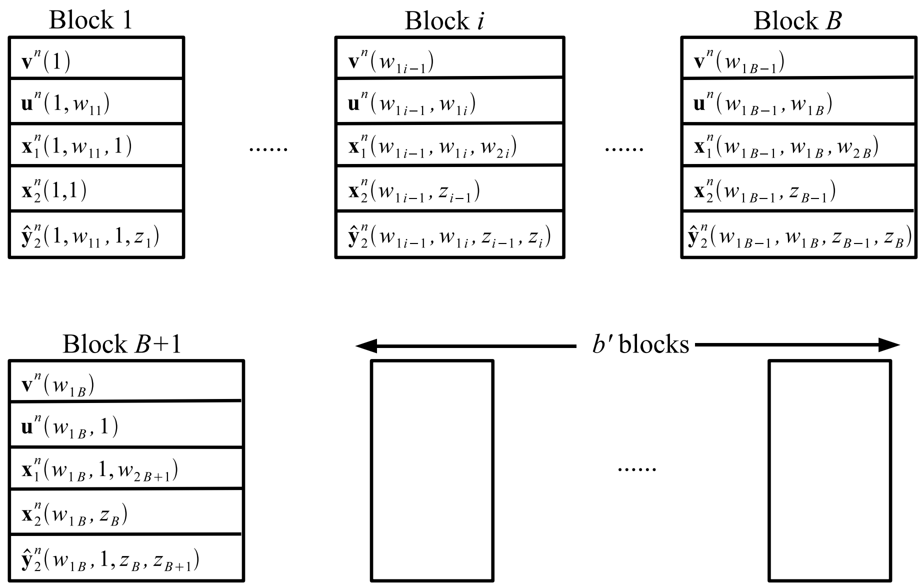

Figure 2.2: Encoding at the transmitter and relay

relay to the receiver so that the receiver can start decoding backwards starting from block $B+1$.

The auxiliary random vector V^N carries information w_{1i-1} that the relay has decoded from the previous block while the auxiliary random vector U^N carries the additional information w_{1i} that the relay can decode from the current block. The index w_{2i} ranges over 1 to $2^{NR''}$ and represents the information that the relay cannot decode. On the other hand, the receiver can decode w_{2i} with the help of the estimate $\hat{y}_2^N$. The index z_i varies over 1 to $2^{N\hat{R}}$ and represents the estimate that the relay intends to communicate to the receiver. The decoding and compression at the relay proceeds as follows:

1. Starting with block 1, the relay decodes w_{11} and determines the compression index z_1. It then transmits the codeword $x_2^N(w_{11}, z_1)$ in the next block.

2. For block i, $2 \leq i \leq B$, the relay (having already decoded w_{1i-1} and determined the compression index z_{i-1}) decodes w_{1i} and determines the compression index z_i. It then transmits the codeword $x_2^N(w_{1i}, z_i)$ in block $i+1$.

3. After block $B+1$, the relay transmits the index z_{B+1} over b' blocks to the receiver.

The receiver starts decoding only after receiving the last block. The decoding at the receiver proceeds as follows:

1. The receiver first makes use of the last b' blocks to decode z_{B+1}.

2. Starting with block $B+1$, the receiver then decodes w_{1B}, followed by z_B and finally by w_{2B+1}.

3. For block i, $2 \leq i \leq B$, the receiver (having already decoded w_{1i} and z_i) proceeds to decode w_{1i-1}, followed by z_{i-1} and then finally by w_{2i}.

The following theorem establishes an achievable rate for this strategy:

Theorem 2.1. *For any relay channel ($\mathcal{X}_1 \times \mathcal{X}_2$, $p(y_2, y_3|x_1, x_2)$, $\mathcal{Y}_2 \times \mathcal{Y}_3$), the following rate is achievable:*

$$R_4 = \sup \left\{ \min \left\{ \begin{array}{c} I(U; Y_2|VX_2Q) + I\left(X_1; \hat{Y}_2 Y_3|UX_2Q\right), \\ I(UV; Y_3|Q) + I\left(X_1; \hat{Y}_2 Y_3|UX_2Q\right) \end{array} \right\} \right\} \tag{2.24}$$

where the supremum is taken over all joint probability density functions of the form (2.13) and subject to the constraint

$$I(X_2; Y_3|UVQ) \geq I\left(\hat{Y}_2; Y_2|UX_2Y_3Q\right). \tag{2.25}$$

Proof. We consider only the probability of error in each block as the total average probability of error can be upper bounded by the sum of the decoding error probabilities at each step, under the assumption that no error propagation from the previous steps has occurred [34]. We will describe in detail the random codebook generation for blocks 1 to $B+1$. After block $B+1$, we use a new codebook in order to reliably transmit z_{B+1} from the relay to the receiver. We will not describe the codebook generation for the last b' blocks in detail as it follows directly from the random codebook generation for a point-to-point channel.

Codebook generation: Fix the probability density function (2.13). We construct the following codebooks independently for all blocks i, $i = 1, 2, ..., B+1$. However, for economy of notation, we will not label the codewords with their block. The reason we generate new codebooks for each block is to guarantee statistical independence between different blocks for random coding arguments. The random codewords to be used in each block are generated independently as follows:

1. Generate a at random N-sequence, q^N, drawn according to

$$p_{Q^N}\left(q^N\right) = \prod_{n=1}^{N} p_Q\left(q_n\right).$$

2. Generate at random $2^{NR'}$ i.i.d. N-sequences, v^N, each drawn according to

$$p_{V^N|Q^N}\left(v^N|q^N\right) = \prod_{n=1}^{N} p_{V|Q}\left(v_n|q_n\right).$$

   Label these $v^N\left(w_p\right), w_p \in \left\{1, 2, ..., 2^{NR'}\right\}$.

3. For each codeword $v^N\left(w_p\right)$, generate $2^{N\hat{R}}$ conditionally independent N-sequences, x_2^N, each drawn according to

$$p_{X_2^N|V^N Q^N}\left(x_2^N|v^N\left(w_p\right), q^N\right) = \prod_{n=1}^{N} p_{X_2|VQ}\left(x_{2n}|v_n\left(w_p\right), q_n\right).$$

   Label these $x_2^N\left(w_p, z_p\right)$, $z_p \in \left\{1, 2, ..., 2^{N\hat{R}}\right\}$.

4. For each codeword $v^N\left(w_p\right)$, generate $2^{NR'}$ conditionally independent N-sequences, u^N, each drawn according to

$$p_{U^N|V^N Q^N}\left(u^N|v^N\left(w_p\right), q^N\right) = \prod_{n=1}^{N} p_{U|VQ}\left(u_n|v_n\left(w_p\right), q_n\right).$$

Label these $u^N(w_p, w)$, $w \in \left\{1, 2, ..., 2^{NR'}\right\}$.

5. For each codeword $u^N(w_p, w)$ and for each codeword $x_2^N(w_p, z_p)$, generate $2^{N\hat{R}}$ conditionally independent N-sequences, $\hat{y}_2^N$, each drawn according to

$$p_{\hat{Y}_2^N|X_2^N U^N Q^N}\left(\hat{y}_2^N|x_2^N(w_p, z_p), u^N(w_p, w), q^N\right)$$
$$= \prod_{n=1}^{N} p_{\hat{Y}_2|X_2UQ}\left(\hat{y}_{2n}|x_{2n}(w_p, z_p), u_n(w_p, w), q_n\right).$$

Label these $\hat{y}_2^N(w_p, w, z_p, z)$, $z = \left\{1, 2, ..., 2^{N\hat{R}}\right\}$.

6. For each codeword $u^N(w_p, w)$, generate $2^{NR''}$ conditionally independent N-sequences, x_1^N, each drawn according to

$$p_{X_1^N|U^N Q^N}\left(x_1^N|u^N(w_p, w), q^N\right) = \prod_{n=1}^{N} p_{X_1|UQ}\left(x_{1n}|u_n(w_p, w), q_n\right).$$

Label these $x_1^N(w_p, w, w_n)$, $w_n \in \left\{1, 2, ..., 2^{NR''}\right\}$.

Encoding and decoding at the relay for block i, $1 \leq i \leq B$:

In block i, the relay would already have decoded w_{1i-1} from the previous block $i-1$. The relay then determines $\hat{w}_{1i}$ such that

$$\left(q^N, u^N(w_{1i-1}, \hat{w}_{1i}), v^N(w_{1i-1}), x_2^N(w_{1i-1}, z_{i-1}), y_2^N(i)\right) \in A_\epsilon^{(N)}(Q, U, V, X_2, Y_2).$$

For sufficiently large N, $\hat{w}_{1i} = w_{1i}$ with arbitrarily high probability (see [25, Chap. 14])if

$$R' < I(U; Y_2|VX_2Q). \tag{2.26}$$

Next, the relay determines z_i such that

$$\left(q^N, u^N(w_{1i-1}, w_{1i}), x_2^N(w_{1i-1}, z_{i-1}), \hat{y}_2^N(w_{1i-1}, w_{1i}, z_{i-1}, z_i), y_2^N(i)\right)$$

$$\in A_\epsilon^{(N)}\left(Q, U, X_2, \hat{Y}_2, Y_2\right).$$

For sufficiently large N, such a z_i will exist with arbitrarily high probability if (2.27)

$$\hat{R} > I\left(Y_2; \hat{Y}_2|UX_2Q\right). \tag{2.27}$$

is satisfied.

Encoding at the relay for block i, $B+2 \leq i \leq B+b^{'}+1$:

After block $B+1$, the relay transmits the last compression index z_{B+1} to the receiver over the last $b^{'}$ blocks.

Decoding at the receiver for the last $b^{'}$ blocks:

The receiver starts decoding only after receiving the last block $y_3^N\left(B+b^{'}+1\right)$. It then makes use of the last $b^{'}$ blocks to decode z_{B+1}. It can be easily shown that if $\max_{x_1 \in \mathcal{X}_1} \max_{p(x_2)} I\left(X_2; Y_3|X_1 = x_1\right) > 0$, z_{B+1} can be transmitted from the relay to the receiver in $b^{'}$ blocks where

$$b^{'} = \left\lceil \frac{I(X_2\hat{Y}_2; Y_3|UV)}{\max_{x_1 \in \mathcal{X}_1} \max_{p(x_2)} I\left(X_2; Y_3|X_1 = x_1\right)} \right\rceil.$$

Since $b^{'}$ is a fixed integer, $\frac{B}{B+b^{'}+1}$ can be made arbitrarily close to unity by choosing B to be large. Hence, the overall rate of transmission will only be reduced by an insignificant amount due to the transmission of the last $b^{'}$ blocks. The overall probability of error can then be made arbitrarily small by allowing $N \to \infty$. If $\max_{x_1 \in \mathcal{X}_1} \max_{p(x_2)} I\left(X_2; Y_3|X_1 = x_1\right) = 0$, the capacity of the relay channel is simply given by that of the reversely degraded relay channel $\max_{x_2 \in \mathcal{X}_2} \max_{p(x_1)} I\left(X_1; Y_3|X_2 = x_2\right)$.

Decoding at the Receiver for block i, $2 \leq i \leq B+1$:

Next, the receiver starts decoding from block $B+1$ and proceeds backwards to block 2. In decoding block i, the receiver has already decoded w_{1i} and z_i accurately from block $i+1$. The receiver then determines the unique $\hat{w}_{1i-1}$ such that

$$\left(q^N, v^N\left(\hat{w}_{1i-1}\right), u^N\left(\hat{w}_{1i-1}, w_{1i}\right), y_3^N(i)\right) \in A_\epsilon^{(N)}\left(Q, V, U, Y_3\right).$$

For sufficiently large N, $\hat{w}_{1i-1} = w_{1i-1}$ with arbitrarily high probability if

$$R^{'} < I\left(UV; Y_3|Q\right). \tag{2.28}$$

Next, it searches for the unique $\hat{z}_{i-1}$ such that

$$\begin{aligned}\left(q^N, v^N\left(w_{1i-1}\right), u^N\left(w_{1i-1}, w_{1i}\right), x_2^N\left(w_{1i-1}, \hat{z}_{i-1}\right), \hat{y}_2^N\left(w_{1i-1}, w_{1i}, \hat{z}_{i-1}, z_i\right), y_3^N(i)\right) \\ \in A_\epsilon^{(N)}\left(Q, V, U, X_2, \hat{Y}_2, Y_3\right).\end{aligned}$$

For sufficiently large N, $\hat{z}_{i-1} = z_{i-1}$ with arbitrarily high probability if

$$\hat{R} < I\left(X_2\hat{Y}_2; Y_3|UVQ\right). \tag{2.29}$$

Finally, the receiver searches for the unique $\hat{w}_{2i}$ such that

$$\begin{aligned}\Big(q^N, v^N\left(w_{1i-1}\right), u^N\left(w_{1i-1}, w_{1i}\right), x_1^N\left(w_{1i-1}, w_{1i}, \hat{w}_{2i}\right), x_2^N\left(w_{1i-1}, z_{i-1}\right), \\ \hat{y}_2^N\left(w_{1i-1}, w_{1i}, z_{i-1}, z_i\right), y_3^N(i)\Big) \in A_\epsilon^{(N)}\left(Q, V, U, X_1, X_2, \hat{Y}_2, Y_3\right).\end{aligned}$$

For sufficiently large N, $\hat{w}_{2i} = w_{2i}$ with arbitrarily high probability if

$$R^{''} < I\left(X_1; \hat{Y}_2Y_3|UX_2Q\right). \tag{2.30}$$

We consider the following:

$$\begin{aligned}I\left(\hat{Y}_2; Y_2Y_3|UX_2Q\right) &= I\left(\hat{Y}_2; Y_2|UX_2Q\right) + I\left(\hat{Y}_2; Y_3|UX_2Y_2Q\right) \\ &= I\left(\hat{Y}_2; Y_2|UX_2Q\right).\end{aligned}$$

From (2.26) and (2.30), we obtain the first term of (2.24). From (2.28) and (2.30), we obtain the second term of (2.24). From (2.27) and (2.29), we obtain the following constraint:

$$
\begin{aligned}
& I\left(Y_2;\hat{Y}_2|UX_2Q\right) \le I\left(X_2\hat{Y}_2;Y_3|UVQ\right) \\
\Rightarrow\quad & I\left(Y_2;\hat{Y}_2|UX_2Q\right) \le I\left(X_2;Y_3|UVQ\right) + I\left(\hat{Y}_2;Y_3|UX_2Q\right) \\
\Rightarrow\; & I\left(\hat{Y}_2;Y_2|UX_2Y_3Q\right) \le I\left(X_2;Y_3|UVQ\right).
\end{aligned}
$$

□

Achievable Rate for the Gaussian Relay Channel

We set $Q \triangleq \emptyset$ and assume $\left(U,V,X_1,X_2,\hat{Y}_2\right)$ to be zero-mean, jointly Gaussian random variables of the same form as (2.15) and (2.18). The parameters α, β and γ are as defined in (2.16). We can show that

$$
\begin{aligned}
I\left(UV;Y_3\right) &= h\left(Y_3\right) - h\left(Y_3|UV\right) \\
&= \frac{1}{2}\log_2\left(1+\frac{h_1^2\overline{\alpha}P_1 + h_2^2\overline{\gamma}P_2 + 2h_1h_2\sqrt{\overline{\alpha}\overline{\beta}\overline{\gamma}P_1P_2}}{h_1^2\alpha P_1 + h_2^2\gamma P_2 + \sigma_3^2}\right).
\end{aligned} \tag{2.31}
$$

We obtain from (2.19) and (2.20) the following:

$$
\begin{aligned}
I\left(X_1;\hat{Y}_2Y_3|UX_2\right) &= \frac{1}{2}\log_2\left(1+\alpha P_1\left(\frac{h_1^2}{\sigma_3^2}+\frac{h_0^2}{\sigma_2^2+\sigma_W^2}\right)\right) \\
I\left(U;Y_2|VX_2\right) &= \frac{1}{2}\log_2\left(1+\frac{h_0^2\beta\overline{\alpha}P_1}{h_0^2\alpha P_1+\sigma_2^2}\right).
\end{aligned}
$$

Next, let us consider

$$
\begin{aligned}
I\left(Y_2;\hat{Y}_2|UX_2Y_3\right) &= h\left(\hat{Y}_2|UX_2Y_3\right) - h\left(\hat{Y}_2|UX_2Y_2Y_3\right) \\
&= h\left(h_0W_1 + Z_2 + Z_W|h_1W_1 + Z_3\right) - h\left(Z_W\right) \\
&= \frac{1}{2}\log_2\left(1+\frac{\alpha P_1\left(h_1^2\sigma_2^2 + h_0^2\sigma_3^2\right)+\sigma_2^2\sigma_3^2}{\sigma_W^2\left(\alpha h_1^2P_1+\sigma_3^2\right)}\right).
\end{aligned} \tag{2.32}
$$

We also consider

$$\begin{aligned}I(X_2;Y_3|UV) &= h(Y_3|UV) - h(Y_3|UVX_2)\\ &= h(h_1W_1 + h_2W_2 + Z_3) - h(h_1W_1 + Z_3)\\ &= \frac{1}{2}\log_2\left(1 + \frac{h_2^2\gamma P_2}{h_1^2\alpha P_1 + \sigma_3^2}\right).\end{aligned} \tag{2.33}$$

Finally, from (2.32) and (2.33), we obtain

$$\sigma_W^2 \geq \frac{\alpha h_1^2 P_1\sigma_2^2 + \alpha h_0^2 P_1\sigma_3^2 + \sigma_2^2\sigma_3^2}{\gamma h_2^2 P_2}. \tag{2.34}$$

2.3.6 SimBack Decoding Strategy

In this section, we exploit the use of *simultaneous decoding* to obtain a new achievable rate for the discrete memoryless relay channel. The codebook generation is exactly the same as in the proof of Thm. 2.1. However, instead of performing sequential decoding at the receiver, we perform simultaneous decoding. In [15], it was shown that the use of simultaneous decoding results in superior performance compared to sequential decoding for the IFC. Hence, in the Simback decoding strategy, instead of decoding $\hat{w}_{1i-1}$ before decoding $\hat{z}_{i-1}$, we decode $\hat{z}_{i-1}$ and $\hat{w}_{1i-1}$ simultaneously. The following theorem establishes an achievable rate for this strategy:

Theorem 2.2. *For any relay channel ($\mathcal{X}_1 \times \mathcal{X}_2$, $p(y_2y_3|x_1x_2)$, $\mathcal{Y}_2 \times \mathcal{Y}_3$), the following rate is achievable:*

$$R_5 = \sup\left\{\min\left\{\begin{array}{l} I\left(X_1;\hat{Y}_2Y_3|UX_2Q\right) + I(U;Y_2|VX_2Q),\\ I(X_1X_2;Y_3|Q) - I\left(Y_2;\hat{Y}_2|UX_1X_2Y_3Q\right)\end{array}\right\}\right\} \tag{2.35}$$

where the supremum is taken over all joint probability density functions of the form (2.13) and subject to the constraint

$$I(X_2;Y_3|UVQ) \geq I\left(\hat{Y}_2;Y_2|UX_2Y_3Q\right). \tag{2.36}$$

Proof. Similar to the proof of Thm. 2.1, a sequence of B messages $w_{1i} \times w_{2i} = \left[1, 2^{NR^{'}}\right] \times \left[1, 2^{NR^{''}}\right]$, $i = 1, 2, ..., B$ will be sent over the channel in $N\left(B + b^{'} + 1\right)$ transmissions. Again, we consider only the probability of error in each decoding step as the total average probability of error can be upper bounded by the sum of the decoding error probabilities at each step, under the assumption that no error propagation from the previous steps has occurred.

Codebook generation:

The codebook to be used in each block are generated exactly as the codebook generation in the proof of Thm. 2.1.

Encoding and decoding at the relay for block i, $1 \leq i \leq B$:

The encoding and decoding at the relay for block i is carried out in exactly the same manner as the SeqBack decoding strategy.

Encoding at the relay for block i, $B + 2 \leq i \leq B + b^{'} + 1$:

The encoding at the relay after block $B + 1$ is carried out in exactly the same manner as the SeqBack decoding strategy.

Decoding at the receiver for the last $b^{'}$ blocks:

The decoding at the receiver for the last $b^{'}$ blocks is carried out in exactly the same manner as the SeqBack decoding strategy.

Decoding at the receiver for block i, $2 \leq i \leq B + 1$:

Since w_{1i} and z_i have been decoded accurately from block $i + 1$, the receiver determines the unique $\hat{w}_{1i-1}$ and $\hat{z}_{i-1}$ such that

$$\Big(q^N, v^N\left(\hat{w}_{1i-1}\right), u^N\left(\hat{w}_{1i-1}, w_{1i}\right), x_2^N\left(\hat{w}_{1i-1}, \hat{z}_{i-1}\right), \\ \hat{y}_2^N\left(\hat{w}_{1i-1}, w_{1i}, \hat{z}_{i-1}, z_i\right), y_3^N\left(i\right)\Big) \in A_\epsilon^{(N)}\left(Q, V, U, X_2, \hat{Y}_2, Y_3\right)$$

For sufficiently large N, $\hat{w}_{1i-1} = w_{1i-1}$ and $\hat{z}_{i-1} = z_{i-1}$ with arbitrarily high probability if

$$R' + \hat{R} < I\left(UX_2\hat{Y}_2; Y_3|Q\right), \tag{2.37}$$

$$\hat{R} < I\left(X_2\hat{Y}_2; Y_3|UVQ\right). \tag{2.38}$$

Finally, the receiver searches for the unique $\hat{w}_{2i}$ such that

$$\Big(q^N, v^N(w_{1i-1}), u^N(w_{1i-1}, w_{1i}), x_1^N(w_{1i-1}, w_{1i}, \hat{w}_{2i}), x_2^N(w_{1i-1}, z_{i-1}),$$
$$\hat{y}_2^N(w_{1i-1}, w_{1i}, z_{i-1}, z_i), y_3^N(i)\Big) \in A_\epsilon^{(N)}\left(Q, V, U, X_1, X_2, \hat{Y}_2, Y_3\right).$$

For sufficiently large n, $\hat{w}_{2i} = w_{2i}$ with arbitrarily high probability if

$$R'' < I\left(X_1; \hat{Y}_2Y_3|UX_2Q\right). \tag{2.39}$$

In [15], the authors show that simultaneous decoding performs better than sequential decoding for the IFC. Hence, in the SimBack decoding strategy, instead of decoding $\hat{w}_{1i-1}$ before decoding $\hat{z}_{i-1}$, we decode $\hat{z}_{i-1}$ and $\hat{w}_{1i-1}$ simultaneously. From (2.27), (2.37) and (2.39), we obtain the following:

$$\begin{aligned} R_5 &< I\left(UX_2\hat{Y}_2; Y_3|Q\right) - I\left(Y_2; \hat{Y}_2|UX_2Q\right) + I\left(X_1; \hat{Y}_2Y_3|UX_2Q\right) \\ &= I(X_1X_2; Y_3|Q) - I\left(\hat{Y}_2; Y_2|UX_1X_2Y_3Q\right). \end{aligned}$$

From (2.26) and (2.39), we obtain

$$R_5 < I(U; Y_2|VX_2Q) + I\left(X_1; \hat{Y}_2Y_3|UX_2Q\right).$$

From (2.27) and (2.38), we obtain the constraint

$$I(X_2; Y_3|UVQ) \geq I\left(\hat{Y}_2; Y_2|UX_2Y_3Q\right).$$

□

Achievable Rate for the Gaussian Relay Channel

We again assume $Q \triangleq \emptyset$ and $\left(U, V, X_1, X_2, \hat{Y}_2\right)$ to be zero-mean, jointly Gaussian random variables of the same form as (2.15) and (2.18). The parameters α, β, and γ are as defined in (2.16). We obtain from Appendix A.3 the following:

$$
\begin{aligned}
I\left(X_1; \hat{Y}_2 Y_3 | U X_2\right) &= \frac{1}{2}\log_2\left(1 + \alpha P_1\left(\frac{h_1^2}{\sigma_3^2} + \frac{h_0^2}{\sigma_2^2 + \sigma_W^2}\right)\right)\\
I\left(U; Y_2 | V X_2\right) &= \frac{1}{2}\log_2\left(1 + \frac{h_0^2 \beta\left(1-\alpha\right)P_1}{h_0^2 \alpha P_1 + \sigma_2^2}\right)\\
I\left(X_1 X_2; Y_3\right) &= \frac{1}{2}\log_2\left(1 + \frac{h_1^2 P_1 + h_2^2 P_2 + 2h_1 h_2\sqrt{\overline{\alpha}\overline{\beta}\overline{\gamma} P_1 P_2}}{\sigma_3^2}\right)\\
I\left(\hat{Y}_2; Y_2 | U X_1 X_2 Y_3\right) &= \frac{1}{2}\log_2\left(1 + \frac{\sigma_2^2}{\sigma_W^2}\right).
\end{aligned}
$$

We also obtain from (2.34) the following constraint:

$$
\sigma_W^2 \geq \frac{\alpha h_1^2 P_1 \sigma_2^2 + \alpha h_0^2 P_1 \sigma_3^2 + \sigma_2^2 \sigma_3^2}{\gamma h_2^2 P_2}.
$$

2.3.7 Sliding Window Decoding Strategy

In this section, we consider a sliding window decoding strategy that achieves the same rate as SeqBack decoding strategy, i.e, it achieves Thm. 2.1. For the sliding window decoding strategy, a sequence of B messages $w_{1i} \times w_{2i} = \left[1, 2^{NR'}\right] \times \left[1, 2^{NR''}\right]$, $i = 1, 2, ..., B$ will be sent over the channel in $N\left(B+1\right)$ transmissions. However, instead of decoding backwards after receiving the last block, the receiver starts decoding block i, after receiving block $i+2$. Moreover, the last b' blocks that was necessary for both SeqBack and SimBack decoding strategies to transmit z_{B+1} from the relay to the receiver is unnecessary for the sliding window decoding strategy. We also modified the sender's transmission in block 1 slightly from that of SeqBack and SimBack decoding strategies. For the sliding window decoding strategy, instead of transmitting $x_1^n\left(1, w_{11}, 1\right)$ in block

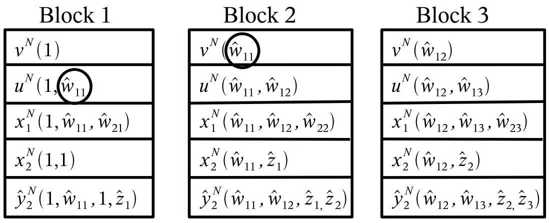

Figure 2.3: Decoding of w_{11}

1, we transmit $x_1^n(1, w_{11}, w_{21})$. This is simply to facilitate the computation of the error probabilities at the receiver. Again, we consider only the probability of error in each decoding step as the total average probability of error can be upper bounded by the sum of the decoding error probabilities at each step, under the assumption that no error propagation from the previous steps has occurred.

Codebook generation:

The codebook to be used in each block are generated independently and exactly as the codebook generation in the proof of Thm. 2.1.

Encoding and decoding at the relay for block i:

The encoding and decoding at the relay for block i is carried out in exactly the same manner as the SeqBack decoding strategy.

Decoding at the receiver:

We assume that the receivers use typical sequence decoder. The conditions follow from standard random coding arguments. We note that since the codebooks have been generated independently for consecutive blocks, statistical independence is maintained between the blocks. We will just take a look at the decoding of the parameters for the first block. Decoding of the rest of the blocks follow exactly. Referring to Fig. 2.3, the receiver decodes w_{11} by using a sliding window of the two past received blocks $y_3^N(1)$ and $y_3^N(2)$. The receiver determines $\hat{w}_{11}$ such

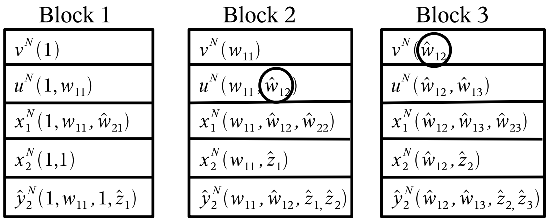

Figure 2.4: Decoding of w_{12}

that

$$\left(q_1^N, v_1^N(1), u_1^N(1, \hat{w}_{11}), x_{21}^N(1,1), y_{31}^N\right) \in A_\epsilon^{(N)}(Q, V, U, X_2, Y_3)$$

and

$$\left(q_2^N, v_2^N(\hat{w}_{11}), y_{32}^N\right) \in A_\epsilon^{(N)}(Q, V, Y_3).$$

This can be decoded with arbitrarily small probability of error as long as N is sufficiently large and

$$R^{'} < I(U; Y_3 | V X_2 Q) + I(V; Y_3 | Q). \tag{2.40}$$

Referring to Fig. 2.4, the receiver next decodes w_{12} by using a sliding window of the two blocks $y_3^N(2)$ and $y_3^N(3)$. The receiver determines $\hat{w}_{12}$ such that

$$\left(q_2^N, v_2^N(w_{11}), u_2^N(w_{11}, \hat{w}_{12}), y_{32}^N\right) \in A_\epsilon^{(N)}(Q, V, U, Y_3)$$

and

$$\left(q_3^N, v_3^N(\hat{w}_{12}), y_{33}^N\right) \in A_\epsilon^{(N)}(Q, V, Y_3).$$

This can be decoded with arbitrarily small probability of error as long as N is sufficiently large and

$$\begin{aligned} R^{'} &< I(U; Y_3 | V Q) + I(V; Y_3 | Q) \\ &= I(UV; Y_3 | Q). \end{aligned} \tag{2.41}$$

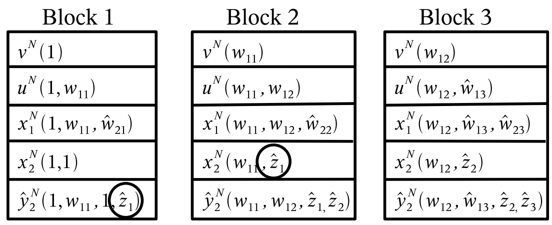

Figure 2.5: Decoding of z_1

We note that (2.41) is a tighter constraint than (2.40) from the following inequalities:

$$\begin{aligned}
&I(U;Y_3|VX_2Q) + I(V;Y_3|Q)\\
&= H(U|VX_2Q) - H(U|VX_2Y_3Q) + I(V;Y_3|Q)\\
&= H(U|VQ) - H(U|VX_2Y_3Q) + I(V;Y_3|Q)\\
&\geq H(U|VQ) - H(U|VY_3Q) + I(V;Y_3|Q)\\
&= I(UV;Y_3|Q). \qquad (2.42)
\end{aligned}$$

Hence, we will choose R' to satisfy (2.41) rather than (2.40). Moreover, the constraint (2.40) only occurs in the decoding of the first block. Referring to Fig. 2.5, the receiver decodes z_1 using a sliding window of the two past received blocks $y_3^N(1)$ and $y_3^N(2)$. The receiver determines $\hat{z}_1$ such that

$$\left(q_1^N, u_1^N(1, w_{11}), x_{21}^N(1,1), \hat{y}_{21}^N(1, w_{11}, 1, \hat{z}_1), y_{31}^N\right) \in A_\epsilon^{(N)}\left(Q, U, X_2, \hat{Y}_2, Y_3\right)$$

and

$$\left(q_2^N, v_2^N(w_{11}), u_2^N(w_{11}, w_{12}), x_{22}^N(w_{11}, \hat{z}_1), y_{32}^N\right) \in A_\epsilon^{(N)}(Q, V, U, X_2, Y_3).$$

This can be decoded with arbitrarily small probability of error as long as N is

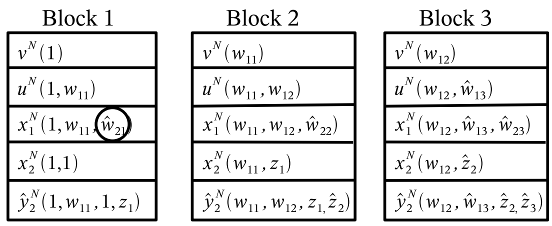

Figure 2.6: Decoding of w_{21}

sufficiently large and

$$\begin{aligned}\hat{R} &< I\left(\hat{Y}_2; Y_3 | U X_2 Q\right) + I\left(X_2; Y_3 | U V Q\right) \\ &= I\left(X_2 \hat{Y}_2; Y_3 | U V Q\right). \end{aligned} \tag{2.43}$$

Referring to Fig. 2.6, the receiver decodes w_{21} using only the first received block $y_3^N(1)$. The receiver determines $\hat{w}_{21}$ such that

$$\begin{aligned}(& q_1^N, u_1^N(1, w_{11}), x_{21}^N(1,1), x_{11}^N(1, w_{11}, \hat{w}_{21}), \\ & \hat{y}_{21}^N(1, w_{11}, 1, z_1), y_{31}^N) \in A_\epsilon^{(N)}\left(Q, U, X_1, X_2, \hat{Y}_2, Y_3\right).\end{aligned}$$

This can be decoded with arbitrarily small probability of error as long as N is sufficiently large and

$$R'' < I\left(X_1; \hat{Y}_2 Y_3 | U X_2 Q\right). \tag{2.44}$$

We note that (2.26), (2.27), (2.41), (2.43) and (2.44) give us the same constraints as SeqBack decoding. Hence, we see that sliding window decoding strategy can also achieve the rate given by Thm. 2.1.

Remark 2.4. Decoding the parameters in a different order gives us a different rate. For example, sequentially decoding w_{11} (using the past received blocks y_{31}^N and y_{32}^N), followed by z_1 (using the past received blocks y_{31}^N and y_{32}^N), followed by w_{21} (using the past received block y_{31}^N) gives us the rate of Cover & El Gamal's

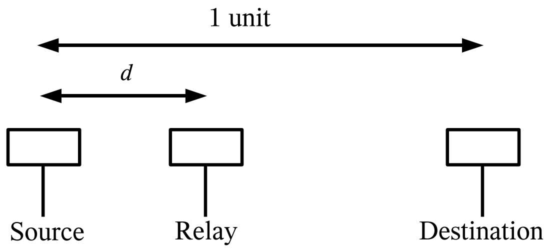

Figure 2.7: Linear configuration for the Gaussian relay channel

generalized strategy. There could potentially be many different ways to decode all the parameters but many of them would give us either the rate of Cover & El Gamal's generalized strategy or the rates given by our generalized strategies.

2.4 Numerical Computations

In this section, we numerically compute the rates for the various strategies described in the previous section, i.e., cut-set upper bound (R_{U}), decode-and-forward (R_1), compress-and-forward (R_2), the generalized lower bound of Cover & El Gamal (R_3), the SeqBack decoding strategy (R_4), and the SimBack decoding strategy (R_5). The physical setup is the Gaussian relay channel shown in Fig. 2.7. Here, the nodes are collinear, the distance between the source and the destination is 1 unit, and the distance between the source and the relay is d. The quantities h_0, h_1 and h_2 are given by

$$h_0 = \frac{1}{d}, \quad h_1 = 1, \quad h_2 = \frac{1}{|1-d|}. \tag{2.45}$$

In all our computations, we have assumed zero-mean, jointly Gaussian random variables. Even though this may not necessarily be optimal, it allows us to compare the rates achieved by the various strategies for this restricted class of probability distributions.

Remark 2.5. From Fig. 2.8 and Fig. 2.9, we note that as the relay node gets closer to the source, i.e., as d decreases, the rates of all the generalized strategies

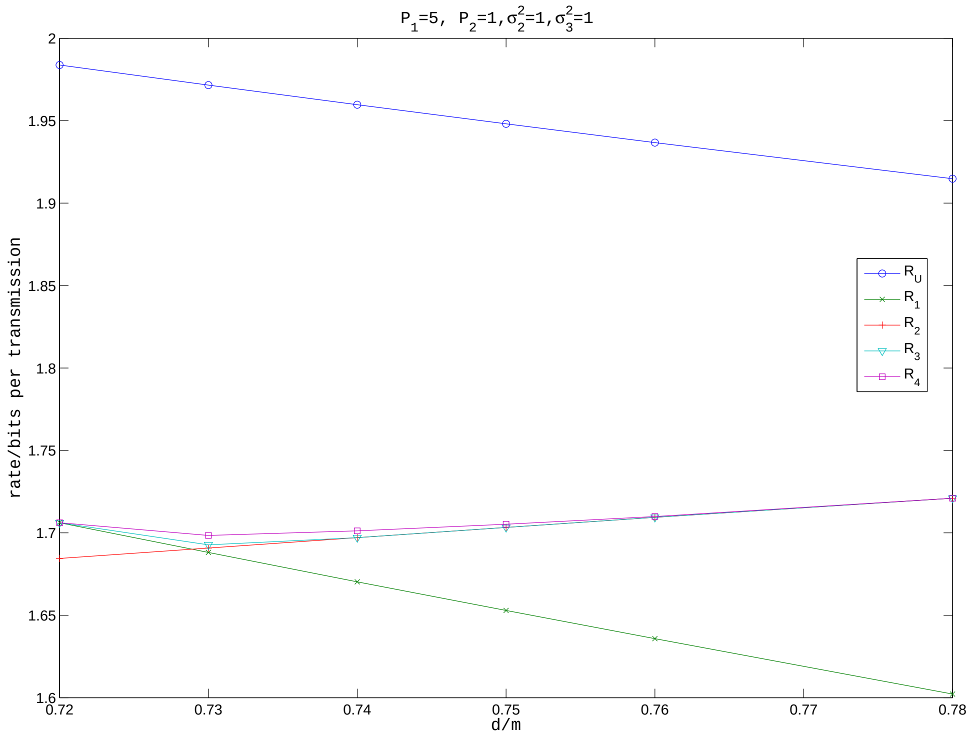

Figure 2.8: Comparison of achievable rates for the relay channel for various coding strategies

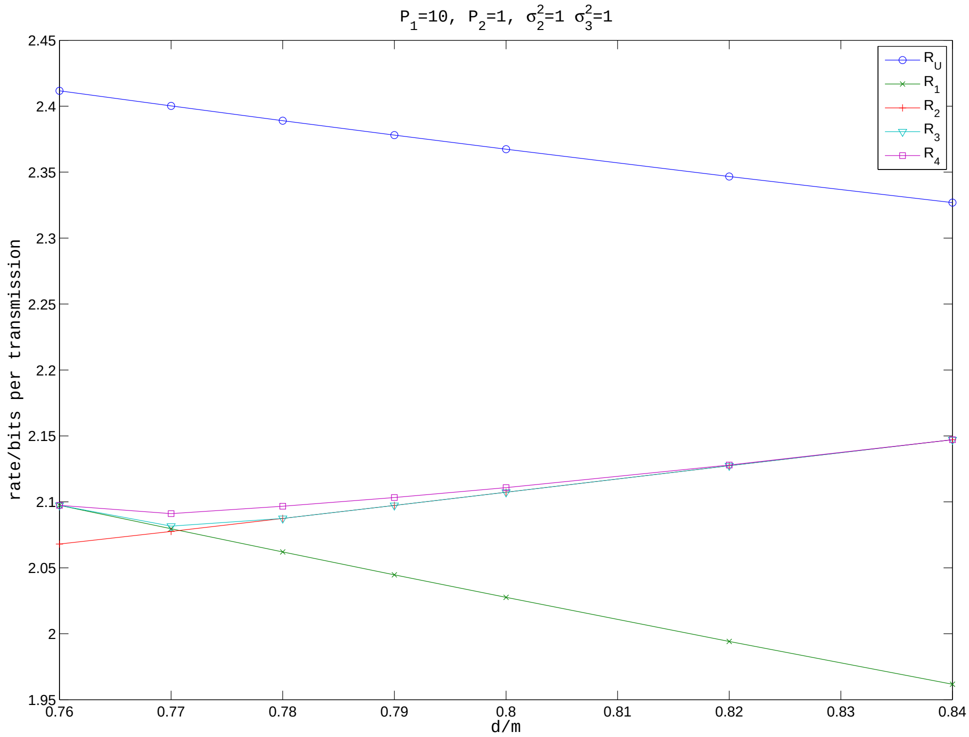

Figure 2.9: Comparison of achievable rates for the relay channel for various coding strategies

coincide with that of the decode-and-forward strategy. Conversely, as the relay node gets closer to the destination, i.e., as d increases, the rates of all the generalized strategies coincide with that of the compress-and-forward strategy. This coincides with the observation in [35], where the authors observed that decode-and-forward performs better as the relay moves toward the source while compress-and-forward performs better as the relay moves toward the destination. In fact, decode-and-forward achieves the capacity when the relay is at the source, while compress-and-forward achieves the capacity when the relay is at the destination. The generalized strategies will offer no improvement over decode-and-forward or compress-and-forward when either one of the two strategies dominates.

Remark 2.6. We also observe that for certain values of h_0, h_1 and h_2, our strategies outperform the decode-and-forward strategy, the compress-and-forward strategy and the generalized strategy of Cover & El Gamal. In general, our strategies outperform the generalized strategy of Cover & El Gamal in regions where the decode-and-forward strategy performs almost as well as the compress-and-forward strategy, i.e., neither decode-and-forward strategy nor compress-and-forward strategy dominates.

2.5 Comparison of the generalized strategies for the relay channel

In this section, we compare the performance of the various generalized strategies.

2.5.1 SeqBack decoding and Simback decoding strategy

From the previous section, we observe that the SeqBack decoding strategy and the SimBack decoding strategy perform equally well for the Gaussian relay channel with the given parameters. In fact, both strategies perform equally well when the constraint $I\left(\hat{Y}_2; Y_2|UX_2Y_3Q\right) \leq I\left(X_2; Y_3|UVQ\right)$ holds with no slack. We can see this from the second term of (2.24)

$$\begin{aligned}
&I\left(X_1;\hat{Y}_2Y_3|UX_2Q\right)+I\left(UV;Y_3|Q\right)\\
&=I\left(X_1;Y_3|UX_2Q\right)+I\left(UV;Y_3|Q\right)+I\left(X_1;\hat{Y}_2|UX_2Y_3Q\right)\\
&=I\left(X_1X_2;Y_3|Q\right)-I\left(X_2;Y_3|UVQ\right)+I\left(X_1;\hat{Y}_2|UX_2Y_3Q\right)\\
&=I\left(X_1X_2;Y_3|Q\right)-I\left(\hat{Y}_2;Y_2|UX_2Y_3Q\right)+I\left(X_1;\hat{Y}_2|UX_2Y_3Q\right)\\
&=I\left(X_1X_2;Y_3|Q\right)-I\left(Y_2;\hat{Y}_2|UX_1X_2Y_3Q\right).
\end{aligned}$$

This is the second term of (2.35). For the Gaussian relay channel with $Q \triangleq \emptyset$ and assuming zero-mean, jointly Gaussian random variables, the rate maximizing distribution is always such that the constraint holds with no slack for both the SeqBack decoding strategy and the SimBack decoding strategy. Hence, both our strategies perform equally well. In general, one might suspect that $R_4 < R_5$ since it is possible that the rate maximizing probability distribution has slackness in the constraint. In this section, we answer this question affirmatively in the following theorem:

Theorem 2.3. *The rate given by Thm. 2.1 is the same as that given by Thm. 2.2, i.e,* $R_4 = R_5$*.*

Proof. Assume that the rate maximizing probability distribution for Thm. 2.2 is as follows:

$$\begin{aligned}
&p_1(q)\,p_1(v|q)\,p_1(u|v,q)\,p_1(x_1|u,q)\\
&\qquad\cdot p_1(x_2|v,q)\,p(y_2,y_3|x_1,x_2)\,p_1(\hat{y}_2|u,x_2,y_2,q)\,. \qquad (2.46)
\end{aligned}$$

Let the joint distribution of the set of random variables $\left(Q^1V^1U^1X_1^1X_2^1Y_2^1\hat{Y}_2^1Y_3^1\right)$ be given by (2.46). Next, let the joint distribution of the set of random variables

$\left(Q^2U^2X_1^2X_2^2Y_2^2\hat{Y}_2^2Y_3^2\right)$, where $V^2 = X_2^2$, be given by

$$p_2\left(q,u,x_1,x_2,y_2,\hat{y}_2,y_3\right) = \sum_{v\in\mathcal{V}} p_1\left(q\right)p_1\left(v|q\right)p_1\left(u|v,q\right)p_1\left(x_1|u,q\right) \cdot p_1\left(x_2|v,q\right)p\left(y_2,y_3|x_1,x_2\right)p_1\left(\hat{y}_2|u,x_2,y_2,q\right). \quad (2.47)$$

Let the random variable I range over $\{1,2\}$, where $0 \leq \Pr\left(I=1\right) = \alpha \leq 1$ and $\Pr\left(I=2\right) = 1-\alpha$. Furthermore, we define $Q \triangleq \left(Q^I, I\right)$, $V \triangleq V^I$, $U \triangleq U^I$, $X_1 \triangleq X_1^I$, $X_2 \triangleq X_2^I$, $Y_2 \triangleq Y_2^I$, $\hat{Y}_2 \triangleq \hat{Y}_2^I$, and $Y_3 \triangleq Y_3^I$ for SeqBack decoding. Essentially, the SeqBack decoding strategy employs a time-sharing strategy between the two codebooks generated by the two probability distributions (2.46) and (2.47) (see also [33, Appendix A]). Next, we need to set an appropriate value for α. If $I\left(X_2^1;Y_3^1|U^1V^1Q^1\right) = 0$, set $\alpha = 0$. Otherwise, we set α as follows:

$$\alpha = \frac{I\left(Y_2^1;\hat{Y}_2^1|U^1X_2^1Y_3^1Q^1\right)}{I\left(X_2^1;Y_3^1|U^1V^1Q^1\right)}. \quad (2.48)$$

For the first term of SeqBack decoding, we obtain

$$\begin{aligned} & I\left(X_1;\hat{Y}_2Y_3|UX_2Q\right) + I\left(U;Y_2|VX_2Q\right) \\ &= I\left(X_1^1;\hat{Y}_2^1Y_3^1|U^1X_2^1Q^1\right) + \alpha\cdot I\left(U^1;Y_2^1|V^1X_2^1Q^1\right) \\ &\quad + (1-\alpha)\cdot I\left(U^2;Y_2^2|V^2X_2^2Q^2\right) \\ &= I\left(X_1^1;\hat{Y}_2^1Y_3^1|U^1X_2^1Q^1\right) + \alpha\cdot I\left(U^1;Y_2^1|V^1X_2^1Q^1\right) \\ &\quad + (1-\alpha)\cdot I\left(U^1;Y_2^1|X_2^1Q^1\right). \end{aligned} \quad (2.49)$$

This is greater than the first term of Thm. 2.2 as seen from the following inequalities:

$$\begin{aligned} I\left(U^1;Y_2^1|X_2^1Q^1\right) &= H\left(Y_2^1|X_2^1Q^1\right) - H\left(Y_2^1|U^1X_2^1Q^1\right) \\ &\geq H\left(Y_2^1|V^1X_2^1Q^1\right) - H\left(Y_2^1|U^1X_2^1Q^1\right) \end{aligned}$$

$$= H\left(Y_2^1|V^1X_2^1Q^1\right) - H\left(Y_2^1|V^1U^1X_2^1Q^1\right)$$
$$= I\left(U^1;Y_2^1|V^1X_2^1Q^1\right). \tag{2.50}$$

Since we have

$$\begin{aligned} & I\left(X_2;Y_3|UVQ\right) - I\left(Y_2;\hat{Y}_2|UX_2Y_3Q\right) \\ &= \alpha I\left(X_2^1;Y_3^1|U^1V^1Q^1\right) - I\left(Y_2;\hat{Y}_2|UX_2Y_3Q\right) \\ &= \alpha I\left(X_2^1;Y_3^1|U^1V^1Q^1\right) - I\left(Y_2^1;\hat{Y}_2^1|U^1X_2^1Y_3^1Q^1\right) \\ &= 0 \end{aligned} \tag{2.51}$$

we obtain for the second term (2.24) of SeqBack decoding

$$I\left(X_1X_2;Y_3|Q\right) - I\left(Y_2;\hat{Y}_2|UX_1X_2Y_3Q\right) \tag{2.52}$$
$$- \left\{I\left(X_2;Y_3|UVQ\right) - I\left(Y_2;\hat{Y}_2|UX_2Y_3Q\right)\right\}$$
$$= I\left(X_1^1X_2^1;Y_3^1|Q^1\right) - I\left(Y_2^1;\hat{Y}_2^1|U^1X_1^1X_2^1Y_3^1Q^1\right) \tag{2.53}$$
$$- \left\{I\left(X_2;Y_3|UVQ\right) - I\left(Y_2;\hat{Y}_2|UX_2Y_3Q\right)\right\}$$
$$= I\left(X_1^1X_2^1;Y_3^1|Q^1\right) - I\left(Y_2^1;\hat{Y}_2^1|U^1X_1^1X_2^1Y_3^1Q^1\right). \tag{2.54}$$

Hence, the second term in (2.24) is always equal to the second term in (2.35). Therefore, all rates achievable by Thm. 2.2 are achievable by Thm. 2.1. Since $R_5 \geq R_4$, we have $R_4 = R_5$. □

2.5.2 SimBack decoding and generalized strategy of Cover & El Gamal

We first cast the achievable rates for SimBack decoding and the generalized strategy of Cover & El Gamal into appropriate forms for comparison. We note that all three generalized strategies are subjected to a constraint, (2.14) for the generalized strategy of Cover & El Gamal, (2.25) for SeqBack decoding, and (2.36) for SimBack decoding. We may express the rates in different forms without the constraints (2.14), (2.25), or (2.36). Since the rates achievable by both SeqBack decoding and SimBack decoding are equal, we may write the rate achievable by both strategies in the following form:

Lemma 2.4. *The rate achievable by both SeqBack and SimBack decoding strategies is given by*

$$R_5 = \sup \left\{ \min \left\{ \begin{array}{c} I\left(X_1X_2;Y_3|Q\right) - I\left(Y_2;\hat{Y}_2|UX_1X_2Y_3Q\right), \\ I\left(X_1;\hat{Y}_2Y_3|UX_2Q\right) + I\left(U;Y_2|VX_2Q\right), \\ I\left(X_1;\hat{Y}_2Y_3|UX_2Q\right) + I\left(U;Y_2|VX_2Q\right) \\ +I\left(X_2;Y_3|UVQ\right) - I\left(Y_2;\hat{Y}_2|UX_2Y_3Q\right) \end{array} \right\} \right\} \tag{2.55}$$

where the supremum is taken over all joint probability distribution functions of the form (2.13).

Proof. We first note that the first two terms of Lem. 2.4 is the same as the two terms of Thm. 2.2. Hence, we see that the rate achievable by Thm. 2.2 is achievable by Lem. 2.4 since $I\left(X_2;Y_3|UVQ\right) - I\left(Y_2;\hat{Y}_2|UX_2Y_3Q\right) \geq 0$ from the constraint (2.36) of Thm. 2.2.

We will show that that the rate achievable by Lem. 2.4 is also achievable by Thm. 2.2. We note that if the rate maximizing probability distribution for Lem. 2.4 is such that $I\left(X_2;Y_3|UVQ\right) - I\left(Y_2;\hat{Y}_2|UX_2Y_3Q\right) \geq 0$, this rate is also achievable by Thm. 2.2.

Hence, we may assume that the rate maximizing probability distribution for Lem. 2.4 is such that $I\left(X_2;Y_3|UVQ\right) - I\left(Y_2;\hat{Y}_2|UX_2Y_3Q\right) < 0$ and is as follows:

$$\begin{aligned} &p_1\left(q\right)p_1\left(v|q\right)p_1\left(u|v,q\right)p_1\left(x_1|u,q\right) \\ &\qquad \cdot p_1\left(x_2|v,q\right)p\left(y_2,y_3|x_1,x_2\right)p_1\left(\hat{y}_2|u,x_2,y_2,q\right). \end{aligned} \tag{2.56}$$

Let the joint distribution of the set of random variables $\left(Q^1V^1U^1X_1^1X_2^1Y_2^1\hat{Y}_2^1Y_3^1\right)$

be given by (2.56). Next, let the joint distribution of the set of random variables $(Q^2V^2U^2X_1^2X_2^2Y_2^2Y_3^2)$, where $\hat{Y}_2^2 \triangleq \emptyset$, be given by

$$p_2\left(q,v,u,x_1,x_2,y_2,y_3\right) \\ = p_1\left(q\right)p_1\left(v|q\right)p_1\left(u|v,q\right)p_1\left(x_1|u,q\right)p_1\left(x_2|v,q\right)p\left(y_2,y_3|x_1,x_2\right). \quad (2.57)$$

Now, let the random variable I range over $\{1,2\}$, where $0 \leq \Pr\left(I=1\right)=\alpha \leq 1$ and $\Pr\left(I=2\right)=1-\alpha$. Furthermore, we define $Q \triangleq \left(Q^I,I\right)$, $V \triangleq V^I$, $U \triangleq U^I$, $X_1 \triangleq X_1^I$, $X_2 \triangleq X_2^I$, $Y_2 \triangleq Y_2^I$, $\hat{Y}_2 \triangleq \hat{Y}_2^I$, and $Y_3 \triangleq Y_3^I$ for SimBack decoding. Next, we need to set an appropriate value for α. We set α as follows:

$$\alpha = \frac{I\left(X_2^1;Y_3^1|U^1V^1Q^1\right)}{I\left(Y_2^1;\hat{Y}_2^1|U^1X_2^1Y_3^1Q^1\right)}. \quad (2.58)$$

We see that the second term of Thm. 2.2 is greater than the first term of Lem. 2.4 as follows:

$$\begin{aligned} & I\left(X_1X_2;Y_3|Q\right) - I\left(Y_2;\hat{Y}_2|UX_1X_2Y_3Q\right) \\ & = I\left(X_1^1X_2^1;Y_3^1|Q^1\right) - \alpha \cdot I\left(Y_2^1;\hat{Y}_2^1|U^1X_1^1X_2^1Y_3^1Q^1\right) \\ & \geq I\left(X_1^1X_2^1;Y_3^1|Q^1\right) - I\left(Y_2^1;\hat{Y}_2^1|U^1X_1^1X_2^1Y_3^1Q^1\right). \end{aligned} \quad (2.59)$$

Moreover, we see that the first term of Thm. 2.2 is greater than the last term of Lem. 2.4 from (2.60) as follows:

$$\begin{aligned} & I\left(X_1^1;Y_3^1|U^1X_2^1Q^1\right) + \alpha \cdot I\left(X_1^1;\hat{Y}_2^1|U^1X_2^1Y_3^1Q^1\right) + I\left(U^1;Y_2^1|V^1X_2^1Q^1\right) \\ & - \left\{ \begin{array}{l} I\left(X_1^1;\hat{Y}_2^1Y_3^1|U^1X_2^1Q^1\right) + I\left(U^1;Y_2^1|V^1X_2^1Q^1\right) \\ +I\left(X_2^1;Y_3^1|U^1V^1Q^1\right) - I\left(Y_2^1;\hat{Y}_2^1|U^1X_2^1Y_3^1Q^1\right) \end{array} \right\} \\ & = I\left(X_1^1;Y_3^1|U^1X_2^1Q^1\right) + \alpha \cdot I\left(Y_2^1;\hat{Y}_2^1|U^1X_2^1Y_3^1Q^1\right) \\ & \quad - \alpha \cdot I\left(Y_2^1;\hat{Y}_2^1|U^1X_1^1X_2^1Y_3^1Q^1\right) + I\left(U^1;Y_2^1|V^1X_2^1Q^1\right) \end{aligned}$$

$$
\begin{aligned}
&-\left\{\begin{array}{l} I\left(X_1^1;\hat{Y}_2^1Y_3^1|U^1X_2^1Q^1\right)+I\left(U^1;Y_2^1|V^1X_2^1Q^1\right)\\ +I\left(X_2^1;Y_3^1|U^1V^1Q^1\right)-I\left(Y_2^1;\hat{Y}_2^1|U^1X_2^1Y_3^1Q^1\right)\end{array}\right\}\\
&=I\left(X_1^1;Y_3^1|U^1X_2^1Q^1\right)+I\left(X_2^1;Y_3^1|U^1V^1Q^1\right)\\
&\quad-\alpha\cdot I\left(Y_2^1;\hat{Y}_2^1|U^1X_1^1X_2^1Y_3^1Q^1\right)+I\left(U^1;Y_2^1|V^1X_2^1Q^1\right)\\
&\quad-\left\{\begin{array}{l} I\left(X_1^1;\hat{Y}_2^1Y_3^1|U^1X_2^1Q^1\right)+I\left(U^1;Y_2^1|V^1X_2^1Q^1\right)\\ +I\left(X_2^1;Y_3^1|U^1V^1Q^1\right)-I\left(Y_2^1;\hat{Y}_2^1|U^1X_2^1Y_3^1Q^1\right)\end{array}\right\}\\
&=I\left(Y_2^1;\hat{Y}_2^1|U^1X_1^1X_2^1Y_3^1Q^1\right)-\alpha\cdot I\left(Y_2^1;\hat{Y}_2^1|U^1X_1^1X_2^1Y_3^1Q^1\right)\geq 0. \qquad (2.60)
\end{aligned}
$$

Hence, all rates achievable by Lem. 2.4 are achievable by Thm. 2.2. Since all rates achievable by Thm. 2.2 is also achievable by Lem. 2.4, the two rates are in fact equivalent. □

Remark 2.7. We obtain the same rate given by Lem. 2.4 if we decode all unknown parameters (q^N, v^N, u^N, x_2^N, $\hat{y}_2^N$, x_1^N) in a single block simultaneously.

