ABSTRACT

Title of dissertation: CORRELATION AND COOPERATION IN NETWORK INFORMATION THEORY

Wei Kang
Doctor of Philosophy, 2008

Dissertation directed by: Professor Şennur Ulukuş
Department of Electrical and Computer Engineering

Recent developments in wireless ad-hoc and sensor networking motivates the investigation of sophisticated phenomena that arise in such networks from an information theoretic point of view. In this dissertation, we focus on two of these phenomena: correlation and cooperation. In wireless networks, correlation mainly originates from the correlated observations of different users, while cooperation is enabled by the wireless medium, which lets third-party users obtain part of the information from the transmitter in order to help deliver it to the destination.

We first study the effects of source correlation in multi-user networks. More specifically, we study the distributed source and channel coding problem for correlated sources, e.g., multiple access channel with correlated sources and multi-terminal rate-distortion problem. In these problems, it is often needed to characterize the joint probability distribution of a pair of random variables satisfying an n-letter Markov chain. An exact characterization of such probability distributions is intractable. We propose a new data processing inequality, which provides us a single-letter necessary

condition for the n-letter Markov chain. Our new data processing inequality yields outer bounds for the multiple access channel with correlated sources and the multi-terminal rate-distortion region.

Next, we investigate the role of correlation in cooperative multi-user networks. We consider the basic three-node relay channel, which is the simplest model for cooperative communications. We propose a new coding scheme for the relay channel, which is in the form of block Markov coding and is based on preserving the correlation in the channel inputs from the transmitter and the relay. The analysis of the error events provides us with three conditions containing mutual information expressions involving infinite letters of the underlying random process. We lower bound these mutual informations to obtain three single-letter conditions. We show that the achievable rates with the classical compress-and-forward (CAF) scheme is a special case of the achievable rates in our new coding scheme. We therefore conclude that our proposed coding scheme yields potentially larger rates than the CAF scheme.

Finally, we focus on the diamond channel, which is a four-node cooperative communication network. We study a special class of diamond channels, which consists of a transmitter, a noisy relay and a noiseless relay, and a destination. We determine the capacity of this class of diamond channels by providing an achievable scheme and a converse. The capacity we show is strictly smaller than the cut-set bound. Our result also shows the optimality of a combination of decode-and-forward (DAF) and CAF at the noisy relay node. This is the first example where a combination of DAF and CAF is shown to be capacity achieving. We also uncover a duality between this diamond channel coding problem and the Kaspi-Berger source coding problem.

Correlation and Cooperation in Network Information Theory

by

Wei Kang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:

Professor Şennur Ulukuş, Chair/Advisor
Professor Prakash Narayan
Professor Alexander Barg
Professor Nuno Martins
Professor Lawrence Washington

DEDICATION

To my parents Zhuang Kang and Jiaying An and my wife Nan Liu

ACKNOWLEDGMENTS

I first thank my advisor Professor Şennur Ulukuş. She brought me into the area of information theory and provided me with guidance as well as freedom through my Ph.D. studies. I thank her for always encouraging me to work on theoretically challenging problems. I also thank her for always being patient and cheering me up no matter how stuck I got or even failed after months or years of effort. I highly appreciate the environment she fosters in our group, which I believe is very healthy and helpful for theoretical and creative research.

I also like to thank Professors Prakash Narayan, Alexander Barg, Nuno Martins, and Lawrence Washington for serving in my Ph.D. dissertation committee and providing many valuable comments and suggestions. I especially thank Professors Prakash Narayan and Alexander Barg for organizing the information and coding theory seminar, from which I not only broadened my range of knowledge, but also learned to appreciate theoretically elegant research.

I would like to thank my fellow students, Onur, Alkan, Shabnam, Ravi, Bahadir, Jing, Raef, Ersen, Sirin, Prasanth, Brooke, etc. The discussions, chats and laughters with them made my life in Maryland much more fun.

Finally, I would like to thank my family. I thank my parents for always being loving, caring and encouraging. Even in the most difficult times, they never lost their confidence and belief in me. I can always rely on their advices and support. I thank my uncle and aunt for taking care of me all these years and sharing their knowledge of life with me. I thank my wife Nan the most. The love, trust and appreciation she provides make my life happy and meaningful. In addition to being my love and soul-mate, she is also my most helpful colleague. She makes me a better person both in life and in research. Without her, I could not reach this far.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

The field of information theory entered the multi-user era in 1961 when Shannon studied the two-way channel [34]. The area of multi-user information theory blossomed in 1970s with celebrated results, such as the capacity region of the multiple access channel by Ahlswede [1] and Liao [28] in 1971, the lossless distributed source coding region by Slepian and Wolf [35] in 1973, the capacity region of the degraded broadcast channel by Bergmans [5] and Gallager [19] in 1974, etc. However, most problems in multi-user information theory remain open today, including the capacity regions of the two way channel, general broadcast channel, interference channel, relay channel, and the lossy distributed source coding region, etc.

As we entered the new century, the theoretical research in multi-user information theory has had a stronger connection with, and therefore, has been stimulated by, the thriving practical developments in wireless communication networks. For example, in cellular networks and wireless LANs, which are mature commercial wireless communication technologies, the uplinks and downlinks can be modeled as multiple access and broadcast channels, which have been studied extensively in multi-user information theory. Recently, the research emphasis has shifted to more complicated wireless networks, e.g., ad hoc and sensor networks. The complication in the structure of such wireless networks introduces more sophisticated phenomena to be considered in a theoretical context, e.g., correlation, cooperation, etc. Correlation in these scenarios mainly comes from the correlated observations of different users. For example, in sensor networks, neighboring sensors obtain correlated observations. Cooperation, on the other hand, is enabled by the wireless medium, which lets third-party users obtain part of the information from the transmitter in order to help deliver it to the destination.

We begin our work by studying the effects of source correlation in multi-user information theory. More specifically, we study the distributed source and channel coding problem for correlated sources, e.g., multiple access channel with correlated sources and multi-terminal rate-distortion region. There has been a significant amount of effort directed towards solving the multi-terminal rate distortion problem since the milestone paper of Wyner and Ziv [42] on the rate-distortion function of a single source with side information at the decoder, which is a special case of the multi-terminal rate-distortion problem. Among all the attempts made on this difficult problem, the notable works by Tung [37] and Housewright [22] (see also [4]) provide inner and outer bounds for the rate-distortion region. A more recent progress on this problem has been made by Wagner and Anantharam in [39], where a tighter outer bound was given. The multiple access channel with correlated sources was first studied by Cover, El Gamal and Salehi in [9] (a simpler proof was given in [2]), where an achievable region expressed by single-letter entropies and mutual informations was given. This achievable region was shown to be suboptimal by Dueck [16]. Cover, El Gamal and

Salehi [9] also provided a capacity result with both achievability and converse in the form of some incomputable n-letter mutual informations.

In distributed source and channel coding for correlated sources, it is often needed to characterize the joint probability distribution of a pair of random variables satisfying an n-letter Markov chain. An exact characterization of such probability distributions is intractable. In Chapter 2, we propose a new data processing inequality, which provides us a single-letter necessary condition for the n-letter Markov chain. Our new data processing inequality yields outer bounds for the multiple access channel with correlated sources and the multi-terminal rate-distortion region.

Next, we study the role of correlation in cooperative multi-user networks. We focus on the relay channel. As the simplest model for cooperative communications, relay channel has attracted plenty of attention since 1971, when it was introduced by van der Meulen [38]. In 1979, Cover and El Gamal proposed two major coding schemes for the relay channel [8]. These two schemes are widely known as Decode-And-Forward (DAF) and Compress-And-Forward (CAF) today; see [25] for a recent review. In our work, we focus on the CAF scheme. In CAF, correlation is created between the transmitter and the relay, through the channel between them, and this correlation is utilized to improve the achievable rates. The shortcoming of the CAF scheme is that the correlation offered by the block coding structure is not utilized effectively, since in each block, the channel inputs from the transmitter and the relay are independent, as the transmitter sends the message only once. We know that the essence of good coding schemes in multi-user systems with correlated sources (e.g., [2, 9]) is to preserve the correlation of the sources in the channel inputs.

Motivated by this basic observation, in Chapter 3, we propose a new coding scheme for the relay channel, which is in the form of block Markov coding and preserves the correlation in the channel inputs from the transmitter and the relay. At the decoding stage, we perform joint decoding for the entire B blocks. The analysis of the error events provides us three conditions containing mutual information expressions involving infinite letters of the underlying random process. To obtain a computable result, we lower bound these mutual informations by noting some Markov structure in the underlying random process. This operation gives us three conditions to be satisfied by the achievable rates which involve eleven variables. We finish our analysis by revisiting the CAF scheme. First, we develop an equivalent representation for the achievable rates given in [8] for the CAF scheme. We then show that this equivalent representation for the achievable rates for the CAF scheme is a special case of the achievable rates in our new coding scheme, which is obtained by a special selection of the eleven random variables mentioned above. We therefore conclude that our proposed coding scheme yields potentially larger rates than the CAF scheme. More importantly, our new coding scheme creates more possibilities, and therefore a spectrum of new achievable schemes for the relay channel through the selection of the underlying probability distribution.

We then focus on the diamond channel, a relatively simple cooperative network, which consists of four nodes: one transmitter, two relays and one receiver. The diamond channel was first proposed by Schein in his Ph.D. dissertation [32]. The diamond channel may be viewed as more complicated than the standard three-node relay channel as it contains one more node; however, it may also be viewed as simpler than the standard relay channel as it does not have a direct link from the source to the destination, simplifying the temporal aspects of the coding problem. In addition, the diamond channel may be viewed as the most simple generalization of the standard relay channel to multi-relay ad-hoc wireless communication channels. The capacity of the general diamond channel is an open problem. Schein [32] studied several special classes of diamond channels. Among them, we will focus on a special class of diamond channels, in which the channels from the two relays to the receiver are specified as two finite-rate, noiseless, orthogonal links, and one of the branches of the broadcast channel from the transmitter to the two relays is noiseless. Schein proposed two achievability schemes for this class of diamond channels without showing the optimality of them.

In Chapter 4, we prove the capacity of this class of diamond channels by providing an achievable scheme and a converse. The capacity we show is strictly smaller than the cut-set bound. Our result also shows the optimality of a combination of DAF and CAF at the noisy relay node. This is the first example where a combination of DAF and CAF is shown to be capacity achieving. Finally, we note that there exists a duality between this diamond channel coding problem and the Kaspi-Berger source coding problem.

The rest of this dissertation is organized as follows. In Chapter 2, we discuss our work on distributed coding of correlated sources, where we develop a new data processing inequality and apply it to distributed source and channel coding. In Chapter 3, we present our new coding scheme for the relay channel. In Chapter 4, we present our work on the capacity of a class of diamond channels. We present our conclusions in Chapter 5.

Chapter 2

A New Data Processing Inequality and Its Applications in Distributed Source and Channel Coding

2.1 Problem Formulation

In this chapter, we consider a pair of correlated discrete source sequences with length n, $(U^n, V^n) = \{(U_1, V_1), \ldots, (U_n, V_n)\}$, which are independent and identically distributed (i.i.d.) in time, i.e.,

$$p(u^n, v^n) = \prod_{i=1}^{n} p(u_i, v_i) \tag{2.1}$$

and

$$p(u_i, v_i) = p(u, v), \qquad i = 1, \ldots, n \tag{2.2}$$

where the single-letter joint distribution $p(u, v)$ is defined on the alphabet $\mathcal{U} \times \mathcal{V}$. Let (X_1, X_2) be two random variables such that (X_1, X_2, U^n, V^n) satisfies

$$p(x_1, x_2, u^n, v^n) = p(u^n, v^n)p(x_1|u^n)p(x_2|v^n) \tag{2.3}$$

or equivalently[1],

$$X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$$

This Markov chain appears in some problems involving the distributed coding of correlated sources. For example, in distributed rate-distortion problem [4, 22, 37], (X_1, X_2) is used to reconstruct, $(\hat{U}^n, \hat{V}^n)$, an estimate of the sources (U^n, V^n), and in the problem of multiple access channel with correlated sources [2, 9], (X_1, X_2) is sent though a multiple access channel in one channel use. Although these specific problems have been studied separately in their own contexts, the common nature of these problems, the distributed coding of correlated sources, enables us to conduct a general study, which will be applicable to these specific problems.

The study of the converse proofs of (or the necessary conditions for) the above specific problems raises the following question. We know that the correlation between (X_1, X_2) is limited, if a single-letter Markov chain $X_1 \longrightarrow U \longrightarrow V \longrightarrow X_2$ is to be satisfied. With the help of more letters of the sources, i.e., $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$ with n larger than 1, the correlation between (X_1, X_2) may increase. The question here is how correlated (X_1, X_2) can be, when n increases. More specifically, can they be arbitrarily correlated? To answer this question, we need to determine the set of all "valid" joint probability distributions $p(x_1, x_2)$, if $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$ is to be satisfied for some n, i.e., for given source pair (U, V), we need to determine the

[1] $X_1 = f_1(U^n)$ and $X_2 = f_2(V^n)$ is a degenerate case.

following set[2]

$$\mathcal{S}_{X_1X_2} \triangleq$$

$$\{p(x_1,x_2)\!:\!\exists n\!\in\!\mathbb{N}^{+}, p(x_1|u^n), p(x_2|v^n), \text{ s.t. } p(x_1,x_2) = \sum_{u^n,v^n} p(x_1|u^n)p(u^n,v^n)p(x_2|v^n)\} \tag{2.4}$$

with $p(u^n, v^n)$ satisfying (2.1) and (2.2).

We note that it is practically impossible to exhaust the elements in the set $\mathcal{S}_{X_1X_2}$ by searching over all conditional distribution pairs $(p(x_1|u^n), p(x_2|v^n))$ for all possible positive integer n. In other words, determining the set of all possible probability distributions $p(x_1, x_2)$ satisfying the n-letter Markov chain $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$, i.e., the set $\mathcal{S}_{X_1X_2}$, seems computationally intractable. To avoid this problem, we seek a single-letter necessary condition for the n-letter Markov chain $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$. The resulting set, say $\mathcal{S}'_{X_1X_2}$, characterized by this computable single-letter constraints, will contain the target set $\mathcal{S}_{X_1X_2}$.

The most intuitive necessary condition for a Markov chain is the data processing inequality [11, p. 32], i.e., if $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$, then

$$I(X_1; X_2) \leq I(U^n; V^n) = nI(U; V) \tag{2.5}$$

Since $I(U^n; V^n)$ increases linearly with n, the constraint in (2.5) will be loose when n is sufficiently large. Although the data processing inequality in its usual form does not

[2]We are also interested in determining the set of all "valid" probability distributions $p(x_1, x_2, u_1, v_1)$, if this Markov chain constraint is to be satisfied.

prove useful in this problem, we will still use the basic methodology of employing a data processing inequality to find a necessary condition for the n-letter Markov chain under consideration. For this, we will introduce a new measure of correlation, and develop a new data processing inequality based on this new measure of correlation.

Spectral method has been instrumental in the study of some properties of pairs of correlated random variables, especially, those of i.i.d. sequences of pairs of correlated random variables, e.g., common information in [41] and isomorphism in [29]. In this chapter, we use spectral method to introduce a new data processing inequality, which provides a single-letter necessary condition for the joint distributions satisfying the n-letter Markov chain.

2.2 Main Results

2.2.1 Some Preliminaries

In this section, we provide some basic results which will be used in our later development. The concepts used here are originally introduced by Witsenhausen in [41] in the context of operator theory. Here, we focus on the finite alphabet case, and derive our results in matrix form.

We first introduce our matrix notation for probability distributions. For a pair of discrete random variables X and Y, which take values in $\mathcal{X}$ and $\mathcal{Y}$, respectively, the $|\mathcal{X}| \times |\mathcal{Y}|$ joint probability distribution matrix P_{XY} is defined as

$$P_{XY}(i,j) \triangleq Pr(X = x_i, Y = y_j) \tag{2.6}$$

where $P_{XY}(i,j)$ denotes the (i,j)-th element of the matrix P_{XY}. The marginal distribution matrix of a random variable X, P_X, is defined as a diagonal matrix with

$$P_X(i,i) \triangleq Pr(X = x_i) \tag{2.7}$$

and the vector-form marginal distribution, p_X, is defined as[3]

$$p_X(i) \triangleq Pr(X = x_i) \tag{2.8}$$

or equivalently $p_X = P_X\mathbf{e}$, where $\mathbf{e}$ is the vector of all ones. p_X can also be defined as $p_X \triangleq P_{XY}$ for some degenerate random variable Y whose alphabet size $|\mathcal{Y}|$ is equal to one. For convenience, we define

$$p_X^{\frac{1}{2}} \triangleq P_X^{\frac{1}{2}}\mathbf{e} \tag{2.9}$$

For conditional distributions, we define matrix $P_{XY|z}$ as

$$P_{XY|z}(i,j) \triangleq Pr(X = x_i, Y = y_j | Z = z) \tag{2.10}$$

The vector-form conditional distribution $p_{X|z}$ is defined as

$$p_{X|z}(i) \triangleq Pr(X = x_i | Z = z) \tag{2.11}$$

or equivalently, $p_{X|z} \triangleq P_{XY|z}$ for some degenerate random variable Y whose alphabet

[3] In this chapter, we only consider the case where p_X is a positive vector.

size $|\mathcal{Y}|$ is equal to one.

We define a new matrix, $\tilde{P}_{XY}$, which will play an important role in the rest of the chapter, as

$$\tilde{P}_{XY} \triangleq P_X^{-\frac{1}{2}} P_{XY} P_Y^{-\frac{1}{2}} \tag{2.12}$$

Since $p_X \triangleq P_{XY}$ for some degenerate random variable Y whose alphabet size $|\mathcal{Y}|$ is equal to one, we define

$$\tilde{p}_X = P_X^{-\frac{1}{2}} P_{XY} P_Y^{-\frac{1}{2}} = P_X^{-\frac{1}{2}} p_X = p_X^{\frac{1}{2}} \tag{2.13}$$

The counterparts for conditional distributions, $\tilde{P}_{XY|z}$ and $\tilde{p}_{X|y}$, can be defined similarly.

A valid joint distribution matrix, P_{XY}, is a matrix whose entries are non-negative and sum to 1. Due to this constraint, not every matrix will qualify as a $\tilde{P}_{XY}$ corresponding to a joint distribution matrix as defined in (2.12). A necessary and sufficient condition for $\tilde{P}_{XY}$ to correspond to a joint distribution matrix is given in Theorem 2.2.1 below, which identifies the spectral properties of $\tilde{P}_{XY}$. Before stating the theorem, we provide a lemma and a definition regarding stochastic matrices, which will be used in the proof of the theorem.

Definition 2.2.1 *[6, p. 48] A square matrix T of order n is called (row) stochastic if*

$$T(i,j) \geq 0 \qquad i,j = 1,\dots,n, \qquad\qquad \sum_{j=1}^{n} T(i,j) = 1 \qquad i = 1,\dots,n \tag{2.14}$$

Lemma 2.2.1 *[6, p. 49] The spectral radius of a stochastic matrix is* 1*. A non-negative matrix T is stochastic if and only if* $\mathbf{e}$ *is an eigenvector of T corresponding to the eigenvalue* 1*.*

Theorem 2.2.1 *Assume a pair of given marginal distributions P_X and P_Y. A non-negative matrix P is a joint distribution matrix with marginal distributions P_X and P_Y, i.e., $P\mathbf{e} = p_X \triangleq P_X\mathbf{e}$ and $P^T\mathbf{e} = p_Y \triangleq P_Y\mathbf{e}$, if and only if the singular value decomposition (SVD) of the non-negative matrix $\tilde{P}$, which is defined as $\tilde{P} \triangleq P_X^{-\frac{1}{2}} P P_Y^{-\frac{1}{2}}$ satisfies*

$$\tilde{P} = M\Lambda N^T = p_X^{\frac{1}{2}}(p_Y^{\frac{1}{2}})^T + \sum_{i=2}^{l} \lambda_i \boldsymbol{\mu}_i \boldsymbol{\nu}_i^T \tag{2.15}$$

where $M \triangleq [\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_l]$ and $N \triangleq [\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_l]$ are two matrices such that $M^T M = I$ and $N^T N = I$, $\Lambda \triangleq \mathrm{diag}[\lambda_1, \ldots, \lambda_l]$ and $l = \min(|\mathcal{X}|, |\mathcal{Y}|)$; $\boldsymbol{\mu}_1 = p_X^{\frac{1}{2}}$, $\boldsymbol{\nu}_1 = p_Y^{\frac{1}{2}}$, and $\lambda_1 = 1 \geq \lambda_2 \geq \cdots \geq \lambda_l \geq 0$. That is, all of the singular values of $\tilde{P}$ are between 0 *and* 1*, the largest singular value of $\tilde{P}$ is* 1*, and the corresponding left and right singular vectors are $p_X^{\frac{1}{2}}$ and $p_Y^{\frac{1}{2}}$.*

Proof: We begin with the "if" part. We want to show that for any non-negative matrix P where the corresponding $\tilde{P} \triangleq P_X^{-\frac{1}{2}} P P_Y^{-\frac{1}{2}}$ satisfies (2.15), P is a joint distribution matrix with marginal distributions p_X and p_Y. Let $\tilde{P}$ satisfy (2.15), then

$$\begin{aligned}
P\mathbf{e} &= P_X^{\frac{1}{2}} \tilde{P} P_Y^{\frac{1}{2}} \mathbf{e} \\
&= P_X^{\frac{1}{2}} \left(p_X^{\frac{1}{2}} (p_Y^{\frac{1}{2}})^T + \sum_{i=2}^{l} \lambda_i \boldsymbol{\mu}_i \boldsymbol{\nu}_i^T \right) p_Y^{\frac{1}{2}} \\
&= P_X^{\frac{1}{2}} p_X^{\frac{1}{2}} (p_Y^{\frac{1}{2}})^T p_Y^{\frac{1}{2}} + P_X^{\frac{1}{2}} \sum_{i=2}^{l} \lambda_i \boldsymbol{\mu}_i \boldsymbol{\nu}_i^T \boldsymbol{\nu}_1 \\
&= p_X \qquad (2.16)
\end{aligned}$$

Similarly, $P^T\mathbf{e} = p_Y$. Thus, the non-negative matrix P is a joint distribution matrix with marginal distributions p_X and p_Y.

Now we turn to the "only if" part. We want to show that for any joint distribution matrix P with marginal distributions p_X and p_Y, (2.15) should be satisfied. We consider a joint distribution P with marginal distributions p_X and p_Y. We need to show that the singular values of $\tilde{P}$ lie in $[0, 1]$, the largest singular value is equal to 1, and $p_X^{\frac{1}{2}}$ and $p_Y^{\frac{1}{2}}$, respectively, are the left and right singular vectors corresponding to the singular value 1. To this end, we first construct a Markov chain $X \longrightarrow Y \longrightarrow Z$ with $P_{XY} = P_{ZY} = P$ (this construction comes from [41]). Note that this also implies $P_X = P_Z$, $\tilde{P}_{XY} = \tilde{P}_{ZY} = \tilde{P}$, and $P_{X|Y} = P_{Z|Y}$. The special structure of the constructed Markov chain provides the following:

$$\begin{aligned}
P_{X|Z} &= P_{X|Y}P_{Y|Z} \\
&= P_{X|Y}P_{Y|X} \\
&= PP_Y^{-1}P^TP_X^{-1} \\
&= P_X^{\frac{1}{2}}(P_X^{-\frac{1}{2}}PP_Y^{-\frac{1}{2}})(P_Y^{-\frac{1}{2}}P^TP_X^{-\frac{1}{2}})P_X^{-\frac{1}{2}} \\
&= P_X^{\frac{1}{2}}\tilde{P}\tilde{P}^TP_X^{-\frac{1}{2}} \qquad (2.17)
\end{aligned}$$

which implies that the matrix $P_{X|Z}$ is similar to the matrix $\tilde{P}\tilde{P}^T$ [20, p. 44]. Therefore, all the eigenvalues of $P_{X|Z}$ are the eigenvalues of $\tilde{P}\tilde{P}^T$ as well, and if $\boldsymbol{\nu}$ is a left eigenvector of $P_{X|Z}$ corresponding to an eigenvalue λ, then $P_X^{\frac{1}{2}}\boldsymbol{\nu}$ is a left eigenvector of $\tilde{P}\tilde{P}^T$ corresponding to the same eigenvalue.

We note that $P_{X|Z}^T$ is a stochastic matrix, therefore, from Lemma 2.2.1, $\mathbf{e}$ is a left eigenvector of $P_{X|Z}$ corresponding the eigenvalue 1, which is equal to the spectral radius of $P_{X|Z}$. Since $P_{X|Z}$ is similar to $\tilde{P}\tilde{P}^T$, we have that $p_X^{\frac{1}{2}}$ is a left eigenvector of $\tilde{P}\tilde{P}^T$ with eigenvalue 1, and all the eigenvalues of $\tilde{P}\tilde{P}^T$ lie in $[-1, 1]$. In addition, $\tilde{P}\tilde{P}^T$ is a symmetric positive semi-definite matrix, which implies that the eigenvalues of $\tilde{P}\tilde{P}^T$ are real and non-negative. Since the eigenvalues of $\tilde{P}\tilde{P}^T$ are non-negative, and the largest eigenvalue is equal to 1, we conclude that all of the eigenvalues of $\tilde{P}\tilde{P}^T$ lie in the interval $[0, 1]$.

The singular values of $\tilde{P}$ are the square roots of the eigenvalues of $\tilde{P}\tilde{P}^T$, and the left singular vectors of $\tilde{P}$ are the eigenvectors of $\tilde{P}\tilde{P}^T$. Thus, the singular values of

$\tilde{P}$ lie in $[0,1]$, the largest singular value is equal to 1, and $p_X^{\frac{1}{2}}$ is a left singular vector corresponding to the singular value 1. The corresponding right singular vector is

$$\boldsymbol{\nu}_1^T = \boldsymbol{\mu}_1^T \tilde{P} = (p_X^{\frac{1}{2}})^T P_X^{-\frac{1}{2}} P P_Y^{-\frac{1}{2}} = \mathbf{e}^T P P_Y^{-\frac{1}{2}} = p_Y^T P_Y^{-\frac{1}{2}} = (p_Y^{\frac{1}{2}})^T \tag{2.18}$$

which concludes the proof. ■

This theorem implies that there is a one-to-one mapping between all joint distribution matrices P and all non-negative matrices $\tilde{P}$ satisfying (2.15). It is easy to see from (2.12) that there is a corresponding $\tilde{P}$ for every P. Conversely, any given non-negative matrix $\tilde{P}$ satisfying (2.15) gives a unique pair of marginal distributions (P_X, P_Y), which is specified by the left and right positive singular vectors corresponding to its largest singular value[4]. Then, from (2.12), using $\tilde{P}$ and (P_X, P_Y) given by its singular vectors, we obtain a corresponding P as

$$P = P_X^{\frac{1}{2}} \tilde{P} P_Y^{\frac{1}{2}} \tag{2.19}$$

Because of this one-to-one relationship, exploring all possible joint distribution matrices P is equivalent to exploring all possible non-negative matrices $\tilde{P}$ satisfying (2.15).

Here, $\lambda_2, \ldots, \lambda_l$ can be viewed as a group of quantities, which measures the correlation between random variables X and Y. We note that when $\lambda_2 = \cdots = \lambda_l = 1$, X and Y are fully correlated, and, when $\lambda_2 = \cdots = \lambda_l = 0$, X and Y are indepen-

[4]We observe that there may exist multiple singular values equal to 1, but $\boldsymbol{\mu}_1$ and $\boldsymbol{\nu}_1$ are the only non-negative singular vectors.

dent. In all the cases between these two extremes, X and Y are arbitrarily correlated. Moreover, Witsenhausen showed that X and Y have a common data if and only if $\lambda_2 = 1$ [41]. In the next section, we will propose a new data processing inequality with respect to these new measures of correlation, $\lambda_2, \ldots, \lambda_l$. By utilizing this new data processing inequality, we will provide a single-letter necessary condition for the n-letter Markov chain $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$.

2.2.2 A New Data Processing Inequality

In this section, first, we introduce a new data processing inequality in the following theorem. Here, we provide a lemma that will be used in the proof of the theorem.

Lemma 2.2.2 *[21, p. 178] For matrices A and B*

$$\lambda_i(AB) \leq \lambda_i(A)\lambda_1(B) \tag{2.20}$$

where $\lambda_i(\cdot)$ denotes the i-th largest singular value of a matrix.

Theorem 2.2.2 *If $X \longrightarrow Y \longrightarrow Z$, then*

$$\lambda_i(\tilde{P}_{XZ}) \leq \lambda_i(\tilde{P}_{XY})\lambda_2(\tilde{P}_{YZ}) \leq \lambda_i(\tilde{P}_{XY}) \tag{2.21}$$

where $i = 2, \ldots, \mathrm{rank}(\tilde{P}_{XZ})$.

Proof: From the structure of the Markov chain, and from the definition of $\tilde{P}_{XY}$ in (2.12), we have

$$\begin{aligned}\tilde{P}_{XZ} &= P_X^{-\frac{1}{2}} P_{XZ} P_Z^{-\frac{1}{2}} \\ &= P_X^{-\frac{1}{2}} P_{XY} P_Y^{-\frac{1}{2}} P_Y^{-\frac{1}{2}} P_{YZ} P_Z^{-\frac{1}{2}} \\ &= \tilde{P}_{XY} \tilde{P}_{YZ} \end{aligned} \tag{2.22}$$

Using (2.15) for $\tilde{P}_{XZ}$, we obtain

$$\tilde{P}_{XZ} = p_X^{\frac{1}{2}} (p_Z^{\frac{1}{2}})^T + \sum_{i=2}^{l} \lambda_i(\tilde{P}_{XZ}) \boldsymbol{\mu}_i(\tilde{P}_{XZ}) \boldsymbol{\nu}_i(\tilde{P}_{XZ})^T \tag{2.23}$$

and applying (2.15) to $\tilde{P}_{XY}$ and $\tilde{P}_{YZ}$ yields

$$\begin{aligned}&\tilde{P}_{XY} \tilde{P}_{YZ} \\ &= \left(p_X^{\frac{1}{2}} (p_Y^{\frac{1}{2}})^T + \sum_{i=2}^{l} \lambda_i(\tilde{P}_{XY}) \boldsymbol{\mu}_i(\tilde{P}_{XY}) \boldsymbol{\nu}_i(\tilde{P}_{XY})^T \right) \times \\ &\quad \times \left(p_Y^{\frac{1}{2}} (p_Z^{\frac{1}{2}})^T + \sum_{i=2}^{l} \lambda_i(\tilde{P}_{YZ}) \boldsymbol{\mu}_i(\tilde{P}_{YZ}) \boldsymbol{\nu}_i(\tilde{P}_{YZ})^T \right) \\ &= p_X^{\frac{1}{2}} (p_Z^{\frac{1}{2}})^T + \left(\sum_{i=2}^{l} \lambda_i(\tilde{P}_{XY}) \boldsymbol{\mu}_i(\tilde{P}_{XY}) \boldsymbol{\nu}_i(\tilde{P}_{XY})^T \right) \left(\sum_{i=2}^{l} \lambda_i(\tilde{P}_{YZ}) \boldsymbol{\mu}_i(\tilde{P}_{YZ}) \boldsymbol{\nu}_i(\tilde{P}_{YZ})^T \right) \end{aligned} \tag{2.24}$$

where the two cross-terms vanish because $p_Y^{\frac{1}{2}}$ plays the roles of both $\boldsymbol{\nu}_1(\tilde{P}_{XY})$ and $\boldsymbol{\mu}_1(\tilde{P}_{YZ})$, and therefore, $p_Y^{\frac{1}{2}}$ is orthogonal to both $\boldsymbol{\nu}_i(\tilde{P}_{XY})$ and $\boldsymbol{\mu}_j(\tilde{P}_{YZ})$, for all $i, j \neq 1$.

Using (2.22) and equating (2.23) and (2.24), we obtain

$$\sum_{i=2}^{l} \lambda_i(\tilde{P}_{XZ})\boldsymbol{\mu}_i(\tilde{P}_{XZ})\boldsymbol{\nu}_i(\tilde{P}_{XZ})^T = \left(\sum_{i=2}^{l} \lambda_i(\tilde{P}_{XY})\boldsymbol{\mu}_i(\tilde{P}_{XY})\boldsymbol{\nu}_i(\tilde{P}_{XY})^T\right)\left(\sum_{i=2}^{l} \lambda_i(\tilde{P}_{YZ})\boldsymbol{\mu}_i(\tilde{P}_{YZ})\boldsymbol{\nu}_i(\tilde{P}_{YZ})^T\right) \tag{2.25}$$

The proof is completed by applying Lemma 2.2.2 to (2.25) and also by noting that $\lambda_2(\tilde{P}_{YZ}) \leq 1$ from Theorem 2.2.1. ■

Theorem 2.2.2 is a new data processing inequality in the sense that the processing from Y to Z reduces the correlation measure λ_i, i.e., the correlation between X and Z, $\lambda_i(\tilde{P}_{XZ})$, is less than or equal to the correlation measure between X and Y, $\lambda_i(\tilde{P}_{XY})$. We note that this theorem is similar to the data processing inequality in [11, p. 32] except instead of mutual information, we use $\lambda_i(\tilde{P}_{XY})$ as the correlation measure. In the sequel, we will show that this new data processing inequality helps us develop a necessary condition for the n-letter Markov chain while the data processing inequality in its usual form [11, p. 32] is not useful in this context.

