Investigation on the Compress-and-Forward Relay Scheme

by

Jie Zhang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The relay channel plays an integral role in network communication systems. An intermediate node acts as a relay to facilitate the communication between the source and the destination. If the rate of codewords is less than the capacity of the source-relay link, the relay can decode the source's messages and forward them to the destination. On the contrary, if the rate of codewords is greater than the capacity of the source-relay link, the relay cannot decode the messages. Nevertheless, the relay can still compress its observations and then send them to the destination. Obviously, if the relay-destination link is of a capacity high enough such that the relay's observations can be losslessly sent to the destination, then the maximum message rate can be achieved as if the relay and the destination can jointly decode. However, when the relay-destination link is of a limited capacity such that the relay's observation cannot be losslessly forwarded to the destination, then what is the maximum achievable rate from the source to the destination? This problem was formulated by Cover in another perspective [7], *i.e.*, what is the minimum rate of the relay-destination link such that the maximum message rate can be achieved?

We try to answer this Cover's problem in this thesis. First, a sufficient rate to achieve the maximum message rate can be obtained by Slepian-Wolf coding, which gives us an upper bound on the optimal relay-destination link rate. In this thesis, we show that under some channel conditions, this sufficient condition is also necessary, which implies that Slepian-Wolf coding is already optimal. Hence, the upper bound meets exactly the minimum value of the required rate. In our approach, we start with the standard converse proof. First, we present a necessary condition for achieving the maximum message rate in the single-letter form. Following the condition, we derive a theorem, which is named as "single-letter criterion". The "single-letter criterion" can be easily utilized to verify different channels. Then we show that for two special cases: when the source-relay link and the source-destination link of the relay channel are both binary symmetric channels (BSCs), and when they are both binary erasure channels (BECs), Slepian-Wolf coding is optimal in achieving the maximum message rate. Moreover, the maximum message rates of these two special channels are also calculated in this thesis.

Acknowledgements

I would like to specially thank my supervisors Prof. Liang-liang Xie and Prof. En-hui Yang. When I began my graduate study in ECE department two years ago, I did not have much knowledge in the field of information theory. I have encountered countless obstacles and difficulties in the past two years. Prof. Liang-liang Xie and Prof. En-hui Yang have been very patient, helpful for giving me suggestions and encouragement through the research. I am very graceful to have met such benevolent and caring supervisors. They not only give me constructive suggestions in study, but also guide me to find the right direction of my whole career.

I would like to thank my committee members, Prof. Zhou Wang and Prof. Pin-han Ho, for their valuable suggestions on my thesis.

I would also like to thank to all my colleagues for their continuous supports. I would like to thank Dake He and Jin Meng for their encouragement and suggestions in my study. I would like to thank Lin Zheng who is like my elder sister and gives me lots of help.

I would like to thank all the friends and families, who have helped me and supported me in one way or another for all these years. I would like to thank Fei Huo who spent lots of time in helping me with the writing correction of my thesis.

Finally, I would also like to thank my parents and my little brother, for their genuine love, and endless support in my life and study. For all these years, they have always believed in me and have taken very good care of me, lead me to where I am today.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problem and Motivation

A relay channel [20] models communication between one sender and one receiver helped by one or more intermediate nodes. It can also be regarded as a combination of broadcast channel [4] and multiple access channel [2]. Because of its special behaviour, the general relay capacity is difficult to characterize. Lots of work have been done to study the properties of the relay channel in both the discrete case [18, 28] and the Gaussian case [14, 23]. Some results are given under certain assumptions [6, 1, 15, 10, 12, 17, 26, 27]. In this thesis, we will investigate the relay channel directly following Cover's open problem [7] on the capacity of the relay channel.

1.1.1 Cover's Open Problem on the Capacity of the Relay Channel

First, we introduce Cover's open problem [7]. Consider a simple discrete memoryless relay channel as shown in Figure 1.1, where X is a codeword, Y_1 is the relay's observation, Y_2 is the destination's observation, and R_0 is the error free communication rate from the relay to the destination, by which, the relay can forward its observation to the destination. In this thesis, we call R_0 the relay rate. Here Y_1, Y_2 are conditionally independent and conditionally identically distributed given X, and hence $p(y_1, y_2|x) = p(y_1|x)p(y_2|x)$. Furthermore,

- $W_1 \in \{1, ..., 2^{nR}\}$ is the message we wish to reliably send over this relay channel;

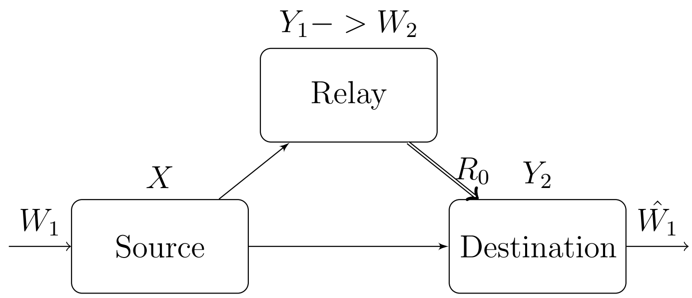

Figure 1.1: The relay channel with a digital link

- a $(2^{nR}, n)$ code for this channel is obtained from an encoding function $f_1 : \{1, ..., 2^{nR}\} \rightarrow \mathcal{X}^n$, and $X^n = f_1(W_1)$;
- $W_2 \in \{1, ..., 2^{nR_0}\}$ is the input of the relay-destination link with $W_2 = f_2(Y_1^n)$, where $f_2 : \mathcal{Y}_1^n \rightarrow \{1, ..., 2^{nR_0}\}$ is a relay encoding function;
- $\hat{W}_1 = g(Y_2^n, W_2)$ is the estimation of the message W_1, where $g : \{1, ..., 2^{nR_0}\} \times \mathcal{Y}_2^n \rightarrow \{1, ..., 2^{nR}\}$ is a decoding function.