Following exactly along the same lines, we can express the first term (2.12) of the generalized strategy of Cover & El Gamal to be of the same form as the first term of Lem. 2.4. We state the following version of the generalized strategy of Cover & El Gamal as a lemma below:

Lemma 2.5. *The rate achievable by Cover & El Gamal's strategy is given by*

$$
R_3=\sup\left\{\min\left\{\begin{array}{c} I\left(X_1X_2;Y_3|Q\right)-I\left(Y_2;\hat{Y}_2|UX_1X_2Y_3Q\right),\\ I\left(X_1;\hat{Y}_2Y_3|UX_2Q\right)+I\left(U;Y_2|VX_2Q\right),\\ I\left(X_1;\hat{Y}_2Y_3|UX_2Q\right)+I\left(U;Y_2|VX_2Q\right)\\ +I\left(X_2;Y_3|VQ\right)-I\left(Y_2;\hat{Y}_2|UX_2Y_3Q\right)\end{array}\right\}\right\} \qquad (2.61)
$$

where the supremum is taken over all joint probability distribution functions of the form (2.13).

Proof. Follows exactly along the lines of the proof of Thm. 2.3 and Lem. 2.4. □

Remark 2.8. Proof of Lem. 2.5 above makes use of previous techniques used in the proofs of Thm. 2.3 and Lem. 2.4. We note that the first two terms of Lem. 2.5 is of the same form as that given in [14, Thm. 7]. However, the third term of Lem. 2.5 has combined the constraint of [14, Thm. 7] with the second term of [14, Thm. 7].

Remark 2.9. Comparing SeqBack and SimBack decoding with the generalized strategy of Cover & El Gamal, we note that the first two terms of Lem. 2.4 and Lem. 2.5 are exactly the same. However, the last term of SeqBack/SimBack decoding strategy is more relaxed than that of Cover & El Gamal's strategy as seen from the following inequalities:

$$\begin{aligned} I\left(X_2; Y_3|UVQ\right) &= H\left(X_2|UVQ\right) - H\left(X_2|UVY_3Q\right) \\ &= H\left(X_2|VQ\right) - H\left(X_2|UVY_3Q\right) \\ &\geq H\left(X_2|VQ\right) - H\left(X_2|VY_3Q\right) \\ &= I\left(X_2; Y_3|VQ\right). \end{aligned} \tag{2.62}$$

We readily see that $R_3 \leq R_5$. In the previous section, we showed an improvement for both SeqBack decoding and SimBack decoding compared to the generalized strategy of Cover & El Gamal by looking at the Gaussian relay channel (Strictly speaking, the rates for all the generalized strategies apply only to discrete memoryless relay channels) and restricting the input distribution to the class of zero-mean Gaussian random variables. For certain parameters of the Gaussian relay channel, we find that $R_5 > R_3$ for this restricted class of Gaussian input distributions. Hence, we conjecture that in general $R_5 > R_3$.

Chapter 3

Relay Channel with General Alphabets

3.1 Introduction

In the previous chapter, we presented three different generalized strategies for the discrete memoryless relay channel. One of the strategies uses sequential backward (SeqBack) decoding, the second strategy exploits simultaneous backward (SimBack) decoding, while the third strategy employs a sliding-window decoding strategy. It was also shown that all three strategies achieve the same rate and was also shown to contain the rate of Cover & El Gamal's generalized strategy. The proof for our generalized strategies and Cover & El Gamal's generalized strategy require the use of strong typicality to invoke Berger's Markov Lemma, [25, Lemma 14.8.1], [41]. However, strong typicality does not apply to continuous random variables. In [35, Remark 30], Kramer, Gastpar & Gupta made a comment that the Markov lemma can be generalized along the lines of [38], and thereby show that [14, Thm. 6] can be applied to the Gaussian relay channel. However, the input distributions must be restricted to the class of Gaussian distributions and no rigorous proof of the coding theorem is given.

In this chapter, we extend the rate of the SeqBack decoding strategy to relay channels with general alphabets. As we have already noted, the proof of the coding theorem given in the previous chapter relies heavily on the discreteness of the channel. We quote an additional problem noted by Wyner in extending results from discrete channels to continuous channels in the case of source coding with side information at the decoder [42].

"*In many other situations in the Shannon theory, such proofs can be easily adapted to the non-discrete case by finding appropriate discrete approximations to non-discrete random objects. In the present problem, however, this approach is particularly difficult. Among the reasons is the following: Let* X*,* Y*,* Z *be a Markov chain of non-discrete random variables, i.e.,* X*,* Z *are conditionally independent given* Y*. Let* $\tilde{X}$*,* $\tilde{Y}$*, and* $\tilde{Z}$ *be finite approximations to* X*,* Y*, and* Z*, respectively. Then* $\tilde{X}$*,* $\tilde{Y}$*, and* $\tilde{Z}$ *is not necessarily a chain.*"

For the generalized decoding strategies, Markov chains also play a vital role in the proof. It is clear that we must also proceed with care.

3.1.1 Outline

This chapter develops a generalized decoding strategy (SeqBack) for the relay channel with general alphabets. We prove the main coding theorem using weak typicality rather than strong typicality. We modify the proof of the extended Markov lemma for Gaussian sources by Oohama [38]. However, without being limited to Gaussian input distributions, our result extends to all input distributions with well-defined probability densities (with appropriate σ-finite measures that will be defined later on). This chapter is organized as follows:

- In Section 3.2, we define the mathematical model for the relay channel under study and the required technical tools. We also review some basic results for jointly typical sequences for random variables with probability densities.
- Section 3.3 summarizes the main result of this chapter.

- Section 3.4 gives an overview of the preprocessing necessary at the relay so that joint typicality for random variables with densities is sufficient for the proof of our coding theorem.
- In Section 3.5, we describe in detail the encoding at the source, decoding at the relay and preprocessing at the relay. We also describe the SeqBack decoding strategy at the receiver and compute in detail the probability of error events at both the relay and receiver.

3.2 Model and Preliminaries

We first describe the mathematical model of the relay channel with standard alphabets. Throughout the chapter, we assume random variables take values from a standard space [43, Sec. 2]. A measurable space $(\mathcal{A}, \mathcal{F}_A)$ is standard (or a Borel space) if $\mathcal{A}$ is a Borel subset of a complete separable metric space and $\mathcal{F}_A$ is the class of Borel subsets of $\mathcal{A}$. For standard alphabets, regular conditional probability distributions always exist. An additional property of standard spaces is that the Cartesian product of standard spaces is also a standard space.

3.2.1 Relay Channel Model

A memoryless relay channel consists of a source input space $(\mathcal{X}_1, \mathcal{F}_{X_1})$, a relay channel input space $(\mathcal{X}_2, \mathcal{F}_{X_2})$, a relay output space $(\mathcal{Y}_2, \mathcal{F}_{Y_2})$, a destination output space $(\mathcal{Y}_3, \mathcal{F}_{Y_3})$, and a regular conditional probability distribution $\Upsilon(E_{Y_2Y_3}|x_1, x_2)$: $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$, $E_{Y_2Y_3} \in \mathcal{F}_{Y_2} \times \mathcal{F}_{Y_3}$.

An $\left(2^{NR}, N\right)$ code for the relay channel is composed of a set of integers $\mathcal{M} = \left\{1, 2, ..., 2^{NR}\right\}$, an encoder which is a (measurable) mapping

$$\Psi : \left\{1, 2, ..., 2^{NR}\right\} \to \mathcal{X}_1^N$$

a set of relay functions $\{\Phi_n\}_{n=1}^{n=N}$ such that

$$\Phi_n : \mathcal{Y}_2^n \to \mathcal{X}_n, \;\; 1 \leq n \leq N,$$

and a decoder

$$\Phi_{N+1} : \mathcal{Y}_3^N \to \left\{1, 2, ..., 2^{NR}\right\}.$$

The relay is causal in nature. Therefore, the relay channel input is allowed to depend only on past observations $y_{21}, y_{22}, ..., y_{2n-1}$. In addition, the channel is assumed to be memoryless. Hence, if the input sequences are x_1^N and x_2^N, the output sequences are governed by the Cartesian product measure

$$\Upsilon^N\left(.|x_1^N, x_2^N\right) = \prod_{n=1}^{n=N} \Upsilon\left(.|x_{1n}, x_{2n}\right)$$

We will denote the marginal conditional distributions $\Upsilon^N\left(\mathcal{Y}_2^N \times .|x_1^N, x_2^N\right)$ and $\Upsilon^N\left(. \times \mathcal{Y}_3^N|x_1^N, x_2^N\right)$ by the same notation as above, where the context will make it clear. If the message $m \in \mathcal{M}$ is sent, let $\lambda(m)$ denote the probability of error. The average probability of error is defined by

$$P_e^{(N)} = \frac{1}{2^{NR}} \sum_m \lambda(m).$$

The probability of error is calculated under the uniform distribution over the codewords $m \in \mathcal{M}$. The rate R is said to be achievable by the relay channel if there exists a sequence of $\left(2^{NR}, N\right)$ codes with $P_e^{(N)} \to 0$ as $N \to \infty$. The capacity C_{R} is the supremum of the set of achievable rates.

3.2.2 Entropy, Conditional Entropy, and Mutual Information

Let X and Y be random variables taking values in standard spaces $(\mathcal{X}, \mathcal{F}_X)$ and $(\mathcal{Y}, \mathcal{F}_Y)$. Let P_{XY} and M_{XY} be two distributions on $(\mathcal{X} \times \mathcal{Y}, \mathcal{F}_X \times \mathcal{F}_Y)$ and assume that $M_{XY} >> P_{XY}$ ($M_{XY} >> P_{XY}$ means that M_{XY} dominates P_{XY}, i.e., for each $E_{XY} \in \mathcal{F}_X \times \mathcal{F}_Y$, $M_{XY}(E_{XY}) = 0$ implies $P_{XY}(E_{XY}) = 0$). Let P_Y and M_Y denote the induced marginal distributions. Define the densities (Radon-Nikodym derivatives):

$$f_{XY} = \frac{dP_{XY}}{dM_{XY}}, \quad f_Y = \frac{dP_Y}{dM_Y}. \tag{3.1}$$

Define also the conditional density

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{XY}(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0 \\ 1 & \text{otherwise.} \end{cases} \tag{3.2}$$

Supposing that these densities exist, we define the relative entropy as

$$H_{P||M}(Y) = \int f_Y(y) \log \frac{1}{f_Y(y)} dM_Y(y) \tag{3.3}$$

and the conditional relative entropy as

$$H_{P||M}(X|Y) = \int f_{XY}(x,y) \log \frac{1}{f_{X|Y}(x|y)} dM_{XY}(x,y). \tag{3.4}$$

Hence, the following chain rule applies for conditional relative entropies:

$$H_{P||M}(X,Y) = H_{P||M}(Y) + H_{P||M}(X|Y). \tag{3.5}$$

We then define the mutual information as follows:

$$I(X;Y) = H_{P||M}(X) + H_{P||M}(Y) - H_{P||M}(X,Y). \tag{3.6}$$

Remark 3.1. Such a definition of mutual information is certainly more restrictive than the usual definition of mutual information for general alphabets by Dobrushin [44] or the definition of mutual information in terms of the divergence of the joint distribution P_{XY} of the random variable X and Y and their product distribution $P_X \times P_Y$ [26, eq. 5.5.1]. We can readily find examples where we might have $\infty + \infty - \infty$ in the RHS of (3.6). However, such a definition allows us to readily extend results for the discrete relay channels to relay channels with general alphabets using standard typical set decoding arguments and a modification of Oohama's Markov lemma. Throughout the chapter, we assume that all the quantities defined above are $< \infty$.

Given a third random variable Z taking values in a standard space $(\mathcal{Z}, \mathcal{F}_Z)$. Suppose that $M_{XYZ} >> P_{XYZ}$ (absolute continuity implies absolute continuity for the restrictions), define the conditional mutual information as

$$I\left(X;Y|Z\right) = H_{P||M}\left(Y|Z\right) + H_{P||M}\left(X|Z\right) - H_{P||M}\left(XY|Z\right). \tag{3.7}$$

If $X \to Y \to Z$ form a Markov chain under M and P, we have [26, Cor. 5.3.3]

$$f_{X|YZ}\left(x|yz\right) = f_{X|Y}\left(x|y\right) \tag{3.8}$$

in which case $H_{P||M}\left(X|Y\right) = H_{P||M}\left(X|YZ\right)$ and $I\left(X;Z|Y\right) = 0$. Finally, we define $P_{X\times Z|Y}$ as follows:

$$P_{X\times Z|Y}\left(E_X \times E_Z \times E_Y\right) = \int_{E_Y} P_{X|Y}\left(E_X|y\right) P_{Z|Y}\left(E_Z|y\right) dP_Y\left(y\right) \tag{3.9}$$

where $E_X \in \mathcal{F}_X$, $E_Y \in \mathcal{F}_Y$, and $E_Z \in \mathcal{F}_Z$.

3.2.3 Jointly typical sequences

Let us assume a standard space $(\mathcal{X}, \mathcal{F}_X)$ together with a σ-finite measure M. For a $\mathcal{F}_X$-set A, we define $\mathrm{vol}\left(A\right)$ as follows:

$$\mathrm{vol}\left(A\right) = \int_A dM. \tag{3.10}$$

We now extend the AEP for densities to a form that we will use to prove our coding theorems. Let $(X_1, X_2, ..., X_k)$ denote a finite collection of standard alphabets taking values in a standard space. Let S denote an ordered subset of these random variables. Suppose that the joint distribution P for $(X_1, X_2, ..., X_k)$ has a density $f(x_1, x_2, ..., x_k)$ with respect to a σ-finite measure M. Hence, the joint distribution P_S for S has a density $f_S(s)$ with respect to the restricted σ-finite measure M_S. Consider N independent copies of S, we then have

$$f_{S^N}\left(s^N\right) = \prod_{n=1}^{n=N} f_S(s_n), \quad s^N \in \mathcal{S}^N.$$

Lemma 3.1.

1. *For $\epsilon > 0$ and sufficiently large N, there exists a $\mathcal{F}_{X_1^N} \times \mathcal{F}_{X_2^N} ... \times \mathcal{F}_{X_k^N}$-measurable set $A_\epsilon^{(N)}(X_1, X_2, ..., X_k)$ that satisfies*

$$\left\{\left(x_1^N, x_2^N, ..., x_k^N\right) : \left|-\frac{1}{N}\log f_{S^N}\left(s^N\right) - H_{P||M}(S)\right| < \epsilon, \forall S \subseteq \{X_1, X_2, ..., X_k\}\right\}.$$

   and

$$P_{S^N}\left(A_\epsilon^{(N)}(S)\right) \geq 1-\epsilon, \quad \forall S \subseteq \{X_1, X_2, ..., X_k\}$$

2. $s^N \in A_\epsilon^{(N)}(S) \Rightarrow \left|-\frac{1}{N}\log f_{S^N}\left(s^N\right) - H_{P||M}(S)\right| < \epsilon$

3. $(1-\epsilon)\, 2^{N\left(H_{P||M}(S)-\epsilon\right)} \leq \text{vol}\left(A_\epsilon^{(N)}(S)\right) \leq 2^{N\left(H_{P||M}(S)+\epsilon\right)}$

4. *Let S_1 and S_2 be two subsets of $\{X_1, X_2, ..., X_k\}$. If $\left(s_1^N, s_2^N\right) \in A_\epsilon^{(N)}(S_1, S_2)$, then*

$$2^{-N\left(H_{P||M}(S_1|S_2)+2\epsilon\right)} \leq f_{S_1^N|S_2^N}\left(s_1^N|s_2^N\right) \leq 2^{-N\left(H_{P||M}(S_1|S_2)-2\epsilon\right)}.$$

5. *Let S_1 and S_2 be two subsets of $\{X_1, X_2, ..., X_k\}$. If s_2^N is an element in $A_\epsilon^{(N)}(S_1, S_2)$, for any $\epsilon > 0$, define $A_\epsilon^{(N)}\left(S_1|s_2^N\right)$ to be the $\mathcal{F}_{S_1^N}$-measurable set of s_1^N sequences that are jointly ϵ-typical with a particular s_2^N sequence,*

then for sufficiently large N, we have

$$\mathrm{vol}\left(A_\epsilon^{(N)}\left(S_1|s_2^N\right)\right) \leq 2^{N\left(H_{P||M}(S_1|S_2)+2\epsilon\right)}.$$

Proof. 1) follows from directly from [45, Thm. 1]. 2) follows directly from the definition of $A_\epsilon^{(N)}(S)$. 3) follows from

$$\begin{aligned}
1-\epsilon &\leq P_{S^N}\left(A_\epsilon^{(N)}(S)\right) \\
&= \int_{A_\epsilon^{(N)}(S)} dP_{S^N} \\
&= \int_{A_\epsilon^{(N)}(S)} f_{S^N}\left(s^N\right) dM_{S^N} \\
&\leq \int_{A_\epsilon^{(N)}(S)} 2^{-N\left(H_{P||M}(S)-\epsilon\right)} dM_{S^N} \\
&= 2^{-N\left(H_{P||M}(S)-\epsilon\right)} \int_{A_\epsilon^{(N)}(S)} dM_{S^N} \\
&= 2^{-N\left(H_{P||M}(S)-\epsilon\right)} \mathrm{vol}\left(A_\epsilon^{(N)}(S)\right)
\end{aligned} \tag{3.11}$$

and

$$\begin{aligned}
1 &\geq P_{S^N}\left(A_\epsilon^{(N)}(S)\right) \\
&= \int_{A_\epsilon^{(N)}(S)} dP_{S^N} \\
&= \int_{A_\epsilon^{(N)}(S)} f_{S^N}\left(s^N\right) dM_{S^N} \\
&\geq \int_{A_\epsilon^{(N)}(S)} 2^{-N\left(H_{P||M}(S)+\epsilon\right)} dM_{S^N} \\
&= 2^{-N\left(H_{P||M}(S)+\epsilon\right)} \int_{A_\epsilon^{(N)}(S)} dM_{S^N} \\
&= 2^{-N\left(H_{P||M}(S)+\epsilon\right)} \mathrm{vol}\left(A_\epsilon^{(N)}(S)\right)
\end{aligned} \tag{3.12}$$

and 4) follows, since if $\left(s_1^N, s_2^N\right) \in A_\epsilon^{(N)}(S_1, S_2)$, we have $2^{-N\left(H_{P||M}(S_2)+\epsilon\right)} \leq f_{S_2^N}\left(s_2^N\right) \leq 2^{-N\left(H_{P||M}(S_2)-\epsilon\right)}$ and also $2^{-N\left(H_{P||M}(S_1,S_2)+\epsilon\right)} \leq f_{S_1^N S_2^N}\left(s_1^N, s_2^N\right) \leq$

$2^{-N\left(H_{P||M}(S_1,S_2)-\epsilon\right)}$. Therefore, we obtain

$$2^{-N\left(H_{P||M}(S_1|S_2)+2\epsilon\right)} \leq \frac{f_{S_1^N S_2^N}\left(s_1^N, s_2^N\right)}{f_{S_2^N}\left(s_2^N\right)} \leq 2^{-N\left(H_{P||M}(S_1|S_2)-2\epsilon\right)}.$$

Finally, 5) follows from [26, Lem. 5.3.2], we have

$$\begin{aligned} 1 &= \int_{\mathcal{S}_1^N} f_{S_1^N|S_2^N}\left(s_1^N|s_2^N\right) dM_{S_1^N|S_2^N} \\ &\geq \int_{A_\epsilon^{(N)}\left(S_1|s_2^N\right)} f_{S_1^N|S_2^N}\left(s_1^N|s_2^N\right) dM_{S_1^N|S_2^N} \\ &\geq \int_{A_\epsilon^{(N)}\left(S_1|s_2^N\right)} 2^{-N\left(H_{P||M}(S_1|S_2)+2\epsilon\right)} dM_{S_1^N|S_2^N} \\ &\geq 2^{-N\left(H_{P||M}(S_1|S_2)+2\epsilon\right)} \int_{A_\epsilon^{(N)}\left(S_1|s_2^N\right)} dM_{S_1^N|S_2^N} \\ &= 2^{-N\left(H_{P||M}(S_1|S_2)+2\epsilon\right)} \mathrm{vol}\left(A_\epsilon^{(N)}\left(S_1|s_2^N\right)\right). \end{aligned} \tag{3.13}$$

□

Remark 3.2. We see that the joint AEP for discrete random variables also holds for random variables with densities. When M is the counting measure, $H_{P||M}(X)$ is the entropy of the discrete random variable X and $\mathrm{vol}\left(A_\epsilon^{(N)}(X)\right)$ denotes the number of elements in the set $A_\epsilon^{(N)}(X)$. When M is the Lebesgue measure, $H_{P||M}(X)$ is the differential entropy of the continuous random variable X and $\mathrm{vol}\left(A_\epsilon^{(N)}(X)\right)$ is the smallest volume set with probability $\geq 1-\epsilon$.

3.3 Summary of Main Results

We now summarize the main contributions of this chapter. First, we state our main result in Thm. 3.2 below.

Theorem 3.2. *For any relay channel* $(\mathcal{X}_1 \times \mathcal{X}_2, \Upsilon(.|x_1, x_2), \mathcal{Y}_2 \times \mathcal{Y}_3)$*, the following* *rate:*

$$R_{\mathrm{CMG}} = \sup\left\{\min\left\{\begin{array}{c} I\left(X_1;\hat{Y}_2Y_3|UX_2Q\right) + I\left(UV;Y_3|Q\right), \\ I\left(X_1;\hat{Y}_2Y_3|UX_2Q\right) + I\left(U;Y_2|VX_2Q\right) \end{array}\right\}\right\} \tag{3.14}$$

subject to the constraint

$$I\left(X_2\hat{Y}_2;Y_3|UVQ\right) \geq I\left(Y_2;\hat{Y}_2|UX_2Q\right). \tag{3.15}$$

is achievable, where the supremum is taken over all densities of joint distributions $P_{QVUX_1X_2\hat{Y}_2Y_2Y_3}$ with a dominating σ-finite measure $M_{QVUX_1X_2\hat{Y}_2Y_2Y_3}$. Both the probability distribution and the marginal dominating σ-finite measure must factor as follows:

$$\begin{aligned} &P_{QVUX_1X_2\hat{Y}_2Y_2Y_3}\left(E_Q \times E_V \times E_U \times E_{X_1} \times E_{X_2} \times E_{\hat{Y}_2} \times E_{Y_2} \times E_{Y_3}\right) \\ &= \int_{\left(E_Q\times E_V\times E_U\times E_{X_1}\times E_{X_2}\times E_{\hat{Y}_2}\times E_{Y_2}\times E_{Y_3}\right)} dP_{QVUX_1X_2\hat{Y}_2Y_2Y_3} \\ &= \int_{E_Q}\int_{E_V}\int_{E_U}\int_{E_{X_1}}\int_{E_{X_2}}\int_{E_{Y_2}\times E_{Y_3}}\int_{E_{\hat{Y}_2}} dP_{\hat{Y}_2|UX_2Y_2Q}dP_{Y_2Y_3|X_1X_2}dP_{X_2|VQ}dP_{X_1|UQ} \\ &\qquad \times dP_{U|VQ}dP_{V|Q}dP_Q \end{aligned} \tag{3.16}$$

where $E_Q \in \mathcal{Q}$, $E_V \in \mathcal{V}$, $E_U \in \mathcal{U}$, $E_{X_1} \in \mathcal{X}_1$, $E_{X_2} \in \mathcal{X}_2$, $E_{\hat{Y}_2} \in \hat{\mathcal{Y}}_2$, $E_{Y_2} \in \mathcal{Y}_2$, and $E_{Y_3} \in \mathcal{Y}_3$. In addition, the marginal dominating σ-finite measure must factor as follows:

$$M_{QVUX_2Y_2} = M_{U\times X_2Y_2|QV} \tag{3.17}$$

$$M_{QVUY_3} = M_{UV\times Y_3|Q} \tag{3.18}$$

$$M_{QVUX_2\hat{Y}_2Y_3} = M_{X_2\hat{Y}_2\times Y_3|QVU} \tag{3.19}$$

$$M_{QVUX_1X_2\hat{Y}_2Y_3} = M_{X_1\times VX_2\hat{Y}_2Y_3|QU} \tag{3.20}$$

$$M_{QUX_2Y_2\hat{Y}_2} = M_{Y_2\times\hat{Y}_2|UX_2Q}. \tag{3.21}$$

Proof. Refer to Section 3.4 and Section 3.5. □

- Since absolute continuity ($M_{XYZ} >> P_{XYZ}$) implies absolute continuity for the restrictions (e.g. $M_{XY} >> P_{XY}, M_X >> P_X, ...$) and from [26, Cor. 5.3.3], we have from Thm. 3.2

$$M_{QVUX_2Y_2} = M_{U\times X_2Y_2|QV} >> P_{U\times X_2Y_2|QV} >> P_{QVUX_2Y_2} \tag{3.22}$$

$$M_{QVUY_3} = M_{UV\times Y_3|Q} >> P_{UV\times Y_3|Q} >> P_{QVUY_3} \tag{3.23}$$

$$M_{QVUX_2\hat{Y}_2Y_3} = M_{X_2\hat{Y}_2\times Y_3|QVU} >> P_{X_2\hat{Y}_2\times Y_3|QVU} >> P_{QVUX_2\hat{Y}_2Y_3} \tag{3.24}$$

$$M_{QVUX_1X_2\hat{Y}_2Y_3} = M_{X_1\times VX_2\hat{Y}_2Y_3|QU} >> P_{X_1\times VX_2\hat{Y}_2Y_3|QU} >> P_{QVUX_1X_2\hat{Y}_2Y_3} \tag{3.25}$$

$$M_{QUX_2Y_2\hat{Y}_2} = M_{Y_2\times\hat{Y}_2|UX_2Q} >> P_{Y_2\times\hat{Y}_2|UX_2Q} >> P_{QUX_2Y_2\hat{Y}_2}. \tag{3.26}$$

- Thm. 3.2 was derived for the SeqBack decoding strategy in the previous chapter for the discrete memoryless relay channel using strong typicality. In this chapter, we prove Thm. 3.2 using joint typicality proved in Lem. 3.1 and therefore extend this result to relay channels with general (standard) alphabets. Our proof follows along the same lines of Oohama's proof of the extended Markov lemma for Gaussian sources [38]. However, it is not necessary to restrict the input probability distributions to the class of Gaussian distributions as in [35, Remark 30]. We however need to restrict our attention to the class of probability distributions with well-defined probability densities (Rem. 3.1).

- Even though the requirement of a dominating σ-finite measure (which factors as in Thm. 3.2) may seem restrictive, Thm. 3.2 allows us to obtain rates for a fairly large class of relay channels. By setting the dominating measure to be the counting measure, we immediately obtain achievable rates for the discrete memoryless relay channel. By setting the dominating measure to be the Lebesgue measure, we obtain achievable rates for the Gaussian relay channel with well-defined continuous input probability density functions. We may also obtain achievable rates for mixed input distributions by setting the dominating measure to be the Lebesgue measure plus the counting measure. For all these cases, conditions (3.22)-(3.26) are clearly satisfied.

3.4 Preprocessing at the Relay and Codebook generation

The codebook generation is essentially the same as the one devised in the previous chapter. The only change is the compression or preprocessing at the relay. We first take a look at the Markov Lemma by Berger which is at the heart of past proofs of achievable rate regions of generalized strategies. The following Lemma is quoted without proof in [25, Lem. 14.8.1]:

Lemma 3.3 (Markov Lemma). *Let (X, Y, Z) form a Markov chain $X \leftrightarrow Y \leftrightarrow Z$. If for a given $\left(y^N, z^N\right) \in A_\epsilon^{*(N)}(Y, Z)$, X^N is drawn according to $\prod_{i=1} p\left(x_i|y_i\right)$ then $\Pr\left\{\left(X^N, y^N, z^N\right) \in A_\epsilon^{*(N)}(X, Y, Z)\right\} > 1-\epsilon$ for N sufficiently large.*

Proof. Refer to [41]. □

Remark 3.3. When y^N is jointly strongly typical with z^N and x^N is jointly strongly typical with y^N, it does not necessarily follow that all three are jointly strongly typical. The Markovity of $X \leftrightarrow Y \leftrightarrow Z$ is a sufficient condition to ensure that all three sequences are jointly strongly typical.

When we set $Q \triangleq \emptyset$ in (3.16), we note that $(V, X_1, Y_3) \leftrightarrow (U, X_2, Y_2) \leftrightarrow \hat{Y}_2$ form a Markov chain. Since the relay cannot decode x_1^N and it does not know the receiver output y_3^N, the relay can only ensure that $\left(u^N, x_2^N, y_2^N, \hat{y}_2^N\right) \in A_\epsilon^{*(N)}\left(U, X_2, Y_2, \hat{Y}_2\right)$. Due to the Markov lemma for strong typicality, this condition is sufficient to ensure that with a high probability $\left(v^N, u^N, x_1^N, x_2^N, \hat{y}_2^N, y_3^N\right) \in A_\epsilon^{*(N)}\left(V, U, X_1, X_2, \hat{Y}_2, Y_3\right)$. Hence, the use of ϵ-strongly typical sequences was necessary in previous proofs of the coding theorems for generalized relay strategies. In this chapter, we modify the compression at the relay along the lines of Oohama [38] in order to prove the coding theorem using ϵ-weakly typical sequences. Hence, Thm. 3.2 can be readily extended to relay channels with continuous alphabets. We first need the following two definitions:

Definition 3.1.

$$\Gamma\left(q^N, u^N, x_2^N, y_2^N, \hat{y}_2^N\right) = \int_{A_\epsilon^{(N)}\left(V, X_1, Y_3 | q^N, u^N, x_2^N, y_2^N, \hat{y}_2^N\right)} dP_{V^N X_1^N Y_3^N | Q^N U^N X_2^N Y_2^N} \tag{3.27}$$

Definition 3.2. The $\mathcal{F}_{Q^N} \times \mathcal{F}_{U^N} \times \mathcal{F}_{X_2^N} \times \mathcal{F}_{Y_2^N} \times \hat{\mathcal{Y}}_{Q^N}$-measurable set $S_{\lambda,\epsilon}^{(N)}\left(Q, U, X_2, Y_2, \hat{Y}_2\right)$ satisfies the following:

$$\begin{aligned} & S_{\lambda,\epsilon}^{(N)}\left(Q, U, X_2, Y_2, \hat{Y}_2\right) \\ & = \Big\{ \left(q^N, u^N, x_2^N, y_2^N, \hat{y}_2^N\right) \in \mathcal{Q}^N \times \mathcal{U}^N \times \mathcal{X}_2^N \times \mathcal{Y}_2^N \times \hat{\mathcal{Y}}_2^N : \\ & \qquad\qquad \Gamma\left(q^N, u^N, x_2^N, y_2^N, \hat{y}_2^N\right) \geq 1 - \lambda \Big\} \end{aligned} \tag{3.28}$$

where $\lambda \in (0, 1)$.

Remark 3.4. At the relay, we determine $\hat{y}_2^N$ such that $\left(q^N, u^N, x_2^N, y_2^N, \hat{y}_2^N\right) \in S_{\lambda,\epsilon}^{(N)}\left(Q, U, X_2, Y_2, \hat{Y}_2\right)$, instead of $\left(q^N, u^N, x_2^N, y_2^N, \hat{y}_2^N\right) \in A_\epsilon^{*(N)}\left(Q, U, X_2, Y_2, \hat{Y}_2\right)$. This condition ensures that with a high probability $\left(q^N, v^N, u^N, x_1^N, x_2^N, \hat{y}_2^N, y_3^N\right) \in A_\epsilon^{(N)}\left(Q, V, U, X_1, X_2, \hat{Y}_2, Y_3\right)$.

3.4.1 Codebook Construction, Preprocessing, and Termination

The messages $w_{1b} \in \mathcal{W}_1$ and $w_{2b} \in \mathcal{W}_2$, where $\mathcal{W}_1 = \left\{1, 2, ..., 2^{NR'}\right\}$, $\mathcal{W}_2 = \left\{1, 2, ..., 2^{NR''}\right\}$ and $b = 1, 2, ..., B$, will be sent over the relay channel in $B + 1$

blocks, each of N transmissions. Similarly, the estimate of the relay $z_b \in \mathcal{Z}$, where $\mathcal{Z} = \left\{1, 2, ..., 2^{N\hat{R}}\right\}$ and $b = 1, 2, ..., B$, will be sent by the relay in $B+1$ blocks, each of N transmissions. We generate a different codebook independently for each block. Finally, the compression index z_{B+1} will be sent over the relay channel in the last $b^{'}$ blocks similar to that described in the previous chapter.

We first fix the probability distribution $P_{QVUX_1X_2\hat{Y}_2Y_2Y_3}$ in (3.16) and generate the codebook as in the previous chapter. For completeness sake, we reproduce the codebook generation for the first $B+1$ blocks here. (However, similar to our description for the SeqBack decoding strategy in the last chapter, we will not describe the codebook generation for the last $b^{'}$ blocks.) In each of the $B+1$ blocks, the codebook is constructed independently as follows:

1. Generate at random one i.i.d. N-sequence $q^N = (q_1, q_2, ..., q_N)$, drawn according to the distribution

$$P_{Q^N}(.) = \prod_{n=1}^{n=N} P_Q(.) .$$

2. Generate at random $2^{NR^{'}}$ i.i.d. N-sequences $v^N = (v_1, v_2, ..., v_N)$, drawn according to the conditional distribution

$$P_{V^N|Q^N}\left(.|q^N\right) = \prod_{n=1}^{n=N} P_{V|Q}(.|q_n) .$$

   Label them $v^N(w_{1\text{p}}), w_{1\text{p}} \in \mathcal{W}_1$.

3. For each codeword $v^N(w_{1\text{p}})$, generate $2^{N\hat{R}}$ conditionally independent $x_2^N = (x_{21}, x_{22}, ..., x_{2N})$, each drawn according to the conditional distribution

$$P_{X_2^N|Q^NV^N}\left(.|q^N, v^N(w_{1\text{p}})\right) = \prod_{n=1}^{n=N} P_{X_2|QV}(.|q_n, v_n(w_{1\text{p}})) .$$

   Label them $x_2^N(w_{1\text{p}}, z_\text{p})$, $z_\text{p} \in \mathcal{Z}$.

4. For each codeword $v^N(w_{1\mathrm{p}})$, generate $2^{nR'}$ conditionally independent N-sequences $u^N = (u_1, u_2, ..., u_N)$, each drawn according to the conditional distribution

$$P_{U^N|Q^N V^N}\left(.|q^N, v^N(w_{1\mathrm{p}})\right) = \prod_{n=1}^{n=N} P_{U|QV}\left(.|q_n, v_n(w_{1\mathrm{p}})\right).$$

   Label them $u^N(w_{1\mathrm{p}}, w_1)$, $w_1 \in \mathcal{W}_1$.

5. For each codeword $u^N(w_{1\mathrm{p}}, w_1)$ and each of the corresponding codeword $x_2^N(w_{1\mathrm{p}}, z_{\mathrm{p}})$, generate $2^{n\hat{R}}$ conditionally independent N-sequences $\hat{y}_2^N = (\hat{y}_{21}, \hat{y}_{22}, ..., \hat{y}_{2N})$, each drawn according to the conditional distribution

$$\begin{aligned} &P_{\hat{Y}_2^N|Q^N U^N X_2^N}\left(.|q^N, u^N(w_{1\mathrm{p}}, w_1), x_2^N(w_{1\mathrm{p}}, z_{\mathrm{p}})\right) \\ &= \prod_{n=1}^{n=N} P_{\hat{Y}_2|QUX_2}\left(.|q_n, u_n(w_{1\mathrm{p}}, w_1), x_{2n}(w_{1\mathrm{p}}, z_{\mathrm{p}})\right). \end{aligned}$$

   Label them $\hat{y}_2^N(w_{1\mathrm{p}}, w_1, z_{\mathrm{p}}, z)$, $z \in \mathcal{Z}$.

6. For each codeword $u^N(w_{1\mathrm{p}}, w_1)$, generate $2^{nR''}$ conditionally independent N-sequences $x_1^N = (x_{11}, ..., x_{1N})$ according to the conditional distribution

$$P_{X_1^N|Q^N, U^N}\left(.|q^N, u^N(w_{1\mathrm{p}}, w_1)\right) = \prod_{n=1}^{n=N} P_{X_1|QU}\left(.|q_n, u_n(w_{1\mathrm{p}}, w_1)\right).$$

   Label them $x_1^N(w_{1\mathrm{p}}, w_1, w_2)$, $w_2 \in \mathcal{W}_2$.

We represent all the $2^{N\hat{R}}$ $\hat{y}_2^N$ codewords generated for a particular q^N, u^N and x_2^N by C_1. We need a mapping function after generating C_1 to map $\left(q^N, u^N, x_2^N, y_2^N\right)$ to a codeword $\hat{y}_2^N$ in C_1. Let us define a (measurable) mapping $\Theta : \mathcal{Q}^N \times \mathcal{U}^N \times \mathcal{X}_2^N \times \mathcal{Y}_2^N \mapsto \hat{\mathcal{Y}}_2$ for block b at the relay as follows:

$$\Theta\left(q^N, u^N(i,j), x_2^N(i,l), y_2^N(b)\right)$$
$$= \begin{cases} \hat{y}_2^N(i,j,l,m) & \text{if there exists an integer } m \geq 2 \text{ such that} \\ & \left(q^N, u^N(i,j), x_2^N(i,l), y_2^N(b), \hat{y}_2^N(i,j,l,m)\right) \\ & \qquad \in S_{\lambda,\epsilon}^N\left(Q, U, X_2, Y_2, \hat{Y}_2\right) \\ & \text{and } \hat{y}_2^N\left(i,j,l,m'\right) \neq \hat{y}_2^N(i,j,l,m), \text{ where } 1 < m' < m \\ \hat{y}_2^N(i,j,l,1) & \text{if no such } \hat{y}_2^N \text{ exists.} \end{cases} \tag{3.29}$$

We define a similar (measurable) mapping $\Theta_1 : \mathcal{Q}^N \times \mathcal{U}^N \times \mathcal{X}_2^N \times \mathcal{Y}_2^N \mapsto \mathcal{Z}$ for block b at the relay as follows:

$$\Theta_1\left(q^N, u^N(i,j), x_2^N(i,l), y_2^N(b)\right)$$
$$= \begin{cases} m & \text{if there exists an integer } m \geq 2 \text{ such that} \\ & \left(q^N, u^N(i,j), x_2^N(i,l), y_2^N(b), \hat{y}_2^N(i,j,l,m)\right) \in S_{\lambda,\epsilon}^N\left(Q, U, X_2, Y_2, \hat{Y}_2\right) \\ & \text{and } \hat{y}_2^N\left(i,j,l,m'\right) \neq \hat{y}_2^N(i,j,l,m), \text{ where } 1 < m' < m \\ 1 & \text{if no such } \hat{y}_2^N \text{ exists.} \end{cases} \tag{3.30}$$

Remark 3.5. The mapping function Θ maps $\left(q^N, u^N, x_2^N, y_2^N\right)$ to the codeword $\hat{y}_2^N$ while the mapping function Θ_1 maps $\left(q^N, u^N, x_2^N, y_2^N\right)$ to the index of that codeword. The two functions above denote only one possible mapping. Let Q_Θ and $\mathbb{E}_{Q_\Theta}[.]$ denote the probability measure and expectation based on the randomness of the choice of the two functions.

3.5 Computation of Probabilities of error

The encoding at the source and decoding cum preprocessing at the relay for each block proceeds as described in the previous chapter. Decoding and encoding at the relay for time b, $1 \leq b \leq B+1$, proceeds as follows:

- From the decoding of block $b-1$, assuming no errors has propagated from the previous decoding, the relay knows w_{1b-1}. It also knows z_{b-1} since this is determined at the relay. The relay then chooses $\hat{w}_{1b}$ such that

$$\left(q^N, v^N\left(w_{1b-1}\right), u^N\left(w_{1b-1}, \hat{w}_{1b}\right), x_2^N\left(w_{1b-1}, z_{b-1}\right), y_2^N(b)\right) \in A_\epsilon^{(N)}\left(Q, V, U, X_2, Y_2\right). \quad (3.31)$$

  For block 1, both the source and relay assumes $w_{10} = 1$ and $z_0 = 1$. The source transmits $x_1^N(1, w_{11}, 1)$ while the relay transmits $x_2^N(1,1)$. After time $B+1$, the relay transmits the compressed index z_{B+1} over the last b' blocks.

- Assuming that the relay decodes w_{1b} accurately, it now determines z_b using the preprocessing function such that

$$z_b = \Theta_1\left(q^N, u^N\left(w_{1b-1}, w_{1b}\right), x_2^N\left(w_{1b-1}, z_{b-1}\right), y_2^N(b)\right). \quad (3.32)$$

- The relay then transmits $x_2^N(w_{1b}, z_b)$ at time $b+1$.

3.5.1 Error Events at the Relay

We give a detailed computation of each of the error events at the relay. First, let us define the following $\mathcal{F}_{Y_2^N}$-measurable set at the relay:

$$\mathcal{K}_{\hat{j}} = \left\{y_2^N : \left(q^N, v^N(i), u^N\left(i, \hat{j}\right), x_2^N(i,l), y_2^N\right) \in A_\epsilon^{(N)}\left(Q, V, U, X_2, Y_2\right)\right\}. \quad (3.33)$$

We denote the probability measure on measurable events taking place at the relay in block b by $P_{\mathrm{r},b}$, the probability measure on measurable events taking place at the receiver in block b by $P_{\mathrm{d},b}$, and the probability measure on the indices sent in time b by P_b. Assuming no errors have propagated from the previous block, we have the probability of error at the relay in block b given by

$$P_{\mathrm{r},b}\left(\mathcal{E}\right) = \sum_{i=1}^{2^{NR'}} \sum_{j=1}^{2^{NR'}} \sum_{k=1}^{k=2^{NR''}} \sum_{l=1}^{2^{N\hat{R}}} P_b\left(i,j,k,l\right) \Upsilon^N \left(\mathcal{K}_j^c \bigcup \cup_{\hat{j}\neq j} \mathcal{K}_{\hat{j}} | x_1^N\left(i,j,k\right), x_2^N\left(i,l\right)\right) \tag{3.34}$$

Since the indices (i,j,k,l) sent are independent of the codebook and mapping function generated for block b, we obtain

$$\begin{aligned}\overline{P_{\mathrm{r},b}\left(\mathcal{E}\right)} &= \overline{\Upsilon^N\left(\mathcal{K}_j^c | x_1^N\left(i,j,k\right), x_2^N\left(i,l\right)\right)} + \sum_{\hat{j}\neq j} \overline{\Upsilon^N\left(\mathcal{K}_{\hat{j}} | x_1^N\left(i,j,k\right), x_2^N\left(i,l\right)\right)} \\ &\leq \epsilon + 2^{-N\left(I(U;Y_2|VX_2Q)-R'-6\epsilon\right)}\end{aligned} \tag{3.35}$$

Refer to Appendix B.1 for the detailed computations. Hence, the following condition is sufficient to ensure that (3.34) when averaged over the ensemble of codewords and mapping functions for block b tend to 0 as $N \to \infty$:

$$R' \leq I\left(U;Y_2|VX_2Q\right). \tag{3.36}$$

3.5.2 Error events for SeqBack Decoding Strategy

Next, we consider the SeqBack decoding strategy. For backward decoding strategies, the receiver starts decoding only after receiving the last block $y_3^N\left(B+b'+1\right)$. The receiver first uses the last b' blocks to decode z_{B+1}. Then it starts decoding each block proceeding backwards from block $B+1$ to block 2. Consider block b, $2 \leq b \leq B+1$, as shown in Figure 3.1. From the decoding of block $b+1$,

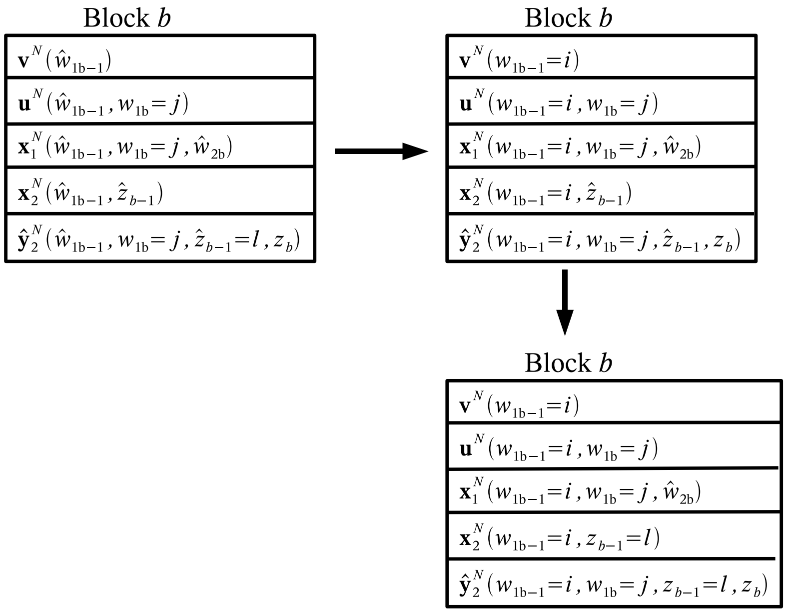

Figure 3.1: SeqBack Decoding Strategy

assuming no error has occurred in the decoding of the previous block, the receiver knows $w_{1b} = j$ and z_b. The receiver then determines the unique w_{1b-1}, z_{b-1} and w_{2b} in sequential order.

- First, the receiver chooses the unique $\hat{w}_{1b-1}$ such that

$$\left(q^N, v^N\left(\hat{w}_{1b-1}\right), u^N\left(\hat{w}_{1b-1}, j\right), y_3^N\left(b\right)\right) \in A_\epsilon^{(N)}\left(Q, V, U, Y_3\right). \tag{3.37}$$

  An error is declared if none such or more than one such is found.

- Assuming that the receiver has decoded the unique $w_{1b-1} = i$, the receiver then chooses the unique $\hat{z}_{b-1}$ such that

$$\begin{aligned}&\left(\ q^N, v^N\left(i\right), u^N\left(i, j\right), x_2^N\left(i, \hat{z}_{b-1}\right), \hat{y}_2^N\left(i, j, \hat{z}_{b-1}, z_b\right), y_3^N\left(b\right)\ \right)\\ &\qquad\qquad\qquad\qquad \in A_\epsilon^{(N)}\left(Q, V, U, X_2, \hat{Y}_2, Y_3\right).\end{aligned}$$

  An error is declared if none such or more than one such is found.

- Assuming that the receiver has decoded the unique $z_{b-1} = l$, finally, the receiver chooses the unique $\hat{w}_{2b}$ such that

$$\left(q^N, v^N(i), u^N(i,j), x_1^N(i,j,\hat{w}_{2b})\, x_2^N(i,l), \hat{y}_2^N(i,j,l,z_b), y_3^N(b) \right) \in A_\epsilon^{(N)}\left(Q,V,U,X_1,X_2,\hat{Y}_2,Y_3\right).$$

An error is declared if none such or more than one such is found.

Let us first define the following $\mathcal{F}_{Y_2^N} \times \mathcal{F}_{Y_3^N}$ measurable sets:

$$\mathcal{L}_{\hat{i}} = \left\{ \left(y_2^N, y_3^N\right) : \left(q^N, v^N\left(\hat{i}\right), u^N\left(\hat{i}, j\right), y_3^N(b)\right) \in A_\epsilon^{(N)}(Q,V,U,Y_3) \right\}, \tag{3.38}$$

$$\mathcal{L}_{\hat{l}} = \left\{ \left(y_2^N, y_3^N\right) : \left(q^N, v^N(i), u^N(i,j), x_2^N\left(i,\hat{l}\right), \hat{y}_2^N\left(i,j,\hat{l},z_b\right), y_3^N(b)\right) \in A_\epsilon^{(N)}\left(Q,V,U,X_2,\hat{Y}_2,Y_3\right) \right\}, \tag{3.39}$$

$$\mathcal{L}_{\hat{k}} = \left\{ \left(y_2^N, y_3^N\right) : \left(q^N, v^N(i), u^N(i,j), x_2^N(i,l), \hat{y}_2^N(i,j,l,z_b), x_1^N\left(i,j,\hat{k}\right), y_3^N(b)\right) \in A_\epsilon^{(N)}\left(Q,V,U,X_1,X_2,\hat{Y}_2,Y_3\right) \right\}, \tag{3.40}$$

$$\mathcal{J}_d = \left\{ \left(y_2^N, y_3^N\right) : \left(q^N, v^N(i), u^N(i,j), x_1^N(i,j,k), x_2^N(i,l), y_2^N(b), \Theta\left(q^N, u^N(i,j), x_2^N(i,l), y_2^N(b)\right), y_3^N(b)\right) \notin A_\epsilon^{(N)}\left(Q,V,U,X_1,X_2,Y_2,\hat{Y}_2,Y_3\right) \right\}. \tag{3.41}$$

Assuming no errors have propagated from the previous decodings, we have the probability of error at the receiver in block b given by

$$
\begin{aligned}
&P_{\mathrm{d},b}\left(\mathcal{E}\right)\\
&=\sum_{i=1}^{2^{NR'}}\sum_{j=1}^{2^{NR'}}\sum_{k=1}^{k=2^{NR''}}\sum_{l=1}^{2^{N\hat{R}}}P_b\left(i,j,k,l\right)\\
&\qquad\times\Upsilon^N\left(\mathcal{L}_i^c\bigcup\mathcal{L}_l^c\bigcup\mathcal{L}_k^c\bigcup\cup_{\hat{i}\neq i}\mathcal{L}_{\hat{i}}\bigcup\cup_{\hat{l}\neq i}\mathcal{L}_{\hat{l}}\bigcup\cup_{\hat{k}\neq k}\mathcal{L}_{\hat{k}}|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)\\
&\leq\sum_{i=1}^{2^{NR'}}\sum_{j=1}^{2^{NR'}}\sum_{k=1}^{k=2^{NR''}}\sum_{l=1}^{2^{N\hat{R}}}P_b\left(i,j,k,l\right)\\
&\qquad\times\Upsilon^N\left(\mathcal{J}_d\bigcup\cup_{\hat{i}\neq i}\mathcal{L}_{\hat{i}}\bigcup\cup_{\hat{l}\neq i}\mathcal{L}_{\hat{l}}\bigcup\cup_{\hat{k}\neq i}\mathcal{L}_{\hat{k}}|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right). \qquad (3.42)
\end{aligned}
$$

Since the indices (i,j,k,l) are independent of the codebook and mapping function generated for block b, we have

$$
\begin{aligned}
\overline{P_{\mathrm{d},b}\left(\mathcal{E}\right)}&\leq\overline{\Upsilon^N\left(\mathcal{J}_d|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)}+\sum_{\hat{i}\neq i}\overline{\Upsilon^N\left(\mathcal{L}_{\hat{i}}|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)}\\
&\quad+\sum_{\hat{l}\neq l}\overline{\Upsilon^N\left(\mathcal{L}_{\hat{l}}|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)}+\sum_{\hat{k}\neq k}\overline{\Upsilon^N\left(\mathcal{L}_{\hat{k}}|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)}\\
&\leq\overline{\Upsilon^N\left(\mathcal{J}_d|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)}\\
&\quad+2^{-N\left(I(UV;Y_3|Q)-R'-6\epsilon\right)}+2^{-N\left(I\left(X_2\hat{Y}_2;Y_3|UVQ\right)-\hat{R}-6\epsilon\right)}\\
&\quad+2^{-N\left(I\left(X_1;\hat{Y}_2Y_3|UX_2Q\right)-R''-4\epsilon\right)} \qquad (3.43)
\end{aligned}
$$

Refer to Appendix B.1 for details of the proof for the bounds of the second, third and fourth term. To bound the first term, we need the following theorem:

Theorem 3.4. *For any positive μ and ϵ, there exists an integer $N_1 = r(\mu,\epsilon)$ such that for $N \geq N_1$, we have*

$$\overline{\Upsilon^N\left(\mathcal{J}_d|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)}\leq 3\mu \qquad (3.44)$$

if

$$\hat{R}>I\left(Y_2;\hat{Y}_2|UX_2Q\right)+3\epsilon. \qquad (3.45)$$

Proof. Refer to Appendix B.2 for the proof. □

Remark 3.6. Thm. 3.4 is a variant of Lem. 3.3 (Berger's Markov Lemma). Assuming $X \leftrightarrow Y \leftrightarrow Z$ form a Markov chain, and y^N and z^N are jointly typical, a sufficient condition that with a high probability x^N is jointly typical with both y^N and z^N is to choose a x^N such that

$$P_{Z^N|Y^N}\left(A_\epsilon^{(N)}\left(Z|x^N, y^N\right)|y^N\right) \geq 1 - \lambda \tag{3.46}$$

where λ is as defined in (3.28). Hence, restriction to the set $S_{\lambda,\epsilon}^{(N)}(X, Y)$ is sufficient to ensure that x^N, y^N and z^N are in the jointly typical set $A_\epsilon^{(N)}(X, Y, Z)$ with probability greater than $1-\lambda$. The need for strong typicality is unnecessary for this result.

Finally, from (3.43) and (3.45), it is easy to see that the following conditions imply that each of the terms tends to 0 as $N \to \infty$:

$$\hat{R} \geq I\left(Y_2; \hat{Y}_2|UX_2Q\right), \tag{3.47}$$

$$\hat{R} \leq I\left(X_2\hat{Y}_2; Y_3|UVQ\right), \tag{3.48}$$

$$R^{'} \leq I\left(UV; Y_3|Q\right), \tag{3.49}$$

$$R^{''} \leq I\left(X_1; \hat{Y}_2Y_3|UX_2Q\right). \tag{3.50}$$

To complete the proof, we note that when averaged over all codebooks and mapping functions for block b, both the terms $P_{\mathrm{r},b}(\mathcal{E})$ and $P_{\mathrm{d},b}(\mathcal{E})$ tends to zero as $N \to \infty$ as long as the above conditions coupled with (3.36) are satisfied. Hence, there exists at least one codebook and mapping function for block b such that both the terms tend to 0 as $N \to \infty$. From (3.47) and (3.48), we obtain the constraint of Thm. 3.2. From (3.49) and (3.50), we obtain the first term of Thm. 3.2. From (3.36) and (3.50), we obtain the second term of Thm. 3.2. This completes the proof of Thm. 3.2.