2.2.3 A Necessary Condition for the n-letter Markov Chain

Now, we switch our attention to i.i.d. sequences of correlated sources. Let (U^n, V^n) be a pair of i.i.d. (in time) sequences, where each letter of these sequences satisfies a joint distribution P_{UV}. Thus, the joint distribution of the sequences is $P_{U^nV^n} = P_{UV}^{\otimes n}$, where $A^{\otimes 1} \triangleq A$, $A^{\otimes k} \triangleq A \otimes A^{\otimes(k-1)}$, and $\otimes$ denotes the Kronecker product of matrices [20].

From (2.12), we know that

$$P_{UV} = P_U^{\frac{1}{2}} \tilde{P}_{UV} P_V^{\frac{1}{2}} \tag{2.26}$$

Then,

$$P_{U^nV^n} = P_{UV}^{\otimes n} = (P_U^{\frac{1}{2}} \tilde{P}_{UV} P_V^{\frac{1}{2}})^{\otimes n} = (P_U^{\frac{1}{2}})^{\otimes n} \tilde{P}_{UV}^{\otimes n} (P_V^{\frac{1}{2}})^{\otimes n} \tag{2.27}$$

We also have $P_{U^n} = P_U^{\otimes n}$ and $P_{V^n} = P_V^{\otimes n}$. Thus,

$$\begin{aligned}
\tilde{P}_{U^nV^n} &\triangleq P_{U^n}^{-\frac{1}{2}} P_{U^nV^n} P_{V^n}^{-\frac{1}{2}} \\
&= (P_U^{-\frac{1}{2}})^{\otimes n} (P_U^{\frac{1}{2}})^{\otimes n} \tilde{P}_{UV}^{\otimes n} (P_V^{\frac{1}{2}})^{\otimes n} (P_V^{-\frac{1}{2}})^{\otimes n} \\
&= \tilde{P}_{UV}^{\otimes n} \qquad (2.28)
\end{aligned}$$

Now, applying SVD to $\tilde{P}_{U^nV^n}$, we have

$$\tilde{P}_{U^nV^n} = M_n \Lambda_n N_n^T = \tilde{P}_{UV}^{\otimes n} = M^{\otimes n} \Lambda^{\otimes n} (N^{\otimes n})^T \tag{2.29}$$

From the uniqueness of the SVD, we know that $M_n = M^{\otimes n}$, $\Lambda_n = \Lambda^{\otimes n}$ and $N_n = N^{\otimes n}$. Then, the ordered singular values of $\tilde{P}_{U^nV^n}$ are

$$\{1, \lambda_2(\tilde{P}_{UV}), \ldots, \lambda_2(\tilde{P}_{UV}), \ldots\}$$

where the second through the $n+1$-st singular values are all equal to $\lambda_2(\tilde{P}_{UV})$.

From Theorem 2.2.2, we know that if $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$, then, for $i = 2$,

$\ldots$, $\min(|\mathcal{X}_1|, |\mathcal{X}_2|)$,

$$\lambda_i(\tilde{P}_{X_1X_2}) \leq \lambda_2(\tilde{P}_{X_1U^n})\lambda_i(\tilde{P}_{U^nV^n})\lambda_2(\tilde{P}_{V^nX_2}) \tag{2.30}$$

We showed above that $\lambda_i(\tilde{P}_{U^nV^n}) \leq \lambda_2(\tilde{P}_{UV})$ for $i \geq 2$. Therefore, for $i = 2, \ldots,$ $\min(|\mathcal{X}_1|, |\mathcal{X}_2|)$, we have

$$\lambda_i(\tilde{P}_{X_1X_2}) \leq \lambda_2(\tilde{P}_{X_1U^n})\lambda_2(\tilde{P}_{UV})\lambda_2(\tilde{P}_{V^nX_2}) \tag{2.31}$$

From Theorem 2.2.1, we know that $\lambda_2(\tilde{P}_{X_1U^n}) \leq 1$ and $\lambda_2(\tilde{P}_{V^nX_2}) \leq 1$.

Based on the above discussion, we have the following theorem.

Theorem 2.2.3 *If $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$, then, for $i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|)$,*

$$\lambda_i(\tilde{P}_{X_1X_2}) \leq \lambda_2(\tilde{P}_{UV}) \tag{2.32}$$

We note that for a given finite n, the above theorem can be strengthened by developing a tighter upper bound for $\lambda_2(\tilde{P}_{X_1U^n})$ and $\lambda_2(\tilde{P}_{V^nX_2})$ in (2.31). However, we will show in Appendix 2.A that the smallest upper bound for $\lambda_2(\tilde{P}_{X_1U^n})$ and $\lambda_2(\tilde{P}_{V^nX_2})$ that is valid for all $n \in \mathbb{N}^+$ is equal to 1.

Theorem 2.2.3 provides a single-letter necessary condition for the n-letter Markov chain $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$ on the joint probability distribution $p(x_1, x_2)$. The set characterized by this single-leter condition is defined as follows.

$$\mathcal{S}'_{X_1X_2} \triangleq \{p(x_1, x_2) : \lambda_i(\tilde{P}_{X_1X_2}) \leq \lambda_2(\tilde{P}_{UV}),\ \text{for } i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|)\} \tag{2.33}$$

From Theorem 2.2.3, we have

$$\mathcal{S}_{X_1X_2} \subseteq \mathcal{S}'_{X_1X_2} \tag{2.34}$$

where $\mathcal{S}_{X_1X_2}$ is defined in (2.4).

Theorem 2.2.3 also answers the question we posed in Section 2.1. Our question was whether (X_1, X_2) can be arbitrarily correlated, when we allow n to take any value in $\mathbb{N}^+$. Theorem 2.2.3 shows that (X_1, X_2) cannot be arbitrarily correlated, as the correlation measures between (X_1, X_2), $\lambda_i(\tilde{P}_{X_1X_2})$, are upper bounded by, $\lambda_2(\tilde{P}_{UV})$, the second correlation measure of the single-letter sources (U, V), no matter what value n takes.

As we mentioned in Section 2.1, the data processing inequality in its usual form [11, p. 32] is not helpful in this problem, while our new data processing inequality, i.e., Theorem 2.2.2, provides a single-letter necessary condition for this n-letter Markov chain. The main reason for this difference is that while the mutual information, $I(U^n; V^n)$, the correlation measure in the original data processing inequality, increases linearly with n, $\lambda_i(\tilde{P}_{U^nV^n})$, the correlation measure in our new data processing inequality, is bounded as n increases, and therefore, makes the problem more tractable.

Theorem 2.2.3 is valid for all discrete random variables. To illustrate the utility and also the limitations of Theorem 2.2.3, we will study a binary example in detail in Appendix 2.B. In this example, (U, V, X_1, X_2) are all binary random variables. For this specific binary example, we will apply Theorem 2.2.3 to obtain a necessary condition for the n-letter Markov chain. Moreover, the special structure of this binary example will enable us to provide a sharper necessary condition than the one given in Theorem 2.2.3. We will compare these two necessary conditions and a sufficient condition for this binary example.

2.2.4 General Result

Theorem 2.2.3 in Section 2.2.3 provides a necessary condition for joint probability distributions $p(x_1, x_2)$, which satisfy the Markov chain $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$. In certain specific problems, e.g., multiple access channel with correlated sources and multi-terminal rate-distortion problem, in addition to $p(x_1, x_2)$, the distributions of (X_1, X_2) conditioned on parts of the n-letter sources may be needed[5], e.g., $p(x_1, x_2|u_1, v_1)$. In this section, we will develop a result similar to that in Theorem 2.2.3 for conditional distributions.

For a pair of i.i.d. sequences (U^n, V^n) of length n, we define $\underline{U}$ as an arbitrary subset of $\{U_1, \ldots, U_n\}$, i.e.,

$$\underline{U} \triangleq \{U_{i_1}, \ldots, U_{i_l}\} \subset \{U_1, \ldots, U_n\} \tag{2.35}$$

and similarly,

$$\underline{V} \triangleq \{V_{j_1}, \ldots, V_{j_k}\} \subset \{V_1, \ldots, V_n\} \tag{2.36}$$

In the following theorem, we propose an upper bound for $\lambda_i(\tilde{P}_{X_1X_2|\underline{u}\underline{v}})$, when $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$ is satisfied.

[5]The reader may wish to consult Sections 2.3 and 2.4 for further motivations to consider conditional probability distributions.

Theorem 2.2.4 *Let* (U^n, V^n) *be a pair of i.i.d. sequences of length* n*, and let the random variables* X_1, X_2 *satisfy* $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$*. Then, for* $i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|)$,

$$\lambda_i(\tilde{P}_{X_1X_2|\underline{uv}}) \leq \lambda_2(\tilde{P}_{UV}) \tag{2.37}$$

where $\underline{U} \subset \{U_1, \ldots, U_n\}$ *and* $\underline{V} \subset \{V_1, \ldots, V_n\}$.

Proof: We consider a special case of $(\underline{U}, \underline{V})$ as follows. We define $\underline{U} \triangleq \{U_1, \ldots, U_l\}$ and $\underline{V} \triangleq \{V_1, \ldots, V_m, V_{l+1}, \ldots, V_{l+k-m}\}$. We also define the complements of $\underline{U}$ and $\underline{V}$ as: $\underline{U}^c \triangleq \{U_1, \ldots, U_n\} \backslash \underline{U}$ and $\underline{V}^c \triangleq \{V_1, \ldots, V_n\} \backslash \underline{V}$. If $\underline{U}$ and $\underline{V}$ take other forms, we can transform them to the form we defined above by permutations. We know that

$$p(x_1, x_2, \underline{u}^c, \underline{v}^c | \underline{u}, \underline{v}) = p(x_1 | \underline{u}^c, \underline{u}, \underline{v}) p(\underline{u}^c, \underline{v}^c | \underline{u}, \underline{v}) p(x_2 | \underline{v}^c, \underline{v}, \underline{u}) \tag{2.38}$$

In other words, given $\underline{U} = \underline{u}$ and $\underline{V} = \underline{v}$, $(X_1, \underline{U}^c, \underline{V}^c, X_2)$ form a Markov chain. Thus, from (2.22),

$$\tilde{P}_{X_1X_2|\underline{uv}} = \tilde{P}_{X_1\underline{U}^c|\underline{uv}} \tilde{P}_{\underline{U}^c\underline{V}^c|\underline{uv}} \tilde{P}_{\underline{V}^cX_2|\underline{uv}} \tag{2.39}$$

Furthermore,

$$\tilde{P}_{\underline{U}^c\underline{V}^c|\underline{uv}} = \tilde{p}^T_{V^l_{m+1}|u^l_{m+1}} \otimes \tilde{p}_{U^{l+k-m}_{l+1}|v^{l+k-m}_{l+1}} \otimes \tilde{P}_{U^n_{l+k-m+1}V^n_{l+k-m+1}} \tag{2.40}$$

As mentioned earlier, a vector marginal distribution can be viewed as a joint distribution matrix with a degenerate random variable whose alphabet size is equal to

1. Since the rank of a vector is 1, from Theorem 2.2.1, the sole singular value of $\tilde{p}_{V_{m+1}^l|u_{m+1}^l}$ (and of $\tilde{p}_{U_{l+1}^{l+k-m}|v_{l+1}^{l+k-m}}$) is equal to 1. Then,

$$\lambda_i(\tilde{P}_{\underline{U}^c\underline{V}^c|\underline{uv}}) = \lambda_i(\tilde{P}_{U_{l+k-m+1}^n V_{l+k-m+1}^n}) \tag{2.41}$$

Combining (2.21), (2.39), and (2.41), we obtain

$$\lambda_i(\tilde{P}_{X_1X_2|\underline{uv}}) \leq \lambda_2(\tilde{P}_{UV}) \tag{2.42}$$

which completes the proof. ■

Now we focus on the conditional distribution $p(x_1, x_2|u_1, u_2)$. We are interested in the set of all possible conditional distributions $p(x_1, x_2|u, v)$ satisfying $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$, i.e., the following set

$$\mathcal{S}_{X_1X_2|UV} \triangleq \left\{ p(x_1, x_2|u_1, v_1) : \begin{array}{l} \exists n \in \mathbb{N}^+, p(x_1|u^n), p(x_2|v^n), \text{ s.t.} \\ p(x_1, x_2|u_1, v_1) = \frac{\sum_{u_2,\ldots,u_n,v_2\ldots,v_n} p(x_1|u^n)p(x_2|v^n)p(u^n,v^n)}{p(u_1,v_1)} \end{array} \right\} \tag{2.43}$$

with $p(u^n, v^n)$ satisfying (2.1) and (2.2). Here, we simplify the notation by omitting the subscripts in U_1 and V_1 in $\mathcal{S}_{X_1X_2|UV}$, i.e., We note that $p(x_1, x_2)$, $p(x_1, x_2|u_1)$ and

$p(x_1, x_2|v_1)$ are all functions of $p(x_1, x_2|u_1, v_1)$ for given $p(u_1, v_1)$, i.e.,

$$p(x_1, x_2) = \sum_{u_1, v_1} p(x_1, x_2|u_1, v_1)p(u_1, v_1) \tag{2.44}$$

$$p(x_1, x_2|u_1) = \sum_{v_1} p(x_1, x_2|u_1, v_1)p(u_1, v_1) \tag{2.45}$$

$$p(x_1, x_2|v_1) = \sum_{u_1} p(x_1, x_2|u_1, v_1)p(u_1, v_1) \tag{2.46}$$

Thus, $\lambda_i(\tilde{P}_{X_1X_2})$, $\lambda_i(\tilde{P}_{X_1X_2|U_1})$, $\lambda_i(\tilde{P}_{X_1X_2|V_1})$, as well as $\lambda_i(\tilde{P}_{X_1X_2|U_1V_1})$ are all functions of $p(x_1, x_2|u_1, v_1)$ for given $p(u_1, v_1)$. We define the set $\mathcal{S}'_{X_1X_2|UV}$ as follows

$$\mathcal{S}'_{X_1X_2|U_1V_1} \triangleq \left\{ p(x_1, x_2|u_1, v_1) : \begin{array}{ll} \lambda_i(\tilde{P}_{X_1X_2}) \leq \lambda_2(\tilde{P}_{UV}) & i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|) \\ \lambda_i(\tilde{P}_{X_1X_2|U_1}) \leq \lambda_2(\tilde{P}_{UV}) & i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|) \\ \lambda_i(\tilde{P}_{X_1X_2|V_1}) \leq \lambda_2(\tilde{P}_{UV}) & i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|) \\ \lambda_i(\tilde{P}_{X_1X_2|U_1V_1}) \leq \lambda_2(\tilde{P}_{UV}) & i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|) \end{array} \right\} \tag{2.47}$$

By applying Theorem 2.2.4 on $p(x_1, x_2)$, $p(x_1, x_2|u_1)$, $p(x_1, x_2|v_1)$ and $p(x_1, x_2|u_1, v_1)$, respectively, we obtain

$$\mathcal{S}_{X_1X_2|UV} \subseteq \mathcal{S}'_{X_1X_2|UV} \tag{2.48}$$

2.3 Example I: Multiple Access Channel with Correlated Sources

The problem of determining the capacity region of the multiple access channel with correlated sources can be formulated as follows. Given a pair of i.i.d. correlated sources (U,V) described by the joint probability distribution $p(u,v)$, and a discrete, memoryless, multiple access channel characterized by the transition probability $p(y|x_1,x_2)$, what are the necessary and sufficient conditions for the reliable transmission of n samples of the sources through the channel, in n channel uses, as $n \longrightarrow \infty$?

2.3.1 Existing Results

The multiple access channel with correlated sources was studied by Cover, El Gamal and Salehi in [9] (a simpler proof was given in [2]), where an achievable region expressed by single-letter entropies and mutual informations was given as follows.

Theorem 2.3.1 *[9] A source (U,V) with joint distribution $p(u,v)$ can be sent with arbitrarily small probability of error over a multiple access channel characterized by $p(y|x_1,x_2)$, if there exist probability mass functions $p(s)$, $p(x_1|u,s)$, $p(x_2|v,s)$, such that*

$$H(U|V) < I(X_1;Y|X_2,V,S) \tag{2.49}$$

$$H(V|U) < I(X_2;Y|X_1,U,S) \tag{2.50}$$

$$H(U,V|W) < I(X_1,X_2;Y|W,S) \tag{2.51}$$

$$H(U,V) < I(X_1,X_2;Y) \tag{2.52}$$

where

$$p(s,u,v,x_1,x_2,y) = p(s)p(u,v)p(x_1|u,s)p(x_2|v,s)p(y|x_1,x_2) \tag{2.53}$$

and

$$w = f(u) = g(v) \tag{2.54}$$

is the common information in the sense of Gacs and Korner (see [41]).

The above region can be simplified if there is no common information between U and V as follows [9]

$$H(U|V) < I(X_1; Y|X_2, V) \tag{2.55}$$

$$H(V|U) < I(X_2; Y|X_1, U) \tag{2.56}$$

$$H(U, V) < I(X_1, X_2; Y) \tag{2.57}$$

where

$$p(u, v, x_1, x_2, y) = p(u, v)p(x_1|u)p(x_2|v)p(y|x_1, x_2) \tag{2.58}$$

This achievable region was shown to be suboptimal by Dueck [16].

Cover, El Gamal and Salehi [9] also provided a capacity result with both achievability and converse in the form of some incomputable n-letter mutual informations. Their result is restated in the following theorem.

Theorem 2.3.2 *[9] The correlated sources (U, V) can be communicated reliably over the discrete memoryless multiple access channel $p(y|x_1, x_2)$ if and only if*

$$[H(U|V), H(V|U), H(U, V)] \in \bigcup_{n=1}^{\infty} \mathcal{C}_n \tag{2.59}$$

where

$$\mathcal{C}_n = \left\{ [R_1, R_2, R_3] : \begin{array}{c} R_1 < \frac{1}{n} I(X_1^n; Y^n | X_2^n, V^n) \\ R_2 < \frac{1}{n} I(X_2^n; Y^n | X_1^n, U^n) \\ R_3 < \frac{1}{n} I(X_1^n, X_2^n; Y^n) \end{array} \right\} \tag{2.60}$$

for some

$$p(u^n, v^n, x_1^n, x_2^n, y^n) = p(x_1^n | u^n) p(x_2^n | v^n) \prod_{i=1}^{n} p(u_i, v_i) \prod_{i=1}^{n} p(y_i | x_{1i}, x_{2i}) \tag{2.61}$$

i.e., for some X_1^n and X_2^n that satisfy the Markov chain $X_1^n \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2^n$.

Some recent results on the transmission of correlated sources over multiple access channels can be found in [27, 30].

2.3.2 A New Outer Bound

We propose a new outer bound for the multiple access channel with correlated sources as follows.

Theorem 2.3.3 *If a pair of i.i.d. sources (U, V) with joint distribution $p(u, v)$ can be transmitted reliably through a discrete, memoryless, multiple access channel characterized by $p(y|x_1, x_2)$, then*

$$H(U|V) \leq I(X_1; Y | X_2, U, Q) \tag{2.62}$$

$$H(V|U) \leq I(X_2; Y | X_1, V, Q) \tag{2.63}$$

$$H(U, V) \leq I(X_1, X_2; Y | Q) \tag{2.64}$$

where random variables X_1, X_2 and Q are such that

$$p(x_1,x_2,y,u,v,q) = p(q)p(u,v)p(y|x_1,x_2)p(x_1,x_2|u,v,q) \tag{2.65}$$

and for every given q,

$$p(x_1,x_2|u,v,Q=q) \in \mathcal{S}_{X_1X_2|UV} \tag{2.66}$$

with $\mathcal{S}_{X_1X_2|UV}$ defined in (2.43).

Note that every quantity in this theorem is in the form of a single-letter except the conditional distribution $p(x_1,x_2|u,v,Q=q) \in \mathcal{S}_{X_1X_2|UV}$, which will be relaxed to a single-letter form in the next section.

Proof: Consider a given block code of length n with encoders $f_1 : \mathcal{U}^n \longmapsto \mathcal{X}_1^n$ and $f_2 : \mathcal{V}^n \longmapsto \mathcal{X}_2^n$ and decoder $g : \mathcal{Y}^n \longmapsto \mathcal{U}^n \times \mathcal{V}^n$. From Fano's inequality [11, p. 39], we have

$$H(U^n,V^n|Y^n) \le n\log_2|\mathcal{U}\times\mathcal{V}|P_e + 1 \triangleq n\epsilon_n \tag{2.67}$$

For a code, for which $P_e \longrightarrow 0$, as $n \longrightarrow \infty$, we have $\epsilon_n \longrightarrow 0$. Then,

$$
\begin{aligned}
nH(U|V) &= H(U^n|V^n) \\
&= I(U^n;Y^n|V^n) + H(U^n|Y^n,V^n) \\
&\leq I(U^n;Y^n|V^n) + H(U^n,V^n|Y^n) \\
&\stackrel{1}{\leq} I(U^n;Y^n|V^n) + n\epsilon_n \\
&= H(Y^n|V^n) - H(Y^n|U^n,V^n) + n\epsilon_n \\
&\stackrel{2}{=} H(Y^n|X_2^n,V^n) - H(Y^n|X_1^n,X_2^n,U^n,V^n) + n\epsilon_n \\
&\stackrel{3}{=} H(Y^n|X_2^n,V^n) - H(Y^n|X_1^n,X_2^n) + n\epsilon_n \\
&\stackrel{4}{=} \sum_{i=1}^{n} \Big[H(Y_i|X_2^n,V^n,Y^{i-1}) - H(Y_i|X_{1i},X_{2i}) \Big] + n\epsilon_n \\
&\stackrel{5}{\leq} \sum_{i=1}^{n} \Big[H(Y_i|X_{2i},V_i) - H(Y_i|X_{1i},X_{2i}) \Big] + n\epsilon_n \\
&\stackrel{6}{=} \sum_{i=1}^{n} \Big[H(Y_i|X_{2i},V_i) - H(Y_i|X_{1i},X_{2i},V_i) \Big] + n\epsilon_n \\
&= \sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i},V_i) + n\epsilon_n
\end{aligned}
\tag{2.68}
$$

where

1. from Fano's inequality in (2.67);
2. from the fact that X_1^n is a deterministic function of U^n and X_2^n is a deterministic function of V^n;
3. from $p(y^n|x_1^n,x_2^n,u^n,v^n) = p(y^n|x_1^n,x_2^n)$;
4. from the chain rule and the memoryless nature of the channel;

5. from the property that conditioning reduces entropy;

6. from $p(y_i|x_{1i}, x_{2i}, v_i) = p(y_i|x_{1i}, x_{2i})$.

Using a symmetrical argument, we obtain

$$nH(V|U) \leq \sum_{i=1}^{n} I(X_{2i}; Y_i|X_{1i}, U_i) + n\epsilon_n \tag{2.69}$$

Moreover,

$$\begin{aligned}
nH(U, V) &= H(U^n, V^n) \\
&= I(U^n, V^n; Y^n) + H(U^n, V^n|Y^n) \\
&\leq I(U^n, V^n; Y^n) + n\epsilon_n \\
&\leq I(X_1^n, X_2^n; Y^n) + n\epsilon_n \\
&= H(Y^n) - H(Y^n|X_1^n, X_2^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} \Big[H(Y_i|Y^{i-1}) - H(Y_i|X_{1i}, X_{2i}) \Big] + n\epsilon_n \\
&\leq \sum_{i=1}^{n} \Big[H(Y_i) - H(Y_i|X_{1i}, X_{2i}) \Big] + n\epsilon_n \\
&= \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i) + n\epsilon_n
\end{aligned} \tag{2.70}$$

We note that the following three expressions, $I(X_{1i}; Y_i|X_{2i}, V_i)$, $I(X_{2i}; Y_i|X_{1i}, U_i)$, and $I(X_{1i}, X_{2i}; Y_i)$, only depend on the marginal conditional distribution $p(x_{1i}, x_{2i}|u_i, v_i)$ with given $p(u_i, v_i)$ and $p(y_i|x_{1i}, x_{2i})$. We also note that X_{1i} is a function of U^n and X_{2i} is a function of V^n. Thus $X_{1i} \longrightarrow U^n \longrightarrow V^n \longrightarrow X_{2i}$, and therefore $p(x_{1i}, x_{2i}|u_i, v_i) \in \mathcal{S}_{X_1X_2|UV}$.

We introduce a time-sharing random variable Q [11, p. 397] as follows. Let Q be uniformly distributed on $\{1, \ldots, n\}$ and be independent of U, V, i.e.,

$$p(u, v, q) = p(q)p(u, v) \tag{2.71}$$

Define random variables X_1 and X_2 to be such that

$$p(x_1, x_2|u, v, Q = i) = p(x_{1i}, x_{2i}|u_i, v_i) \tag{2.72}$$

and $p(x_1, x_2|u, v, Q = i) \in \mathcal{S}_{X_1X_2|U_1V_1}$ for all $i = 1, \ldots, n$. Then,

$$\sum_{i=1}^{n} I(X_{1i}; Y_i|X_{2i}, V_i) = nI(X_1; Y|X_2, V, Q) \tag{2.73}$$

$$\sum_{i=1}^{n} I(X_{2i}; Y_i|X_{1i}, U_i) = nI(X_2; Y|X_1, U, Q) \tag{2.74}$$

$$\sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i) = nI(X_1, X_2; Y|Q) \tag{2.75}$$

Combining (2.73), (2.74) and (2.75) with (2.68), (2.69) and (2.70) completes the proof.

■

2.3.3 A New Necessary Condition

It can be shown that the outer bound in Theorem 2.3.3 is equivalent to the following

$$\mathbf{H} \in \mathcal{R}(\mathcal{S}_{X_1X_2|UV}) \triangleq \mathrm{co}\Big\{ \bigcup_{\mathbf{p}\in\mathcal{S}_{X_1X_2|UV}} \mathcal{R}(\mathbf{p}) \Big\} \tag{2.76}$$

where

$$\mathbf{H} \triangleq [H(U|V), H(V|U), H(U,V)] \tag{2.77}$$

$$\mathbf{p} \triangleq p(x_1, x_2|u, v) \tag{2.78}$$

$$\mathcal{R}(\mathbf{p}) \triangleq \left\{ [R_1, R_2, R_3] : \begin{array}{l} R_1 \leq I(X_1; Y|X_2, V) \\ R_2 \leq I(X_2; Y|X_1, U) \\ R_3 \leq I(X_1, X_2; Y) \end{array} \right\} \tag{2.79}$$

and co$\{\cdot\}$ represents the closure of the convex hull of the set argument.

From Section 2.2.4, we know that

$$\mathcal{S}_{X_1X_2|UV} \subseteq \mathcal{S}'_{X_1X_2|UV} \tag{2.80}$$

Then, we obtain a single-letter outer bound for the multiple access channel with correlated sources as follows.

Theorem 2.3.4 *If a pair of i.i.d. sources (U, V) with joint distribution $p(u, v)$ can be transmitted reliably through a discrete, memoryless, multiple access channel characterized by $p(y|x_1, x_2)$, then*

$$H(U|V) \leq I(X_1; Y|X_2, V, Q) \tag{2.81}$$

$$H(V|U) \leq I(X_2; Y|X_1, U, Q) \tag{2.82}$$

$$H(U, V) \leq I(X_1, X_2; Y|Q) \tag{2.83}$$

where random variable Q independent of (U,V), and random variables X_1, X_2 with conditional distribution $p(x_1,x_2|u,v,q)$ are such that, for any given $Q=q$

$$\lambda_i(\tilde{P}_{X_1X_2|q}) \le \lambda_2(\tilde{P}_{UV}) \qquad i=2,\dots,\min(|\mathcal{X}_1|,|\mathcal{X}_2|) \tag{2.84}$$

$$\lambda_i(\tilde{P}_{X_1X_2|u_1q}) \le \lambda_2(\tilde{P}_{UV}) \qquad i=2,\dots,\min(|\mathcal{X}_1|,|\mathcal{X}_2|) \tag{2.85}$$

$$\lambda_i(\tilde{P}_{X_1X_2|v_1q}) \le \lambda_2(\tilde{P}_{UV}) \qquad i=2,\dots,\min(|\mathcal{X}_1|,|\mathcal{X}_2|) \tag{2.86}$$

$$\lambda_i(\tilde{P}_{X_1X_2|u_1v_1q}) \le \lambda_2(\tilde{P}_{UV}) \qquad i=2,\dots,\min(|\mathcal{X}_1|,|\mathcal{X}_2|) \tag{2.87}$$

Equivalently,

$$\mathbf{H} \in \mathcal{R}(\mathcal{S}'_{X_1X_2|UV}) \triangleq \mathrm{co}\Big\{ \bigcup_{\mathbf{p}\in\mathcal{S}'_{X_1X_2|UV}} \mathcal{R}(\mathbf{p}) \Big\} \tag{2.88}$$

2.3.4 Numerical Example

In this section, we give some simple numerical examples to show the improvement our proposed outer bound provides with respect to the cut-set bound [11]. For simplicity, we only consider the sum-rate here. Assume a multiple access channel where the alphabets of X_1, X_2 and Y are all binary, and the channel transition probability matrix $p(y|x_1,x_2)$ is given as

$Y\backslash X_1X_2$	11	10	01	00
1	1	1/2	1/2	0
0	0	1/2	1/2	1

The following is the cut-set bound for the sum-rate, which we provide as a benchmark,

$$H(U,V) < \max_{p(x_1,x_2)} I(X_1,X_2;Y) = 1 \tag{2.89}$$

where the maximization is over all binary bivariate distributions. The maximum is achieved by $P(X_1 = 1, X_2 = 1) = P(X_1 = 0, X_2 = 0) = 1/2$. We note that the cut-set bound does not depend on the source distribution. We specify the single-letter necessary condition we proposed in Section 2.3.3 and obtain the following upper bound on the sum-rate

$$H(U,V) < \max_{p(x_1,x_2):\lambda_2(\tilde{P}_{X_1X_2})\leq\lambda_2(\tilde{P}_{UV})} I(X_1,X_2;Y) \tag{2.90}$$

Note that we are using a weakened version of our outer bound in Theorem 2.3.4. Theorem 2.3.4 restricts probability distribution $p(x_1,x_2,u_1,v_1)$ by imposing four constraints in (2.84), (2.85), (2.86) and (2.87). We weaken our outer bound by imposing only (2.84) on probability distribution $p(x_1,x_2)$.

We also consider the achievable sum-rate proposed in [9]

$$H(U,V) \leq \max_{X_1\longrightarrow U\longrightarrow V\longrightarrow X_2} I(X_1,X_2;Y) \tag{2.91}$$

First, we consider a binary source (U,V) with the following joint distribution

$p(u, v)$

$U \backslash V$	1	0
1	1/3	1/6
0	1/6	1/3

In this case, $H(U, V) = 1.92$. We first note by using the cut-set bound in (2.89) that it is impossible to transmit this source through the given channel reliably. The upper bound we developed in this chapter gives 2/3 for this source. We also note that, for this case, our upper bound coincides with the single-letter achievability expression. Therefore, for this case, our upper bound on sum-rate is tight, as it matches the achievability expression.

Next, we consider a binary source (U, V) with the following joint distribution $p(u, v)$

$U \backslash V$	1	0
1	0	0.1
0	0.1	0.8

In this case, $H(U, V) = 0.92$, the single-letter achievability in (2.91) reaches 0.51 and our upper bound is 0.56. We note that, in this case, the cut-set bound in (2.89) fails to test whether it is possible to have reliable transmission or not, while our upper bound determines conclusively that reliable transmission is not possible.

Finally, we consider a binary source (U, V) with the following joint distribution

$p(u, v)$

$U \backslash V$	1	0
1	0	0.85
0	0.1	0.05

In this case, $H(U, V) = 0.75$, the single-letter achievability expression in (2.91) gives 0.57 and our upper bound is 0.9. We note that the joint entropy of the sources falls into the gap between the achievability expression and our upper bound, which means that we cannot conclude whether it is possible (or not) to transmit these sources through the channel reliably.

2.4 Example II: Multi-terminal Rate-distortion Region

Ever since the milestone paper of Wyner and Ziv [42] on the rate-distortion function of a single source with side information at the decoder, there has been a significant amount of efforts directed towards solving a generalization of this problem, the so called multi-terminal rate-distortion problem. Among all the attempts on this difficult problem, the notable works by Tung [37] and Housewright [22] (see also [4]) provide the inner and outer bounds for the rate-distortion region. A more recent progress on this problem is by Wagner and Anantharam in [39], where a tighter outer bound is given. A very recent result can be found in [33].