The probability of error is given by

$$P_e^{(n)} = P\{\hat{W}_1 \neq W_1\}$$

where W_1 is uniformly distributed over $\{1, ..., 2^{nR}\}$ and

$$p(w_1, y_1, y_2) = 2^{-nR} \prod_{i=1}^{n} p(y_{1i}|x_i(w_1)) \prod_{i=1}^{n} p(y_{2i}|x_i(w_1)).$$

Let $C(R_0)$ be the supremum of the achievable rate R for a given R_0 , that is, the supremum of the rates R for which $P_e^{(n)}$ can be made to tend to zero as $n \rightarrow \infty$.

We note the following facts:

1. $C(0) = \sup_{p(x)} I(X; Y_2)$.
2. $C(\infty) = \sup_{p(x)} I(X; Y_1, Y_2)$.
3. $C(R_0)$ is a nondecreasing function of R_0.

In light of this, Cover proposed the following question: "What is the critical value of R_0 such that $C(R_0)$ first equals $C(\infty)$?". This is equivalent to finding a suitable R_0^* which satisfies:

$$R_0^* := \inf\{R_0 : C(R_0) = C(\infty) = \sup_{p(x)} I(X; Y_1, Y_2)\}. \tag{1.1}$$

It is the smallest relay rate required for the relay-destination communication while the maximum rate $C(\infty) = \sup_{p(x)} I(X; Y_1, Y_2)$ can be achieved.

In this thesis, we are interested in the compress-and-forward scheme. If the compression at the relay is lossless, we can recover Y_1 at the destination from the knowledge of W_2 and Y_2. Moreover, using Slepian-Wolf coding we can get a compression rate $R_0 = H(Y_1|Y_2)$ at the relay by treating Y_2 as the side information. Thus, we have a sufficient relay rate $R_0 = H(Y_1|Y_2)$ which can maximize the rate $C(R_0)$ to be $C(\infty) = \sup_{p(x)} I(X; Y_1, Y_2)$. Then, R_0^* in Equation (1.1) satisfies:

$$R_0^* \leq \min_{p(x)} H(Y_1|Y_2) \text{ and } p(x) \in [0, 1].$$

Our target is to maximize the rate instead of recovering the relay observation at the destination. Hence two questions arise:

1. Slepian-Wolf coding already provides an optimal compression rate for lossless coding [25, 5, 16, 19]. If lossy compression is applied at the relay, can we still find a R_0 to fulfil the requirement of $C(R_0) = C(\infty) = \sup_{p(x)} I(X; Y_1, Y_2)$?

2. Is it possible that a smaller relay rate $R_0 < H(Y_1|Y_2)$ exists, such that $C(R_0) = C(\infty) = \sup_{p(x)} I(X; Y_1, Y_2)$?

To answer these two questions, we will analyze two special relay channels in this thesis.

1.2 Organization of the Thesis and Contribution

The content of this thesis is organized as follows:

In Chapter 2, we will introduce some fundamental blocks in information theory and some key theorems used throughout this thesis. First, some basic definitions and theorems in information theory are given, such as the concepts of entropy, jointly typical sequences and channel capacity theorem. Also Fano's inequality, log sum inequality and data processing inequality are presented. These theorems will play vital roles in my later proofs. Then a quick review of the relay channel, the two different schemes: decode-and-forward and compress-and-forward will be discussed separately. The achievable rate of Slepian-Wolf coding is also given in this chapter.

In Chapter 3, first, we present a sufficient relay rate $R_0 = H(Y_1|Y_2)$ to achieve the maximum message rate obtained by Slepian-Wolf coding. Using Fano's inequality, a necessary condition for achieving the maximum rate is given in the single-letter form. Following the condition, a theorem called "single-letter criterion" is derived, which can be used to verify in certain condition Slepian-Wolf coding is optimal. In the end, we prove $R_0 = H(Y_1|Y_2)$ is the smallest relay rate required for maximizing the rate in two special cases: when the source-relay link and the source-destination link are both BSC channels, and when they are both BEC channels. From now on, we will refer to them as the BSC relay channel and the BEC relay channel respectively.

Chapter 4 will draw some conclusions and give some potential directions for future work.

Chapter 2

Preliminaries

2.1 Basic Knowledge in Information Theory

In this section, we introduce some fundamental building blocks and important theorems in information theory which will be used throughout this thesis. For detailed references, please refer to [8] and [11] .

2.1.1 Basics in Information Theory

Suppose X is a random variable (r.v.) with probability mass function (pmf) $p(x) = Pr\{X = x\}, x \in \mathcal{X}$, $\mathcal{X}$ is the alphabet. In information theory, entropy is used to measure the uncertainty of a r.v.. Mathematically, it is the expectation of the information contained in a message.

Definition 1 *The entropy $H(X)$ of a discrete r.v. X is defined by*

$$H(X) = -\sum_{x\in\mathcal{X}} p(x) \log p(x).$$

The value of entropy is nonnegative, since $0 \leq p(x) \leq 1$.

Joint entropy $H(X, Y)$ is a measure of the uncertainty of a set of r.v.s, *i.e.* $\{X, Y\}$.

Definition 2 *The joint entropy $H(X, Y)$ of a pair of discrete r.v.s (X, Y) with a joint distribution $p(x, y)$ is defined as*

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y).$$

Conditional entropy $H(Y|X)$ is understood as the amount of uncertainty of r.v. Y given the information of r.v. X.

Definition 3 *If $(X,Y) \sim p(x,y)$, then the conditional entropy $H(Y|X)$ is defined as*

$$H(Y|X) = \sum_{x\in\mathcal{X}} p(x)H(Y|X=x) = -\sum_{x\in\mathcal{X}} p(x)\sum_{y\in\mathcal{Y}} p(y|x)\log p(y|x).$$

$H(Y|X)$ refers to the average entropy of Y condition on X, and $H(X|Y) \leq H(X), H(X|X) = 0$.