Chapter 4

On the Interference Channel

4.1 Introduction

The interference channel (IFC) models the situation where M unrelated senders try to communicate their separate messages to M different receivers via a common channel as shown in Fig. 4.1. In this model, there is no cooperation between any of the senders or receivers, and hence, the transmission from each sender to its corresponding receiver is viewed as interference by the other sender-receiver pairs. In this chapter, we limit ourselves to the two-user IFC. The study of the IFC was first initiated by Shannon [2] and was further studied by Ahlswede [3]. In [46],

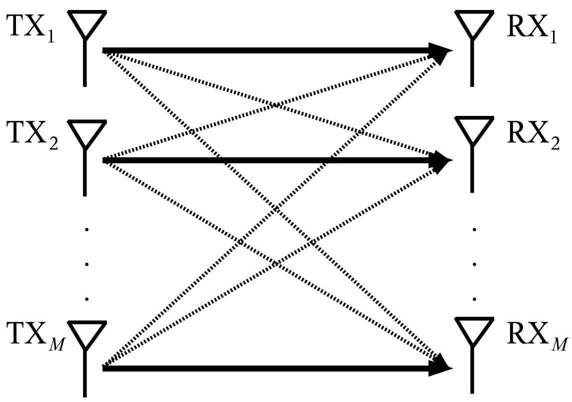

Figure 4.1: An M-user IFC

Carleial determined an improved achievable rate region for the IFC. Later, Han and Kobayashi established the best achievable rate region to date for the general IFC [15]. The capacity region of the IFC has been determined for the following cases:

- the Gaussian IFC with strong interference [15], [47], [48];
- the discrete memoryless IFC with strong interference [49];
- a class of discrete degraded IFCs [50], which includes the discrete additive degraded IFC studied by Benzel [51];
- a class of deterministic IFCs [17].

However, the capacity region of the general discrete memoryless IFC and the general Gaussian IFC remains unsolved. In this chapter, we make progress in the study of IFC by first establishing a simplified description of the Han-Kobayashi rate region for the general IFC. We recently discovered a new coding strategy for the general IFC [16], whose rate region was shown to include that of the Han-Kobayashi rate region. However, it was unknown whether the two rate regions are in fact equivalent. Using the simplified Han-Kobayashi rate region, we prove the equivalence between the two rate regions.

In addition, we make use of our simplified description to prove the capacity region of a new class of IFCs. Finally, we extend this result to the IFC with common information.

4.1.1 Outline

This chapter is organized as follows:

- In Section 4.2, we define the mathematical model for the discrete memoryless IFC and also the Gaussian IFC.
- In Section 4.3, we review the Han-Kobayashi rate region which is the best rate region to date.

- In Section 4.4, we establish the simplified description of the Han-Kobayashi rate region. We also prove the equivalence between the Han-Kobayashi rate region and the Chong-Motani-Garg rate region.
- In Section 4.6, we prove the capacity region of a new class of IFCs using our simplified description.
- Finally, we extend the proof to the same class of IFCs with common information in Section 4.6.3.

4.2 Mathematical Preliminary

A two-user discrete IFC consists of two input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a probability transition function $p(.,.|x_1, x_2)$. The conditional joint probability distribution of the discrete memoryless IFC used without feedback can be factored as

$$p_{Y_1^N Y_2^N|X_1^N X_2^N}\left(y_1^N, y_2^N|x_1^N, x_2^N\right) = \prod_{n=1}^{N} p_{Y_1Y_2|X_1X_2}\left(y_{1n}, y_{2n}|x_{1n}, x_{2n}\right).$$

Since there is no cooperation between the receivers, the capacity region of the discrete memoryless IFC depends only on the conditional marginal distributions

$$p\left(y_1|x_1, x_2\right) = \sum_{y_2 \in \mathcal{Y}_2} p\left(y_1, y_2|x_1, x_2\right), \tag{4.1}$$

$$p\left(y_2|x_1, x_2\right) = \sum_{y_1 \in \mathcal{Y}_1} p\left(y_1, y_2|x_1, x_2\right). \tag{4.2}$$

A $\left(2^{NR_1}, 2^{NR_2}, N\right)$ code for a IFC with independent information consists of two sets of integers $\mathcal{M}_1 = \left\{1, 2, ..., 2^{NR_1}\right\}$ and $\mathcal{M}_2 = \left\{1, 2, ..., 2^{NR_2}\right\}$ called the message sets, two encoding functions

$$\Psi_1 : \mathcal{M}_1 \mapsto \mathcal{X}_1^N \quad \text{and} \quad \Psi_2 : \mathcal{M}_2 \mapsto \mathcal{X}_2^N,$$

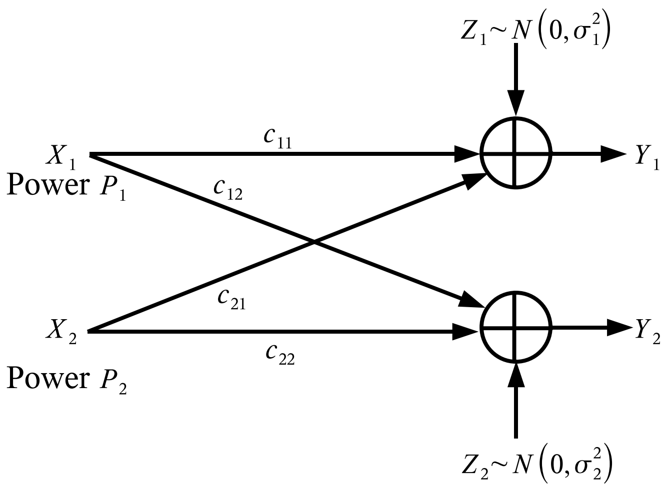

Figure 4.2: The Gaussian IFC

and two decoding functions

$$\Phi_1 : \mathcal{Y}_1^N \mapsto \mathcal{M}_1 \quad \text{and} \quad \Phi_2 : \mathcal{Y}_2^N \mapsto \mathcal{M}_2.$$

The average probability of error is defined as the probability that the decoded message is not equal to the transmitted message, i.e.,

$$P_e^{(N)} = \Pr\left(\Phi_1\left(Y_1^N\right) \neq M_1 \text{ or } \Phi_2\left(Y_2^N\right) \neq M_2\right)$$

where (M_1, M_2) are assumed to be uniformly distributed over $\mathcal{M}_1 \times \mathcal{M}_2$. A rate pair (R_1, R_2) is said to be achievable for the IFC if there exists a sequence of $\left(2^{NR_1}, 2^{NR_2}, N\right)$ codes with $P_e^{(N)} \to 0$ as $N \to \infty$.

4.2.1 Gaussian Interference Channel

The discrete-time additive white Gaussian IFC, shown in Fig. 4.2, is described by

$$Y_1 = c_{11}X_1 + c_{21}X_2 + Z_1$$

$$Y_2 = c_{12}X_1 + c_{22}X_2 + Z_2$$

where the input and output signals are real, the coefficients c_{ij} are real constants, and the noise terms Z_1 and Z_2 are zero-mean Gaussian random variables. Also, the mean value of X_1^2 and X_2^2 cannot exceed P_1 and P_2 respectively, i.e.,

$$\mathbb{E}\left[X_1^2\right] \leq P_1 \quad \text{and} \quad \mathbb{E}\left[X_2^2\right] \leq P_2.$$

In [46], it was shown that any Gaussian IFC can be reduced to a standard form by an appropriate transformation, where $c_{11}^2 = c_{22}^2 = 1$ and $\mathbb{E}\left[Z_1^2\right] = \mathbb{E}\left[Z_2^2\right] = 1$. The capacity region of the Gaussian IFC is not known, except for the case of no interference, where $c_{21}^2 = c_{12}^2 = 0$, for the case of strong interference, where $c_{21}^2 \geq 1$ and $c_{12}^2 \geq 1$, and for the one-sided Gaussian IFC under strong interference, where $c_{12}^2 = 0$ and $c_{21}^2 \geq 1$ or $c_{21}^2 = 0$ and $c_{12}^2 \geq 1$.

4.3 The Han-Kobayashi Region

In [15], Han and Kobayashi introduced 5 auxiliary random variables Q, U_1, W_1, U_2, and W_2, defined on arbitrary finite sets $\mathcal{Q}$, $\mathcal{U}_1$, $\mathcal{W}_1$, $\mathcal{U}_2$, and $\mathcal{W}_2$, respectively. In the Han-Kobayashi coding strategy, sender TX_1 splits the message M_1 into (M_{11}, M_{12}), where $\mathcal{M}_{11} = \left\{1, 2, ..., 2^{NS_1}\right\}$ and $\mathcal{M}_{12} = \left\{1, 2, ..., 2^{NT_1}\right\}$. Similarly, sender TX_2 splits the message M_2 into (M_{21}, M_{22}), where $\mathcal{M}_{21} = \left\{1, 2, ..., 2^{NT_2}\right\}$ and $\mathcal{M}_{22} = \left\{1, 2, ..., 2^{NS_2}\right\}$. This split aims at allowing each of the receivers to decode partial information from its non-intended sender. Hence, M_{12} represents the message intended for receiver RX_1 which can also be decoded by receiver RX_2, and similarly, M_{21} represents the message intended for receiver RX_2 which can also be decoded by receiver RX_1. Here, the auxiliary random variable W_1 serves to carry the message M_{12}, while the auxiliary random variable U_1 serves to carry the message M_{11}. The same applies to the auxiliary random variables

W_2 and U_2. The encoding functions Ψ_1 and Ψ_2 are then given by

$$\Psi_1 : M_1 = (M_{11}, M_{12}) \mapsto \mathcal{X}_1^n \quad \text{and} \quad \Psi_2 : M_2 = (M_{21}, M_{22}) \mapsto \mathcal{X}_2^N,$$

where the function Ψ_1 consists of three separate functions Ψ_{11}, Ψ_{12} and Ψ_{13} defined as follows:

$$\Psi_{11} : M_{11} \mapsto \mathcal{U}_1^N, \quad \Psi_{12} : M_{12} \mapsto \mathcal{W}_1^N \quad \text{and} \quad \Psi_{13} : \mathcal{U}_1^N \times \mathcal{W}_1^N \mapsto \mathcal{X}_1^N.$$

Similarly, Ψ_2 decomposes into the following three components:

$$\Psi_{21} : M_{21} \mapsto \mathcal{W}_2^N \quad \Psi_{22} : M_{22} \mapsto \mathcal{U}_2^N, \quad \text{and} \quad \Psi_{23} : \mathcal{U}_2^N \times \mathcal{W}_2^N \mapsto \mathcal{X}_2^N.$$

In a nutshell, this strategy is basically an application of Cover's superposition coding technique [5] and was first used by Carleial [46] in the context of the Gaussian IFC. Carleial made use of a sequential decoder, otherwise known as the stripping decoder. In this approach, receiver RX_1 decodes either W_1 or W_2 first before decoding U_1, whereas receiver RX_2 decodes either W_1 or W_2 first before decoding U_2. On the other hand, Han and Kobayashi uses the more powerful joint decoder where receiver RX_1 decodes W_1, W_2, and U_1 simultaneously, while receiver RX_2 decodes W_1, W_2, and U_2 simultaneously. In addition, Han and Kobayashi introduced a time-sharing parameter Q instead of using the convex-hull operation. The time-sharing parameter Q includes, as a special case, the TDM/FDM strategy introduced by Carleial [46] for the Gaussian IFC. Next, we state the achievable rate region of Han and Kobayashi, $\mathcal{R}_{\text{HK}}^{o}$, as described in [15].[1]

Let $\mathcal{P}^*$ be the set of probability distributions $P^*(.)$ that factor as

$$\begin{aligned} &P^*(q, u_1, w_1, u_2, w_2, x_1, x_2) \\ &\qquad = p(q)\, p(u_1|q)\, p(w_1|q)\, p(u_2|q)\, p(w_2|q)\, p(x_1|u_1, w_1, q)\, p(x_2|u_2, w_2, q) \end{aligned} \tag{4.3}$$

[1]We use superscript "o" and "c" to differentiate the original description of the Han-Kobayashi region from our compact description.

and where $p(x_1|u_1, w_1, q)$ and $p(x_2|u_2, w_2, q)$ equal either 0 or 1. Suppose we fix $P^*(.)$. $\mathcal{S}_{\mathrm{HK}}(P^*)$ is defined as the set of all (S_1, T_1, S_2, T_2) such that

$$S_1 \le a_{o1}, \tag{4.4}$$

$$T_1 \le b_{o1}, \tag{4.5}$$

$$T_2 \le c_{o1}, \tag{4.6}$$

$$S_1 + T_1 \le d_{o1}, \tag{4.7}$$

$$S_1 + T_2 \le e_{o1}, \tag{4.8}$$

$$T_1 + T_2 \le f_{o1}, \tag{4.9}$$

$$S_1 + T_1 + T_2 \le g_{o1}, \tag{4.10}$$

and

$$S_2 \le a_{o2}, \tag{4.11}$$

$$T_2 \le b_{o2}, \tag{4.12}$$

$$T_1 \le c_{o2}, \tag{4.13}$$

$$S_2 + T_2 \le d_{o2}, \tag{4.14}$$

$$S_2 + T_1 \le e_{o2}, \tag{4.15}$$

$$T_1 + T_2 \le f_{o2}, \tag{4.16}$$

$$S_2 + T_2 + T_1 \le g_{o2}, \tag{4.17}$$

$$-S_1, -T_1, -S_2, -T_2 \le 0, \tag{4.18}$$

where

$$a_{o1} = I(Y_1; U_1|W_1W_2Q), \tag{4.19}$$

$$b_{o1} = I(Y_1; W_1|U_1W_2Q), \tag{4.20}$$

$$c_{o1} = I(Y_1; W_2|U_1W_1Q), \tag{4.21}$$

$$d_{o1} = I(Y_1; U_1W_1|W_2Q), \tag{4.22}$$

$$e_{o1} = I(Y_1; U_1W_2|W_1Q), \tag{4.23}$$

$$f_{o1} = I\left(Y_1; W_1 W_2 | U_1 Q\right), \tag{4.24}$$

$$g_{o1} = I\left(Y_1; U_1 W_1 W_2 | Q\right), \tag{4.25}$$

and

$$a_{o2} = I\left(Y_2; U_2 | W_2 W_1 Q\right), \tag{4.26}$$

$$b_{o2} = I\left(Y_2; W_2 | U_2 W_1 Q\right), \tag{4.27}$$

$$c_{o2} = I\left(Y_2; W_1 | U_2 W_2 Q\right), \tag{4.28}$$

$$d_{o2} = I\left(Y_2; U_2 W_2 | W_1 Q\right), \tag{4.29}$$

$$e_{o2} = I\left(Y_2; U_2 W_1 | W_2 Q\right), \tag{4.30}$$

$$f_{o2} = I\left(Y_2; W_2 W_1 | U_2 Q\right), \tag{4.31}$$

$$g_{o2} = I\left(Y_2; U_2 W_2 W_1 | Q\right). \tag{4.32}$$

Let $\mathcal{R}^o_{\text{HK}}\left(P^*\right)$ be defined as the set of all (R_1, R_2) such that $0 \leq R_1 \leq S_1 + T_1$ and $0 \leq R_2 \leq S_2 + T_2$ where $(S_1, T_1, S_2, T_2) \in \mathcal{S}_{\text{HK}}\left(P^*\right)$. We have the following result:

Theorem 4.1 (Han-Kobayashi)**.** *The set*

$$\mathcal{R}^o_{\text{HK}} = \bigcup_{P^* \in \mathcal{P}^*} \mathcal{R}^o_{\text{HK}}\left(P^*\right) \tag{4.33}$$

is an achievable rate region for the discrete memoryless IFC.

Proof. Refer to [15]. □

4.4 The main result

Our main contribution is the following compact description of the Han-Kobayashi achievable rate region:

Theorem 4.2. *Let $\mathcal{P}^*_1$ be the set of probability distributions $P^*_1(.)$ that factor as*

$$P^*_1\left(q, w_1, w_2, x_1, x_2\right) = p\left(q\right) p\left(x_1, w_1 | q\right) p\left(x_2, w_2 | q\right). \tag{4.34}$$

For a fixed $P_1^* \in \mathcal{P}_1^*$*, let* $\mathcal{R}_{\text{HK}}^c(P_1^*)$ *be the set of* (R_1, R_2) *satisfying*

$$R_1 \leq I(X_1; Y_1|W_2Q) \tag{4.35}$$

$$R_2 \leq I(X_2; Y_2|W_1Q) \tag{4.36}$$

$$R_1 + R_2 \leq I(X_1W_2; Y_1|Q) + I(X_2; Y_2|W_1W_2Q) \tag{4.37}$$

$$R_1 + R_2 \leq I(X_1; Y_1|W_1W_2Q) + I(X_2W_1; Y_2|Q) \tag{4.38}$$

$$R_1 + R_2 \leq I(X_1W_2; Y_1|W_1Q) + I(X_2W_1; Y_2|W_2Q) \tag{4.39}$$

$$2R_1 + R_2 \leq I(X_1W_2; Y_1|Q) + I(X_1; Y_1|W_1W_2Q) + I(X_2W_1; Y_2|W_2Q) \tag{4.40}$$

$$R_1 + 2R_2 \leq I(X_2; Y_2|W_1W_2Q) + I(X_2W_1; Y_2|Q) + I(X_1W_2; Y_1|W_1Q) \tag{4.41}$$

Then we have

$$\mathcal{R}_{\text{HK}}^c = \bigcup_{P_1^* \in \mathcal{P}_1^*} \mathcal{R}_{\text{HK}}^c(P_1^*). \tag{4.42}$$

is an achievable rate region for the IFC. Furthermore, $\mathcal{R}_{\text{HK}}^c = \mathcal{R}_{\text{HK}}^o$ *and the region remains invariant if we impose the following constraints on the cardinalities of the auxiliary sets:*

$$\|\mathcal{W}_1\| \leq \|\mathcal{X}_1\| + 4, \quad \|\mathcal{W}_2\| \leq \|\mathcal{X}_2\| + 4 \quad \text{and} \quad \|\mathcal{Q}\| \leq 7. \tag{4.43}$$

Proof. The Han-Kobayashi rate region given in Thm. 4.1 can be reduced to Lem. 4.3 using Fourier-Motzkin elimination. It is then straightforward to see that $\mathcal{R}_{\text{HK}}^o \subseteq \mathcal{R}_{\text{HK}}^c$. In order to prove that $\mathcal{R}_{\text{HK}}^c \subseteq \mathcal{R}_{\text{HK}}^o$, we make use of Lem. 4.4. The assertion about the cardinalities of $\mathcal{W}_1$, $\mathcal{W}_2$, and $\mathcal{Q}$ follows directly from the application of Caratheodory's theorem to the expressions (4.35)-(4.41). □

Before proceeding to Lem. 4.3, we need to derive a few simple results about $\mathcal{R}_{\text{HK}}^o$. The Han-Kobayashi rate region $\mathcal{R}_{\text{HK}}^o$ was derived by assuming deterministic encoding functions rather than probabilistic functions. Hence, We can write the following:

$$
\begin{aligned}
& I\left(U_1;Y_1|W_1W_2Q\right) \\
& = H\left(Y_1|W_1W_2Q\right) - H\left(Y_1|U_1W_1W_2Q\right) \\
& = H\left(Y_1|W_1W_2Q\right) - H\left(Y_1|X_1U_1W_1W_2Q\right) \\
& = H\left(Y_1|W_1W_2Q\right) - H\left(Y_1|X_1W_1W_2Q\right) \\
& = I\left(X_1;Y_1|W_1W_2Q\right). \qquad (4.44)
\end{aligned}
$$

Following along the same lines, we can write the following equalities:

$$I\left(U_2;Y_2|W_1W_2Q\right) = I\left(X_2;Y_2|W_1W_2Q\right) \qquad (4.45)$$

$$I\left(U_1W_2;Y_1|W_1Q\right) = I\left(X_1W_2;Y_1|W_1Q\right) \qquad (4.46)$$

$$I\left(U_2W_1;Y_2|W_2Q\right) = I\left(X_2W_1;Y_2|W_2Q\right) \qquad (4.47)$$

$$I\left(U_1W_1;Y_1|W_2Q\right) = I\left(X_1;Y_1|W_2Q\right) \qquad (4.48)$$

$$I\left(U_2W_2;Y_2|W_1Q\right) = I\left(X_2;Y_2|W_1Q\right) \qquad (4.49)$$

$$I\left(U_1W_1W_2;Y_1|Q\right) = I\left(X_1W_2;Y_1|Q\right) \qquad (4.50)$$

$$I\left(U_2W_1W_2;Y_2|Q\right) = I\left(X_2W_1;Y_2|Q\right). \qquad (4.51)$$

In addition, it can be shown that for a fixed $P_1^* \in \mathcal{P}_1^*$, there exists a fixed $P^* \in \mathcal{P}^*$ such that

$$P_1^*\left(q,w_1,w_2,x_1,x_2\right) = \sum_{u_1\in\mathcal{U}_1,u_2\in\mathcal{U}_2} P^*\left(q,u_1,u_2,w_1,w_2,x_1,x_2\right). \qquad (4.52)$$

Refer to Appendix C.1. On applying the above equalities together with Fourier-Motzkin elimination, Kobayashi & Han obtained the following result:

Lemma 4.3. *(Kobayashi-Han) For a fixed $P^* \in \mathcal{P}^*$, let $\mathcal{R}^o_{\text{HK}}\left(P^*\right)$ be the set of all rate pairs $\left(R_1,R_2\right)$ satisfying*

$$R_1 \leq I\left(U_1W_1;Y_1|W_2Q\right) \qquad (4.53)$$

$$R_1 \leq I\left(U_1;Y_1|W_1W_2Q\right) + I\left(W_1;Y_2|U_2W_2Q\right) \quad (4.54)$$

$$R_2 \leq I\left(U_2W_2;Y_2|W_1Q\right) \quad (4.55)$$

$$R_2 \leq I\left(U_2;Y_2|W_1W_2Q\right) + I\left(W_2;Y_1|U_1W_1Q\right) \quad (4.56)$$

$$R_1 + R_2 \leq I\left(U_1;Y_1|W_1W_2Q\right) + I\left(U_2W_2W_1;Y_2|Q\right) \quad (4.57)$$

$$R_1 + R_2 \leq I\left(U_2;Y_2|W_1W_2Q\right) + I\left(U_1W_1W_2;Y_1|Q\right) \quad (4.58)$$

$$R_1 + R_2 \leq I\left(U_1W_2;Y_1|W_1Q\right) + I\left(U_2W_1;Y_2|W_2Q\right) \quad (4.59)$$

$$2R_1 + R_2 \leq I\left(U_1;Y_1|W_1W_2Q\right) + I\left(U_1W_1W_2;Y_1|Q\right) + I\left(U_2W_1;Y_2|W_2Q\right) \quad (4.60)$$

$$2R_1 + R_2 \leq 2I\left(U_1;Y_1|W_1W_2Q\right) + I\left(U_2W_1;Y_2|W_2Q\right) + I\left(W_1W_2;Y_2|U_2Q\right) \quad (4.61)$$

$$R_1 + 2R_2 \leq I\left(U_2;Y_2|W_1W_2Q\right) + I\left(U_2W_1W_2;Y_2|Q\right) + I\left(U_1W_2;Y_1|W_1Q\right) \quad (4.62)$$

$$R_1 + 2R_2 \leq 2I\left(U_2;Y_2|W_1W_2Q\right) + I\left(U_1W_2;Y_1|W_1Q\right) + I\left(W_1W_2;Y_1|U_1Q\right). \quad (4.63)$$

Finally, we have

$$\mathcal{R}^o_{\mathrm{HK}} = \bigcup_{P^* \in \mathcal{P}^*} \mathcal{R}^o_{\mathrm{HK}}\left(P^*\right). \quad (4.64)$$

Proof. Refer to [52, Thm. B] or to Appendix C.2. □

The equivalence between $\mathcal{R}^c_{\mathrm{HK}}$ and $\mathcal{R}^o_{\mathrm{HK}}$ emerges from the following lemma:

Lemma 4.4. *For a fixed $P_1^* \in \mathcal{P}_1^*$, there exists a fixed $P^* \in \mathcal{P}^*$ such that $\mathcal{R}^c_{\mathrm{HK}}\left(P_1^*\right) \subseteq \mathcal{R}^o_{\mathrm{HK}}\left(P^*\right) \cup \mathcal{R}^o_{\mathrm{HK}}\left(P^{**}\right) \cup \mathcal{R}^o_{\mathrm{HK}}\left(P^{***}\right)$ where*

$$P_1^*\left(q,w_1,w_2,x_1,x_2\right) = \sum_{u_1 \in \mathcal{U}_1, u_2 \in \mathcal{U}_2} P^*\left(q,u_1,u_2,w_1,w_2,x_1,x_2\right), \quad (4.65)$$

$$P^{**} = \sum_{w_1 \in \mathcal{W}_1} P^*, \quad (4.66)$$

$$P^{***} = \sum_{w_2 \in \mathcal{W}_2} P^*. \quad (4.67)$$

Proof. Suppose (R_1, R_2) is in $\mathcal{R}_{\text{HK}}^{c}(P_1^*)$ but not in $\mathcal{R}_{\text{HK}}^{o}(P^*)$. Then either (4.54), (4.56), (4.61), or (4.63) is violated. If (4.54) is violated, we have

$$\begin{aligned} R_1 &> I(W_1; Y_2|U_2W_2Q) + I(U_1; Y_1|W_1W_2Q) \\ &= I(W_1; Y_2|X_2Q) + I(X_1; Y_1|W_1W_2Q). \end{aligned} \tag{4.68}$$

If (4.61) is violated, we have from (4.57) the following inequality:

$$\begin{aligned} R_1 &> I(U_1; Y_1|W_1W_2Q) + I(U_2W_1; Y_2|W_2Q) + I(W_1W_2; Y_2|U_2Q) \\ &\quad - I(U_2W_1W_2; Y_2|Q) \\ &= I(U_1; Y_1|W_1W_2Q) + I(W_1W_2; Y_2|U_2Q) - I(W_2; Y_2|Q) \\ &= I(U_1; Y_1|W_1W_2Q) + I(W_1; Y_2|W_2U_2Q) + I(W_2; Y_2|U_2Q) - I(W_2; Y_2|Q) \\ &\geq I(U_1; Y_1|W_1W_2Q) + I(W_1; Y_2|W_2U_2Q) \\ &= I(W_1; Y_2|X_2Q) + I(X_1; Y_1|W_1W_2Q). \end{aligned} \tag{4.69}$$

Hence, (4.68) holds true if either (4.54) or (4.61) is violated. By substituting $W_1 = \Phi$ in Lem. 4.3, we see that $\mathcal{R}_{\text{HK}}^{o}(P^{**})$ consists of all rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I(X_1; Y_1|W_2Q), \\ R_2 &\leq I(X_2; Y_2|Q), \\ R_2 &\leq I(W_2; Y_1|X_1Q) + I(X_2; Y_2|W_2Q), \\ R_1 + R_2 &\leq I(X_1W_2; Y_1|Q) + I(X_2; Y_2|W_2Q). \end{aligned}$$

However, from (4.35), we obtain

$$R_1 \leq I(X_1; Y_1|W_2Q),$$

and from (4.68) and (4.38), we obtain

$$R_2 < I\left(X_2; Y_2|Q\right),$$

and from (4.68) and (4.39), we obtain

$$\begin{aligned} R_2 &< I\left(W_2; Y_1|W_1 Q\right) + I\left(X_2; Y_2|W_2 Q\right) \\ &\leq I\left(W_2; Y_1|X_1 Q\right) + I\left(X_2; Y_2|W_2 Q\right), \end{aligned}$$

and from (4.68) and (4.40), we obtain

$$R_1 + R_2 \leq I\left(X_1 W_2; Y_1|Q\right) + I\left(X_2; Y_2|W_2 Q\right).$$

We see that (R_1, R_2) satisfying the above constraints are in $\mathcal{R}^o_{\mathrm{HK}}\left(P^{**}\right)$. The proof for

$$R_2 > I\left(W_2; Y_1|X_1 Q\right) + I\left(X_2; Y_2|W_1 W_2 Q\right)$$

follows exactly along the same lines. It then follows that $\mathcal{R}^c_{\mathrm{HK}}\left(P_1^*\right) \subseteq \mathcal{R}^o_{\mathrm{HK}}\left(P^*\right) \cup \mathcal{R}^o_{\mathrm{HK}}\left(P^{**}\right) \cup \mathcal{R}^o_{\mathrm{HK}}\left(P^{***}\right)$. □

Finally, since $\mathcal{R}^c_{\mathrm{HK}}\left(P_1^*\right) \subseteq \mathcal{R}^o_{\mathrm{HK}}\left(P^*\right) \cup \mathcal{R}^o_{\mathrm{HK}}\left(P^{**}\right) \cup \mathcal{R}^o_{\mathrm{HK}}\left(P^{***}\right)$, it immediately follows that $\mathcal{R}^c_{\mathrm{HK}} \subseteq \mathcal{R}^o_{\mathrm{HK}}$ and since $\mathcal{R}^o_{\mathrm{HK}} \subseteq \mathcal{R}^c_{\mathrm{HK}}$, we obtain our result $\mathcal{R}^c_{\mathrm{HK}} = \mathcal{R}^o_{\mathrm{HK}}$.

4.5 Discussion

In this section, we make a few remarks about our results.

Remark 4.1. Han and Kobayashi made use of the polymatroidal structure underlying the collection of bounds that specify the region $\mathcal{R}^o_{\mathrm{HK}}$, (4.4)-(4.18), to convert them to a set of bounds on R_1, R_2, R_1+R_2, $2R_1+R_2$ and R_1+2R_2 [15, Thm. 4.1]. Even though Thm. 4.2 is just a different description of the Han-Kobayashi rate region, it gives the simplest description of the best rate region to date. From [15, Thm. 4.1], the cardinalities of the auxiliary sets is given by $\|\mathcal{W}_1\| \leq \|\mathcal{X}_1\| + 7$, $\|\mathcal{W}_2\| \leq \|\mathcal{X}_2\| + 7$, $\|\mathcal{U}_1\| \leq \|\mathcal{X}_1\| + 2$, $\|\mathcal{U}_2\| \leq \|\mathcal{X}_2\| + 2$, and $\|\mathcal{Q}\| \leq 11$. Hence, Thm. 4.2 also gives us tighter bounds for the cardinalities of the auxiliary sets. Another interesting observation is that even though the coding technique requires the use of the auxiliary random variables U_1 and U_2, the rate region $\mathcal{R}^c_{\mathrm{HK}}$ does not depend on these auxiliary random variables. Hence, cardinality bounds on $\mathcal{U}_1$ and $\mathcal{U}_2$ are unnecessary.

Remark 4.2. We observe that the Chong-Motani-Garg region, i.e., $\mathcal{R}_{\mathrm{CMG}}$, reported in [16], is equivalent to the Han-Kobayashi region. This equivalence sheds light on the two interesting observations behind our compact description of the Han-Kobayashi region (see Rem. 4.1). We first observe that for receiver RX_1, no decoding error is committed if the message $M_1 = (M_{11}, M_{12})$ is decoded correctly but the message M_{21} is decoded wrongly. The same applies to receiver RX_2. This implies that constraint (4.6) and (4.13) are unnecessary to drive the overall probability of error to ϵ. Moreover, the coding scheme considered in [16] uses only 3 auxiliary random variables Q, W_1, and W_2 defined on arbitrary finite sets $\mathcal{Q}$, $\mathcal{W}_1$, and $\mathcal{W}_2$. The auxiliary random variables W_1 and W_2 now serve as cloud centers that can be distinguished by both receivers. For sender TX_1, instead of generating two independent codebooks with codewords $W_1^N(j)$ and $U_1^N(k)$, for each codeword $W_1^N(j)$, we generate a codebook with codewords $X_1^N(j,k)$, where $j \in \{1, 2, ..., 2^{NT_1}\}$ and $k \in \{1, 2, ..., 2^{NS_1}\}$. This construction renders the constraints (4.5), (4.9), (4.12), and (4.16) unnecessary. Combining these two observations yields the following result:

Lemma 4.5. *Let $\mathcal{S}_{\mathrm{CMG}}(P_1^*)$ be be the set of non-negative rate-tuples (S_1, T_1, S_2, T_2) that satisfy*

$$S_1 \leq a_{o1}, \tag{4.70}$$

$$S_1 + T_1 \leq d_{o1}, \tag{4.71}$$

$$S_1 + T_2 \leq e_{o1}, \tag{4.72}$$

$$S_1 + T_1 + T_2 \leq g_{o1}, \tag{4.73}$$

and

$$S_2 \leq a_{o2}, \tag{4.74}$$

$$S_2 + T_2 \leq d_{o2}, \tag{4.75}$$

$$S_2 + T_1 \leq e_{o2}, \tag{4.76}$$

$$S_2 + T_2 + T_1 \leq g_{o2}, \tag{4.77}$$

$$-S_1, -T_1, -S_2, -T_2 \leq 0, \tag{4.78}$$

where

$$a_{o1} = I\left(Y_1; X_1 | W_1 W_2 Q\right) = I\left(Y_1; U_1 | W_1 W_2 Q\right), \tag{4.79}$$

$$d_{o1} = I\left(Y_1; X_1 | W_2 Q\right) = I\left(Y_1; U_1 W_1 | W_2 Q\right), \tag{4.80}$$

$$e_{o1} = I\left(Y_1; X_1 W_2 | W_1 Q\right) = I\left(Y_1; U_1 W_2 | W_1 Q\right), \tag{4.81}$$

$$g_{o1} = I\left(Y_1; X_1 W_2 | Q\right) = I\left(Y_1; U_1 W_1 W_2 | Q\right), \tag{4.82}$$

and

$$a_{o2} = I\left(Y_2; X_2 | W_2 W_1 Q\right) = I\left(Y_2; U_2 | W_2 W_1 Q\right), \tag{4.83}$$

$$d_{o2} = I\left(Y_2; X_2 | W_1 Q\right) = I\left(Y_2; U_2 W_2 | W_1 Q\right), \tag{4.84}$$

$$e_{o2} = I\left(Y_2; X_2 W_1 | W_2 Q\right) = I\left(Y_2; U_2 W_2 | W_1 Q\right), \tag{4.85}$$

$$g_{o2} = I\left(Y_2; X_2 W_1 | Q\right) = I\left(Y_2; U_2 W_2 W_1 | Q\right). \tag{4.86}$$

Let $\mathcal{R}_{\mathrm{CMG}}\left(P_1^*\right)$ *be defined as the set of all* (R_1, R_2) *such that* $0 \leq R_1 \leq S_1 + T_1$ *and* $0 \leq R_2 \leq S_2 + T_2$ *where* $(S_1, T_1, S_2, T_2) \in \mathcal{S}_{\mathrm{CMG}}\left(P_1^*\right)$*. Then, the set given by*

$$\mathcal{R}_{\mathrm{CMG}} = \bigcup_{P_1^* \in \mathcal{P}_1^*} \mathcal{R}_{\mathrm{CMG}}\left(P_1^*\right) \tag{4.87}$$

is an achievable rate region for the discrete memoryless IFC.

Proof. Refer to Appendix C.4. □

We can see that $\mathcal{R}_{\text{CMG}} = \mathcal{R}^{c}_{\text{HK}}$ through the following simple argument. First, since we can choose a fixed P_1^* such that

$$P_1^* \left(q, w_1, w_2, x_1, x_2\right) = \sum_{u_1 \in \mathcal{U}_1, u_2 \in \mathcal{U}_2} P^* \left(q, u_1, u_2, w_1, w_2, x_1, x_2\right). \tag{4.88}$$

We readily see that $\mathcal{R}^{o}_{\text{HK}} \left(P^*\right) \subseteq \mathcal{R}_{\text{CMG}} \left(P_1^*\right)$, and hence $\mathcal{R}^{o}_{\text{HK}} \subseteq \mathcal{R}_{\text{CMG}}$. The bounds (4.70)-(4.78) can be again be simplified using Fourier-Motzkin elimination to obtain the following result:

Lemma 4.6. *(Han-Kobayashi) For a fixed $P_1^* \in \mathcal{P}_1^*$, let $\mathcal{R}_{\text{CMG}} \left(P_1^*\right)$ be the set of $\left(R_1, R_2\right)$ satisfying*

$$R_1 \leq I\left(X_1; Y_1 | W_2 Q\right) \tag{4.89}$$
$$R_1 \leq I\left(X_1; Y_1 | W_1 W_2 Q\right) + I\left(X_2 W_1; Y_2 | W_2 Q\right) \tag{4.90}$$
$$R_2 \leq I\left(X_2; Y_2 | W_1 Q\right) \tag{4.91}$$
$$R_2 \leq I\left(X_2; Y_2 | W_1 W_2 Q\right) + I\left(X_1 W_2; Y_1 | W_1 Q\right) \tag{4.92}$$
$$R_1 + R_2 \leq I\left(X_1 W_2; Y_1 | Q\right) + I\left(X_2; Y_2 | W_1 W_2 Q\right) \tag{4.93}$$
$$R_1 + R_2 \leq I\left(X_1; Y_1 | W_1 W_2 Q\right) + I\left(X_2 W_1; Y_2 | Q\right) \tag{4.94}$$
$$R_1 + R_2 \leq I\left(X_1 W_2; Y_1 | W_1 Q\right) + I\left(X_2 W_1; Y_2 | W_2 Q\right) \tag{4.95}$$
$$2R_1 + R_2 \leq I\left(X_1 W_2; Y_1 | Q\right) + I\left(X_1; Y_1 | W_1 W_2 Q\right) + I\left(X_2 W_1; Y_2 | W_2 Q\right) \tag{4.96}$$
$$R_1 + 2R_2 \leq I\left(X_2; Y_2 | W_1 W_2 Q\right) + I\left(X_2 W_1; Y_2 | Q\right) + I\left(X_1 W_2; Y_1 | W_1 Q\right). \tag{4.97}$$

Then we have

$$\mathcal{R}_{\text{CMG}} = \bigcup_{P_1^* \in \mathcal{P}_1^*} \mathcal{R}_{\text{CMG}} \left(P_1^*\right). \tag{4.98}$$

is an achievable rate region for the IFC.

Proof. Refer to [52, Thm. D] or to Appendix C.3. □

One can readily see that $\mathcal{R}_{\text{CMG}} \subseteq \mathcal{R}^{c}_{\text{HK}}$ since the Chong-Motani-Garg rate region for the general IFC has two additional constraints. Hence, we see that the two

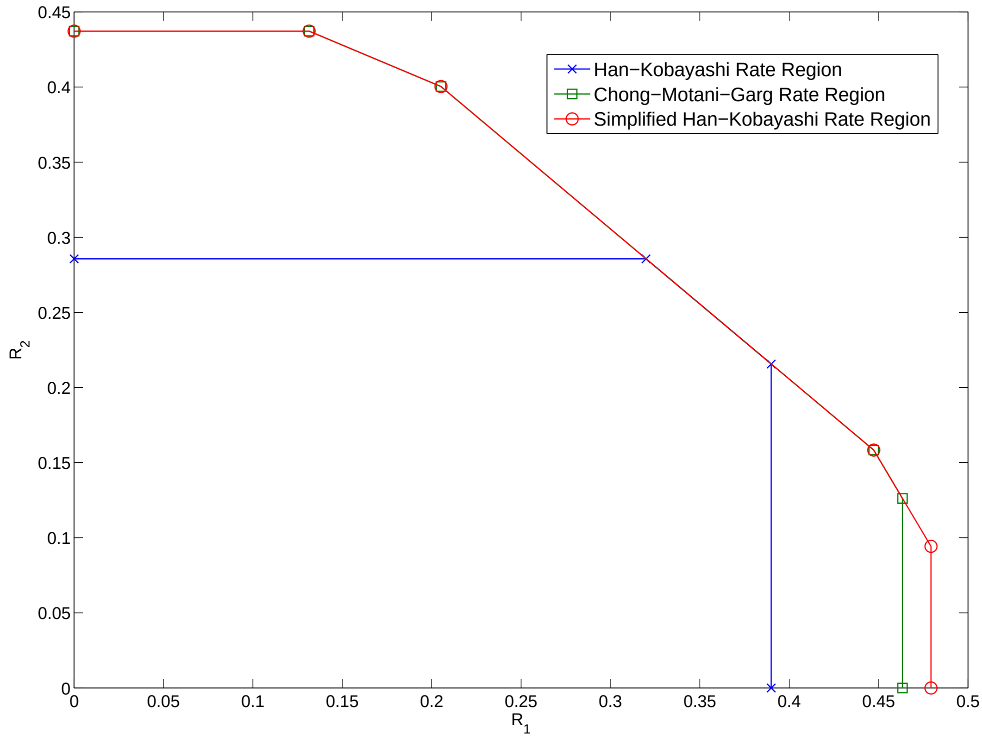

Figure 4.3: An example where $\mathcal{R}^o_{\mathrm{HK}}(P^*) \subsetneq \mathcal{R}_{\mathrm{CMG}}(P^*_1) \subsetneq \mathcal{R}^c_{\mathrm{HK}}(P^*_1)$

rate regions are equivalent, i.e., $\mathcal{R}_{\mathrm{CMG}} = \mathcal{R}^c_{\mathrm{HK}}$.

Remark 4.3. We note that the only differences between $\mathcal{R}^c_{\mathrm{HK}}(P^*_1)$, $\mathcal{R}_{\mathrm{CMG}}(P^*_1)$, and $\mathcal{R}^o_{\mathrm{HK}}(P^*)$ lie only in the bounds for R_1 and R_2. This observation allows for answering the question posed by Kramer in [53] on the existence of $P^* \in \mathcal{P}^*$ such that $\mathcal{R}^o_{\mathrm{HK}}(P^*) \subsetneq \mathcal{R}_{\mathrm{CMG}}(P^*_1)$ for certain IFCs where

$$P^*_1(q, w_1, w_2, x_1, x_2) = \sum_{u_1 \in \mathcal{U}_1, u_2 \in \mathcal{U}_2} P^*(q, u_1, u_2, w_1, w_2, x_1, x_2). \tag{4.99}$$

For the Gaussian IFC, when we set $|\mathcal{Q}| = 1$, we can easily determine parameters where $\mathcal{R}^o_{\mathrm{HK}}(P^*) \subsetneq \mathcal{R}_{\mathrm{CMG}}(P^*_1)$. (The Han-Kobayashi rate region can be directly applied to the Gaussian IFC as it was proven using only weak typicality.) We assume the following customary restriction on the input signals where W_1, W_2,

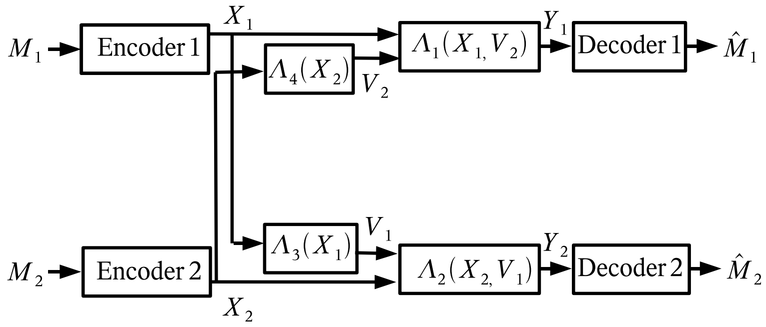

Figure 4.4: The class of deterministic IFC studied by El Gamal and Costa

X_1, and X_2 are Gaussian random variables and

$$\frac{\mathbb{E}\left[W_1^2\right]}{\mathbb{E}\left[X_1^2\right]} = \alpha, \qquad \frac{\mathbb{E}\left[W_2^2\right]}{\mathbb{E}\left[X_2^2\right]} = \beta \tag{4.100}$$

such that $\alpha \in [0,1]$, $\beta \in [0,1]$, $\mathbb{E}\left[X_1^2\right] = P_1$, and $\mathbb{E}\left[X_2^2\right] = P_2$. From Fig. 4.3, when we set $P_1 = P_2 = 1$, $c_{12}^2 = c_{21}^2 = 0.4$, and $\alpha = 0.5$ and $\beta = 0.85$, $\mathcal{R}_{\text{HK}}^{o}\left(P^*\right) \subsetneq \mathcal{R}_{\text{CMG}}\left(P_1^*\right) \subsetneq \mathcal{R}_{\text{HK}}^{c}\left(P_1^*\right)$.

It is interesting to note that there exists fixed distributions satisfying (4.99) where $\mathcal{R}_{\text{HK}}^{o}\left(P^*\right) \subsetneq \mathcal{R}_{\text{CMG}}\left(P_1^*\right) \subsetneq \mathcal{R}_{\text{HK}}^{c}\left(P_1^*\right)$. However, when maximized over all possible distributions, all three descriptions are equivalent, i.e., they describe the same rate region.

4.6 Capacity region of a class of deterministic IFC

We first consider a class of deterministic IFCs (without common information) as shown in Fig. 4.4. The outputs Y_1 and Y_2, and the interferences V_1 and V_2 are deterministic functions of the inputs X_1 and X_2:

$$Y_1 = \Lambda_1\left(X_1, V_2\right), \tag{4.101}$$

$$Y_2 = \Lambda_2\left(V_1, X_2\right), \tag{4.102}$$

$$V_1 = \Lambda_3(X_1), \tag{4.103}$$

$$V_2 = \Lambda_4(X_2). \tag{4.104}$$

In addition, for this class of deterministic IFCs, Y_1 and X_1 must uniquely determine V_2, while Y_2 and X_2 must uniquely determine V_1. Hence, there exist functions Λ_5 and Λ_6 such that we have

$$V_1 = \Lambda_5(X_2, Y_2), \tag{4.105}$$

$$V_2 = \Lambda_6(X_1, Y_1). \tag{4.106}$$

El Gamal and Costa determined the capacity region of the channel of Fig. 4.4 satisfying (4.105) and (4.106) in [17, Thm. 1]. The achievability follows directly from the Han-Kobayashi rate region [15]. It was noted by Kramer [53] that the capacity region of this class of IFCs bears an uncanny resemblance in form to an achievable rate region determined by the authors for the general IFC [22]. It was also established in the previous section that the Han-Kobayashi rate region was in fact equivalent to that of the Chong-Motani-Garg rate region. Even though no new achievable rate region was proven for the IFC, the simplified Han-Kobayashi rate region makes it easier to prove the capacity region for a wider class of IFCs.

For the class of IFCs considered in this chapter, we relax the constraint that the outputs Y_1 and Y_2 be deterministic functions of the inputs X_1 and X_2. Hence, we remove conditions (4.101) and (4.102) imposed by El Gamal and Costa. In addition, we relax the constraints (4.105) and (4.106) to include the case of strong interference.

4.6.1 Channel Model

Consider the class of IFCs shown in Fig. 4.5. The channel itself consists of four finite alphabets $\mathcal{X}_1 = \{1, 2, ..., \|\mathcal{X}_1\|\}$, $\mathcal{X}_2 = \{1, 2, ..., \|\mathcal{X}_2\|\}$, $\mathcal{Y}_1 = \{1, 2, ..., \|\mathcal{Y}_1\|\}$, and $\mathcal{Y}_2 = \{1, 2, ..., \|\mathcal{Y}_2\|\}$, two deterministic functions in agreement with (4.103)

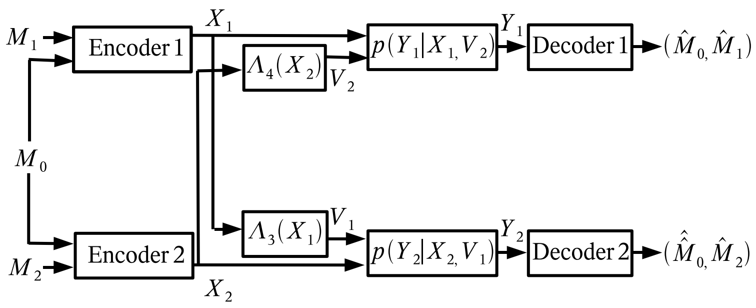

Figure 4.5: The class of IFCs under investigation

and (4.104), and two conditional marginal distributions given by

$$p_1(y_1|x_1v_2) = \sum_{y_2 \in \mathcal{Y}_2} p(y_1y_2|x_1x_2) \tag{4.107}$$

$$p_2(y_2|v_1x_2) = \sum_{y_1 \in \mathcal{Y}_1} p(y_1y_2|x_1x_2). \tag{4.108}$$

Since there is no cooperation between the receivers, the capacity region of the IFC depends only on the conditional marginal distributions. We assume that this channel is memoryless. A $\left(2^{NR_0}, 2^{NR_1}, 2^{NR_2}, N\right)$ code for this channel consists of two encoders

$$\Psi_1 : \left\{1, ..., 2^{NR_0}\right\} \times \left\{1, ..., 2^{NR_1}\right\} \to \mathcal{X}_1^N$$
$$\Psi_2 : \left\{1, ..., 2^{NR_0}\right\} \times \left\{1, ..., 2^{NR_2}\right\} \to \mathcal{X}_2^N$$

and two decoding functions

$$\Phi_1 : \mathcal{Y}_1^N \to \left\{1, ..., 2^{NR_0}\right\} \times \left\{1, ..., 2^{NR_1}\right\}$$
$$\Phi_2 : \mathcal{Y}_2^N \to \left\{1, ..., 2^{NR_0}\right\} \times \left\{1, ..., 2^{NR_2}\right\}.$$

The average probability of error is defined as the probability the decoded message is not equal to the transmitted message, i.e.,

$$P_e^{(N)} = \Pr\left(\Phi_1\left(Y_1^N\right) \neq (M_0, M_1) \text{ or } \Phi_2\left(Y_2^N\right) \neq (M_0, M_2)\right) \tag{4.109}$$

where (M_0, M_1, M_2) is assumed to be uniformly distributed over $\left\{1, 2, ..., 2^{NR_0}\right\} \times \left\{1, 2, ..., 2^{NR_1}\right\} \times \left\{1, 2, ..., 2^{NR_2}\right\}$. A rate triplet (R_0, R_1, R_2) is said to be achievable for the IFC if there exists a sequence of $\left(2^{NR_0}, 2^{NR_1}, 2^{NR_2}, N\right)$ codes with $P_e^{(N)} \to 0$ as $N \to \infty$.

We will first take a look at the capacity region of this class of IFCs without any common information ($\mathcal{M}_0 = \phi$), which is the more commonly studied case, before extending the proof to the case with common information. In addition, we require that the following two conditions:

$$I\left(V_1^N; Y_2^N | X_2^N\right) \geq I\left(V_1^N; Y_1^N | X_2^N\right) \tag{4.110}$$

$$I\left(V_2^N; Y_1^N | X_1^N\right) \geq I\left(V_2^N; Y_2^N | X_1^N\right) \tag{4.111}$$

are satisfied for all product distributions $p\left(x_1^N\right) p\left(x_2^N\right)$ on $\mathcal{X}_1^N \times \mathcal{X}_2^N$. Even though (4.110) and (4.111) are block level constraints, we will show later on that there exist single letter constraints that imply (4.110) and (4.111).

4.6.2 Deterministic IFC Without Common Information

Theorem 4.7. *The capacity region of the IFC shown in Fig. 4.5, without any common information, satisfying conditions (4.110) and (4.111) is the union of all rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leq I\left(X_1; Y_1 | V_2 Q\right) \tag{4.112}$$

$$R_2 \leq I\left(X_2; Y_2 | V_1 Q\right) \tag{4.113}$$

$$R_1 + R_2 \leq I\left(X_1 V_2; Y_1 | Q\right) + I\left(X_2; Y_2 | V_1 V_2 Q\right) \tag{4.114}$$

$$R_1 + R_2 \leq I\left(X_1; Y_1 | V_1 V_2 Q\right) + I\left(X_2 V_1; Y_2 | Q\right) \tag{4.115}$$

$$R_1 + R_2 \leq I\left(X_1V_2;Y_1|V_1Q\right) + I\left(X_2V_1;Y_2|V_2Q\right) \tag{4.116}$$

$$2R_1 + R_2 \leq I\left(X_1V_2;Y_1|Q\right) + I\left(X_1;Y_1|V_1V_2Q\right) + I\left(X_2V_1;Y_2|V_2Q\right) \tag{4.117}$$

$$R_1 + 2R_2 \leq I\left(X_2;Y_2|V_1V_2Q\right) + I\left(X_2V_1;Y_2|Q\right) + I\left(X_1V_2;Y_1|V_1Q\right). \tag{4.118}$$

for all input distributions $p\left(q\right)p\left(x_1|q\right)p\left(x_2|q\right)$. Furthermore, the region remains invariant if we impose the following constraint: $\|\mathcal{Q}\| \leq 7$.

We first give a few examples of IFCs for which Thm. 4.7 gives the capacity before going on the proof. These include the result in [49] and two new channels.

- Discrete memoryless IFC with strong interference: For this class of IFCs, $V_1 \triangleq X_1$ and $V_2 \triangleq X_2$. From [49], we know that if $I\left(X_1;Y_2|X_2\right) \geq I\left(X_1;Y_1|X_2\right)$ and $I\left(X_2;Y_1|X_1\right) \geq I\left(X_2;Y_2|X_1\right)$ for all product distributions on $\mathcal{X}_1 \times \mathcal{X}_2$, conditions (4.110) and (4.111) will be satisfied for all product distributions on $\mathcal{X}_1^n \times \mathcal{X}_2^n$.

- A class of deterministic IFC: If there exists functions Λ_5 and Λ_6 such that (4.105) and (4.106) are satisfied, we see that conditions (4.110) and (4.111) will be satisfied for all product distributions on $\mathcal{X}_1^n \times \mathcal{X}_2^n$. This class of channels includes the class of deterministic IFCs determined by El Gamal and Costa, but without the condition that Y_1 and Y_2 be deterministic functions of (X_1, V_2) and (V_1, X_2), respectively.

Example 4.1. We consider a symmetric, deterministic IFC with the following alphabets: $\mathcal{X}_1 = \{0, 1, 2\}$, $\mathcal{V}_1 = \{0, 1\}$, $\mathcal{Y}_1 = \{0, 1, 2, 3\}$, $\mathcal{X}_2 = \{0, 1, 2\}$, $\mathcal{V}_2 = \{0, 1\}$ and $\mathcal{Y}_2 = \{0, 1, 2, 3\}$. The functions Λ_3 and Λ_4 are given by

$$\Lambda_3\left(X_1 = 0\right) = 0; \quad \Lambda_3\left(X_1 = 1\right) = \Lambda_3\left(X_1 = 2\right) = 1, \tag{4.119}$$

$$\Lambda_4\left(X_2 = 0\right) = 0; \quad \Lambda_4\left(X_2 = 1\right) = \Lambda_4\left(X_2 = 2\right) = 1. \tag{4.120}$$

Consider the transition probability matrices shown in Table 4.1. One can easily check that it is not the class of deterministic IFC studied by El Gamal and Costa since $Y_1 \neq \Lambda_1(X_1, V_2)$ and $Y_2 \neq \Lambda_2(X_2, V_1)$. One can also easily verify that it is not a discrete memoryless IFC under strong interference. Set $p(X_1 = 1) = p(X_2 = 2) = \frac{1}{2}$, $p(X_2 = 0) = \frac{1}{2}$, $p(X_2 = 1) = \frac{1}{4}$, $p(X_2 = 2) = \frac{1}{4}$, $p_1 = \frac{1}{10}$, and $p_2 = \frac{1}{10}$. We then have $I(X_1; Y_2|X_2) = 0$ and $I(X_1; Y_1|X_2) = \frac{1}{2}$. Hence, $I(X_1; Y_2|X_2) \geq I(X_1; Y_1|X_2)$ does not hold for all product distributions $p(X_1)\, p(X_2)$.