The multi-terminal rate-distortion problem can be formulated as follows. Consider a pair of discrete memoryless sources (U, V), with joint distribution $p(u, v)$ defined on the finite alphabet $\mathcal{U} \times \mathcal{V}$. The reconstruction of the sources is built on another finite alphabet $\hat{\mathcal{U}} \times \hat{\mathcal{V}}$. The distortion measures are defined as $d_1 : \mathcal{U} \times \hat{\mathcal{U}} \longmapsto \mathbb{R}^+ \cup \{0\}$ and $d_2 : \mathcal{V} \times \hat{\mathcal{V}} \longmapsto \mathbb{R}^+ \cup \{0\}$. Assume that two distributed encoders are functions $f_1 : \mathcal{U}^n \longmapsto \{1, 2, \dots, M_1\}$ and $f_2 : \mathcal{V}^n \longmapsto \{1, 2, \dots, M_2\}$ and a joint decoder is the function $g : \{1, 2, \dots, M_1\} \times \{1, 2, \dots, M_2\} \longmapsto \hat{\mathcal{U}}^n \times \hat{\mathcal{V}}^n$, where n is a positive integer. A pair of distortion levels $\mathbf{D} \triangleq (D_1, D_2)$ is said to be $\mathbf{R}$-attainable, for some rate pair $\mathbf{R} \triangleq (R_1, R_2)$, if for all $\epsilon > 0$ and $\delta > 0$, there exist, some positive integer n and a set of distributed encoders and joint decoder (f_1, f_2, g) with rates[6] $(\frac{1}{n} \log_2 M_1, \frac{1}{n} \log_2 M_2) = (R_1 + \delta, R_2 + \delta)$, such that the distortion between the sources (U^n, V^n) and the decoder output $(\hat{U}^n, \hat{V}^n)$ satisfies $\big(Ed_1(U^n, \hat{V}^n), Ed_2(V^n, \hat{V}^n)\big) \leq (D_1 + \epsilon, D_2 + \epsilon)$ where $d_1(U^n, \hat{U}^n) \triangleq \frac{1}{n} \sum_{i=1}^n d_1(U_i, \hat{U}_i)$ and $d_2(V^n, \hat{V}^n) \triangleq \frac{1}{n} \sum_{i=1}^n d_2(V_i, \hat{V}_i)$. The problem here is to determine, for a fixed $\mathbf{D}$, the set $\mathcal{R}(\mathbf{D})$ of all rate pairs $\mathbf{R}$, for which $\mathbf{D}$ is $\mathbf{R}$-attainable.

2.4.1 Existing Results

We restate the inner bound provided in [37] and [22] in the following theorem.

Theorem 2.4.1 *[22, 37] $\mathcal{R}(\mathbf{D}) \supseteq \mathcal{R}_{in}(\mathbf{D})$, where $\mathcal{R}_{in}(\mathbf{D})$ is the set of all $\mathbf{R}$ such that there exists a pair of discrete random variables (X_1, X_2), for which the following three conditions are satisfied:*

1. *The joint distribution satisfies*

$$X_1 \longrightarrow U \longrightarrow V \longrightarrow X_2 \tag{2.92}$$

[6] By $(A, B) < (C, D)$, we mean both $A < B$ and $C < D$, and $(A, B) \leq (C, D)$ is defined in the similar manner.

2. The rate pair satisfies

$$R_1 \geq I(U,V;X_1|X_2) \tag{2.93}$$

$$R_2 \geq I(U,V;X_2|X_1) \tag{2.94}$$

$$R_1 + R_2 \geq I(U,V;X_1,X_2) \tag{2.95}$$

3. There exists $\big(\hat{U}(X_1,X_2),\hat{V}(X_1,X_2)\big)$ *such that* $\big(Ed_1(U,\hat{U}),Ed_2(V,\hat{V})\big) \leq \mathbf{D}$*.*

An outer bound is also given in [37] and [22] as follows.

Theorem 2.4.2 *[22,37]* $\mathcal{R}(\mathbf{D}) \subseteq \mathcal{R}_{out,1}(\mathbf{D})$*, where* $\mathcal{R}_{out,1}(\mathbf{D})$ *is the set of all* $\mathbf{R}$ *such that there exists a pair of discrete random variables* (X_1,X_2)*, for which the following three conditions are satisfied:*

1. The joint distribution satisfies

$$X_1 \longrightarrow U \longrightarrow V \tag{2.96}$$

$$U \longrightarrow V \longrightarrow X_2 \tag{2.97}$$

2. The rate pair satisfies

$$R_1 \geq I(U,V;X_1|X_2) \tag{2.98}$$

$$R_2 \geq I(U,V;X_2|X_1) \tag{2.99}$$

$$R_1 + R_2 \geq I(U,V;X_1,X_2) \tag{2.100}$$

3. *There exists* $\big(\hat{U}(X_1, X_2), \hat{V}(X_1, X_2)\big)$ *such that* $\big(Ed_1(U, \hat{U}), Ed_2(V, \hat{V})\big) \leq \mathbf{D}$.

A tighter upper bound was recently proposed by Wagner and Anantharam as follows[7].

Theorem 2.4.3 *[39]* $\mathcal{R}(\mathbf{D}) \subseteq \mathcal{R}_{out,2}(\mathbf{D})$*, where* $\mathcal{R}_{out,2}(\mathbf{D})$ *is the set of all* $\mathbf{R}$ *such that there exists a pair of discrete random variables* (X_1, X_2)*, for which the following three conditions are satisfied:*

1. *The joint distribution satisfies*

$$p(x_1, x_2|u, v) :\exists \text{ random variable } W,$$

$$p(x_1, x_2|u, v) = \sum_w p(w)p(x_1|w, u)p(x_2|w, v) \tag{2.101}$$

This distribution may be represented by the following Markov chain like notation

$$\begin{array}{ccccc} X_1 & \longrightarrow U & \longrightarrow & V \longrightarrow & X_2 \\ & \searrow & & \nearrow & \\ & & W & & \end{array} \tag{2.102}$$

2. *The rate pair satisfies*

$$R_1 \geq I(U, V; X_1|X_2) \tag{2.103}$$

$$R_2 \geq I(U, V; X_2|X_1) \tag{2.104}$$

$$R_1 + R_2 \geq I(U, V; X_1, X_2) \tag{2.105}$$

[7] This is a simplified version of [39] with the assumption that there is no non-trivial hidden source behind (U^n, V^n).

3. *There exists* $\left(\hat{U}(X_1,X_2),\hat{V}(X_1,X_2)\right)$ *such that* $\left(Ed_1(U,\hat{U}),Ed_2(V,\hat{V})\right)\leq\mathbf{D}$.

We note that the above three bounds agree on both the second condition, i.e., the rate constraints in terms of some mutual information expressions, and the third condition, i.e., the reconstruction functions. However, the first condition in these three bounds constraining the underlying probability distributions $p(x_1,x_2|u,v)$ are different. It is easy to see that the Markov chain condition in the inner bound, i.e., $X_1 \longrightarrow U \longrightarrow V \longrightarrow X_2$, implies the Markov chain conditions in the outer bound in Theorem 2.4.3, i.e., (2.102), while (2.102) implies the Markov chain condition in the outer bound in Theorem 2.4.2, i.e., $X_1 \longrightarrow U \longrightarrow V$ and $U \longrightarrow V \longrightarrow X_2$.

2.4.2 A New Outer Bound

We propose a new outer bound for the multi-terminal rate-distortion region as follows.

Theorem 2.4.4 $\mathcal{R}(\mathbf{D})\subseteq\mathcal{R}_{out,3}(\mathbf{D})$, *where* $\mathcal{R}_{out,3}(\mathbf{D})$ *is the set of all* $\mathbf{R}$ *such that there exist some positive integer* n, *and discrete random variables* Q,X_1,X_2 *for which the following three conditions are satisfied:*

1. *The joint distribution satisfies*

$$p(u,v,x_1,x_2,q)=p(q)p(x_1,x_2|u,v,q)p(u,v) \tag{2.106}$$

where for given $Q=q$

$$p(x_1,x_2|u,v,Q=q)\in\mathcal{S}_{X_1X_2|UV} \tag{2.107}$$

with $\mathcal{S}_{X_1X_2|UV}$ defined in (2.43).

2. *The rate pair satisfies*

$$R_1 \geq I(U,V;X_1|X_2,Q) \tag{2.108}$$

$$R_2 \geq I(U,V;X_2|X_1,Q) \tag{2.109}$$

$$R_1 + R_2 \geq I(U,V;X_1,X_2|Q) \tag{2.110}$$

3. *There exists $\big(\hat{U}(X_1,X_2,Q),\hat{V}(X_1,X_2,Q)\big)$ such that $\big(Ed_1(U,\hat{U}),Ed_2(V,\hat{V})\big) \leq$ $\mathbf{D}$.*

Note that every quantity in this theorem is in the form of a single-letter except the conditional distribution $p(x_1,x_2|u,v,Q=q) \in \mathcal{S}_{X_1X_2|U_1V_1}$, which will be relaxed to a single-letter form in the next section. We also note that the outer bound provided in Theorem 2.4.4 contains a time-sharing random variable Q, which is not needed in the two existing outer bounds given in Theorem 2.4.2 and 2.4.3. We will compare our new outer bound with the existing bounds later.

Proof: We consider an arbitrary block code of two distributed encoders and one joint decoder with reconstructions

$$(\hat{U},\hat{V})^n \triangleq \big((\hat{U},\hat{V})_1,\ldots,(\hat{U},\hat{V})_n\big) = \big(g_1(Y,Z),\ldots,g_n(Y,Z)\big) \tag{2.111}$$

where $Y = f_1(U^n)$ and $Z = f_2(V^n)$, such that the distortions satisfy

$$\begin{aligned}
\big(Ed_1(U^n, \hat{V}^n), Ed_2(V^n, \hat{V}^n)\big) &\triangleq \left(\frac{1}{n}E\sum_{i=1}^{n} d_1(U_i, \hat{U}_i), \frac{1}{n}E\sum_{i=1}^{n} d_2(V_i, \hat{V}_i)\right) \\
&= (\Delta_1, \Delta_2) \\
&< (D_1 + \epsilon, D_2 + \epsilon) \qquad (2.112)
\end{aligned}$$

Here, we define $M_1 = |\mathcal{Y}|$ and $M_2 = |\mathcal{Z}|$, where $\mathcal{Y}$ and $\mathcal{Z}$ are alphabets of Y and Z, respectively.

We define the auxiliary random variables $X_{1i} = (Y, U^{i-1})$ and $X_{2i} = (Z, V^{i-1})$. Then, we have

$$
\begin{aligned}
\log_2 M_1 &\geq H(Y) \\
&= I(U^n, V^n; Y) \\
&\overset{1}{\geq} I(U^n, V^n; Y|Z) \\
&= \sum_{i=1}^{n} I(U_i, V_i; Y|Z, U^{i-1}, V^{i-1}) \\
&= \sum_{i=1}^{n} I(U_i, V_i; Y, Z|U^{i-1}, V^{i-1}) - I(U_i, V_i; Z|U^{i-1}, V^{i-1}) \\
&\overset{2}{=} \sum_{i=1}^{n} I(U_i, V_i; Y, Z|U^{i-1}, V^{i-1}) - I(U_i, V_i; Z|V^{i-1}) \\
&= \sum_{i=1}^{n} I(U_i, V_i; Y, Z, U^{i-1}|V^{i-1}) - I(U_i, V_i; U^{i-1}|V^{i-1}) - I(U_i, V_i; Z|V^{i-1}) \\
&\overset{3}{=} \sum_{i=1}^{n} I(U_i, V_i; Y, Z, U^{i-1}|V^{i-1}) - I(U_i, V_i; Z|V^{i-1}) \\
&= \sum_{i=1}^{n} I(U_i, V_i; Y, U^{i-1}|Z, V^{i-1}) \\
&= \sum_{i=1}^{n} I(U_i, V_i; X_{1i}|X_{2i})
\end{aligned} \tag{2.113}
$$

where

1. follows from the fact that $Y \longrightarrow U^n \longrightarrow V^n \longrightarrow Z$. We observe that the equality holds when Y is independent of Z;

2. follows from the fact that

$$p(z|u_i, v_i, v^{i-1}) = p(z|u_i, v_i, u^{i-1}, v^{i-1}) \tag{2.114}$$

3. follows from the memoryless property of the sources.

Using a symmetrical argument, we obtain

$$\log_2 M_2 \geq \sum_{i=1}^{n} I(U_i, V_i; X_{2i}|X_{1i}) \tag{2.115}$$

Moreover,

$$\begin{aligned}
\log_2 M_1 M_2 \geq & H(Y, Z) \\
= & I(U^n, V^n; Y, Z) \\
= & \sum_{i=1}^{n} H(U_i, V_i) - H(U_i, V_i|Y, Z, U^{i-1}, V^{i-1}) \\
= & \sum_{i=1}^{n} I(U_i, V_i; X_{1i}, X_{2i})
\end{aligned} \tag{2.116}$$

We define the reconstruction function as follows

$$(\hat{U}_i, \hat{V}_i) = g'_i(X_{1i}, X_{2i}) = g'_i\big((Y, U^{i-1}), (Z, V^{i-1})\big) = g_i(Y, Z) \tag{2.117}$$

where g_i is defined in (2.111). Then, the expected distortion is

$$\big(Ed_1(U^n, \hat{V}^n), Ed_2(V^n, \hat{V}^n)\big) = \left(\frac{1}{n}\sum_{i=1}^{n} Ed_1(U_i, \hat{U}_i), \frac{1}{n}\sum_{i=1}^{n} Ed_2(V_i, \hat{V}_i)\right) = (\Delta_1, \Delta_2) \tag{2.118}$$

We note that the three mutual information expressions, i.e., $I(U_i, V_i; X_{1i}|X_{2i})$, $I(U_i, V_i; X_{2i}|X_{1i})$, and $I(U_i, V_i; X_{1i}, X_{2i})$, and the two distortion expressions, i.e., $Ed_1(U_i, \hat{U}_i)$ and $Ed_1(U_2, \hat{U}_2)$, only depend on the marginal conditional distribution

$p(x_{1i}, x_{2i}|u_i, v_i)$ and function g'_i with given $p(u_i, v_i)$. We also note that X_{1i} is a function of U^n and X_{2i} is a function of V^n. Thus $X_{1i} \longrightarrow U^n \longrightarrow V^n \longrightarrow X_{2i}$, and therefore $p(x_{1i}, x_{2i}|u_i, v_i) \in \mathcal{S}_{X_1X_2|U_1V_1}$.

We introduce a time-sharing random variable Q, which is uniformly distributed on $\{1, \ldots, n\}$ and independent of U and V, i.e.,

$$p(u, v, q) = p(u, v)p(q) \tag{2.119}$$

Define random variables X_1 and X_2 be such that

$$p(x_{1i}, x_{2i}|u_i, v_i) = p(x_1, x_2|u, v, Q = i) \tag{2.120}$$

and therefore $p(x_1, x_2|u, v, Q = i) \in \mathcal{S}_{X_1X_2|U_1V_1}$ for all $i = 1, \ldots, n$. Then,

$$\sum_{i=1}^{n} I(U_i, V_i; X_{1i}|X_{2i}) = nI(U_1, V_1; X_1|X_2, Q) \tag{2.121}$$

$$\sum_{i=1}^{n} I(U_i, V_i; X_{2i}|X_{1i}) = nI(U_1, V_1; X_2|X_1, Q) \tag{2.122}$$

$$\sum_{i=1}^{n} I(U_i, V_i; X_{1i}, X_{2i}) = nI(U_1, V_1; X_1, X_2|Q) \tag{2.123}$$

Define a reconstruction function $g(X_1, X_2, Q) = (\hat{U}, \hat{V})$ to be such that

$$g(X_1, X_2, Q = i) = g'_i(X_1, X_2) \tag{2.124}$$

Then,

$$\sum_{i=1}^{n} Ed_1(U_i, \hat{U}_i) = nEd_1(U, \hat{U}) = n\Delta_1 \tag{2.125}$$

$$\sum_{i=1}^{n} Ed_1(V_i, \hat{V}_i) = nEd_1(V, \hat{V}) = n\Delta_2 \tag{2.126}$$

So far we have shown that

$$(R_1 + \delta, R_2 + \delta) = (\frac{1}{n}\log_2 M_1, \frac{1}{n}\log_2 M_2) \in R_{\text{out},3}\big((\Delta_1, \Delta_2)\big) \tag{2.127}$$

We know that $(\Delta_1, \Delta_2) \leq (D_1+\epsilon, D_2+\epsilon)$. Because of the monotonicity of the function $R_{\text{out},3}(\cdot)$, we have

$$(R_1+\delta, R_2+\delta) = (\frac{1}{n}\log_2 M_1, \frac{1}{n}\log_2 M_2) \in R_{\text{out},3}\big((\Delta_1, \Delta_2)\big) \subseteq R_{\text{out},3}\big((D_1+\epsilon, D_2+\epsilon)\big) \tag{2.128}$$

Let $\delta \longrightarrow 0$ and $\epsilon \longrightarrow 0$. Due to the continuity of the function $R_{\text{out},3}(\cdot)$, which will be proven in Appendix 2.D, we have [37, 42]

$$(R_1, R_2) \in R_{\text{out},3}\big((D_1, D_2)\big) \tag{2.129}$$

■

Next, we state and prove that our outer bound given in Theorem 2.4.4 is tighter than $\mathcal{R}_{\text{out},2}(\mathbf{D})$ given in Theorem 2.4.3.

Theorem 2.4.5

$$\mathcal{R}_{out,3}(\mathbf{D}) \subseteq \mathcal{R}_{out,2}(\mathbf{D}) \tag{2.130}$$

Proof: We have two proofs for this theorem. We will provide the first proof here and leave the second proof to Section 2.4.4. We prove this theorem by construction. For every (R_1, R_2) point in $\mathcal{R}_{out,3}(\mathbf{D})$, there exist random variables Q, X_1, X_2 satisfying (2.106), (R_1, R_2) pair satisfying (2.108), (2.109) and (2.110), and a reconstruction pair $\big(\hat{U}(X_1, X_2, Q), \hat{V}(X_1, X_2, Q)\big)$ such that $\big(Ed_1(U, \hat{U}), Ed_2(V, \hat{V})\big) \leq \mathbf{D}$. According to [22], let $X_1' = (X_1, Q)$ and $X_2' = (X_2, Q)$. Then, $p(x_1', x_2'|u, v)$ satisfies the condition (2.102). Moreover,

$$R_1 \geq I(U, V; X_1|X_2, Q) = I(U, V; X_1'|X_2') \tag{2.131}$$

and similarly,

$$R_2 \geq I(U, V; X_2|X_1, Q) = I(U, V; X_2'|X_1') \tag{2.132}$$

and finally,

$$\begin{aligned}
R_1 + R_2 &\geq I(U, V; X_1, X_2|Q) \\
&= H(U, V|Q) - H(U, V|X_1, X_2, Q) \\
&\overset{1}{=} H(U, V) - H(U, V|X_1, X_2, Q) \\
&= H(U, V) - H(U, V|X_1', X_2') \\
&= I(U, V; X_1', X_2')
\end{aligned} \tag{2.133}$$

where 1. follows from the fact that Q is independent of (U, V). $(\hat{U}, \hat{V})$ is a function of (X_1, X_2, Q), and, therefore, it is a function of $(X'_1, X'_2) = \big((X_1, Q), (X_2, Q)\big)$.

Hence, for every rate pair $(R_1, R_2) \in \mathcal{R}_{\text{out},3}(\mathbf{D})$, there exist random variables X'_1, X'_2 such that $p(x'_1, x'_2|u_1, v_1)$ satisfies (2.102), (R_1, R_2) pair satisfies the mutual information constraints, and the reconstruction satisfies the distortion constraints. In other words, $(R_1, R_2) \in \mathcal{R}_{\text{out},2}(\mathbf{D})$, proving the theorem. ■

2.4.3 A New Necessary Condition

From Section 2.2.4, we know that

$$\mathcal{S}_{X_1X_2|UV} \subseteq \mathcal{S}'_{X_1X_2|UV} \tag{2.134}$$

Then, we obtain a single-letter outer bound for the multi-terminal rate-distortion region as follows.

Theorem 2.4.6 $\mathcal{R}(\mathbf{D}) \subseteq \mathcal{R}_{out,4}(\mathbf{D})$, *where* $\mathcal{R}_{out,4}(\mathbf{D})$ *is the set of all* $\mathbf{R}$ *such that there exist discrete random variable* Q *independent of* (U, V), *and discrete random variables* X_1, X_2 *for which the following three conditions are satisfied:*

1. *The joint distribution satisfies,*

$$p(u,v, x_1, x_2, q) = p(q)p(x_1, x_2|u, v, q)p(u, v) \tag{2.135}$$

where for given $Q = q$, $p(x_1, x_2|u, v, Q = q)$ satisfies

$$\lambda_i(\tilde{P}_{X_1X_2|q}) \leq \lambda_2(\tilde{P}_{UV}) \qquad i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|) \tag{2.136}$$

$$\lambda_i(\tilde{P}_{X_1X_2|uq}) \leq \lambda_2(\tilde{P}_{UV}) \qquad i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|) \tag{2.137}$$

$$\lambda_i(\tilde{P}_{X_1X_2|vq}) \leq \lambda_2(\tilde{P}_{UV}) \qquad i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|) \tag{2.138}$$

$$\lambda_i(\tilde{P}_{X_1X_2|uvq}) \leq \lambda_2(\tilde{P}_{UV}) \qquad i = 2, \ldots, \min(|\mathcal{X}_1|, |\mathcal{X}_2|) \tag{2.139}$$

i.e., $p(x_1, x_2|u, v, Q = q) \in \mathcal{S}'_{X_1X_2|UV}$ *for every* q.

2. *The rate pair satisfies*

$$R_1 \geq I(U, V; X_1|X_2, Q) \tag{2.140}$$

$$R_2 \geq I(U, V; X_2|X_1, Q) \tag{2.141}$$

$$R_1 + R_2 \geq I(U, V; X_1, X_2|Q) \tag{2.142}$$

3. *There exists* $\big(\hat{U}(X_1, X_2, Q), \hat{V}(X_1, X_2, Q)\big)$ *such that* $\big(Ed_1(U, \hat{U}), Ed_2(V, \hat{V})\big) \leq \mathbf{D}$.

2.4.4 Comparison of the Bounds

All of the inner and outer bounds we discussed above are in general incomputable due to the lack of bounds on the sizes of the alphabets of the involved auxiliary random variables. Thus, we are not able to compare these bounds numerically. In this section, we will establish some relationships between these bounds by comparing the different feasible sets of the probability distributions involved in these bounds.

We begin with the inner bound. Using the time-sharing argument, a convexification of the inner bound $\mathcal{R}_{\text{in}}(\mathbf{D})$ yields another inner bound $\mathcal{R}'_{\text{in}}(\mathbf{D})$, which is larger than $\mathcal{R}_{\text{in}}(\mathbf{D})$. We define the set

$$\mathcal{S}_{\text{in}} \triangleq \{p(x_1, x_2|u, v) : X_1 \longrightarrow U \longrightarrow V \longrightarrow X_2\} \tag{2.143}$$

Then, this new inner bound may be expressed as a function of $\mathcal{S}_{\text{in}}$ and $\mathbf{D}$ as follows,

$$\mathcal{R}_{\text{in}}(\mathbf{D}) \subseteq \mathcal{R}'_{\text{in}}(\mathbf{D}) = \mathcal{F}(\mathcal{S}_{\text{in}}, \mathbf{D}) \subseteq \mathcal{R}(\mathbf{D}) \tag{2.144}$$

where $\mathcal{F}(\mathcal{S}_{\text{in}}, \mathbf{D})$ is defined as,

$$\mathcal{F}(\mathcal{S}_{\text{in}}, \mathbf{D}) \triangleq \bigcup_{\mathbf{p}\in\mathcal{P}(\mathcal{S}_{\text{in}},\mathbf{D})} \mathcal{C}(\mathbf{p}) \tag{2.145}$$

$$\mathbf{p} \triangleq p(x_1, x_2, q|u, v) = p(x_1, x_2|u, v, Q = q)p(q) \tag{2.146}$$

$$\mathcal{P}(\mathcal{S}_{\text{in}}, \mathbf{D}) \triangleq \left\{ \mathbf{p} : \begin{array}{c} \forall q, p(x_1, x_2|u, v, Q = q) \in \mathcal{S}_{\text{in}}; \\ \exists\big(\hat{U}(X_1, X_2, Q), \hat{V}(X_1, X_2, Q)\big), \\ \text{s.t. } \big(Ed_1(U, \hat{U}), Ed_2(V, \hat{V})\big) \leq \mathbf{D} \end{array} \right\} \tag{2.147}$$

$$\mathcal{C}(\mathbf{p}) \triangleq \left\{ (R_1, R_2) : \begin{array}{r} R_1 \geq I(U, V; X_1|X_2, Q) \\ R_2 \geq I(U, V; X_2|X_1, Q) \\ R_1 + R_2 \geq I(U, V; X_1, X_2|Q) \end{array} \right\} \tag{2.148}$$

In [22], it was shown that $\mathcal{R}_{\text{out},1}(\mathbf{D})$ is convex. Thus, the outer bound $\mathcal{R}_{\text{out},1}(\mathbf{D})$

can be represented in terms of function $\mathcal{F}$ as well, i.e.,

$$\mathcal{R}_{\text{out},1}(\mathbf{D}) = \mathcal{F}(\mathcal{S}_{\text{out},1}, \mathbf{D}) \tag{2.149}$$

where

$$\mathcal{S}_{\text{out},1} \triangleq \{p(x_1, x_2|u, v) : X_1 \longrightarrow U \longrightarrow V \text{ and } U \longrightarrow V \longrightarrow X_2\} \tag{2.150}$$

The result by Wagner and Anatharam [39] can also be expressed by using the function $\mathcal{F}$ as

$$\mathcal{R}_{\text{out},2}(\mathbf{D}) = \mathcal{F}(\mathcal{S}_{\text{out},2}, \mathbf{D}) \tag{2.151}$$

where

$$\mathcal{S}_{\text{out},2} \triangleq \{p(x_1, x_2|u, v) : \exists w, p(x_1, x_2, w|u, v) = p(w)p(x_1|w, u)p(x_2|w, v)\} \tag{2.152}$$

From the definition of the function $\mathcal{F}$, we can see that $\mathcal{F}$ is monotone with respect to the set argument when the distortion argument is fixed, i.e.,

$$\mathcal{F}(A, \mathbf{D}) \subseteq \mathcal{F}(B, \mathbf{D}), \qquad \text{if } A \subseteq B \tag{2.153}$$

Therefore, since

$$\mathcal{S}_{\text{in}} \subseteq \mathcal{S}_{\text{out},2} \subseteq \mathcal{S}_{\text{out},1} \tag{2.154}$$

we have

$$\mathcal{R}_{\text{in}}(\mathbf{D}) = \mathcal{F}(\mathcal{S}_{\text{in}}, \mathcal{D}) \subseteq \mathcal{R}_{\text{out},1}(\mathbf{D}) = \mathcal{F}(\mathcal{S}_{\text{out},1}, \mathcal{D}) \subseteq \mathcal{R}_{\text{out},2}(\mathbf{D}) = \mathcal{F}(\mathcal{S}_{\text{out},2}, \mathcal{D}) \quad (2.155)$$

We conclude that the gap between the inner and the outer bounds comes only from the difference between the feasible sets of the probability distributions $p(x_1, x_2|u, v)$.

In Theorem 2.4.5, we have shown $\mathcal{R}_{out,3}(\mathbf{D}) \subseteq \mathcal{R}_{out,2}(\mathbf{D})$. Here, we provide an alternative proof which comes from the comparison of the feasible sets of probability distributions $p(x_1, x_2|u, v)$. We note that $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$ implies the Markov chain like condition in (2.102) by taking $U_1 = U$, $V_1 = V$ and $(U_2^n, V_2^n) = W$, which means that

$$\mathcal{S}_{\text{out},3} \triangleq \mathcal{S}_{X_1X_2|UV} \subseteq \mathcal{S}_{\text{out},2} \quad (2.156)$$

and because of the monotonicity of $\mathcal{F}(\cdot, \mathbf{D})$ in (2.153), we have

$$\mathcal{F}(\mathcal{S}_{\text{out},3}, \mathbf{D}) = \mathcal{R}_{\text{out},3}(\mathbf{D}) \subseteq \mathcal{R}_{\text{out},2}(\mathbf{D}) = \mathcal{F}(\mathcal{S}_{\text{out},2}, \mathbf{D}) \quad (2.157)$$

From Section 2.2.4, we have that

$$\mathcal{S}_{\text{out},3} \subseteq \mathcal{S}_{\text{out},4} \triangleq \mathcal{S}'_{X_1X_2|UV} \quad (2.158)$$

and therefore

$$\mathcal{R}_{\text{out},3}(\mathbf{D}) = \mathcal{F}(\mathcal{S}_{\text{out},3}, \mathbf{D}) \subseteq \mathcal{R}_{\text{out},4}(\mathbf{D}) = \mathcal{F}(\mathcal{S}_{\text{out},4}, \mathbf{D}) \quad (2.159)$$

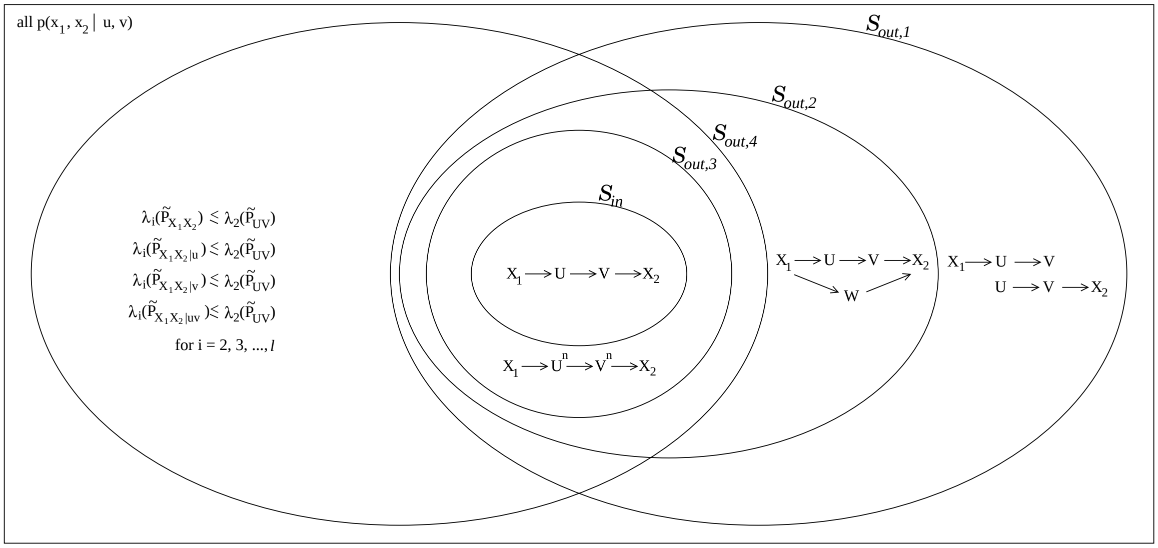

Figure 2.1: Different sets of probability distributions $p(x_1, x_2|u, v)$.

So far, we have not been able to determine whether $\mathcal{S}_{\text{out},4} \subseteq \mathcal{S}_{\text{out},2}$ or $\mathcal{S}_{\text{out},2} \subseteq \mathcal{S}_{\text{out},4}$, however, we know that there exists some probability distribution $p(x_1, x_2|u_1, v_1)$, which belongs to $\mathcal{S}_{\text{out},2}$, but does not belong to $\mathcal{S}_{\text{out},4}$. For example, assume $\lambda_2(\tilde{P}_{UV}) < 1$ and some random variable W independent to (U, V). Let $X_1 = (f_1(U_1), W)$ and $X_2 = (f_2(V_1), W)$. We note that (X_1, X_2, U_1, V_1) satisfies the Markov chain like condition in (2.102), i.e., $p(x_1, x_2|u_1, v_1) \in \mathcal{S}_{\text{out},2}$. But, (X_1, X_2) contains common information W, which means that $\lambda_2(\tilde{P}_{X_1X_2}) = 1 > \lambda_2(\tilde{P}_{UV})$ [41], and therefore, $p(x_1, x_2|u_1, v_1) \notin \mathcal{S}_{\text{out},4}$. Based on this observation, we note that introducing $\mathcal{S}_{\text{out},4}$ helps us rule out some unachievable probability distributions that may exist in $\mathcal{S}_{\text{out},2}$. The relation between different feasible sets of probability distributions $p(x_1, x_2|u_1, v_1)$ is illustrated in Figure 2.1.