Next we introduce the chain rule for joint entropy and it will be used in our proofs.

Theorem 1 *(Chain rule):*

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y).$$

Extending the chain rule to n vector,

$$H(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, ..., X_1).$$

Thus, we have $H(X_1, X_2, ..., X_n) \leq \sum_{i=1}^{n} H(X_i)$, and equality holds if $X_1, X_2, ..., X_n$ are independent.

Mutual information $I(X;Y)$ is the measure of mutual dependence of r.v.s X and Y as follows.

Definition 4 *Consider two r.v.s X and Y with a joint pmf $p(x,y)$ and marginal pmfs $p(x)$ and $p(y)$. The mutual information $I(X;Y)$ is defined as*

$$I(X;Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}.$$

Some properties of mutual information are given:

1. $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$
2. $I(X;X) = H(X) - H(X|X) = H(X)$
3. $I(X;Y) \geq 0$.
4. By the chain rule, mutual information can be expanded:

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z) = I(X;Y) + I(X;Z|Y).$$

The relations of $H(X), H(X,Y), H(X|Y), H(Y|X)$ and $I(X;Y)$ are shown in Figure 2.1.

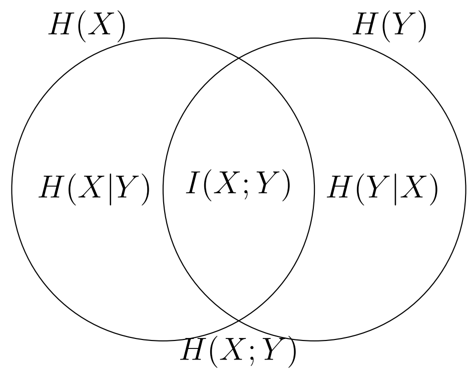

Figure 2.1: Entropy, joint entropy, conditional entropy and mutual information

2.1.2 Markov Chain, Fano's Inequality, Log Sum Inequality and Data Processing Inequality

Definition 5 *R.v.s X, Y, Z are said to form a Markov chain in this order (denoted by $X \to Y \to Z$) if the conditional distribution of Z depends only on Y and is conditionally independent of X. Specifically, X, Y and Z form a Markov chain $X \to Y \to Z$ if the joint probability mass function can be written as*

$$p(x, y, z) = p(x)p(y|x)p(z|y).$$

Some properties of Markov chain are given:

1. $X \to Y \to Z$ if and only if (iff) X and Z are conditionally independent given Y, then

$$p(x, z|y) = p(x|y)p(z|y).$$

2. $X \to Y \to Z$ implies that $Z \to Y \to X$.

3. If $Z = f(Y)$, then $X \to Y \to Z$.

We give an example of Markov chain.

Example 1 *Let X be a r.v. with a distribution $p(x)$. We observe a r.v. Y which is related to X by the conditional distribution $p(y|x)$. From Y, we calculate a function $g(Y) = \hat{X}$, which is an estimate of X. Then $X \to Y \to \hat{X}$ forms a Markov chain.*

Let X, Y and $\hat{X}$ be r.v.s defined in Example 1. Then we define the probability of error P_e to be

$$P_e = P\{\hat{X} \neq X\}.$$

Theorem 2 *(Fano's inequality):*

$$H(X|Y) \leq H(P_e) + P_e \log(|\mathcal{X}| - 1),$$

and $H(P_e) = -P_e \log P_e - (1 - P_e) \log(1 - P_e)$. *This inequality can be weakened to*

$$H(X|Y) \leq 1 + P_e \log|\mathcal{X}|, \textit{ furthermore,}$$
$$P_e \geq \frac{H(X|Y)-1}{\log|\mathcal{X}|}$$

Fano's inequality relates the conditional entropy $H(X|Y)$ to the error probability of estimating X from Y. It is very important for the converse proof of channel coding theorems and also in our work. Another two useful inequalities *log sum inequality* and *data processing inequality* are given in the following:

Theorem 3 *(Log sum inequality): For non-negative numbers,* $a_1, a_2, ..., a_n$ *and* $b_1, b_2, ..., b_n$,

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i\right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

with equality iff $\frac{a_i}{b_i} = const.$

Theorem 4 *(Data processing inequality): If* $X \to Y \to Z$, *then* $I(X;Y) \geq I(X;Z)$.

2.1.3 Typical Set and Jointly Typical Sequences

Asymptotic Equipartition Property (AEP) and joint AEP are the keys in the proofs of shannon coding theorems. AEP is the result of the weak law of large numbers.

Theorem 5 *(AEP): If* $X_1, X_2, ..., X_n$ *are independent and identically distributed (i.i.d.)* $\sim p(x)$, *then*

$$-\frac{1}{n}\log p(X_1, X_2, ..., X_n) \to H(X) \text{ in probability.}$$

Definition 6 *The typical set* $A_\epsilon^{(n)}$ *with respect to* $p(x)$ *is the set of sequences* $(x_1, x_2, ..., x_n) \in \mathcal{X}^n$ *with the following properties:*

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, x_2, ..., x_n) \leq 2^{-n(H(X)-\epsilon)}.$$

And the set $A_\epsilon^{(n)}$ has some properties:

Theorem 6
1. If $(x_1, x_2, ..., x_n) \in A_\epsilon^{(n)}$*, then* $H(X) - \epsilon \leq -\frac{1}{n}\log p(x_1, x_2, ..., x_n) \leq H(X) + \epsilon$.
2. $P\{A_\epsilon^{(n)}\} > 1 - \epsilon$*, for* n *sufficiently large.*
3. $|A_\epsilon^{(n)}| \leq 2^{-n(H(X)+\epsilon)}$*, where* $|A_\epsilon^{(n)}|$ *denotes the number of elements in the set* $A_\epsilon^{(n)}$.
4. $|A_\epsilon^{(n)}| \geq (1-\epsilon)2^{-n(H(X)-\epsilon)}$*, for* n *sufficiently large.*