Table 4.1: Transition probability matrices

$p(Y_1\|X_1, V_2 = 0)$			
	$X_1 = 0$	$X_1 = 1$	$X_1 = 2$
$Y_1 = 0$	p	0	0
$Y_1 = 1$	$1 - p$	0	0
$Y_1 = 2$	0	0	0
$Y_1 = 3$	0	1	1
$p(Y_1\|X_1, V_2 = 1)$			
	$X_1 = 0$	$X_1 = 1$	$X_1 = 2$
$Y_1 = 0$	0	0	1
$Y_1 = 1$	0	p_1	0
$Y_1 = 2$	1	$1 - p_1$	0
$Y_1 = 3$	0	0	0
$p(Y_2\|X_2, V_1 = 0)$			
	$X_2 = 0$	$X_2 = 1$	$X_2 = 2$
$Y_2 = 0$	p	0	0
$Y_2 = 1$	$1 - p$	0	0
$Y_2 = 2$	0	0	0
$Y_2 = 3$	0	1	1
$p(Y_2\|X_2, V_1 = 1)$			
	$X_2 = 0$	$X_2 = 1$	$X_2 = 2$
$Y_2 = 0$	0	0	1
$Y_2 = 1$	0	p_1	0
$Y_2 = 2$	1	$1 - p_1$	0
$Y_2 = 3$	0	0	0

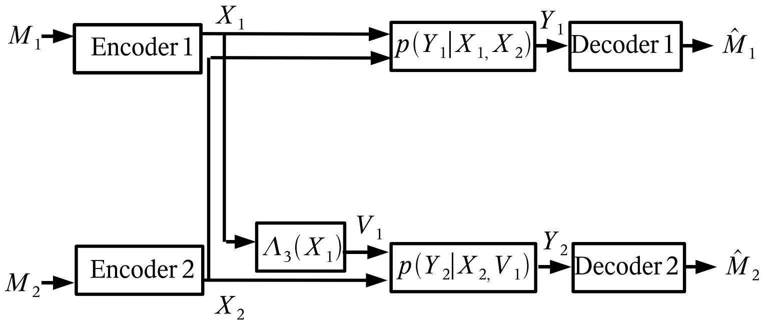

Figure 4.6: Asymmetric IFC

- A semi-deterministic IFC with strong interference: If the inequality $I(X_2; Y_1|X_1) \geq I(X_2; Y_2|X_1)$ is satisfied for all product distributions on $\mathcal{X}_1 \times \mathcal{X}_2$ (hence $V_2 \triangleq X_2$), and there exists a function Λ_5 such that (4.105) is satisfied, we readily see that conditions (4.110) and (4.111) will be satisfied for all product distributions on $\mathcal{X}_1^n \times \mathcal{X}_2^n$. This class of IFCs is a mixture of the IFC with strong interference and the class of deterministic IFCs introduced by El Gamal and Costa and is as shown in Fig. 4.6.

Proof. 1) Achievability: This follows directly from the simplified Han-Kobayashi rate region [22, Thm. 1] with $V_1 \triangleq W_1$ and $V_2 \triangleq W_2$. The assertion about the cardinality of $\|\mathcal{Q}\|$ follows directly from the application of Caratheodory's theorem to the expressions (4.112)-(4.118).

2) Converse: From Fano's inequalities, we obtain

$$H(M_1|Y_1^n) \leq N\epsilon_{1N} \text{ and } H\left(M_2|Y_2^N\right) \leq N\epsilon_{2N}. \tag{4.121}$$

We will make use of the following facts: (a) The data processing inequality. (b) The independence of $\left(X_1^N, V_1^N\right)$ and $\left(X_2^N, V_2^N\right)$. (c) The fact that Y_{1n} depends only on (X_{1n}, V_{2n}) and Y_{2n} depends only on (X_{2n}, V_{1n}). (d) Conditions (4.103) and (4.104). (e) Conditions (4.110) and (4.111). (f) Conditioning reduces entropy.

Let us first consider

$$
\begin{aligned}
NR_1 = H(M_1) &\le I\left(M_1; Y_1^N\right) + N\epsilon_{1N} \\
&\overset{(a)}{\le} I\left(X_1^N; Y_1^N\right) + N\epsilon_{1N} \\
&\overset{(b)}{\le} I\left(X_1^N; Y_1^N | V_2^N\right) + N\epsilon_{1N} \\
&\overset{(c)}{=} \sum_{n=1}^{N} \left[H\left(Y_{1n} | V_2^N Y_1^{n-1}\right) - H\left(Y_{1n} | V_{2n} X_{1n}\right)\right] + N\epsilon_{1N} \\
&\overset{(f)}{\le} \sum_{n=1}^{N} I\left(X_{1n}; Y_{1n} | V_{2n}\right) + N\epsilon_{1N}.
\end{aligned}
\tag{4.122}
$$

Analogously, we may derive an expression for R_2 similar in form to (4.113). Next, let us consider

$$
\begin{aligned}
&N(R_1 + R_2) \\
&\quad = H(M_1) + H(M_2) \\
&\quad \le I\left(M_1; Y_1^N\right) + I\left(M_2; Y_2^N\right) + N\left(\epsilon_{1N} + \epsilon_{2N}\right) \\
&\quad \overset{(a)(b)}{\le} I\left(X_1^N; Y_1^N\right) + I\left(X_2^N; Y_2^N | X_1^N\right) + N\left(\epsilon_{1N} + \epsilon_{2N}\right) \\
&\quad \overset{(d)}{=} I\left(X_1^N; Y_1^N\right) + I\left(V_2^N; Y_2^N | X_1^N\right) + I\left(X_2^N; Y_2^N | X_1^N V_2^N\right) + N\left(\epsilon_{1N} + \epsilon_{2N}\right) \\
&\quad \overset{(e)}{\le} I\left(X_1^N; Y_1^N\right) + I\left(V_2^N; Y_1^N | X_1^N\right) + I\left(X_2^N; Y_2^N | X_1^N V_2^N\right) + N\left(\epsilon_{1N} + \epsilon_{2N}\right) \\
&\quad \overset{(d)}{=} I\left(V_2^N X_1^N; Y_1^N\right) + I\left(X_2^N; Y_2^N | X_1^N V_1^N V_2^N\right) + N\left(\epsilon_{1N} + \epsilon_{2N}\right) \\
&\quad \overset{(c)(f)}{\le} \sum_{n=1}^{N} \left[I\left(V_{2n} X_{1n}; Y_{1n}\right) + I\left(X_{2n}; Y_{2n} | V_{1n} V_{2n}\right)\right] + N\left(\epsilon_{1N} + \epsilon_{2N}\right).
\end{aligned}
\tag{4.123}
$$

Analogously, we may derive an expression for $R_1 + R_2$ similar in form to (4.115). Next, let us also consider

$$
\begin{aligned}
N(R_1 + R_2) &= H(M_1) + H(M_2) \\
&\le I\left(M_1; Y_1^N\right) + I\left(M_2; Y_2^N\right) + N\left(\epsilon_{1N} + \epsilon_{2N}\right) \\
&\overset{(a)}{\le} I\left(X_1^N; Y_1^N\right) + I\left(X_2^N; Y_2^N\right) + N\left(\epsilon_{1N} + \epsilon_{2N}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(d)}{=} I\left(V_1^N;Y_1^N\right)+I\left(X_1^N;Y_1^N|V_1^N\right)+I\left(V_2^n;Y_2^n\right)\\
&\quad +I\left(X_2^N;Y_2^N|V_2^N\right)+N\left(\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(b)}{\leq} I\left(V_1^N;Y_1^N|X_2^N\right)+I\left(X_1^N;Y_1^N|V_1^N\right)+I\left(V_2^N;Y_2^N|X_1^N\right)\\
&\quad +I\left(X_2^N;Y_2^N|V_2^N\right)+N\left(\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(e)}{\leq} I\left(V_1^N;Y_2^N|X_2^N\right)+I\left(X_1^N;Y_1^N|V_1^N\right)+I\left(V_2^N;Y_1^N|X_1^N\right)\\
&\quad +I\left(X_2^N;Y_2^N|V_2^N\right)+N\left(\epsilon_{1N}+\epsilon_{2N}\right)\\
&= I\left(V_2^N X_1^N;Y_1^N|V_1^N\right)+I\left(V_1^N X_2^N;Y_2^N|V_2^N\right)+N\left(\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(c)(f)}{\leq} \sum_{n=1}^{N}\left[I\left(X_{1n}V_{2n};Y_{1n}|V_{1n}\right)+I\left(X_{2n}V_{1n};Y_{2n}|V_{2n}\right)\right]+N\left(\epsilon_{1N}+\epsilon_{2N}\right). \qquad (4.124)
\end{aligned}
$$

Finally, let us consider

$$
\begin{aligned}
&N\left(2R_1+R_2\right)=2H\left(M_1\right)+H\left(M_2\right)\\
&\leq 2I\left(M_1;Y_1^N\right)+I\left(M_2;Y_2^N\right)+N\left(2\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(a)(b)}{\leq} I\left(X_1^N;Y_1^N\right)+I\left(X_1^N;Y_1^N|X_2^N\right)+I\left(X_2^N;Y_2^N\right)+N\left(2\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(d)}{=} I\left(X_1^N;Y_1^N\right)+I\left(V_1^N;Y_1^N|X_2^N\right)+I\left(X_1^N;Y_1^N|X_2^N V_1^N\right)+I\left(V_2^N;Y_2^N\right)\\
&\quad +I\left(X_2^N;Y_2^N|V_2^N\right)+N\left(2\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(b)(e)}{\leq} I\left(X_1^N;Y_1^N\right)+I\left(V_1^N;Y_2^N|X_2^N\right)+I\left(X_1^N;Y_1^N|X_2^N V_1^N\right)+I\left(V_2^N;Y_1^N|X_1^N\right)\\
&\quad +I\left(X_2^N;Y_2^N|V_2^N\right)+N\left(2\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(d)}{=} I\left(V_2^N X_1^N;Y_1^N\right)+I\left(X_2^N V_1^N;Y_2^N|V_2^N\right)+I\left(X_1^N;Y_1^N|X_2^N V_1^N V_2^N\right)\\
&\quad +N\left(2\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(c)(f)}{\leq} \sum_{n=1}^{N}\left[I\left(V_{2n}X_{1n};Y_{1n}\right)+I\left(X_{2n}V_{1n};Y_{2n}|V_{2n}\right)+I\left(X_{1n};Y_{1n}|V_{1n}V_{2n}\right)\right]\\
&\quad +N\left(2\epsilon_{1N}+\epsilon_{2N}\right). \qquad (4.125)
\end{aligned}
$$

Analogously, we may derive an expression for R_1+2R_2 similar in form to (4.118).

Finally, we obtain conditions (4.112)-(4.118) by introducing a time-sharing random variable Q and allowing $N \to \infty$. □

4.6.3 Deterministic IFC with Common Information

Recently, there has also been some research activity into the IFC with common information. In this setting, both transmitters have a common message, in addition to its own private messages, to transmit to both receivers. Maric, Yates, and Kramer [54] established the capacity of the strong IFC with common information. Following this, Jiang, Xin, and Garg [18] determined an achievable rate region for the general IFC with common information. They also established the capacity region of a class of deterministic IFCs, introduced by El Gamal and Costa [17], with common information.

Hence, we next consider the case where both transmitters have common information M_0 which they want to transmit to both receivers. In addition, we require that the following two conditions:

$$I\left(V_1^N; Y_2^N | X_2^N M_0\right) \geq I\left(V_1^N; Y_1^N | X_2^N M_0\right) \tag{4.126}$$

$$I\left(V_2^N; Y_1^N | X_1^N M_0\right) \geq I\left(V_2^N; Y_2^N | X_1^N M_0\right) \tag{4.127}$$

are satisfied for all input probability distributions of the form $p\left(M_0 X_1^N X_2^N\right) = p\left(M_0\right) p\left(X_1^N | M_0\right) p\left(X_2^N | M_0\right)$.

Theorem 4.8. *The capacity region of the IFC shown in Fig. 4.5 satisfying the conditions (4.126) and (4.127) is the union of all rate triplets* (R_0, R_1, R_2) *satisfying*

$$R_1 \leq I\left(X_1; Y_1 | V_0 V_2\right) \tag{4.128}$$

$$R_2 \leq I\left(X_2; Y_2 | V_0 V_1\right) \tag{4.129}$$

$$R_1 + R_2 \leq I\left(X_1 V_2; Y_1 | V_0\right) + I\left(X_2; Y_2 | V_0 V_1 V_2\right) \tag{4.130}$$

$$R_1 + R_2 \leq I\left(X_1; Y_1 | V_0 V_1 V_2\right) + I\left(X_2 V_1; Y_2 | V_0\right) \tag{4.131}$$

$$R_1 + R_2 \leq I(X_1V_2; Y_1|V_0V_1) + I(X_2V_1; Y_2|V_0V_2) \quad (4.132)$$

$$2R_1 + R_2 \leq I(X_1V_2; Y_1|V_0) + I(X_1; Y_1|V_0V_1V_2) + I(X_2V_1; Y_2|V_0V_2) \quad (4.133)$$

$$R_1 + 2R_2 \leq I(X_2; Y_2|V_0V_1V_2) + I(X_2V_1; Y_2|V_0) + I(X_1V_2; Y_1|V_0V_1) \quad (4.134)$$

$$R_0 + R_1 \leq I(V_2X_1; Y_1) \quad (4.135)$$

$$R_0 + R_2 \leq I(V_1X_2; Y_2) \quad (4.136)$$

$$R_0 + R_1 + R_2 \leq I(V_2X_1; Y_1) + I(X_2; Y_2|V_0V_1V_2) \quad (4.137)$$

$$R_0 + R_1 + R_2 \leq I(X_1; Y_1|V_0V_1V_2) + I(V_1X_2; Y_2) \quad (4.138)$$

$$R_0 + 2R_1 + R_2 \leq I(V_2X_1; Y_1) + I(X_1; Y_1|V_1V_2V_0) + I(V_1X_2; Y_2|V_2V_0) \quad (4.139)$$

$$R_0 + R_1 + 2R_2 \leq I(X_2; Y_2|V_1V_2V_0) + I(V_1X_2; Y_2) + I(X_1V_2; Y_1|V_1V_0) \quad (4.140)$$

for all input distributions $p(v_0)\,p(x_1|v_0)\,p(x_2|v_0)$. *Furthermore, the region remains invariant if we impose the following constraint:* $\|\mathcal{V}_0\| \leq \|\mathcal{X}_1\|\,\|\mathcal{X}_2\| + 7$.

Proof. 1) Achievability: This follows by applying Fourier-Motzkin elimination to the conditions of [18, Thm. 1]. Refer to Appendix C.5. We then substitute $V_0 \triangleq W_0$, $V_1 \triangleq W_1$, and $V_2 \triangleq W_2$ in Thm. C.1. Finally, we note that $R_1 \leq I(X_1; Y_1|V_0V_1V_2) + I(V_1X_2; Y_2|V_0V_2)$ is redundant due to our imposed constraint (4.126) as follows:

$$\begin{aligned} I(X_1; Y_1|V_0V_1V_2) + I(V_1X_2; Y_2|V_0V_2) &\geq I(X_1; Y_1|V_0V_1V_2) + I(V_1; Y_2|V_0X_2) \\ &\geq I(X_1; Y_1|V_0V_1V_2) + I(V_1; Y_1|V_0X_2) \\ &= I(X_1; Y_1|V_0V_2). \end{aligned} \quad (4.141)$$

Similarly, $R_2 \leq I(X_2; Y_2|V_0V_1V_2) + I(V_2X_1; Y_1|V_0V_1)$ is redundant due to our imposed constraint (4.127). The assertion about the cardinality of $\|\mathcal{V}_0\|$ follows directly from the application of Caratheodory's theorem to (4.128)-(4.140).

2) Converse: The converse proof of Thm. 4.8 follow closely the converse proof

of Thm. 4.7. We will make use of the following facts: (a) The independence of M_0, M_1, and M_2. (b) The conditional independence of $\left(M_1, V_1^N, X_1^N\right)$ and $\left(M_2, V_2^N, X_2^N\right)$ given M_0. (c) The Markov chains $M_1 \to M_0 X_1^N \to Y_1^N$ and $M_2 \to M_0 X_2^N \to Y_2^N$. (d) The fact that Y_{1n} depends only on (X_{1n}, V_{2n}) and Y_{2n} depends only on (V_{1n}, X_{2n}). (e) Conditions (4.126) and (4.127). (f) Conditioning reduces entropy. In addition, from Fano's inequalities, we obtain $H\left(M_0|Y_1^N\right) \leq N\epsilon_{3N}$ and $H\left(M_0|Y_2^N\right) \leq N\epsilon_{4N}$. First, let us consider

$$
\begin{aligned}
NR_1 = H(M_1) &\overset{(a)}{=} H(M_1 M_0|M_0) \\
&\leq I\left(M_1 M_0; Y_1^N|M_0\right) + N\epsilon_{1N} \\
&\overset{(b)(c)}{\leq} I\left(X_1^N; Y_1^N|V_2^N M_0\right) + N\epsilon_{1N} \\
&\overset{(d)(f)}{\leq} \sum_{n=1}^{N} I(X_{1n}; Y_{1n}|V_{2n} M_0) + N\epsilon_{1N}.
\end{aligned}
\tag{4.142}
$$

Analogously, we may derive an expression for R_2 similar in form to (4.129). Next, let us consider

$$
\begin{aligned}
N(R_1 + R_2) &\overset{(a)}{=} H(M_1 M_0|M_0) + H(M_2 M_0|M_0) \\
&\leq I\left(M_1 M_0; Y_1^N|M_0\right) + I\left(M_2 M_0; Y_2^N|M_0\right) + N(\epsilon_{1N} + \epsilon_{2N}) \\
&\overset{(b)(c)}{\leq} I\left(X_1^N; Y_1^N|M_0\right) + I\left(X_2^N; Y_2^N|X_1^N M_0\right) + N(\epsilon_{1N} + \epsilon_{2N}) \\
&= I\left(X_1^N; Y_1^N|M_0\right) + I\left(V_2^N; Y_2^N|X_1^N M_0\right) + I\left(X_2^N; Y_2^N|X_1^N V_2^N M_0\right) \\
&\quad + N(\epsilon_{1N} + \epsilon_{2N}) \\
&\overset{(e)}{\leq} I\left(X_1^N; Y_1^N|M_0\right) + I\left(V_2^N; Y_1^N|X_1^N M_0\right) + I\left(X_2^N; Y_2^N|X_1^N V_2^N M_0\right) \\
&\quad + N(\epsilon_{1N} + \epsilon_{2N}) \\
&= I\left(V_2^N X_1^N; Y_1^N|M_0\right) + I\left(X_2^N; Y_2^N|V_1^N V_2^N M_0\right) + N(\epsilon_{1N} + \epsilon_{2N}) \\
&\overset{(d)(f)}{\leq} \sum_{n=1}^{N} \left[I(V_{2n} X_{1n}; Y_{1n}|M_0) + I(X_{2n}; Y_{2n}|M_0 V_{1n} V_{2n})\right] + N(\epsilon_{1N} + \epsilon_{2N}).
\end{aligned}
\tag{4.143}
$$

Analogously, we may derive an expression for $R_1 + R_2$ similar in form to (4.131). Next, let us also consider

$$\begin{aligned}
N\left(R_1+R_2\right) &\overset{(a)}{=} H\left(M_1M_0|M_0\right)+H\left(M_2M_0|M_0\right)\\
&\leq I\left(M_1M_0;Y_1^N|M_0\right)+I\left(M_2M_0;Y_2^N|M_0\right)+N\left(\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(c)}{\leq} I\left(X_1^N;Y_1^N|M_0\right)+I\left(X_2^N;Y_2^N|M_0\right)+N\left(\epsilon_{1N}+\epsilon_{2N}\right)\\
&= I\left(V_1^N;Y_1^N|M_0\right)+I\left(X_1^N;Y_1^N|V_1^NM_0\right)+I\left(V_2^N;Y_2^N|M_0\right)\\
&\quad+I\left(X_2^N;Y_2^N|V_2^NM_0\right)+N\left(\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(b)}{\leq} I\left(V_1^N;Y_1^N|X_2^NM_0\right)+I\left(X_1^N;Y_1^N|V_1^NM_0\right)+I\left(V_2^N;Y_2^N|X_1^NM_0\right)\\
&\quad+I\left(X_2^N;Y_2^N|V_2^NM_0\right)+N\left(\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(e)}{\leq} I\left(V_1^N;Y_2^N|X_2^NM_0\right)+I\left(X_1^N;Y_1^N|V_1^NM_0\right)+I\left(V_2^N;Y_1^N|X_1^NM_0\right)\\
&\quad+I\left(X_2^N;Y_2^N|V_2^NM_0\right)+N\left(\epsilon_{1N}+\epsilon_{2N}\right)\\
&= I\left(V_2^NX_1^N;Y_1^N|V_1^NM_0\right)+I\left(V_1^NX_2^N;Y_2^N|V_2^NM_0\right)+N\left(\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(d)(f)}{\leq} \sum_{n=1}^{N}\left[I\left(X_{1n}V_{2n};Y_{1n}|M_0V_{1n}\right)+I\left(X_{2n}V_{1n};Y_{2n}|M_0V_{2n}\right)\right]+N\left(\epsilon_{1N}+\epsilon_{2N}\right).
\end{aligned} \tag{4.144}$$

Next, let us consider

$$\begin{aligned}
N\left(2R_1+R_2\right) &\overset{(a)}{=} 2H\left(M_1M_0|M_0\right)+H\left(M_2|M_0\right)\\
&\leq 2I\left(M_1M_0;Y_1^N|M_0\right)+I\left(M_2M_0;Y_2^N|M_0\right)+N\left(2\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(b)(c)}{\leq} I\left(X_1^N;Y_1^N|M_0\right)+I\left(X_1^N;Y_1^N|X_2^NM_0\right)+I\left(X_2^N;Y_2^N|M_0\right)\\
&\quad+N\left(2\epsilon_{1N}+\epsilon_{2N}\right)\\
&= I\left(X_1^N;Y_1^N|M_0\right)+I\left(V_1^N;Y_1^N|X_2^NM_0\right)+I\left(X_1^N;Y_1^N|X_2^NV_1^NM_0\right)\\
&\quad+I\left(V_2^N;Y_2^N|M_0\right)+I\left(X_2^N;Y_2^N|V_2^NM_0\right)+N\left(2\epsilon_{1N}+\epsilon_{2N}\right)\\
&\overset{(b)(e)}{\leq} I\left(X_1^N;Y_1^N|M_0\right)+I\left(V_1^N;Y_2^N|X_2^NM_0\right)+I\left(X_1^N;Y_1^N|X_2^NV_1^NM_0\right)\\
&\quad+I\left(V_2^N;Y_1^N|X_1^NM_0\right)+I\left(X_2^N;Y_2^N|V_2^NM_0\right)+N\left(2\epsilon_{1N}+\epsilon_{2N}\right)
\end{aligned}$$

$$= I\left(V_2^N X_1^N; Y_1^N | M_0\right) + I\left(X_2^N V_1^N; Y_2^N | V_2^N M_0\right) + I\left(X_1^N; Y_1^N | V_1^N V_2^N M_0\right)$$
$$+ N\left(2\epsilon_{1N} + \epsilon_{2N}\right)$$
$$\overset{(d)(f)}{\leq} \sum_{n=1}^{N} \left[I\left(V_{2n} X_{1n}; Y_{1n} | M_0\right) + I\left(X_{2n} V_{1n}; Y_{2n} | V_{2n} M_0\right) + I\left(X_{1n}; Y_{1n} | V_{1n} V_{2n} M_0\right)\right]$$
$$+ N\left(2\epsilon_{1N} + \epsilon_{2N}\right). \tag{4.145}$$

Analogously, we may derive an expression for $R_1 + 2R_2$ similar in form to (4.134). Next, let us consider

$$\begin{aligned} N\left(R_0 + R_1\right) &= H\left(M_0 M_1\right) \\ &\leq I\left(M_0 M_1; Y_1^N\right) + N\left(\epsilon_{1N} + \epsilon_{3N}\right) \\ &\overset{(c)}{\leq} I\left(M_0 X_1^N; Y_1^N\right) + N\left(\epsilon_{1N} + \epsilon_{3N}\right) \\ &\overset{(f)}{\leq} I\left(M_0 V_2^N X_1^N; Y_1^N\right) + N\left(\epsilon_{1N} + \epsilon_{3N}\right) \\ &\overset{(d)(f)}{\leq} \sum_{n=1}^{N} I\left(X_{1n} V_{2n}; Y_{1n}\right) + N\left(\epsilon_{1N} + \epsilon_{3N}\right). \end{aligned} \tag{4.146}$$

Analogously, we may derive an expression for $R_0 + R_2$ similar in form to (4.136). Next, let us consider

$$N\left(R_0 + R_1 + R_2\right) \overset{(a)}{=} H\left(M_1 M_0\right) + H\left(M_2 M_0 | M_0\right)$$
$$\leq I\left(M_1 M_0; Y_1^N\right) + I\left(M_2 M_0; Y_2^N | M_0\right) + N\left(\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right)$$
$$\overset{(b)(c)}{\leq} I\left(M_0 X_1^N; Y_1^N\right) + I\left(X_2^N; Y_2^N | X_1^N M_0\right) + N\left(\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right)$$
$$= I\left(M_0 X_1^N; Y_1^N\right) + I\left(V_2^N; Y_2^N | X_1^N M_0\right) + I\left(X_2^N; Y_2^N | X_1^N V_2^N M_0\right)$$
$$+ N\left(\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right)$$
$$\overset{(e)}{\leq} I\left(M_0 X_1^N; Y_1^N\right) + I\left(V_2^N; Y_1^N | X_1^N M_0\right) + I\left(X_2^N; Y_2^N | X_1^N V_2^N M_0\right)$$
$$+ N\left(\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right)$$
$$\overset{(c)}{=} I\left(M_0 V_2^N X_1^N; Y_1^N\right) + I\left(X_2^N; Y_2^N | V_1^N V_2^N M_0\right) + N\left(\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right)$$

$$\overset{(d)(f)}{\leq} \sum_{n=1}^{N} \left[I\left(V_{2n} X_{1n}; Y_{1n}\right) + I\left(X_{2n}; Y_{2n} | V_{1n} V_{2n} M_0\right) \right] + N\left(\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right). \tag{4.147}$$

We may analogously derive an expression for $R_0 + R_1 + R_2$ similar in form to the expression (4.138). Finally, let us consider

$$\begin{aligned}
&N\left(R_0 + 2R_1 + R_2\right) \\
&\overset{(a)}{=} H\left(M_1 M_0\right) + H\left(M_1 | M_0\right) + H\left(M_2 | M_0\right) \\
&\leq I\left(M_1 M_0; Y_1^N\right) + I\left(M_1 M_0; Y_1^N | M_0\right) + I\left(M_2 M_0; Y_2^N | M_0\right) \\
&\quad + N\left(2\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right) \\
&\overset{(b)(c)}{\leq} I\left(M_0 X_1^N; Y_1^N\right) + I\left(X_1^N; Y_1^N | X_2^N M_0\right) + I\left(X_2^N; Y_2^N | M_0\right) \\
&\quad + N\left(2\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right) \\
&= I\left(M_0 X_1^N; Y_1^N\right) + I\left(V_1^N; Y_1^N | X_2^N M_0\right) + I\left(X_1^N; Y_1^N | X_2^N V_1^N M_0\right) \\
&\quad + I\left(V_2^N; Y_2^N | M_0\right) + I\left(X_2^N; Y_2^N | V_2^N M_0\right) + N\left(2\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right) \\
&\overset{(b)(e)}{\leq} I\left(M_0 X_1^N; Y_1^N\right) + I\left(V_1^N; Y_2^N | X_2^N M_0\right) + I\left(X_1^N; Y_1^N | X_2^N V_1^N M_0\right) \\
&\quad + I\left(V_2^N; Y_1^N | X_1^N M_0\right) + I\left(X_2^N; Y_2^N | V_2^N M_0\right) + N\left(2\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right) \\
&= I\left(M_0 V_2^N X_1^N; Y_1^N\right) + I\left(X_2^N V_1^N; Y_2^N | V_2^N M_0\right) + I\left(X_1^N; Y_1^N | V_1^N V_2^N M_0\right) \\
&\quad + N\left(2\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right) \\
&\overset{(d)(f)}{\leq} \sum_{n=1}^{N} \left[I\left(V_{2n} X_{1n}; Y_{1n}\right) + I\left(X_{2n} V_{1n}; Y_{2n} | V_{2n} M_0\right) + I\left(X_{1n}; Y_{1n} | V_{1n} V_{2n} M_0\right) \right] \\
&\quad + N\left(2\epsilon_{1N} + \epsilon_{2N} + \epsilon_{3N}\right).
\end{aligned} \tag{4.148}$$

Analogously, we may derive an expression for $R_0 + R_1 + 2R_2$ similar in form to (4.140). Finally, we define $M_0 \triangleq V_0$ and allowing $N \to \infty$, we obtain the conditions in Thm. 4.8. □

For Thm. 4.8, a time-sharing random parameter Q is unnecessary as we may set $V_0 \triangleq \left(V_0', Q\right)$ in Thm. 4.8. Thm. 4.8 includes, but is not restricted to, the class of strong IFCs with common information studied in [54] and also the class of deterministic IFCs with common information studied in [18].

Chapter 5

Capacity Theorems for the "Z"-Channel

5.1 Introduction

We consider the two-user "Z"-channel (ZC), Fig. 5.1, recently introduced by Vishwanath, Jindal, and Goldsmith [13]. The ZC consists of two senders and two receivers. The transmission of sender TX_1 can reach only receiver RX_1, while that of sender TX_2 can reach both receivers. One of the senders transmits information to its intended receiver (without interfering with the unintended receiver), while the other sender transmits information to both receivers. The complete

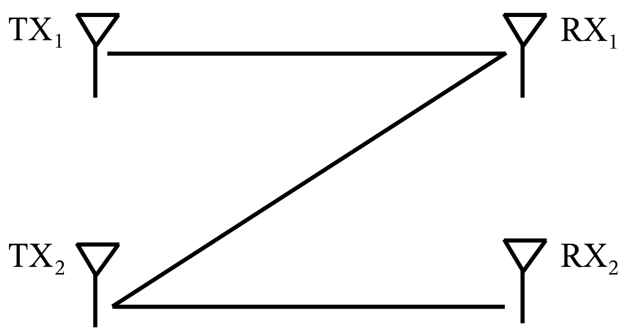

Figure 5.1: The configuration of the ZC

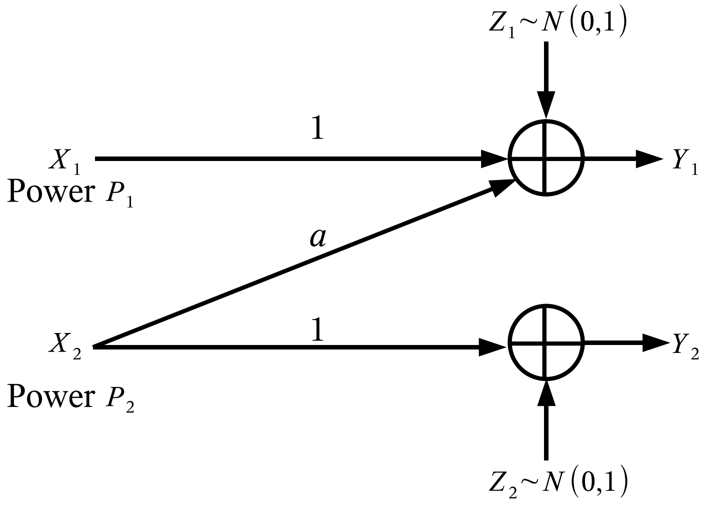

Figure 5.2: Standard form Gaussian ZC

characterization of the discrete memoryless ZC (DMZC) remains unknown to date.

In this chapter, we also study the Gaussian ZC shown in Fig. 5.2. We use the term *weak crossover link gain* to describe the scenario $0 < a^2 < 1$ and the term *strong crossover link gain* to describe the scenario $a^2 \geq 1$. Furthermore, we use the terms *moderately strong crossover link gain* and *very strong crossover link gain* to differentiate between the two scenarios $1 \leq a^2 \leq 1 + P_1$ and $a^2 > 1 + P_1$, respectively. Vishwanath, Jindal, and Goldsmith [13] established an achievable rate region for the Gaussian ZC with very strong crossover link gain. In [55], Liu and Ulukus determined an inner bound and an outer bound to the capacity region of the Gaussian ZC with weak crossover link gain. To date, the capacity region of the Gaussian ZC is only known when the crossover link gain is 1 [55].

In this chapter, we study both the discrete memoryless ZC and the Gaussian ZC. We first establish achievable rates for the general DMZC. The coding strategy uses rate-splitting and superposition coding at the sender with information for both receivers. At the receivers, we use joint decoding. We then specialize the rates obtained to two different types of degraded DMZCs and also derive respective outer bounds to their capacity regions. We show that as long as a certain condition is satisfied, the achievable rate region is the capacity region for one type of degraded DMZC. The results are then extended to the two-user Gaussian ZC with different crossover link gains. We determine an outer bound to the capacity region of the Gaussian ZC with strong crossover link gain and establish the capacity region for moderately strong crossover link gain.

5.1.1 Outline

The outline of the chapter is as follows:

- We first give a mathematical model for the DMZC in Section 5.2. We then describe three different types of degraded ZCs. We also describe the Gaussian ZC model.
- Next, we review past results on the ZC in Section 5.3. We describe a problem in one of the proofs in [13] for the capacity region of one type of degraded DMZC.
- In Section 5.4, we establish an achievable rate region for the general DMZC using rate-splitting and joint decoding.
- In Section 5.5, we specialize the result for the general setting to one type of degraded DMZC. We also determine an outer bound to the capacity region. The result is extended directly to the two-user Gaussian ZC with weak crossover link gain.
- In Section 5.6, we specialize the result for the general setting to another type of degraded DMZC. The result is extended directly to the Gaussian ZC with strong crossover link gain. We also determine respective outer bounds to their capacity regions. We establish the capacity region of the Gaussian ZC with moderately strong crossover link gain. In the discrete case, we show that the achievable rate region is the capacity region if a certain condition is satisfied. Finally, we show that the achievable rate region, determined

in [13], for the Gaussian ZC with very strong crossover link gain can be enlarged.

5.2 Mathematical Preliminaries

A two-user discrete ZC consists of four finite sets $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}_1$, $\mathcal{Y}_2$, and a joint distribution $p(y_1, y_2|x_1, x_2)$, with the conditional marginal distributions given by

$$p(y_1|x_1, x_2) = \sum_{y_2 \in \mathcal{Y}_2} p(y_1, y_2|x_1, x_2), \tag{5.1}$$

$$p(y_2|x_2) = \sum_{y_1 \in \mathcal{Y}_1} p(y_1, y_2|x_1, x_2). \tag{5.2}$$

The ZC is said to be memoryless if

$$p_{Y_1^N Y_2^N|X_1^N X_2^N}\left(y_1^N, y_2^N|x_1^N, x_2^N\right) = \prod_{n=1}^{N} p_{Y_1|X_1X_2Y_2}(y_{1n}|x_{1n}, x_{2n}, y_{2n})\, p_{Y_2|X_2}(y_{2n}|x_{2n}).$$

Throughout the chapter, we assume the ZC to be memoryless. From (5.2), we see that

$$X_1 \to X_2 \to Y_2 \tag{5.3}$$

form a Markov chain. As there is no cooperation between the two receivers, the capacity region of the ZC depends on the joint distribution $p(y_1, y_2|x_1, x_2)$ only through the conditional marginal distributions. In addition, we note that X_1 and Y_2 are independent for all input distributions of the form $p(x_1)\,p(x_2)$.

$$\begin{aligned}\sum_{x_2 \in \mathcal{X}_2}\sum_{y_1 \in \mathcal{Y}_1} p(y_1, y_2|x_1, x_2)\, p(x_1)\, p(x_2) &= p(x_1)\sum_{x_2 \in \mathcal{X}_2} p(y_2|x_2)\, p(x_2) \\ &= p(x_1)\, p(y_2). \end{aligned} \tag{5.4}$$

Similarly, X_1^N and Y_2^N are independent for all input distributions of the form $p\left(x_1^N\right) p\left(x_2^N\right)$. In the ZC, sender TX_1 produces an integer $M_1 \in \left\{1, ..., 2^{NR_1}\right\}$.

Sender TX_2 produces an integer pair $(M_{21}, M_{22}) \in \left\{1, ..., 2^{NR_{21}}\right\} \times \left\{1, ..., 2^{NR_{22}}\right\}$. M_1 denotes the message sender TX_1 intends to transmit to receiver RX_1, M_{21} denotes the message sender TX_2 intends to transmit to receiver RX_1, and M_{22} denotes the message sender TX_2 intends to transmit to receiver RX_2. A (2^{NR_1}, $2^{NR_{21}}$, $2^{NR_{22}}$, N) code for a ZC with independent messages consists of two encoders

$$\Psi_1 : \left\{1, ..., 2^{NR_1}\right\} \to \mathcal{X}_1^N,$$
$$\Psi_2 : \left\{1, ..., 2^{NR_{21}}\right\} \times \left\{1, ..., 2^{NR_{22}}\right\} \to \mathcal{X}_2^N,$$

and two decoders

$$\Phi_1 : \mathcal{Y}_1^N \to \left\{1, ..., 2^{NR_1}\right\} \times \left\{1, ..., 2^{NR_{21}}\right\},$$
$$\Phi_2 : \mathcal{Y}_2^N \to \left\{1, ..., 2^{NR_{22}}\right\}.$$

The average probability of error is defined as the probability that the decoded messages are not equal to the transmitted messages, i.e.,

$$P_e^{(N)} = \Pr\left(\Phi_1\left(Y_1^N\right) \neq (M_1, M_{21}) \text{ or } \Phi_2\left(Y_2^N\right) \neq M_{22}\right).$$

The distributions of M_1, M_{21}, and M_{22} are assumed to be uniform. A rate triplet (R_1, R_{21}, R_{22}) is said to be achievable for the ZC if there exists a sequence of $\left(2^{NR_1}, 2^{NR_{21}}, 2^{NR_{22}}, N\right)$ codes with $P_e^{(N)} \to 0$ as $N \to \infty$.

Willems and Van Der Meulen proved that stochastic encoders and decoders do not increase the capacity region of the discrete memoryless multiple access channel with cribbing encoders [56]. The same argument can be extended to the ZC.

Proposition 5.1. *Stochastic encoders and decoders do not increase the capacity region of the ZC.*

Proof. For stochastic encoders and decoders, we may assume that the encoding and decoding functions are given by

$$x_1^N = \Psi_1\left(M_1, A^{E1}\right) \tag{5.5}$$

$$x_2^N = \Psi_2\left(M_{21}, M_{22}, A^{E2}\right) \tag{5.6}$$

$$\left(\hat{M}_1, \hat{M}_{21}\right) = \Phi_1\left(Y_1^N, A^{D1}\right) \tag{5.7}$$

$$\hat{M}_{22} = \Phi_2\left(Y_2^N, A^{D2}\right) \tag{5.8}$$

where A^{E1}, A^{E2}, A^{D1}, and A^{D2} are random variables independent of each other and all other random variables. Now, define

$$A \triangleq \left(A^{E1}, A^{E2}, A^{D1}, A^{D2}\right) \tag{5.9}$$

where A ranges over $\mathcal{A}$ and $p(.)$ is A's density function. If a (2^{NR_1}, $2^{NR_{21}}$, $2^{NR_{22}}$, N)-code exists for stochastic encoders and decoders, and achieves a probability of error P_e, we then have

$$\begin{aligned} P_e &= \Pr\left\{\left(\hat{M}_1, \hat{M}_{21}, \hat{M}_{22}\right) \neq (M_1, M_{21}, M_{22})\right\} \\ &= \int_{a\in\mathcal{A}} p(A=a)\Pr\left\{\left(\hat{M}_1, \hat{M}_{21}, \hat{M}_{22}\right) \neq (M_1, M_{21}, M_{22})\,|A=a\right\}da. \end{aligned} \tag{5.10}$$

It then readily follows that there must exist an $a \in \mathcal{A}$ such that

$$\Pr\left\{\left(\hat{M}_1, \hat{M}_{21}, \hat{M}_{22}\right) \neq (M_1, M_{21}, M_{22})\,|A=a\right\} \leq P_e. \tag{5.11}$$

Hence, the capacity region of the ZC is unaffected if we assume deterministic encoders and decoders. □

5.2.1 Some useful properties of Markov chains

We state some useful properties of Markov chains that we will use throughout the chapter (see [57, Sec. 1.1.5]).

- Decomposition: $X \to Z \to YW \Rightarrow X \to Z \to Y$
- Weak Union: $X \to Z \to YW \Rightarrow X \to ZW \to Y$
- Contraction: $(X \to Z \to Y) \,\&\, (X \to ZY \to W) \Rightarrow X \to Z \to YW$

5.2.2 Degraded ZC

We first define three types of physically degraded ZCs. A ZC is said to be stochastically degraded if its conditional marginal distributions are the same as that of a physically degraded ZC. Since $\Pr\left(\left(\hat{M}_1, \hat{M}_{21}\right) \neq (M_1, M_{21})\right)$ and $\Pr\left(\hat{M}_{22} \neq M_{22}\right)$ depend only on the conditional marginals $p_1(y_1|x_1, x_2)$ and $p_2(y_2|x_2)$, the capacity region of the stochastically degraded ZC is the same as that of the corresponding physically degraded ZC. In the rest of the chapter, we assume that the ZCs are physically degraded.

Definition 5.1. We define a ZC to be a *degraded* ZC of *type I* if

$$X_2 \to (X_1, Y_2) \to Y_1 \tag{5.12}$$

form a Markov chain.

Remark 5.1. The joint distribution $p(y_1, y_2|x_1, x_2)$ can be written as

$$\begin{aligned} p(y_1, y_2|x_1, x_2) &= p(y_2|x_1, x_2)\, p(y_1|x_1, x_2, y_2) \\ &= p(y_2|x_2)\, p(y_1|x_1, y_2). \end{aligned} \tag{5.13}$$

For the degraded ZC of type I, the following inequality holds:

$$I(W; Y_2) \geq I(W; Y_1|X_1) \tag{5.14}$$

for all input distributions $p(x_1)\, p(w)\, p(x_2|w)$.

Example 5.1. Fig. 5.5 shows a degraded Gaussian ZC of type I. One may easily verify that the two Markov chains given by (5.3) and (5.12) are simultaneously satisfied.

Definition 5.2. We define a ZC to be a *degraded* ZC of *type II* if

$$X_2 \to (X_1, Y_1) \to Y_2 \tag{5.15}$$

form a Markov chain.

Remark 5.2. For the degraded ZC of type II, the joint distribution $p(y_1, y_2|x_1, x_2)$ can be written as

$$\begin{aligned} p(y_1, y_2|x_1, x_2) &= p(y_1|x_1, x_2)\, p(y_2|x_1, x_2, y_1) \\ &= p(y_1|x_1, x_2)\, p(y_2|x_1, y_1)\,. \end{aligned} \tag{5.16}$$

The following inequality holds:

$$I(W; Y_1|X_1) \geq I(W; Y_2) \tag{5.17}$$

for all input distributions $p(x_1)\, p(w)\, p(x_2|w)$.

Example 5.2. Fig. 5.7 shows a degraded Gaussian ZC of type II. One may easily verify that the two Markov chains given by (5.3) and (5.15) are simultaneously satisfied.

Definition 5.3. We define a ZC to be a *degraded* ZC of *type III* if

$$(X_1, X_2) \to Y_1 \to Y_2 \tag{5.18}$$

form a Markov chain.

Remark 5.3. The degraded ZC of type III was first defined in [13] and corresponds to the case where the output of receiver RX_2 (Y_2) is a degraded version of the

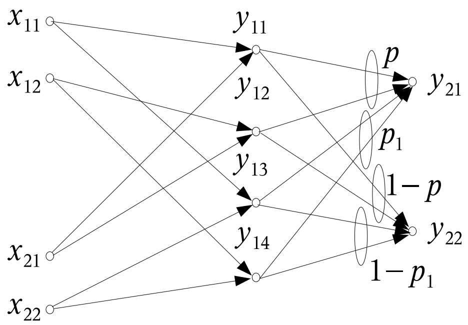

Figure 5.3: An example of a degraded ZC of type III

output of receiver RX_1 (Y_1). By applying the weak union property for Markov chains, we see that the Markov chain $X_2 \to (X_1, Y_1) \to Y_2$ holds for the degraded ZC of type III. Hence, a degraded ZC of type III is also a degraded ZC of type II. However, the converse may not necessarily be true.

Example 5.3. We consider the degraded ZC of type III shown in Fig. 5.3 where $\mathcal{X}_1 = \{x_{11}, x_{12}\}$, $\mathcal{X}_2 = \{x_{21}, x_{22}\}$, $\mathcal{Y}_1 = \{y_{11}, y_{12}, y_{13}, y_{14}\}$, and $\mathcal{Y}_2 = \{y_{21}, y_{22}\}$. We note that receiver RX_1 is able to decode X_1 and X_2 without error. We also have $p\,(Y_2|Y_1 = y_{11}) = p\,(Y_2|Y_1 = y_{12})$ and $p\,(Y_2|Y_1 = y_{13}) = p\,(Y_2|Y_1 = y_{14})$. One may easily verify that the two Markov chains given by (5.3) and (5.18) are simultaneously satisfied.

5.2.3 Gaussian ZC

For a general Gaussian ZC, the inputs and outputs are related by

$$Y_1^* = c_{11}X_1^* + c_{21}X_2^* + Z_1^* \tag{5.19}$$

$$Y_2^* = c_{22}X_2^* + Z_2^*. \tag{5.20}$$

as depicted in Fig. 5.4. The channel outputs and inputs are real-valued and have power constraints $\mathbb{E}\left[|X_1^*|^2\right] \leq P_1^*$ and $\mathbb{E}\left[|X_2^*|^2\right] \leq P_2^*$. Z_1^* and Z_2^* are zero-mean Gaussian random variables with variance σ_1^2 and σ_2^2 respectively. Similar to the

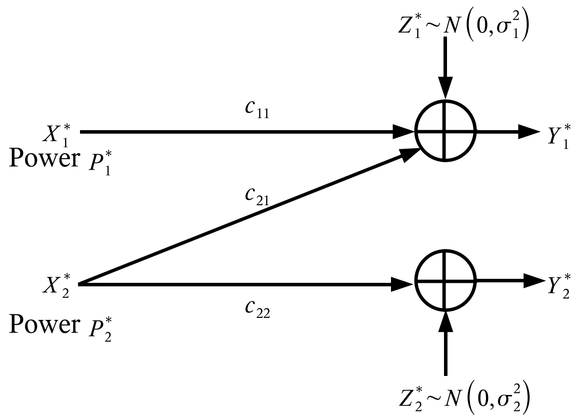

Figure 5.4: General Gaussian ZC

Gaussian IFC, one can use a scaling transformation to convert the Gaussian ZC into its standard form as shown in Fig. 5.2. The inputs and outputs of the standard form Gaussian ZC are related by

$$\begin{aligned} Y_1 &= X_1 + aX_2 + Z_1, \\ Y_2 &= X_2 + Z_2 \end{aligned} \tag{5.21}$$

where

$$\begin{aligned} X_1 &= \frac{c_{11}}{\sigma_1} X_1^*, \quad Y_1 = \frac{Y_1^*}{\sigma_1}, \quad Z_1 = \frac{Z_1^*}{\sigma_1} \\ X_2 &= \frac{c_{22}}{\sigma_2} X_2^*, \quad Y_2 = \frac{Y_2^*}{\sigma_2}, \quad Z_2 = \frac{Z_2^*}{\sigma_2} \end{aligned} \tag{5.22}$$

and the new power constraints and channel gain are

$$P_1 = \frac{c_{11}^2}{\sigma_1^2} P_1^*, \quad P_2 = \frac{c_{22}^2}{\sigma_2^2} P_2^*, \quad a = \frac{c_{12}}{c_{11}} \frac{\sigma_1}{\sigma_2}. \tag{5.23}$$

Equivalent Gaussian ZC with weak crossover link gain ($0 < a^2 < 1$)

In [58], Costa showed that the class of Gaussian ZIFC with weak interference ($a^2 \in (0, 1)$) and the class of degraded Gaussian IFC are equivalent, i.e., for every Gaussian ZIFC with weak interference, there is a degraded Gaussian IFC with

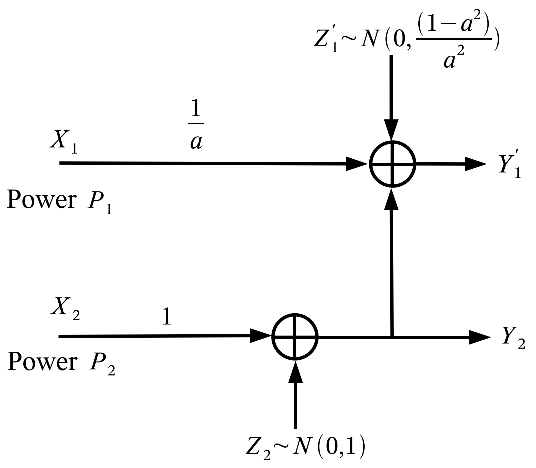

Figure 5.5: Degraded Gaussian ZC of type I

the same capacity region. Using the same arguments as in [58], we can deduce that the class of Gaussian ZC with weak crossover link gain and the class of degraded Gaussian ZC of type I are equivalent, i.e., for every Gaussian ZC with weak crossover gain, there is a degraded Gaussian ZC of type I with the same capacity region. Hence, the capacity region of the channel shown in Fig. 5.5 is equivalent to that of the model shown in Fig. 5.2 when $0 < a^2 < 1$. Hence, an achievable rate region for the degraded DMZC of type I can be readily extended to the Gaussian ZC with weak crossover link gain. The assumption $0 < a^2 < 1$ ensures that the term $\frac{1-a^2}{a^2}$ is non-negative.

Equivalent Gaussian ZC with strong crossover link gain ($a^2 \geq 1$)

Consider the two channels shown in Fig. 5.6. The second channel is equivalent to the first since scaling the output of a channel does not affect its capacity. The channel shown in Fig. 5.7 is equivalent to the channel shown in Fig. 5.6b since they have identical conditional marginal distributions. In Fig. 5.7, the outputs are related to the inputs by

$$
\begin{aligned}
Y_1' &= \frac{X_1}{a} + X_2 + Z_{21}, \\
Y_2' &= X_2 + Z_{21} + Z_{22}
\end{aligned}
\tag{5.24}
$$

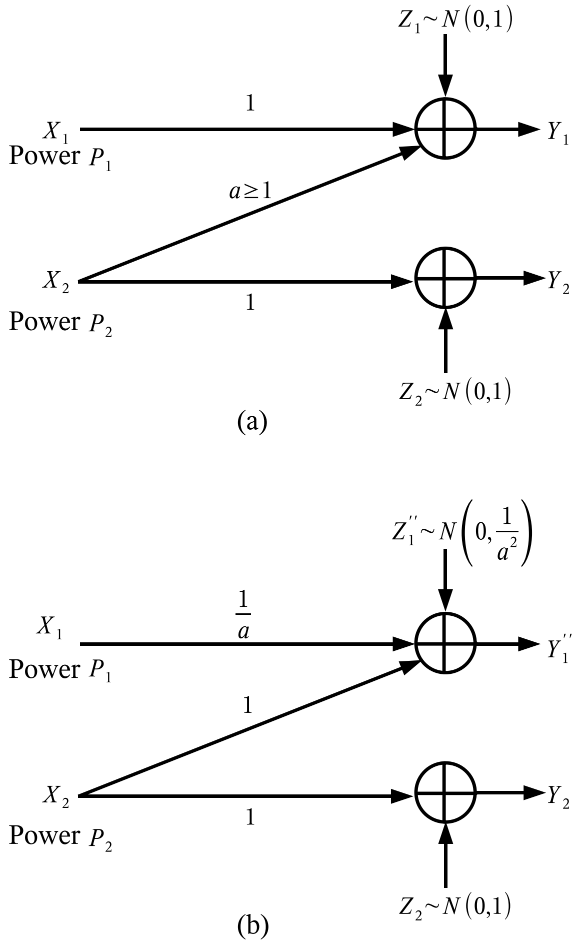

Figure 5.6: Transformation of the Gaussian ZC ($a^2 \geq 1$)

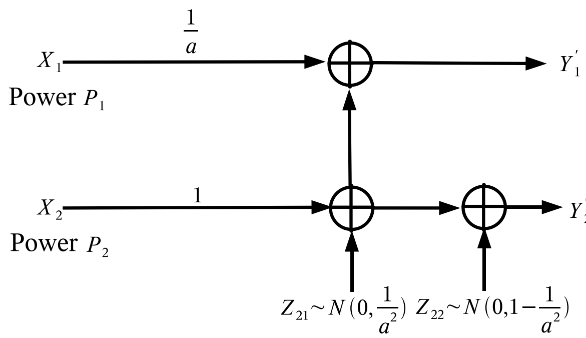

Figure 5.7: A degraded Gaussian ZC of type II

where $Z_{21} \sim \mathcal{N}\left(0, \frac{1}{a^2}\right)$ and $Z_{22} \sim \mathcal{N}\left(0, 1 - \frac{1}{a^2}\right)$. We will make use of this equivalent channel to determine an outer bound to the capacity region of the Gaussian ZC with strong crossover link gain. Here, we have made the assumption that $a^2 \geq 1$ to ensure that the term $1 - \frac{1}{a^2}$ is non-negative. Since the class of Gaussian ZC with strong crossover link gain and the class of degraded Gaussian ZC of type II are equivalent, an achievable rate region for the degraded DMZC of type II can be readily extended to the Gaussian ZC with strong crossover link gain.

5.3 Review of past results

In this section, we review some known results for the ZC.

5.3.1 Degraded ZC of Type I

In [55, Larger Achievable Region 2], Liu and Ulukus determined a lower bound to the capacity region of the Gaussian ZC with weak crossover link gain. This corresponds to the degraded ZC of Type I. Liu and Ulukus make use of rate splitting and successive decoding technique similar to Carleial for the Gaussian IFC [46]. Let us denote the information sender TX_2 intends to transmit to receiver RX_1 by M_{21} and the information sender TX_2 intends to transmit to receiver RX_2 by M_{22}. M_{21} has rate T_{21}. Sender TX_2 splits M_{22} in $[M_{221}, M_{222}]$, where M_{221} and M_{222} have rates S_{22} and T_{22}, respectively. M_{221} represents the information that only receiver RX_2 can decode, while M_{21} and M_{222} represents the information that both receivers can decode.

One strategy is to have receiver RX_1 decode M_{21} followed by M_{222} and finally M_1. Receiver RX_2 decodes M_{21} followed by M_{222} and finally M_{221}. Another strategy is to have receiver RX_1 decode M_{222} followed by M_{21} and finally M_1, while receiver RX_2 decodes M_{222} followed by M_{21} and finally M_{221}. The *Larger Achievable Region 2* determined by Liu and Ulukus is the union of the achievable rate regions of these two strategies for the Gaussian ZC with weak crossover link gain. When this strategy is applied to the degraded DMZC of type I, an achievable rate region is given by the set $\mathcal{R}_{\mathrm{LU}}$, which is the closure of the convex hull of all rate triplets (R_1, R_{21}, R_{22}) satisfying

$$R_1 \leq S_1 \tag{5.25}$$

$$R_{21} \leq T_{21} \tag{5.26}$$

$$R_{22} \leq S_{22} + T_{22} \tag{5.27}$$

where S_1, T_{21}, S_{22} and T_{22} are subject to the constraints

$$T_{21} + T_{22} \leq I\,(W; Y_1) \tag{5.28}$$

$$S_1 \leq I\,(X_1; Y_1|W) \tag{5.29}$$

$$S_{22} \leq I\,(X_2; Y_2|W) \tag{5.30}$$

for all input distributions $p\,(w, x_1, x_2) = p\,(x_1)\,p\,(w, x_2)$. In [55], Liu and Ulukus also determined an outer bound to the capacity region of the Gaussian ZC with weak crossover link gain. By making use of the entropy power inequality, Liu and Ulukus obtained the following theorem:

Theorem 5.2. *[Liu and Ulukus] For the Gaussian ZC with weak crossover link gain ($0 \leq a^2 \leq 1$), the achievable rate triplets (R_1, R_{21}, R_{22}) have to satisfy*

$$R_{21} \leq \gamma\left(\frac{a^2\beta P_2}{a^2\,(1-\beta)\,P_2 + 1}\right) \tag{5.31}$$

$$R_{22} \leq \gamma\,((1-\beta)\,P_2) \tag{5.32}$$

$$R_1 + R_{21} \leq \gamma\left(\frac{a^2\beta P_2 + P_1}{a^2\,(1-\beta)\,P_2 + 1}\right) \tag{5.33}$$

for some $0 \leq \beta \leq 1$ and where $\gamma\,(x) \triangleq \frac{1}{2}\log_2(1+x)$.