Finally, we note that we can obtain a tighter outer bound in terms of the function

$\mathcal{F}(\cdot, \mathbf{D})$ by using a set argument which is the intersection of $\mathcal{S}_{\text{out},2}$ and $\mathcal{S}_{\text{out},4}$, i.e.,

$$\mathcal{R}_{\text{out},2\cap 4}(\mathbf{D}) \triangleq \mathcal{F}(\mathcal{S}_{\text{out},2} \cap \mathcal{S}_{\text{out},4}, \mathbf{D}) \tag{2.160}$$

It is straightforward to see that this outer bound $\mathcal{R}_{\text{out},2\cap 4}(\mathbf{D})$ is in general tighter than the outer bound $\mathcal{F}(\mathcal{S}_{\text{out},2}, \mathbf{D})$.

2.5 Conclusion

In this chapter, we studied the problem of distributed source and channel coding for correlated sources. In the distributed coding on correlated sources, the problem of describing a joint distribution involving an n-letter Markov chain arises. By using a spectral method, we provided a new data processing inequality based on new measures of correlation, which gave us a single-letter necessary condition for the n-letter Markov chain. We applied our results to two specific examples involving distributed coding of correlated sources: the multiple access channel with correlated sources and the multi-terminal rate-distortion region, and proposed two new outer bounds for these two problems.

2.A A Tight Upper Bound for $\lambda_2(\tilde{P}_{X_1U^n})$ over all n

Let $F(n, P_{X_1})$ be the set of all joint distributions for X_1 and U^n with a given marginal distribution for X_1, P_{X_1}. Then, we will show

$$\sup_{F(n,P_{X_1}),\, n=1,2,\ldots} \lambda_2(\tilde{P}_{X_1U^n}) = 1 \tag{2.161}$$

To find $\sup_{F(n,P_{X_1}),\, n=1,2,\ldots} \lambda_2(\tilde{P}_{X_1U^n})$, we need to exhaust the sets $F(n, P_{X_1})$ with $n \geq 1$. In the following, we show that it suffices to check only the asymptotic case.

For any joint distribution $P_{X_1U^n} \in F(n, P_{X_1})$, we attach an independent U, say U_{n+1}, to the existing n-sequence, and get a new joint distribution $P_{X_1U^{n+1}} = P_{X_1U^n} \otimes p_U$, where p_U is the marginal distribution of U in the vector form. By arguments similar to those in Section 2.2.4, we have that $\lambda_i(\tilde{P}_{X_1U^{n+1}}) = \lambda_i(\tilde{P}_{X_1U^n})$. Therefore, for every $P_{X_1U^n} \in F(n, P_{X_1})$, there exists some $P_{X_1U^{n+1}} \in F(n+1, P_{X_1})$, such that $\lambda_i(\tilde{P}_{X_1U^{n+1}}) = \lambda_i(\tilde{P}_{X_1U^n})$. Thus,

$$\sup_{F(n,P_{X_1})} \lambda_2(\tilde{P}_{X_1U^n}) \leq \sup_{F(n+1,P_{X_1})} \lambda_2(\tilde{P}_{X_1U^{n+1}}) \tag{2.162}$$

From (2.162), we see that $\sup_{F(n,P_{X_1})} \lambda_2(\tilde{P}_{X_1U^n})$ is monotonically non-decreasing in n. We also note that $\lambda_2(\tilde{P}_{X_1U^n})$ is upper bounded by 1 for all n, i.e., $\lambda_2(\tilde{P}_{X_1U^n}) \leq 1$. Therefore,

$$\sup_{F(n,P_{X_1}),\, n=1,2,\ldots} \lambda_2(\tilde{P}_{X_1U^n}) = \lim_{n \longrightarrow \infty} \sup_{F(n,P_{X_1})} \lambda_2(\tilde{P}_{X_1U^n}) \tag{2.163}$$

To complete the proof, we need the following lemma.

Lemma 2.A.1 *[41] $\lambda_2(\tilde{P}_{XY}) = 1$ if and only if P_{XY} decomposes. By P_{XY} decomposes, we mean that there exist sets $S_1 \in \mathcal{X}$, $S_2 \in \mathcal{Y}$, such that $P(S_1)$, $P(\mathcal{X} - S_1)$, $P(S_2)$, $P(\mathcal{Y} - S_2)$ are positive, while $P((\mathcal{X} - S_1) \times S_2) = P(S_1 \times (\mathcal{Y} - S_2)) = 0$.*

In the following, we will show by construction that there exists a joint distribution that decomposes asymptotically.

For a given marginal distribution P_{X_1}, we arbitrarily choose a subset S_1 from the alphabet of X_1 with positive $P(S_1)$. We find a set S_2 in the alphabet of U^n such that $P(S_1) = P(S_2)$ if it is possible. Otherwise, we pick S_2 with positive $P(S_2)$ such that $|P(S_1) - P(S_2)|$ is minimized. We denote $\mathcal{L}(n)$ to be the set of all subsets of the alphabet of U^n and we also define $P_{\max} = \max Pr(s)$ for all $s \in \mathcal{U}$. Then, we have

$$\min_{S_2 \subset \mathcal{L}(n)} |P(S_2) - P(S_1)| \leq P_{\max}^n \tag{2.164}$$

We construct a joint distribution for X_1 and U^n as follows. First, we construct the joint distribution P^i corresponding to the case where X_1 and U^n are independent. Second, we rearrange the alphabets of X_1 and U^n and group the sets S_1, $\mathcal{X}_1 - S_1$, S_2 and $\mathcal{U}^n - S_2$ as follows

$$P^i = \begin{bmatrix} P_{11}^i & P_{12}^i \\ P_{21}^i & P_{22}^i \end{bmatrix} \tag{2.165}$$

where P_{11}^i, P_{12}^i, P_{21}^i, P_{22}^i correspond to the sets $S_1 \times S_2$, $S_1 \times (\mathcal{U}^n - S_2)$, $(\mathcal{X}_1 - S_1) \times S_2$, $(\mathcal{X}_1 - S_1) \times (\mathcal{U}^n - S_2)$, respectively. Here, we assume that $P(S_2) \geq P(S_1)$. Then, we scale these four sub-matrices as $P_{11} = \frac{P_{11}^i P(S_1)}{P(S_1)P(S_2)}$, $P_{12} = 0$, $P_{21} = \frac{P_{21}^i (P(S_2) - P(S_1))}{(1-P(S_1))P(S_2)}$,

$P_{22} = \frac{P^i_{22}(1-P(S_2))}{(1-P(S_1))(1-P(S_2))}$, and let

$$P = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \tag{2.166}$$

We note that P is a joint distribution for X_1 and U^n with the given marginal distributions. Next, we move the mass in the sub-matrix P_{21} to P_{11}, which yields

$$P' \triangleq \begin{bmatrix} P'_{11} & 0 \\ 0 & P_{22} \end{bmatrix} = P + E = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} + \begin{bmatrix} E_{11} & 0 \\ -E_{21} & 0 \end{bmatrix} \tag{2.167}$$

where $E_{21} \triangleq P_{21}$, $E_{11} \triangleq \frac{P^i_{11}(P(S_2)-P(S_1))}{P(S_1)P(S_2)}$, and $P'_{11} = \frac{P_{11}P(S_2)}{P(S_1)}$. We denote P'_{X_1} and P'_{U^n} as the marginal distributions of P'. We note that $P'_{U^n} = P_{U^n}$ and $P'_{X_1} = P_{X_1}M$ where M is a scaling diagonal matrix. The elements in the set S_1 are scaled up by a factor of $\frac{P(S_2)}{P(S_1)}$, and those in the set $\mathcal{X}_1 - S_1$ are scaled down by a factor of $\frac{1-P(S_2)}{1-P(S_1)}$. Then,

$$\tilde{P}' = M^{-\frac{1}{2}}\tilde{P} + M^{-\frac{1}{2}}P_{X_1}^{-\frac{1}{2}}EP_{U^n}^{-\frac{1}{2}} \tag{2.168}$$

We will need the following lemmas in the remainder of our derivations. Lemma 2.A.3 can be proved using techniques similar to those in the proof of Lemma 2.A.2 [36].

Lemma 2.A.2 *[36] If $A' = A + E$, then $|\lambda_i(A') - \lambda_i(A)| \leq ||E||_2$, where $||E||_2$ is the spectral norm of E.*

Lemma 2.A.3 *If $A' = MA$, where M is an invertible matrix, then $||M^{-1}||_2^{-1} \leq \lambda_i(A')/\lambda_i(A) \leq ||M||_2$.*

Since P' decomposes, using Lemma 2.A.1, we conclude that $\lambda_2(\tilde{P}') = 1$. We upper bound $||P_{X_1}^{-\frac{1}{2}} E P_{U^n}^{-\frac{1}{2}}||_2$ as follows,

$$||P_{X_1}^{-\frac{1}{2}} E P_{U^n}^{-\frac{1}{2}}||_2 \leq ||P_{X_1}^{-\frac{1}{2}} E P_{U^n}^{-\frac{1}{2}}||_F \tag{2.169}$$

where $||\cdot||_F$ is the Frobenius norm. Combining (2.165) and (2.167), we have

$$||P_{X_1}^{-\frac{1}{2}} E P_{U^n}^{-\frac{1}{2}}||_F \leq \frac{(P(S_2) - P(S_1))}{P_1' P(S_2)} ||P_{X_1}^{-\frac{1}{2}} P^i P_{U^n}^{-\frac{1}{2}}||_F \tag{2.170}$$

where $P_1' \triangleq \min(P(S_1), 1 - P(S_1))$. Since P^i corresponds to the independent case, we have $||P_{X_1}^{-\frac{1}{2}} P^i P_{U^n}^{-\frac{1}{2}}||_F = 1$ from (2.15). Then, from (2.164), (2.169) and (2.170), we obtain

$$||P_{X_1}^{-\frac{1}{2}} E P_{U^n}^{-\frac{1}{2}}||_2 \leq c_1 P_{\max}^n \tag{2.171}$$

where $c_1 \triangleq \frac{1}{P_1' P(S_2)}$.

From Lemma 2.2.2, we have

$$||M^{-\frac{1}{2}} P_{X_1}^{-\frac{1}{2}} E P_{U^n}^{-\frac{1}{2}}||_2 = |\lambda_1(M^{-\frac{1}{2}} P_{X_1}^{-\frac{1}{2}} E P_{U^n}^{-\frac{1}{2}})| \leq \left(\frac{1 - P(S_1)}{1 - P(S_2)}\right)^{\frac{1}{2}} c_1 P_{\max}^n \triangleq c_2 P_{\max}^n \tag{2.172}$$

From Lemma 2.A.2, we have

$$1 - c_2 P_{\max}^n \leq \lambda_2(M^{-\frac{1}{2}} \tilde{P}) \leq 1 + c_2 P_{\max}^n \tag{2.173}$$

We upper bound $||M^{\frac{1}{2}}||_2$ as follows

$$||M^{\frac{1}{2}}||_2 = \sqrt{\frac{P(S_2)}{P(S_1)}} \leq 1 + \sqrt{\frac{P(S_2) - P(S_1)}{P(S_1)}} \leq 1 + \frac{P_{\max}^{n/2}}{\sqrt{P(S_1)}} \triangleq 1 + c_3 P_{\max}^{n/2} \quad (2.174)$$

Similarly, $||M^{-\frac{1}{2}}||_2^{-1} \geq 1 - c_4 P_{\max}^{n/2}$. From Lemma 2.A.3, we have

$$(1 - c_4 P_{\max}^{n/2}) \leq \frac{\lambda_2(\tilde{P})}{\lambda_2(M^{-\frac{1}{2}}\tilde{P})} \leq (1 + c_3 P_{\max}^{n/2}) \quad (2.175)$$

Since P is a joint distribution matrix, from Theorem 2.2.1, we know that $\lambda_2(\tilde{P}) \leq 1$. Therefore, we have

$$(1 - c_4 P_{\max}^{n/2})(1 - c_2 P_{\max}^{n}) \leq \lambda_2(\tilde{P}) \leq 1 \quad (2.176)$$

When $P_{\max} < 1$, corresponding to the non-trivial case, $\lim_{n \longrightarrow \infty} P_{\max}^{n/2} = 0$, and using (2.163), (2.161) follows.

The case $P(S_2) < P(S_1)$ can be proved similarly. ■

2.B An Illustrative Binary Example

In this section, we will study a specific binary example in detail. The aims of this study are, first, to ilustrate the single-letter necessary condition we proposed for the n-letter Markov chain in Section 2.2.3, second, to develop a sharper necessary condition in this specific case, and finally, to compare different necessary conditions and a sufficient condition in this specific example.

The binary example under consideration is as follows. Let U, V, X_1 and X_2 be binary random variables, which take values from $\{0, 1\}$. We assume that (U, V) are a pair of binary symmetric sources, i.e.,

$$Pr(U = 0) = Pr(U = 1) = Pr(V = 0) = Pr(V = 1) = \frac{1}{2} \tag{2.177}$$

From (2.12) and (2.15), we have

$$\tilde{P}_{UV} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} + \lambda_2(\tilde{P}_{UV}) \boldsymbol{\mu}_2(\tilde{P}_{UV}) \boldsymbol{\nu}_2(\tilde{P}_{UV})^T \tag{2.178}$$

Here we focus on the symmetric case, i.e.,

$$\boldsymbol{\mu}_2(\tilde{P}_{UV}) = \boldsymbol{\nu}_2(\tilde{P}_{UV}) = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \tag{2.179}$$

In addition, we assume the following marginal distributions for X_1 and X_2,

$$p_{X_1} = \begin{bmatrix} a^2 \\ 1 - a^2 \end{bmatrix} \tag{2.180}$$

$$p_{X_2} = \begin{bmatrix} b^2 \\ 1 - b^2 \end{bmatrix} \tag{2.181}$$

where $0 \leq a, b \leq 1$. Then, from (2.12) and (2.15), we have

$$\tilde{P}_{X_1X_2} = \begin{bmatrix} a \\ \sqrt{1-a^2} \end{bmatrix} \begin{bmatrix} b & \sqrt{1-b^2} \end{bmatrix} + \lambda_2(\tilde{P}_{X_1X_2})\boldsymbol{\mu}_2(\tilde{P}_{X_1X_2})\boldsymbol{\nu}_2(\tilde{P}_{X_1X_2})^T \quad (2.182)$$

We note that

$$\boldsymbol{\mu}_2(\tilde{P}_{X_1X_2})\boldsymbol{\nu}_2(\tilde{P}_{X_1X_2})^T = \sigma \begin{bmatrix} \sqrt{1-a^2} \\ -a \end{bmatrix} \begin{bmatrix} \sqrt{1-b^2} & -b \end{bmatrix} \quad (2.183)$$

where $\sigma \in \{1, -1\}$. For the simplicity of the derivation in the sequel, we let $\lambda = \sigma\lambda_2(\tilde{P}_{X_1,X_2})$. Then, we have

$$\tilde{P}_{X_1X_2} = \begin{bmatrix} a \\ \sqrt{1-a^2} \end{bmatrix} \begin{bmatrix} b & \sqrt{1-b^2} \end{bmatrix} + \lambda \begin{bmatrix} \sqrt{1-a^2} \\ -a \end{bmatrix} \begin{bmatrix} \sqrt{1-b^2} & -b \end{bmatrix} \quad (2.184)$$

From Theorem 2.2.1, we know that the entries of $\tilde{P}_{X_1X_2}$ are non-negative, i.e.,

$$\tilde{P}_{X_1X_2} = \begin{bmatrix} ab + \lambda\sqrt{(1-a^2)(1-b^2)} & a\sqrt{1-b^2} - \lambda b\sqrt{1-a^2} \\ b\sqrt{1-a^2} - \lambda a\sqrt{1-b^2} & \sqrt{(1-a^2)(1-b^2)} + \lambda ab \end{bmatrix} \geq 0 \quad (2.185)$$

which implies that

$$-\xi_2 \leq \lambda \leq \xi_1 \quad (2.186)$$

where

$$\xi_1 \triangleq \frac{\min(a^2, b^2)\min(1-a^2, 1-b^2)}{ab\sqrt{(1-a^2)(1-b^2)}} \leq 1 \tag{2.187}$$

$$\xi_2 \triangleq \frac{\min(1-a^2, b^2)\min(a^2, 1-b^2)}{ab\sqrt{(1-a^2)(1-b^2)}} \leq 1 \tag{2.188}$$

From Theorem 2.2.3, we have

$$-\lambda_2(\tilde{P}_{UV}) \leq \lambda \leq \lambda_2(\tilde{P}_{UV}) \tag{2.189}$$

Thus, from above, we have

$$-\min(\xi_2, \lambda_2(\tilde{P}_{UV})) \leq \lambda \leq \min(\xi_1, \lambda_2(\tilde{P}_{UV})) \tag{2.190}$$

A sharper bound in this special case can be obtained as follows.

Theorem 2.B.1 *If $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$, and (X_1, X_2, U^n, V^n) satisfies the above settings, then for sufficiently large n,*

$$-\min\left(\xi_2, \lambda_2(\tilde{P}_{UV})\frac{1+\xi_2}{2}\right) \leq \lambda \leq \min\left(\xi_1, \lambda_2(\tilde{P}_{UV})\frac{1+\xi_1}{2}\right) \tag{2.191}$$

The proof of Theorem 2.B.1 is given in Appendix 2.C.

The bound in (2.191) is tighter than the one in (2.190) because $\xi_1 \leq 1$ and therefore $\frac{1+\xi_1}{2} \leq 1$. A similar argument holds for the other side of the inequality as well.

In the above derivation, we provided two necessary conditions for the n-letter Markov chain $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$, where $n \longrightarrow \infty$, in this special case of binary random variables. In other words, we provided two outer bounds for λ, where the joint distributions $p(x_1, x_2, u^n, v^n)$ satisfy the n-letter Markov chain $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$ with $n \longrightarrow \infty$ and satisfy the fixed marginal distributions given in (2.180) and (2.181).

For reference, we give a sufficient condition for $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$, or equivalently, an inner bound for λ satisfying this n-letter Markov chain. This inner bound is obtained by noting that if (X_1, X_2) satisfies $X_1 \longrightarrow U \longrightarrow V \longrightarrow X_2$, then it satisfies $X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2$. In this case, using Theorem 2.2.1 we have

$$\lambda = \lambda_L \lambda_2(\tilde{P}_{UV}) \lambda_R \tag{2.192}$$

where λ_L and λ_R are such that

$$\tilde{P}_{X_1U} \triangleq \begin{bmatrix} a \\ \sqrt{1-a^2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} + \lambda_L \begin{bmatrix} \sqrt{1-a^2} \\ -a \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \geq 0 \tag{2.193}$$

$$\tilde{P}_{VX_2} \triangleq \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} b & \sqrt{1-b^2} \end{bmatrix} + \lambda_R \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{1-b^2} & -b \end{bmatrix} \geq 0 \tag{2.194}$$

Due to the non-negativity of the matrices $\tilde{P}_{X_1U}$ and $\tilde{P}_{VX_2}$, we have

$$-\frac{\min(a^2, 1-a^2)}{a\sqrt{1-a^2}} \le \lambda_L \le \frac{\min(a^2, 1-a^2)}{a\sqrt{1-a^2}} \tag{2.195}$$

$$-\frac{\min(b^2, 1-b^2)}{b\sqrt{1-b^2}} \le \lambda_R \le \frac{\min(b^2, 1-b^2)}{b\sqrt{1-b^2}} \tag{2.196}$$

Thus, we have

$$-\lambda_2(\tilde{P}_{UV})\xi_3 \le \lambda \le \lambda_2(\tilde{P}_{UV})\xi_3 \tag{2.197}$$

where

$$\xi_3 \triangleq \frac{\min(a^2, 1-a^2)\min(b, 1-b^2)}{ab\sqrt{(1-a^2)(1-b^2)}} \tag{2.198}$$

Then, combining (2.190), (2.191), and (2.197), we have the two outer bounds and one inner bound for λ as follows

$$\lambda_2(\tilde{P}_{UV})\xi_3 \le \sup_{X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2} \lambda \le \min(\xi_1, \lambda_2(\tilde{P}_{UV})\frac{1+\xi_1}{2}) \le \min(\xi_1, \lambda_2(\tilde{P}_{UV})) \tag{2.199}$$

$$-\min(\xi_2, \lambda_2(\tilde{P}_{UV})) \le -\min(\xi_2, \lambda_2(\tilde{P}_{UV})\frac{1+\xi_2}{2}) \le \inf_{X_1 \longrightarrow U^n \longrightarrow V^n \longrightarrow X_2} \lambda \le -\lambda_2(\tilde{P}_{UV})\xi_3 \tag{2.200}$$

We illustrate these three bounds with $\lambda_2(\tilde{P}_{UV}) = 0.5$ in Figure 2.2.

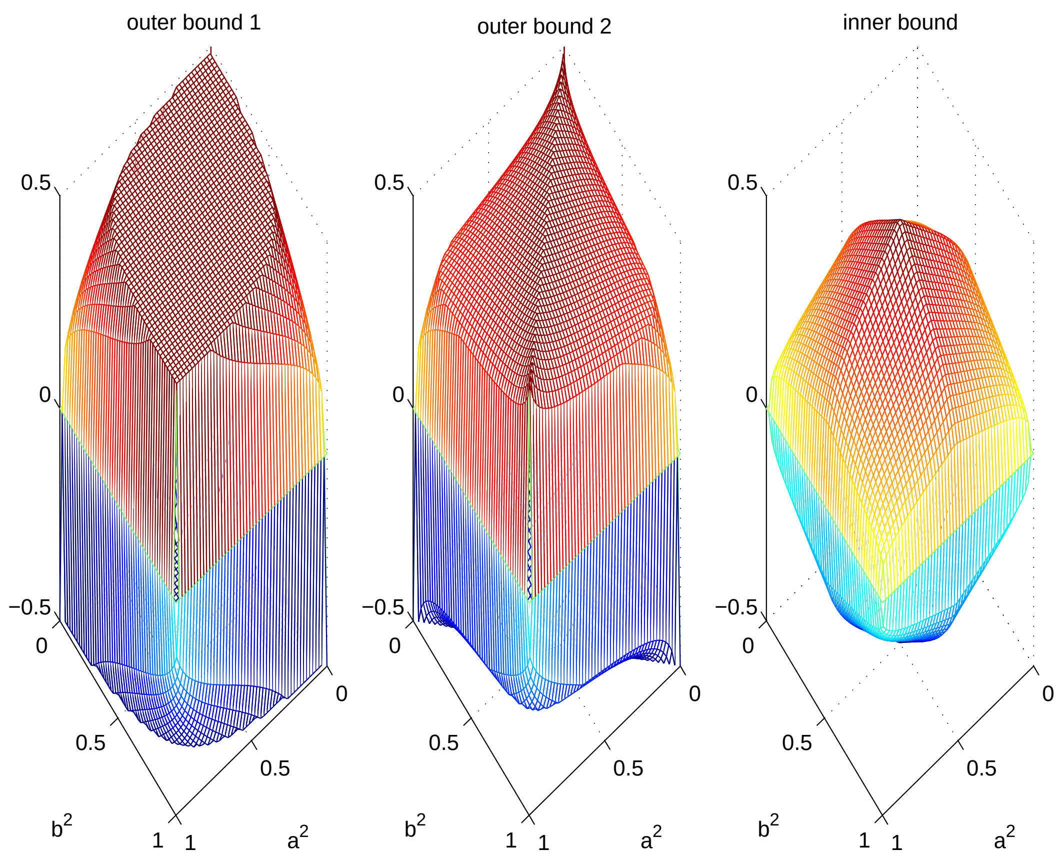

Figure 2.2: (i) Outer bound 1, (ii) outer bound 2, and (iii) inner bound for λ.

2.C Proof of Theorem 2.B.1

From (2.178), we know

$$\tilde{P}_{UV} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} + \lambda_2(\tilde{P}_{UV}) \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \tag{2.201}$$

From (2.29), we know

$$\tilde{P}_{U^nV^n} = \tilde{P}_{UV}^{\otimes n} = \frac{1}{2^n} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} + \sum_{i=2}^{2^n} \lambda_2(\tilde{P}_{UV})^{l_i} \boldsymbol{\mu}_i(\tilde{P}_{U^nV^n}) \boldsymbol{\nu}_i^T(\tilde{P}_{U^nV^n}) \tag{2.202}$$

where $l_i \in \{1, 2, \ldots, n\}$, for $i = 2, \ldots, 2^n$. Due to the symmetric structure of $\tilde{P}_{U^nV^n}$, we have

$$\boldsymbol{\mu}_i(\tilde{P}_{U^nV^n}) = \boldsymbol{\nu}_i(\tilde{P}_{U^nV^n}), \qquad i = 2, \ldots, 2^n \tag{2.203}$$

We also have

$$\tilde{P}_{X_1U^n} = \frac{1}{2^{n/2}} \begin{bmatrix} a \\ \sqrt{1-a^2} \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} + \begin{bmatrix} \sqrt{1-a^2} \\ -a \end{bmatrix} \mathbf{c}^T \tag{2.204}$$

where $\mathbf{c}$ is the product of the second singular value and the second right singular vector of $\tilde{P}_{X_1U^n}$. Similarly,

$$\tilde{P}_{V^nX_2} = \frac{1}{2^{n/2}} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} b & \sqrt{1-b^2} \end{bmatrix} + \mathbf{d} \begin{bmatrix} \sqrt{1-b^2} & -b \end{bmatrix} \tag{2.205}$$

From (2.24), we know that

$$\begin{aligned} \tilde{P}_{X_1X_2} =& \tilde{P}_{X_1U^n}\tilde{P}_{U^nV^n}\tilde{P}_{V^nX_2} \\ =& \begin{bmatrix} a \\ \sqrt{1-a^2} \end{bmatrix} \begin{bmatrix} b & \sqrt{1-b^2} \end{bmatrix} + \\ & \begin{bmatrix} \sqrt{1-a^2} \\ -a \end{bmatrix} \mathbf{c}^T \left(\sum_{i=2}^{2^n} \lambda_2(\tilde{P}_{UV})^{l_i} \boldsymbol{\mu}_i(\tilde{P}_{U^nV^n}) \boldsymbol{\nu}_i(\tilde{P}_{U^nV^n}) \right) \mathbf{d} \begin{bmatrix} \sqrt{1-b^2} & -b \end{bmatrix} \end{aligned} \tag{2.206}$$

Thus, we conclude that,

$$\lambda = \mathbf{c}^T \left(\sum_{i=2}^{2^n} \lambda_2(\tilde{P}_{UV})^{l_i} \boldsymbol{\mu}_i(\tilde{P}_{U^nV^n}) \boldsymbol{\nu}_i^T(\tilde{P}_{U^nV^n}) \right) \mathbf{d} \tag{2.207}$$

Consider the following optimization problem,

$$\max \quad \lambda = \max_{\mathbf{c},\mathbf{d}} \quad \mathbf{c}^T \left(\sum_{i=2}^{2^n} \lambda_2(\tilde{P}_{UV})^{l_i} \boldsymbol{\mu}_i(\tilde{P}_{U^nV^n}) \boldsymbol{\nu}_i^T(\tilde{P}_{U^nV^n}) \right) \mathbf{d} \tag{2.208}$$

We define

$$\gamma_i \triangleq \mathbf{c}^T \boldsymbol{\mu}_i(\tilde{P}_{U^nV^n}), \quad i = 2, \ldots, 2^n \tag{2.209}$$

$$\delta_i \triangleq \mathbf{d}^T \boldsymbol{\nu}_i(\tilde{P}_{U^nV^n}), \quad i = 2, \ldots, 2^n \tag{2.210}$$

Then,

$$\lambda = \sum_{i=2}^{2^n} \lambda_2(\tilde{P}_{UV})^{l_i} \gamma_i \delta_i \tag{2.211}$$

We partition the set $\{2, \ldots, 2^n\}$ into two disjoint subsets, $\mathcal{L}^+$ and $\mathcal{L}^-$, such that

$$i \in \begin{cases} \mathcal{L}^+ & \text{if } \gamma_i \delta_i \geq 0 \\ \mathcal{L}^- & \text{if } \gamma_i \delta_i < 0 \end{cases} \qquad i = 1, \ldots, 2^n \tag{2.212}$$

Hence,

$$\begin{aligned}
\lambda &= \sum_{i \in \mathcal{S}^+} \lambda_2(\tilde{P}_{UV})^{l_i} \gamma_i \delta_i + \sum_{i \in \mathcal{S}^-} \lambda_2(\tilde{P}_{UV})^{l_i} \gamma_i \delta_i \\
&\overset{1}{\leq} \lambda_2(\tilde{P}_{UV}) \sum_{i \in \mathcal{S}^+} \gamma_i \delta_i \\
&\overset{2}{\leq} \frac{\lambda_2(\tilde{P}_{UV})}{4} \sum_{i \in \mathcal{S}^+} (\gamma_i + \delta_i)^2 \\
&\overset{3}{\leq} \frac{\lambda_2(\tilde{P}_{UV})}{4} \sum_{i=2}^{2^n} (\gamma_i + \delta_i)^2 \\
&\overset{4}{=} \frac{\lambda_2(\tilde{P}_{UV})}{4} (\mathbf{c} + \mathbf{d})^T (\mathbf{c} + \mathbf{d}) \\
&= \frac{\lambda_2(\tilde{P}_{UV})}{2} \left(\frac{\mathbf{c}^T \mathbf{c} + \mathbf{d}^T \mathbf{d}}{2} + \mathbf{c}^T \mathbf{d} \right) \\
&\overset{5}{\leq} \frac{\lambda_2(\tilde{P}_{UV})}{2} (1 + \mathbf{c}^T \mathbf{d})
\end{aligned} \tag{2.213}$$

where

1. because of the definition of $\mathcal{L}^+$ and $\mathcal{L}^-$ in (2.212) and $0 \leq \lambda_2(\tilde{P}_{UV}) \leq 1$;

2. because for non-negative $\gamma_i\delta_i$,

$$(\gamma_i - \delta_i)^2 = \gamma_i^2 + \delta_i^2 - 2\gamma_i\delta_i \geq 0 \tag{2.214}$$

   Hence, by adding $4\gamma_i\delta_i$ to both sides of the above inequality, we have

$$(\gamma_i + \delta_i)^2 \geq 4\gamma_i\delta_i \tag{2.215}$$

3. due to the fact that $(\gamma_i + \delta_i)^2$ is non-negative for $i \in \mathcal{L}^-$;

4. comes from the following derivation

$$\begin{aligned}
\sum_{i=2}^{2^n}(\gamma_i + \delta_i)^2 &= \sum_{i=2}^{2^n}\left(\mathbf{c}^T\boldsymbol{\mu}_i(\tilde{P}_{U^nV^n}) + \mathbf{d}^T\boldsymbol{\nu}_i(\tilde{P}_{U^nV^n})\right)^2 \\
&= \sum_{i=2}^{2^n}\left((\mathbf{c}+\mathbf{d})^T\boldsymbol{\mu}_i(\tilde{P}_{U^nV^n})\right)^2 \\
&\overset{(a)}{=} \sum_{i=1}^{2^n}\left((\mathbf{c}+\mathbf{d})^T\boldsymbol{\mu}_i(\tilde{P}_{U^nV^n})\right)^2 \\
&= (\mathbf{c}+\mathbf{d})^T MM^T(\mathbf{c}+\mathbf{d}) \\
&\overset{(b)}{=} (\mathbf{c}+\mathbf{d})^T(\mathbf{c}+\mathbf{d})
\end{aligned} \tag{2.216}$$

   where

   (a) because both the vectors $\mathbf{c}$ and $\mathbf{d}$ are within the subspace spanned by singular vectors $[\boldsymbol{\mu}_2(\tilde{P}_{U^nV^n}), \cdots, \boldsymbol{\mu}_{2^n}(\tilde{P}_{U^nV^n})]$, thus

$$(\mathbf{c}+\mathbf{d})^T \boldsymbol{\mu}_1(\tilde{P}_{U^nV^n}) = 0 \tag{2.217}$$

(b) because

$$MM^T = I \tag{2.218}$$

5. because $\mathbf{c}^T\mathbf{c} = \lambda_2(\tilde{P}_{X_1U^n})^2$ and $\mathbf{d}^T\mathbf{d} = \lambda_2(\tilde{P}_{V^nX_2})^2$ and from Theorem 2.2.1, we know that the square of λ_2 is less than or equal to 1.

From the above discussion, we conclude that

$$\max \quad \lambda \leq \max_{\mathbf{c},\mathbf{d}} \quad \frac{\lambda_2(\tilde{P}_{UV})}{2}(1+\mathbf{c}^T\mathbf{d}) \tag{2.219}$$

Thus, we can upper bound λ by $\max_{\mathbf{c},\mathbf{d}} \frac{\lambda_2(\tilde{P}_{UV})}{2}(1+\mathbf{c}^T\mathbf{d})$.