Definition 7 *The set* $A_\epsilon^{(n)}$ *of jointly typical sequences* $\{(x^n, y^n)\}$ *with respect to the distribution* $p(x, y)$ *is the set of* n*-sequences with empirical entropies* ϵ *-close to the true entropies, i.e.,*

$$A_\epsilon^{(n)} = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n :$$

$$|-\tfrac{1}{n}\log p(x^n) - H(X)| < \epsilon,$$

$$|-\tfrac{1}{n}\log p(y^n) - H(Y)| < \epsilon,$$

$$|-\tfrac{1}{n}\log p(x^n, y^n) - H(X, Y)| < \epsilon\},$$

where

$$p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i).$$

Theorem 7 *(Joint AEP): Let* (X^n, Y^n) *be sequences of length* n *drawn i.i.d. according to* $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$*. Then*
1. $P((X^n, Y^n) \in A_\epsilon^{(n)}) \to 1$ *as* $n \to \infty$.
2. $|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$.

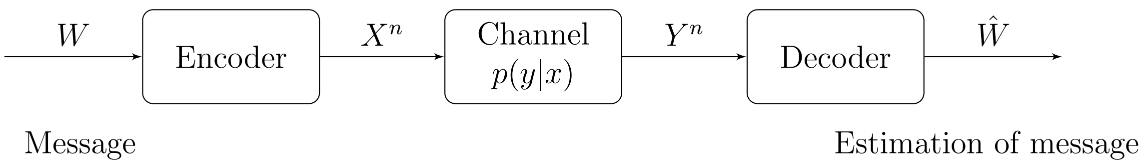

Figure 2.2: A communication channel

2.1.4 Channel Capacity and Channel Coding Theorem

Figure 2.2 illustrates the basic model of a communication system. A message W, drawn from the index set $\{1, 2, ..., M\}$, generates the codeword $X^n(W)$, which is received as a random sequence $Y^n \sim p(y^n|x^n)$ by the receiver. The receiver then estmates the index W by an appropriate decoding rule $\hat{W} = g(Y^n)$. An error occurs if $\hat{W}$ is not the same index W that was transmitted. Channel capacity is the tightest upper bound on the amount of information that can be *reliably transmitted* over a communications channel. *Reliable transmission* means the message transmitted through the channel can be recovered with arbitrarily small error probability [11]. Some basic definitions are given before we give the mathematical definition of channel capacity.

Definition 8 *A discrete channel, denoted by $(\mathcal{X}, p(y|x), \mathcal{Y})$, consists of two finite sets $\mathcal{X}$ and $\mathcal{Y}$ and a collection of pmfs $p(y|x)$, one for each $x \in \mathcal{X}$, such that for every x and y, $p(y|x) \geq 0$, and for every x, $\sum_y p(y|x) = 1$, with the interpretation that X is the input and Y is the output of the channel.*

Definition 9 *The n-th extension of the discrete memoryless channel (DMC) is the channel $(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y}^n)$, where*

$$p(y_k|x^k, y^{k-1}) = p(y_k|x_k), k = 1, 2, \cdots, n,$$

where $x^k = x_1 \cdots x_{k-1}x_k$, and $y^{k-1} = y_1 \cdots y_{k-2}y_{k-1}$. If there is no feedback in the channel model, i.e., if the past output symbols do not effect the input symbols, namely, $p(x_k|x^{k-1}, y^{k-1}) = p(x_k|x^{k-1})$, then we have the following reduced form for DMC

$$p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i).$$

Definition 10 *We define the "information" channel capacity of a DMC as*

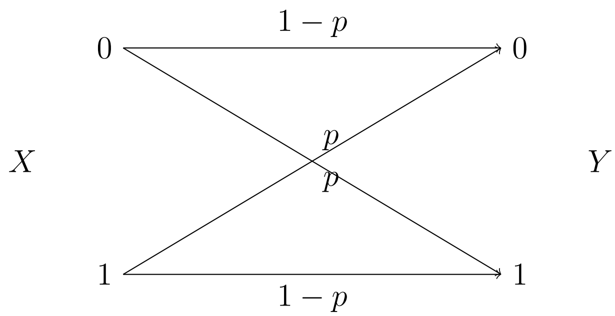

Figure 2.3: BSC channel

$$C = max_{p(x)} I(X;Y)$$

where the maximum is taken over all possible input distributions $p(x)$*.*

Example 2 *(Capacity of BSC channel (Figure 2.3)): BSC is a channel with binary input and binary output and error probability* p*. If* $X \in \{0,1\}$ *is the input r.v. and* $Y \in \{0,1\}$ *is the output r.v., then the channel is characterized by the condition probabilities:* $P(Y=0|X=0) = P(Y=1|X=1) = 1-p$ *and* $P(Y=1|X=0) = P(Y=0|X=1) = p$*, assuming* $0 \le p < 1/2$*. Calculate the mutual information:*

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_x p(x) H(Y|X=x) \\ &= H(Y) - \sum_x p(x) H(p) \\ &= H(Y) - H(p) \\ &\le 1 - H(p), \end{aligned}$$

equality is achieved when Y*'s distribution is uniform, equivalently, the input distribution is uniform. Hence the information capacity of a BSC channel with parameter* p *is*

C = 1- H(p) bits.