Proof. The proof can be found in [55, Thm. 2]. □

Remark 5.4. This outer bound includes the best outer bound to the capacity region of the Gaussian ZIFC under weak interference derived by Kramer [59, Thm. 2]. Kramer makes use of a proposition of Sato for a degraded interference channel, while Liu and Ulukus derived this using the entropy power inequality. To see the equivalence between the two, we can ignore the constraint for R_{21} since $R_{21} = 0$ for an interference channel. Hence, for the Gaussian ZIFC under weak interference, the achievable rate pair (R_1, R_2) have to satisfy

$$R_1 \leq \gamma \left(\frac{a^2 \beta P_2 + P_1}{a^2 (1-\beta) P_2 + 1} \right) \tag{5.34}$$

$$R_2 \leq \gamma \left((1-\beta) P_2 \right) \tag{5.35}$$

for some $0 \leq \beta \leq 1$. This is in fact the outer bound determined by Kramer for the capacity region of the degraded Gaussian IFC, which is equivalent to that of the Gaussian ZIFC under weak interference.

5.3.2 Degraded ZC of Type III

It was stated in [13] that the capacity region of a degraded DMZC of type III is the closure of the convex hull of all triplets (R_1, R_{21}, R_{22}) subject to

$$R_1 \leq I(X_1; Y_1 | X_2) \tag{5.36}$$

$$R_{21} \leq I(X_2; Y_1 | W X_1) \tag{5.37}$$

$$R_1 + R_{21} \leq I(X_1 X_2; Y_1 | W) \tag{5.38}$$

$$R_{22} \leq I(W; Y_2) \tag{5.39}$$

for some input distributions $p(w, x_1, x_2) = p(x_1) p(w, x_2)$.

Remark 5.5. The rates given by (5.36)-(5.39) can readily be seen to be achievable. Since the output of receiver RX_2 (Y_2) is a degraded version of receiver RX_1 (Y_1), we can use superposition coding at sender TX_2, where the auxiliary random variable U represents the information to be transmitted from sender TX_2 to receiver RX_2. Unfortunately, this achievable rate may not be the outer bound in

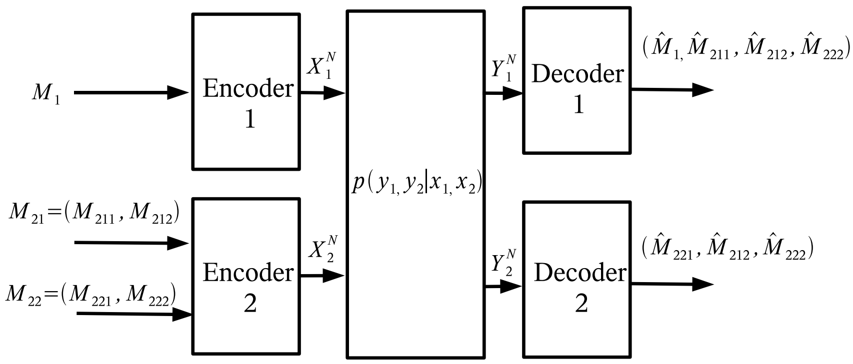

Figure 5.8: Encoding and Decoding for the ZC

general due to the following problem in the converse.

In [13], the authors define $W_n = (M_{22}, Y_{11}, Y_{12}, ..., Y_{1n-1})$ and state that $W_n \rightarrow X_{2n} \rightarrow (Y_{1n}Y_{2n})$ form a Markov chain. However, this is not necessarily the case as W_n may contain some information about Y_{1n} that is not in X_{2n}. We first observe that W_n contains all the past outputs of receiver RX_1 until time $n-1$. Moreover, the current output of receiver RX_1 (Y_{1n}) is dependent on the current input of sender TX_1 and sender TX_2 $(p(y_1|x_1, x_2))$. Hence, the Markov chain should be given by $W_n \rightarrow (X_{1n}X_{2n}) \rightarrow (Y_{1n}Y_{2n})$. Therefore, in the derivation of the outer bound, the input distribution $p(w, x_1, x_2)$ may not be equal to $p(x_1)\, p(w, x_2)$ as specified in [13].

5.4 Achievable rate region for the DMZC

Similar to Carleial's treatment of the interference channel [46], we make use of rate splitting and superposition coding. Transmitter 2 splits M_{21} in $[M_{211}, M_{212}]$, where M_{211} and M_{212} have rates S_{21} and T_{21}, respectively. Similarly, Transmitter 2 splits M_{22} in $[M_{221}, M_{222}]$, where M_{221} and M_{222} have rates S_{22} and T_{22}, respectively. Referring to Fig. 5.8, M_{211} represents the information that only receiver RX_1 can decode, while M_{221} represents the information that only receiver RX_2 can decode. M_{212} and M_{222} represents the information that both receivers can decode.

Carleial suggested the use of sequential decoding at the receivers for the interference channel. In [15], Han and Kobayashi refined Carleial's method by using a joint decoder superior to sequential decoding for the interference channel. Rather than using the convex-hull operation, they added a time-sharing random variable Q. Following the ideas of Han and Kobayashi, we use a joint decoder at the receivers and also include a time-sharing random variable Q. We first describe the codebook generation, encoding at the transmitters, and decoding at the receivers before describing our main result in Thm. 5.3.

5.4.1 Random Codebook Construction

We first fix the following input probability distribution:

$$\begin{aligned} &p\left(q, x_1, w, u_1, u_2, x_2\right) \\ &\quad = p\left(q\right) p\left(x_1|q\right) p\left(w|q\right) p\left(u_1|w, q\right) p\left(u_2|w, q\right) p\left(x_2|w, u_1, u_2, q\right). \end{aligned} \tag{5.40}$$

The auxiliary r.v. W carries the common information M_{212} and M_{222}, the auxiliary r.v. U_1 carries the information M_{211}, while the auxiliary r.v. U_2 carries the information M_{221}. The codebook is constructed as follows:

1. Generate one N-sequence $q^N = (q_1, ..., q_N)$, drawn according to

$$p\left(q^N\right) = \prod_{n=1}^{n=N} p_Q\left(q_n\right).$$

2. Generate 2^{NS_1} conditionally independent N-sequences $x_1^N = (x_{11}, ..., x_{1N})$,

each drawn according to

$$p_{X_1^N|Q^N}\left(x_1^N|q^N\right) = \prod_{n=1}^{n=N} p_{X_1|Q}\left(x_{1n}|q_n\right).$$

Label them $x_1^N(m_1)$, $m_1 \in \left\{1, 2, ..., 2^{NS_1}\right\}$.

3. Next, we generate $2^{N(T_{21}+T_{22})}$ conditionally independent N-sequences $W^N = (w_1, ..., w_N)$, each drawn according to

$$p_{W^N|Q^N}\left(w^N|q^N\right) = \prod_{n=1}^{n=N} p_{W|Q}\left(w_n|q_n\right).$$

Label them $w^N(m_{212}, m_{222})$, $m_{212} \in \left\{1, 2, ..., 2^{NT_{21}}\right\}$, $m_{222} \in \left\{1, 2, ..., 2^{NT_{22}}\right\}$.

4. For the codeword q^N and each of the codewords $w^N(m_{212}, m_{222})$, generate $2^{NS_{21}}$ conditionally independent N-sequences $u_1^N = (u_{11}, ..., u_{1N})$, each drawn according to

$$p_{U_1^N|W^NQ^N}\left(u_1^N|w^N(m_{212}, m_{222}), q^N\right) = \prod_{n=1}^{n=N} p_{U_1|WQ}\left(u_{1n}|w_n(m_{212}, m_{222}), q_n\right).$$

Label them $u_1^N(m_{211}, m_{212}, m_{222})$, $m_{211} \in \left\{1, 2, ..., 2^{NS_{21}}\right\}$.

5. For the codeword q^N and each of the codewords $w^N(m_{212}, m_{222})$, generate $2^{NS_{22}}$ conditionally independent N-sequences $u_2^N = (u_{21}, ..., u_{2N})$, each drawn according to

$$p_{U_2^N|W^NQ^N}\left(u_2^N|w^N(m_{212}, m_{222}), q^N\right) = \prod_{n=1}^{n=N} p_{U_2|WQ}\left(u_{2n}|w_n(m_{212}, m_{222}), q_n\right).$$

Label them $u_2^N(m_{221}, m_{212}, m_{222})$, $m_{221} \in \left\{1, 2, ..., 2^{NS_{22}}\right\}$.

6. Finally, for the codeword q^N and each of the codewords $w^N(m_{212}, m_{222})$, $u_1^N(m_{211}, m_{212}, m_{222})$, and $u_2^N(m_{221}, m_{212}, m_{222})$, generate one N-sequence $x_2^N{=}(x_{21}, ..., x_{2N})$, drawn according to

$$p_{X_2^N|U_1^N U_2^N W^N Q^N}\left(x_2^N | u_1^N(m_{211}, m_{212}, m_{222}), u_2^N(m_{221}, m_{212}, m_{222}), w^N(m_{212}, m_{222}), q^N \right)$$
$$= \prod_{n=1}^{n=N} p_{X_2|U_1 U_2 W Q}\left(x_{2n} | u_{1n}(m_{211}, m_{212}, m_{222}), u_{2n}(m_{221}, m_{212}, m_{222}), w_n(m_{212}, m_{222}), q_n \right).$$

   Label them $x_2^N(m_{211}, m_{212}, m_{221}, m_{222})$.

5.4.2 Encoding and Decoding

To send the index m_1, sender TX_1 sends the codeword $x_1^N(m_1)$. To send the pair (m_{211}, m_{212}) to receiver RX_1 and the pair (m_{221}, m_{222}) to receiver RX_2, sender TX_2 sends the codeword $x_2^N(m_{211}, m_{212}, m_{221}, m_{222})$. For decoding, receiver RX_1 determines the unique $(\hat{m}_1, \hat{m}_{211}, \hat{m}_{212}, \hat{m}_{222})$ such that

$$\left(q^N, w^N(\hat{m}_{212}, \hat{m}_{222}), u_1^N(\hat{m}_{211}, \hat{m}_{212}, \hat{m}_{222}), x_1^N(\hat{m}_1), y_1^N\right) \in A_\epsilon^{(N)}(Q, W, U_1, X_1, Y_1). \quad (5.41)$$

For the other decoder, receiver RX_2 determines the unique $(\hat{m}_{221}, \hat{m}_{212}, \hat{m}_{222})$ such that

$$\left(q^N, w^N(\hat{m}_{212}, \hat{m}_{222}), u_2^N(\hat{m}_{221}, \hat{m}_{212}, \hat{m}_{222}), y_2^N\right) \in A_\epsilon^{(N)}(Q, W, U_2, Y_2). \quad (5.42)$$

5.4.3 Main Result

We may then state the main result below.

Theorem 5.3. *An achievable rate region for sending information over the DMZC* *is given by the set* $\mathcal{R}_{\mathrm{G}}$*, which is the closure of all rate triplets* (R_1, R_{21}, R_{22}) *satisfying*

$$R_1 \leq S_1 \tag{5.43}$$

$$R_{21} \leq S_{21} + T_{21} \tag{5.44}$$

$$R_{22} \leq S_{22} + T_{22} \tag{5.45}$$

where S_1*,* S_{21}*,* S_{22}*,* T_{21}*, and* T_{22} *are subject to the following constraints:*

$$S_1 + S_{21} + T_{21} + T_{22} \leq I\left(X_1 W U_1; Y_1|Q\right) \tag{5.46}$$

$$S_{21} + T_{21} + T_{22} \leq I\left(W U_1; Y_1|X_1 Q\right) \tag{5.47}$$

$$S_1 + S_{21} \leq I\left(X_1 U_1; Y_1|W Q\right) \tag{5.48}$$

$$S_1 \leq I\left(X_1; Y_1|W U_1 Q\right) \tag{5.49}$$

$$S_{21} \leq I\left(U_1; Y_1|X_1 W Q\right) \tag{5.50}$$

$$S_{22} + T_{21} + T_{22} \leq I\left(W U_2; Y_2|Q\right) \tag{5.51}$$

$$S_{22} \leq I\left(U_2; Y_2|W Q\right) \tag{5.52}$$

for all input distributions of the form (5.40).

Proof. Refer to Appendix D.1. □

It is easy to see that $\mathcal{R}_{\mathrm{G}}$ is convex. In addition, we note that Thm. 5.3 is not limited to the ZC. It also applies to the general two-sender two-receiver channel (without the constraint in (5.3)) where one sender has information to transmit to both receivers, while the other sender has information to transmit to only one receiver. Next, we show that $\mathcal{R}_{\mathrm{G}}$ includes the capacity regions of the multiple access channel and the degraded broadcast channel. It also includes the best known achievable rate region for the ZIFC.

Remark 5.6. We obtain the multiple access channel when $R_{22} = 0$. By setting $S_{22} = T_{21} = T_{22} = 0$, $R_1 = S_1$, $R_{21} = S_{21}$, $Q \triangleq W \triangleq U_2 \triangleq \emptyset$ and $U_1 \triangleq X_2$, we obtain the capacity of the multiple access channel, which is the closure of the convex hull of all rate pairs (R_1, R_{21}) satisfying

$$R_1 \leq I(X_1; Y_1 | X_2) \tag{5.53}$$

$$R_{21} \leq I(X_2; Y_1 | X_1) \tag{5.54}$$

$$R_1 + R_{21} \leq I(X_1 X_2; Y_2) \tag{5.55}$$

for some input distributions $p(x_1, x_2) = p(x_1) p(x_2)$.

Remark 5.7. We obtain the broadcast channel if Y_1 is independent of the input X_1. If Y_2 is a degraded version of Y_1, we obtain the degraded broadcast channel. By setting $S_{22} = T_{21} = S_1 = R_1 = 0$, $R_{21} = S_{21}$, $R_{22} = T_{22}$, $U_2 \triangleq Q \triangleq \emptyset$, and $U_1 \triangleq X_2$, we obtain the capacity region of the degraded broadcast channel, which is the closure of the convex hull of all rate pairs (R_{21}, R_{22}) satisfying

$$R_{21} \leq I(X_2; Y_1 | W) \tag{5.56}$$

$$R_{22} \leq I(W; Y_2) \tag{5.57}$$

for some input distributions $p(w, x_2) = p(w) p(x_2 | w)$.

Remark 5.8. We obtain the ZIFC when $R_{21} = 0$. By setting $S_{21} = T_{21} = 0$, $U_1 \triangleq \emptyset$, and $U_2 \triangleq X_2$, we obtain the Han-Kobayashi rate region (the best rate region to date) for the ZIFC which is the closure of all rate pairs (R_1, R_{22}) satisfying

$$R_1 \leq S_1 \tag{5.58}$$

$$R_{22} \leq S_{22} + T_{22} \tag{5.59}$$

where S_1, S_{22}, and T_{22} are subject to the following constraints:

$$S_1 + T_{22} \leq I(X_1 W; Y_1 | Q) \tag{5.60}$$

$$T_{22} \leq I(W; Y_1 | X_1 Q) \tag{5.61}$$

$$S_1 \leq I(X_1; Y_1 | W Q) \tag{5.62}$$

$$S_{22} + T_{22} \leq I(X_2; Y_2|Q) \tag{5.63}$$

$$S_{22} \leq I(X_2; Y_2|WQ) \tag{5.64}$$

for some input probability distributions of the following form:

$$p(q, w, x_1, x_2) = p(q)\, p(x_1|q)\, p(w|q)\, p(x_2|w, q)\,. \tag{5.65}$$

By using Fourier-Motzkin elimination, we can reduce this to a set of bounds containing only R_1 and R_{22} (Refer to [59], [22]).

5.5 Rate Regions for the Degraded DMZC of Type I

As we have mentioned in Section 5.2, the capacity region of a Gaussian ZC with weak crossover link gain is equivalent to that of a degraded Gaussian ZC of type I. We shall first determine an achievable rate region for the degraded DMZC of type I. We note that receiver RX_2 is able to decode all the information meant for receiver RX_1. Hence, we may set $S_{21} = 0$. We are then able to establish the following lemma:

Lemma 5.4. *An achievable rate region for sending information over the degraded DMZC of type I* $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1 x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ *is given by the set* $\mathcal{R}_1$*, which is the closure of all rate triplets* (R_1, R_{21}, R_{22}) *satisfying*

$$R_1 \leq S_1 \tag{5.66}$$

$$R_{21} \leq T_{21} \tag{5.67}$$

$$R_{22} \leq S_{22} + T_{22} \tag{5.68}$$

where S_1, S_{21}, S_{22}, and T_{22} are subject to the following constraints:

$$S_1 + T_{21} + T_{22} \leq I\left(WX_1; Y_1|Q\right) \tag{5.69}$$

$$T_{21} + T_{22} \leq I\left(W; Y_1|X_1Q\right) \tag{5.70}$$

$$S_1 \leq I\left(X_1; Y_1|WQ\right) \tag{5.71}$$

$$S_{22} \leq I\left(X_2; Y_2|WQ\right) \tag{5.72}$$

for all input probability distributions of the form (5.65). Furthermore, the region is unchanged if we impose the following constraints on the cardinalities of the auxiliary sets:

$$\|\mathcal{W}\| \leq \|\mathcal{X}_2\| + 2 \quad \textit{and} \quad \|\mathcal{Q}\| \leq 4. \tag{5.73}$$

Proof. Set $S_{21} = 0$, $U_1 \triangleq W$, and $U_2 \triangleq X_2$ in Thm. 5.3. We note that for a degraded DMZC of type I, $I\left(W; Y_1|X_1\right) \leq I\left(W; Y_2\right)$ for all input distributions $p\left(x_1\right)p\left(w\right)p\left(x_2|w\right)$. This implies that

$$\begin{aligned} I\left(W; Y_1|X_1Q\right) + I\left(X_2; Y_2|WQ\right) &\leq I\left(W; Y_2|Q\right) + I\left(X_2; Y_2|WQ\right) \\ &= I\left(X_2; Y_2|Q\right). \end{aligned} \tag{5.74}$$

Hence, the following constraint:

$$S_{22} + T_{21} + T_{22} \leq I\left(X_2; Y_2|Q\right) \tag{5.75}$$

is redundant for a degraded DMZC of type I. The assertions about the cardinalities of $\mathcal{W}$ and $\mathcal{Q}$ follow from the application of Caratheodory's theorem to the expressions (5.69)-(5.72). □

Remark 5.9. By observing that $I\left(W; Y_1\right) \leq I\left(W; Y_1|X_1\right)$, we readily see that the achievable rate region of Lem. 5.4 will always include the achievable rate region determined by Liu and Ulukus, i.e. $\mathcal{R}_{\mathrm{LU}} \subseteq \mathcal{R}_1$.

5.5.1 Outer bound to the capacity region of the degraded DMZC of type I

The following is an outer bound to the capacity region of the degraded DMZC of type I:

Theorem 5.5. *The set of rate triplets* (R_1, R_{21}, R_{22}) *satisfying*

$$R_{21} \leq I(W; Y_1 | X_1 Q) \tag{5.76}$$

$$R_{22} \leq I(X_2; Y_2 | W Q) \tag{5.77}$$

$$R_1 + R_{21} \leq I(W X_1; Y_1 | Q) \tag{5.78}$$

for some input probability distributions of the form $p(q, w, x_1, x_2) = p(q)\, p(x_1|q)\, p(w|q)\, p(x_2|w, q)$ *constitutes an outer bound to the capacity region of the degraded DMZC of type I. Furthermore, the region is unchanged if we impose the following constraints on the cardinalities of the auxiliary sets:*

$$\|\mathcal{W}\| \leq \|\mathcal{X}_2\| + 1 \quad \textit{and} \quad \|\mathcal{Q}\| \leq 3. \tag{5.79}$$

Proof. Refer to Appendix D.2. □

5.5.2 Achievable Rate Region for the Gaussian ZC with Weak Crossover Link Gain $\left(0 < a^2 < 1\right)$

We have already established an achievable rate region for the degraded DMZC of type I. Lem. 5.4 can then be readily extended to a Gaussian ZC with weak crossover link gain.

Corollary 5.6. *For* $0 < a^2 < 1$*, an achievable rate region for the Gaussian ZC is given by the set* $\mathcal{R}_2$*, which is the closure of the convex hull of all rate triplets*

(R_1, R_{21}, R_{22}) *satisfying*

$$R_1 \le S_1 \tag{5.80}$$

$$R_{21} \le T_{21} \tag{5.81}$$

$$R_{22} \le S_{22} + T_{22} \tag{5.82}$$

where S_1*,* T_{21}*,* S_{22}*, and* T_{22} *are subject to the constraints*

$$S_1 + T_{21} + T_{22} \le \gamma\left(\frac{a^2\beta P_2 + P_1}{a^2(1-\beta)P_2 + 1}\right) \tag{5.83}$$

$$T_{21} + T_{22} \le \gamma\left(\frac{a^2\beta P_2}{a^2(1-\beta)P_2 + 1}\right) \tag{5.84}$$

$$S_1 \le \gamma\left(\frac{P_1}{a^2(1-\beta)P_2 + 1}\right) \tag{5.85}$$

$$S_{22} \le \gamma\left(\left(1-\beta\right)P_2\right) \tag{5.86}$$

for any $0 \le \beta \le 1$ *and where* $\gamma(x) \triangleq \frac{1}{2}\log_2(1+x)$*.*

Proof. The proof follows directly from Lem. 5.4 with $\|\mathcal{Q}\| = 1$, $X_2 = W + V$ where W, V, and X_1 are independent Gaussian random variables, and $\beta = \frac{\mathbb{E}(W^2)}{\mathbb{E}(X_2^2)}$. □

5.6 Rate Regions for the Degraded DMZC of Type II

As we have mentioned in Section 5.2, the capacity region of a Gaussian ZC with strong crossover link gain is equivalent to that of a degraded Gaussian ZC of type II. Hence, we shall first determine an achievable rate region for the degraded DMZC of type II. In addition, the achievable rate region in Lem. 5.7 is also applicable to the degraded DMZC of type III.

Lemma 5.7. *An achievable rate region for sending information over the degraded DMZC of type II and type III* $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ *is given by the* *set $\mathcal{R}_3$, which is the closure of all triplets (R_1, R_{21}, R_{22}) satisfying*

$$R_{21} \leq I(X_2; Y_1 | W X_1 Q) \quad (5.87)$$

$$R_{22} \leq I(W; Y_2 | Q) \quad (5.88)$$

$$R_1 \leq I(X_1; Y_1 | X_2 Q) \quad (5.89)$$

$$R_1 + R_{21} \leq I(X_1 X_2; Y_1 | W Q) \quad (5.90)$$

$$R_1 + R_{21} + R_{22} \leq I(X_1 X_2; Y_1 | Q) \quad (5.91)$$

for all input probability distributions of the form (5.65). Furthermore, the region is unchanged if we impose the following constraints on the cardinalities of the auxiliary sets:

$$\|\mathcal{W}\| \leq \|\mathcal{X}_2\| + 2 \quad \textit{and} \quad \|\mathcal{Q}\| \leq 5. \quad (5.92)$$

Proof. Set $T_{21} = S_{22} = 0$, $R_1 = S_1$, $R_{21} = S_{21}$, $R_{22} = T_{22}$, $U_2 \triangleq W$, and $U_1 \triangleq X_2$ in Thm. 5.3. Since

$$I(W; Y_2 | Q) + I(X_2; Y_1 | W X_1 Q) \leq I(X_2; Y_1 | X_1 Q). \quad (5.93)$$

for a degraded DMZC of type II and type III, the constraint

$$R_{21} + R_{22} \leq I(X_2; Y_1 | X_1 Q) \quad (5.94)$$

is redundant. The assertions about the cardinalities of $\mathcal{W}$ and $\mathcal{Q}$ follow directly from the application of Caratheodory's theorem to the expressions (5.87)-(5.91).

□

5.6.1 Outer bound to the capacity region of the degraded DMZC of type II and type III

The following is an outer bound to the capacity region of the degraded DMZC of type II and type III:

Theorem 5.8. *The set of rate triplets* (R_1, R_{21}, R_{22}) *satisfying*

$$R_{21} \leq I\left(X_2; Y_1 | W X_1 Q\right) \tag{5.95}$$

$$R_{22} \leq I\left(W; Y_2 | Q\right) \tag{5.96}$$

$$R_1 \leq I\left(X_1; Y_1 | X_2 Q\right) \tag{5.97}$$

$$R_1 + R_{21} + R_{22} \leq I\left(X_1 X_2; Y_1 | Q\right) \tag{5.98}$$

for some input probability distributions of the form $p\left(q, w, x_1, x_2\right) = p\left(q\right) p\left(x_1|q\right) p\left(w|q\right) p\left(x_2|w, q\right)$ *constitutes an outer bound to the capacity region of the degraded DMZC of type II and type III. Furthermore, the region is unchanged if we impose the following constraints on the cardinalities of the auxiliary sets:*

$$\|\mathcal{W}\| \leq \|\mathcal{X}_2\| + 1 \quad \textit{and} \quad \|\mathcal{Q}\| \leq 4. \tag{5.99}$$

Proof. Refer to Appendix D.3. □

We note that the outer bound of Thm. 5.8 has one less constraint than the achievable rate region of Lem. 5.7. A natural question is under what conditions do the inner bound and outer bound meet. This is given in the following theorem below:

Theorem 5.9. *The capacity region of the class of DMZC of type II, coupled with the condition that* $I\left(W; Y_1\right) \leq I\left(W; Y_2\right)$ *for all input distributions of the form* $p\left(w, x_1, x_2\right) = p\left(x_1\right) p\left(w, x_2\right)$*, is the set* $\mathcal{R}_3'$*, which is the closure of the set of rate triplets satisfying (5.95)-(5.98) for some input probability distributions of the form* $p\left(q, w, x_1, x_2\right) = p\left(q\right) p\left(x_1|q\right) p\left(w|q\right) p\left(x_2|w, q\right)$*. Furthermore, the region is unchanged if we impose the same constraints on the cardinalities of the auxiliary sets as (5.99).*

Proof. Let us first assume that a certain rate triplet $\left(R_1^{'}, R_{21}^{'}, R_{22}^{'}\right)$ satisfies (5.95)-(5.98) for a fixed input distribution

$$p_1\left(q, w, x_1, x_2\right) = p_1\left(q\right) p_1\left(x_1|q\right) p_1\left(w|q\right) p_1\left(x_2|w, q\right). \tag{5.100}$$

Let the joint distribution of the set of random variables $(Q^1 W^1 X_1^1 X_2^1)$ be given by (5.100). Let the joint distribution of the set of random variables $(Q^2 W^2 X_1^2 X_2^2)$, where $W^2 \triangleq \emptyset$, be given by

$$\begin{aligned} p_2\left(q, x_1, x_2\right) &= \sum_{w \in \mathcal{W}} p_1\left(q, w, x_1, x_2\right) \\ &= p_1\left(q\right) p_1\left(x_1|q\right) p_1\left(x_2|q\right). \end{aligned} \tag{5.101}$$

Now, let the random variable I range over $\{1, 2\}$, where $0 \leq \Pr\left(I = 1\right) = \alpha \leq 1$ and $\Pr\left(I = 2\right) = 1 - \alpha$. Furthermore, we define $X_1 \triangleq X_1^I$, $X_2 \triangleq X_2^I$, $W \triangleq W^I$, and $Q \triangleq \left(Q^I, I\right)$. Next, we need to set an appropriate value for α. If $I\left(W^1; Y_2^1|Q^1\right) = 0$, set $\alpha = 0$. Otherwise, we set α as follows:

$$\alpha = \frac{R_{22}^{'}}{I\left(W^1; Y_2^1|Q^1\right)}. \tag{5.102}$$

We note that $R_{21}^{'}$ satisfies

$$\begin{aligned} R_{21}^{'} &\leq I\left(X_2^1; Y_1^1|W^1 X_1^1 Q^1\right) \\ &= \alpha I\left(X_2^1; Y_1^1|W^1 X_1^1 Q^1\right) + \left(1 - \alpha\right) I\left(X_2^1; Y_1^1|W^1 X_1^1 Q^1\right) \\ &\leq \alpha I\left(X_2^1; Y_1^1|W^1 X_1^1 Q^1\right) + \left(1 - \alpha\right) I\left(X_2^1; Y_1^1|X_1^1 Q^1\right) \\ &= \alpha I\left(X_2^1; Y_1^1|W^1 X_1^1 Q^1\right) + \left(1 - \alpha\right) I\left(X_2^2; Y_1^2|X_1^2 Q^2\right) \\ &= I\left(X_2; Y_1|W X_1 Q\right). \end{aligned} \tag{5.103}$$

We also note that $R_1^{'} + R_{21}^{'}$ satisfies

$$R_1^{'} + R_{21}^{'} \leq I\left(X_1^1 X_2^1; Y_1^1|Q^1\right) - R_{22}^{'}$$

$$
\begin{aligned}
&= I\left(X_1X_2;Y_1|Q\right) - R_{22}^{'} \\
&= I\left(X_1X_2;Y_1|Q\right) - \alpha I\left(W^1;Y_2^1|Q^1\right) \\
&= I\left(X_1X_2;Y_1|Q\right) - \alpha I\left(W^1;Y_2^1|Q^1\right) - (1-\alpha) I\left(W^2;Y_2^2|Q^2\right) \\
&= I\left(X_1X_2;Y_1|Q\right) - I\left(W;Y_2|Q\right) \\
&\leq I\left(X_1X_2;Y_1|Q\right) - I\left(W;Y_1|Q\right) \\
&= I\left(X_1X_2;Y_1|WQ\right) \qquad (5.104)
\end{aligned}
$$

if $I(W;Y_1) \leq I(W;Y_2)$ for all input probability distributions $p(x_1)\,p(w)\,p(x_2|w)$. We see that the same rate triplet $\left(R_1^{'}, R_{21}^{'}, R_{22}^{'}\right)$ satisfies

$$R_{21}^{'} \leq I\left(X_2;Y_1|WX_1Q\right) \qquad (5.105)$$

$$R_{22}^{'} \leq I\left(W;Y_2|Q\right) \qquad (5.106)$$

$$R_1^{'} \leq I\left(X_1;Y_1|X_2Q\right) \qquad (5.107)$$

$$R_1^{'} + R_{21}^{'} \leq I\left(X_1X_2;Y_1|WQ\right) \qquad (5.108)$$

$$R_1^{'} + R_{21}^{'} + R_{22}^{'} \leq I\left(X_1X_2;Y_1|Q\right). \qquad (5.109)$$

Hence, all rate triplets in the set $\mathcal{R}_3^{'}$ are achievable. □

The region $\mathcal{R}_3^{'}$ is in fact also the capacity region of a certain class of degraded DMZC of type I.

Theorem 5.10. *$\mathcal{R}_3^{'}$ is the capacity region of the class of degraded DMZC of type I with Y_2 being a deterministic function of X_1 and Y_1, i.e., $Y_2 = \Lambda(X_1, Y_1)$.*

Proof. Since $Y_2 = \Lambda(X_1, Y_1)$, we note that $X_2 \rightarrow (X_1, Y_1) \rightarrow Y_2$ form a Markov chain. In fact, this special class of ZC is a degraded DMZC of both type I and type II. It is easy to verify that for the degraded DMZC of type I, $I(W;Y_1) \leq I(W;Y_2)$ for all input distributions $p(x_1)\,p(w)\,p(x_2|w)$. □

5.6.2 Achievable Rate Region for the Gaussian ZC with Strong Crossover Link Gain $\left(a^2 \geq 1\right)$

So far, we have established an achievable rate region for the degraded DMZC of type II and type III. Since the capacity region of the Gaussian ZC with strong crossover link gain corresponds to that of a degraded Gaussian ZC of type II, we see that Lem. 5.7 is readily applicable with obvious modifications.

Corollary 5.11. *For $a^2 \geq 1$, an achievable rate region for the Gaussian ZC is given by the set $\mathcal{R}_4$, which is the closure of the convex hull of all rate triplets (R_1, R_{21}, R_{22}) satisfying*

$$R_{21} \leq \gamma\left(a^2 \beta P_2\right) \tag{5.110}$$

$$R_{22} \leq \gamma\left(\frac{(1-\beta)\, P_2}{1+\beta P_2}\right) \tag{5.111}$$

$$R_1 \leq \gamma\left(P_1\right) \tag{5.112}$$

$$R_1 + R_{21} \leq \gamma\left(a^2 \beta P_2 + P_1\right) \tag{5.113}$$

$$R_1 + R_{21} + R_{22} \leq \gamma\left(a^2 P_2 + P_1\right). \tag{5.114}$$

for any $0 \leq \beta \leq 1$.

Proof. The proof follows directly from Lem. 5.7 with $\|\mathcal{Q}\| = 1$. We also assume that $X_2 = W + V$ where W, V, and X_1 are independent, zero-mean Gaussian random variables and where $\beta = \frac{\mathbb{E}(V^2)}{\mathbb{E}(X_2^2)}$. □

Remark 5.10. Corollary 5.11 was derived in [13] for the Gaussian ZC with very strong crossover link gain. We note that the last constraint (5.114) is redundant for the Gaussian ZC with very strong crossover link gain.

5.6.3 Outer Bound to the Capacity Region of the Gaussian ZC with Strong Crossover Link Gain $\left(a^2 \geq 1\right)$

In the previous section, we derived an achievable rate region for the Gaussian ZC with strong crossover link gain. Next, we proceed to establish an outer bound to the capacity region of the Gaussian ZC with strong crossover link gain. We make use of the equivalent channel shown in Fig. 5.7 and Shannon's entropy power inequality to derive an outer bound.

Theorem 5.12. *For a Gaussian ZC with power constraints P_1 and P_2, and $a^2 \geq 1$, any achievable rate triplet (R_1, R_{21}, R_{22}) has to satisfy*

$$R_{21} \leq \gamma\left(a^2\beta P_2\right) \tag{5.115}$$

$$R_{22} \leq \gamma\left(\frac{(1-\beta)\,P_2}{1+\beta P_2}\right) \tag{5.116}$$

$$R_1 \leq \gamma\left(P_1\right) \tag{5.117}$$

$$R_1 + R_{21} + R_{22} \leq \gamma\left(a^2 P_2 + P_1\right). \tag{5.118}$$

for some $0 \leq \beta \leq 1$.

Proof. Refer to Appendix D.4. □

5.6.4 Capacity Region of the Gaussian ZC with Moderately Strong Crossover Link Gain $(1 \leq a^2 \leq 1 + P_1)$

We have derived an achievable rate region and an outer bound for the Gaussian ZC when $a^2 \geq 1$. In this section, we show that the achievable rate region coincides with the outer bound when the crossover link gain is moderately strong, i.e, when $1 \leq a^2 \leq 1 + P_1$.

Theorem 5.13. *The capacity region of the Gaussian ZC with moderately strong crossover link gain is given by the closure of all rate triplets (R_1, R_{21}, R_{22}) satisfying (5.115)-(5.118) for some* $0 \leq \beta \leq 1$.

Proof. Let us first assume a particular rate triplet $\left(R_1^{'}, R_{21}^{'}, R_{22}^{'}\right)$ satisfies (5.115)-(5.118) for $\beta = \beta_0$. Next, let us set β_1 as follows:

$$\begin{aligned} \beta_1 &= \frac{\frac{1+P_2}{2^{2R_{22}^{'}}} - 1}{P_2} \\ \Longrightarrow R_{22}^{'} &= \gamma\left(\frac{(1-\beta_1)\,P_2}{1+\beta_1 P_2}\right). \end{aligned} \tag{5.119}$$

We note that $\beta_0 \leq \beta_1 \leq 1$ since $R_{22}^{'} \leq \gamma\left(\frac{(1-\beta_0)P_2}{1+\beta_0 P_2}\right)$. Let us consider the last constraint given by (5.118). We obtain

$$\begin{aligned} R_1^{'} + R_{21}^{'} &\leq \frac{1}{2}\log_2\left(1 + a^2 P_2 + P_1\right) - \frac{1}{2}\log_2\left(\frac{1+P_2}{1+\beta_1 P_2}\right) \\ &= \frac{1}{2}\log_2\left(\frac{1 + a^2 P_2 + P_1 + \beta_1 P_2 + a^2\beta_1 {P_2}^2 + \beta_1 P_1 P_2}{1+P_2}\right) \\ &= \gamma\left(a^2\beta_1 P_2 + P_1 + \frac{(1-\beta_1)\,P_2\,(a^2 - 1 - P_1)}{1+P_2}\right) \\ &\leq \gamma\left(a^2\beta_1 P_2 + P_1\right), \quad a^2 \leq 1 + P_1. \end{aligned} \tag{5.120}$$

We see that the same rate triplet $\left(R_1^{'}, R_{21}^{'}, R_{22}^{'}\right)$ also satisfies (5.110)-(5.114) for $\beta = \beta_1$.

$$R_{21}^{'} \leq \gamma\left(a^2\beta_0 P_2\right) \leq \gamma\left(a^2\beta_1 P_2\right), \quad \beta_1 \geq \beta_0 \tag{5.121}$$

$$R_{22}^{'} = \gamma\left(\frac{(1-\beta_1)\,P_2}{1+\beta_1 P_2}\right) \tag{5.122}$$

$$R_1^{'} \leq \gamma\left(P_1\right) \tag{5.123}$$

$$R_1^{'} + R_{21}^{'} \leq \gamma\left(a^2\beta_1 P_2 + P_1\right) \tag{5.124}$$

$$R_1^{'} + R_{21}^{'} + R_{22}^{'} \leq \gamma\left(a^2 P_2 + P_1\right). \tag{5.125}$$

Hence, any rate triplet in the outer bound is achievable. □

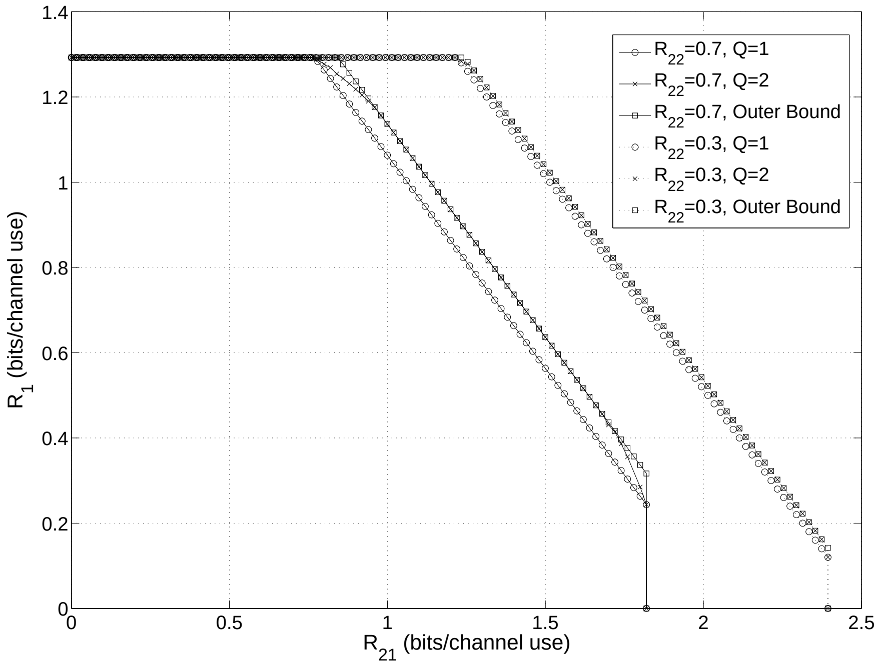

Figure 5.9: Numerical Computations ($P_1 = 5$, $P_2 = 5$, $a^2 = 9$, $R_{22} = 0.3/0.7$)

5.6.5 Achievable Rates for the Gaussian ZC with Very Strong Crossover Link Gain ($a^2 \geq P_1 + 1$)

In [13], Vishwanath, Jindal, and Goldsmith determined an achievable rate region for very strong crossover link gain using superposition coding at sender TX_2 and successive decoding at receiver RX_1. In fact, the achievable rate region of Vishwanath, Jindal, and Goldsmith corresponds to that of Corollary 5.11 with very strong crossover link gain. However, their technique does not apply to the case of moderately strong crossover link gain. This is because their successive decoding method would require receiver RX_1 to be able to decode all the information intended for receiver RX_2. This is possible only with very strong crossover link gain.

We have already determined the capacity of the Gaussian ZC with moderately strong crossover link gain. A very natural question that comes to mind is whether Corollary 5.11 also gives us the capacity region of the Gaussian ZC with very strong crossover link gain. Our experience with the Gaussian ZIFC under very strong interference may influence one to think that Corollary 5.11 would also give us the capacity region of the Gaussian ZC with very strong crossover link gain. However, in this section, we show that this is not the case in general. In fact, this is suggested by the time-sharing random variable Q in the converse proof of [13]. We can enlarge the achievable rate region of Corollary 5.11 for the Gaussian ZC with very strong crossover link gain by allowing $\|\mathcal{Q}\| > 1$. We could theoretically compute an achievable rate region for larger values of $\|\mathcal{Q}\|$ but for computational reasons, we restrict attention to $\|\mathcal{Q}\| = 2$.

Corollary 5.14. *For $a^2 \geq 1 + P_1$, an achievable rate region for the Gaussian ZC is given by the set $\mathcal{R}_5$, which is the closure of the convex hull of all (R_1, R_{21}, R_{22}) triplets satisfying*

$$R_{21} \leq \lambda.\gamma\left(\frac{a^2\beta\rho P_2}{\lambda}\right) + \overline{\lambda}.\gamma\left(\frac{a^2\sigma\overline{\rho}P_2}{\overline{\lambda}}\right) \tag{5.126}$$

$$R_{22} \leq \lambda.\gamma\left(\frac{\overline{\beta}\rho P_2}{\lambda + \beta\rho P_2}\right) + \overline{\lambda}.\gamma\left(\frac{\overline{\sigma}\rho P_2}{\overline{\lambda} + \sigma\overline{\rho}P_2}\right) \tag{5.127}$$

$$R_1 \leq \lambda.\gamma\left(\frac{\alpha P_1}{\lambda}\right) + \overline{\lambda}.\gamma\left(\frac{\overline{\alpha}P_1}{\overline{\lambda}}\right) \tag{5.128}$$

$$R_1 + R_{21} \leq \lambda.\gamma\left(\frac{a^2\beta\rho P_2 + \alpha P_1}{\lambda}\right) + \overline{\lambda}.\gamma\left(\frac{a^2\sigma\overline{\rho}P_2 + \overline{\alpha}P_1}{\overline{\lambda}}\right) \tag{5.129}$$

$$R_1 + R_{21} + R_{22} \leq \lambda.\gamma\left(\frac{a^2\rho P_2 + \alpha P_1}{\lambda}\right) + \overline{\lambda}.\gamma\left(\frac{a^2\overline{\rho}P_2 + \overline{\alpha}P_1}{\overline{\lambda}}\right). \tag{5.130}$$

for any $0 \leq \lambda, \alpha, \beta, \rho, \sigma \leq 1$.

Proof. The result follows directly from Lem. 5.7 with $\|\mathcal{Q}\| = 2$. We assume that $X_2^N = W^N + V^N$ where W^N, V^N, and X_1^N are independent. During a fraction λ of the time, the symbols of X_1^N, W^N, and V^N are Gaussian distributed with zero mean, and variances $\frac{\alpha P_1}{\lambda}$, $\frac{\overline{\beta}\rho P_2}{\lambda}$ and $\frac{\beta\rho P_2}{\lambda}$, respectively:

$$X_{1n} \sim \mathcal{N}\left(0, \frac{\alpha P_1}{\lambda}\right), \quad W_n \sim \mathcal{N}\left(0, \frac{\overline{\beta}\rho P_2}{\lambda}\right),$$

$$V_n \sim \mathcal{N}\left(0, \frac{\beta\rho P_2}{\lambda}\right), \quad 0 \leq \alpha, \beta, \rho \leq 1, \quad n = 1, 2, ..., N\lambda \tag{5.131}$$

and during the remaining fraction $\overline{\lambda} \triangleq 1 - \lambda$ of the time:

$$\begin{aligned} X_{1n} &\sim \mathcal{N}\left(0, \frac{\overline{\alpha} P_1}{\overline{\lambda}}\right), \quad W_n \sim \mathcal{N}\left(0, \frac{\overline{\sigma}\rho P_2}{\overline{\lambda}}\right), \\ V_n &\sim \mathcal{N}\left(0, \frac{\sigma\overline{\rho} P_2}{\overline{\lambda}}\right), \quad 0 \leq \sigma \leq 1, \quad n = N\lambda + 1, ..., N \end{aligned} \tag{5.132}$$

which ensures that the power constraints are satisfied. □

Remark 5.11. Fig. 5.9 shows numerical computations of the achievable rates for the Gaussian ZC with $P_1 = 5$, $P_2 = 5$, $a^2 = 9$ ($a^2 > 1 + P_1$). Instead of plotting rate triplets (R_1, R_{21}, R_{22}), we fix $R_{22} = \{0.3, 0.7\}$ and plot the rate pair (R_1, R_{21}). From Fig. 5.9, we see that when R_{22} is fixed, Corollary 5.11 gives rate pairs (R_1, R_{21}) that correspond to a Gaussian multiple-access channel. However, we see that when we increase $\|\mathcal{Q}\|$ from 1 to 2, Corollary 5.14 gives an achievable rate region that is even larger than that of Corollary 5.11 for the Gaussian ZC with very strong crossover link gain. Moreover, we note that for the parameters chosen, setting $\|\mathcal{Q}\| = 2$ suffices to achieve the capacity for most rate triplets. In general, Corollary 5.11 is not the capacity region of the Gaussian ZC with very strong crossover link gain.

Remark 5.12. However, Corollary 5.11 gives us the capacity region of the Gaussian ZIFC under strong interference. We can ignore the constraint for R_{21} since $R_{21} = 0$ for an interference channel. By setting $\beta = 0$, we obtain the capacity region of the Gaussian ZIFC [15] under strong interference.

Chapter 6

Conclusion and future work

We have taken an information-theoretic look at three non-centralized multi-user communication systems: the relay channel, the interference channel (IFC), and the "Z"-channel (ZC).

For the general relay channel, we introduced and studied three new generalized strategies: The first strategy makes use of sequential backward decoding, the second strategy makes use of simultaneous backward decoding, and the third strategy makes use of sliding window decoding. The advantage of the sliding window decoding strategy is that the receiver can start decoding information without waiting for the last block to be transmitted. However, backward decoding strategies simplify proofs for achievable rates. Assuming zero-mean, jointly Gaussian random variables, all three strategies give higher achievable rates than Cover & El Gamal's generalized strategy for certain parameters of the Gaussian relay channel. In fact, we show that all three of our strategies achieve the same rate. Interestingly, a change in the decoding order resulted in a new achievable rate for the relay channel as shown by the sliding window decoding strategy. This suggests the variation of decoding order in order to obtain new achievable rates for other channels.

We have also extended the rate achievable for SeqBack decoding to the relay channel with standard alphabets. Future research for the relay channel should look into extending the result for even more general cases (less restrictions on the input probability distributions). This could be done by quantization of the relay channel with general alphabets. In other words, one must prove that there exists a suitable partitioning of the alphabets (which can be made finer and finer) such that arbitrarily small probabilities of error can be achieved using the quantized relay channel [60]. Another possible method is to prove Feinstein's Lemma [61] for relay channels with general alphabets. Hence, the mutual information can be expressed directly in the form of divergence.

For the general IFC, a simplified description of the Han-Kobayashi rate region was established. Using this result, we proved the equivalence between the Han-Kobayashi rate region and the Chong-Motani-Garg region. Moreover, a tighter bound for the cardinality of the time-sharing auxiliary random variable also emerged from our simplified description. We then make use of our simplified description to establish the capacity region of a class of discrete memoryless IFC before extending the result to the same class of IFCs, where both transmitters now have a common message to transmit. Interestingly, the simplified description of the Han-Kobayashi rate region first started off with a new coding strategy based on the broadcast channel. Even though no new achievable rate region was obtained, this work lead to a simplified description. This suggests the use of different coding strategies to simplify rate regions already established for other channels. This work also revealed the importance of Fourier-Motzkin elimination in removing redundant inequalities in the description of rate regions.

For the two-user ZC, we studied both the discrete memoryless ZC and the Gaussian ZC. We established achievable rates for the general discrete memoryless ZC. We then specialized the rates obtained to two different types of degraded discrete memoryless ZCs and also derived respective outer bounds to their capacity regions. We showed that as long as a certain condition is satisfied, the achievable rate region is the capacity region for one type of degraded discrete memoryless ZC. The results were then extended to the two-user Gaussian ZC with different crossover link gains. We determined an outer bound to the capacity region of the Gaussian ZC with strong crossover link gain and established the capacity region for moderately strong crossover link gain.

Future research for the both the IFC and the ZC should look at incorporating Marton's strategy for the broadcast channel [8] into the coding strategies.