From (2.12), we know that $\tilde{P}_{X_1U^n}$ is a non-negative matrix, i.e.,

$$\tilde{P}_{X_1U^n} = \begin{bmatrix} \frac{1}{2^{n/2}}a\mathbf{e}^T + \sqrt{1-a^2}\mathbf{c}^T \\ \frac{1}{2^{n/2}}\sqrt{1-a^2}\mathbf{e}^T - a\mathbf{c}^T \end{bmatrix} \geq \mathbf{0} \tag{2.220}$$

where $\mathbf{e}$ is defined as a vector where all its elements are equal to 1, and for matrix A and B, by $\mathbf{A} \geq \mathbf{B}$, we mean all the entries of the matrix $\mathbf{A}-\mathbf{B}$ are non-negative. This property implies that

$$\frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}}\mathbf{e} \geq \bar{\mathbf{c}} \triangleq \frac{1}{2^{n/2}}\frac{a}{\sqrt{1-a^2}}\mathbf{e}+\mathbf{c} \geq \mathbf{0} \tag{2.221}$$

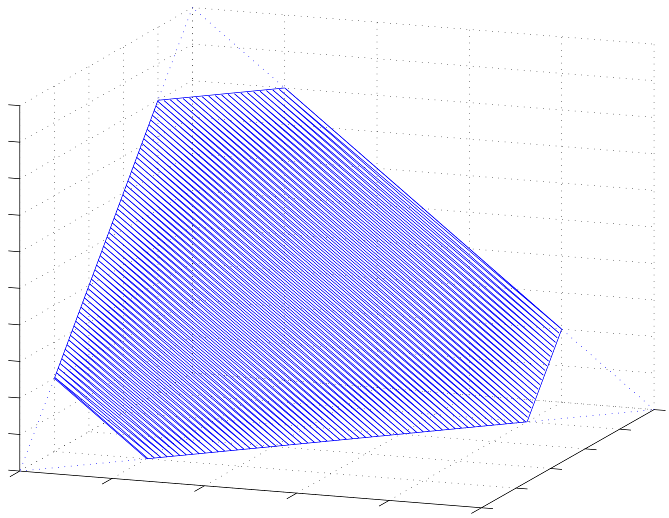

Figure 2.3: Subset of simplex satisfying (2.221).

We know that $\mathbf{c}$ is orthogonal to $\mathbf{e}$, i.e.,

$$\mathbf{c}^T\mathbf{e} = \sum_{i=1}^{2^n} c_i = 0 \tag{2.222}$$

Hence, we see that the vector $\bar{\mathbf{c}}$ is on the hyperplane that contains the point $\frac{1}{2^{n/2}}\frac{a}{\sqrt{1-a^2}}\mathbf{e}$ and is orthogonal to the vector $\mathbf{e}$. On the other hand, (2.221) shows that each coordinate of $\bar{\mathbf{c}}$ is non-negative and less than or equal to $\frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}}$. Thus, the vector $\bar{\mathbf{c}}$ lies on a subset of simplex. See Figure 2.3 for a three-dimension illustration.

By a symmetric argument, we have

$$\frac{1}{2^{n/2}}\frac{1}{b\sqrt{1-b^2}}\mathbf{e} \geq \bar{\mathbf{d}} \triangleq \frac{1}{2^{n/2}}\frac{b}{\sqrt{1-b^2}}\mathbf{e} + \mathbf{d} \geq \mathbf{0} \tag{2.223}$$

Since $\bar{\mathbf{c}} \triangleq \frac{1}{2^{n/2}} \frac{a}{\sqrt{1-a^2}} \mathbf{e} + \mathbf{c}$ and $\bar{\mathbf{d}} \triangleq \frac{1}{2^{n/2}} \frac{b}{\sqrt{1-b^2}} \mathbf{e} + \mathbf{d}$,

$$\begin{aligned}
\bar{\mathbf{c}}^T \bar{\mathbf{d}} &= \left(\frac{1}{2^{n/2}} \frac{a}{\sqrt{1-a^2}} \mathbf{e} + \mathbf{c} \right)^T \left(\frac{1}{2^{n/2}} \frac{b}{\sqrt{1-b^2}} \mathbf{e} + \mathbf{d} \right) \\
&= \frac{ab}{\sqrt{(1-a^2)(1-b^2)}} + \frac{1}{2^{n/2}} \frac{a}{\sqrt{1-a^2}} \mathbf{e}^T \mathbf{d} + \frac{1}{2^{n/2}} \frac{b}{\sqrt{1-b^2}} \mathbf{e}^T \mathbf{c} + \mathbf{c}^T \mathbf{d} \\
&= \frac{ab}{\sqrt{(1-a^2)(1-b^2)}} + \mathbf{c}^T \mathbf{d} \qquad (2.224)
\end{aligned}$$

Then,

$$\max_{\mathbf{c},\mathbf{d}} \quad \mathbf{c}^T \mathbf{d} = \max_{\bar{\mathbf{c}},\bar{\mathbf{d}}} \quad \bar{\mathbf{c}}^T \bar{\mathbf{d}} - \frac{ab}{\sqrt{(1-a^2)(1-b^2)}} \qquad (2.225)$$

The feasible sets of $\bar{\mathbf{c}}$ and $\bar{\mathbf{d}}$ are defined as follows,

$$\mathcal{C} \triangleq \left\{ \mathbf{x} : \frac{1}{2^{n/2}} \frac{1}{a\sqrt{1-a^2}} \mathbf{e} \geq \mathbf{x} \geq \mathbf{0} \quad \text{and} \quad \mathbf{e}^T \mathbf{x} = 2^{n/2} \frac{a}{\sqrt{1-a^2}} \right\} \qquad (2.226)$$

$$\mathcal{D} \triangleq \left\{ \mathbf{x} : \frac{1}{2^{n/2}} \frac{1}{b\sqrt{1-b^2}} \mathbf{e} \geq \mathbf{x} \geq \mathbf{0} \quad \text{and} \quad \mathbf{e}^T \mathbf{x} = 2^{n/2} \frac{b}{\sqrt{1-b^2}} \right\} \qquad (2.227)$$

Consider the following optimization problem

$$\max_{\bar{\mathbf{c}} \in \mathcal{C}, \bar{\mathbf{d}} \in \mathcal{D}} \quad \bar{\mathbf{c}}^T \bar{\mathbf{d}} \qquad (2.228)$$

In the following, we will show that there exist $\mathcal{C}' \subseteq \mathcal{C}$ and $\mathcal{D}' \subseteq \mathcal{D}$ such that

$$\max_{\bar{\mathbf{c}} \in \mathcal{C}, \bar{\mathbf{d}} \in \mathcal{D}} \quad \bar{\mathbf{c}}^T \bar{\mathbf{d}} = \max_{\bar{\mathbf{c}} \in \mathcal{C}', \bar{\mathbf{d}} \in \mathcal{D}'} \quad \bar{\mathbf{c}}^T \bar{\mathbf{d}} \qquad (2.229)$$

If we assume that

$$\max_{\bar{\mathbf{c}}\in\mathcal{C}} \quad \bar{\mathbf{c}}^T\bar{\mathbf{d}} = \max_{\bar{\mathbf{c}}\in\mathcal{C}'} \quad \bar{\mathbf{c}}^T\bar{\mathbf{d}} \qquad \forall\bar{\mathbf{d}}\in\mathcal{D} \tag{2.230}$$

$$\max_{\bar{\mathbf{d}}\in\mathcal{D}} \quad \bar{\mathbf{c}}^T\bar{\mathbf{d}} = \max_{\bar{\mathbf{d}}\in\mathcal{D}'} \quad \bar{\mathbf{c}}^T\bar{\mathbf{d}} \qquad \forall\bar{\mathbf{c}}\in\mathcal{C} \tag{2.231}$$

and we also assume that the set $\mathcal{C}'$ ($\mathcal{D}'$ respectively) does not depend on the value of $\bar{\mathbf{d}}$ ($\bar{\mathbf{c}}$), then we have

$$\begin{aligned}\max_{\bar{\mathbf{c}}\in\mathcal{C},\bar{\mathbf{d}}\in\mathcal{D}} \quad \bar{\mathbf{c}}^T\bar{\mathbf{d}} &= \max_{\bar{\mathbf{c}}\in\mathcal{C}} \quad \max_{\bar{\mathbf{d}}\in\mathcal{D}} \quad \bar{\mathbf{c}}^T\bar{\mathbf{d}} \\ &\overset{1}{=} \max_{\bar{\mathbf{c}}\in\mathcal{C}} \quad \max_{\bar{\mathbf{d}}\in\mathcal{D}'} \quad \bar{\mathbf{c}}^T\bar{\mathbf{d}} \\ &\overset{2}{=} \max_{\bar{\mathbf{d}}\in\mathcal{D}'} \quad \max_{\bar{\mathbf{c}}\in\mathcal{C}} \quad \bar{\mathbf{c}}^T\bar{\mathbf{d}} \\ &\overset{3}{=} \max_{\bar{\mathbf{d}}\in\mathcal{D}'} \quad \max_{\bar{\mathbf{c}}\in\mathcal{C}'} \quad \bar{\mathbf{c}}^T\bar{\mathbf{d}} \\ &= \max_{\bar{\mathbf{c}}\in\mathcal{C}',\bar{\mathbf{d}}\in\mathcal{D}'} \quad \bar{\mathbf{c}}^T\bar{\mathbf{d}}\end{aligned} \tag{2.232}$$

where

1. because of (2.231);
2. because we assume that the set $\mathcal{D}'$ does not depend on the value of $\bar{\mathbf{c}}$;
3. because of (2.230).

Now we need to show our assumptions, (2.230) and (2.231), are valid, for which we need the following lemma.

Lemma 2.C.1 *[7, p. 722] Let $\mathcal{C}$ be a convex subset of $\mathbb{R}^n$, and let $\mathcal{C}^*$ be the set of minima of a concave function $f : \mathcal{C} \longmapsto \mathbb{R}$ over $\mathcal{C}$. If $\mathcal{C}$ is closed and contains at least one extreme point, and $\mathcal{C}^*$ is nonempty, then $\mathcal{C}^*$ contains some extreme point of $\mathcal{C}$.*

Here the extreme point is defined as follows:

Definition 2.C.1 *[7, p. 721] A vector $\mathbf{x}$ is said to be an* extreme point *of a convex set $\mathcal{C}$ if $\mathbf{x}$ belongs to $\mathcal{C}$ and there do not exist vectors $\mathbf{y} \in \mathcal{C}$ and $\mathbf{z} \in \mathcal{C}$, with $\mathbf{y} \neq \mathbf{x}$ and $\mathbf{z} \neq \mathbf{x}$, and a scalar $\alpha \in (0, 1)$ such that $\mathbf{x} = \alpha\mathbf{y} + (1 - \alpha)\mathbf{z}$. An equivalent definition is that $\mathbf{x}$ cannot be expressed as a convex combination of some vectors of $\mathcal{C}$, all of which are different from $\mathbf{x}$.*

Thus, if we assume

$$\mathcal{C}' \triangleq \{\text{extreme points of } \mathcal{C}\} \tag{2.233}$$

$$\mathcal{D}' \triangleq \{\text{extreme points of } \mathcal{D}\} \tag{2.234}$$

(2.230) and (2.231) will be satisfied. We observe that the set $\mathcal{C}'$ (respectively, the set $\mathcal{D}'$), which consists of all the extreme points in the set $\mathcal{C}$ (in the set $\mathcal{D}$), does not depend on the value of $\bar{\mathbf{d}}$ ($\bar{\mathbf{c}}$).

Next, we determine the extreme point set $\mathcal{C}'$ in the following lemma.

Lemma 2.C.2 *The set $\mathcal{C}'$ consists of all the vectors, each of which contains $2^n a^2$ non-zero entries with value $\frac{1}{2^{n/2}} \frac{1}{a\sqrt{1-a^2}}$, when n is sufficiently large.*

Proof: We define the set $\mathcal{C}''$ as the set where each element contains $2^n a^2$ non-zero entries equal to $\frac{1}{2^{n/2}} \frac{1}{a\sqrt{1-a^2}}$. It is easy to see that every vector in $\mathcal{C}''$ is within the set $\mathcal{C}$.

We need to show that any vector in the set $\mathcal{C}$ is a convex combination of some vectors in $\mathcal{C}''$. This can be proven by induction. It is easy to see that, if a vector such that $2^n - 1$ out of 2^n entries take values from $\{0, \frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}}\}$, the last entry will converge to 0, when n goes to infinity. Let $\mathbf{s} \in \mathcal{C}$ such that l out of 2^n entries take values in $(0, \frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}})$. Then, we choose any 2 out of these l entries, which are equal to α and β, respectively. If $\alpha + \beta \leq \frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}}$, then

$$\begin{aligned}
&\begin{bmatrix} \cdots & \alpha & \cdots & \beta & \cdots \end{bmatrix} \\
&\qquad = \frac{\beta}{\alpha+\beta}\begin{bmatrix} \cdots & 0 & \cdots & \alpha+\beta & \cdots \end{bmatrix} + \frac{\alpha}{\alpha+\beta}\begin{bmatrix} \cdots & \alpha+\beta & \cdots & 0 & \cdots \end{bmatrix}
\end{aligned} \tag{2.235}$$

If $\alpha + \beta \geq \frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}}$, then

$$\begin{aligned}
&\begin{bmatrix} \cdots & \alpha & \cdots & \beta & \cdots \end{bmatrix} \\
&\qquad = \frac{\frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}} - \beta}{\frac{2}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}} - \alpha - \beta}\begin{bmatrix} \cdots & \frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}} & \cdots & \alpha+\beta-\frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}} & \cdots \end{bmatrix} \\
&\qquad\quad + \frac{\frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}} - \alpha}{\frac{2}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}} - \alpha - \beta}\begin{bmatrix} \cdots & \alpha+\beta-\frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}} & \cdots & \frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}} & \cdots \end{bmatrix}
\end{aligned} \tag{2.236}$$

which means that $\mathbf{s}$ can be expressed as a convex combination of two vectors. These two vectors belong to set $\mathcal{C}$ and both of them have $l-1$ out of 2^n entries takes value in $(0, \frac{1}{2^{n/2}}\frac{1}{a\sqrt{1-a^2}})$. By induction, we can show that every vector in set $\mathcal{C}$ can be expressed as a convex combination of some vectors in $\mathcal{C}''$. On the other hand, it is easy to see that any vector $\mathbf{s}$ in $\mathcal{C}''$ cannot be expressed as a convex combination of some vectors in the set $\mathcal{C}$ other than $\mathbf{s}$ itself. Thus we conclude that $\mathcal{C}' = \mathcal{C}''$. ■

Similarly, the set $\mathcal{D}'$ consists all the vectors, each of which contains $2^n b^2$ non-zero entries with value $\frac{1}{2^{n/2}}\frac{1}{b\sqrt{1-b^2}}$. Then,

$$\max_{\bar{\mathbf{c}}\in\mathcal{C},\bar{\mathbf{d}}\in\mathcal{D}} \bar{\mathbf{c}}^T\bar{\mathbf{d}} = \max_{\bar{\mathbf{c}}\in\mathcal{C}',\bar{\mathbf{d}}\in\mathcal{D}'} \bar{\mathbf{c}}^T\bar{\mathbf{d}} = \min(a^2,b^2)\frac{1}{a\sqrt{1-a^2}}\frac{1}{b\sqrt{1-b^2}} \tag{2.237}$$

and,

$$\begin{aligned}\max_{\mathbf{c},\mathbf{d}} \mathbf{c}^T\mathbf{d} &= \max_{\bar{\mathbf{c}}\in\mathcal{C},\bar{\mathbf{d}}\in\mathcal{D}} \bar{\mathbf{c}}^T\bar{\mathbf{d}} - \frac{ab}{\sqrt{(1-a^2)(1-b^2)}}\\ &= \min(a^2,b^2)\frac{1}{ab\sqrt{(1-a^2)(1-b^2)}} - \frac{ab}{\sqrt{(1-a^2)(1-b^2)}}\\ &= \min(a^2,b^2)\min(1-a^2,1-b^2)\frac{1}{ab\sqrt{(1-a^2)(1-b^2)}}\end{aligned} \tag{2.238}$$

Hence,

$$\lambda \le \lambda_2(\tilde{P}_{UV})\frac{1+\mathbf{c}^T\mathbf{d}}{2} \le \lambda_2(\tilde{P}_{UV})\frac{1+\frac{\min(a^2,b^2)\min(1-a^2,1-b^2)}{ab\sqrt{(1-a^2)(1-b^2)}}}{2} \tag{2.239}$$

The lower bound of λ can be derived in a similar manner. We rewrite (2.206) in the following form

$$\tilde{P}_{X_1X_2} = \begin{bmatrix} a \\ \sqrt{1-a^2} \end{bmatrix}\begin{bmatrix} b & \sqrt{1-b^2} \end{bmatrix} + (-\lambda)\begin{bmatrix} \sqrt{1-a^2} \\ -a \end{bmatrix}\begin{bmatrix} -\sqrt{1-b^2} & b \end{bmatrix} \tag{2.240}$$

By the same arguments as above, we obtain

$$-\lambda \leq \lambda_2(\tilde{P}_{UV}) \frac{1 + \frac{\min(1-a^2, b^2)\min(a^2, 1-b^2)}{ab\sqrt{(1-a^2)(1-b^2)}}}{2} \tag{2.241}$$

Combining (2.239) and (2.241), we have

$$-\lambda_2(\tilde{P}_{UV}) \frac{1 + \frac{\min(1-a^2, b^2)\min(a^2, 1-b^2)}{ab\sqrt{(1-a^2)(1-b^2)}}}{2} \leq \lambda \leq \lambda_2(\tilde{P}_{UV}) \frac{1 + \frac{\min(a^2, b^2)\min(1-a^2, 1-b^2)}{ab\sqrt{(1-a^2)(1-b^2)}}}{2} \tag{2.242}$$

■

2.D Some Properties of Function $\mathcal{F}$

Function $\mathcal{F}(\cdot, \cdot)$ has two arguments, the probability set argument and the distortion argument. We recall the definition of $\mathcal{F}$ as follows.

$$\mathcal{F}(\mathcal{S}, \mathbf{D}) \triangleq \bigcup_{\mathbf{p} \in \mathcal{P}(\mathcal{S}, \mathbf{D})} \mathcal{C}(\mathbf{p}) \tag{2.243}$$

$$\mathbf{p} \triangleq p(x_1, x_2, q|u, v) = p(x_1, x_2|u, v, Q = q)p(q) \tag{2.244}$$

$$\mathcal{P}(\mathcal{S}, \mathbf{D}) \triangleq \left\{ \mathbf{p} : \begin{array}{c} \forall q, p(x_1, x_2|u, v, Q = q) \in \mathcal{S}_{\text{in}}; \\ \exists \big(\hat{U}(X_1, X_2, Q), \hat{V}(X_1, X_2, Q)\big), \\ \text{s.t. } \big(Ed_1(U, \hat{U}), Ed_2(V, \hat{V})\big) \leq \mathbf{D} \end{array} \right\} \tag{2.245}$$

$$\mathcal{C}(\mathbf{p}) \triangleq \left\{ (R_1, R_2) : \begin{array}{r} R_1 \geq I(U, V; X_1|X_2, Q) \\ R_2 \geq I(U, V; X_2|X_1, Q) \\ R_1 + R_2 \geq I(U, V; X_1, X_2|Q) \end{array} \right\} \tag{2.246}$$

From the definition, we note that for the probability set argument, if $A \subseteq B$, then

$$\mathcal{P}(A, \mathbf{D}) \subseteq \mathcal{P}(B, \mathbf{D}) \tag{2.247}$$

and therefore

$$\mathcal{F}(A, \mathbf{D}) = \bigcup_{\mathbf{p} \in \mathcal{P}(A,\mathbf{D})} \mathcal{C}(\mathbf{p}) \subseteq \mathcal{F}(B, \mathbf{D}) = \bigcup_{\mathbf{p} \in \mathcal{P}(B,\mathbf{D})} \mathcal{C}(\mathbf{p}) \tag{2.248}$$

which means that function $\mathcal{F}$ is monotone in the probability set argument.

Similarly, if $\mathbf{D}_1 \leq \mathbf{D}_2$, then

$$\mathcal{P}(\mathbf{S}, \mathbf{D}_1) \subseteq \mathcal{P}(\mathbf{S}, \mathbf{D}_2) \tag{2.249}$$

and therefore

$$\mathcal{F}(\mathbf{S}, \mathbf{D}_1) = \bigcup_{\mathbf{p} \in \mathcal{P}(\mathbf{S},\mathbf{D}_1)} \mathcal{C}(\mathbf{p}) \subseteq \mathcal{F}(\mathbf{S}, \mathbf{D}_2) = \bigcup_{\mathbf{p} \in \mathcal{P}(\mathbf{S},\mathbf{D}_2)} \mathcal{C}(\mathbf{p}) \tag{2.250}$$

which means that function $\mathcal{F}$ is monotone in the distortion argument.

Consider two distortions $\mathbf{D}_1$ and $\mathbf{D}_2$ such that

$$\mathbf{D} = \lambda \mathbf{D}_1 + (1 - \lambda)\mathbf{D}_2 \tag{2.251}$$

where $0 \leq \lambda \leq 1$. Assume rate pairs $\mathbf{R}_1 \in \mathcal{F}(\mathcal{S}, \mathbf{D}_1)$ and $\mathbf{R}_2 \in \mathcal{F}(\mathcal{S}, \mathbf{D}_2)$. We note that there exists $\mathbf{p}_1 \in \mathcal{P}(\mathbf{S}, \mathbf{D}_1)$ and $\mathbf{p}_2 \in \mathcal{P}(\mathbf{S}, \mathbf{D}_2)$ such that $\mathbf{R}_1 \in \mathcal{C}(\mathbf{p}_1)$

and $\mathbf{R}_2 \in \mathcal{C}(\mathbf{p}_2)$. We define a binary random variable Λ with $Pr(\Lambda = 1) = \lambda$ and $Pr(\Lambda = 2) = 1 - \lambda$ and we define $Q' = (Q, \Lambda)$ and $\mathbf{p} \triangleq p(x_1, x_2, q'|u, v)$, where

$$p(x_1, x_2, q, \Lambda = i|u, v) = p_i(x_1, x_2, q|u, v) \qquad i = 1, 2. \tag{2.252}$$

It is easy to check that $\mathbf{R} \triangleq \lambda \mathbf{R}_1 + (1 - \lambda)\mathbf{R}_2 \in \mathcal{C}(\mathbf{p})$ and $p \in \mathcal{P}(\mathbf{S}, \mathbf{D})$. Thus,

$$\mathbf{R} \in \mathcal{F}(\mathcal{S}, \mathbf{D}) \tag{2.253}$$

i.e., $\mathcal{F}$ is convex in the distortion argument.

By a similar argument, we can show that if $\mathbf{R}_1$ and $\mathbf{R}_2$ are both in the set $\mathcal{F}(\mathcal{S}, \mathbf{D})$, then $\mathbf{R} \triangleq \lambda \mathbf{R}_1 + (1 - \lambda)\mathbf{R}_2 \in \mathcal{F}(\mathcal{S}, \mathbf{D})$, i.e., $\mathcal{F}(\mathcal{S}, \mathbf{D})$ is a convex set.

Finally, we will show the continuity of $\mathcal{F}(\mathcal{S}, \mathbf{D})$. We assume that $\mathcal{S}$ includes the conditional probability corresponding to the deterministic case where $X_1 = U$ and $X_2 = V$. In this case, $\mathcal{F}(\mathcal{S}, \mathbf{0})$ is inner bounded by the Slepian-Wolf region. Due the the monotonicity of $\mathcal{F}(\mathcal{S}, \mathbf{D})$ in $\mathbf{D}$, the boundary of $\mathcal{F}(\mathcal{S}, \mathbf{D})$ for any $\mathbf{D}$ lies outside of the Slepian-Wolf region. We also note that for any point on the boundary of $\mathcal{F}(\mathcal{S}, \mathbf{D})$, the distance between this point and the Slepian-Wolf region is upper bounded by a finite number, say l, where the distance here is the Euclidean distance in two-dimansional space, and therefore, the distance between this point and $\mathcal{F}(\mathcal{S}, (D_1 - a, D_2 - a))$ with $0 < a < \min(D_1, D_2)$ is also upper bounded by l. Because of the convexity of $\mathcal{F}(\mathcal{S}, \mathbf{D})$ in $\mathbf{D}$, the distance between this point and $\mathcal{F}(\mathcal{S}, (D_1 - \epsilon, D_2 - \epsilon))$ with $\epsilon < \alpha$ is upper bounded by $\frac{\epsilon l}{\alpha}$, which proves the continuity of $\mathcal{F}(\mathcal{S}, \mathbf{D})$.

Chapter 3

A New Achievable Scheme for the Relay Channel

3.1 Introduction

As the simplest model for cooperative communications, relay channel has attracted plenty of attention since 1971, when it was first introduced by van der Meulen [38]. In 1979, Cover and El Gamal proposed two major coding schemes for the relay channel [8]. These two schemes are widely known as Decode-And-Forward (DAF) and Compress-And-Forward (CAF) today; see [25] for a recent review. These two coding schemes represent two different types of cooperation. In DAF, the cooperation is relatively obvious, where the relay decodes the message from the transmitter, and the transmitter and the relay cooperatively transmit the constructed common information to the receiver in the next block. In CAF, the cooperation spirit is less easy to recognize, as the message is sent by the transmitter only once. However, the relay cooperates with the transmitter by compressing and sending its signal to the receiver. The rate gains in these achievable schemes are due to the fact that, through the channel from the transmitter to the relay, *correlation* is created between the transmitter and the relay, and this correlation is utilized to improve the rates.

In the DAF scheme, correlation is created and then utilized in a block Markov coding structure. More specifically, a *full* correlation is created by decoding the message fully at the relay, which enables the transmitter and the relay to create any kind of joint distribution for the channel inputs in the next block. The shortcoming of the DAF scheme is that by forcing the relay to decode the message in its entirety, it limits the overall achievable rate by the rate from the transmitter to the relay. In contrast, by not forcing a full decoding at the relay, the CAF scheme does not limit the overall rate by the rate from the transmitter to the relay, and may yield higher overall rates. The shortcoming of the CAF scheme, on the other hand, is that the correlation offered by the block coding structure is not utilized effectively, since in each block the channel inputs X and X_1 from the transmitter and the relay are independent, as the transmitter sends the message only once.

However, the essence of good coding schemes in multi-user systems with correlated sources (e.g., [2, 9]) is to preserve the correlation of the sources in the channel inputs. Motivated by this basic observation, in this chapter, we propose a new coding scheme for the relay channel, that is based on the idea of preserving the correlation in the channel inputs from the transmitter and the relay. We will show that our new coding scheme may be viewed as a more general version of the CAF scheme, and therefore, our new coding scheme may potentially yield larger rates than the CAF scheme. Our proposed scheme can be further combined with the DAF scheme to yield rates that are potentially larger than those offered by both DAF and CAF schemes, similar in spirit to [8, Theorem 7].

Our new achievability scheme for the relay channel may be viewed as a variation of the coding scheme of Ahlswede and Han [2] for the multiple access channel with a correlated helper. In our work, we view the relay as the helper because the receiver does not need to decode the information sent by the relay. Also, we note that the relay is a *correlated helper* as the communication channel from the transmitter to the relay provides relay for free a correlated version of the signal sent by the transmitter. The key aspects of the Ahlswede-Han [2] scheme are: to preserve the correlation between the channel inputs of the transmitter and the helper (relay), and for the receiver to decode a "virtual" source, a compressed version of the helper, but not the entire signal of the helper.

Our new coding scheme is in the form of block Markov coding, as in [8,10,40]. The transmitter uses a superposition Markov code, similar to the one used in the DAF scheme [8], except in the random codebook generation stage, a method similar to the one in [9] is used in order to preserve the correlation between the blocks. Thus, in each block, the fresh information message is mapped into a codeword conditioned on the codeword of the previous block. Therefore, the overall codebook at the transmitter has a tree structure, where the codewords in block l emanate from the codewords in block $l-1$. The depth of the tree is $B-1$. A similar strategy is applied at the relay side where the compressed version of the received signal is mapped into a two-block-long codeword conditioned on the codeword of the previous block. Therefore, the overall codebook at the relay has a tree structure as well. As a result of this coding strategy, we successfully preserve the correlation between the channel inputs of the transmitter and the relay. However, unlike the DAF scheme where a *full* correlation is acquired through decoding at the relay, our scheme provides only a *partially* correlated helper at the relay by not trying to decode the transmitter's signal fully. From [2, 9], we note that the channel inputs are correlated through the virtual sources in our case, and therefore, the channel inputs between the consecutive blocks are correlated. This correlation between the blocks will surely hurt the achievable rate. The correlation between the blocks is the price we pay for preserving the correlation between the channel inputs of the transmitter and the relay within any given block.

At the decoding stage, we perform joint decoding for the entire B blocks after all of the B blocks have been received, which is different compared with the DAF and CAF schemes. The reason for performing joint decoding at the receiver is that due to the correlation between the blocks, decoding at any time before the end of all the B blocks would decrease the achievable rate. We note that joint decoding increases the decoding complexity and the delay as compared to DAF and CAF, though neither of these is a major concern in an information theoretic context. The only problem with the joint decoding strategy is that it makes the analysis difficult as it requires the evaluation of some mutual information expressions involving the joint probability distributions of up to B blocks of codes, where B is very large.

The analysis of the error events provides us three conditions containing mutual information expressions involving infinite letters of the underlying random process. Evaluation of these mutual information expressions is very difficult, if not impossible. To obtain a computable result, we lower bound these mutual informations by noting some Markov structure in the underlying random process. This operation gives us three conditions to be satisfied by the achievable rates. These conditions involve eleven variables, the two channel inputs from the transmitter and the relay, the two channel outputs at the relay and the receiver and the compressed version of the channel output at the relay, in two consecutive blocks, and the channel input from the transmitter in the previous block.

We finish our analysis by revisiting the CAF scheme. We develop an equivalent representation for the achievable rates given in [8] for the CAF scheme. We then show that this equivalent representation for the achievable rates for the CAF scheme is a special case of the achievable rates in our new coding scheme, which is obtained by a special selection of the eleven variables mentioned above. We therefore conclude that our proposed coding scheme yields potentially larger rates than the CAF scheme. More importantly, our new coding scheme creates more possibilities, and therefore a spectrum of new achievable schemes for the relay channel through the selection of the underlying probability distribution, and yields the well-known CAF scheme as a special case, corresponding to a particular selection of the underlying probability distribution.

3.2 The Relay Channel

Consider a relay channel with finite input alphabets $\mathcal{X}$, $\mathcal{X}_1$ and finite output alphabets $\mathcal{Y}$, $\mathcal{Y}_1$, characterized by the transition probability $p(y, y_1|x, x_1)$. An n-length block code for the relay channel $p(y, y_1|x, x_1)$ consists of encoders f, f_i, $i = 1, \ldots, n$ and a decoder g

$$f : \mathcal{M} \longrightarrow \mathcal{X}^n$$

$$f_i : \mathcal{Y}_1^{i-1} \longrightarrow \mathcal{X}_1, \qquad i = 1, \ldots, n$$

$$g : \mathcal{Y}^n \longrightarrow \mathcal{M}$$

where the encoder at the transmitter sends $x^n = f(m)$ into the channel, where $m \in \mathcal{M} \triangleq \{1, 2, \ldots, M\}$; the encoder at the relay at the ith channel instance sends $x_{1i} = f_i(y_1^{i-1})$ into the channel; the decoder outputs $\hat{m} = g(y^n)$. The average probability of error is defined as

$$P_e = \frac{1}{M} \sum_{m \in \mathcal{M}} Pr(\hat{m} \neq m | m \text{ is transmitted}) \tag{3.1}$$

A rate R is achievable for the relay channel $p(y, y_1|x, x_1)$ if for every $0 < \epsilon < 1$, $\eta > 0$, and every sufficiently large n, there exists an n-length block code (f, f_i, g) with $P_e \leq \epsilon$ and $\frac{1}{n} \ln M \geq R - \eta$.

3.3 A New Achievability Scheme for the Relay Channel

We adopt a block Markov coding scheme, similar to the DAF and CAF schemes. We have overall B blocks. In each block, we transmit codewords of length n. We denote the variables in the lth block with a subscript of $[l]$. We denote n-letter codewords transmitted in each block with a superscript of n. Following the standard relay channel literature, we denote the (random) signals transmitted by the transmitter and the relay by X and X_1, the signals received at the receiver and the relay by Y and Y_1, and the compressed version of Y_1 at the relay by $\hat{Y}_1$. The realizations of these random signals will be denoted by lower-case letters. For example, the n-letter signals transmitted by the transmitter and the relay in the lth block will be represented by $x_{[l]}^n$ and $x_{1[l]}^n$.

Consider the following discrete time stationary Markov process $G_{[l]} \triangleq (X, \hat{Y}_1, X_1, y, Y_1)_{[l]}$ for $l = 0, 1, \ldots, B$, with the transition probability distribution

$$p\left((x, \hat{y}_1, x_1, y, y_1)_{[l]} | (x, \hat{y}_1, x_1, y, y_1)_{[l-1]}\right) \\ = p(x_{[l]}|x_{[l-1]}) p(y_{1[l]}, y_{[l]}|x_{[l]}, x_{1[l]}) p(x_{1[l]}|\hat{y}_{1[l-1]}) p(\hat{y}_{1[l]}|y_{1[l]}, x_{1[l]}) \tag{3.2}$$

The codebook generation and the encoding scheme for the lth block, $l = 1, \ldots, B-1$, are as follows.