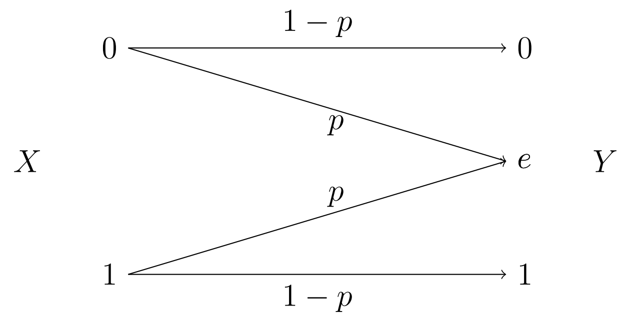

Figure 2.4: BEC channel

Example 3 *(Capacity of BEC (Figure 2.4)): BEC is a channel with binary input and binary output and erasure probability p. Let $X \in \{0, 1\}$ be the input r.v. and $Y \in \{0, 1, e\}$ be the output r.v., where e is the erasure symbol. Then the channel is characterized by the condition probabilities: $P(Y = 0|X = 0) = P(Y = 1|X = 1) = 1 - p, P(Y = 1|X = 0) = P(Y = 0|X = 1) = 0$ and $P(Y = e|X = 0) = P(Y = e|X = 1) = p$, assuming $0 \le p < 1/2$. Suppose $P(X = 0) = \pi$, then $P(Y = 0) = \pi(1 - p), P(Y = e) = p$ and $P(Y = 1) = (1 - \pi)(1 - p)$. Calculate the mutual information:*

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(Y) - H(p) \\
&= H(\pi(1-p), p, (1-\pi)(1-p)) - H(p) \\
&= (1-p)H(\pi) + H(p) - H(p) \\
&= (1-p)H(\pi) \\
&\le 1-p,
\end{aligned}
$$

where $H(\pi(1-p), p, (1-\pi)(1-p)) = -\pi(1-p)\log\pi(1-p) - p\log p - (1-\pi)(1-p)\log(1-\pi)(1-p)$, and equality is achieved when $\pi = 1/2$, i.e., input distribution is uniform. Hence the information capacity of a BEC channel with parameter p is

$$C = 1\text{- } p \text{ bits.}$$

Definition 11 *An (M, n) code for a channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of the following:*

- *An index set $1, 2, ..., M$.*

- *An encoding function* $f : \{1, 2, ..., M\} \to \mathcal{X}^n$*, yielding codewords* $X^n(1), X^n(2), .., X^n(M)$*. The set of codewords is called the codebook.*
- *A decoding function*

$$g : y^n \to \{1, 2, ..., M\},$$

is a deterministic rule which assigns a guess to each possible received vector.

Let $\lambda_i = P(g(Y^n) \neq i | X^n = X^n(i)) = \sum_{y^n} p(y^n|x^n(i))I(g(y^n) \neq i)$ *be the conditional probability of error given that index* i *was sent, where* $I(.)$ *is the indicator function.* $\lambda^n = max_{\{i \in 1,2,...,M\}} \lambda_i$ *is the maximal probability of error and* $P_e^{(n)} = \frac{1}{M}\sum_{i=1}^{M} \lambda_i$ *is average probability of error.*

Theorem 8 *(The channel coding theorem): All rates below capacity* C *are achievable. Specifically, for every rate* $R < C$*, there exists a sequence of* $(2^{nR}, n)$ *codes with maximum probability of error* $\lambda^{(n)} \to 0$*.*

Conversely, any sequence of $(2^{nR}, n)$ *codes with* $\lambda^n \to 0$ *must have* $R \leq C$*.*

The channel coding theorem is perhaps the basic and most important theorem in information theory: channel capacity is achievable. Claude Shannon used a number of new ideas in his original 1948 paper[24] to prove this. These ideas can be summarized as follows.

- Allowing an arbitrarily small but non-zero probability of error,
- using the channel many times in succession, so that the law of large numbers comes into effect, and
- calculating the average of the probability of error over a random choice of codebooks, which symmetrizes the probability, and which can then be used to show the existence of at least one good code.

2.2 The Relay Channel

Fig 2.5 shows the general model of the relay channel with only one relay node. The channel consists of four finite sets $\{\mathcal{X}, \mathcal{X}_1, \mathcal{Y}_1, \mathcal{Y}_2\}$ and a collection of pmfs $p(y_1, y_2|x, x_1)$. x is the

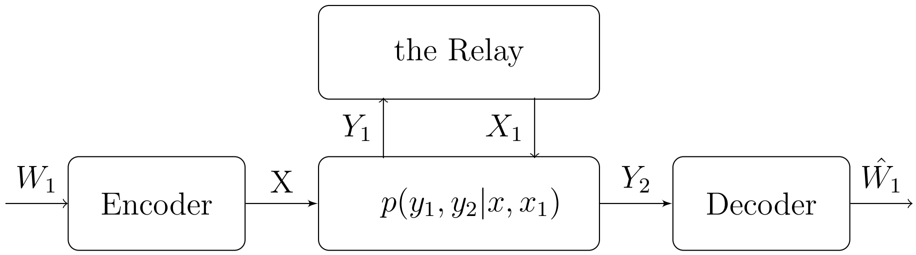

Figure 2.5: The general relay channel

input to the channel, y_2 is the observation of the destination, y_1 is the relay's observation and x_1 is the input symbol for the relay-destination link chosen by the relay. The general relay channel capacity is difficult to determine. Results are only known for a few special cases, e.g., the physical degraded relay channel [8, 6] and the Gaussian relay channel [14, 23, 1](asymptotic capacity). In order to investigate the relay capacity, two important coding schemes have been developed in [6]. One is the decode-and-forward scheme, the other is the compress-and-forward scheme.

2.2.1 Decode-and-Forward

In the decode-and-forward scheme, block Markov encoding is used. The relay decodes the source's message in the current block, and then forwards it to the destination in the next block.