Appendix A

Proof of Theorems in Chapter 2

A.1 Derivation of (2.4)

The proof for (2.6) follows exactly along the same lines. Hence, we only show the explicit derivation of (2.4). Let $\alpha = \frac{\mathbb{E}[W^2]}{P_1}$. The relay output Y_2 is given by

$$\begin{aligned} Y_2 &= h_0 X_1 + Z_2 \\ &= a h_0 X_2 + h_0 W + Z_2. \end{aligned} \tag{A.1}$$

The destination output Y_3 is given by

$$\begin{aligned} Y_3 &= h_1 X_1 + h_2 X_2 + Z_3 \\ &= \left(a h_1 + h_2\right) X_2 + h_1 W + Z_3. \end{aligned} \tag{A.2}$$

Since Y_3 is a zero-mean Gaussian random variable, the variance of Y_3 is given by

$$\begin{aligned} \mathbb{E}\left(Y_3^2\right) &= h_1^2 \mathbb{E}\left[X_1^2\right] + h_2^2 \mathbb{E}\left[X_2^2\right] + 2 h_1 h_2 \mathbb{E}\left[X_1 X_2\right] + \mathbb{E}\left[Z_3^2\right] \\ &= h_1^2 P_1 + h_2^2 P_2 + 2 h_1 h_2 \sqrt{\left(1-\alpha\right) P_1 P_2} + \sigma_3^2. \end{aligned} \tag{A.3}$$

Hence, we have

$$h\left(Y_3\right)=\frac{1}{2}\log_2\left(2\pi e\left(h_1^2P_1+h_2^2P_2+2h_1h_2\sqrt{\left(1-\alpha\right)P_1P_2}+\sigma_3^2\right)\right). \qquad \text{(A.4)}$$

We may compute the first term of the cut-set upper bound as follows:

$$\begin{aligned}I\left(X_1X_2;Y_3\right)&=h\left(Y_3\right)-h\left(Y_3|X_1X_2\right)\\&=h\left(Y_3\right)-h\left(Z_3\right)\\&=\frac{1}{2}\log_2\left(1+\frac{h_1^2P_1+h_2^2P_2+2h_1h_2\sqrt{\left(1-\alpha\right)P_1P_2}}{\sigma_3^2}\right).\end{aligned}$$

Next, let us consider

$$\begin{aligned}h\left(Y_2Y_3|X_2\right)&=h\left(h_0W+Z_2,h_1W+Z_3\right)\\&=\frac{1}{2}\log_2\left(\left(2\pi e\right)^2\begin{vmatrix}h_0^2\alpha P_1+\sigma_2^2 & h_0h_1\alpha P_1\\ h_0h_1\alpha P_1 & h_1^2\alpha P_1+\sigma_3^2\end{vmatrix}\right)\\&=\frac{1}{2}\log_2\left(\left(2\pi e\right)^2\left(\alpha P_1\left(h_0^2\sigma_3^2+h_1^2\sigma_2^2\right)+\sigma_2^2\sigma_3^2\right)\right). \qquad \text{(A.5)}\end{aligned}$$

Finally, we may compute the second term of the cut-set upper bound as follows:

$$\begin{aligned}I\left(X_1;Y_2Y_3|X_2\right)&=h\left(Y_2Y_3|X_2\right)-h\left(Y_2Y_3|X_1X_2\right)\\&=h\left(Y_2Y_3|X_2\right)-h\left(Z_2Z_3\right)\\&=\frac{1}{2}\log_2\left(1+\alpha P_1\left(\frac{h_0^2}{\sigma_2^2}+\frac{h_1^2}{\sigma_3^2}\right)\right).\end{aligned}$$

A.2 Derivation of (2.10) and (2.11)

The relay output and the destination output is given by (2.1) and (2.2) respectively. We have

$$I\left(X_2;Y_3\right)=h\left(Y_3\right)-h\left(Y_3|X_2\right)$$

$$= h\left(Y_3\right) - h\left(h_1 X_1 + Z_3\right)$$
$$= \frac{1}{2}\log_2\left(1 + \frac{h_2^2 P_2}{h_1^2 P_1 + \sigma_3^2}\right). \tag{A.6}$$

Since $I\left(Y_2; \hat{Y}_2 | X_2 Y_3\right) = h\left(Y_2 | X_2 Y_3\right) - h\left(Y_2 | X_2 \hat{Y}_2 Y_3\right)$, let us consider

$$\begin{aligned} h\left(Y_2|X_2Y_3\right) &= h\left(h_0X_1 + Z_2|h_1X_1 + Z_3\right) \\ &= h\left(h_0X_1 + Z_2, h_1X_1 + Z_3\right) - h\left(h_1X_1 + Z_3\right) \\ &= \frac{1}{2}\log_2\left(2\pi e\frac{\begin{vmatrix} h_0^2P_1 + \sigma_2^2 & h_0h_1P_1 \\ h_0h_1P_1 & h_1^2P_1 + \sigma_3^2 \end{vmatrix}}{h_1^2P_1 + \sigma_3^2}\right) \\ &= \frac{1}{2}\log_2\left(2\pi e\frac{h_1^2P_1\sigma_2^2 + h_0^2P_1\sigma_3^2 + \sigma_2^2\sigma_3^2}{h_1^2P_1 + \sigma_3^2}\right). \end{aligned} \tag{A.7}$$

We also consider

$$\begin{aligned} h\left(Y_2|X_2\hat{Y}_2Y_3\right) &= h\left(h_0X_1 + Z_2|h_1X_1 + Z_3, h_0X_1 + Z_2 + Z_W\right) \\ &= \frac{1}{2}\log_2\left(2\pi e\frac{\begin{vmatrix} h_0^2P_1 + \sigma_2^2 & h_0h_1P_1 & h_0^2P_1 + \sigma_2^2 \\ h_0h_1P_1 & h_1^2P_1 + \sigma_3^2 & h_0h_1P_1 \\ h_0^2P_1 + \sigma_2^2 & h_0h_1P_1 & h_0^2P_1 + \sigma_2^2 + \sigma_W^2 \end{vmatrix}}{\begin{vmatrix} h_1^2P_1 + \sigma_3^2 & h_0h_1P_1 \\ h_0h_1P_1 & h_0^2P_1 + \sigma_2^2 + \sigma_W^2 \end{vmatrix}}\right) \\ &= \frac{1}{2}\log_2\left(2\pi e\frac{h_1^2P_1\sigma_2^2\sigma_W^2 + h_0^2P_1\sigma_3^2\sigma_W^2 + \sigma_2^2\sigma_3^2\sigma_W^2}{h_1^2P_1\sigma_2^2 + h_0^2P_1\sigma_3^2 + \sigma_2^2\sigma_3^2 + h_1^2P_1\sigma_W^2 + \sigma_3^2\sigma_W^2}\right). \end{aligned} \tag{A.8}$$

From (A.6), (A.7), and (A.8), we can show that the constraint (2.9) is satisfied if

$$\sigma_W^2 \geq \frac{h_1^2P_1\sigma_2^2 + h_0^2P_1\sigma_3^2 + \sigma_2^2\sigma_3^2}{h_2^2P_2}.$$

Finally, we can compute the rate as follows:

$$\begin{aligned}
I\left(X_1;\hat{Y}_2Y_3|X_2\right) &= h\left(\hat{Y}_2Y_3|X_2\right) - h\left(\hat{Y}_2Y_3|X_1X_2\right)\\
&= h\left(h_0X_1 + Z_2 + Z_W, h_1X_1 + Z_3\right) - h\left(Z_2 + Z_W, Z_3\right)\\
&= \frac{1}{2}\log_2\left(\frac{\begin{vmatrix} h_0^2P_1 + \sigma_2^2 + \sigma_W^2 & h_0h_1P_1 \\ h_0h_1P_1 & h_1^2P_1 + \sigma_3^2 \end{vmatrix}}{\begin{vmatrix} \sigma_2^2 + \sigma_W^2 & 0 \\ 0 & \sigma_3^2 \end{vmatrix}}\right)\\
&= \frac{1}{2}\log_2\left(1 + P_1\left(\frac{h_1^2}{\sigma_3^2} + \frac{h_0^2}{\sigma_2^2 + \sigma_W^2}\right)\right).
\end{aligned}$$

A.3 Derivation of (2.19)-(2.23)

We can compute $I\left(X_1;\hat{Y}_2Y_3|UX_2\right)$ as follows:

$$\begin{aligned}
I\left(X_1;\hat{Y}_2Y_3|UX_2\right) &= h\left(\hat{Y}_2Y_3|UX_2\right) - h\left(\hat{Y}_2Y_3|UX_1X_2\right)\\
&= h\left(h_0W_1 + Z_2 + Z_W, h_1W_1 + Z_3\right) - h\left(Z_2 + Z_W, Z_3\right)\\
&= \frac{1}{2}\log_2\left(\frac{\begin{vmatrix} h_0^2\alpha P_1 + \sigma_2^2 + \sigma_W^2 & h_0h_1\alpha P_1 \\ h_0h_1\alpha P_1 & h_1^2\alpha P_1 + \sigma_3^2 \end{vmatrix}}{\begin{vmatrix} \sigma_2^2 + \sigma_W^2 & 0 \\ 0 & \sigma_3^2 \end{vmatrix}}\right)\\
&= \frac{1}{2}\log_2\left(1 + \alpha P_1\left(\frac{h_1^2}{\sigma_3^2} + \frac{h_0^2}{\sigma_2^2 + \sigma_W^2}\right)\right).
\end{aligned}$$

We can compute $I\left(U;Y_2|VX_2\right)$ as follows:

$$\begin{aligned}
I\left(U;Y_2|VX_2\right) &= h\left(Y_2|VX_2\right) - h\left(Y_2|UVX_2\right)\\
&= h\left(bh_0W_0 + h_0W_1 + Z_2\right) - h\left(h_0W_1 + Z_2\right)\\
&= \frac{1}{2}\log_2\left(1 + \frac{h_0^2\beta\overline{\alpha}P_1}{h_0^2\alpha P_1 + \sigma_2^2}\right).
\end{aligned}$$

We can compute $I\left(X_1X_2;Y_3\right)$ as follows:

$$\begin{aligned}
I\left(X_1X_2;Y_3\right) &= h\left(Y_3\right) - h\left(Y_3|X_1X_2\right) \\
&= h\left(Y_3\right) - h\left(Z_3\right) \\
&= \frac{1}{2}\log_2\left(1 + \frac{h_1^2P_1 + h_2^2P_2 + 2h_1h_2\sqrt{\overline{\alpha}\overline{\beta}\overline{\gamma}P_1P_2}}{\sigma_3^2}\right).
\end{aligned}$$

We can compute $I\left(Y_2;\hat{Y}_2|UX_1X_2Y_3\right)$ as follows:

$$\begin{aligned}
I\left(Y_2;\hat{Y}_2|UX_1X_2Y_3\right) &= I\left(Z_2 + Z_W; Z_2\right) \\
&= \frac{1}{2}\log_2\left(1 + \frac{\sigma_2^2}{\sigma_W^2}\right).
\end{aligned}$$

Next, let us consider

$$h\left(h_0W_1 + Z_2 + Z_W|h_1W_1 + Z_3\right) = \frac{1}{2}\log_2\left(2\pi e\frac{\begin{vmatrix} h_0^2\alpha P_1 + \sigma_2^2 + \sigma_W^2 & h_0h_1\alpha P_1 \\ h_0h_1\alpha P_1 & h_1^2\alpha P_1 + \sigma_3^2 \end{vmatrix}}{h_1^2\alpha P_1 + \sigma_3^2}\right) \tag{A.9}$$

and

$$\begin{aligned}
&h\left(h_0W_1 + Z_2 + Z_W|h_0W_1 + Z_2, h_1W_1 + Z_3\right) \\
&= \frac{1}{2}\log_2\left(2\pi e\frac{\begin{vmatrix} h_0^2\alpha P_1 + \sigma_2^2 + \sigma_W^2 & h_0^2\alpha P_1 + \sigma_2^2 & h_0h_1\alpha P_1 \\ h_0^2\alpha P_1 + \sigma_2^2 & h_0^2\alpha P_1 + \sigma_2^2 & h_0h_1\alpha P_1 \\ h_0h_1\alpha P_1 & h_0h_1\alpha P_1 & h_1^2\alpha P_1 + \sigma_3^2 \end{vmatrix}}{\begin{vmatrix} h_0^2\alpha P_1 + \sigma_2^2 & h_0h_1\alpha P_1 \\ h_0h_1\alpha P_1 & h_1^2\alpha P_1 + \sigma_3^2 \end{vmatrix}}\right).
\end{aligned} \tag{A.10}$$

From (A.9) and (A.10), we have

$$\begin{aligned} I\left(\hat{Y}_2; Y_2|UX_2Y_3\right) &= I\left(h_0W_1 + Z_2 + Z_W; h_0W_1 + Z_2|h_1W_1 + Z_3\right) \\ &= h\left(h_0W_1 + Z_2 + Z_W|h_1W_1 + Z_3\right) \\ &\quad - h\left(h_0W_1 + Z_2 + Z_W|h_0W_1 + Z_2, h_1W_1 + Z_3\right) \\ &= \frac{1}{2}\log_2\left(1 + \frac{\alpha P_1\left(h_1^2\sigma_2^2 + h_0^2\sigma_3^2\right) + \sigma_2^2\sigma_3^2}{\sigma_W^2\left(\alpha h_1^2 P_1 + \sigma_3^2\right)}\right). \end{aligned} \tag{A.11}$$

We also compute $I\left(X_2; Y_3|V\right)$ as follows:

$$\begin{aligned} I\left(X_2; Y_3|V\right) &= h\left(Y_3|V\right) - h\left(Y_3|V, X_2\right) \\ &= h\left(bh_1W_0 + h_1W_1 + h_2W_2 + Z_3\right) - h\left(bh_1W_0 + h_1W_1 + Z_3\right) \\ &= \frac{1}{2}\log_2\left(1 + \frac{h_2^2\gamma P_2}{h_1^2\overline{\alpha}\beta P_1 + h_1^2\alpha P_1 + \sigma_3^2}\right). \end{aligned} \tag{A.12}$$

Finally, from (A.11) and (A.12), the constraint (2.14) is satisfied if

$$\sigma_W^2 \geq \frac{\left[\alpha h_1^2 P_1\sigma_2^2 + \sigma_3^2\left(\alpha h_0^2 P_1 + \sigma_2^2\right)\right]\left[\left(\beta - \alpha\beta + \alpha\right)h_1^2 P_1 + \sigma_3^2\right]}{\gamma h_2^2 P_2\left(\alpha h_1^2 P_1 + \sigma_3^2\right)}.$$

Appendix B

Proof of Theorems in Chapter 3

B.1 Detailed Computation of the Probabilities of error

We may bound $\overline{\Upsilon^N\left(\mathcal{K}_j^c|x_1^N(i,j,k),x_2^N(i,l)\right)}$ as follows:

$$\begin{aligned}
&\int \Upsilon^N\left(\left\{y_2^N:\left(q^N,v^N,u^N,x_2^N,y_2^N\right)\notin A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)\right\}|x_1^N,x_2^N\right)dP_{Q^NU^NV^NX_1^NX_2^N}\\
&=\int\left[\int_{\left\{y_2^N:\left(q^N,v^N,u^N,x_2^N,y_2^N\right)\notin A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)\right\}}dP_{Y_2^N|X_1^NX_2^N}\right]dP_{Q^NU^NV^NX_1^NX_2^N}\\
&=\int_{\left\{\left(q^N,u^N,v^N,x_1^N,x_2^N,y_2^N\right):\left(q^N,v^N,u^N,x_2^N,y_2^N\right)\notin A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)\right\}}dP_{Q^NU^NV^NX_1^NX_2^NY_2^N}\\
&=\int_{\left(A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)\right)^c}\left[\int dP_{X_1^N|Q^NU^NV^NX_2^NY_2^N}\right]dP_{Q^NU^NV^NX_2^NY_2^N}\\
&=\int_{\left(A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)\right)^c}dP_{Q^NU^NV^NX_2^NY_2^N}\\
&=P_{Q^NU^NV^NX_2^NY_2^N}\left(\left(A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)\right)^c\right)\\
&\le\epsilon.
\end{aligned} \tag{B.1}$$

We may bound the term $\overline{\Upsilon^N\left(\mathcal{K}_{\hat{j}}|x_1^N(i,j,k),x_2^N(i,l)\right)}$ by

$$\begin{aligned}
&\int \Upsilon^N\left(\left\{y_2^N:\left(q^N,v^N,u^{N\prime},x_2^N,y_2^N\right)\in A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)\right\}|x_1^N,x_2^N\right)dP_{Q^NU^NU^{N\prime}V^NX_1^NX_2^N}\\
&=\int\left[\int_{\left\{y_2^N:\left(q^N,v^N,u^{N\prime},x_2^N,y_2^N\right)\in A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)\right\}}dP_{Y_2^N|X_1^NX_2^N}\right]dP_{Q^NU^NU^{N\prime}V^NX_1^NX_2^N}\\
&=\int_{\left\{\left(q^N,u^N,u^{N\prime},v^N,x_1^N,x_2^N,y_2^N\right):\left(q^N,v^N,u^{N\prime},x_2^N,y_2^N\right)\in A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)\right\}}dP_{Q^NU^NU^{N\prime}V^NX_1^NX_2^NY_2^N}\\
&=\int_{A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)}\left[\int dP_{U^NX_1^N|Q^NV^NX_2^NY_2^N}\right]f\left(q^N,v^N\right)f\left(u^{N\prime}|q^N,v^N\right)f\left(x_2^N,y_2^N|q^N,v^N\right)
\end{aligned}$$

$$
\begin{aligned}
& \times dM_{Q^N V^N U^{N'} X_2^N Y_2^N} \\
&= \int_{A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)} f\left(q^N, v^N\right) f\left(u^{N'}|q^N, v^N\right) f\left(x_2^N, y_2^N|q^N, v^N\right) dM_{Q^N V^N U^{N'} X_2^N Y_2^N} \\
&\leq \int_{A_\epsilon^{(N)}(Q,V,U,X_2,Y_2)} 2^{-N\left(H_{P||M}(QV)-\epsilon\right)} 2^{-N\left(H_{P||M}(U|QV)-2\epsilon\right)} 2^{-N\left(H_{P||M}(X_2Y_2|QV)-2\epsilon\right)} \\
& \times dM_{Q^N V^N U^{N'} X_2^N Y_2^N} \\
&\overset{(a)}{=} 2^{-N\left(H_{P||M}(QV)-\epsilon\right)} 2^{-N\left(H_{P||M}(U|QV)-2\epsilon\right)} 2^{-N\left(H_{P||M}(X_2Y_2|QV)-2\epsilon\right)} \mathrm{vol}\left(A_\epsilon^{(N)}\left(Q,V,U,X_2,Y_2\right)\right) \\
&\leq 2^{-N\left(H_{P||M}(QV)-\epsilon\right)} 2^{-N\left(H_{P||M}(U|QV)-2\epsilon\right)} 2^{-N\left(H_{P||M}(X_2Y_2|QV)-2\epsilon\right)} 2^{N\left(H_{P||M}(QVUX_2Y_2)+\epsilon\right)} \\
&= 2^{-N(I(U;Y_2|VX_2Q)-6\epsilon)} \qquad \text{(B.2)}
\end{aligned}
$$

where $P_{QVUU'X_1X_2} = P_{U'\times UX_1X_2|QV}$ and (a) follows from $M_{Q^N V^N U^N X_2^N Y_2^N} = M_{U^N \times X_2^N Y_2^N|Q^N V^N} = M_{Q^N V^N U^{N'} X_2^N Y_2^N}$. Next, $\overline{\Upsilon^N\left(\mathcal{L}_{\hat{i}}|x_1^N(i,j,k), x_2^N(i,l)\right)}$ may be bounded by

$$
\begin{aligned}
& \int \Upsilon^N\left(\left\{y_3^N : \left(q^N, v^{N'}, u^{N'}, y_3^N\right) \in A_\epsilon^{(N)}\left(Q,V,U,Y_3\right)\right\}|x_1^N, x_2^N\right) dP_{Q^N V^{N'} U^{N'} V^N U^N X_1^N X_2^N} \\
&= \int \left[\int_{\left\{y_3^N : \left(q^N, v^{N'}, u^{N'}, y_3^N\right) \in A_\epsilon^{(N)}(Q,V,U,Y_3)\right\}} dP_{Y_3^N|X_1^N X_2^N}\right] dP_{Q^N V^{N'} U^{N'} V^N U^N X_1^N X_2^N} \\
&= \int_{\left\{\left(q^N, v^{N'}, u^{N'}, v^N, u^N, x_1^N, x_2^N, y_3^N\right) : \left(q^N, v^{N'}, u^{N'}, y_3^N\right) \in A_\epsilon^{(N)}(Q,V,U,Y_3)\right\}} dP_{Q^N V^{N'} U^{N'} V^N U^N X_1^N X_2^N Y_3^N} \\
&= \int_{A_\epsilon^{(N)}(Q,V,U,Y_3)} \left[\int dP_{V^N U^N X_1^N X_2^N|Q^N Y_3^N}\right] dP_{Q^N V^{N'} U^{N'} Y_3^N} \\
&= \int_{A_\epsilon^{(N)}(Q,V,U,Y_3)} f\left(q^N\right) f\left(v^{N'}, u^{N'}|q^N\right) f\left(y_3^N|q^N\right) dM_{Q^N V^{N'} U^{N'} Y_3^N} \\
&\leq \int_{A_\epsilon^{(N)}(Q,V,U,Y_3)} 2^{-N\left(H_{P||M}(Q)-\epsilon\right)} 2^{-N\left(H_{P||M}(Y_3|Q)-2\epsilon\right)} 2^{-N\left(H_{P||M}(VU|Q)-2\epsilon\right)} dM_{Q^N V^{N'} U^{N'} Y_3^N} \\
&\overset{a}{=} 2^{-N\left(H_{P||M}(Q)-\epsilon\right)} 2^{-N\left(H_{P||M}(Y_3|Q)-2\epsilon\right)} 2^{-N\left(H_{P||M}(VU|Q)-2\epsilon\right)} \mathrm{vol}\left(A_\epsilon^{(N)}\left(Q,V,U,Y_3\right)\right) \\
&\leq 2^{-N\left(H_{P||M}(Q)-\epsilon\right)} 2^{-N\left(H_{P||M}(Y_3|Q)-2\epsilon\right)} 2^{-N\left(H_{P||M}(VU|Q)-2\epsilon\right)} 2^{N\left(H_{P||M}(QVUY_3)+\epsilon\right)} \\
&= 2^{-N(I(UV;Y_3|Q)-6\epsilon)} \qquad \text{(B.3)}
\end{aligned}
$$

where $P_{Q^N V^{N'} U^{N'} V^N U^N X_1^N X_2^N} = P_{V^{N'} U^{N'} \times V^N U^N X_1^N X_2^N|Q^N}$ and (a) follows from $M_{Q^N V^N U^N Y_3^N} = M_{U^N V^N \times Y_3^N|Q^N} = M_{Q^N V^{N'} U^{N'} Y_3^N}$. In order to bound the term $\overline{\Upsilon^N\left(\mathcal{L}_{\hat{l}}|x_1^N(i,j,k), x_2^N(i,l)\right)}$, we observe that since we are averaging over all codebooks and mapping functions for block b, $z_b = \Theta_1\left(q^N, u^N(i,j), x_2^N(i,l), y_2^N\right)$ may be set to any fixed index m_0. Hence, we obtain

$$
\begin{aligned}
& \overline{\Upsilon^N\left(\left\{\left(y_2^N, y_3^N\right) : \left(q^N, v^N(i), u^N(i,j), x_2^N\left(i,\hat{l}\right), \hat{y}_2^N\left(i,j,\hat{l},z_b\right), y_3^N(b)\right) \in A_\epsilon^{(N)}\right\}|x_1^N(i,j,k), x_2^N(i,l)\right)} \\
&= \overline{\Upsilon^N\left(\left\{\left(y_2^N, y_3^N\right) : \left(q^N, v^N(i), u^N(i,j), x_2^N\left(i,\hat{l}\right), \hat{y}_2^N\left(i,j,\hat{l},m_0\right), y_3^N(b)\right) \in A_\epsilon^{(N)}\right\}|x_1^N(i,j,k), x_2^N(i,l)\right)} \\
&= \overline{\Upsilon^N\left(\mathcal{Y}_2 \times \left\{y_3^N : \left(q^N, v^N(i), u^N(i,j), x_2^N\left(i,\hat{l}\right), \hat{y}_2^N\left(i,j,\hat{l},m_0\right), y_3^N(b)\right) \in A_\epsilon^{(N)}\right\}|x_1^N(i,j,k), x_2^N(i,l)\right)} \\
&= \int \Upsilon^N\left(\left\{y_3^N : \left(q^N, v^N, u^N, x_2^{N'}, \hat{y}_2^{N'}, y_3^N\right) \in A_\epsilon^{(N)}\left(Q,V,U,X_2,\hat{Y}_2,Y_3\right)\right\}|x_1^N, x_2^N\right) \\
& \times dP_{Q^N V^N U^N X_1^N X_2^N X_2^{N'} \hat{Y}_2^{N'}}
\end{aligned}
$$

$$
\begin{aligned}
&= \int \left[\int_{\left\{ y_3^N : \left(q^N, v^N, u^N, x_2^{N\prime}, \hat{y}_2^{N\prime}, y_3^N \right) \in A_\epsilon^{(N)}(Q,V,U,X_2,\hat{Y}_2,Y_3) \right\}} dP_{Y_3^N | X_1^N X_2^N} \right] dP_{Q^N V^N U^N X_1^N X_2^N X_2^{N\prime} \hat{Y}_2^{N\prime}} \\
&= \int_{\left(q^N, v^N, u^N, x_1^N, x_2^N, x_2^{N\prime}, \hat{y}_2^{N\prime}, y_3^N \right) : \left(q^N, v^N, u^N, x_2^{N\prime}, \hat{y}_2^{N\prime}, y_3^N \right) \in A_\epsilon^{(N)}(Q,V,U,X_2,\hat{Y}_2,Y_3)} \\
&\qquad \times dP_{Q^N V^N U^N X_1^N X_2^N X_2^{N\prime} \hat{Y}_2^{N\prime} Y_3^N} \\
&= \int_{A_\epsilon^{(N)}(Q,V,U,X_2,\hat{Y}_2,Y_3)} \left[\int dP_{X_1^N X_2^N | Q^N V^N U^N Y_3^N} \right] dP_{Q^N V^N U^N X_2^{N\prime} \hat{Y}_2^{N\prime} Y_3^N} \\
&= \int_{A_\epsilon^{(N)}(Q,V,U,X_2,\hat{Y}_2,Y_3)} f\left(q^N, v^N, u^N \right) f\left(y_3^N | q^N, v^N, u^N \right) f\left(x_2^{N\prime}, \hat{y}_2^{N\prime} | q^N, v^N, u^N \right) \\
&\qquad \times dM_{Q^N V^N U^N X_2^{N\prime} \hat{Y}_2^{N\prime} Y_3^N} \\
&\leq \int_{A_\epsilon^{(N)}(Q,V,U,X_2,\hat{Y}_2,Y_3)} 2^{-N\left(H_{P||M}(QVU) - \epsilon \right)} 2^{-N\left(H_{P||M}(Y_3|QVU) - 2\epsilon \right)} 2^{-N\left(H_{P||M}(X_2 \hat{Y}_2 | QVU) - 2\epsilon \right)} \\
&\qquad \times dM_{Q^N V^N U^N X_2^{N\prime} \hat{Y}_2^{N\prime} Y_3^N} \\
&\overset{(a)}{=} 2^{-N\left(H_{P||M}(QVU) - \epsilon \right)} 2^{-N\left(H_{P||M}(Y_3|QVU) - 2\epsilon \right)} 2^{-N\left(H_{P||M}(X_2 \hat{Y}_2 | QVU) - 2\epsilon \right)} \mathrm{vol}\left(A_\epsilon^{(N)}\left(Q, V, U, X_2, \hat{Y}_2, Y_3 \right) \right) \\
&\leq 2^{-N\left(H_{P||M}(QVU) - \epsilon \right)} 2^{-N\left(H_{P||M}(Y_3|QVU) - 2\epsilon \right)} 2^{-N\left(H_{P||M}(X_2 \hat{Y}_2 | QVU) - 2\epsilon \right)} 2^{N\left(H_{P||M}(QVUX_2 \hat{Y}_2 Y_3) + \epsilon \right)} \\
&= 2^{-N\left(I(X_2 \hat{Y}_2 ; Y_3 | VUQ) - 6\epsilon \right)}
\end{aligned} \tag{B.4}
$$

where $P_{Q^N V^N U^N X_1^N X_2^N X_2^{N\prime} \hat{Y}_2^{N\prime}} = P_{X_1^N X_2^N \times X_2^{N\prime} \hat{Y}_2^{N\prime} | Q^N V^N U^N}$ and (a) from the fact that $M_{Q^N V^N U^N X_2^N \hat{Y^N}_2 Y_3^N} = M_{X_2^N \hat{Y^N}_2 \times Y_3^N | Q^N V^N U^N} = M_{Q^N V^N U^N X_2^{N\prime} \hat{Y}_2^{N\prime} Y_3^N}$. Finally, we may bound $\overline{\Upsilon^N \left(\mathcal{L}_{\hat{k}} | x_1^N(i,j,k), x_2^N(i,l) \right)}$ as follows:

$$
\begin{aligned}
&\overline{\Upsilon^N \left(\mathcal{L}_{\hat{k}} | x_1^N(i,j,k), x_2^N(i,l) \right)} \\
&= \int \Upsilon^N \left(\left\{ \left(y_2^N, y_3^N \right) : \left(q^N, v^N, u^N, x_2^N, \Theta\left(q^N, u^N, x_2^N, y_2^N \right), x_1^{N\prime}, y_3^N \right) \in A_\epsilon^{(N)} \right\} | x_1^N, x_2^N \right) \\
&\qquad \times dP_{\mathbf{C}_1 Q^N V^N U^N X_1^N X_1^{N\prime} X_2^N} \\
&= \int \left[\int_{\left\{ (y_2^N, y_3^N) : \left(q^N, v^N, u^N, x_2^N, \Theta(q^N, u^N, x_2^N, y_2^N), x_1^{N\prime}, y_3^N \right) \in A_\epsilon^{(N)} \right\}} dP_{Y_2^N Y_3^N | X_1^N X_2^N} \right] \\
&\qquad \times dP_{\mathbf{C}_1 Q^N V^N U^N X_1^N X_1^{N\prime} X_2^N} \\
&= \int_{\substack{\left(q^N, v^N, u^N, x_1^N, x_1^{N\prime}, x_2^N, C_1, y_3^N \right) \\ : \left(q^N, v^N, u^N, x_2^N, \Theta(q^N, u^N, x_2^N, y_2^N), x_1^{N\prime}, y_3^N \right) \in A_\epsilon^{(N)}}} dP_{\mathbf{C}_1 Q^N V^N U^N X_1^N X_1^{N\prime} X_2^N Y_2^N Y_3^N} \\
&= \int \left[\int_{A_\epsilon^{(N)}\left(X_1 | q^N, v^N, u^N, x_2^N, \Theta(q^N, u^N, x_2^N, y_2^N), y_3^N \right)} dP_{X_1^{N\prime} | Q^N U^N} \right] dP_{\mathbf{C}_1 Q^N V^N U^N X_1^N X_2^N Y_2^N Y_3^N} \\
&= \int \left[\int_{A_\epsilon^{(N)}\left(X_1 | q^N, v^N, u^N, x_2^N, \Theta(q^N, u^N, x_2^N, y_2^N), y_3^N \right)} f\left(x_1^{N\prime} | q^N, u^N \right) dM_{X_1^{N\prime} | Q^N U^N} \right] \\
&\qquad \times dP_{\mathbf{C}_1 Q^N V^N U^N X_1^N X_2^N Y_2^N Y_3^N}
\end{aligned} \tag{B.5}
$$

where $P_{\mathbf{C}_1 Q^N V^N U^N X_1^N X_1^{N\prime} X_2^N} = P_{X_1^{N\prime} \times \mathbf{C}_1 V^N X_1^N X_2^N | Q^N U^N}$. Since $f\left(x_1^{N\prime}|q^N, u^N\right) = f\left(x_1^N|q^N, u^N\right)$ and $M_{X_1^{N\prime}|Q^N U^N} = M_{X_1^N|Q^N V^N U^N X_2^N \hat{Y^N}_2 Y_3^N}$, we then have

$$\begin{aligned}
&\overline{\Upsilon^N\left(\mathcal{L}_{\hat{k}}|x_1^N(i,j,k), x_2^N(i,l)\right)} \\
&\leq \int \left[2^{-N\left(H_{P||M}(X_1|QU)-2\epsilon\right)} \int_{A_\epsilon^{(N)}\left(X_1|q^N, v^N, u^N, x_2^N, \Theta(q^N, u^N, x_2^N, y_2^N), y_3^N\right)} dM_{X_1^N|Q^N V^N U^N X_2^N \hat{Y^N}_2 Y_3^N}\right] \\
&\qquad \times dP_{\mathbf{C}_1 Q^N V^N U^N X_1^N X_2^N Y_2^N Y_3^N} \\
&\leq \int \left[2^{-N\left(H_{P||M}(X_1|QU)-2\epsilon\right)} 2^{N\left(H_{P||M}(X_1|QVUX_2\hat{Y}_2 Y_3)+2\epsilon\right)}\right] dP_{\mathbf{C}_1 Q^N V^N U^N X_1^N X_2^N Y_2^N Y_3^N} \\
&\leq 2^{-N\left(I(X_1;\hat{Y}_2 Y_3|QUX_2)-4\epsilon\right)}.
\end{aligned} \tag{B.6}$$

B.2 Proof of Thm. 3.4

We define the following $\mathcal{F}_{Y_3^N} \times \mathcal{F}_{X_1^N} \times \mathcal{F}_{V_1^N}$-measurable set for the sake of simplification of notation:

$$\mathcal{K}_{\left(q^N, u^N, x_2^N, y_2^N, C_1\right)} = A_\epsilon^{(N)}\left(Y_3 X_1 V | q^N, u^N, x_2^N, y_2^N, \Theta\left(q^N, u^N, x_2^N, y_2^N\right)\right). \tag{B.7}$$

Averaged over all codewords and mapping functions for block b, we obtain

$$\begin{aligned}
&\overline{\Upsilon^N\left(\mathcal{J}_d|x_1^N(i,j,k), x_2^N(i,l)\right)} \\
&= \mathbb{E}_{Q\Theta}\Bigg[\int \Upsilon^N\left(\left\{\left(y_2^N, y_3^N\right) : \left(q^N, v^N, u^N, x_1^N, x_2^N, \Theta\left(q^N, u^N, x_2^N, y_2^N\right), y_2^N, y_3^N\right) \notin A_\epsilon^{(N)}\right\} | x_1^N, x_2^N\right) \\
&\qquad \times dP_{Q^N U^N V^N X_1^N X_2^N \mathbf{C}_1}\Bigg] \\
&= \mathbb{E}_{Q\Theta}\left[\int_{\substack{(q^N, v^N, u^N, x_1^N, x_2^N, C_1, y_2^N, y_3^N): \\ (q^N, v^N, u^N, x_1^N, x_2^N, \Theta(q^N, u^N, x_2^N, y_2^N), y_2^N, y_3^N) \notin A_\epsilon^{(N)}}} dP_{Q^N U^N V^N X_1^N X_2^N \mathbf{C}_1 Y_2^N Y_3^N}\right] \\
&= \mathbb{E}_{Q\Theta}\left[\int \left[\int \left[\int_{\mathcal{K}(q^N, u^N, x_2^N, y_2^N, C_1)^c} dP_{Y_3^N X_1^N V^N | Q^N U^N X_2^N}\right] dP_{Y_2^N|Q^N U^N X_2^N}\right] dP_{Q^N U^N X_2^N \mathbf{C}_1}\right] \\
&= \mathbb{E}_{Q\Theta}\Bigg[\int \Bigg[\int_{\substack{y_2^N : \left\{(q^N, u^N, x_2^N, y_2^N, \Theta(q^N, u^N, x_2^N, y_2^N)) \right. \\ \left. \in S_{\lambda,\epsilon}^{(N)}(Q, U, X_2, Y_2, \hat{Y}_2)\right\}}} \left[\int_{\left(\mathcal{K}_{(q^N, u^N, x_2^N, y_2^N, C_1)}\right)^c} dP_{Y_3^N X_1^N V^N | Q^N U^N X_2^N}\right] \\
&\qquad \times dP_{Y_2^N|Q^N U^N X_2^N}\Bigg] dP_{Q^N U^N X_2^N \mathbf{C}_1}\Bigg]
\end{aligned}$$

$$+\mathbb{E}_{Q_\Theta}\left[\int\left[\int_{\substack{y_2^N:\{(q^N,u^N,x_2^N,y_2^N,\Theta(q^N,u^N,x_2^N,y_2^N))\\ \notin S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)\}}}\left[\int_{\left(\mathcal{K}_{\left(q^N,u^N,x_2^N,y_2^N,C_1\right)}\right)^c} dP_{Y_3^N X_1^N V^N|Q^N U^N X_2^N}\right]\right.\right.$$
$$\left.\left.\times\, dP_{Y_2^N|Q^N U^N X_2^N}\right] dP_{Q^N U^N X_2^N \mathbf{C}_1}\right] \tag{B.8}$$

For the first term, by definition, $\forall\left(q^N,u^N,x_2^N,y_2^N,\hat{y}_2^N\right):\left(q^N,u^N,x_2^N,y_2^N,\hat{y}_2^N\right)\in S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)$, we have

$$\int_{\mathcal{K}_{\left(q^N,u^N,x_2^N,y_2^N,C_1\right)}} dP_{Y_3^N X_1^N V^N|Q^N U^N X_2^N}\geq 1-\lambda$$
$$\Longrightarrow \int_{\left(\mathcal{K}_{\left(q^N,u^N,x_2^N,y_2^N,C_1\right)}\right)^c} dP_{Y_3^N X_1^N V^N|Q^N U^N X_2^N}\leq \lambda. \tag{B.9}$$

Hence, we may now bound the first term as follows:

$$\mathbb{E}_{Q_\Theta}\left[\int\left[\int_{\substack{y_2^N:\{(q^N,u^N,x_2^N,y_2^N,\Theta(q^N,u^N,x_2^N,y_2^N))\\ \in S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)\}}}\left[\int_{\left(\mathcal{K}_{\left(q^N,u^N,x_2^N,y_2^N,C_1\right)}\right)^c} dP_{Y_3^N X_1^N V^N|Q^N U^N X_2^N}\right]\right.\right.$$
$$\left.\left.\times\, dP_{Y_2^N|Q^N U^N X_2^N}\right] dP_{Q^N U^N X_2^N \mathbf{C}_1}\right]$$
$$\leq \mathbb{E}_{Q_\Theta}\left[\int\left[\int_{\substack{y_2^N:\{(q^N,u^N,x_2^N,y_2^N,\Theta(q^N,u^N,x_2^N,y_2^N))\\ \in S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)\}}}\lambda\ dP_{Y_2^N|Q^N U^N X_2^N}\right] dP_{Q^N U^N X_2^N \mathbf{C}_1}\right]$$
$$\leq \mathbb{E}_{Q_\Theta}\left[\int\left[\int\lambda\ dP_{Y_2^N|Q^N U^N X_2^N}\right] dP_{Q^N U^N X_2^N \mathbf{C}_1}\right]\leq\lambda. \tag{B.10}$$

Before considering the second term, let $\mathcal{T}\left(C_1|q^N,u^N,x_2^N\right)$ denote the set of all $\mathcal{F}_{Y_2^N}$-measurable sequences y_2^N such that there is at least one codeword $\hat{y}_2^N$ in $\mathcal{C}_1$, excluding the first codeword which serves as a dummy codeword, where $\left(q^N,u^N,x_2^N,y_2^N,\hat{y}_2^N\right)\in S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)$. For the second term, let us consider the following:

$$\int\left[\int\left[\int_{\substack{y_2^N:\{(q^N,u^N,x_2^N,y_2^N,\Theta(q^N,u^N,x_2^N,y_2^N))\\ \notin S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)\}}} dP_{Y_2^N|Q^N U^N X_2^N}\right] dP_{\mathbf{C}_1|Q^N U^N X_2^N}\right] dP_{Q^N U^N X_2^N}$$
$$=\int\left[\int\left[\int_{y_2^N:y_2^N\notin\mathcal{T}(\mathcal{C}_1|q^N,u^N,x_2^N)} dP_{Y_2^N|Q^N U^N X_2^N}\right] dP_{\mathbf{C}_1|Q^N U^N X_2^N}\right] dP_{Q^N U^N X_2^N}$$
$$=\int\left[\int\left[\int_{\mathcal{C}_1:y_2^N\notin\mathcal{T}(\mathcal{C}_1|q^N,u^N,x_2^N)} dP_{\mathbf{C}_1|Q^N U^N X_2^N}\right] dP_{Y_2^N|Q^N U^N X_2^N}\right] dP_{Q^N U^N X_2^N}. \tag{B.11}$$

Next, we may bound the innermost term as follows:

$$\begin{aligned}
&\int_{\mathcal{C}_1: y_2^N \notin \mathcal{T}(C_1|q^N,u^N,x_2^N)} dP_{\mathbf{C}_1|Q^N U^N X_2^N} \\
&= P_{\mathbf{C}_1|Q^N U^N X_2^N}\left(\left\{C_1 : y_2^N \notin \mathcal{T}\left(\mathcal{C}_1|q^N,u^N,x_2^N\right)\right\}|q^N,u^N,x_2^N\right) \\
&= \left[1 - P_{\hat{Y}_2^N|Q^N U^N X_2^N}\left(\left\{\hat{y}_2^N : \left(q^N,u^N,x_2^N,y_2^N,\hat{y}_2^N\right) \in S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)\right\}|q^N,u^N,x_2^N\right)\right]^{2^{NR_0}-1} \\
&= \left[1 - \int_{\left\{\hat{y}_2^N:(q^N,u^N,x_2^N,y_2^N,\hat{y}_2^N)\in S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)\right\}} dP_{\hat{Y}_2^N|Q^N U^N X_2^N}\right]^{2^{NR_0}-1} \\
&= \left[1 - \int_{\left\{\hat{y}_2^N:(q^N,u^N,x_2^N,y_2^N,\hat{y}_2^N)\in S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)\right\}} f\left(\hat{y}_2^N|q^N,u^N,x_2^N\right) dM_{\hat{Y}_2^N|Q^N U^N X_2^N}\right]^{2^{NR_0}-1}. \quad \text{(B.12)}
\end{aligned}$$

Since $S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right) \subseteq A_\epsilon^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)$, we may lower bound the term $f\left(\hat{y}_2^N|q^N,u^N,x_2^N\right)$ by

$$\begin{aligned}
f\left(\hat{y}_2^N|q^N,u^N,x_2^N,y_2^N\right) &= \frac{f\left(\hat{y}_2^N,y_2^N|q^N,u^N,x_2^N\right)}{f\left(y_2^N|q^N,u^N,x_2^N\right)} \frac{f\left(\hat{y}_2^N|q^N,u^N,x_2^N\right)}{f\left(\hat{y}_2^N|q^N,u^N,x_2^N\right)} \\
&\leq f\left(\hat{y}_2^N|q^N,u^N,x_2^N\right) 2^{N\left(I\left(Y_2;\hat{Y}_2|UX_2Q\right)+3\epsilon\right)} \\
\Rightarrow f\left(\hat{y}_2^N|q^N,u^N,x_2^N\right) &\geq f\left(\hat{y}_2^N|q^N,u^N,x_2^N,y_2^N\right) 2^{-N\left(I\left(Y_2;\hat{Y}_2|UX_2Q\right)+3\epsilon\right)}. \quad \text{(B.13)}
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
&\int_{\mathcal{C}_1: y_2^N \notin \mathcal{T}(C_1|q^N,u^N,x_2^N)} dP_{\mathbf{C}_1|Q^N U^N X_2^N} \\
&= \left[1 - 2^{-N\left(I(Y_2;\hat{Y}_2|UX_2Q)+3\epsilon\right)} \int_{\left\{\hat{y}_2^N:(q^N,u^N,x_2^N,y_2^N,\hat{y}_2^N)\in S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)\right\}} f\left(\hat{y}_2^N|q^N,u^N,x_2^N,y_2^N\right) \right. \\
&\qquad \left. \times\, dM_{\hat{Y}_2^N|Q^N U^N X_2^N}\right]^{2^{NR_0}-1} \\
&= 1 - \int_{\left\{\hat{y}_2^N:(q^N,u^N,x_2^N,y_2^N,\hat{y}_2^N)\in S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)\right\}} f\left(\hat{y}_2^N|q^N,u^N,x_2^N,y_2^N\right) dM_{\hat{Y}_2^N|Q^N U^N X_2^N} \\
&\quad + e^{-2^{-N\left(I(Y_2;\hat{Y}_2|UX_2Q)+3\epsilon\right)}\left(2^{NR_0}-1\right)}. \quad \text{(B.14)}
\end{aligned}$$

Let us denote the last term by $\alpha^{(N)}$. We see that $\alpha^{(N)}$ goes to zero double exponentially fast with N as long as $R_0 > I\left(Y_2;\hat{Y}_2|UX_2Q\right) + 3\epsilon$. Next, let us consider the following:

$$\int \left[\int_{\left\{\hat{y}_2^N:(q^N,u^N,x_2^N,y_2^N,\hat{y}_2^N)\in S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)\right\}} f\left(\hat{y}_2^N|q^N,u^N,x_2^N,y_2^N\right) dM_{\hat{Y}_2^N|Q^N U^N X_2^N}\right] \times f\left(q^N,u^N,x_2^N,y_2^N\right) dM_{Q^N U^N X_2^N Y_2^N}$$

$$= \int_{S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)} f\left(q^N, u^N, x_2^N, y_2^N, \hat{y}_2^N\right) dM_{Q^N U^N X_2^N Y_2^N \hat{Y}_2^N}$$
$$= \int_{S_{\lambda,\epsilon}^{(N)}(Q,U,X_2,Y_2,\hat{Y}_2)} dP_{Q^N U^N X_2^N Y_2^N \hat{Y}_2^N}. \tag{B.15}$$

Next, let us consider the following set of inequalities:

$$\begin{aligned}
1-\epsilon &\leq \int_{A_\epsilon^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)} dP_{Q^N U^N X_2^N Y_2^N \hat{Y}_2^N} \\
&= \int_{A_\epsilon^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right) \cap S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)} dP_{Q^N U^N X_2^N Y_2^N \hat{Y}_2^N} \\
&\quad + \int_{A_\epsilon^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right) \cap S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)^c} dP_{Q^N U^N X_2^N Y_2^N \hat{Y}_2^N} \\
&\leq \int_{S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)} dP_{Q^N U^N X_2^N Y_2^N \hat{Y}_2^N} \\
&\quad + (1-\lambda)\left\{1 - \int_{S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)} dP_{Q^N U^N X_2^N Y_2^N \hat{Y}_2^N}\right\} \\
\Longrightarrow 1 - \frac{\epsilon}{\lambda} &\leq \int_{S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)} dP_{Q^N U^N X_2^N Y_2^N \hat{Y}_2^N}.
\end{aligned} \tag{B.16}$$

Hence, we may bound (B.11) as follows:

$$\int \left[\int \left[\int_{\substack{y_2^N : \left\{ \left(q^N, u^N, x_2^N, y_2^N, \Theta\left(q^N, u^N, x_2^N, y_2^N\right)\right) \\ \notin S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)\right\}}} dP_{Y_2^N | Q^N U^N X_2^N} \right] dP_{\mathbf{C}_1 | Q^N U^N X_2^N} \right] dP_{Q^N U^N X_2^N}$$
$$\leq \frac{\epsilon}{\lambda} + \alpha^{(N)}. \tag{B.17}$$

Finally, we may bound the second term of (B.8) as follows:

$$\mathbb{E}_{Q_\Theta}\left[\int \left[\int_{\substack{y_2^N : \left\{ \left(q^N, u^N, x_2^N, y_2^N, \Theta\left(q^N, u^N, x_2^N, y_2^N\right)\right) \\ \notin S_{\lambda,\epsilon}^{(N)}\left(Q,U,X_2,Y_2,\hat{Y}_2\right)\right\}}} \left[\int_{\left(\mathcal{K}_{\left(q^N, u^N, x_2^N, y_2^N, C_1\right)}\right)^c} dP_{Y_3^N X_1^N V^N | Q^N U^N X_2^N} \right] \times dP_{Y_2^N | Q^N U^N X_2^N} \right] dP_{Q^N U^N X_2^N \mathbf{C}_1} \right]$$
$$\begin{aligned}
&\leq \mathbb{E}_{Q_\Theta}\left[\frac{\epsilon}{\lambda} + \alpha^{(N)}\right] \\
&= \frac{\epsilon}{\lambda} + \mathbb{E}_{Q_\Theta}\left[\alpha^{(N)}\right] \\
&= \frac{\epsilon}{\lambda} + \overline{\alpha}^{(N)}
\end{aligned} \tag{B.18}$$

where $\overline{\alpha}^{(N)} = \mathbb{E}_{Q_\Theta}\left[\alpha^{(N)}\right]$. Combining the first and second term in (B.8), $\overline{\Upsilon^N\left(\mathcal{J}_d|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)}$ may be bounded as follows:

$$\overline{\Upsilon^N\left(\mathcal{J}_d|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)} \leq \frac{\epsilon}{\lambda} + \lambda + \overline{\alpha}^{(N)}. \tag{B.19}$$

For $\mu \in (0,1)$, put $\epsilon = \mu^2$ and $\lambda = \mu$, we obtain

$$\begin{aligned}\overline{\Upsilon^N\left(\mathcal{J}_d|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)} &\leq \frac{\epsilon}{\lambda} + \lambda + \overline{\alpha}^{(N)} \\ &\leq 2\mu + \overline{\alpha}^{(N)}. \end{aligned} \tag{B.20}$$

It can be seen that for each $\mu \in (0,1)$ and for every $\epsilon > 0$, $\overline{\alpha}^{(N)}$ decays double exponentially fast to zero as $N \to \infty$. Hence, for any $\mu \in (0,1)$, there exists an integer $N_1 = r\left(\mu,\epsilon\right)$ such that $\overline{\alpha}^{(N)} \leq \mu$ for $N \geq N_1$. We then obtain

$$\overline{\Upsilon^N\left(\mathcal{J}_d|x_1^N\left(i,j,k\right),x_2^N\left(i,l\right)\right)} \leq 3\mu \tag{B.21}$$

for $N \geq N_1$.

Appendix C

Proof of Theorems in Chapter 4

C.1 Proof of existence of conditional probability distributions and deterministic encoding functions achieving same marginal probability distributions

Let us assume a fixed $P_1^* \in \mathcal{P}_1^*$ where

$$P_1^*(q, w_1, w_2, x_1, x_2) = p_Q(q)\, p_{W_1|Q}(w_1|q)\, p_{X_1|W_1Q}(x_1|w_1, q)\, p_{W_2|Q}(w_2|q)\, p_{X_2|W_2Q}(x_2|w_2, q)\,. \quad \text{(C.1)}$$

We have to find conditional probability distributions $p_{U_1|Q}(u_1|q)$, $p_{U_2|Q}(u_2|q)$, $p_{X_1|U_1W_1Q}(x_1|u_1, w_1, q)$, and $p_{X_2|U_2W_2Q}(x_2|u_2, w_2, q)$ such that

$$p_{X_1|W_1Q}(x_1|w_1, q) = \sum_{u_1 \in \mathcal{U}_1} p_{X_1|U_1W_1Q}(x_1|u_1, w_1, q)\, p_{U_1|Q}(u_1|q) \quad \text{(C.2)}$$

$$p_{X_2|W_2Q}(x_2|w_2, q) = \sum_{u_2 \in \mathcal{U}_2} p_{X_2|U_2W_2Q}(x_2|u_2, w_2, q)\, p_{U_2|Q}(u_2|q) \quad \text{(C.3)}$$

and where both $p(x_1|u_1,w_1,q)$ and $p(x_2|u_2,w_2,q)$ equals either 0 or 1 $\forall q \in \mathcal{Q}$, $\forall u_1 \in \mathcal{U}_1$, $\forall u_2 \in \mathcal{U}_2$, $\forall w_1 \in \mathcal{W}_1$, $\forall w_2 \in \mathcal{W}_2$, $\forall x_1 \in \mathcal{X}_1$, and $\forall x_2 \in \mathcal{X}_2$, i.e., x_i is a deterministic function of q_i, u_i, and w_i $(i = 1, 2)$.

The following algorithm allows us to find the conditional probability distributions $p_{U_1|Q}(u_1|q)$ and $p_{X_1|U_1W_1Q}(x_1|u_1,w_1,q)$. We assume $\mathcal{Q} = \left\{q_1, q_2, ..., q_{\|\mathcal{Q}\|}\right\}$, $\mathcal{W}_1 = \left\{w_{11}, w_{12}, ..., w_{1\|\mathcal{W}_1\|}\right\}$, and $\mathcal{X}_1 = \left\{x_{11}, x_{12}, ..., x_{1\|\mathcal{X}_1\|}\right\}$.

Algorithm C.1.

for $l = 1$ to $\|\mathcal{Q}\|$

set $j = 1$

initialize $r_{mn} = p(x_{1n}|w_{1m}, q_l)$, $m = 1, 2, ...\|\mathcal{W}_1\|$ and $n = 1, 2, ...\|\mathcal{X}_1\|$

while $\left(\max_{m,n} r_{mn}\right) > 0$

$p(u_{1j}|q_l) = \min_{m,n} (r_{mn})^+$

for $m = 1$ to $\|\mathcal{W}_1\|$

$n' = \arg\min_{n} (r_{mn})^+$

$r_{mn'} \leftarrow r_{mn'} - p(u_{1j}|q_l)$

$p(x_{1n'}|u_{1j}, w_{1m}, q_l) = 1$

$p(x_{1n}|u_{1j}, w_{1m}, q_l) = 0$, $n = 1, 2, ...\|\mathcal{X}_1\|$ and $n \neq n'$

end

$j \leftarrow j + 1$

end

$J_l = j - 1$

end

$\|\mathcal{U}_1\| = \max_{l} J_l$

Here, $(x)^+$ denotes the following function:

$$(x)^+ = \begin{cases} x & \text{if } x > 0 \\ +\infty & \text{if } x \leq 0 \end{cases} . \tag{C.4}$$

Since

$$\sum_{n=1}^{\|\mathcal{X}_1\|} p\left(x_{1n}|w_{1m},q_l\right) = 1, \quad 1 \leq m \leq \|\mathcal{W}_1\|, 1 \leq l \leq \|\mathcal{Q}\|, \tag{C.5}$$

it is easy to verify that Algorithm C.1 gives us a conditional probability distribution $p_{U_1|Q}\left(u_1|q\right)$ where

$$\sum_{j=1}^{J_l} p\left(u_{1j}|q_l\right) = 1, \quad 1 \leq l \leq \|\mathcal{Q}\|. \tag{C.6}$$

It is also easy to verify that Algorithm C.1 will always give us $p_{U_1|Q}\left(u_1|q\right)$ and $p_{X_1|U_1W_1Q}\left(x_1|u_1,w_1,q\right)$ such that

$$p\left(x_{1n}|w_{1m},q_l\right) = \sum_{j=1}^{J_l} p\left(x_{1n}|u_{1j},w_{1m},q_l\right) p\left(u_{1j}|q_l\right)$$
$$, 1 \leq n \leq \|\mathcal{X}_1\|,\ 1 \leq m \leq \|\mathcal{W}_1\|,\ 1 \leq l \leq \|\mathcal{Q}\| \tag{C.7}$$

and where $p\left(x_{1n}|u_{1j},w_{1m},q_l\right)$ equals either 0 or 1 for $1 \leq n \leq \|\mathcal{X}_1\|,\ 1 \leq j \leq J_l,\ 1 \leq m \leq \|\mathcal{W}_1\|,\ 1 \leq l \leq \|\mathcal{Q}\|$. The same algorithm can be applied to determine $p_{U_2|Q}\left(u_2|q\right)$ and $p_{X_2|U_2W_2Q}\left(x_2|u_2,w_2,q\right)$.

C.2 Proof of Lem. 4.3

To obtain the projected achievable rate region $\mathcal{R}_{\text{HK}}\left(P^*\right)$ using Fourier-Motzkin elimination, we need the following additional inequalities:

$$R_1 - S_1 - T_1 \leq 0, \tag{C.8}$$

$$R_2 - S_2 - T_2 \leq 0, \tag{C.9}$$

$$-R_1 \leq 0, \tag{C.10}$$

$$-R_2 \leq 0. \tag{C.11}$$

In addition, it is easy to verify that the following information theoretic inequalities between the bound constants, $a_{oi}, ..., g_{oi}$, $i = 1, 2$, hold:

$$
\begin{aligned}
a_{oi}, b_{oi} \leq& d_{oi} \leq a_{oi} + b_{oi}, \\
a_{oi}, c_{oi} \leq& e_{oi} \leq a_{oi} + c_{oi}, \\
b_{oi}, c_{oi} \leq& f_{oi} \leq b_{oi} + c_{oi}, \\
d_{oi}, e_{oi}, f_{oi} \leq& g_{oi} \leq c_{oi} + d_{oi}, b_{oi} + e_{oi}, a_{oi} + f_{oi}.
\end{aligned} \tag{C.12}
$$

Eliminate S_1: First collect all the inequalities not involving S_1 among all the inequalities to obtain

$$T_1 \leq b_{o1}, \tag{C.13}$$

$$T_2 \leq c_{o1}, \tag{C.14}$$

$$T_1 + T_2 \leq f_{o1}, \tag{C.15}$$

$$-T_1 \leq 0, \tag{C.16}$$

$$S_2 \leq a_{o2}, \tag{C.17}$$

$$T_2 \leq b_{o2}, \tag{C.18}$$

$$T_1 \leq c_{o2}, \tag{C.19}$$

$$S_2 + T_2 \leq d_{o2}, \tag{C.20}$$

$$S_2 + T_1 \leq e_{o2}, \tag{C.21}$$

$$T_1 + T_2 \leq f_{o2}, \tag{C.22}$$

$$S_2 + T_2 + T_1 \leq g_{o2}, \tag{C.23}$$

$$-S_2 \leq 0, \tag{C.24}$$

$$-T_2 \leq 0, \tag{C.25}$$

$$-R_1 \leq 0, \tag{C.26}$$

$$-R_2 \leq 0, \tag{C.27}$$

$$R_2 - S_2 - T_2 \leq 0. \tag{C.28}$$

Next, collect all the S_1 with positive coefficients to obtain

$$S_1 \leq a_{o1}, \tag{C.29}$$

$$S_1 + T_1 \leq d_{o1}, \tag{C.30}$$

$$S_1 + T_2 \leq e_{o1}, \tag{C.31}$$

$$S_1 + T_1 + T_2 \leq g_{o1}. \tag{C.32}$$

Furthermore, collect all the S_1 with negative coefficients to obtain

$$-S_1 \leq 0, \tag{C.33}$$

$$R_1 - S_1 - T_1 \leq 0. \tag{C.34}$$

We eliminate S_1 by adding each inequality from (C.29)-(C.32) and each inequality from (C.33),(C.34) to obtain inequalities not involving S_1:

$$0 \leq a_{o1}, \tag{C.35}$$

$$R_1 - T_1 \leq a_{o1}, \tag{C.36}$$

$$T_1 \leq d_{o1}, \tag{C.37}$$

$$R_1 \leq d_{o1}, \tag{C.38}$$

$$T_2 \leq e_{o1}, \tag{C.39}$$

$$R_1 + T_2 - T_1 \leq e_{o1}, \tag{C.40}$$

$$T_1 + T_2 \leq g_{o1}, \tag{C.41}$$

$$R_1 + T_2 \leq g_{o1}. \tag{C.42}$$

It is clear that (C.35) is redundant, (C.37) is redundant due to (C.13), (C.39) is redundant due to (C.14), and (C.41) is redundant due to (C.15).