Random codebook generation: Let $(x_{[l-1]}^n(m_{[l-1]}), x_{1[l-1]}^n, y_{1[l-1]}^n, y_{[l-1]}^n)$ denote the transmitted and the received signals in the $(l-1)$st block, where $m_{[l-1]}$ is the message sent by the transmitter in the $(l-1)$st block. An illustration of the codebook structure is shown in Figure 3.1.

1. For each $x_{[l-1]}^n(m_{[l-1]})$ sequence, generate M sequences, where $x_{[l]}^n(m_{[l]})$, the $m_{[l]}$th sequence, is generated independently according to $\prod_{i=1}^n p(x_{i[l]}|x_{i[l-1]})$. Here, every codeword in the $(l-1)$st block expands into a codebook in the lth block. This expansion is indicated by a directed cone from $x_{[l-1]}^n$ to $x_{[l]}^n$ in Figure 3.1.

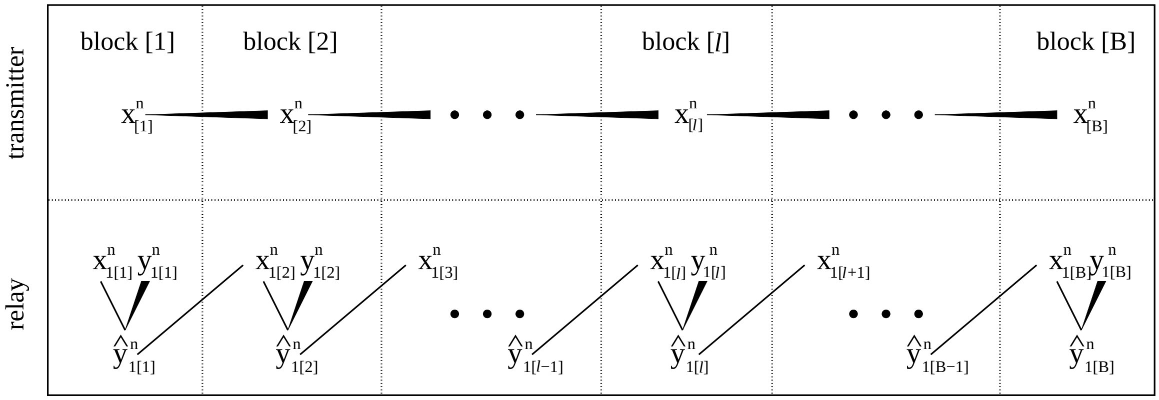

Figure 3.1: Codebook structure.

2. For each $x^n_{1[l-1]}$ sequence, generate L $\hat{Y}^n_{1[l-1]}$ sequences independently uniformly distributed in the conditional strong typical set[1] $\mathcal{T}_\delta(x^n_{1[l-1]})$ with respect to the distribution $p(\hat{y}_{1[l-1]}|x_{1[l-1]})$. If $\frac{1}{n}\ln L > I(Y_{1[l-1]};\hat{Y}_{1[l-1]}|X_{1[l-1]})$, for any given $y^n_{1[l-1]}$ sequence, there exists one $\hat{y}^n_{1[l-1]}$ sequence with high probability when n is sufficiently large such that $(y^n_{1[l-1]}, \hat{y}^n_{1[l-1]}, x^n_{1[l-1]})$ are jointly typical according to the probability distribution $p(y_{1[l-1]}, \hat{y}_{1[l-1]}, x_{1[l-1]})$. Denote this $\hat{y}^n_{1[l-1]}$ as $\hat{y}^n_{1[l-1]}(y^n_{1[l-1]}, x^n_{1[l-1]})$. Here, the quantization from $y^n_{1[l-1]}$ to $\hat{y}^n_{1[l-1]}$, parameterized by $x^n_{1[l-1]}$, is indicated in Figure 3.1 by a directed cone from $y^n_{1[l-1]}$ to $\hat{y}^n_{1[l-1]}$, with a straight line from $x^n_{1[l-1]}$ for the parameterization.

3. For each $\hat{y}^n_{1[l-1]}$, generate one $x^n_{1[l]}$ sequence according to $\prod_{i=1}^n p(x_{1i[l]}|\hat{y}_{1i[l-1]})$. This one-to-one mapping is indicated by a straight line between $\hat{y}^n_{1[l-1]}$ and $x^n_{1[l]}$ in Figure 3.1.

Encoding: Let $m_{[l]}$ be the message to be sent in this block. If we assume that $(x^n_{[l-1]}(m_{[l-1]}), x^n_{1[l-1]})$ are sent and $y^n_{1[l-1]}$ is received in the previous block, we choose

[1]Strong typical set and conditional strong typical set are defined in [13, Definition 1.2.8, 1.2.9]. For the sake of simplicity, we omit the subscript which is used to indicate the underlying distribution in [13].

$(x^n_{[l]}(m_{[l]}),\ \hat{y}^n_{1[l-1]}(y^n_{1[l-1]}, x^n_{1[l-1]}),\ x^n_{1[l]})$ according to the code generation method described above and transmit $(x^n_{[l]}(m_{[l]}),\ x^n_{1[l]})$. In the first block, we assume a virtual 0th block, where $(x^n_{[0]}, x^n_{1[0]}, \hat{y}^n_{1[0]})$, as well as $x^n_{1[1]}$, are known by the transmitter, the relay and the receiver. In the Bth block, the transmitter randomly generates one $x^n_{[B]}$ sequence according to $\prod_{i=1}^n p(x_{i[B]}|x_{i[B-1]})$ and sends it into the channel. The relay, after receiving $y^n_{1[B]}$, randomly generates one $\hat{y}^n_{1[B]}$ sequence according to $\prod_{i=1}^n p(\hat{y}_{1i[B]}|y_{1i[B]}, x_{1i[B]})$. We assume that the transmitter and the relay reliably transmit $x^n_{[B]}$ and $\hat{y}^n_{1[B]}$ to the receiver using the next b blocks, where b is some finite positive integer. We note that $B+b$ blocks are used in our scheme, while only the first $B-1$ blocks carry the message. Thus, the final achievable rate is $\frac{B-1}{B+b}\frac{1}{n}\ln M$ which converges to $\frac{1}{n}\ln M$ for sufficiently large B since b is finite.

Decoding: After receiving B blocks of y^n sequences, i.e., $y^n_{[1]}, \ldots, y^n_{[B]}$, and assuming $x^n_{1[1]}$, $x^n_{[B]}$ and $\hat{y}^n_{1[B]}$ are known at the receiver, we seek $x^n_{[1]}, \ldots, x^n_{[B-1]}$, $\hat{y}^n_{1[1]}, \ldots, \hat{y}^n_{1[B-1]}$, $x^n_{1[2]}, \ldots, x^n_{1[B]}$, such that

$$\left(x^n_{[1]}, \ldots, x^n_{[B]}, \hat{y}^n_{1[1]}, \ldots, \hat{y}^n_{1[B]}, x^n_{1[1]}, \ldots, x^n_{1[B]}, y^n_{[1]}, \ldots, y^n_{[B]}\right) \in \mathcal{T}_\delta$$

according to the stationary distribution of the Markov process $G_{[l]}$ in (3.2).

The differences between our scheme and the CAF scheme are as follows. At the transmitter side, in our scheme, the fresh message $m_{[l]}$ is mapped into the codeword $x^n_{[l]}$ conditioned on the codeword of the previous block $x^n_{[l-1]}$, while in the CAF scheme, $m_{[l]}$ is mapped into $x^n_{[l]}$, which is generated independent of $x^n_{[l-1]}$. At the relay side, in our scheme, the compressed received signal $\hat{y}^n_{1[l-1]}$ is mapped into the codeword

$x^n_{1[l]}$, which is generated according to $p(x_{1[l]}|\hat{y}_{1[l-1]})$, while in the CAF scheme, $x^n_{1[l]}$ is generated independent of $\hat{y}^n_{1[l-1]}$. The aim of our design is to preserve the correlation built in the $(l-1)$st block in the channel inputs of the lth block. At the decoding stage, we perform joint decoding for the entire B blocks after all of the B blocks have been received, while in the CAF scheme, the decoding of the message of the $(l-1)$st block is performed at the end of the lth block.

Probability of error: When n is sufficiently large, the probability of error can be made arbitrarily small when the following conditions are satisfied.

1. For all j such that $1 \leq j \leq B-1$,

$$\begin{aligned}&\frac{1}{n}(B-j)\ln M + (B-j)I(\hat{Y}_{1[l]};Y_{1[l]}|X_{1[l]},X_{[l]})\\&\qquad < I(X^{[B-1]}_{[j]},\hat{Y}^{[B-1]}_{1[j]},X^{[B]}_{1[j+1]};Y^{[B]}_{[j]},\hat{Y}_{1[B]},X_{[B]}|X_{[j-1]},X_{1[j]})\end{aligned} \tag{3.3}$$

2. For all j,k such that $1 \leq j < k \leq B-1$,

$$\begin{aligned}&\frac{1}{n}(B-j)\ln M + (B-k)I(\hat{Y}_{1[l]};Y_{1[l]}|X_{1[l]},X_{[l]})\\&< I(X^{[B-1]}_{[j]},\hat{Y}^{[B-1]}_{1[k]},X^{[B]}_{1[k+1]};Y^{[B]}_{[j]},\hat{Y}_{1[B]},X_{1[B]},\hat{Y}^{[k-1]}_{1[j]},X^{[k]}_{1[j+1]}|X_{[j-1]},X_{[j]})\end{aligned} \tag{3.4}$$

3. For all j,k such that $1 \leq k < j \leq B-1$,

$$\begin{aligned}&(j-k)I(\hat{Y}_{1[l]};Y_{1[l]}|X_{1[l]},X_{[l]}) + \frac{1}{n}(B-j)\ln M + (B-j)I(\hat{Y}_{1[l]};Y_{1[l]}|X_{1[l]},X_{[l]})\\&\qquad < I(X^{[B-1]}_{[j]},\hat{Y}^{[B-1]}_{1[k]},X^{[B]}_{1[k+1]};Y^{[B]}_{[k]},\hat{Y}_{1[B]},X_{[B]}|X^{[j-1]}_{[k]},X_{1[k]})\end{aligned} \tag{3.5}$$

where the subscript $[l]$ on the left hand sides of (3.3), (3.4) and (3.5) indicates that the corresponding random variables belong to a generic sample $g_{[l]}$ of the underlying random process in (3.2). The details of the calculation of the probability of error where these conditions are obtained can be found in Appendix 3.A. The derivation uses standard techniques from information theory, such as counting error events, etc.

In the above conditions, we used the notation $A_{[j]}^{[B]}$ as a shorthand to denote the sequence of random variables $A_{[j]}, A_{[j+1]}, \ldots, A_{[B]}$. Consequently, we note that the mutual informations on the right hand sides of (3.3), (3.4) and (3.5) contain vectors of random variables whose lengths go up to B, where B is very large. In order to simplify the conditions in (3.3), (3.4) and (3.5), we lower bound the mutual information expressions on the right hand sides of (3.3), (3.4) and (3.5) by those that involve random variables that belong to up to three blocks. The detailed derivation of the following lower bounding operation can be found in Appendix 3.B. The derivation uses standard techniques from information theory, such as the chain rule of mutual information, and exploiting the Markov structure of the involved random variables.

1. For all j such that $1 \leq j \leq B-1$,

$$(B-j)\left(\frac{1}{n}\ln M + I(\hat{Y}_{1[l]}; Y_{1[l]}|X_{1[l]}, X_{[l]})\right)$$
$$< (B-j)I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]}|X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \qquad (3.6)$$

2. For all j, k such that $1 \leq j < k \leq B-1$,

$$(k-j)\frac{1}{n}\ln M + (B-k)\left(\frac{1}{n}\ln M + I(\hat{Y}_{1[l]}; Y_{1[l]}|X_{1[l]}, X_{[l]})\right)$$
$$< (k-j)I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]}|X_{1[l]}, Y_{[l-1]}, \hat{Y}_{1[l-1]}, X_{1[l-1]}, X_{[l-2]})$$
$$+ (B-k)I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]}|X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \tag{3.7}$$

3. For all j, k such that $1 \leq k < j \leq B-1$,

$$(j-k)I(\hat{Y}_{1[l]}; Y_{1[l]}|X_{1[l]}, X_{[l]}) + (B-j)\left(\frac{1}{n}\ln M + I(\hat{Y}_{1[l]}; Y_{1[l]}|X_{1[l]}, X_{[l]})\right)$$
$$< (j-k)I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]}|X_{[l]}, X_{[l-1]}, X_{1[l-1]}, Y_{[l-1]})$$
$$+ (B-j)I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]}|X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \tag{3.8}$$

We can further derive sufficient conditions for the above three conditions in (3.6), (3.7) and (3.8) as follows. We define the following quantities:

$$C_1 \triangleq \frac{1}{n}\ln M + I(\hat{Y}_{1[l]}; Y_{1[l]}|X_{1[l]}, X_{[l]}) \tag{3.9}$$
$$C_2 \triangleq \frac{1}{n}\ln M \tag{3.10}$$
$$C_3 \triangleq I(\hat{Y}_{1[l]}; Y_{1[l]}|X_{1[l]}, X_{[l]}) \tag{3.11}$$
$$D_1 \triangleq I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]}|X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \tag{3.12}$$
$$D_2 \triangleq I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]}|X_{1[l]}, Y_{[l-1]}, \hat{Y}_{1[l-1]}, X_{1[l-1]}, X_{[l-2]}) \tag{3.13}$$
$$D_3 \triangleq I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]}|X_{[l]}, X_{[l-1]}, X_{1[l-1]}, Y_{[l-1]}) \tag{3.14}$$

Then, the sufficient conditions in (3.6), (3.7) and (3.8) can also be written as,

1. For all j such that $1 \leq j \leq B-1$,

$$(B-j)C_1 < (B-j)D_1 \tag{3.15}$$

2. For all j, k such that $1 \leq j < k \leq B-1$,

$$(k-j)C_2 + (B-k)C_1 < (k-j)D_2 + (B-k)D_1 \tag{3.16}$$

3. For all j, k such that $1 \leq k < j \leq B-1$,

$$(j-k)C_3 + (B-j)C_1 < (j-k)D_3 + (B-j)D_1 \tag{3.17}$$

We note that the above conditions are implied by the following three conditions,

$$C_1 < D_1 \tag{3.18}$$

$$C_2 < D_2 \tag{3.19}$$

$$C_3 < D_3 \tag{3.20}$$

or in other words, by,

$$R - \eta \leq \frac{1}{n} \ln M < I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{1[l]}, Y_{[l-1]}, \hat{Y}_{1[l-1]}, X_{1[l-1]}, X_{[l-2]}) \tag{3.21}$$

$$I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) < I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[l]}, X_{[l-1]}, X_{1[l-1]}, Y_{[l-1]}) \tag{3.22}$$

$$R - \eta + I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) < I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \tag{3.23}$$

The expressions in (3.21), (3.22) and (3.23) give sufficient conditions to be satisfied by the rate in order for the probability of error to become arbitrarily close to zero. We note that these conditions depend on variables used in three consecutive blocks, l, $l-1$ and $l-2$. With this development, we obtain the main result of this chapter which is stated in the following theorem.

Theorem 3.3.1 *The rate R is achievable for the relay channel, if the following conditions are satisfied*

$$R \leq I(Y, \hat{Y}_1; X | X_1, \tilde{\hat{Y}}_1, \tilde{Y}, \tilde{X}_1, \tilde{\tilde{X}}) \tag{3.24}$$

$$I(\hat{Y}_1; Y_1 | X_1, X) < I(Y; \hat{Y}_1, X_1 | X, \tilde{Y}, \tilde{X}, \tilde{X}_1) \tag{3.25}$$

$$R + I(\hat{Y}_1; Y_1 | X_1, X) \leq I(Y; \hat{Y}_1, X_1, X | \tilde{Y}, \tilde{X}_1, \tilde{\tilde{X}}) \tag{3.26}$$

where

$$\tilde{\tilde{X}} \longrightarrow (\tilde{X}, \tilde{\hat{Y}}_1, \tilde{X}_1, \tilde{Y}, \tilde{Y}_1) \longrightarrow (X, \hat{Y}_1, X_1, Y, Y_1) \tag{3.27}$$

$$p(x, \hat{y}_1, x_1, y, y_1, \tilde{x}) = p(\tilde{x}, \tilde{\hat{y}}_1, \tilde{x}_1, \tilde{y}, \tilde{y}_1, \tilde{\tilde{x}}) \tag{3.28}$$

$$p(x, \hat{y}_1, x_1, y, y_1 | \tilde{x}, \tilde{\hat{y}}_1, \tilde{x}_1, \tilde{y}, \tilde{y}_1) = p(x|\tilde{x}) p(x_1|\tilde{\hat{y}}_1) p(y_1, y|x, x_1) p(\hat{y}_1|y_1, x_1) \tag{3.29}$$

In the above theorem, the notations $\tilde{\ }$ and $\tilde{\tilde{\ }}$ are used to denote the signals belonging to the previous block and the block before the previous block, respectively, with respect to a reference block. Therefore, we see that the achievable rate in the relay channel, using our proposed coding scheme, needs to satisfy three conditions that involve mutual information expressions calculated using eleven variables which satisfy the Markov chain constraint in (3.27), the marginal distribution constraint in (3.28), and the additional inter-block probability distribution constraint in (3.29).

In the next section, we will revisit the well-known CAF scheme proposed in [8]. First, we will develop an equivalent representation for the well-known representation of the achievable rate in the CAF scheme. We will then show that the rates achievable by the CAF scheme can be achieved with our proposed scheme by choosing a certain special structure for the joint probability distribution of the eleven random variables in Theorem 3.3.1 while still satisfying the three conditions in (3.27), (3.28) and (3.29).

3.4 Revisiting the Compress-And-Forward (CAF) Scheme

In [8], the achievable rates for the CAF are characterized as in the following theorem.

Theorem 3.4.1 ([8]) *The rate R is achievable for the relay channel, if the following conditions are satisfied*

$$R \leq I(X;Y,\hat{Y}_1|X_1) \tag{3.30}$$

$$I(Y_1;\hat{Y}_1|X_1,Y) < I(X_1;Y) \tag{3.31}$$

where

$$p(x,x_1,y,y_1,\hat{y}_1) = p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|y_1,x_1) \tag{3.32}$$

In the following theorem, we present three equivalent forms for the rate achievable by the CAF scheme.

Theorem 3.4.2 *The following three conditions are equivalent.*

1. *For some $p(x,x_1,y,y_1,\hat{y}_1) = p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|y_1,x_1)$*

$$R - I(X;\hat{Y}_1|X_1) \leq I(X;Y|\hat{Y}_1,X_1) \tag{3.33}$$

$$I(Y_1;\hat{Y}_1|X_1) < I(\hat{Y}_1;Y|X_1) + I(X_1;Y) \tag{3.34}$$

2. *For some $p(x,x_1,y,y_1,\hat{y}_1) = p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|y_1,x_1)$*

$$R - I(X;\hat{Y}_1|X_1) \leq I(X;Y|\hat{Y}_1,X_1) \tag{3.35}$$

$$R - I(X;\hat{Y}_1|X_1) + I(Y_1;\hat{Y}_1|X_1) \leq I(X,\hat{Y}_1;Y|X_1) + I(X_1;Y) \tag{3.36}$$

3. For some $p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|y_1, x_1)$

$$R - I(X; \hat{Y}_1|X_1) \leq I(X; Y|\hat{Y}_1, X_1) \tag{3.37}$$

$$I(\hat{Y}_1; Y_1|X_1, X) < I(\hat{Y}_1; Y|X_1, X) + I(X_1; Y|X) \tag{3.38}$$

$$R - I(X; \hat{Y}_1|X_1) + I(Y_1; \hat{Y}_1|X_1) \leq I(X, \hat{Y}_1; Y|X_1) + I(X_1; Y) \tag{3.39}$$

The proof of the above theorem is given in Appendix 3.C.

We rewrite the final equivalent representation in (3.37), (3.38) and (3.39) in the following more compact form in order to compare the rates achievable with our proposed scheme and the rates achievable with the CAF scheme in the next section.

$$R \leq I(X; Y, \hat{Y}_1|X_1) \tag{3.40}$$

$$I(\hat{Y}_1; Y_1|X_1, X) < I(\hat{Y}_1, X_1; Y|X) \tag{3.41}$$

$$R + I(Y_1; \hat{Y}_1|X_1, X) \leq I(X, \hat{Y}_1, X_1; Y) \tag{3.42}$$

3.5 Comparison of the Achievable Rates with Our Scheme and with the CAF Scheme

We note that the conditions on the achievable rates with our scheme given in Theorem 3.3.1, i.e., (3.24), (3.25), (3.26), are very similar to the final equivalent form for the conditions on the achievable rates with the CAF scheme, i.e., (3.40), (3.41), (3.42), except for two differences. First, the channel inputs of the transmitter and the relay, i.e., X and X_1, in our proposed scheme can be correlated, while in the CAF scheme they are independent, and second, in our scheme there are some extra random variables, which mutual information expressions are conditioned on, e.g., $\tilde{X}, \tilde{X}_1, \tilde{Y}, \tilde{\hat{Y}}_1, \tilde{\tilde{X}}$. These two differences come from our coding scheme where we introduced correlation between the channel inputs of the transmitter and the relay in a block, and between the variables across the blocks. The correlation between the channel inputs from the transmitter and the relay in any block is an advantage, as for channels which favor correlation, this translates into higher rates. However, the correlation across the blocks is a disadvantage as it decreases the efficiency of transmission, and therefore the achievable rates. In fact, the price we pay for the correlation between the channel inputs in any given block is precisely the correlation we have created across the blocks. For a given correlation structure, it is not clear which of these two opposite effects will overcome the other. That is, the rate of our scheme for a certain correlated distribution may be lower or higher than the rate of the CAF scheme. However, we note that the CAF scheme can be viewed as a special case of our proposed scheme by choosing an independent distribution, i.e., by choosing the following conditional distribution in (3.29)

$$p(x, \hat{y}_1, x_1, y, y_1|\tilde{x}, \tilde{\hat{y}}_1, \tilde{x}_1, \tilde{y}, \tilde{y}_1) = p(x)p(x_1)p(y_1, y|x, x_1)p(\hat{y}_1|x_1, y_1) \tag{3.43}$$

In this case, the expressions in Theorem 3.3.1, i.e., (3.24), (3.25), (3.26), degenerate into the third equivalent form for the CAF scheme in Theorem 3.4.2, i.e., (3.40), (3.41), (3.42). The above observation implies that the maximum achievable rate with our proposed scheme over all possible distributions is not less than the achievable rate of the CAF scheme. Thus, we can claim that this chapter offers more choices in the achievability scheme than the CAF scheme, and that these choices may potentially yield larger achievable rates than those offered by the CAF scheme.

3.6 Conclusion

In this chapter, we proposed a new achievability scheme for the general relay channel. This coding scheme is in the form of a block Markov code. The transmitter uses a superposition Markov code. The relay compresses the received signal and maps the compressed version of the received signal into a codeword conditioned on the codeword of the previous block. The receiver performs joint decoding after it has received all of the B blocks. We showed that this coding scheme can be viewed as a generalization of the well-known CAF scheme proposed by Cover and El Gamal. Our coding scheme provides options for preserving the correlation between the channel inputs of the transmitter and the relay, which is not possible in the CAF scheme. Thus, our proposed scheme may potentially yield a larger achievable rate than the CAF scheme.

3.A Appendix: Probability of Error Calculation

The average probability of decoding error can be expressed as follows,

$$P_e = Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2 \cap E_1^c) \tag{3.44}$$

where

$$E_1 \triangleq \left(x^n_{[1,\dots,B]}, \hat{y}^n_{1[1,\dots,B]}, x^n_{1[1,\dots,B]}, y^n_{[1,\dots,B]}\right) \notin \mathcal{T}_\delta \tag{3.45}$$

$$E_2 \triangleq \bigcup_{\left(\bar{x}^n_{[1,\dots,B]}, \bar{\hat{y}}^n_{1[1,\dots,B-1]}\right) \neq \left(x^n_{[1,\dots,B]}, \hat{y}^n_{1[1,\dots,B-1]}\right)} \left(\bar{x}^n_{[1,\dots,B]}, \bar{\hat{y}}^n_{1[1,\dots,B]}, \bar{x}^n_{1[1,\dots,B]}, y^n_{[1,\dots,B]}\right) \in \mathcal{T}_\delta \tag{3.46}$$

where $(\bar{x}^n_{[1,\dots,B]}, \bar{\hat{y}}^n_{1[1,\dots,B-1]}, \bar{x}^n_{1[2,\dots,B]})$ is another codeword that is generated according to the rules of our scheme.

From (3.2), we note the following Markov properties:

1. conditioned on $(\hat{Y}_{1[l]}, X_{[l]}, X_{1[l]})$, $Y_{[l]}$ is independent of $G_{[\dots,l-1]}$ and $G_{[l,\dots]}$;

2. conditioned on $(X_{[l-1]}, \hat{Y}_{1[l-1]})$, $G_{[l,\dots]}$ is independent of $G_{[\dots,l-1]}$.

Here, and in the sequel, subscript $[l]$ refers to a generic block within overall B blocks.

$Pr(E_1)$ can be upper bounded as follows:

$$\begin{aligned} & Pr(E_1) \leq \\ & \sum_{l=1}^{B} \left(Pr\left((x^n_{[l]}, x^n_{1[l]}, y^n_{[l]}, y^n_{1[l]}, g^n_{[\dots,l-1]}) \notin \mathcal{T}_\delta | g^n_{[\dots,l-1]} \in \mathcal{T}_\delta\right)\right. \\ & \left. + \, Pr\left((\hat{y}^n_{1[l]}, x^n_{[l]}, x^n_{1[l]}, y^n_{[l]}, y^n_{1[l]}, g^n_{[\dots,l-1]}) \notin \mathcal{T}_\delta | (x^n_{[l]}, x^n_{1[l]}, y^n_{[l]}, y^n_{1[l]}, g^n_{[\dots,l-1]}) \in \mathcal{T}_\delta\right)\right) \end{aligned} \tag{3.47}$$

From the way the code is generated, we have

$$Pr\left((x^n_{[l]}, x^n_{1[l]}, y^n_{[l]}, y^n_{1[l]}, g^n_{[\dots,l-1]}) \notin \mathcal{T}_\delta | g^n_{[\dots,l-1]} \in \mathcal{T}_\delta\right) \leq \epsilon \tag{3.48}$$

The compression from $y_{1[l]}^n$ to $\hat{y}_{1[l]}^n$ is a conditional version of a rate-distortion code. If $R' > I(Y_1; \hat{Y}_1 | X_1)$, then, when n is sufficiently large, we have

$$Pr\left((\hat{y}_{1[l]}^n, x_{[l]}^n, x_{1[l]}^n, y_{[l]}^n, y_{1[l]}^n, g_{[\dots,l-1]}^n) \notin \mathcal{T}_\delta | (x_{[l]}^n, x_{1[l]}^n, y_{[l]}^n, y_{1[l]}^n, g_{[\dots,l-1]}^n) \in \mathcal{T}_\delta\right) \leq \epsilon \quad (3.49)$$

Thus,

$$Pr(E_1) \leq 2B\epsilon \quad (3.50)$$

Now we switch to the error event E_2.

$$\begin{aligned}
&Pr(E_2 \cap E_1^c) \\
&= \sum_{\left(x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B]}^n, x_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n\right) \in \mathcal{T}_\delta} p(x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B]}^n, x_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n) \\
&\qquad \times Pr\left(E_2 | (x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B]}^n, x_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n) \text{ sent}\right) \\
&\leq \max_{\left(x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B]}^n, x_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n\right) \in \mathcal{T}_\delta} Pr\left(E_2 | (x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B]}^n, x_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n) \text{ sent}\right)
\end{aligned} \quad (3.51)$$

From our proposed coding scheme, we note that the codebooks at both transmitter and relay have tree structures with $B-1$ stages. A correct codeword $x_{[1,\dots,B-1]}^n$ can be viewed as a path in the tree-structured codebook at the transmitter. Similarly, for the codeword $\hat{y}_{1[1,\dots,B-1]}^n$ at the relay. An error occurs when we diverge from the correct path at a certain stage in the tree. Thus, the error event E_2 can be decomposed as

$$E_2 = \bigcup_{\substack{j=2,\ldots,B-1 \\ k=2,\ldots,B-1}} \bigcup_{\substack{\left(\bar{x}^n_{[1]},\ldots,\bar{x}^n_{[j-1]},\bar{\hat{y}}^n_{1[1]},\ldots,\bar{\hat{y}}^n_{1[k-1]}\right)=\left(x^n_{[1]},\ldots,x^n_{[j-1]},\hat{y}^n_{1[1]},\ldots,\hat{y}^n_{1[k-1]}\right) \\ \left(\bar{x}^n_{[j]},\bar{\hat{y}}^n_{1[k]}\right)\neq\left(x^n_{[j]},\hat{y}^n_{1[k]}\right)}}$$

$$\left(\bar{x}^n_{[1]},\ldots,\bar{x}^n_{[B]},\bar{\hat{y}}^n_{1[1]},\ldots,\bar{\hat{y}}^n_{1[B]},\bar{x}^n_{1[1]},\ldots,\bar{x}^n_{1[B]},y^n_{[1]},\ldots,y^n_{[B]}\right) \in \mathcal{T}_\delta \tag{3.52}$$

where each term in the union in the above equation represents the error event that results when we diverge from the correct paths at the jth stage at the transmitter and at the kth stage at the relay.

Let us define $\mathcal{F}_1$ to be the set consisting of all feasible codeword pairs $(x^n_{[j]}, \hat{y}^n_{1[j]})$ for the jth block for a given $x^n_{[j-1]}$ and $x^n_{1[j]}$. Then, we have

$$\begin{aligned}
F_1 \triangleq |\mathcal{F}_1| &\leq M \exp(n(H(\hat{Y}_{1[j]}|X_{[j]}, X_{1[j]}) + 2\epsilon)) \frac{L}{(1-\epsilon)\exp(n(H(\hat{Y}_{1[j]}|X_{1[j]}) - 2\epsilon))} \\
&\leq M \exp(n(H(\hat{Y}_{1[j]}|X_{[j]}, X_{1[j]}) + 2\epsilon)) \frac{\exp(n(I(\hat{Y}_{1[j]};Y_{1[j]}|X_{1[j]}) + \epsilon))}{(1-\epsilon)\exp(n(H(\hat{Y}_{1[j]}|X_{1[j]}) - 2\epsilon))} \\
&\leq M \exp(n(I(\hat{Y}_{1[j]};Y_{1[j]}|X_{1[j]}, X_{[j]}) + 6\epsilon))
\end{aligned} \tag{3.53}$$

We also define $\mathcal{F}_2$ to be the set consisting of all feasible codewords $x^n_{[j]}$ for the jth block for a given $x^n_{[j-1]}$. Then,

$$F_2 \triangleq |\mathcal{F}_2| = M \tag{3.54}$$

Similarly, we define $\mathcal{F}_3$ to be the set consisting of all feasible codewords $\hat{y}^n_{1[j]}$ for the

jth block for a given $x^n_{[j]}$ and $x^n_{1[j]}$. Then,

$$\begin{aligned} F_3 \triangleq |\mathcal{F}_3| &\leq L \frac{\exp(n(H(\hat{Y}_{1[j]}|X_{1[j]}, X_{[j]}) + 2\epsilon))}{(1-\epsilon)\exp(n(H(\hat{Y}_{1[j]}|X_{1[j]}) - 2\epsilon))} \\ &\leq \exp(n(I(\hat{Y}_{1[j]}; Y_{1[j]}|X_{1[j]}, X_{[j]}) + 6\epsilon)) \end{aligned} \tag{3.55}$$

We define the error event E_{2jk}

$$E_{2jk} \triangleq \bigcup_{\substack{\left(\bar{x}^n_{[1]},\ldots,\bar{x}^n_{[j-1]},\bar{\hat{y}}^n_{1[1]},\ldots,\bar{\hat{y}}^n_{1[k-1]}\right)=\left(x^n_{[1]},\ldots,x^n_{[j-1]},\hat{y}^n_{1[1]},\ldots,\hat{y}^n_{1[k-1]}\right) \\ \left(\bar{x}^n_{[j]},\bar{\hat{y}}^n_{1[k]}\right)\neq\left(x^n_{[j]},\hat{y}^n_{1[k]}\right)}} \left(\bar{x}^n_{[1]},\ldots,\bar{x}^n_{[B]},\bar{\hat{y}}^n_{1[1]},\ldots,\bar{\hat{y}}^n_{1[B]},\bar{x}^n_{1[1]},\ldots,\bar{x}^n_{1[B]},y^n_{[1]},\ldots,y^n_{[B]}\right) \in \mathcal{T}_\delta \tag{3.56}$$

Then, we have

$$Pr(E_2 \cap E_1^c) \leq \sum_{j=2}^{B-1}\sum_{k=2}^{B-1} Pr(E_{2jk} \cap E_1^c) \tag{3.57}$$

and

$$Pr(E_{2jk} \cap E_1^c) \leq |\mathcal{A}_{jk}| \max_{(\bar{x}^n_{[1]},\ldots,\bar{x}^n_{[B-1]},\bar{\hat{y}}^n_{1[1]},\ldots,\bar{\hat{y}}^n_{1[B-1]})\in\mathcal{A}_{jk}} P_1(\bar{x}^n_{[1]},\ldots,\bar{x}^n_{[B-1]},\bar{\hat{y}}^n_{1[1]},\ldots,\bar{\hat{y}}^n_{1[B-1]}) \tag{3.58}$$

where

$$\mathcal{A}_{jk} \triangleq \left\{ \begin{array}{rl} \text{codeword } (\bar{x}^n_{[1]}, \ldots, \bar{x}^n_{[B-1]}, \bar{\hat{y}}^n_{1[1]}, \ldots, & \bar{\hat{y}}^n_{1[B-1]}) : \\ \left(\bar{x}^n_{[1]}, \ldots, \bar{x}^n_{[j-1]}, \bar{\hat{y}}^n_{1[1]}, \ldots, \bar{\hat{y}}^n_{1[k-1]}\right) & = \left(x^n_{[1]}, \ldots, x^n_{[j-1]}, \hat{y}^n_{1[1]}, \ldots, \hat{y}^n_{1[k-1]}\right) \\ \left(\bar{x}^n_{[j]}, \bar{\hat{y}}^n_{1[k]}\right) & \neq \left(x^n_{[j]}, \hat{y}^n_{1[k]}\right) \end{array} \right\} \tag{3.59}$$

$$\begin{aligned} &P_1(\bar{x}^n_{[1]}, \ldots, \bar{x}^n_{[B-1]}, \bar{\hat{y}}^n_{1[1]}, \ldots, \bar{\hat{y}}^n_{1[B-1]}) \\ &\quad \triangleq Pr((\bar{x}^n_{[1]}, \ldots, \bar{x}^n_{[B]}, \bar{\hat{y}}^n_{1[1]}, \ldots, \bar{\hat{y}}^n_{1[B]}, \bar{x}^n_{1[1]}, \ldots, \bar{x}^n_{1[B]}, y^n_{[1]}, \ldots, y^n_{[B]}) \in \mathcal{T}_\delta) \end{aligned} \tag{3.60}$$

given $\left(x^n_{[1]}, \ldots, x^n_{[B]}, \hat{y}^n_{1[1]}, \ldots, \hat{y}^n_{1[B]}, x^n_{1[1]}, \ldots, x^n_{1[B]}, y^n_{[1]}, \ldots, y^n_{[B]}\right) \in \mathcal{T}_\delta$.