Theorem 9 *(Decode-and-Forward): The achievable rate for the relay channel by decode-and-forward is*

$$R < max_{p(x,x_1)} min\{I(X;Y_1|X_1), I(X,X_1;Y_2)\}$$

for some $p(x, x_1)$.

2.2.2 Compress-and-Forward

In our work, we focus on the compress-and-forward scheme. The relay helps the destination to decode the original message by compressing and forwarding its observation to the destination. Some important work have been done on investigating the optimal compress-and-forward scheme [26, 27, 13].

Theorem 10 *(Compress-and-Forward): The achievable rate R of the relay channel by compress-and-forward is:*

$$R < I(X; \hat{Y}_1, Y_2 | X_1)$$

with the constraint:

$$I(X_1; Y_2) > I(Y_1; \hat{Y}_1 | X_1, Y_2)$$

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

2.3 Slepian-Wolf Coding

2.3.1 Coding Procedure Using Random Binning

The coding procedure [8] using random binning is very popular. Here we present the scheme of encoding of single source X^n as shown in Figure 2.6.

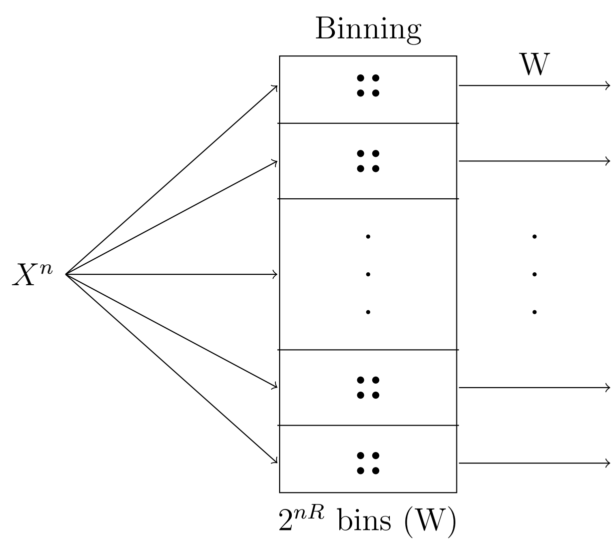

Figure 2.6: Random binning

Suppose there are 2^{nR} bins, all the sequences X^n are randomly thrown into the bins. The set of sequences are assigned to the same index $W \in \{1, 2, ..., 2^{nR}\}$, if they are laying in the same bin. For decoding the source from the bin index W, we look for a typical X^n sequence in the bin. If there is one and only one typical sequence in the bin, we declare the sequence $\hat{X}^n$ to be the estimation of the source sequence. Error will occur if there is none or more than one typical sequence in this bin. However if the bin number is large enough (much larger than the total number of typical sequences), then the probability that there is more than one typical sequence in each bin is very small. Thus the error probability is arbitrarily small.

Let $f(X^n) = W$ be the bin index corresponding to X^n and $g(.)$ is the decoding function that $\hat{X}^n = g(f(X^n))$ is the estimation of X^n. Next we introduce the conventional method to analyze the probability of error [8] (averaged over the random choice of codes f):

$$
\begin{aligned}
P(g(f(X)) &\neq X) \leq P(X) \notin A_\epsilon^{(n)}) + \sum_x P(\exists x' \neq x : x' \in A_\epsilon^{(n)}, f(x') = f(x))p(x) \\
&\leq \epsilon + \sum_x \sum_{x' \in A_\epsilon^{(n)}, x' \neq x} P(f(x') = f(x))p(x) \\
&\leq \epsilon + \sum_x \sum_{x' \in A_\epsilon^{(n)}} 2^{-nR} p(x) \\
&= \epsilon + \sum_{x' \in A_\epsilon^{(n)}} 2^{-nR} \sum_x p(x) \\
&\leq \epsilon + \sum_{x' \in A_\epsilon^{(n)}} 2^{-nR} \\
&\leq \epsilon + 2^{n(H(X)+\epsilon)} 2^{-nR} \\
&\leq 2\epsilon
\end{aligned}
$$

if $R > H(X) + \epsilon$ and n is sufficiently large.

2.3.2 Achievable Rate of Slepian-Wolf Coding

Slepian-Wolf coding [25, 5, 19, 16] is a lossless source coding scheme used to encode two correlated sources. In this thesis we consider the basic form of Slepian-Wolf coding, encoding source $\{Y_{1i}\}_{i=1}^{\infty}$ with side information $\{Y_{2i}\}_{i=1}^{\infty}$ available at the decoder as shown in Fig 2.7.

Assume $\{(Y_{1i}, Y_{2i})\}_{i=1}^{\infty}$ is a jointly memoryless source with joint probability distribution $P_{Y_1Y_2}$ over $\mathcal{Y}_1^n \times \mathcal{Y}_2^n$. Random binning scheme is used in the coding procedure.

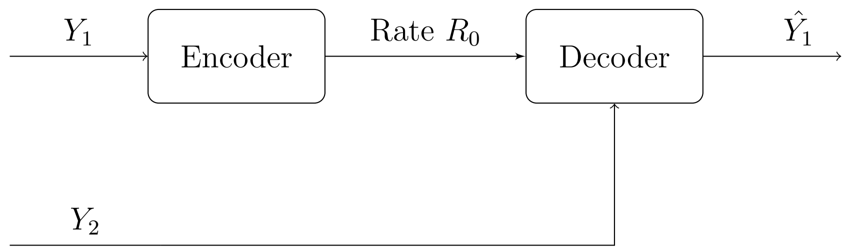

Figure 2.7: Source coding with side information at decoder

Random code generation: Independently assign every $y_1^n \in \mathcal{Y}_1^n$ to one of 2^{nR_0} bins according to a uniform distribution on $\{1, 2, ..., 2^{nR_0}\}$. Thus we have a mapping:

$$f : \mathcal{Y}_1^n \to \{1, 2, ..., 2^{nR_0}\}$$
$$\text{and } f(y_1^n) = w_2 \in \{1, 2, ..., 2^{nR_0}\} \text{ for } y_1^n \in \mathcal{Y}_1^n$$