Eliminate S_2: First collect all the inequalities not involving S_2 among all the non-redundant inequalities to obtain

$$T_1 \le b_{o1}, \tag{C.43}$$

$$T_2 \le c_{o1}, \tag{C.44}$$

$$T_1 + T_2 \le f_{o1}, \tag{C.45}$$

$$-T_1 \le 0, \tag{C.46}$$

$$T_2 \le b_{o2}, \tag{C.47}$$

$$T_1 \le c_{o2}, \tag{C.48}$$

$$T_1 + T_2 \le f_{o2}, \tag{C.49}$$

$$-T_2 \le 0, \tag{C.50}$$

$$-R_1 \le 0, \tag{C.51}$$

$$-R_2 \le 0, \tag{C.52}$$

$$R_1 - T_1 \le a_{o1}, \tag{C.53}$$

$$R_1 \le d_{o1}, \tag{C.54}$$

$$R_1 + T_2 - T_1 \le e_{o1}, \tag{C.55}$$

$$R_1 + T_2 \le g_{o1}. \tag{C.56}$$

Next, collect all the S_2 with positive coefficients to obtain

$$S_2 \le a_{o2}, \tag{C.57}$$

$$S_2 + T_2 \le d_{o2}, \tag{C.58}$$

$$S_2 + T_1 \le e_{o2}, \tag{C.59}$$

$$S_2 + T_2 + T_1 \le g_{o2}. \tag{C.60}$$

Furthermore, collect all the S_2 with negative coefficients to obtain

$$-S_2 \le 0, \tag{C.61}$$

$$R_2 - S_2 - T_2 \le 0. \tag{C.62}$$

We eliminate S_2 by adding each inequality from (C.57)-(C.60) and each inequality from (C.61),(C.62) to obtain inequalities not involving S_2:

$$0 \leq a_{o2}, \tag{C.63}$$

$$R_2 - T_2 \leq a_{o2}, \tag{C.64}$$

$$T_2 \leq d_{o2}, \tag{C.65}$$

$$R_2 \leq d_{o2}, \tag{C.66}$$

$$T_1 \leq e_{o2}, \tag{C.67}$$

$$R_2 + T_1 - T_2 \leq e_{o2}, \tag{C.68}$$

$$T_2 + T_1 \leq g_{o2}, \tag{C.69}$$

$$R_2 + T_1 \leq g_{o2}. \tag{C.70}$$

It is clear that (C.63) is redundant, (C.65) is redundant due to (C.47), (C.67) is redundant due to (C.48), and (C.69) is redundant due to (C.49).
Eliminate T_1: First collect all the inequalities not involving T_1 among all the inequalities above to obtain

$$T_2 \leq c_{o1}, \tag{C.71}$$

$$T_2 \leq b_{o2}, \tag{C.72}$$

$$-T_2 \leq 0, \tag{C.73}$$

$$-R_1 \leq 0, \tag{C.74}$$

$$-R_2 \leq 0, \tag{C.75}$$

$$R_1 \leq d_{o1}, \tag{C.76}$$

$$R_1 + T_2 \leq g_{o1}, \tag{C.77}$$

$$R_2 - T_2 \leq a_{o2}, \tag{C.78}$$

$$R_2 \leq d_{o2}. \tag{C.79}$$

Next, collect all the T_1 with positive coefficients to obtain

$$T_1 \leq b_{o1}, \tag{C.80}$$

$$T_1 + T_2 \leq f_{o1}, \tag{C.81}$$

$$T_1 \leq c_{o2}, \tag{C.82}$$

$$T_1 + T_2 \leq f_{o2}, \tag{C.83}$$

$$R_2 + T_1 - T_2 \leq e_{o2}, \tag{C.84}$$

$$R_2 + T_1 \leq g_{o2}. \tag{C.85}$$

Furthermore, collect all the T_1 with negative coefficients to obtain

$$-T_1 \leq 0, \tag{C.86}$$

$$R_1 - T_1 \leq a_{o1}, \tag{C.87}$$

$$R_1 + T_2 - T_1 \leq e_{o1}. \tag{C.88}$$

We eliminate T_1 by adding each inequality from (C.80)-(C.85) and each inequality from (C.86)-(C.88) to obtain inequalities not involving T_1:

$$0 \leq b_{o1}, \tag{C.89}$$

$$R_1 \leq a_{o1} + b_{o1}, \tag{C.90}$$

$$R_1 + T_2 \leq b_{o1} + e_{o1}, \tag{C.91}$$

$$T_2 \leq f_{o1}, \tag{C.92}$$

$$R_1 + T_2 \leq a_{o1} + f_{o1}, \tag{C.93}$$

$$R_1 + 2T_2 \leq e_{o1} + f_{o1}, \tag{C.94}$$

$$0 \leq c_{o2}, \tag{C.95}$$

$$R_1 \leq a_{o1} + c_{o2}, \tag{C.96}$$

$$R_1 + T_2 \leq e_{o1} + c_{o2}, \tag{C.97}$$

$$T_2 \leq f_{o2}, \tag{C.98}$$

$$R_1 + T_2 \leq a_{o1} + f_{o2}, \tag{C.99}$$

$$R_1 + 2T_2 \leq e_{o1} + f_{o2}, \tag{C.100}$$

$$R_2 - T_2 \leq e_{o2}, \tag{C.101}$$

$$R_1 + R_2 - T_2 \leq a_{o1} + e_{o2}, \tag{C.102}$$

$$R_1 + R_2 \leq e_{o1} + e_{o2}, \tag{C.103}$$

$$R_2 \leq g_{o2}, \tag{C.104}$$

$$R_1 + R_2 \leq a_{o1} + g_{o2}, \tag{C.105}$$

$$R_1 + R_2 + T_2 \leq e_{o1} + g_{o2}. \tag{C.106}$$

It is clear that (C.89) and (C.95) are redundant, (C.90) is redundant due to (C.76) ($a_{o1} + b_{o1} \geq d_{o1}$), (C.91) is redundant due to (C.77) ($b_{o1} + e_{o1} \geq g_{o1}$), (C.92) is redundant due to (C.71), (C.93) is redundant due to (C.77) ($a_{o1} + f_{o1} \geq g_{o1}$), (C.98) is redundant due to (C.72), (C.101) is redundant due to (C.78), and (C.104) is redundant due to (C.79).

Eliminate T_2: First collect all the inequalities not involving T_2 among all the inequalities above to obtain

$$-R_1 \leq 0, \tag{C.107}$$

$$-R_2 \leq 0, \tag{C.108}$$

$$R_1 \leq d_{o1}, \tag{C.109}$$

$$R_2 \leq d_{o2}, \tag{C.110}$$

$$R_1 \leq a_{o1} + c_{o2}, \tag{C.111}$$

$$R_1 + R_2 \leq e_{o1} + e_{o2}, \tag{C.112}$$

$$R_1 + R_2 \leq a_{o1} + g_{o2}. \tag{C.113}$$

Next, collect all the T_2 with positive coefficients to obtain

$$T_2 \leq c_{o1}, \tag{C.114}$$

$$T_2 \leq b_{o2}, \tag{C.115}$$

$$R_1 + T_2 \leq g_{o1}, \tag{C.116}$$

$$R_1 + 2T_2 \leq e_{o1} + f_{o1}, \tag{C.117}$$

$$R_1 + T_2 \leq e_{o1} + c_{o2}, \tag{C.118}$$

$$R_1 + T_2 \leq a_{o1} + f_{o2}, \tag{C.119}$$

$$R_1 + 2T_2 \leq e_{o1} + f_{o2}, \tag{C.120}$$

$$R_1 + R_2 + T_2 \leq e_{o1} + g_{o2}. \tag{C.121}$$

Furthermore, collect all the T_2 with negative coefficients to obtain

$$-T_2 \leq 0, \tag{C.122}$$

$$R_2 - T_2 \leq a_{o2}, \tag{C.123}$$

$$R_1 + R_2 - T_2 \leq a_{o1} + e_{o2}. \tag{C.124}$$

We eliminate T_2 by adding each inequality from (C.114)-(C.121) and each inequality from (C.122)-(C.124) to obtain inequalities not involving T_2:

$$0 \leq c_{o1}, \tag{C.125}$$

$$R_2 \leq a_{o2} + c_{o1}, \tag{C.126}$$

$$R_1 + R_2 \leq a_{o1} + c_{o1} + e_{o2}, \tag{C.127}$$

$$0 \leq b_{o2}, \tag{C.128}$$

$$R_2 \leq a_{o2} + b_{o2}, \tag{C.129}$$

$$R_1 + R_2 \leq a_{o1} + b_{o2} + e_{o2}, \tag{C.130}$$

$$R_1 \leq g_{o1}, \tag{C.131}$$

$$R_1 + R_2 \leq a_{o2} + g_{o1}, \tag{C.132}$$

$$2R_1 + R_2 \leq a_{o1} + g_{o1} + e_{o2}, \tag{C.133}$$

$$R_1 \leq e_{o1} + f_{o1}, \tag{C.134}$$

$$R_1 + 2R_2 \leq 2a_{o2} + e_{o1} + f_{o1}, \tag{C.135}$$

$$3R_1 + 2R_2 \leq 2a_{o1} + e_{o1} + f_{o1} + 2e_{o2}, \tag{C.136}$$

$$R_1 \leq e_{o1} + c_{o2}, \tag{C.137}$$

$$R_1 + R_2 \leq e_{o1} + a_{o2} + c_{o2}, \tag{C.138}$$

$$2R_1 + R_2 \leq a_{o1} + e_{o1} + c_{o2} + e_{o2}, \tag{C.139}$$

$$R_1 \leq a_{o1} + f_{o2}, \tag{C.140}$$

$$R_1 + R_2 \leq a_{o1} + a_{o2} + f_{o2}, \tag{C.141}$$

$$2R_1 + R_2 \leq 2a_{o1} + e_{o2} + f_{o2}, \tag{C.142}$$

$$R_1 \leq e_{o1} + f_{o2}, \tag{C.143}$$

$$R_1 + 2R_2 \leq e_{o1} + 2a_{o2} + f_{o2}, \tag{C.144}$$

$$3R_1 + 2R_2 \leq 2a_{o1} + e_{o1} + 2e_{o2} + f_{o2}, \tag{C.145}$$

$$R_1 + R_2 \leq e_{o1} + g_{o2}, \tag{C.146}$$

$$R_1 + 2R_2 \leq a_{o2} + g_{o2} + e_{o1}, \tag{C.147}$$

$$2R_1 + 2R_2 \leq a_{o1} + e_{o1} + e_{o2} + g_{o2}. \tag{C.148}$$

It is clear that (C.125) and (C.128) are redundant, (C.127) is redundant due to (C.112) ($a_{o1} + c_{o1} \geq e_{o1}$), (C.129) is redundant due to (C.110) ($a_{o2} + b_{o2} \geq d_{o2}$), (C.130) is redundant due to (C.113) ($b_{o2} + e_{o2} \geq g_{o2}$), (C.131) is redundant due to (C.109), (C.134) is redundant due to (C.109) ($e_{o1} + f_{o1} \geq d_{o1}$), (C.136) is redundant due to (C.112) and (C.133) ($a_{o1} + f_{o1} \geq g_{o1}$), (C.137) is redundant due to (C.111), (C.138) is redundant due to (C.112) ($a_{o2} + c_{o2} \geq e_{o2}$), (C.139) is redundant due to (C.111) and (C.112), (C.140) is redundant due to (C.111), (C.141) is redundant due to (C.113) ($a_{o2} + f_{o2} \geq g_{o2}$), (C.143) is redundant due to (C.111), (C.144) is redundant due to (C.147) ($a_{o2} + f_{o2} \geq g_{o2}$), (C.145) is redundant due to (C.112) and (C.142), (C.146) is redundant due to (C.112), and (C.148) is redundant due to (C.112) and (C.113). Finally, we obtain the following inequalities not involving T_2:

$$R_1 \le d_{o1}, \tag{C.149}$$

$$R_1 \le a_{o1} + c_{o2}, \tag{C.150}$$

$$R_2 \le d_{o2}, \tag{C.151}$$

$$R_2 \le a_{o2} + c_{o1}, \tag{C.152}$$

$$R_1 + R_2 \le a_{o1} + g_{o2}, \tag{C.153}$$

$$R_1 + R_2 \le a_{o2} + g_{o1}, \tag{C.154}$$

$$R_1 + R_2 \le e_{o1} + e_{o2}, \tag{C.155}$$

$$2R_1 + R_2 \le a_{o1} + g_{o1} + e_{o2}, \tag{C.156}$$

$$2R_1 + R_2 \le 2a_{o1} + e_{o2} + f_{o2}, \tag{C.157}$$

$$R_1 + 2R_2 \le a_{o2} + g_{o2} + e_{o1}, \tag{C.158}$$

$$R_1 + 2R_2 \le 2a_{o2} + e_{o1} + f_{o1}, \tag{C.159}$$

$$-R_1 \le 0, \tag{C.160}$$

$$-R_2 \le 0. \tag{C.161}$$

C.3 Proof of Lem. 4.3

To obtain the projected achievable rate region $\mathcal{R}_{\mathrm{CMG}}\left(P_1^*\right)$ using Fourier-Motzkin elimination, we need the following additional inequalities:

$$R_1 - S_1 - T_1 \le 0, \tag{C.162}$$

$$R_2 - S_2 - T_2 \le 0, \tag{C.163}$$

$$-R_1 \le 0, \tag{C.164}$$

$$-R_2 \le 0. \tag{C.165}$$

Eliminate S_1: First collect all the inequalities not involving S_1 among all the inequalities to obtain

$$-T_1 \leq 0, \tag{C.166}$$

$$S_2 \leq a_{o2}, \tag{C.167}$$

$$S_2 + T_2 \leq d_{o2}, \tag{C.168}$$

$$S_2 + T_1 \leq e_{o2}, \tag{C.169}$$

$$S_2 + T_2 + T_1 \leq g_{o2}, \tag{C.170}$$

$$-S_2 \leq 0, \tag{C.171}$$

$$-T_2 \leq 0, \tag{C.172}$$

$$R_2 - S_2 - T_2 \leq 0, \tag{C.173}$$

$$-R_1 \leq 0, \tag{C.174}$$

$$-R_2 \leq 0. \tag{C.175}$$

Next, collect all the S_1 with positive coefficients to obtain

$$S_1 \leq a_{o1}, \tag{C.176}$$

$$S_1 + T_1 \leq d_{o1}, \tag{C.177}$$

$$S_1 + T_2 \leq e_{o1}, \tag{C.178}$$

$$S_1 + T_1 + T_2 \leq g_{o1}. \tag{C.179}$$

Furthermore, collect all the S_1 with negative coefficients to obtain

$$-S_1 \leq 0, \tag{C.180}$$

$$R_1 - S_1 - T_1 \leq 0. \tag{C.181}$$

We eliminate S_1 by adding each inequality from (C.176)-(C.179) and each inequality from (C.180),(C.181) to obtain inequalities not involving S_1:

$$0 \leq a_{o1}, \tag{C.182}$$

$$R_1 - T_1 \leq a_{o1}, \tag{C.183}$$

$$T_1 \leq d_{o1}, \tag{C.184}$$

$$R_1 \leq d_{o1}, \tag{C.185}$$

$$T_2 \leq e_{o1}, \tag{C.186}$$

$$R_1 + T_2 - T_1 \leq e_{o1}, \tag{C.187}$$

$$T_1 + T_2 \leq g_{o1}, \tag{C.188}$$

$$R_1 + T_2 \leq g_{o1}. \tag{C.189}$$

It is clear that (C.182) is redundant.

Eliminate S_2: First collect all the inequalities not involving S_2 among all the inequalities above to obtain

$$-T_1 \leq 0, \tag{C.190}$$

$$-T_2 \leq 0, \tag{C.191}$$

$$-R_1 \leq 0, \tag{C.192}$$

$$-R_2 \leq 0, \tag{C.193}$$

$$R_1 - T_1 \leq a_{o1}, \tag{C.194}$$

$$T_1 \leq d_{o1}, \tag{C.195}$$

$$R_1 \leq d_{o1}, \tag{C.196}$$

$$T_2 \leq e_{o1}, \tag{C.197}$$

$$R_1 + T_2 - T_1 \leq e_{o1}, \tag{C.198}$$

$$T_1 + T_2 \leq g_{o1}, \tag{C.199}$$

$$R_1 + T_2 \leq g_{o1}. \tag{C.200}$$

Next, collect all the S_2 with positive coefficients to obtain

$$S_2 \le a_{o2}, \tag{C.201}$$

$$S_2 + T_2 \le d_{o2}, \tag{C.202}$$

$$S_2 + T_1 \le e_{o2}, \tag{C.203}$$

$$S_2 + T_2 + T_1 \le g_{o2}. \tag{C.204}$$

Furthermore, collect all the S_2 with negative coefficients to obtain

$$-S_2 \le 0, \tag{C.205}$$

$$R_2 - S_2 - T_2 \le 0. \tag{C.206}$$

We eliminate S_2 by adding each inequality from (C.201)-(C.204) and each inequality from (C.205),(C.206) to obtain inequalities not involving S_2:

$$0 \le a_{o2}, \tag{C.207}$$

$$R_2 - T_2 \le a_{o2}, \tag{C.208}$$

$$T_2 \le d_{o2}, \tag{C.209}$$

$$R_2 \le d_{o2}, \tag{C.210}$$

$$T_1 \le e_{o2}, \tag{C.211}$$

$$R_2 + T_1 - T_2 \le e_{o2}, \tag{C.212}$$

$$T_2 + T_1 \le g_{o2}, \tag{C.213}$$

$$R_2 + T_1 \le g_{o2}. \tag{C.214}$$

It is clear that (C.207) is redundant.

Eliminate T_1: First collect all the inequalities not involving T_1 among all the inequalities above to obtain

$$-T_2 \le 0, \tag{C.215}$$

$$-R_1 \leq 0, \tag{C.216}$$

$$-R_2 \leq 0, \tag{C.217}$$

$$R_1 \leq d_{o1}, \tag{C.218}$$

$$T_2 \leq e_{o1}, \tag{C.219}$$

$$R_1 + T_2 \leq g_{o1}, \tag{C.220}$$

$$R_2 - T_2 \leq a_{o2}, \tag{C.221}$$

$$T_2 \leq d_{o2}, \tag{C.222}$$

$$R_2 \leq d_{o2}. \tag{C.223}$$

Next, collect all the T_1 with positive coefficients to obtain

$$T_1 \leq d_{o1}, \tag{C.224}$$

$$T_1 + T_2 \leq g_{o1}, \tag{C.225}$$

$$T_1 \leq e_{o2}, \tag{C.226}$$

$$R_2 + T_1 - T_2 \leq e_{o2}, \tag{C.227}$$

$$T_2 + T_1 \leq g_{o2}, \tag{C.228}$$

$$R_2 + T_1 \leq g_{o2}. \tag{C.229}$$

Furthermore, collect all the T_1 with negative coefficients to obtain

$$-T_1 \leq 0, \tag{C.230}$$

$$R_1 - T_1 \leq a_{o1}, \tag{C.231}$$

$$R_1 + T_2 - T_1 \leq e_{o1}. \tag{C.232}$$

We eliminate T_1 by adding each inequality from (C.224)-(C.229) and each inequality from (C.230)-(C.232) to obtain inequalities not involving T_1:

$$0 \leq d_{o1}, \tag{C.233}$$

$$R_1 \leq a_{o1} + d_{o1}, \tag{C.234}$$

$$R_1 + T_2 \leq d_{o1} + e_{o1}, \tag{C.235}$$

$$T_2 \leq g_{o1}, \tag{C.236}$$

$$R_1 + T_2 \leq a_{o1} + g_{o1}, \tag{C.237}$$

$$R_1 + 2T_2 \leq e_{o1} + g_{o1}, \tag{C.238}$$

$$0 \leq e_{o2}, \tag{C.239}$$

$$R_1 \leq a_{o1} + e_{o2}, \tag{C.240}$$

$$R_1 + T_2 \leq e_{o1} + e_{o2}, \tag{C.241}$$

$$R_2 - T_2 \leq e_{o2}, \tag{C.242}$$

$$R_1 + R_2 - T_2 \leq a_{o1} + e_{o2}, \tag{C.243}$$

$$R_1 + R_2 \leq e_{o1} + e_{o2}, \tag{C.244}$$

$$T_2 \leq g_{o2}, \tag{C.245}$$

$$R_1 + T_2 \leq a_{o1} + g_{o2}, \tag{C.246}$$

$$R_1 + 2T_2 \leq e_{o1} + g_{o2}, \tag{C.247}$$

$$R_2 \leq g_{o2}, \tag{C.248}$$

$$R_1 + R_2 \leq a_{o1} + g_{o2}, \tag{C.249}$$

$$R_1 + R_2 + T_2 \leq e_{o1} + g_{o2}. \tag{C.250}$$

It is clear that (C.233) and (C.239) are redundant, (C.234) is redundant due to (C.218), (C.235) is redundant due to (C.220) ($d_{o1} + e_{o1} \geq g_{o1}$), (C.236) is redundant due to (C.219), (C.237) is redundant due to (C.220), (C.242) is redundant due to (C.221), (C.245) is redundant due to (C.222), and (C.248) is redundant due to (C.223).

Eliminate T_2: First collect all the inequalities not involving T_2 among all the inequalities above to obtain

$$-R_1 \leq 0, \tag{C.251}$$

$$-R_2 \leq 0, \tag{C.252}$$

$$R_1 \leq d_{o1}, \tag{C.253}$$

$$R_2 \leq d_{o2}, \tag{C.254}$$

$$R_1 \leq a_{o1} + e_{o2}, \tag{C.255}$$

$$R_1 + R_2 \leq e_{o1} + e_{o2}, \tag{C.256}$$

$$R_1 + R_2 \leq a_{o1} + g_{o2}. \tag{C.257}$$

Next, collect all the T_2 with positive coefficients to obtain

$$T_2 \leq e_{o1}, \tag{C.258}$$

$$T_2 \leq d_{o2}, \tag{C.259}$$

$$R_1 + T_2 \leq g_{o1}, \tag{C.260}$$

$$R_1 + 2T_2 \leq e_{o1} + g_{o1}, \tag{C.261}$$

$$R_1 + T_2 \leq e_{o1} + e_{o2}, \tag{C.262}$$

$$R_1 + T_2 \leq a_{o1} + g_{o2}, \tag{C.263}$$

$$R_1 + 2T_2 \leq e_{o1} + g_{o2}, \tag{C.264}$$

$$R_1 + R_2 + T_2 \leq e_{o1} + g_{o2}. \tag{C.265}$$

Furthermore, collect all the T_2 with negative coefficients to obtain

$$-T_2 \leq 0, \tag{C.266}$$

$$R_2 - T_2 \leq a_{o2}, \tag{C.267}$$

$$R_1 + R_2 - T_2 \leq a_{o1} + e_{o2}. \tag{C.268}$$

We eliminate T_2 by adding each inequality from (C.258)-(C.265) and each inequality from (C.266)-(C.268) to obtain inequalities not involving T_2:

$$0 \leq e_{o1}, \tag{C.269}$$

$$R_2 \leq a_{o2} + e_{o1}, \tag{C.270}$$

$$R_1 + R_2 \leq a_{o1} + e_{o1} + e_{o2}, \tag{C.271}$$

$$0 \leq d_{o2}, \tag{C.272}$$

$$R_2 \leq a_{o2} + d_{o2}, \tag{C.273}$$

$$R_1 + R_2 \leq a_{o1} + d_{o2} + e_{o2}, \tag{C.274}$$

$$R_1 \leq g_{o1}, \tag{C.275}$$

$$R_1 + R_2 \leq a_{o2} + g_{o1}, \tag{C.276}$$

$$2R_1 + R_2 \leq a_{o1} + g_{o1} + e_{o2}, \tag{C.277}$$

$$R_1 \leq e_{o1} + g_{o1}, \tag{C.278}$$

$$R_1 + 2R_2 \leq 2a_{o2} + e_{o1} + g_{o1}, \tag{C.279}$$

$$3R_1 + 2R_2 \leq 2a_{o1} + e_{o1} + g_{o1} + 2e_{o2}, \tag{C.280}$$

$$R_1 \leq e_{o1} + e_{o2}, \tag{C.281}$$

$$R_1 + R_2 \leq e_{o1} + a_{o2} + e_{o2}, \tag{C.282}$$

$$2R_1 + R_2 \leq a_{o1} + e_{o1} + 2e_{o2}, \tag{C.283}$$

$$R_1 \leq a_{o1} + g_{o2}, \tag{C.284}$$

$$R_1 + R_2 \leq a_{o1} + a_{o2} + g_{o2}, \tag{C.285}$$

$$2R_1 + R_2 \leq 2a_{o1} + e_{o2} + g_{o2}, \tag{C.286}$$

$$R_1 \leq e_{o1} + g_{o2}, \tag{C.287}$$

$$R_1 + 2R_2 \leq e_{o1} + 2a_{o2} + g_{o2}, \tag{C.288}$$

$$3R_1 + 2R_2 \leq 2a_{o1} + e_{o1} + 2e_{o2} + g_{o2}, \tag{C.289}$$

$$R_1 + R_2 \leq e_{o1} + g_{o2}, \tag{C.290}$$

$$R_1 + 2R_2 \leq a_{o2} + g_{o2} + e_{o1}, \tag{C.291}$$

$$2R_1 + 2R_2 \leq a_{o1} + e_{o1} + e_{o2} + g_{o2}. \tag{C.292}$$

It is clear that (C.269) and (C.272) are redundant, (C.271) is redundant due to (C.256), (C.273) is redundant due to (C.254), (C.274) is redundant due to (C.254) and (C.255), (C.275) is redundant due to (C.253), (C.278) is redundant due to (C.253), (C.279) is redundant due to (C.270) and (C.276), (C.280) is redundant due to (C.256) and (C.277), (C.281) is redundant due to (C.256), (C.282) is redundant due to (C.256), (C.283) is redundant due to (C.255) and (C.256), (C.284) is redundant due to (C.255), (C.285) is redundant due to (C.257), (C.286) is redundant due to (C.255) and (C.257), (C.287) is redundant due to (C.255), (C.288) is redundant due to (C.291), (C.289) is redundant due to (C.255), (C.256), and (C.257), (C.290) is redundant due to (C.256), and (C.292) is due to (C.256) and (C.257). Finally, we obtain the following inequalities not involving T_2:

$$R_1 \leq d_{o1}, \tag{C.293}$$

$$R_1 \leq a_{o1} + e_{o2}, \tag{C.294}$$

$$R_2 \leq d_{o2}, \tag{C.295}$$

$$R_2 \leq a_{o2} + e_{o1}, \tag{C.296}$$

$$R_1 + R_2 \leq a_{o1} + g_{o2}, \tag{C.297}$$

$$R_1 + R_2 \leq a_{o2} + g_{o1}, \tag{C.298}$$

$$R_1 + R_2 \leq e_{o1} + e_{o2}, \tag{C.299}$$

$$2R_1 + R_2 \leq a_{o1} + g_{o1} + e_{o2}, \tag{C.300}$$

$$R_1 + 2R_2 \leq a_{o2} + g_{o2} + e_{o1}, \tag{C.301}$$

$$-R_1 \leq 0, \tag{C.302}$$

$$-R_2 \leq 0. \tag{C.303}$$

C.4 Proof of Lem. 4.5

Codebook Generation

Generate a codeword q^N of length N, generating each element i.i.d according to $\prod_{n=1}^{N} p_Q(q_n)$. For the codeword q^N, generate 2^{NT_1} conditionally independent codewords $w_1^N(j)$, $j \in \{1, 2, ..., 2^{NT_1}\}$, generating each element i.i.d according to $\prod_{n=1}^{N} p_{W_1|Q}(w_{1n}|q_n)$. For the codeword q^N, and each of the codewords $w_1^N(j)$, generate 2^{NS_1} conditionally independent codewords $x_1^N(j,k)$, $k \in \{1,2,...,2^{NS_1}\}$, generating each element i.i.d according to $\prod_{n=1}^{N} p_{X_1|W_1Q}(x_{1n}|q_n, w_{1n}(j))$. For the codeword q^N, generate 2^{NT_2} conditionally independent codewords $w_2^N(l)$, $l \in \{1,2,...,2^{NT_2}\}$, generating each element i.i.d according to $\prod_{n=1}^{N} p_{W_2|Q}(w_{2n}|q_n)$. For the codeword q^N, and each of the codewords $w_2^N(l)$, generate 2^{NS_2} conditionally independent codewords $x_2^N(l,m)$, $m \in \{1,2,...,2^{NS_2}\}$, generating each element i.i.d according to $\prod_{n=1}^{N} p_{X_2|W_2Q}(x_{2n}|q_n, w_{2n}(l))$.

Encoding

For encoder 1, to send the codeword pair (j,k), send the corresponding codeword $x_1^N(j,k)$. For encoder 2, to send the codeword pair (l,m), send the corresponding codeword $x_2^N(l,m)$.

Decoding

Receiver 1 determines the unique $\left(\hat{j},\hat{k}\right)$ and a $\hat{l}$ such that

$$\left(w_1^N\left(\hat{j}\right), x_1^N\left(\hat{j},\hat{k}\right), w_2^N\left(\hat{l}\right), y_1^N\right) \in A_\epsilon^{(N)}\left(W_1, X_1, W_2, Y_1\right). \tag{C.304}$$

Receiver 2 determines the unique $\left(\hat{\hat{l}},\hat{m}\right)$ and a $\hat{\hat{j}}$ such that

$$\left(w_2^N\left(\hat{\hat{l}}\right), x_2^N\left(\hat{\hat{l}},\hat{m}\right), w_1^N\left(\hat{\hat{j}}\right), y_2^N\right) \in A_\epsilon^{(N)}\left(W_2, X_2, W_1, Y_2\right). \tag{C.305}$$

Analysis of the Probability of Error

We consider only the decoding error of probability for receiver RX_1. The same analysis applies for receiver RX_2. By the symmetry of the random code construction, the conditional probability of error does not depend on which pair of indexes is sent. Thus the conditional probability of error is the same as the unconditional probability of error. So, without loss of generality, we assume that $(j,k) = (1,1)$

and $(l, m) = (1, 1)$ was sent.

We have an error if the correct codewords, $\left\{w_1^N(1), x_1^N(1,1), w_2^N(1)\right\}$ are not jointly typical with the received sequence. If incorrect codewords ($w_1^N\left(\hat{j}\right)$, $x_1^N\left(\hat{j},\hat{k}\right)$, $w_2^N\left(\hat{l}\right)$) are jointly typical with the received codeword, i.e., $\hat{j} \neq 1$ or $\hat{k} \neq 1$, an error is also declared. However, no error is declared if ($w_1^N(1)$, $x_1^N(1,1)$, $w_2\left(\hat{l} \neq 1\right)$) are jointly typical with the received sequence. Define the following event:

$$E_{jkl} = \left\{\left(w_1^N(j), x_1^N(j,k), w_2^N(l), y_1^N\right) \in A_\epsilon^{(N)}\right\}. \tag{C.306}$$

Then by the union of events bound,

$$\begin{aligned}
P_e^{(N)} &= P\left(E_{111}^c \bigcup \cup_{(j,k)\neq(1,1)} E_{jkl}\right) \\
&\leq P\left(E_{111}^c\right) + \sum_{j\neq1,k=1,l=1} P\left(E_{j11}\right) + \sum_{j\neq1,k=1,l\neq1} P\left(E_{j1l}\right) \\
&\quad + \sum_{j=1,k\neq1,l=1} P\left(E_{1k1}\right) + \sum_{j=1,k\neq1,l\neq1} P\left(E_{1kl}\right) \\
&\quad + \sum_{j\neq1,k\neq1,l=1} P\left(E_{jk1}\right) + \sum_{j\neq1,k\neq1,l\neq1} P\left(E_{jkl}\right) \\
&\leq P\left(E_{111}^c\right) + 2^{NT_1}2^{-N(I(X_1;Y_1|W_2Q)-4\epsilon)} \\
&\quad + 2^{N(T_1+T_2)}2^{-N(I(W_2X_1;Y_1|Q)-4\epsilon)} \\
&\quad + 2^{NS_1}2^{-N(I(X_1;Y_1|W_1W_2Q)-4\epsilon)} \\
&\quad + 2^{N(S_1+T_2)}2^{-N(I(W_2X_1;Y_1|W_1Q)-4\epsilon)} \\
&\quad + 2^{N(S_1+T_1)}2^{-N(I(X_1;Y_1|W_2Q)-4\epsilon)} \\
&\quad + 2^{N(S_1+T_1+T_2)}2^{-N(I(W_2X_1;Y_1|Q)-4\epsilon)}.
\end{aligned} \tag{C.307}$$

Since $\epsilon > 0$ is arbitrary, the conditions of Lem. 4.5 imply that each term tends to 0 as $N \to \infty$. The above bound shows that the average probability of error, averaged over all choices of codebooks in the random code construction, is arbitrarily small. Hence there exists at least one code $\mathcal{C}^*$ with arbitrarily small probability of error. We only consider the error probability of receiver RX_1. For $j \neq 1$, we have

$$
\begin{aligned}
&P\left(E_{j11}\right)=P\left(\left(q^{N}, w_{1}^{N}(j), x_{1}^{N}(j, 1), w_{2}^{N}(1), y_{1}^{N}\right) \in A_{\epsilon}^{(N)}\right) \\
&=\sum_{\left(q^{N}, w_{1}^{N}, x_{1}^{N}, w_{2}^{N}, y_{1}^{N}\right) \in A_{\epsilon}^{(N)}} p\left(w_{1}^{N}, x_{1}^{N} | q^{N}\right) p\left(w_{2}^{N}, y_{1}^{N} | q^{N}\right) p\left(q^{N}\right) \\
&\leq\left|A_{\epsilon}^{(n)}\right| 2^{-N\left(H\left(W_{1} X_{1} | Q\right)-\epsilon\right)} 2^{-N\left(H\left(W_{2} Y_{1} | Q\right)-\epsilon\right)} 2^{-N(H(Q)-\epsilon)} \\
&\leq 2^{-N\left(H\left(W_{1} X_{1} | Q\right)+H\left(W_{2} Y_{1} | Q\right)+H(Q)-H\left(Q W_{1} X_{1} W_{2} Y_{1}\right)-4 \epsilon\right)} \\
&=2^{-N\left(I\left(X_{1} ; Y_{1} | Q W_{2}\right)-4 \epsilon\right)} .
\end{aligned}
$$

For $j \neq 1$, $k \neq 1$ we have

$$
\begin{aligned}
&P\left(E_{jk1}\right)=P\left(\left(q^{N}, w_{1}^{N}(j), x_{1}^{N}(j, k), w_{2}^{N}(1), y_{1}^{N}\right) \in A_{\epsilon}^{(N)}\right) \\
&=\sum_{\left(q^{N}, w_{1}^{N}, x_{1}^{N}, w_{2}^{N}, y_{1}^{N}\right) \in A_{\epsilon}^{(N)}} p\left(w_{1}^{N}, x_{1}^{N} | q^{N}\right) p\left(w_{2}^{N}, y_{1}^{N} | q^{N}\right) p\left(q^{N}\right) \\
&\leq\left|A_{\epsilon}^{(n)}\right| 2^{-N\left(H\left(W_{1} X_{1} | Q\right)-\epsilon\right)} 2^{-N\left(H\left(W_{2} Y_{1} | Q\right)-\epsilon\right)} 2^{-N(H(Q)-\epsilon)} \\
&\leq 2^{-N\left(H\left(W_{1} X_{1} | Q\right)+H\left(W_{2} Y_{1} | Q\right)+H(Q)-H\left(Q W_{1} X_{1} W_{2} Y_{1}\right)-4 \epsilon\right)} \\
&=2^{-N\left(I\left(X_{1} ; Y_{1} | Q W_{2}\right)-4 \epsilon\right)} .
\end{aligned}
$$

For $k \neq 1$ we have

$$
\begin{aligned}
&P\left(E_{1k1}\right)=P\left(\left(q^{N}, w_{1}^{N}(1), x_{1}^{N}(1, k), w_{2}^{N}(1), y_{1}^{N}\right) \in A_{\epsilon}^{(N)}\right) \\
&=\sum_{\left(q^{N}, w_{1}^{N}, x_{1}^{N}, w_{2}^{N}, y_{1}^{N}\right) \in A_{\epsilon}^{(N)}} p\left(x_{1}^{N} | q^{N} w_{1}^{N}\right) p\left(w_{2}^{N} y_{1}^{N} | q^{N} w_{1}^{N}\right) p\left(q^{N} w_{1}^{N}\right) \\
&\leq\left|A_{\epsilon}^{(n)}\right| 2^{-N\left(H\left(X_{1} | Q W_{1}\right)-\epsilon\right)} 2^{-N\left(H\left(W_{2} Y_{1} | Q W_{1}\right)-\epsilon\right)} 2^{-N\left(H\left(Q W_{1}\right)-\epsilon\right)} \\
&\leq 2^{-N\left(H\left(X_{1} | Q W_{1}\right)+H\left(W_{2} Y_{1} | Q W_{1}\right)+H\left(Q W_{1}\right)-H\left(Q W_{1} X_{1} W_{2} Y_{1}\right)-4 \epsilon\right)} \\
&=2^{-N\left(I\left(X_{1} ; Y_{1} | Q W_{1} W_{2}\right)-4 \epsilon\right)} .
\end{aligned}
$$

For $j \neq 1$, $l \neq 1$ we have

$$
P\left(E_{j1l}\right)=P\left(\left(q^{N}, w_{1}^{N}(j), x_{1}^{N}(j, 1), w_{2}^{N}(l), y_{1}^{N}\right) \in A_{\epsilon}^{(N)}\right)
$$

$$
\begin{aligned}
&= \sum_{\left(q^N, w_1^N, x_1^N, w_2^N, y_1^N\right) \in A_\epsilon^{(N)}} p\left(w_1^N x_1^N w_2^N | q^N\right) p\left(y_1^N | q^N\right) p\left(q^N\right) \\
&\leq \left|A_\epsilon^{(n)}\right| 2^{-N(H(W_1 X_1 W_2|Q)-\epsilon)} 2^{-N(H(Y_1|Q)-\epsilon)} 2^{-N(H(Q)-\epsilon)} \\
&\leq 2^{-N(H(W_1 X_1 W_2|Q)+H(Y_1|Q)+H(Q)-H(QW_1 X_1 W_2 Y_1)-4\epsilon)} \\
&\leq 2^{-N(I(X_1 W_2;Y_1|Q)-4\epsilon)}.
\end{aligned}
$$

For $j \neq 1$, $k \neq 1$, $l \neq 1$ we have

$$
\begin{aligned}
P\left(E_{jkl}\right) &= P\left(\left(q^N, w_1^N(j), x_1^N(j,k), w_2^N(l), y_1^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N, w_1^N, x_1^N, w_2^N, y_1^N\right) \in A_\epsilon^{(N)}} p\left(w_1^N x_1^N w_2^N | q^N\right) p\left(y_1^N | q^N\right) p\left(q^N\right) \\
&\leq \left|A_\epsilon^{(n)}\right| 2^{-N(H(W_1 X_1 W_2|Q)-\epsilon)} 2^{-N(H(Y_1|Q)-\epsilon)} 2^{-N(H(Q)-\epsilon)} \\
&\leq 2^{-N(H(W_1 X_1 W_2|Q)+H(Y_1|Q)+H(Q)-H(QW_1 X_1 W_2 Y_1)-4\epsilon)} \\
&= 2^{-N(I(X_1 W_2;Y_1|Q)-4\epsilon)}.
\end{aligned}
$$

For $k \neq 1$, $l \neq 1$ we have

$$
\begin{aligned}
P\left(E_{1kl}\right) &= P\left(\left(q^N, w_1^N(1), x_1^N(1,k), w_2^N(l), y_1^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N, w_1^N, x_1^N, w_2^N, y_1^N\right) \in A_\epsilon^{(N)}} p\left(x_1^N w_2^N | q^N w_1^N\right) p\left(y_1^N | q^N w_1^N\right) p\left(q^N w_1^N\right) \\
&\leq \left|A_\epsilon^{(N)}\right| 2^{-N(H(X_1 W_2|QW_1)-\epsilon)} 2^{-N(H(Y_1|QW_1)-\epsilon)} 2^{-N(H(QW_1)-\epsilon)} \\
&= 2^{-N(H(X_1 W_2|QW_1)+H(Y_1|QW_1)+H(QW_1)-H(QW_1 X_1 W_2 Y_1)-4\epsilon)} \\
&\leq 2^{-N(I(X_1 W_2;Y_1|QW_1)-4\epsilon)}.
\end{aligned}
$$

C.5 Proof of the Achievability of Thm. 4.8

Let $\mathcal{P}_2^*$ be the set of probability distributions $P_2^*(.)$ that factor as

$$P_2^*\left(w_0, w_1, w_2, x_1, x_2, y_1, y_2\right)$$

$$= p(w_0)\, p(w_1|w_0)\, p(w_2|w_0)\, p(x_1|w_1, w_0)\, p(x_2|w_2, w_0)\, p(y_1, y_2|x_1, x_2) \quad \text{(C.308)}$$

Suppose we fix $P_2^*(.)$. Let $\mathcal{S}_{\text{JXG}}(P_2^*)$ be defined as the set of all $(R_0, S_1, T_1, S_2, T_2)$ such that

$$S_1 \le a_{c1}, \quad \text{(C.309)}$$

$$S_1 + T_1 \le b_{c1}, \quad \text{(C.310)}$$

$$S_1 + T_2 \le c_{c1}, \quad \text{(C.311)}$$

$$S_1 + T_1 + T_2 \le d_{c1}, \quad \text{(C.312)}$$

$$R_0 + S_1 + T_1 + T_2 \le e_{c1}, \quad \text{(C.313)}$$

and

$$S_2 \le a_{c2}, \quad \text{(C.314)}$$

$$S_2 + T_2 \le b_{c2}, \quad \text{(C.315)}$$

$$S_2 + T_1 \le c_{c2}, \quad \text{(C.316)}$$

$$S_2 + T_1 + T_2 \le d_{c2}, \quad \text{(C.317)}$$

$$R_0 + S_2 + T_1 + T_2 \le e_{c2}, \quad \text{(C.318)}$$

$$-R_0, -S_1, -T_1, -S_2, -T_2 \le 0, \quad \text{(C.319)}$$

$$\text{(C.320)}$$

where

$$a_{c1} = I(X_1; Y_1|W_0W_1W_2), \quad \text{(C.321)}$$

$$b_{c1} = I(X_1; Y_1|W_0W_2), \quad \text{(C.322)}$$

$$c_{c1} = I(W_2X_1; Y_1|W_0W_1), \quad \text{(C.323)}$$

$$d_{c1} = I(W_2X_1; Y_1|W_0), \quad \text{(C.324)}$$

$$e_{c1} = I(W_0W_2X_1; Y_1), \quad \text{(C.325)}$$

and

$$a_{c2} = I\left(X_2; Y_2 | W_0 W_1 W_2\right), \tag{C.326}$$

$$b_{c2} = I\left(X_2; Y_2 | W_0 W_1\right), \tag{C.327}$$

$$c_{c2} = I\left(W_1 X_2; Y_2 | W_0 W_2\right), \tag{C.328}$$

$$d_{c2} = I\left(W_1 X_2; Y_2 | W_0\right), \tag{C.329}$$

$$e_{c2} = I\left(W_0 W_1 X_2; Y_2\right). \tag{C.330}$$

Let $\mathcal{R}_{\mathrm{JXG}}\left(P_2^*\right)$ be the set of (R_0, R_1, R_2) such that $0 \leq R_1 \leq S_1 + T_1$ and $0 \leq R_2 \leq S_2 + T_2$ for some $(R_0, S_1, T_1, S_2, T_2) \in \mathcal{S}_{\mathrm{JXG}}\left(P_2^*\right)$. We have the following result:

Theorem C.1 (Jiang-Xin-Garg)**.** *The set*

$$\mathcal{R}_{\mathrm{JXG}} = \bigcup_{P_2^* \in \mathcal{P}_2^*} \mathcal{R}_{\mathrm{JXG}}\left(P_2^*\right) \tag{C.331}$$

is an achievable rate region for the discrete memoryless IFC.

Proof. Refer to [18]. □

To obtain the projected achievable rate region $\mathcal{R}_{\mathrm{JXG}}\left(P_2^*\right)$ using the Fourier-Motzkin elimination, we need the following additional inequalities:

$$R_1 - S_1 - T_1 \leq 0, \tag{C.332}$$

$$R_2 - S_2 - T_2 \leq 0, \tag{C.333}$$

$$-R_1 \leq 0, \tag{C.334}$$

$$-R_2 \leq 0. \tag{C.335}$$

In addition, it is easy to verify that the following information theoretic inequalities between the bound constants, $a_{ci}, ..., e_{ci}$, $i = 1, 2$, hold:

$$\begin{aligned} &a_{ci} \leq b_{ci}, c_{ci} \leq d_{ci} \leq e_{ci}, \\ &d_{ci} \leq b_{ci} + c_{ci}. \end{aligned} \tag{C.336}$$

Eliminate S_1: First collect all the inequalities not involving S_1 among all the inequalities to obtain

$$S_2 \leq a_{c2}, \tag{C.337}$$

$$S_2 + T_2 \leq b_{c2}, \tag{C.338}$$

$$S_2 + T_1 \leq c_{c2}, \tag{C.339}$$

$$S_2 + T_1 + T_2 \leq d_{c2}, \tag{C.340}$$

$$R_0 + S_2 + T_1 + T_2 \leq e_{c2}, \tag{C.341}$$

$$-R_0 \leq 0, \tag{C.342}$$

$$-T_1 \leq 0, \tag{C.343}$$

$$-S_2 \leq 0, \tag{C.344}$$

$$-T_2 \leq 0, \tag{C.345}$$

$$R_2 - S_2 - T_2 \leq 0, \tag{C.346}$$

$$-R_1 \leq 0, \tag{C.347}$$

$$-R_2 \leq 0. \tag{C.348}$$

Next, collect all the S_1 with positive coefficients to obtain

$$S_1 \leq a_{c1}, \tag{C.349}$$

$$S_1 + T_1 \leq b_{c1}, \tag{C.350}$$

$$S_1 + T_2 \leq c_{c1}, \tag{C.351}$$

$$S_1 + T_1 + T_2 \leq d_{c1}, \tag{C.352}$$

$$R_0 + S_1 + T_1 + T_2 \leq e_{c1}. \tag{C.353}$$

Furthermore, collect all the S_1 with negative coefficients to obtain

$$-S_1 \leq 0, \tag{C.354}$$

$$R_1 - S_1 - T_1 \leq 0. \tag{C.355}$$

We eliminate S_1 by adding each inequality from (C.349)-(C.353) and each inequality from (C.354), (C.355) to obtain inequalities not involving S_1:

$$0 \le a_{c1}, \tag{C.356}$$
$$T_1 \le b_{c1}, \tag{C.357}$$
$$T_2 \le c_{c1}, \tag{C.358}$$
$$T_1 + T_2 \le d_{c1}, \tag{C.359}$$
$$R_0 + T_1 + T_2 \le e_{c1}, \tag{C.360}$$
$$R_1 - T_1 \le a_{c1}, \tag{C.361}$$
$$R_1 \le b_{c1}, \tag{C.362}$$
$$R_1 + T_2 - T_1 \le c_{c1}, \tag{C.363}$$
$$R_1 + T_2 \le d_{c1}, \tag{C.364}$$
$$R_0 + R_1 + T_2 \le e_{c1}. \tag{C.365}$$

It is clear that (C.356) is redundant.

Eliminate S_2: First collect all the inequalities not involving S_2 among all the inequalities to obtain

$$-R_0 \le 0, \tag{C.366}$$
$$-T_1 \le 0, \tag{C.367}$$
$$-T_2 \le 0, \tag{C.368}$$
$$-R_1 \le 0, \tag{C.369}$$
$$-R_2 \le 0, \tag{C.370}$$
$$T_1 \le b_{c1}, \tag{C.371}$$
$$T_2 \le c_{c1}, \tag{C.372}$$
$$T_1 + T_2 \le d_{c1}, \tag{C.373}$$
$$R_0 + T_1 + T_2 \le e_{c1}, \tag{C.374}$$

$$R_1 - T_1 \leq a_{c1}, \tag{C.375}$$

$$R_1 \leq b_{c1}, \tag{C.376}$$

$$R_1 + T_2 - T_1 \leq c_{c1}, \tag{C.377}$$

$$R_1 + T_2 \leq d_{c1}, \tag{C.378}$$

$$R_0 + R_1 + T_2 \leq e_{c1}. \tag{C.379}$$

Next, collect all the S_2 with positive coefficients to obtain

$$S_2 \leq a_{c2}, \tag{C.380}$$

$$S_2 + T_2 \leq b_{c2}, \tag{C.381}$$

$$S_2 + T_1 \leq c_{c2}, \tag{C.382}$$

$$S_2 + T_1 + T_2 \leq d_{c2}, \tag{C.383}$$

$$R_0 + S_2 + T_1 + T_2 \leq e_{c2}. \tag{C.384}$$

Furthermore, collect all the S_2 with negative coefficients to obtain

$$-S_2 \leq 0, \tag{C.385}$$

$$R_2 - S_2 - T_2 \leq 0. \tag{C.386}$$

We eliminate S_2 by adding each inequality from (C.380)-(C.384) and each inequality from (C.385), (C.386) to obtain inequalities not involving S_2:

$$0 \leq a_{c2}, \tag{C.387}$$

$$T_2 \leq b_{c2}, \tag{C.388}$$

$$T_1 \leq c_{c2}, \tag{C.389}$$

$$T_1 + T_2 \leq d_{c2}, \tag{C.390}$$

$$R_0 + T_1 + T_2 \leq e_{c2}, \tag{C.391}$$

$$R_2 - T_2 \leq a_{c2}, \tag{C.392}$$

$$R_2 \leq b_{c2}, \tag{C.393}$$

$$R_2 + T_1 - T_2 \leq c_{c2}, \tag{C.394}$$

$$R_2 + T_1 \leq d_{c2}, \tag{C.395}$$

$$R_0 + R_2 + T_1 \leq e_{c2}. \tag{C.396}$$

It is clear that (C.387) is redundant.

Eliminate T_1: First collect all the inequalities not involving T_1 among all the inequalities to obtain

$$-R_0 \leq 0, \tag{C.397}$$

$$-T_2 \leq 0, \tag{C.398}$$

$$-R_1 \leq 0, \tag{C.399}$$

$$-R_2 \leq 0, \tag{C.400}$$

$$T_2 \leq c_{c1}, \tag{C.401}$$

$$R_1 \leq b_{c1}, \tag{C.402}$$

$$R_1 + T_2 \leq d_{c1}, \tag{C.403}$$

$$R_0 + R_1 + T_2 \leq e_{c1}, \tag{C.404}$$

$$T_2 \leq b_{c2}, \tag{C.405}$$

$$R_2 - T_2 \leq a_{c2}, \tag{C.406}$$

$$R_2 \leq b_{c2}. \tag{C.407}$$

Next, collect all the T_1 with positive coefficients to obtain

$$T_1 \leq b_{c1}, \tag{C.408}$$

$$T_1 + T_2 \leq d_{c1}, \tag{C.409}$$

$$R_0 + T_1 + T_2 \leq e_{c1}, \tag{C.410}$$

$$T_1 \leq c_{c2}, \tag{C.411}$$

$$T_1 + T_2 \leq d_{c2}, \tag{C.412}$$

$$R_0 + T_1 + T_2 \leq e_{c2}, \tag{C.413}$$

$$R_2 + T_1 - T_2 \leq c_{c2}, \tag{C.414}$$

$$R_2 + T_1 \leq d_{c2}, \tag{C.415}$$

$$R_0 + R_2 + T_1 \leq e_{c2}. \tag{C.416}$$

Furthermore, collect all the T_1 with negative coefficients to obtain

$$-T_1 \leq 0, \tag{C.417}$$

$$R_1 - T_1 \leq a_{c1}, \tag{C.418}$$

$$R_1 + T_2 - T_1 \leq c_{c1}. \tag{C.419}$$

We eliminate T_1 by adding each inequality from (C.408)-(C.416) and each inequality from (C.417)-(C.419) to obtain inequalities not involving T_1:

$$0 \leq b_{c1}, \tag{C.420}$$

$$T_2 \leq d_{c1}, \tag{C.421}$$

$$R_0 + T_2 \leq e_{c1}, \tag{C.422}$$

$$0 \leq c_{c2}, \tag{C.423}$$

$$T_2 \leq d_{c2}, \tag{C.424}$$

$$R_0 + T_2 \leq e_{c2}, \tag{C.425}$$

$$R_2 - T_2 \leq c_{c2}, \tag{C.426}$$

$$R_2 \leq d_{c2}, \tag{C.427}$$

$$R_0 + R_2 \leq e_{c2}, \tag{C.428}$$

$$R_1 \leq a_{c1} + b_{c1}, \tag{C.429}$$

$$R_1 + T_2 \leq a_{c1} + d_{c1}, \tag{C.430}$$

$$R_0 + R_1 + T_2 \leq a_{c1} + e_{c1}, \tag{C.431}$$

$$R_1 \leq a_{c1} + c_{c2}, \tag{C.432}$$

$$R_1 + T_2 \leq a_{c1} + d_{c2}, \tag{C.433}$$

$$R_0 + R_1 + T_2 \leq a_{c1} + e_{c2}, \tag{C.434}$$

$$R_1 + R_2 - T_2 \leq a_{c1} + c_{c2}, \tag{C.435}$$

$$R_1 + R_2 \leq a_{c1} + d_{c2}, \tag{C.436}$$

$$R_0 + R_1 + R_2 \leq a_{c1} + e_{c2}, \tag{C.437}$$

$$R_1 + T_2 \leq c_{c1} + b_{c1}, \tag{C.438}$$

$$R_1 + 2T_2 \leq c_{c1} + d_{c1}, \tag{C.439}$$

$$R_0 + R_1 + 2T_2 \leq c_{c1} + e_{c1}, \tag{C.440}$$

$$R_1 + T_2 \leq c_{c1} + c_{c2}, \tag{C.441}$$

$$R_1 + 2T_2 \leq c_{c1} + d_{c2}, \tag{C.442}$$

$$R_0 + R_1 + 2T_2 \leq c_{c1} + e_{c2}, \tag{C.443}$$

$$R_1 + R_2 \leq c_{c1} + c_{c2}, \tag{C.444}$$

$$R_1 + R_2 + T_2 \leq c_{c1} + d_{c2}, \tag{C.445}$$

$$R_0 + R_1 + R_2 + T_2 \leq c_{c1} + e_{c2}. \tag{C.446}$$

It is clear that (C.420) and (C.423) are redundant, (C.421) is redundant due to (C.401), (C.424) is redundant due to (C.405), (C.426) is redundant due to (C.406), (C.427) is redundant due to (C.407), (C.429) is redundant due to (C.402), (C.430) is redundant due to (C.403), (C.431) is redundant due to (C.404), and (C.438) is redundant due to (C.403) ($c_{c1} + b_{c1} \geq d_{c1}$).