In order to have the probability of such error events go to zero, we need the following conditions to hold.

When $j = k$, from the structure of the block Markov code and (3.53), we have

$$|\mathcal{A}_{jk}| = F_1^{B-j} \leq M^{B-j} \exp(n(B-j)(I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) + 6\epsilon)) \tag{3.61}$$

and

$$
\begin{aligned}
&P_1(\bar{x}^n_{[1]}, \ldots, \bar{x}^n_{[B-1]}, \bar{\hat{y}}^n_{1[1]}, \ldots, \bar{\hat{y}}^n_{1[B-1]}) \\
&\quad \leq \exp(n(H(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[j]}, X^{[B]}_{1[j+1]} | Y^{[B]}_{[j]}, \hat{Y}_{1[B]}, X_{[B]}, X_{[j-1]}, X_{1[j]}) + 2\epsilon)) \\
&\qquad \times \exp(-n(H(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[j]}, X^{[B]}_{1[j+1]} | X_{[j-1]}, X_{1[j]}) - 2\epsilon)) \\
&\quad = \exp(n(-I(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[j]}, X^{[B]}_{1[j+1]}; Y^{[B]}_{[j]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, X_{1[j]}) + 4\epsilon))
\end{aligned} \tag{3.62}
$$

When $j < k$, we have

$$
|\mathcal{A}_{jk}| = F_2^{k-j} F_1^{B-k} \leq M^{B-j} \exp(n(B-k)(I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) + 6\epsilon)) \tag{3.63}
$$

and

$$
\begin{aligned}
&P_1(\bar{x}^n_{[1]}, \ldots, \bar{x}^n_{[B-1]}, \bar{\hat{y}}^n_{1[1]}, \ldots, \bar{\hat{y}}^n_{1[B-1]}) \\
&\leq \exp(n(H(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[k]}, X^{[B]}_{1[k+1]} | Y^{[B]}_{[j]}, \hat{Y}_{1[B]}, X_{[B]}, \hat{Y}^{[k-1]}_{1[j]}, X_{[j-1]}, X^{[k]}_{1[j]}) + 2\epsilon)) \\
&\quad \times \exp(-n(H(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[k]}, X^{[B]}_{1[k+1]} | X_{[j-1]}, X_{1[j]}) - 2\epsilon)) \\
&= \exp(n(-I(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[k]}, X^{[B]}_{1[k+1]}; Y^{[B]}_{[j]}, \hat{Y}_{1[B]}, X_{[B]}, \hat{Y}^{[k-1]}_{1[j]}, X^{[k]}_{1[j+1]} | X_{[j-1]}, X_{1[j]}) + 4\epsilon))
\end{aligned} \tag{3.64}
$$

When $j > k$, we have

$$
\begin{aligned}
|\mathcal{A}_{jk}| = F_3^{j-k} F_1^{B-j} &\leq \exp(n(j-k)(I(\hat{Y}_{1[j]}; Y_{1[j]} | X_{1[j]}, X_{[j]}) + 6\epsilon)) \\
&\qquad \times M_l^{B-k} \exp(n(B-k)(I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) + 6\epsilon))
\end{aligned} \tag{3.65}
$$

and

$$
\begin{aligned}
&P_1(\bar{x}^n_{[1]}, \dots, \bar{x}^n_{[B-1]}, \bar{\hat{y}}^n_{1[1]}, \dots, \bar{\hat{y}}^n_{1[B-1]}) \\
&\quad \le \exp(n(H(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[k]}, X^{[B]}_{1[k+1]} | Y^{[B]}_{[k]}, \hat{Y}_{1[B]}, X_{[B]}, X^{[j-1]}_{k]}, X_{1[k]}) + 2\epsilon)) \\
&\qquad \times \exp(-n(H(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[k]}, X^{[B]}_{1[k+1]} | X^{[j-1]}_{[k]}, X_{1[k]}) - 2\epsilon)) \\
&\quad = \exp(n(-I(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[k]}, X^{[B]}_{1[k+1]}; Y^{[B]}_{[k]}, \hat{Y}_{1[B]}, X_{[B]} | X^{[j-1]}_{[k]}, X_{1[k]}) + 4\epsilon)) \qquad (3.66)
\end{aligned}
$$

Thus, when n is sufficiently large, using (3.58) and (3.61) through (3.66), we have

$$
Pr(E_{2jk} \cap E_1^c) \le \epsilon, \qquad j, k = 2, \dots, B-1 \qquad (3.67)
$$

if the following conditions are satisfied:

1. For all j such that $1 \le j \le B-1$,

$$
\begin{aligned}
&\frac{1}{n}(B-j)\ln M + (B-j) I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \\
&\qquad < I(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[j]}, X^{[B]}_{1[j+1]}; Y^{[B]}_{[j]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, X_{1[j]}) \qquad (3.68)
\end{aligned}
$$

2. For all j, k such that $1 \le j < k \le B-1$,

$$
\begin{aligned}
&\frac{1}{n}(B-j)\ln M + (B-k) I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \\
&< I(X^{[B-1]}_{[j]}, \hat{Y}^{[B-1]}_{1[k]}, X^{[B]}_{1[k+1]}; Y^{[B]}_{[j]}, \hat{Y}_{1[B]}, X_{1[B]}, \hat{Y}^{[k-1]}_{1[j]}, X^{[k]}_{1[j+1]} | X_{[j-1]}, X_{[j]}) \quad (3.69)
\end{aligned}
$$

3. For all j, k such that $1 \leq k < j \leq B-1$,

$$(j-k)I(\hat{Y}_{1[l]}; Y_{1[l]}|X_{1[l]}, X_{[l]}) + \frac{1}{n}(B-j)\ln M + (B-j)I(\hat{Y}_{1[l]}; Y_{1[l]}|X_{1[l]}, X_{[l]})$$
$$< I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[k]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]}|X_{[k]}^{[j-1]}, X_{1[k]}) \tag{3.70}$$

Therefore, we have

$$P_e = Pr(E_1) + Pr(E_2 \cap E_1^c) \leq (2B + B^2)\epsilon \tag{3.71}$$

When n is sufficiently large, $(2B + B^2)\epsilon$ can be made arbitrarily small.

3.B Appendix: Lower Bounding the Mutual Informations in (3.3), (3.4), (3.5)

For the right hand side of (3.3), we have

$$
\begin{aligned}
& I(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]}; Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, X_{1[j]}) \\
& \overset{1}{=} \sum_{l=j}^{B-1} I(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]}; Y_{[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) \\
& \quad + I(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& \overset{2}{=} I(Y_{[j]}; X_{[j]}, \hat{Y}_{1[j]} | X_{1[j]}, X_{[j-1]}) + \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{1[B]}, X_{[B-1]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& \overset{3}{=} \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]} | X_{1[B]}, X_{[B-1]}) \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{1[B]}, X_{[B-1]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& \overset{4}{\geq} \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) \\
& \quad + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]} | X_{1[B]}, X_{[B-1]}, X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& \quad + I(Y_{[B]}; X_{1[B]}, X_{[B-1]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& = \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) \\
& \quad + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]}, X_{[B-1]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& \overset{5}{=} \sum_{l=j+1}^{B} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) \\
& \overset{6}{\geq} (B-j) I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]})
\end{aligned}
\tag{3.72}
$$

where

1. follows from the chain rule;
2. because of Markov properties 1 and 2;
3. because of the stationarity of the random process and the property that conditioning reduces entropy;
4. because of Markov property 2;
5. because of Markov property 1;
6. because of Markov property 2 and the stationarity of the random process.

For the right hand side of (3.4), we have

$$
\begin{aligned}
& I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j+1]}^{[k]} | X_{[j-1]}, X_{1[j]}) \\
& \stackrel{1}{=} I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[j]}, \hat{Y}_{1[j]} | X_{[j-1]}, X_{1[j]}) \\
& \quad + \sum_{l=j+1}^{k-1} I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[l-1]}, X_{1[j]}^{[l-1]}) \\
& \quad + I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[k]}, X_{1[k]} | X_{[j-1]}, Y_{[j]}^{[k-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k-1]}) \\
& \quad + \sum_{l=k+1}^{B-1} I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \quad + I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, Y_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \stackrel{2}{\geq} I(X_{[j]}; Y_{[j]}, \hat{Y}_{1[j]} | X_{[j-1]}, X_{1[j]}) + \sum_{l=j+1}^{k-1} I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[l-1]}, X_{1[j]}^{[l]}) \\
& \quad + I(X_{[k]}, \hat{Y}_{1[k]}; Y_{[k]} | X_{[j-1]}, Y_{[j]}^{[k-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \quad + \sum_{l=k+1}^{B-1} I(X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]}; Y_{[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \quad + I(X_{[B-1]}, X_{1[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, Y_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \stackrel{3}{=} \sum_{l=j+1}^{k-1} I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[l-1]}, X_{1[j]}^{[l]}) \\
& \quad + \sum_{l=k+1}^{B-1} I(X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]}; Y_{[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \quad + I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{[B-1]}, X_{1[B]}) \\
& \quad + I(X_{[B]}, \hat{Y}_{1[B]}; Y_{[B]} | X_{[j-1+B-k]}, Y_{[j+B-k]}^{[B-1]}, \hat{Y}_{1[j+B-k]}^{[B-1]}, X_{1[j+B-k]}^{[B]}) \\
& \quad + I(X_{[B-1]}, X_{1[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, Y_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \stackrel{4}{\geq} \sum_{l=j+1}^{k-1} I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[l-1]}, X_{1[j]}^{[l]})
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{l=k+1}^{B-1} I(X_{[l]},\hat{Y}_{1[l]},X_{1[l]};Y_{[l]}|X_{[j-1]},Y_{[j]}^{[l-1]},\hat{Y}_{1[j]}^{[k-1]},X_{1[j]}^{[k]})\\
&+I(X_{[B]};Y_{[B]},\hat{Y}_{1[B]}|X_{1[B]},S)+I(X_{[B]},\hat{Y}_{1[B]},X_{1[B]};Y_{[B]}|S)\\
\overset{5}{\geq}\ &(k-j)I(X_{[l]};Y_{[l]},\hat{Y}_{1[l]}|X_{1[l]},Y_{[l-1]},\hat{Y}_{1[l-1]},X_{1[l-1]},X_{[l-2]})\\
&+(B-k)I(Y_{[l]};X_{[l]},\hat{Y}_{1[l]},X_{1[l]}|X_{[l-2]},X_{1[l-1]},Y_{[l-1]}) \qquad (3.73)
\end{aligned}
$$

where

$$
S \triangleq (X_{[j-1+B-k]},Y_{[j+B-k]}^{[B-1]},\hat{Y}_{1[j+B-k]}^{[B-1]},X_{1[j+B-k]}^{[B-1]},X_{[j-1]},Y_{[j]}^{[B-1]},\hat{Y}_{1[j]}^{[k-1]},X_{1[j]}^{[k]}) \qquad (3.74)
$$

and

1. follows from the chain rule;
2. because of Markov properties 1 and 2;
3. because of the stationarity of the random process;

4. because of the following derivation

$$
\begin{aligned}
& I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{[B-1]}, X_{1[B]}) \\
& \quad + I(X_{[B]}, \hat{Y}_{1[B]}; Y_{[B]} | X_{[j-1+B-k]}, Y^{[B-1]}_{[j+B-k]}, \hat{Y}^{[B-1]}_{1[j+B-k]}, X^{[B]}_{1[j+B-k]}) \\
& \quad + I(X_{[B-1]}, X_{1[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, Y^{[B-1]}_{[j]}, \hat{Y}^{[k-1]}_{1[j]}, X^{[k]}_{1[j]}) \\
& \geq I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{[B-1]}, X_{1[B]}, S) + I(X_{[B]}, \hat{Y}_{1[B]}; Y_{[B]} | X_{1[B]}, S) \\
& \quad + I(X_{[B-1]}, X_{1[B]}; Y_{[B]}, \hat{Y}_{1[B]} | S) \\
& \geq I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{[B-1]}, X_{1[B]}, S) + I(X_{[B]}, \hat{Y}_{1[B]}; Y_{[B]} | X_{1[B]}, S) \\
& \quad + I(X_{[B-1]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{1[B]}, S) + I(X_{1[B]}; Y_{[B]} | S) \\
& = I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{1[B]}, S) + I(X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]}; Y_{[B]} | S)
\end{aligned} \tag{3.75}
$$

5. because of Markov property 1 and 2 and the stationarity of the random process.

For the right hand side of (3.5), we have

$$
\begin{aligned}
& I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[k]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[k]}^{[j-1]}, X_{1[k]}) \\
& \stackrel{1}{=} \sum_{l=k}^{B-1} I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[l]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[B-1]}) \\
& \stackrel{2}{\geq} I(Y_{[k]}; \hat{Y}_{1[k]} | X_{[k]}, X_{1[k]}) + \sum_{l=k+1}^{j-1} I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[k]}^{[l]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + I(Y_{[j]}; X_{[j]}, \hat{Y}_{1[j]}, X_{1[j]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[j-1]}) \\
& \quad + \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{[j]}^{[B-1]}, X_{1[B]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[B-1]}) \\
& \stackrel{3}{=} \sum_{l=k+1}^{j-1} I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[k]}^{[l]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[l-1]}) + I(Y_{[B]}; \hat{Y}_{1[B]} | X_{[B]}, X_{1[B]}) \\
& \quad + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]} | X_{[k+B-j]}^{[B-1]}, X_{1[k+B-j]}, Y_{[k+B-j]}^{[B-1]}) \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{[j]}^{[B-1]}, X_{1[B]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[B-1]}) \\
& \stackrel{4}{\geq} \sum_{l=k+1}^{j-1} I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[j]}^{[l]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + I(Y_{[B]}; \hat{Y}_{1[B]}, X_{1[B]} | X_{[j]}^{[B]}, S') + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]} | S') \\
& \stackrel{5}{\geq} (j-k) I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[l]}, X_{[l-1]}, X_{1[l-1]}, Y_{[l-1]}) \\
& \quad + (B-j) I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \qquad (3.76)
\end{aligned}
$$

where

$$S' \triangleq (X_{1[k+B-j]}, Y_{[k]}^{[B-1]}, X_{[k]}^{[j-1]}, X_{1[k]}) \tag{3.77}$$

and

1. follows from the chain rule;

2. because of Markov properties 1 and 2;

3. because of the stationarity of the random process;

4. because of the following derivation

$$\begin{aligned}
&I(Y_{[B]}; \hat{Y}_{1[B]}|X_{[B]}, X_{1[B]}) + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]}|X_{[k+B-j]}^{[B-1]}, X_{1[k+B-j]}, Y_{[k+B-j]}^{[B-1]}) \\
&\quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{[j]}^{[B-1]}, X_{1[B]}|X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[B-1]}) \\
&\geq I(Y_{[B]}; \hat{Y}_{1[B]}|X_{[B]}, X_{1[B]}, S') + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]}|X_{[j]}^{[B-1]}, S') \\
&\quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{[j]}^{[B-1]}, X_{1[B]}|S') \\
&= I(Y_{[B]}; \hat{Y}_{1[B]}|X_{[B]}, X_{1[B]}, S') + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]}|X_{[j]}^{[B-1]}, S') \\
&\quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{1[B]}|X_{[j]}^{[B-1]}, S') + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{[j]}^{[B-1]}|S') \\
&\geq I(Y_{[B]}; \hat{Y}_{1[B]}|X_{[B]}, X_{1[B]}, X_{[j]}^{[B-1]}, S') + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]}|X_{[j]}^{[B-1]}, S') \\
&\quad + I(Y_{[B]}; X_{1[B]}|X_{[B]}, X_{[j]}^{[B-1]}, S') + I(Y_{[B]}; X_{[j]}^{[B-1]}|S') \\
&= I(Y_{[B]}; \hat{Y}_{1[B]}, X_{1[B]}|X_{[j]}^{[B]}, S') + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]}|S')
\end{aligned} \tag{3.78}$$

5. because of Markov property 1 and 2 and the stationarity of the random process.

3.C Appendix: Proof of Theorem 3.4.2

First, we note that condition 1 is equivalent to the expression in Theorem 3.4.1. We also note that condition 2 is seemingly weaker than condition 1 because (3.36) is implied by (3.33) and (3.34), and condition 3 is seemingly stronger than condition 2 because condition 3 consists of every element in condition 2 plus (3.38). Even though they seem different, these three conditions are indeed equivalent. The equivalence of conditions 2 and 3 is shown in [2]. Here, we use a similar proof technique to show the equivalence of conditions 1 and 2 as follows[2]. For a given distribution $p(x, x_1, y, y_1, \hat{y}_1)$, condition 1 is stronger than condition 2, which means that an arbitrary rate R satisfying condition 1 will also satisfy condition 2. Conversely, for a rate R satisfying condition 2, if (3.34) is satisfied, then condition 1 is satisfied. If (3.34) is not satisfied, i.e.,

$$I(Y_1; \hat{Y}_1|X_1) \geq I(\hat{Y}_1; Y|X_1) + I(X_1; Y) \tag{3.79}$$

we know that $R \in [0, R^*]$, where

$$R^* - I(X; \hat{Y}_1|X_1) \leq I(X; Y|\hat{Y}_1, X_1) \tag{3.80}$$

$$R^* - I(X; \hat{Y}_1|X_1) + I(Y_1; \hat{Y}_1|X_1) = I(X, \hat{Y}_1; Y|X_1) + I(X_1; Y) \tag{3.81}$$

[2]A similar result is given in [14] by means of time-sharing.

That is, R^* is defined such that (3.36) is satisfied with equality. We may rewrite (3.80) and (3.81) as

$$R^* \leq I(X;Y|X_1) + I(X;\hat{Y}_1|Y,X_1) \tag{3.82}$$

$$R^* = I(X,X_1;Y) - I(Y_1;\hat{Y}_1|X,X_1,Y) \tag{3.83}$$

We define a new random variable $\hat{Y}_1'$ such that $\hat{Y}_1'$ has the same marginal distribution as $\hat{Y}_1$ and $\hat{Y}_1' \to \hat{Y}_1 \to (Y_1,X,X_1,Y)$. Due to the continuity of mutual information, there exists a choice of $\hat{Y}_1'$ such that $I(X;\hat{Y}_1'|Y,X_1) = A$ for any $A \in [0, I(X;\hat{Y}_1|Y,X_1)]$. If $R^* - I(X;Y|X_1) > 0$, we choose $\hat{Y}_1'$ such that $R^* = I(X;Y|X_1) + I(X;\hat{Y}_1'|Y,X_1)$. We note that, in this case, $I(Y_1;\hat{Y}_1|X,X_1,Y) \geq I(Y_1;\hat{Y}_1'|X,X_1,Y)$. Thus,

$$R^* = I(X;Y|X_1) + I(X;\hat{Y}_1'|Y,X_1) \tag{3.84}$$

$$R^* \leq I(X,X_1;Y) - I(Y_1;\hat{Y}_1'|X,X_1,Y) \tag{3.85}$$

which means that R^* satisfies condition 1 with joint distribution $p(x,x_1,y,y_1,\hat{y}_1')$ and so does any $R \leq R^*$. If $R^* - I(X;Y|X_1) \leq 0$, we choose $\hat{Y}_1'$ independent of $(\hat{Y}_1,X,X_1,Y_1,Y)$. In this case,

$$R^* \leq I(X;Y|X_1) + I(X;\hat{Y}_1'|Y,X_1) = I(X;Y|X_1) \tag{3.86}$$

$$0 = I(Y_1;\hat{Y}_1'|X_1) \leq I(\hat{Y}_1';Y|X_1) + I(X_1;Y) \tag{3.87}$$

Therefore, in this case, R^* satisfies condition 1 with joint distribution $p(x, x_1, y, y_1, \hat{y}_1')$ and so does any $R \leq R^*$.

As we mentioned above the equivalence between condition 2 and 3 is shown in [2]. For completeness, we restate their proof here as follows. For a given distribution $p(x, x_1, y, y_1, \hat{y}_1)$, condition 3 is stronger than condition 2, which means that an arbitrary rate R satisfying condition 3 will also satisfy condition 2. Conversely, for a rate R satisfying condition 2, if (3.38) is satisfied, then condition 3 is satisfied. If (3.38) is not satisfied, i.e., the following inequalities are satisfied

$$R - I(X; \hat{Y}_1|X_1) \leq I(X; Y|\hat{Y}_1, X_1) \tag{3.88}$$

$$I(\hat{Y}_1; Y_1|X_1, X) \geq I(\hat{Y}_1; Y|X_1, X) + I(X_1; Y|X) \tag{3.89}$$

$$R - I(X; \hat{Y}_1|X_1) + I(Y_1; \hat{Y}_1|X_1) \leq I(X, \hat{Y}_1; Y|X_1) + I(X_1; Y) \tag{3.90}$$

then the following inequalities are satisfied also, since we simply drop the first inequality,

$$I(\hat{Y}_1; Y_1|X_1, X) \geq I(\hat{Y}_1; Y|X_1, X) + I(X_1; Y|X) \tag{3.91}$$

$$R - I(X; \hat{Y}_1|X_1) + I(Y_1; \hat{Y}_1|X_1) \leq I(X, \hat{Y}_1; Y|X_1) + I(X_1; Y) \tag{3.92}$$

By combining (3.91) and (3.92), we have

$$
\begin{aligned}
R \leq & I(X;\hat{Y}_1|X_1) - I(Y_1;\hat{Y}_1|X_1) + I(\hat{Y}_1;Y_1|X_1,X) \\
& + I(X,\hat{Y}_1;Y|X_1) + I(X_1;Y) - I(\hat{Y}_1;Y|X_1,X) - I(X_1;Y|X) \\
\leq & I(X;Y|X_1) - (I(X_1;Y|X) - I(X_1;Y)) \\
\leq & I(X;Y|X_1)
\end{aligned} \tag{3.93}
$$

which implies condition 3, i.e., (3.37), (3.38) and (3.39), with $\hat{Y}_1$ set to be a constant.

Chapter 4

Capacity of a Class of Diamond Channels

4.1 Problem Statement and the Result

The diamond channel was first introduced by Schein in 2001 [32]. The diamond channel consists of one transmitter, two relays and a receiver, where the transmitter and the two relays form a broadcast channel as the first stage and the two relays and the receiver form a multiple access channel as the second stage. The capacity of the diamond channel in its most general form is open. Schein explored several special cases of the diamond channel, one of which [32, Section 3.5] is specified as follows (see Figure 4.1). The multiple access channel consists of two orthogonal links with rate constraints R_1 and R_2, respectively. The broadcast channel contains a noisy branch and a noiseless branch, i.e., with input X and two outputs X and Y. We refer to the relay node receiving Y as the noisy relay and the relay node receiving X as the noiseless relay. Schein provided two achievable schemes for this class of diamond channels. In this chapter, we will prove the capacity of this special class of diamond channels.

The formal definition of the problem is as follows. Consider a channel with in-

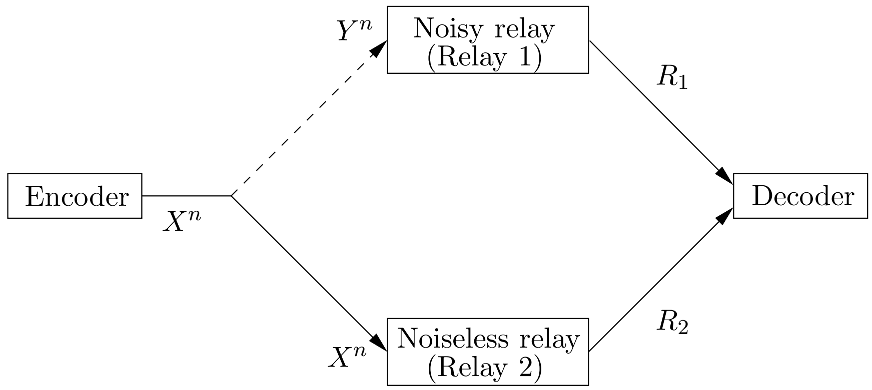

Figure 4.1: The diamond channel.

put alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, which is characterized by the transition probability $p(y|x)$. Assume an n-length block code consisting of (f, g, h, φ) where

$$f : \{1, 2, \ldots, M\} \mapsto \mathcal{X}^n \tag{4.1}$$

$$g : \mathcal{Y}^n \mapsto \{1, 2, \ldots, |g|\} \tag{4.2}$$

$$h : \{1, 2, \ldots, M\} \mapsto \{1, 2, \ldots, |h|\} \tag{4.3}$$

$$\varphi : \{1, 2, \ldots, |g|\} \times \{1, 2, \ldots, |h|\} \mapsto \{1, 2, \ldots, M\} \tag{4.4}$$

Here f denotes the encoding function at the transmitter, g and h denote the processing functions at the noisy and noiseless relays, respectively, and φ denotes the decoding function at the receiver.

The encoder sends $x^n = f(m)$ into the channel, where $m \in \{1, 2, \ldots, M\}$. The decoder reconstructs $\hat{m} = \varphi(g(Y^n), h(m))$. The average probability of error is defined as

$$P_e \triangleq \frac{1}{M} \sum_{m=1}^{M} Pr(\hat{m} \neq m | m \text{ is sent}) \tag{4.5}$$

The rate triple (R, R_1, R_2) is achievable if for every $0 < \epsilon < 1$, $\eta > 0$ and every sufficiently large n, there exists an n-length block code (f, g, h, φ), such that $P_e \leq \epsilon$ and

$$\frac{1}{n} \ln M \geq R - \eta \tag{4.6}$$

$$\frac{1}{n} \ln |g| \leq R_1 + \eta \tag{4.7}$$

$$\frac{1}{n} \ln |h| \leq R_2 + \eta \tag{4.8}$$

The following theorem characterizes the capacity of the class of diamond channels considered in this chapter.

Theorem 4.1.1 *The rate triple (R, R_1, R_2) is achievable in the above channel if and only if the following conditions are satisfied*

$$R \leq I(U;Y) + H(X|U) \tag{4.9}$$

$$R_1 \geq I(Z;Y|U,X) \tag{4.10}$$

$$R_2 \geq H(X|Z,U) \tag{4.11}$$

$$R_1 + R_2 \geq R + I(Y;Z|X,U) \tag{4.12}$$

for some joint distribution

$$p(u,z,x,y) = p(u,x)p(y|x)p(z|u,y) \tag{4.13}$$

with cardinalities of alphabets satisfying

$$|\mathcal{U}| \leq |\mathcal{X}| + 4 \tag{4.14}$$

$$|\mathcal{Z}| \leq |\mathcal{U}||\mathcal{Y}| + 3 \leq |\mathcal{X}||\mathcal{Y}| + 4|\mathcal{X}| + 3 \tag{4.15}$$

4.2 The Achievability

Assume a given joint distribution

$$p(u, z, x, y) = p(u, x)p(y|x)p(z|u, y) \tag{4.16}$$

and consider that the information theoretic quantities on the right hand sides of (4.9), (4.10), (4.11) and (4.12) are evaluated with this fixed joint probability distribution.

Consider a message W with rate R. If $R \leq H(X|Z, U)$, reliable transmission can be achieved by letting $g(Y^n) = \phi$ (constant) and $h(W) = W$, i.e., by sending the message through the noiseless relay. Thus, we will only consider the case where

$$H(X|Z, U) < R \leq I(U; Y) + H(X|U) \tag{4.17}$$

We will show that the message can be reliably transmitted with a pair of functions (g, h) such that $(\frac{1}{n} \ln |g|, \frac{1}{n} \ln |h|)$ lies in the inverse pentagon[1] with corners a and b in Figure 4.2. However, we instead prove reliable transmission with $(\frac{1}{n} \ln |g|, \frac{1}{n} \ln |h|)$

[1]By "inverse pentagon" with corner points a and b, we mean the region in the (R_1, R_2) space that is to the "north-east" of line segment $[a, b]$. More specifically, this is the region described by inequalities in (4.10), (4.11) and (4.12).

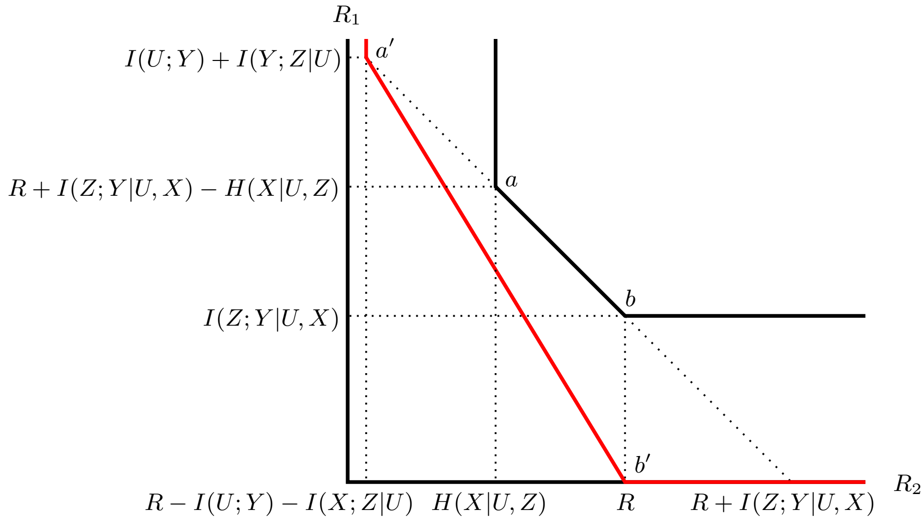

Figure 4.2: Rate region of (R_1, R_2) when $H(X|U,Z) \leq R \leq I(U;Y) + I(X;Z|U)$.

lying in the inverse pentagon with corners a' and b', which contains the inverse pentagon with corners a and b and thus imposes a stronger condition to prove. It is straightforward to have reliable transmission with the rate pair at point b' by letting $g(Y^n) = \phi$ (constant) and $h(W) = W$. Thus, it remains to prove that reliable transmission is possible with the rate pair at point a', i.e.,

$$R_1 = I(U;Y) + I(Y;Z|U) \tag{4.18}$$

$$R_2 = R - I(U;Y) - I(X;Z|U) \tag{4.19}$$

Let us assume that the message W is decomposed as $W = (W_a, W_b, W_c)$. For a positive number ϵ, let us define

$$M_a \triangleq |W_a| = \exp(n(I(U;Y) - 3\epsilon)) \tag{4.20}$$

$$M_b \triangleq |W_b| = \frac{M}{M_a M_c} = \exp(\ln M - n(I(U;Y) + I(X;Z|U) + 6\epsilon)) \tag{4.21}$$

$$M_c \triangleq |W_c| = \exp(n(I(X;Z|U) - 3\epsilon)) \tag{4.22}$$

Random codebook generation: We use a superpostion code structure. The size of the inner code is M_a. For each inner codeword, we independently generate M_b outer codes. The size of each outer code is M_c.