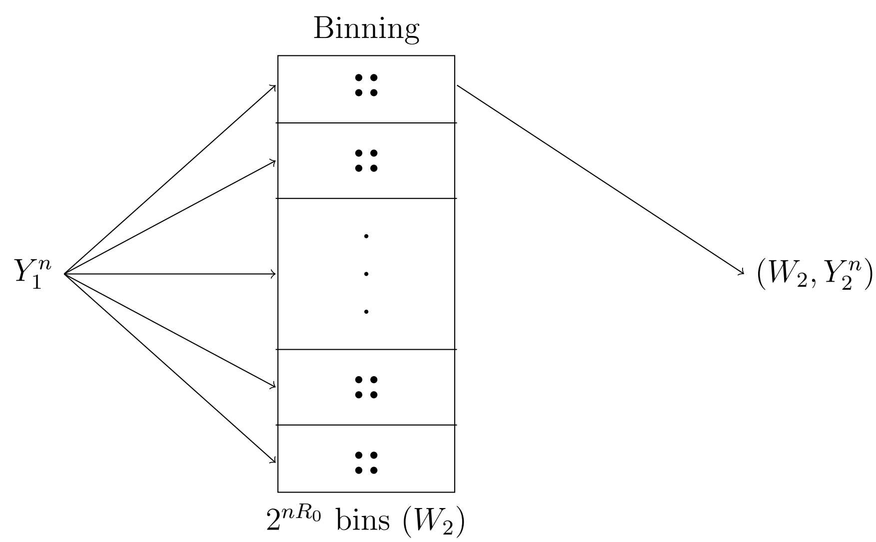

Figure 2.8: Binning scheme

Encoding: The sender sends the index w_2 of the bin in which $y_1^n \in \mathcal{Y}_1^n$ belongs to (Fig 2.8).
Decoding: Upon receiving w_2 and side information y_2^n, the decoder decodes the message by looking for y_1^n in bin w_2, such that $(y_1^n, y_2^n) \in A_\epsilon^{(n)}(Y_1, Y_2)$. If there is none or more than

one y_1^n in one bin, it declares an error.
Similarly analyzing the error probability as shown in Section 2.3.1, we have: the probability of error $P_e \rightarrow 0$, if $R_0 > H(Y_1|Y_2)$.

Chapter 3

Optimal Relay Rate for the BSC Relay Channel and the BEC Relay Channel

3.1 A Sufficient Relay Rate

In this section, we present a sufficient relay rate R_0 to maximize $C(R_0)$ by applying Slepian-Wolf coding.

The simple model (Figure 1.1) of discrete memoryless relay channel with a digital link from the relay to the destination is studied in this thesis.

For any $(2^{nR}, n)$ code of the relay channel, there are:

1. W_1 is the message drawn from the index set $\{1, 2, ..., 2^{nR}\}$;
2. f_1: $\{1, 2, ..., 2^{nR}\} \to \mathcal{X}^n$ is the encoding function at the sender, yielding codewords $X^n(1), X^n(2), ..., X^n(2^{nR})$;
3. Y_1^n is the relay's observation;
4. $f_2 : \mathcal{Y}_1^n \to \{1, 2, ..., 2^{nR_0}\}$ is the encoding function at the relay, and $W_2 = f_2(Y_1^n)$ is the input of the relay-destination link, $W_2 \in \{1, 2, ..., 2^{nR_0}\}$, $Y_1^n \in \mathcal{Y}_1^n$. w_2 is the realization of it and we also refer it as bin number in random binning scheme;
5. Y_2^n is the destination's observation;

6. $\hat{W}_1 = g(W_2, Y_2^n)$ is the estimation of message W_1, where g is the decoding function.

Y_1, Y_2 are conditionally independent given X, *i.e.*, $p(y_1, y_2|x) = p(y_1|x)p(y_2|x)$. Since W_2 is compressed of Y_1, so R_0 is also referred to as the compression rate of the relay. Our work is presented by looking for the minimum R_0 such that $C(R_0) = C(\infty) = \max_{p(x)} I(X; Y_1, Y_2)$.

Applying Slepian-Wolf coding in the classical compress-and-forward scheme, the compression rate

$$R_0 = \min_{p(x)} H(Y_1|Y_2) \text{ and } p(x) \in [0, 1],$$

is sufficient such that $C(R_0) = C(\infty) = \max_{p(x)} I(X; Y_1, Y_2)$. Thus, the minimum relay rate R_0^* in Equation (1.1) satisfies:

$$R_0^* \leq \min_{p(x)} H(Y_1|Y_2) \text{ and } p(x) \in [0, 1].$$

In the following sections, we will give a necessary condition to achieve the maximum rate $\max_{p(x)} I(X; Y_1, Y_2)$ in the single-letter form. Following this condition, we will derive a theorem named as "single-letter criterion", which states, in certain condition $\min_{p(x)} H(Y_1|Y_2)$ is the minimum relay rate as shown in Equation (1.1). Finally, we will prove that for the BSC relay channel and the BEC relay channel, the minimum relay rate R_0^* is equal to $\min_{p(x)} H(Y_1|Y_2)$.

3.2 Single-letter Criterion

Before the detailed discussion in this section, first we specify an assumption and all the later proofs and results are based on this assumption.

Assumption: Let $X^n = \{X_1, X_2, \cdots, X_n\}$ denote the random codeword that is uniformly chosen from the codebook used by the source. We assume $X_1, X_2, ..., X_n$ are i.i.d..