Eliminate T_2: First collect all the inequalities not involving T_2 among all the inequalities to obtain

$$-R_0 \leq 0, \tag{C.447}$$

$$-R_1 \leq 0, \tag{C.448}$$

$$-R_2 \leq 0, \tag{C.449}$$

$$R_1 \leq b_{c1}, \tag{C.450}$$

$$R_2 \leq b_{c2}, \tag{C.451}$$

$$R_0 + R_2 \leq e_{c2}, \tag{C.452}$$

$$R_1 \leq a_{c1} + c_{c2}, \tag{C.453}$$

$$R_1 + R_2 \leq a_{c1} + d_{c2}, \tag{C.454}$$

$$R_0 + R_1 + R_2 \leq a_{c1} + e_{c2}, \tag{C.455}$$

$$R_1 + R_2 \leq c_{c1} + c_{c2}. \tag{C.456}$$

(C.457)

Next, collect all the T_2 with positive coefficients to obtain

$$T_2 \leq c_{c1}, \tag{C.458}$$

$$R_1 + T_2 \leq d_{c1}, \tag{C.459}$$

$$R_0 + R_1 + T_2 \leq e_{c1}, \tag{C.460}$$

$$T_2 \leq b_{c2}, \tag{C.461}$$

$$R_0 + T_2 \leq e_{c1}, \tag{C.462}$$

$$R_0 + T_2 \leq e_{c2}, \tag{C.463}$$

$$R_1 + T_2 \leq a_{c1} + d_{c2}, \tag{C.464}$$

$$R_0 + R_1 + T_2 \leq a_{c1} + e_{c2}, \tag{C.465}$$

$$R_1 + 2T_2 \leq c_{c1} + d_{c1}, \tag{C.466}$$

$$R_0 + R_1 + 2T_2 \leq c_{c1} + e_{c1}, \tag{C.467}$$

$$R_1 + T_2 \leq c_{c1} + c_{c2}, \tag{C.468}$$

$$R_1 + 2T_2 \leq c_{c1} + d_{c2}, \tag{C.469}$$

$$R_0 + R_1 + 2T_2 \leq c_{c1} + e_{c2}, \tag{C.470}$$

$$R_1 + R_2 + T_2 \leq c_{c1} + d_{c2}, \tag{C.471}$$

$$R_0 + R_1 + R_2 + T_2 \leq c_{c1} + e_{c2}. \tag{C.472}$$

Furthermore, collect all the T_2 with negative coefficients to obtain

$$-T_2 \leq 0, \tag{C.473}$$

$$R_2 - T_2 \leq a_{c2}, \tag{C.474}$$

$$R_1 + R_2 - T_2 \leq a_{c1} + c_{c2}. \tag{C.475}$$

$$\tag{C.476}$$

We eliminate T_2 by adding each inequality from (C.458)-(C.472) and each inequality from (C.473)-(C.475) to obtain inequalities not involving T_2:

$$0 \leq c_{c1}, \tag{C.477}$$

$$R_1 \leq d_{c1}, \tag{C.478}$$

$$R_0 + R_1 \leq e_{c1}, \tag{C.479}$$

$$0 \leq b_{c2}, \tag{C.480}$$

$$R_0 \leq e_{c1}, \tag{C.481}$$

$$R_0 \leq e_{c2}, \tag{C.482}$$

$$R_1 \leq a_{c1} + d_{c2}, \tag{C.483}$$

$$R_0 + R_1 \leq a_{c1} + e_{c2}, \tag{C.484}$$

$$R_1 \leq c_{c1} + d_{c1}, \tag{C.485}$$

$$R_0 + R_1 \leq c_{c1} + e_{c1}, \tag{C.486}$$

$$R_1 \leq c_{c1} + c_{c2}, \tag{C.487}$$

$$R_1 \leq c_{c1} + d_{c2}, \tag{C.488}$$

$$R_0 + R_1 \leq c_{c1} + e_{c2}, \tag{C.489}$$

$$R_1 + R_2 \leq c_{c1} + d_{c2}, \tag{C.490}$$

$$R_0 + R_1 + R_2 \leq c_{c1} + e_{c2}, \tag{C.491}$$

$$R_2 \leq c_{c1} + a_{c2}, \tag{C.492}$$

$$R_1 + R_2 \leq d_{c1} + a_{c2}, \tag{C.493}$$

$$R_0 + R_1 + R_2 \le e_{c1} + a_{c2}, \tag{C.494}$$

$$R_2 \le a_{c2} + b_{c2}, \tag{C.495}$$

$$R_0 + R_2 \le e_{c1} + a_{c2}, \tag{C.496}$$

$$R_0 + R_2 \le a_{c2} + e_{c2}, \tag{C.497}$$

$$R_1 + R_2 \le a_{c1} + a_{c2} + d_{c2}, \tag{C.498}$$

$$R_0 + R_1 + R_2 \le a_{c1} + a_{c2} + e_{c2}, \tag{C.499}$$

$$R_1 + 2R_2 \le c_{c1} + d_{c1} + 2a_{c2}, \tag{C.500}$$

$$R_0 + R_1 + 2R_2 \le c_{c1} + e_{c1} + 2a_{c2}, \tag{C.501}$$

$$R_1 + R_2 \le c_{c1} + a_{c2} + c_{c2}, \tag{C.502}$$

$$R_1 + 2R_2 \le c_{c1} + 2a_{c2} + d_{c2}, \tag{C.503}$$

$$R_0 + R_1 + 2R_2 \le c_{c1} + 2a_{c2} + e_{c2}, \tag{C.504}$$

$$R_1 + 2R_2 \le c_{c1} + a_{c2} + d_{c2}, \tag{C.505}$$

$$R_0 + R_1 + 2R_2 \le c_{c1} + a_{c2} + e_{c2}, \tag{C.506}$$

$$R_1 + R_2 \le a_{c1} + c_{c1} + c_{c2}, \tag{C.507}$$

$$2R_1 + R_2 \le a_{c1} + d_{c1} + c_{c2}, \tag{C.508}$$

$$R_0 + 2R_1 + R_2 \le a_{c1} + e_{c1} + c_{c2}, \tag{C.509}$$

$$R_1 + R_2 \le a_{c1} + b_{c2} + c_{c2}, \tag{C.510}$$

$$R_0 + R_1 + R_2 \le a_{c1} + e_{c1} + c_{c2}, \tag{C.511}$$

$$R_0 + R_1 + R_2 \le a_{c1} + c_{c2} + e_{c2}, \tag{C.512}$$

$$2R_1 + R_2 \le 2a_{c1} + c_{c2} + d_{c2}, \tag{C.513}$$

$$R_0 + 2R_1 + R_2 \le 2a_{c1} + c_{c2} + e_{c2}, \tag{C.514}$$

$$3R_1 + 2R_2 \le 2a_{c1} + c_{c1} + d_{c1} + 2c_{c2}, \tag{C.515}$$

$$R_0 + 3R_1 + 2R_2 \le 2a_{c1} + c_{c1} + e_{c1} + 2c_{c2}, \tag{C.516}$$

$$2R_1 + R_2 \le a_{c1} + c_{c1} + 2c_{c2}, \tag{C.517}$$

$$3R_1 + 2R_2 \le 2a_{c1} + c_{c1} + 2c_{c2} + d_{c2}, \tag{C.518}$$

$$R_0 + 3R_1 + 2R_2 \le 2a_{c1} + c_{c1} + 2c_{c2} + e_{c2}, \tag{C.519}$$

$$2R_1 + 2R_2 \leq a_{c1} + c_{c1} + c_{c2} + d_{c2}, \tag{C.520}$$

$$R_0 + 2R_1 + 2R_2 \leq a_{c1} + c_{c1} + c_{c2} + e_{c2}. \tag{C.521}$$

It is clear that (C.477) and (C.480) are redundant, (C.478) is redundant due to (C.450), (C.481) is redundant due to (C.479), (C.482) is redundant due to (C.452), (C.483) is redundant due to (C.453), (C.484) is redundant due to (C.479), (C.485) is redundant due to (C.493), (C.486) is redundant due to (C.479), (C.487) is redundant due to (C.453), (C.488) is redundant due to (C.453), (C.489) is redundant due to (C.479), (C.490) is redundant due to (C.454), (C.491) is redundant due to (C.455), (C.495) is redundant due to (C.451), (C.496) is redundant due to (C.452), (C.497) is redundant due to (C.452), (C.498) is redundant due to (C.454), (C.499) is redundant due to (C.455), (C.500) is redundant due to (C.492) and (C.493), (C.501) is redundant due to (C.492) and (C.494), (C.502) is redundant due to (C.456), (C.503) is redundant due to (C.505), (C.504) is redundant due to (C.506), (C.507) is redundant due to (C.456), (C.510) is redundant due to (C.454) ($b_{c2} + c_{c2} \geq d_{c2}$), (C.511) is redundant due to (C.509), (C.512) is redundant due to (C.455), (C.513) is redundant due to (C.453) and (C.454), (C.514) is redundant due to (C.453) and (C.454), (C.515) is redundant due to (C.456) and (C.508), (C.516) is redundant due to (C.456) and (C.509), (C.517) is redundant due to (C.453) and (C.456), (C.518) is redundant due to (C.453), (C.454), and (C.456) (C.519) is redundant due to (C.453), (C.455), and (C.456), (C.520) is redundant due to (C.454) and (C.456), and (C.521) is redundant due to (C.455) and (C.456). Hence, we obtain the following inequalities not involving T_2:

$$R_1 \leq b_{c1}, \tag{C.522}$$

$$R_1 \leq a_{c1} + c_{c2}, \tag{C.523}$$

$$R_2 \leq b_{c2}, \tag{C.524}$$

$$R_2 \leq c_{c1} + a_{c2}, \tag{C.525}$$

$$R_0 + R_2 \leq e_{c2}, \tag{C.526}$$

$$R_0 + R_1 \leq e_{c1}, \tag{C.527}$$

$$R_1 + R_2 \leq a_{c1} + d_{c2}, \tag{C.528}$$

$$R_1 + R_2 \leq c_{c1} + c_{c2}. \tag{C.529}$$

$$R_1 + R_2 \leq d_{c1} + a_{c2}, \tag{C.530}$$

$$R_1 + 2R_2 \leq c_{c1} + a_{c2} + d_{c2}, \tag{C.531}$$

$$2R_1 + R_2 \leq a_{c1} + d_{c1} + c_{c2}, \tag{C.532}$$

$$R_0 + R_1 + R_2 \leq a_{c1} + e_{c2}, \tag{C.533}$$

$$R_0 + R_1 + R_2 \leq e_{c1} + a_{c2}, \tag{C.534}$$

$$R_0 + R_1 + 2R_2 \leq c_{c1} + a_{c2} + e_{c2}, \tag{C.535}$$

$$R_0 + 2R_1 + R_2 \leq a_{c1} + e_{c1} + c_{c2}, \tag{C.536}$$

$$-R_0 \leq 0, \tag{C.537}$$

$$-R_1 \leq 0, \tag{C.538}$$

$$-R_2 \leq 0. \tag{C.539}$$

Appendix D

Proof of Theorems in Chapter 5

D.1 Proof of Thm. 5.11

By the symmetry of the random code generation, the conditional probability of error does not depend on which indices are sent. Therefore, we may assume that the message

$$(m_1, (m_{211}, m_{212}), (m_{221}, m_{222})) = (1, (1, 1), (1, 1))$$

is sent. Let $P(.)$ denote the conditional probability that $(1, (1, 1), (1, 1))$ is sent. For receiver RX_1, we define the following events:

$$E_{ijkm} = \Big\{ \left(q^N, x_1^N(i), w^N(k, m), u_1^N(j, k, m), y_1^N\right) \in A_\epsilon^{(N)}(Q, X_1, W, U_1, Y_1) \Big\}. \quad \text{(D.1)}$$

Then we can bound the probability of error as follows:

$$P_e^{(N)}(1) = P\left(E_{1111}^c \bigcup \cup_{(i,j,k,m)\neq(1,1,1,1)} E_{ijkm}\right)$$
$$\leq P(E_{1111}^c) + \sum_{\substack{i\neq1,j\neq1,\\(k,m)\neq(1,1)}} P(E_{ijkm}) + \sum_{\substack{i=1,j\neq1,\\(k,m)\neq(1,1)}} P(E_{1jkm})$$

$$+ \sum_{\substack{i\neq 1, j=1, \\ (k,m)\neq(1,1)}} P\left(E_{i1km}\right) + \sum_{\substack{i\neq 1, j\neq 1, \\ (k,m)=(1,1)}} P\left(E_{ij11}\right) + \sum_{\substack{i\neq 1, j=1, \\ (k,m)=(1,1)}} P\left(E_{i111}\right)$$
$$+ \sum_{\substack{i=1, j\neq 1, \\ (k,m)=(1,1)}} P\left(E_{1j11}\right) + \sum_{\substack{i=1, j=1, \\ (k,m)\neq(1,1)}} P\left(E_{11km}\right). \tag{D.2}$$

For $(i,j,(k,m)) \neq (1,1,(1,1))$, we have

$$\begin{aligned}
P\left(E_{ijkm}\right) &= P\left(\left(q^N, x_1^N(i), w^N(k,m), u_1^N(j,k,m), y_1^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N, x_1^N, w^N, u_1^N, y_1^N\right) \in A_\epsilon^{(N)}} p\left(q^N\right) p\left(x_1^N, w^N, u_1^N | q^N\right) p\left(y_1^N | q^N\right) \\
&\leq \left\| A_\epsilon^{(N)} \right\| 2^{-N(H(Q)-\epsilon+H(X_1WU_1|Q)-2\epsilon+H(Y_1|Q)-2\epsilon)} \\
&\leq 2^{-N(H(Q)+H(X_1WU_1|Q)+H(Y_1|Q)-H(QX_1WU_1Y_1)-6\epsilon)} \\
&= 2^{-N(I(X_1WU_1;Y_1|Q)-6\epsilon)}.
\end{aligned} \tag{D.3}$$

For $(j,(k,m)) \neq (1,(1,1))$, we have

$$\begin{aligned}
P\left(E_{1jkm}\right) &= P\left(\left(q^N, x_1^N(1), w^N(k,m), u_1^N(j,k,m), y_1^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N, x_1^N, w^N, u_1^N, y_1^N\right) \in A_\epsilon^{(N)}} p\left(q^N, x_1^N\right) p\left(w^N, u_1^N | q^N\right) p\left(y_1^N | x_1^N, q^N\right) \\
&\leq \left\| A_\epsilon^{(N)} \right\| 2^{-N(H(QX_1)-\epsilon+H(WU_1|Q)-2\epsilon+H(Y_1|X_1Q)-2\epsilon)} \\
&\leq 2^{-N(H(QX_1)+H(WU_1|Q)+H(Y_1|X_1Q)+H(QX_1WU_1Y_1)-6\epsilon)} \\
&= 2^{-N(I(WU_1;Y_1|X_1Q)-6\epsilon)}.
\end{aligned} \tag{D.4}$$

For $(i,(k,m)) \neq (1,(1,1))$, we have

$$\begin{aligned}
P\left(E_{i1km}\right) &= P\left(\left(q^N, x_1^N(i), w^N(k,m), u_1^N(1,k,m), y_1^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N, x_1^N, w^N, u_1^N, y_1^N\right) \in A_\epsilon^{(N)}} p\left(q^N\right) p\left(x_1^N, w^N, u_1^N | q^N\right) p\left(y_1^N | q^N\right) \\
&\leq \left\| A_\epsilon^{(N)} \right\| 2^{-N(H(Q)-\epsilon+H(X_1WU_1|Q)-2\epsilon+H(Y_1|Q)-2\epsilon)} \\
&\leq 2^{-N(H(Q)+H(X_1WU_1|Q)+H(Y_1|Q)-H(QX_1WU_1Y_1)-6\epsilon)}
\end{aligned}$$

$$= 2^{-N(I(X_1WU_1;Y_1|Q)-6\epsilon)}. \tag{D.5}$$

For $(i,j) \neq (1,1)$, we have

$$\begin{aligned}
P\left(E_{ij11}\right) &= P\left(\left(q^N, x_1^N(i), w^N(1,1), u_1^N(j,1,1), y_1^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N, x_1^N, w^N, u_1^N, y_1^N\right) \in A_\epsilon^{(N)}} p\left(q^N, w^N\right) p\left(x_1^N, u_1^N | w^N, q^N\right) p\left(y_1^N | w^N, q^N\right) \\
&\leq \left\| A_\epsilon^{(N)} \right\| 2^{-N(H(QW)-\epsilon+H(X_1U_1|WQ)-2\epsilon+H(Y_1|WQ)-2\epsilon)} \\
&\leq 2^{-N(H(QW)+H(X_1U_1|WQ)+H(Y_1|WQ)-H(QX_1WU_1Y_1)-6\epsilon)} \\
&= 2^{-N(I(X_1U_1;Y_1|WQ)-6\epsilon)}.
\end{aligned} \tag{D.6}$$

For $i \neq 1$, we have

$$\begin{aligned}
P\left(E_{i111}\right) &= P\left(\left(q^N, x_1^N(i), w^N(1,1), u_1^N(1,1,1), y_1^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N, x_1^N, w^N, u_1^N, y_1^N\right) \in A_\epsilon^{(N)}} p\left(q^N, w^N, u_1^N\right) p\left(x_1^N | q^N\right) p\left(y_1^N | w^N, u_1^N, q^N\right) \\
&\leq \left\| A_\epsilon^{(N)} \right\| 2^{-N(H(QWU_1)-\epsilon+H(X_1|Q)-2\epsilon+H(Y_1|WU_1Q)-2\epsilon)} \\
&\leq 2^{-N(H(QWU_1)+H(X_1|Q)+H(Y_1|WU_1Q)-H(QX_1WU_1Y_1)-6\epsilon)} \\
&= 2^{-N(I(X_1;Y_1|WU_1Q)-6\epsilon)}.
\end{aligned} \tag{D.7}$$

For $j \neq 1$, we have

$$\begin{aligned}
P\left(E_{1j11}\right) &= P\left(\left(q^N, x_1^N(1), w^N(1,1), u_1^N(j,1,1), y_1^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N, x_1^N, w^N, u_1^N, y_1^N\right) \in A_\epsilon^{(N)}} p\left(q^N, w^N, x_1^N\right) p\left(u_1^N | w^N, q^N\right) p\left(y_1^N | x_1^N, w^N, q^N\right) \\
&\leq \left\| A_\epsilon^{(N)} \right\| 2^{-N(H(QWX_1)-\epsilon+H(U_1|WQ)-2\epsilon+H(Y_1|X_1WQ)-2\epsilon)} \\
&\leq 2^{-N(H(QWX_1)+H(U_1|WQ)+H(Y_1|X_1WQ)-H(QX_1WU_1Y_1)-6\epsilon)} \\
&= 2^{-N(I(U_1;Y_1|X_1WQ)-6\epsilon)}.
\end{aligned} \tag{D.8}$$

For $(k,m) \neq (1,1)$, we have

$$\begin{aligned}
P\left(E_{11km}\right) &= P\left(\left(q^N, x_1^N(1), w^N(k,m), u_1^N(1,k,m), y_1^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N, x_1^N, w^N, u_1^N, y_1^N\right) \in A_\epsilon^{(N)}} p\left(q^N, x_1^N\right) p\left(w^N, u_1^N | q^N\right) p\left(y_1^N | x_1^N, q^N\right) \\
&\leq \left\| A_\epsilon^{(N)} \right\| 2^{-N(H(QX_1)-\epsilon+H(WU_1|Q)-2\epsilon+H(Y_1|X_1Q)-2\epsilon)} \\
&\leq 2^{-N(H(QX_1)+H(WU_1|Q)+H(Y_1|X_1Q)-H(QX_1WU_1Y_1)-6\epsilon)} \\
&= 2^{-N(I(WU_1;Y_1|X_1Q)-6\epsilon)}.
\end{aligned} \tag{D.9}$$

We may then bound the probability of error at receiver RX_1 as follows:

$$\begin{aligned}
P_e^{(N)}(1) \leq P\left(E_{1111}^c\right) &+ 2^{N(S_1+S_{21}+T_{21}+T_{22})} 2^{-N(I(X_1WU_1;Y_1|Q)-6\epsilon)} \\
&+ 2^{N(S_{21}+T_{21}+T_{22})} 2^{-N(I(WU_1;Y_1|X_1Q)-6\epsilon)} \\
&+ 2^{N(S_1+T_{21}+T_{22})} 2^{-N(I(X_1WU_1;Y_1|Q)-6\epsilon)} \\
&+ 2^{N(S_1+S_{21})} 2^{-N(I(X_1U_1;Y_1|WQ)-6\epsilon)} \\
&+ 2^{NS_1} 2^{-N(I(X_1;Y_1|WU_1Q)-6\epsilon)} \\
&+ 2^{NS_{21}} 2^{-N(I(U_1;Y_1|X_1WQ)-6\epsilon)} \\
&+ 2^{N(T_{21}+T_{22})} 2^{-N(I(WU_1;Y_1|X_1Q)-6\epsilon)}.
\end{aligned} \tag{D.10}$$

For receiver RX_2, we define the following events:

$$E_{lkm} = \left(\left(q^N, w^N(k,m), u_2^N(l,k,m), y_2^N\right) \in A_\epsilon^{(N)}(Q, W, U_2, Y_2)\right). \tag{D.11}$$

Then we can bound the probability of error as follows:

$$\begin{aligned}
P_e^{(N)}(2) &= P\left(E_{111}^c \bigcup \cup_{(l,k,m)\neq(1,1,1)} E_{lkm}\right) \\
&\leq P\left(E_{111}^c\right) + \sum_{l\neq 1,(k,m)\neq(1,1)} P\left(E_{lkm}\right) \\
&\quad + \sum_{\substack{l=1,\\(k,m)\neq(1,1)}} P\left(E_{1km}\right) + \sum_{\substack{l\neq 1,\\(k,m)=(1,1)}} P\left(E_{l11}\right).
\end{aligned} \tag{D.12}$$

For $(l,(k,m)) \neq (1,(1,1))$, we have

$$\begin{aligned}
P(E_{lkm}) &= P\left(\left(q^N, w^N(k,m), u_2^N(l,k,m), y_2^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N,w^N,u_2^N,y_2^N\right)\in A_\epsilon^{(N)}} p\left(q^N\right) p\left(w^N, u_2^N|q^N\right) p\left(y_2^N|q^N\right) \\
&\leq \left\|A_\epsilon^{(N)}\right\| 2^{-N(H(Q)-\epsilon)} 2^{-N(H(WU_2|Q)-2\epsilon)} 2^{-N(H(Y_2|Q)-2\epsilon)} \\
&\leq 2^{-N(H(Q)+H(WU_2|Q)+H(Y_2|Q)-H(QWU_2Y_2)-6\epsilon)} \\
&= 2^{-N(I(WU_2;Y_2|Q)-6\epsilon)}. \qquad \text{(D.13)}
\end{aligned}$$

For $(k,m) \neq (1,1)$, we have

$$\begin{aligned}
P(E_{1km}) &= P\left(\left(q^N, w^N(k,m), u_2^N(1,k,m), y_2^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N,w^N,u_2^N,y_2^N\right)\in A_\epsilon^{(N)}} p\left(q^N\right) p\left(w^N, u_2^N|q^N\right) p\left(y_2^N|q^N\right) \\
&\leq \left\|A_\epsilon^{(N)}\right\| 2^{-N(H(Q)-\epsilon)} 2^{-N(H(WU_2|Q)-2\epsilon)} 2^{-N(H(Y_2|Q)-2\epsilon)} \\
&\leq 2^{-N(H(Q)+H(WU_2|Q)+H(Y_2|Q)-H(QWU_2Y_2)-6\epsilon)} \\
&= 2^{-N(I(WU_2;Y_2|Q)-6\epsilon)}. \qquad \text{(D.14)}
\end{aligned}$$

For $l \neq 1$, we have

$$\begin{aligned}
P(E_{l11}) &= P\left(\left(q^N, w^N(1,1), u_2^N(l,1,1), y_2^N\right) \in A_\epsilon^{(N)}\right) \\
&= \sum_{\left(q^N,w^N,u_2^N,y_2^N\right)\in A_\epsilon^{(N)}} p\left(q^N, w^N\right) p\left(u_2^N|w^N, q^N\right) p\left(y_2^N|w^N, q^N\right) \\
&\leq \left\|A_\epsilon^{(N)}\right\| 2^{-N(H(QW)-\epsilon+H(U_2|WQ)-2\epsilon+H(Y_2|WQ)-2\epsilon)} \\
&\leq 2^{-N(H(QW)+H(U_2|WQ)+H(Y_2|WQ)-H(QWU_2Y_2)-6\epsilon)} \\
&= 2^{-N(I(U_2;Y_2|WQ)-6\epsilon)}. \qquad \text{(D.15)}
\end{aligned}$$

We may then bound the probability of error at receiver RX_2 as follows:

$$P_e^{(N)}(2) \leq P\left(E_{111}^c\right) + 2^{N(S_{22}+T_{21}+T_{22})} 2^{-N(I(WU_2;Y_2|Q)-6\epsilon)}$$

$$+ 2^{N(T_{21}+T_{22})} 2^{-N(I(WU_2;Y_2|Q)-6\epsilon)}$$
$$+ 2^{NS_{22}} 2^{-N(I(U_2;Y_2|WQ)-6\epsilon)}. \tag{D.16}$$

Since $\epsilon > 0$ is arbitrary, the conditions of Thm. 5.11 ensure that each of the terms in (D.10) and (D.16) tends to 0 as $N \to \infty$.

D.2 Proof of Thm. 5.5

By Fano's inequality, we have

$$H\left(M_{21}|Y_1^N\right) \leq N\epsilon_{1N} \tag{D.17}$$
$$H\left(M_{22}|Y_2^N\right) \leq N\epsilon_{2N} \tag{D.18}$$
$$H\left(M_1|Y_1^N\right) \leq N\epsilon_{3N} \tag{D.19}$$

where $\epsilon_{1N}, \epsilon_{2N}, \epsilon_{3N} \to 0$ as $N \to \infty$. We first bound R_{21} as follows:

$$\begin{aligned}
NR_{21} &= I\left(M_{21};Y_1^N\right) + H\left(M_{21}|Y_1^N\right) \\
&\leq I\left(M_{21};Y_1^N\right) + N\epsilon_{1N} \\
&= H\left(M_{21}\right) - H\left(M_{21}|Y_1^N\right) + N\epsilon_{1N} \\
&\overset{(a)}{=} H\left(M_{21}|X_1^N\left(M_1\right)\right) - H\left(M_{21}|Y_1^N\right) + N\epsilon_{1N} \\
&\leq H\left(M_{21}|X_1^N\right) - H\left(M_{21}|X_1^N Y_1^N\right) + N\epsilon_{1N} \\
&= I\left(M_{21};Y_1^N|X_1^N\right) + N\epsilon_{1N} \\
&= \sum_{n=1}^{n=N} I\left(M_{21};Y_{1n}|X_1^N Y_1^{n-1}\right) + N\epsilon_{1N} \\
&= \sum_{n=1}^{n=N} H\left(Y_{1n}|X_1^N Y_1^{n-1}\right) - H\left(Y_{1n}|X_1^N Y_1^{n-1} M_{21}\right) + N\epsilon_{1N} \\
&\leq \sum_{n=1}^{n=N} H\left(Y_{1n}|X_{1n}\right) - H\left(Y_{1n}|X_1^N Y_1^{n-1} M_{21} Y_2^{n-1}\right) + N\epsilon_{1N}
\end{aligned}$$

$$\overset{(b)}{=}\sum_{n=1}^{n=N} H\left(Y_{1n}|X_{1n}\right)-H\left(Y_{1n}|X_1^N M_{21} Y_2^{n-1}\right)+N\epsilon_{1N}$$
$$\overset{(c)}{=}\sum_{n=1}^{n=N} H\left(Y_{1n}|X_{1n}\right)-H\left(Y_{1n}|X_{1n} M_{21} Y_2^{n-1}\right)+N\epsilon_{1N}$$
$$=\sum_{n=1}^{n=N} H\left(Y_{1n}|X_{1n}\right)-H\left(Y_{1n}|X_{1n} W_n\right)+N\epsilon_{1N}$$
$$=\sum_{n=1}^{n=N} I\left(W_n;Y_{1n}|X_{1n}\right)+N\epsilon_{1N}. \tag{D.20}$$

where we define the random variable $W_n=\left(M_{21},Y_2^{n-1}\right)$ for all n, (a) follows from the fact that since M_{21} and M_1 are independent, so are M_{21} and $X_1^N\left(M_1\right)$, and (b) follows from the fact that $\left(M_{21}X_{1n}^N Y_{1n}\right)\rightarrow\left(X_1^{n-1}Y_2^{n-1}\right)\rightarrow Y_1^{n-1}$ form a Markov chain. This is due to the memoryless property of the channel and the fact that for any i, Y_{1i} depends only on Y_{2i} and X_{1i} (refer to (5.13)). Finally, (c) follows from the fact that $\left(X_1^{n-1}X_{1n+1}^N\right)\rightarrow\left(M_{21}Y_2^{n-1}X_{1n}\right)\rightarrow Y_{1n}$ form a Markov chain. We can prove this using the functional dependence graph technique introduced in [62]. Alternatively, we first note the following Markov chain:

$$\left(X_1^{n-1}X_{1n+1}^N W_n\right)\rightarrow\left(X_{1n}Y_{2n}\right)\rightarrow Y_{1n} \tag{D.21}$$

which follows from the fact that Y_{1n} depends only on Y_{2n} and X_{1n}. Using the weak union property, we obtain the following Markov chain:

$$\left(X_1^{n-1}X_{1n+1}^N\right)\rightarrow\left(X_{1n}W_n Y_{2n}\right)\rightarrow Y_{1n}. \tag{D.22}$$

Next, we note that X_1^N and Y_2^N are independent. Hence, (W_n,Y_{2n}) is independent of X_1^N. Coupled with the contraction property [57], we obtain the following Markov chain:

$$\left(X_1^{n-1}X_{1n+1}^N\right)\rightarrow X_{1n}\rightarrow\left(W_n Y_{2n} Y_{1n}\right). \tag{D.23}$$

Finally, using the weak union property and the decomposition property [57], we obtain $\left(X_1^{n-1}X_{1n+1}^N\right)\rightarrow\left(W_n X_{1n}\right)\rightarrow Y_{1n}$ as desired. Next, we bound R_{22} as follows:

$$\begin{aligned}
NR_{22} &= I\left(M_{22};Y_2^N|M_{21}\right) + H\left(M_{22}|Y_2^N M_{21}\right) \\
&= I\left(M_{21}M_{22};Y_2^N|M_{21}\right) + H\left(M_{22}|Y_2^N M_{21}\right) \\
&\le I\left(X_2^N;Y_2^N|M_{21}\right) + N\epsilon_{2N} \\
&= \sum_{n=1}^{n=N} I\left(X_2^N;Y_{2n}|M_{21}Y_2^{n-1}\right) + N\epsilon_{2N} \\
&= \sum_{n=1}^{n=N} H\left(Y_{2n}|M_{21}Y_2^{n-1}\right) - H\left(Y_{2n}|M_{21}Y_2^{n-1}X_2^N\right) + N\epsilon_{2N} \\
&\overset{(a)}{=} \sum_{n=1}^{n=N} H\left(Y_{2n}|M_{21}Y_2^{n-1}\right) - H\left(Y_{2n}|M_{21}Y_2^{n-1}X_{2n}\right) + N\epsilon_{2N} \\
&= \sum_{n=1}^{n=N} H\left(Y_{2n}|W_n\right) - H\left(Y_{2n}|W_n X_{2n}\right) + N\epsilon_{2N} \\
&= \sum_{n=1}^{n=N} I\left(X_{2n};Y_{2n}|W_n\right) + N\epsilon_{2N}
\end{aligned} \tag{D.24}$$

where (a) follows immediately from the Markov chain given by $\left(X_2^{n-1}X_{2n+1}^N\right) \to (W_n X_{2n}) \to Y_{2n}$. We first note the following Markov chain:

$$\left(X_2^{n-1}X_{2n+1}^N W_n\right) \to (X_{1n}X_{2n}) \to Y_{1n}Y_{2n}. \tag{D.25}$$

Using the weak union property, we obtain

$$\left(X_2^{n-1}X_{2n+1}^N\right) \to (W_n X_{1n}X_{2n}) \to Y_{1n}Y_{2n}. \tag{D.26}$$

Using the fact that $W_n X_2^N$ and X_1^N are independent, and applying the contraction property, we obtain

$$\left(X_2^{n-1}X_{2n+1}^N\right) \to (W_n X_{2n}) \to (X_{1n}Y_{1n}Y_{2n}). \tag{D.27}$$

Applying the decomposition property, we obtain the desired Markov chain

$$\left(X_2^{n-1}X_{2n+1}^N\right) \rightarrow (W_n X_{2n}) \rightarrow Y_{2n}.$$

Finally, we bound $R_{21} + R_1$ as follows:

$$\begin{aligned}
N\left(R_{21}+R_1\right) &= I\left(M_1 M_{21}; Y_1^N\right) + H\left(M_{21}|Y_1^N\right) + H\left(M_1|Y_1^N M_{21}\right) \\
&\leq I\left(M_{21}X_1^N; Y_1^N\right) + N\epsilon_{1N} + N\epsilon_{3N} \\
&= \sum_{n=1}^{n=N} I\left(M_{21}X_1^N; Y_{1n}|Y_1^{n-1}\right) + N\epsilon_{1N} + N\epsilon_{3N} \\
&= \sum_{n=1}^{n=N} H\left(Y_{1n}|Y_1^{n-1}\right) - H\left(Y_{1n}|X_1^N Y_1^{n-1} M_{21}\right) + N\epsilon_{1N} + N\epsilon_{3N} \\
&\leq \sum_{n=1}^{n=N} H\left(Y_{1n}\right) - H\left(Y_{1n}|X_1^N Y_1^{n-1} M_{21} Y_2^{n-1}\right) + N\epsilon_{1N} + N\epsilon_{3N} \\
&= \sum_{n=1}^{n=N} H\left(Y_{1n}\right) - H\left(Y_{1n}|X_1^N M_{21} Y_2^{n-1}\right) + N\epsilon_{1N} + N\epsilon_{3N} \\
&= \sum_{n=1}^{n=N} H\left(Y_{1n}\right) - H\left(Y_{1n}|X_{1n} M_{21} Y_2^{n-1}\right) + N\epsilon_{1N} + N\epsilon_{3N} \\
&= \sum_{n=1}^{n=N} H\left(Y_{1n}\right) - H\left(Y_{1n}|X_{1n} W_n\right) + N\epsilon_{1N} + N\epsilon_{3N} \\
&= \sum_{n=1}^{n=N} I\left(W_n X_{1n}; Y_{1n}\right) + N\epsilon_{1N} + N\epsilon_{3N}.
\end{aligned} \tag{D.28}$$

By the Markovity of $W_n \rightarrow (X_{1n}X_{2n}) \rightarrow (Y_{1n}Y_{2n})$ and the independence of (W_n, X_{2n}) and X_{1n}, we observe that

$$p\left(w_n, x_{1n}, x_{2n}, y_{1n}, y_{2n}\right) = p\left(w_n, x_{2n}\right) p\left(x_{1n}\right) p\left(y_{1n}, y_{2n}|x_{1n}, x_{2n}\right).$$

By introducing a time-sharing random variable Q similar to the proof for the converse of the capacity region of the multiple access channel [25, Pg. 402], we obtain Thm. 5.5. The assertions about the cardinalities of $\mathcal{W}$ and $\mathcal{Q}$ follow directly from the application of Caratheodory's theorem to the expressions (5.76)-(5.78).

D.3 Proof of Thm. 5.8

By Fano's inequality, we again have

$$H\left(M_{21}|Y_1^N\right) \leq N\epsilon_{1N} \tag{D.29}$$

$$H\left(M_{22}|Y_2^N\right) \leq N\epsilon_{2N} \tag{D.30}$$

$$H\left(M_1|Y_1^N\right) \leq N\epsilon_{3N} \tag{D.31}$$

where $\epsilon_{1N}, \epsilon_{2N}, \epsilon_{3N} \to 0$ as $N \to \infty$. We first bound R_{21} as follows:

$$\begin{aligned}
NR_{21} &= I\left(M_{21};Y_1^N\right) + H\left(M_{21}|Y_1^N\right) \\
&\leq I\left(M_{21};Y_1^N\right) + N\epsilon_{1N} \\
&\overset{(a)}{=} H\left(M_{21}|M_{22}X_1^N\left(M_1\right)\right) - H\left(M_{21}|Y_1^N\right) + N\epsilon_{1N} \\
&\leq H\left(M_{21}|M_{22}X_1^N\right) - H\left(M_{21}|M_{22}X_1^N Y_1^N\right) + N\epsilon_{1N} \\
&= I\left(M_{21};Y_1^N|M_{22}X_1^N\right) + N\epsilon_{1N} \\
&\leq H\left(Y_1^N|M_{22}X_1^N\right) - H\left(Y_1^N|M_{21}M_{22}X_1^N X_2^N\right) + N\epsilon_{1N} \\
&= H\left(Y_1^N|M_{22}X_1^N\right) - H\left(Y_1^N|M_{22}X_1^N X_2^N\right) + N\epsilon_{1N} \\
&= \sum_{n=1}^{n=N} H\left(Y_{1n}|X_1^N M_{22} Y_1^{n-1}\right) - H\left(Y_{1n}|X_{1n}X_{2n}\right) + N\epsilon_{1N} \\
&\overset{(b)}{=} \sum_{n=1}^{n=N} H\left(Y_{1n}|X_1^N M_{22} Y_1^{n-1} Y_2^{n-1}\right) - H\left(Y_{1n}|X_{1n}X_{2n}\right) + N\epsilon_{1N} \\
&\leq \sum_{n=1}^{n=N} H\left(Y_{1n}|X_{1n} M_{22} Y_2^{n-1}\right) - H\left(Y_{1n}|X_{1n}X_{2n}\right) + N\epsilon_{1N} \\
&= \sum_{n=1}^{n=N} H\left(Y_{1n}|X_{1n}W_n\right) - H\left(Y_{1n}|X_{1n}X_{2n}W_n\right) + N\epsilon_{1N} \\
&\leq \sum_{n=1}^{n=N} I\left(X_{2n};Y_{1n}|W_n X_{1n}\right) + N\epsilon_{1N}
\end{aligned} \tag{D.32}$$

where we define the random variable $W_n = \left(M_{22}, Y_2^{n-1}\right)$ for all n, (a) follows from the fact that since M_{21}, M_{22}, and M_1 are independent, so are M_{21}, M_{22}, and $X_1^N(M_1)$, and (b) follows from the fact that $Y_2^{n-1} \to \left(X_1^{n-1}Y_1^{n-1}\right) \to \left(M_{22}X_{1n}^N Y_{1n}\right)$ form a Markov chain. This follows from the discrete memoryless property of the channel and the fact that for any i, Y_{2i} depends only on X_{1i} and Y_{1i} (refer to (5.16)). Next, we bound R_{22} as follows:

$$\begin{aligned}
NR_{22} &= I\left(M_{22}; Y_2^N\right) + H\left(M_{22}|Y_2^N\right) \\
&\leq I\left(M_{22}; Y_2^N\right) + N\epsilon_{2N} \\
&= \sum_{n=1}^{N} I\left(M_{22}; Y_{2n}|Y_2^{n-1}\right) + N\epsilon_{2N} \\
&\leq \sum_{n=1}^{N} I\left(M_{22}Y_2^{n-1}; Y_{2n}\right) + N\epsilon_{2N} \\
&= \sum_{n=1}^{N} I\left(W_n; Y_{2n}\right) + N\epsilon_{2N}.
\end{aligned} \tag{D.33}$$

Next, we bound R_1 as follows:

$$\begin{aligned}
NR_1 &= I\left(M_1; Y_1^N\right) + H\left(M_1|Y_1^N\right) \\
&\leq I\left(X_1^N; Y_1^N\right) + N\epsilon_{3N} \\
&= H\left(X_1^N\right) - H\left(X_1^N|Y_1^N\right) + N\epsilon_{3N} \\
&= H\left(X_1^N|X_2^N\right) - H\left(X_1^N|Y_1^N\right) + N\epsilon_{3N} \\
&\leq H\left(X_1^N|X_2^N\right) - H\left(X_1^N|X_2^N Y_1^N\right) + N\epsilon_{3N} \\
&= I\left(X_1^N; Y_1^N|X_2^N\right) + N\epsilon_{3N} \\
&= \sum_{n=1}^{n=N} H\left(Y_{1n}|X_2^N Y_1^{n-1}\right) - H\left(Y_{1n}|X_2^N Y_1^{n-1} X_1^N\right) + N\epsilon_{3N} \\
&\leq \sum_{n=1}^{n=N} H\left(Y_{1n}|X_{2n}\right) - H\left(Y_{1n}|X_{1n}X_{2n}\right) + N\epsilon_{3N} \\
&= \sum_{n=1}^{n=N} I\left(X_{1n}; Y_{1n}|X_{2n}\right) + N\epsilon_{3N}.
\end{aligned} \tag{D.34}$$

For the degraded discrete memoryless ZC of type II, we have

$$I\left(M_{22}; Y_1^N | X_1^N\right) \geq I\left(M_{22}; Y_2^N\right) \tag{D.35}$$

from the data processing inequality and the fact that $M_{22} \to X_2^N \to \left(X_1^N Y_1^N\right) \to Y_2^N$ form a Markov chain. The above inequality similarly holds for the discrete memoryless ZC of type III. To bound $R_1 + R_{21} + R_{22}$, we have

$$\begin{aligned}
&N\left(R_1 + R_{21} + R_{22}\right) \\
&\quad = I\left(M_{22}; Y_2^N\right) + H\left(M_{22}|Y_2^N\right) + I\left(M_1; Y_1^N\right) \\
&\qquad + H\left(M_1|Y_1^N\right) + I\left(M_{21}; Y_1^N\right) + H\left(M_{21}|Y_1^N\right) \\
&\quad \leq I\left(M_{22}; Y_1^N|X_1^N\right) + I\left(X_1^N; Y_1^N\right) \\
&\qquad + I\left(M_{21}; Y_1^N|M_{22}X_1^N\right) + N\epsilon_{1N} + N\epsilon_{2N} + N\epsilon_{3N} \\
&\quad \leq I\left(M_{21}M_{22}X_1^N; Y_1^N\right) + N\epsilon_{1N} + N\epsilon_{2N} + N\epsilon_{3N} \\
&\quad \leq I\left(X_1^N X_2^N; Y_1^N\right) + N\epsilon_{1N} + N\epsilon_{2N} + N\epsilon_{3N} \\
&\quad \leq \sum_{n=1}^{n=N} H\left(Y_{1n}|Y_1^{n-1}\right) - H\left(Y_{1n}|X_{1n}X_{2n}\right) + N\epsilon_{1N} + N\epsilon_{2N} + N\epsilon_{3N} \\
&\quad \leq \sum_{n=1}^{n=N} I\left(X_{1n}X_{2n}; Y_{1n}\right) + N\epsilon_{1N} + N\epsilon_{2N} + N\epsilon_{3N}.
\end{aligned} \tag{D.36}$$

By the Markovity of $W_n \to (X_{1n}X_{2n}) \to (Y_{1n}Y_{2n})$ and the independence of (W_n, X_{2n}) and X_{1n}, we observe again that

$$p\left(w_n, x_{1n}, x_{2n}, y_{1n}, y_{2n}\right) = p\left(w_n, x_{2n}\right) p\left(x_{1n}\right) p\left(y_{1n}, y_{2n}|x_{1n}, x_{2n}\right).$$

Finally, we obtain Thm. 5.8, by introducing a time-sharing random variable Q. The assertions about the cardinalities of $\mathcal{W}$ and $\mathcal{Q}$ follow directly from the application of Caratheodory's theorem to the expressions (5.95)-(5.98).

D.4 Proof of Thm. 5.12

We determine an outer bound to the capacity region of the equivalent Gaussian ZC with strong crossover link gain as shown in Fig. 5.7. By Fano's inequality, we have

$$H\left(M_{21}|Y_1'^N\right) \leq N\epsilon_{1N} \tag{D.37}$$

$$H\left(M_{22}|Y_2'^N\right) \leq N\epsilon_{2N} \tag{D.38}$$

$$H\left(M_1|Y_1'^N\right) \leq N\epsilon_{3N} \tag{D.39}$$

where $\epsilon_{1N}, \epsilon_{2N}, \epsilon_{3N} \to 0$ as $N \to \infty$. We first bound the term $H\left(M_{22}|Y_1'^N M_1\right) = H\left(M_{22}|X_2^N + Z_{21}^N\right)$. From the following Markov chain:

$$(M_{21}, M_{22}) \to X_2^N \to X_2^N + Z_{21}^N \to X_2^N + Z_{21}^N + Z_{22}^N \tag{D.40}$$

we have by the data processing inequality and Fano's inequality

$$\begin{aligned}
I\left(M_{22}; X_2^N + Z_{21}^N\right) &\geq I\left(M_{22}; X_2^N + Z_{21}^N + Z_{22}^N\right) \\
\Longrightarrow H\left(M_{22}|X_2^N + Z_{21}^N\right) &\leq H\left(M_{22}|X_2^N + Z_{21}^n + Z_{22}^N\right) \\
&= H\left(M_{22}|Y_2'^N\right) \\
&\leq N\epsilon_{2N}.
\end{aligned} \tag{D.41}$$

Next, we bound the following term $h\left(Y_1'^N|M_1 M_{22}\right)$. Consider the following inequalities,

$$\begin{aligned}
\frac{N}{2}\log_2\left(\frac{2\pi e}{a^2}\right) &= h\left(Z_{21}^N|M_1 M_{21} M_{22}\right) \\
&= h\left(\frac{1}{a}X_1^N\left(M_1\right) + X_2^N\left(M_{21}, M_{22}\right) + Z_{21}^N|M_1 M_{21} M_{22}\right) \\
&= h\left(Y_1'^N|M_1 M_{21} M_{22}\right) \\
&\leq h\left(Y_1'^N|M_1 M_{22}\right)
\end{aligned}$$

$$\le h\left(Y_1'^N|M_1\right)$$
$$\le \frac{N}{2}\log_2\left((2\pi e)\left(\frac{a^2P_2+1}{a^2}\right)\right). \quad \text{(D.42)}$$

Thus, there exists a $\beta \in [0,1]$, such that

$$h\left(Y_1'^N|M_1M_{22}\right) = h\left(X_2^N\left(M_{21},M_{22}\right)+Z_{21}^N|M_{22}\right)$$
$$= \frac{N}{2}\log_2\left((2\pi e)\left(\frac{a^2\beta P_2+1}{a^2}\right)\right). \quad \text{(D.43)}$$

We next obtain a lower bound for $h\left(X_2^n+Z_{21}^n+Z_{22}^n|M_{22}\right)$ by making use of the entropy power inequality

$$2^{\frac{2}{N}h\left(X_2^N+Z_{21}^N+Z_{22}^N|M_{22}\right)} \ge 2^{\frac{2}{N}h\left(X_2^N+Z_{21}^N|M_{22}\right)}+2^{\frac{2}{N}h\left(Z_{22}^N\right)}$$
$$= (2\pi e)\left(\beta P_2+1\right)$$
$$\Rightarrow h\left(X_2^N+Z_{21}^N+Z_{22}^N|M_{22}\right) \ge \frac{N}{2}\log_2\left((2\pi e)\left(\beta P_2+1\right)\right). \quad \text{(D.44)}$$

We can now bound R_{21} as follows:

$$NR_{21} = H\left(M_{21}\right)$$
$$= I\left(M_{21};Y_1'^N|M_1M_{22}\right)+H\left(M_{21}|Y_1'^NM_1M_{22}\right)$$
$$\le I\left(M_{21};Y_1'^N|M_1M_{22}\right)+H\left(M_{21}|Y_1'^N\right)$$
$$\le I\left(M_{21};Y_1'^N|M_1M_{22}\right)+N\epsilon_{1N}$$
$$= h\left(Y_1'^N|M_1M_{22}\right)-h\left(Y_1'^N|M_1M_{21}M_{22}\right)+N\epsilon_{1N}$$
$$= h\left(Y_1'^N|M_1M_{22}\right)-h\left(Z_{21}^N\right)+N\epsilon_{1N}$$
$$= \frac{N}{2}\log_2\left((2\pi e)\left(\frac{a^2\beta P_2+1}{a^2}\right)\right)-\frac{N}{2}\log_2\left(\frac{2\pi e}{a^2}\right)+N\epsilon_{1N}$$
$$= \frac{N}{2}\log_2\left(a^2\beta P_2+1\right)+N\epsilon_{1N}. \quad \text{(D.45)}$$

We bound R_{22} as follows:

$$
\begin{aligned}
NR_{22} &= H\left(M_{22}\right) \\
&= I\left(M_{22}; Y_2'^N\right) + H\left(M_{22}|Y_2'^N\right) \\
&\leq I\left(M_{22}; Y_2'^N\right) + N\epsilon_{2N} \\
&= h\left(Y_2'^N\right) - h\left(Y_2'^N|M_{22}\right) + N\epsilon_{2N} \\
&= h\left(X_2^N + Z_{21}^N + Z_{22}^N\right) - h\left(X_2^N + Z_{21}^N + Z_{22}^N|M_{22}\right) + N\epsilon_{2N} \\
&\leq \frac{N}{2}\log_2\left(1 + \frac{(1-\beta)\,P_2}{1+\beta P_2}\right) + N\epsilon_{2N}.
\end{aligned} \tag{D.46}
$$

We then bound R_1 as follows:

$$
\begin{aligned}
NR_1 &= H\left(M_1\right) \\
&= I\left(M_1; Y_1'^N|M_{21}M_{22}\right) + H\left(M_1|Y_1'^N M_{21}M_{22}\right) \\
&\leq I\left(M_1; Y_1'^N|M_{21}M_{22}\right) + H\left(M_1|Y_1'^N\right) \\
&\leq I\left(M_1; Y_1'^N|M_{21}M_{22}\right) + N\epsilon_{3N} \\
&\leq h\left(Y_1'^N|M_{21}M_{22}\right) - h\left(Y_1'^N|M_1M_{21}M_{22}\right) + N\epsilon_{3N} \\
&= h\left(\frac{X_1^N}{a} + Z_{21}^N\right) - h\left(Z_{21}^N\right) + N\epsilon_{3N} \\
&\leq \frac{N}{2}\log_2\left(1 + P_1\right) + N\epsilon_{3N}.
\end{aligned} \tag{D.47}
$$

Finally, we bound the term $R_1 + R_{21} + R_{22}$ as follows:

$$
\begin{aligned}
N\left(R_1 + R_{21} + R_{22}\right) &= H\left(M_1M_{21}M_{22}\right) \\
&= I\left(M_1M_{21}M_{22}; Y_1'^N\right) + H\left(M_1|Y_1'^N\right) \\
&\quad + H\left(M_{21}|Y_1'^N M_1\right) + H\left(M_{22}|Y_1'^N M_1M_{21}\right) \\
&\leq I\left(M_1M_{21}M_{22}; Y_1'^N\right) + H\left(M_1|Y_1'^N\right) \\
&\quad + H\left(M_{21}|Y_1'^N\right) + H\left(M_{22}|Y_1'^N M_1\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leq I\left(M_1 M_{21} M_{22}; Y_1^{\prime N}\right) + N\epsilon_{1N} + N\epsilon_{2N} + N\epsilon_{3N} \\
&= h\left(\frac{X_1^N}{a} + X_2^N + Z_{21}^N\right) - h\left(Z_{21}^N\right) \\
&\quad + N\epsilon_{1N} + N\epsilon_{2N} + N\epsilon_{3N} \\
&\leq \frac{N}{2}\log_2\left(1 + a^2 P_2 + P_1\right) + N\epsilon_{1N} + N\epsilon_{2N} + N\epsilon_{3N}. \quad \text{(D.48)}
\end{aligned}
$$

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 1948.

[2] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, vol. 1. Berkeley, CA: Univ. California Press, 1961, pp. 611–644.

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

[4] H. Liao, "Multiple access channels," Ph.D. dissertation.

[5] T. M. Cover, "An achievable rate region for the broadcasting channel," *IEEE Trans. Inf. Theory*, vol. 21, pp. 399–404, July 1975.

[6] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredaci Informaccii*, vol. 10(3), pp. 3–14, 1974.

[7] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 20, pp. 279–280, Mar. 1974.

[8] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.

[9] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[10] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, p. 26582668, Oct. 2003.

[11] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, p. 18751892, Sept. 2004.

[12] C. K. Toh, *Ad Hoc Mobile Wireless Networks.* Prentice Hall, 2002.

[13] S. Vishwanath, N. Jindal, and A. Goldsmith, "The "Z" channel," in *IEEE Global Telecommunications Conference*, vol. 3, San Francisco, CA, Dec 2003, pp. 1726–1730.

[14] T. M. Cover and A. A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sept 1979.

[15] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[16] H. F. Chong, Mehul Motani and Hari Krishna Garg, "A comparison of two achievable rate regions for the interference channel," *UCSD-ITA*, Feb. 2006.

[17] A. A. E. Gamal and M. H. M. Costa, "The capacity region of a class of deterministic interference channels," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 343–346, March 1982.

[18] J. Jiang, Y. Xin, and H. K. Garg, "Interference channels with common information," *IEEE Trans. Inf. Theory*, submitted for publication. [Online]. Available: http://arxiv.org/PS_cache/cs/pdf/0610/

[19] H. F. Chong, M. Motani, and H. K. Garg, "Generalized backward decoding strategies for the relay channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 394 – 401, Jan. 2007.

[20] H. F. Chong, Mehul Motani and Hari Krishna Garg, "On achievable rates for relay channels," in *UCSD Information Theory and Applications Workshop, 2007*, San Diego, CA.

[21] H. F. Chong, M. Motani, and H. K. Garg, "Relay channel with general alphabets," *IEEE Trans. Inf. Theory*, p. Submitted.

[22] H. F. Chong, M. Motani, H. K. Garg, and H. E. Gamal, "On a Simplification of the Han-Kobayashi Rate Region for the Interference Channel," *IEEE Trans. Inf. Theory*, submitted for publication.

[23] ——, "The Capacity Region of a Class of Interference Channels," in *Information Theory, 2007. Proceedings. IEEE International Symposium on.*

[24] H. F. Chong, M. Motani, and H. K. Garg, "Capacity Theorems for the "Z"-channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1348–1365, 2007.

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley Interscience, 1991.

[26] R. M. Gray, *Entropy and Information Theory.* Springer Verlag, 2000.

[27] D. Kroning, in *Lecture notes presented at 251-0247-00L Formal Verification class, ETH, Zurich*, 2004/2005. [Online]. Available: http://www.inf.ethz.ch/personal/kroening/classes/251-0247-00/ws2004-2005/fourier-motzkin.pdf

[28] E. C. van der Meulen, "Transmission of information in a T-terminal discrete memoryless channel," Ph.D. dissertation, Univ. of California, Berkeley, CA, 1968.

[29] ——, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[30] H. Sato, "Information transmission through a channel with relay," The Aloha System, University of Hawaii, Honolulu, Tech. Rep. B76-7, Mar. 1976.

[31] R. C. King, "Multiple access channels with generalized feedback," Ph.D. dissertation, Stanford University, Mar. 1978.

[32] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 841–850, Nov. 1982.

[33] F. M. J. Willems, "Information theoretical results for the discrete memoryless multiple access channel," Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, Oct. 1992.

[34] C. M. Zeng, F. Kuhlmann, and A. Buzo, "Achievability proof of some multiuser channel coding theorems using backward decoding," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1160–1165, Nov. 1989.

[35] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.

[36] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–11, Jan. 1976.

[37] T. M. Cover, and Y. H. Kim, "Capacity of a class of deterministic relay channel," in *Information Theory, 2004. Proceedings. IEEE International Symposium on*, Nice, Frace, June 2007, pp. 591–595.

[38] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1912–1922, Nov. 1997.

[39] J. A. Thomas, "Feedback can at most double gaussian multiple access channel capacity," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 711–716, 1987.

[40] A. E. Gamal, M. Mohseni and S. Zahedi, "Bounds on capacity and minimum enery-per-bit for awgn relay channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.

[41] T. Berger, "Multiterminal source coding," in *Lecture notes presented at the 1977 CISM Summer School, Udine, Italy*, July 18-20 1977, pp. 171–231.

[42] A. D. Wyner, "The rate-distortion function for source-coding with side information at the decoder-ii: General sources," *Inform. Contr.*, vol. 38, pp. 60–80, Jul. 1978.

[43] R. M. Gray, *Probability, Random Processes, and Ergodic Properties.* Springer Verlag, 2006.

[44] R. L. Dobrushin, "A general formulation of the fundamental shannon theorem in information theory," *Translation in Transactions Amer. Math. Soc.*, vol. 33, pp. 323–438, 1959.

[45] A. R. Barron, "The strong ergodic theorem for densities: Generalized shannon-mcmillan-breiman theorem," *Ann. of Prob.*, vol. 13, no. 4, pp. 1292–1303, 1985.

[46] A. B. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 60–70, Jan. 1978.

[47] ——, "A case where interference does not reduce capacity," *IEEE Trans. Inf. Theory*, vol. 21, no. 5, pp. 569–570, Sept. 1975.

[48] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Inf. Theory*, vol. 27, no. 6, pp. 786–788, Nov. 1981.

[49] M. H. M. Costa and A. A. E. Gamal, "The capacity region of the discrete memoryless interference channel with strong interference," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 710–711, Sept. 1987.

[50] N. Liu and S. Ulukus, "The Capacity Region of a Class of Discrete Degraded Interference Channels," in *Proc. of the 39th Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Sept. 2006.

[51] R. Benzel, "The capacity region of a class of discrete additive degraded interference channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 228–231, March 1979.

[52] K. Kobayashi and T. S. Han, "A further consideration of the hk and cmg regions for the interference channel," *UCSD-ITA*, Jan. 2007.

[53] G. Kramer, "Review of rate regions for interference channels," *International Zurich Seminar*, Feb. 2006.

[54] I. Maric, R. D. Yates, and G. Kramer, "The Capacity Region of the strong interference channel with common information," in *Proc. of the 39th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2005.

[55] N. Liu and S. Ulukus, "On the capacity region of the Gaussian Z-channel," in *IEEE Global Telecommunications Conference*, vol. 1, Dallas, TX, Nov 2004, pp. 415–419.

[56] F. M. J. Willems and E. C. V. D. Meulen, "The Discrete Memoryless Multiple-Access Channel with Cribbing Encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[57] J. Pearl, *Causality: Models, Reasoning, and Inference*, 1st ed. Cambridge University Press, 2000, ch. 1.

[58] M. H. M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 5, pp. 607–615, 1985.

[59] G. Kramer, "Outer bounds on the capacity of Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 581–586, Mar. 2004.

[60] I. Csiszar, "Arbitrarily varying channels with general alphabets and states," *IEEE Trans. Inf. Theory*, vol. 38, no. 6, pp. 1725–1742, Nov 1992.

[61] T. T. Kadota, "Generalizations of Feinstein's fundamental lemma," *IEEE Trans. Inf. Theory*, vol. 16, pp. 791–792, Nov 1970.

[62] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan 2003.