- Independently generate M_a sequences, $u^n(1), u^n(2), \ldots, u^n(M_a)$, according to $\prod_{i=1}^n p(u_i)$ where $p(u_i) = p(u)$, for $i = 1, 2, \ldots, n$.

- For $u^n(j)$, $j = 1, 2, \ldots, M_a$, independently generate M_b codebooks, $\mathcal{C}(j,1), \mathcal{C}(j,2)$, $\ldots$, $\mathcal{C}(j, M_b)$.

- In the codebook $\mathcal{C}(j,k)$, $j = 1, 2, \ldots, M_a$, $k = 1, 2, \ldots, M_b$, independently generate M_c codewords $x^n(j,k,1), x^n(j,k,2), \ldots, x^n(j,k,M_c)$ according to $\prod_{i=1}^n p(x_i| U_i = u_i(j))$, where $p(x_i|U = u_i(j)) = p(x|u)$, for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, M_a$, $k = 1, 2, \ldots, M_b$.

There will be no overlapping codebooks with high probability when n is sufficiently large, because

$$\frac{1}{n} \ln M_b M_c < H(X|U) \tag{4.23}$$

Encoding at the transmitter: Let $W = (W_a, W_b, W_c)$ be the message. We send codeword $X^n = f(W_a, W_b, W_c) \triangleq x^n(W_a, W_b, W_c)$ into the channel.

Processing at the noisy relay: First, after having received Y^n, seek

$$\hat{U}^n = u^n(\hat{W}_a) \in \{u^n(1), u^n(2), \dots, u^n(M_a)\} \tag{4.24}$$

such that

$$(\hat{U}^n, Y^n) \in \mathcal{T}^n_{[UY]} \tag{4.25}$$

where the definition of strong typical set can be found in [13, Section 1.2]. If there is not any such $\hat{U}^n$, then let $\hat{U}^n$ be an arbitrary sequence in $\{u^n(1), u^n(2), \dots, u^n(M_a)\}$. Secondly, construct a conditional rate distortion code according to $\prod_{i=1}^n p(z_i, y_i|\hat{u}_i)$ with encoding function $g'(Y^n, \hat{U}^n)$ and $|g'| = L = \exp(n(I(Y;Z|U) + \tau))$. Finally send $\hat{U}^n$ and $Z^n \triangleq g'(Y^n, \hat{U}^n)$ to the destination, i.e.,

$$g(Y^n) = (\hat{U}^n, Z^n) \tag{4.26}$$

where

$$|g| = M_a \times L \le \exp(n(I(U;Y) + I(X;Z|U) + \tau - 3\epsilon)) \tag{4.27}$$

Processing at the noiseless relay: Let $h(f(W_a, W_c, W_b)) = W_b$ where

$$|h| = M_b = \exp(\ln M - n(I(U;Y) + I(X;Z|U) + 6\epsilon)) \tag{4.28}$$

Decoding: Decoder collects $(\hat{U}^n, Z^n)$ from the noisy relay and W_b from the noiseless relay. The decoder seeks a codeword $x^n(W_a, W_b, i)$ from the codebook $\mathcal{C}(W_a, W_b)$ such that

$$(x^n(\hat{W}_a, W_b, i), Z^n) \in \mathcal{T}^n_{[XZ|U]}(\hat{U}^n) \tag{4.29}$$

Probability of error: The error occurs when $(\hat{U}, \hat{X}) \neq (U, X)$. The average probability of error can be decomposed into

$$Pr(E) \leq Pr(E_1 \cup E_2 \cup E_3) = Pr(E_1) + Pr(E_2 \cap E_1^c) + Pr(E_3 \cap E_1^c \cap E_2^c) \tag{4.30}$$

where

$$E_1 \triangleq (U^n, X^n, Y^n, Z^n) \notin \mathcal{T}^n_{[UXYZ]} \tag{4.31}$$

$$E_2 \triangleq \bigcup_{\bar{u}^n \neq U^n, \bar{u}^n \in \{u^n(1), u^n(2), \ldots, u^n(M_a)\}} (\bar{u}^n, Y^n) \in \mathcal{T}^n_{[UY]} \tag{4.32}$$

$$E_3 \triangleq \bigcup_{\bar{x}^n \neq X^n, \bar{x}^n \in \mathcal{C}(W_a, W_b)} (\bar{x}^n, Z^n) \in \mathcal{T}^n_{[XZ|U]}(U^n) \tag{4.33}$$

We note that

$$\begin{aligned} Pr(E_1) \leq & Pr(U^n \notin \mathcal{T}^n_{[U]}) + Pr((Y^n, Z^n) \notin \mathcal{T}^n_{[YZ|U]}(U^n)) + \\ & Pr(X^n \notin \mathcal{T}^n_{[X|YZU]}(Y^n, Z^n, U^n)) \end{aligned} \tag{4.34}$$

where

- U^n is generated in an i.i.d. fashion with probability $p(u)$. Thus, when n is sufficiently large, we have

$$Pr(U^n \notin \mathcal{T}^n_{[U]}) \leq \epsilon \tag{4.35}$$

- Z^n is a conditional rate distortion code for Y^n conditioned on U^n. Thus, when n is sufficiently large, $L = \exp(nI(Y;Z|U) + \tau)$, and $U^n \in \mathcal{T}^n_{[U]}$, we have

$$Pr((Y^n, Z^n) \notin \mathcal{T}^n_{[YZ|U]}(U^n)) \leq \epsilon \tag{4.36}$$

- X^n can be viewed as being generated according to an i.i.d. conditional probability $p(x|u,y)$ with respect to (U^n, Y^n). Thus, when n is sufficiently large and $(Y^n, Z^n, U^n) \in \mathcal{T}^n_{[YZU]}$,

$$Pr(X^n \notin \mathcal{T}^n_{[X|YZU]}(Y^n, Z^n, U^n)) \leq \epsilon \tag{4.37}$$

From the above calculation, we have

$$Pr(E_1) = Pr((U^n, X^n, Y^n, Z^n) \notin \mathcal{T}^n_{[UXZ]}) \leq 3\epsilon \tag{4.38}$$

For the second error event, we note that $M_a = \exp(n(I(U;Y) - 3\epsilon)$ and

$$\begin{aligned}
Pr(E_2 \cap E_1^c) &= Pr\left(\bigcup_{\bar{u}^n \neq U^n, \bar{u}^n \in \{u^n(1), u^n(2), \ldots, u^n(M_a)\}} (\bar{u}^n, Y^n) \in \mathcal{T}^n_{[UY]} | (Y^n) \in \mathcal{T}^n_{[Y]} \right) \\
&\leq \sum_{i=1}^{M_a} Pr((u^n(i), Y^n) \in \mathcal{T}^n_{[UY]} | Y^n \in \mathcal{T}^n_{[Y]}) \\
&\leq M_a Pr(u^n(i) \in \mathcal{T}^n_{[U|Y]}(Y^n)) \\
&\leq M_a \exp(-nH(U) + n\epsilon) \exp(nH(U|Y) + n\epsilon) \\
&= \exp(-n\epsilon) \\
&\leq \epsilon
\end{aligned} \tag{4.39}$$

for sufficiently large n. We note that $M_c = \exp(n(I(X;Z|U) - 3\epsilon)$, then

$$\begin{aligned}
Pr(E_3 \cap E_1^c) &= Pr\left(\bigcup_{\bar{x}^n \neq X^n, \bar{x}^n \in \mathcal{C}(W_a, W_b)} (\bar{x}^n, Z^n) \in \mathcal{T}^n_{[XZ|U]}(U^n) | (Z^n, U) \in \mathcal{T}^n_{[ZU]} \right) \\
&\leq \sum_{i=1}^{M_c} Pr((x(M_a, M_b, i), Z^n) \in \mathcal{T}^n_{[XZ|U]}(U^n) | (Z^n, U^n) \in \mathcal{T}^n_{[ZU]}) \\
&\leq M_c Pr(x(M_a, M_b, i) \in \mathcal{T}^n_{[X|ZU]}(Y^n)) \\
&\leq M_c \exp(-nH(X|U) + n\epsilon) \exp(nH(X|Z, U) + n\epsilon) \\
&= \exp(-n\epsilon) \\
&\leq \epsilon
\end{aligned} \tag{4.40}$$

for sufficiently large n. Thus, the average probability error is upper bounded as

$$Pr(E) \leq 3\epsilon + \epsilon + \epsilon = 5\epsilon \tag{4.41}$$

which goes to zero when n goes to infinity.

4.3 The Converse

Define $Z_i \triangleq g$ and $U_i \triangleq (Y^{i-1}, X_{i+1}^n)$. We note that

$$p(u_i, x_i, y_i, z_i) = p(u_i, x_i)p(y_i|x_i)p(z_i|y_i, u_i) \tag{4.42}$$

We have

$$\begin{aligned}
\ln M &= H(X^n) \\
&= \sum_{i=1}^{n} H(X_i|X_{i+1}^n) \\
&\leq \sum_{i=1}^{n} I(Y^{i-1}; Y_i) + H(X_i|X_{i+1}^n) \\
&= \sum_{i=1}^{n} I(Y^{i-1}, X_{i+1}^n; Y_i) - I(X_{i+1}^n; Y_i|Y^{i-1}) + H(X_i|Y^{i-1}, X_{i+1}^n) + I(Y^{i-1}; X_i|X_{i+1}^n) \\
&\stackrel{1}{=} \sum_{i=1}^{n} I(Y^{i-1}, X_{i+1}^n; Y_i) + H(X_i|Y^{i-1}, X_{i+1}^n) \\
&= \sum_{i=1}^{n} I(U_i; Y_i) + H(X_i|U_i)
\end{aligned} \tag{4.43}$$

where

1. Because of the following equality [12, Lemma 7]

$$\sum_{i=1}^{n} I(X_{i+1}^n; Y_i|Y^{i-1}) = \sum_{i=1}^{n} I(Y^{i-1}; X_i|X_{i+1}^n) \tag{4.44}$$

We have

$$
\begin{aligned}
\ln |g| &\geq H(g) \\
&\geq H(g|h) \\
&\geq H(g|h) - H(g|h, Y^n) \\
&= I(g; Y^n|h) \\
&= \sum_{i=1}^{n} I(g; Y_i|h, Y^{i-1}) \\
&= \sum_{i=1}^{n} I(g, X_{i+1}^n; Y_i|h, Y^{i-1}) - I(X_{i+1}^n; Y_i|g, h, Y^{i-1}) \\
&\stackrel{1}{=} \sum_{i=1}^{n} I(g, X_{i+1}^n; Y_i|h, Y^{i-1}) - I(Y^{i-1}; X_i|g, h, X_{i+1}^n) \\
&\geq \sum_{i=1}^{n} I(g, X_{i+1}^n; Y_i|h, Y^{i-1}) - H(X_i|g, h, X_{i+1}^n) \\
&= -H(X^n|g, h) + \sum_{i=1}^{n} I(g, X_{i+1}^n; Y_i|h, Y^{i-1}) \\
&\stackrel{2}{\geq} \sum_{i=1}^{n} I(g, X_{i+1}^n; Y_i|h, Y^{i-1}) - \epsilon \\
&\geq \sum_{i=1}^{n} I(g; Y_i|h, Y^{i-1}, X_{i+1}^n) - \epsilon \\
&\stackrel{3}{\geq} \sum_{i=1}^{n} I(g; Y_i|h, Y^{i-1}, X_{i+1}^n, X_i) - \epsilon \\
&\stackrel{4}{=} \sum_{i=1}^{n} I(g; Y_i|Y^{i-1}, X_{i+1}^n, X_i) - \epsilon \\
&= \sum_{i=1}^{n} I(Z_i; Y_i|U_i, X_i) - \epsilon \qquad (4.45)
\end{aligned}
$$

where

1. Because of the following equality [12, Lemma 7]

$$\sum_{i=1}^{n} I(X_{i+1}^{n};Y_i|g,h,Y^{i-1}) = \sum_{i=1}^{n} I(Y^{i-1};X_i|g,h,X_{i+1}^{n}) \tag{4.46}$$

2. Due to Fano's inequality.

3. g is a deterministic function of Y^n. Due to the memoryless property, we have

$$H(g|Y_i,h,Y^{i-1},X_{i+1}^{n},X_i) = H(g|Y_i,h,Y^{i-1},X_{i+1}^{n}) \tag{4.47}$$

4. g is a deterministic function of Y^n and h is a deterministic function of X^n. Due to the memoryless property, we have

$$H(g|h,Y^{i-1},X_{i+1}^{n},X_i) = H(g|Y^{i-1},X_{i+1}^{n},X_i) \tag{4.48}$$

$$H(g|h,Y^{i-1},X_{i+1}^{n},X_i,Y_i) = H(g|Y^{i-1},X_{i+1}^{n},X_i,Y_i) \tag{4.49}$$

We have

$$
\begin{aligned}
\ln |h| &\geq H(h|g) \\
&\geq I(h; X^n|g) \\
&= H(X^n|g) - H(X^n|g, h) \\
&\overset{1}{\geq} H(X^n|g) - n\epsilon \\
&= \sum_{i=1}^{n} H(X_i|X_{i+1}^n, g) - \epsilon \\
&\geq \sum_{i=1}^{n} H(X_i|Y^{i-1}, X_{i+1}^n, g) - \epsilon \\
&= \sum_{i=1}^{n} H(X_i|U_i, Z_i) - \epsilon
\end{aligned}
\tag{4.50}
$$

where

1. Due to Fano's inequality.

We have

$$
\begin{aligned}
\ln|g| + \ln|h| &\geq H(g,h) \\
&\geq I(g,h;X^n,Y^n) \\
&\geq I(X^n;g,h) + I(Y^n;g,h|X^n) \\
&= H(X^n) - H(X^n|g,h) + I(Y^n;g,h|X^n) \\
&\stackrel{1}{\geq} \ln M - n\epsilon + I(Y^n;g,h|X^n) \\
&\stackrel{2}{=} \ln M - n\epsilon + I(Y^n;g|X^n) \\
&= \ln M + \sum_{i=1}^{n} -\epsilon + I(Y_i;g|X^n,Y^{i-1}) \\
&\stackrel{3}{=} \ln M + \sum_{i=1}^{n} -\epsilon + I(Y_i;g|X_i,Y^{i-1},X_{i+1}^n) \\
&= \ln M + \sum_{i=1}^{n} -\epsilon + I(Y_i;Z_i|X_i,U_i)
\end{aligned}
\tag{4.51}
$$

1. Due to Fano's inequality.

2. h is a deterministic function of X^n

3. g is a deterministic function of Y^n. Due to the memoryless property, we have

$$H(g|X_i,Y^{i-1},X_{i+1}^n,X^{i-1}) = H(g|X_i,Y^{i-1},X_{i+1}^n) \tag{4.52}$$

$$H(g|Y_i,X_i,Y^{i-1},X_{i+1}^n,X^{i-1}) = H(g|Y_i,X_i,Y^{i-1},X_{i+1}^n) \tag{4.53}$$

We note that $\frac{1}{n}\ln M \geq R-\eta$, $\frac{1}{n}\ln|g| \leq R_1+\eta$ and $\frac{1}{n}\ln|h| \leq R_2+\eta$, for an arbitrary $\eta>0$. Assume $\epsilon \to 0$, then from (4.43), (4.45), (4.50) and (4.51), we have

$$R \leq \frac{1}{n}\sum_{i=1}^{n} I(U_i;Y_i) + H(X_i|U_i) \tag{4.54}$$

$$R_1 \geq \frac{1}{n}\sum_{i=1}^{n} I(Z_i;Y_i|U_i,X_i) \tag{4.55}$$

$$R_2 \geq \frac{1}{n}\sum_{i=1} H(X_i|U_i,Z_i) \tag{4.56}$$

$$R_1+R_2 \geq R + \frac{1}{n}\sum_{i=1}^{n} I(Y_i;Z_i|X_i,U_i) \tag{4.57}$$

Define a time-sharing random variable Q, which is uniformly distributed on $\{1,2,\ldots,n\}$. Also define a set of random variables $(X,Y,\tilde{U},\tilde{Z})$ such that

$$Pr(X=x,Y=y,\tilde{U}=u,\tilde{Z}=z|Q=i) = p(X_i=x,Y_i=y,U_i=u,Z_i=z), \qquad i=1,2,\ldots,n \tag{4.58}$$

Define $U=(\tilde{U},Q)$ and $Z=(\tilde{Z},Q)$, then

$$\begin{aligned} R &\leq \frac{1}{n}\sum_{i=1}^{n} I(U_i;Y_i) + H(X_i|U_i) \\ &= I(\tilde{U};Y|Q) + H(X|\tilde{U},Q) \\ &\leq I(\tilde{U},Q;Y) + H(X|\tilde{U},Q) \\ &= I(U;Y) + H(X|U) \end{aligned} \tag{4.59}$$

$$
\begin{aligned}
R_1 &\geq \frac{1}{n}\sum_{i=1}^{n} I(Z_i;Y_i|U_i,X_i) \\
&= I(\tilde{Z};Y|\tilde{U},Q,X) \\
&= I(Z;Y|U,X)
\end{aligned} \tag{4.60}
$$

$$
\begin{aligned}
R_2 &\geq \frac{1}{n}\sum_{i=1} H(X_i|U_i,Z_i) \\
&= H(X|\tilde{U},\tilde{Z},Q) \\
&= H(X|U,Z)
\end{aligned} \tag{4.61}
$$

$$
\begin{aligned}
R_1 + R_2 &\geq R + \frac{1}{n}\sum_{i=1}^{n} I(Y_i;Z_i|X_i,U_i) \\
&= R + I(\tilde{Z};Y|\tilde{U},X,Q) \\
&= R + I(Z;Y|U,X)
\end{aligned} \tag{4.62}
$$

where (4.59), (4.60), (4.61) and (4.62) are the same as (4.9), (4.10), (4.11) and (4.12), concluding the proof.

Finally, we note that the bounds on the cardinalities of the alphabets in (4.14) and (4.15) can be proven in a way similar to [23, Appendix D].

4.4 Remarks

We have several remarks regarding this result as follows:

1. The capacity is strictly smaller than the cut-set bound [11], because first

$$R \leq R_1 + R_2 - I(Y; Z|U, X) \tag{4.63}$$

An operational interpretation is that when the noisy relay cannot fully decode the message, or in other words, when the noisy relay cannot remove the noise completely, the data going through the link from the noisy relay to the receiver contains noise. Thus, the useful information flowing through the multiple access cut will be strictly less than $R_1 + R_2$. Secondly, we note that

$$R \leq I(U; Y) + H(X|U) \leq H(X) \tag{4.64}$$

An operational interpretation is that when the noisy relay decodes the message with a positive rate, the rate of information flowing through the broadcast cut becomes strictly less than $H(X)$.

Consider the following example. Let X and Y be binary and

$$Y = X \oplus W \tag{4.65}$$

where the sum is a modulo-2 sum and W has a Bernoulli distribution with entropy 0.5 bits. We assume $R_1 = R_2 = 0.5$ bits. The cut-set bound in this example is 1 bit, which is not achievable. Because if R is equal to 1 bit, we have,

$$R = I(U;Y) + H(X|U) = H(X) = 1 \tag{4.66}$$

then, U has to be independent of X and Y. Also, we have

$$R = R_1 + R_2 - I(Y;Z|U,X) = R_1 + R_2 = 1 \tag{4.67}$$

then, Z has to be independent of X and Y if U is independent of X and Y. However, if U and Z are independent of X and Y, we arrive at the following contradiction,

$$0.5 = R_2 \geq H(X|Z,U) = H(X) = 1 \tag{4.68}$$

which means that the cut-set bound is not achievable in this example. We note that, even in this binary example where $|\mathcal{X}| = |\mathcal{Y}| = 2$, the cardinalities of the auxiliary random variables U and Z are $|\mathcal{U}| \leq 6$ and $|\mathcal{Z}| \leq 15$. These large cardinality bounds make it practically impossible to evaluate the capacity of this diamond channel. However, we note that, even though we were not able to compute the exact value of the capacity in this example, we were able to conclude that the capacity is strictly less than the cut-set bound, which is 1 bit.

We know that the capacity of a diamond channel with four orthogonal links is equal to the cut-set bound in this channel. Our result shows that introducing the broadcast node will reduce the capacity of this all-orthogonal diamond channel. Networks with broadcast nodes have been studied recently from different perspectives, e.g., information theory and network coding [15,26,31]. We note that our diamond channel model is a simple example of a general network with a broadcast node. Thus, we conclude that the cut-set bound in general is not tight in networks with broadcast nodes.

2. The processing at the noisy relay includes two operations: decode the inner code U^n and compress the channel output Y^n to Z^n conditioned on U^n. This processing is essentially the same as Theorem 7 in [8], i.e., combination of DAF and CAF. DAF [8, Theorem 1] has been shown to be optimal in the degraded relay channel [8]. Partial DAF, a special case of [8, Theorem 7] without compression, has been shown to be optimal in semi-deterministic relay channel [17] and the relay channel with orthogonal transmitter-relay link [18]. Recently, CAF [8, Theorem 6] has been shown to be optimal in two special relay channels [3, 24]. To our knowledge, we are the first to show the optimality of the combination of DAF and CAF in some specific channel, even though the channel we consider is not a three-node relay channel in the strict sense, i.e., as in [8].

3. If we assume $R = H(X) - R_0$, then Theorem 4.1.1 can be rewritten as follows

$$R \leq I(U;Y) + H(X|U) \quad \longleftrightarrow \quad R_0 \geq I(U;X|Y) \tag{4.69}$$

$$R_1 \geq I(Z;Y|U,X) \quad \longleftrightarrow \quad R_1 \geq I(Z;Y|U,X) \tag{4.70}$$

$$R_2 \geq H(X|Z,U) \quad \longleftrightarrow \quad R_2 \geq I(X;X|Z,U) \tag{4.71}$$

$$R_1 + R_2 \geq R + I(Y;Z|X,U) \quad \longleftrightarrow \quad R_0 + R_1 + R_2 \geq I(X,Y;U,X,Z) \tag{4.72}$$

for some joint distribution

$$p(u, z, x, y) = p(u, x)p(y|x)p(z|u, y) \tag{4.73}$$

We note that the right hand sides of (4.69), (4.70), (4.71) and (4.72) in addition to the distribution constraint in (4.73) are the same as the rate region of the rate-distortion problem studied by Kaspi and Berger as shown in Figure 4.3 [23, Theorem 2.1, Case C].

This duality between our diamond channel coding problem and the Kaspi-Berger source coding problem is similar to the duality between the single-user channel coding problem and the Slepian-Wolf source coding problem [13, Section 3.1] by viewing the codebook information in the channel coding problem as the information sent to all the terminals in the source coding problem, e.g., the information with rate R_0 in Figure 4.3. Thus, the achievability of our diamond channel coding problem can be obtained from the achievability of Kaspi-Berger source coding problem, in the same way that the achievability of the multiple access channel coding problem can be obtained from the achievability of fork network coding problem [13, Section 3.2].

4.5 Conclusion

In this chapter, we studied a special class of diamond channels which was introduced by Schein in 2001. In this special class, each diamond channel consists of a trans-

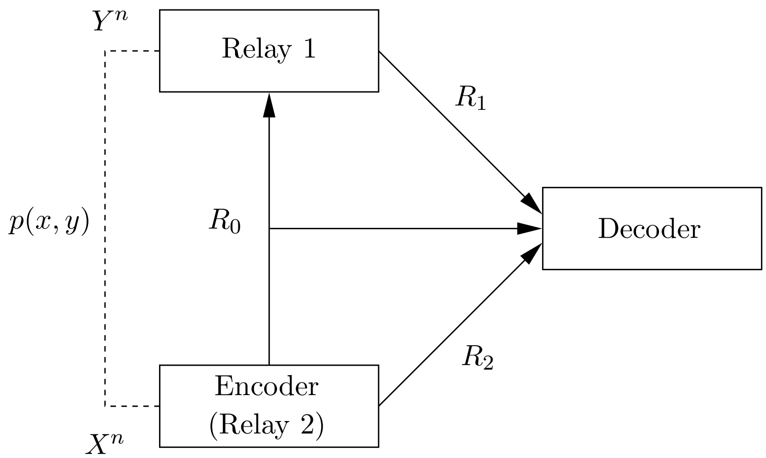

Figure 4.3: Kaspi-Berger rate distortion problem.

mitter, a noisy relay, a noiseless relay and a receiver. We proved the capacity of this class of diamond channels by providing an achievable scheme and a converse. The capacity we showed is strictly smaller than the cut-set bound. Our result also shows the optimality of a combination of DAF and CAF at the noisy relay node. This is the first example where a combination of DAF and CAF is shown to be capacity achieving. Finally, we noted that there exists a duality between this diamond channel coding problem and the Kaspi-Berger source coding problem.

Chapter 5

Conclusion

In this dissertation, we studied correlation and cooperation, two important phenomena that arise in the context of multi-user information theory. In wireless networks, correlation mainly originates from the correlated observations of different users, while cooperation is enabled by the wireless medium, which lets third-party users obtain part of the information from the transmitter in order to help deliver it to the destination.

We first studied the effects of source correlation in multi-user networks. More specifically, we studied the distributed source and channel coding problem for correlated sources, e.g., multiple access channel with correlated sources and multi-terminal rate-distortion problem. In these problems, it is often needed to characterize the joint probability distribution of a pair of random variables satisfying an n-letter Markov chain. An exact characterization of such probability distributions is intractable. We proposed a new data processing inequality, which provided us a single-letter necessary condition for the n-letter Markov chain. Our new data processing inequality yielded outer bounds for the the multiple access channel with correlated sources and multi-terminal rate-distortion region.

Next, we investigated the role of correlation in cooperative multi-user networks. We considered the basic three-node relay channel, which is the simplest model for cooperative communications. We proposed a new coding scheme for the relay channel, which is in the form of block Markov coding and is based on preserving the correlation in the channel inputs from the transmitter and the relay. The analysis of the error events provided us with three conditions containing mutual information expressions involving infinite letters of the underlying random process. We lower bounded these mutual informations to obtain three single-letter conditions. We showed that the achievable rates with the classical CAF scheme is a special case of the achievable rates in our new coding scheme. We therefore concluded that our proposed coding scheme yields potentially larger rates than the CAF scheme.

Finally, we focused on the diamond channel, which is a four-node cooperative communication network. We studied a special class of diamond channels, which consists of a transmitter, a noisy relay and a noiseless relay, and a destination. We determined the capacity of this class of diamond channels by providing an achievable scheme and a converse. The capacity we showed is strictly smaller than the cut-set bound. Our result also showed the optimality of a combination of DAF and CAF at the noisy relay node. This is the first example where a combination of DAF and CAF is shown to be capacity achieving. We also uncovered a duality between this diamond channel coding problem and the Kaspi-Berger source coding problem.

BIBLIOGRAPHY

[1] R. Ahlswede. Multi-way communication channels. In *Proc. Second International Symp. on Information Transmission*, Tsahkadsor, Armenia, USSR, 1971.

[2] R. Ahlswede and T. S. Han. On source coding with side information via a multiple-access channel and related problems in multi-user information theory. *IEEE Trans. Inform. Theory*, 29(3):396–412, 1983.

[3] M. Aleksic, P. Razaghi, and W. Yu. Capacity of a class of modulo-sum relay channels. *Submitted to IEEE Transactions on Information Theory*, 2007, http://arxiv.org/pdf/0704.3591.

[4] T. Berger. *Multi-terminal source coding.* Chapter in The Information Theory Approach to Communications (G. Longo, ed.), Springer-Verlag, 1978.

[5] P. P. Bergmans. A simple converse for broadcast channels with additive white Gaussian noise. *IEEE Trans. Inform. Theory*, 20:279–280, 1974.

[6] A. Berman and R. J. Plemmons. *Nonnegative matrices in the mathematical sciences.* Academic Press, 1979.

[7] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 2 edition, 1999.

[8] T. M. Cover and A. El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inform. Theory*, 25:572–584, Sep. 1979.

[9] T. M. Cover, A. El Gamal, and M. Salehi. Multiple access channel with arbitrarily correlated sources. *IEEE Trans. Inform. Theory*, 26:648–657, Nov. 1980.

[10] T. M. Cover and C. S. K. Leung. An achievable rate region for the multiple-access channel with feedback. *IEEE Trans. Inform. Theory*, 27(3):292–298, 1981.

[11] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.

[12] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. Inform. Theory*, 24(3):339–348, 1978.

[13] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[14] R. Dabora and S. Servetto. On the role of estimate-and-forward with time-sharing in cooperative communications. *Submitted to IEEE Transactions on Information Theory*, 2006, http://cn.ece.cornell.edu/publications/papers/20060529/pp.pdf.

[15] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros. Capacity of wireless erasure networks. *IEEE Trans. Inform. Theory*, 52(3):789–804, 2006.

[16] G. Dueck. A note on the multiple access channel with correlated sources. *IEEE Trans. Inform. Theory*, 27:232–235, Mar. 1981.

[17] A. El Gamal and M. Aref. The capacity of the semideterministic relay channel. *IEEE Trans. Inform. Theory*, 28(3):536, 1982.

[18] A. El Gamal and S. Zahedi. Capacity of a class of relay channels with orthogonal components. *IEEE Trans. Inform. Theory*, 51(5):1815–1817, 2005.

[19] R. G. Gallager. Capacity and coding for degraded broadcast channels. *Problems of information transmission*, pages 185–193, Jul.- Sep. 1974.

[20] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge, 1985.

[21] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge, 1991.

[22] K. B. Housewright. *Source Coding Studies for Multiterminal Systems*. PhD thesis, University of California, Los Angeles, 1977.

[23] A. H. Kaspi and T. Berger. Rate-distortion for correlated sources with partially separated encoders. *IEEE Trans. Inform. Theory*, 28(6):828–840, 1982.

[24] Y. H. Kim. Capacity of a class of deterministic relay channels. *IEEE Trans. Inform. Theory*, 53(3):1328–1329, 2008.

[25] G. Kramer, M. Gastpar, and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Trans. Inform. Theory*, 51(9):3037–3063, September 2005.

[26] G. Kramer, S. M. S. Tabatabaei Yazdi, and S. A. Savari. Network coding on line networks with broadcast. In *Proc. Conf. Inf. Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2008.

[27] A. Lapidoth and S. Tinguely. Sending a bi-variate Gaussian source over a Gaussian MAC. In *Proc. IEEE International Symp. on Information Theory (ISIT)*, Seattle, WA, Jul. 2006.

[28] H. Liao. *Multiple-access channel*. PhD thesis, University of Hawaii, 1972.

[29] K. Marton. The structure of isomorphisms of discrete memoryless correlated sources. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 56(3):317–327, 1981.

[30] S. S. Pradhan, S. Choi, and K. Ramchandran. A graph-based framework for transmission of correlated sources over multiple access channels. *IEEE Trans. Inform. Theory*, 53(12):4583–4604, 2007.

[31] N. Ratnakar and G. Kramer. The multicast capacity of deterministic relay network with no interference. *IEEE Trans. Inform. Theory*, 52(6):2425–2432, 2006.

[32] B. E. Schein. *Distributed Coordination in Network Information Theory*. PhD thesis, Massachusetts Institute of Technology, 2001.

[33] S. Servetto. The region of achievable rates for multiterminal source coding. *Submitted to the IEEE Transactions on Information Theory*, 2006, http://cn.ece.cornell.edu/publications/papers/20060228/pp1.pdf.

[34] C. E. Shannon. Two-way communication channels. In *Proc. 4th Berkeley Symp. Prob. Stat.*, Berkeley, CA, 1961.

[35] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inform. Theory*, 49(4):471–480, 1973.

[36] G. W. Stewart. On the early history of the singular value decomposition. *SIAM Review*, 35:551–566, Dec. 1993.

[37] S. Y. Tung. *Multiterminal Source Coding*. PhD thesis, Cornell University, 1978.

[38] E. C. van der Meulen. Three-terminal communication channels. *Adv. App. Prob.*, 3:120–154, 1971.

[39] A. B. Wagner and V. Anantharam. An improved outer bound for the multiterminal source coding problem. *IEEE Trans. Inform. Theory*, 54(5):1919–1937, 2008.

[40] F. M. J. Willems. The feedback capacity for a class of discrete memoryless multiple access channels. *IEEE Trans. Inform. Theory*, 28(1):93–95, 1982.

[41] H. S. Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal on Applied Mathematics*, 28:100–113, Jan. 1975.

[42] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inform. Theory*, 22(1):1–10, 1976.