In this section, we will use Fano's inequality, data processing inequality and log sum inequality to derive a necessary condition for achieving the maximum rate $\max_{p(x)} I(X; Y_1, Y_2)$

in the single-letter form. For a $(2^{nR}, n)$ code defined in Section 3.1, we have

$$\begin{aligned}
nR &= H(W_1) = H(W_1|Y_2^n, W_2) + I(W_1; Y_2^n, W_2) \\
&\overset{(a)}{\leq} n\epsilon_n + I(W_1; Y_2^n, W_2) \\
&\overset{(b)}{\leq} n\epsilon_n + I(X^n; Y_2^n, W_2) \\
&= n\epsilon_n + I(X^n; Y_1^n, W_2, Y_2^n) - I(X^n; Y_1^n|W_2, Y_2^n) \\
&= n\epsilon_n + I(X^n; Y_1^n, Y_2^n) - I(X^n; Y_1^n|W_2, Y_2^n),
\end{aligned}$$

where (a) follows from Fano's inequality, (b) follows from data processing inequality.

$$\begin{aligned}
I(X^n; Y_1^n, Y_2^n) &= H(Y_1^n, Y_2^n) - H(Y_1^n, Y_2^n|X^n) \\
&= H(Y_1^n, Y_2^n) - H(Y_1^n|X^n) - H(Y_2^n|X^n) \\
&\leq \sum_{i=1}^{n} H(Y_{1i}, Y_{2i}) - \sum_{i=1}^{n} H(Y_{1i}|X_i) - \sum_{i=1}^{n} H(Y_{2i}|X_i) \\
&= \sum_{i=1}^{n} H(Y_{1i}, Y_{2i}) - \sum_{i=1}^{n} H(Y_{1i}, Y_{2i}|X_i) \\
&= \sum_{i=1}^{n} I(X_i; Y_{1i}, Y_{2i}) \\
&\leq n \sup_{p(x)} I(X; Y_1, Y_2) \\
\Rightarrow R &\leq \epsilon_n + \frac{I(X^n; Y_1^n, Y_2^n) - I(X^n; Y_1^n|W_2, Y_2^n)}{n} \\
\Rightarrow R &\leq \epsilon_n + \sup_{p(x)} I(X; Y_1, Y_2) - \frac{I(X^n; Y_1^n|W_2, Y_2^n)}{n} \\
&\to \sup_{p(x)} I(X; Y_1, Y_2) - \frac{I(X^n; Y_1^n|W_2, Y_2^n)}{n}, n \to \infty. \qquad (3.1)
\end{aligned}$$

From Equation (3.1), we obtain the following proposition:

Proposition 1 *In order to achieve the maximum rate* $C(\infty) = \sup_{p(x)} I(X; Y_1, Y_2)$*, it is necessary to ensure* $\frac{I(X^n; Y_1^n | W_2, Y_2^n)}{n} \to 0$*, as* $n \to \infty$*.*

Before we derive a "single-letter criterion", we first define the following notations:

$Y_{1i}^- = (Y_{11}, Y_{12}, ..., Y_{1(i-1)}), Y_{1i}^+ = (Y_{1(i+1)}, Y_{1(i+2)}, ..., Y_{1n}), Y_{2i}^- = (Y_{21}, Y_{22}, ..., Y_{2(i-1)}), Y_{2i}^+ = (Y_{2(i+1)}, Y_{2(i+2)}, ..., Y_{2n})$.

Let us first show the following three lemmas.

Lemma 1 $I(X^n; Y_1^n | W_2, Y_2^n) \geq \sum_{i=1}^{n} I(X_i; Y_{1i} | U_i, Y_{2i})$*, for* $U_i = (W_2, Y_{1i}^-, Y_{2i}^-, Y_{2i}^+)$*.*

Proof:

$$
\begin{aligned}
& I(X^n; Y_1^n | W_2, Y_2^n) \\
= \;& H(Y_1^n | W_2, Y_2^n) - H(Y_1^n | X^n, W_2, Y_2^n) \\
= \;& \sum_{i=1}^{n} [H(Y_{1i} | Y_{1i}^-, W_2, Y_2^n) - H(Y_{1i} | Y_{1i}^-, X^n, W_2, Y_2^n)] \\
= \;& \sum_{i=1}^{n} [H(Y_{1i} | Y_{1i}^-, W_2, Y_{2i}^-, Y_{2i}^+, Y_{2i}) - H(Y_{1i} | Y_{1i}^-, X^n, W_2, Y_{2i}^-, Y_{2i}^+, Y_{2i})] \\
= \;& \sum_{i=1}^{n} [H(Y_{1i} | U_i, Y_{2i}) - H(Y_{1i} | X^n, U_i, Y_{2i})] \\
= \;& \sum_{i=1}^{n} [H(Y_{1i} | U_i, Y_{2i}) - H(Y_{1i} | X_i^-, X_i, X_i^+, U_i, Y_{2i})] \\
\geq \;& \sum_{i=1}^{n} [H(Y_{1i} | U_i, Y_{2i}) - H(Y_{1i} | X_i, U_i, Y_{2i})] \\
= \;& \sum_{i=1}^{n} I(X_i; Y_{1i} | U_i, Y_{2i}),
\end{aligned}
\tag{3.2}
$$

in which $U_i = (W_2, Y_{1i}^-, Y_{2i}^-, Y_{2i}^+)$.

Lemma 2 *$U_i \to Y_{1i} \to X_i$ is a Markov chain, for $U_i = (W_2, Y_{1i}^-, Y_{2i}^-, Y_{2i}^+)$.*

To prove Lemma 2, we need to show $p(u_i, x_i|y_{1i}) = p(u_i|y_{1i})p(x_i|y_{1i})$. Since $p(u_i, x_i|y_{1i}) = p(u_i|y_{1i})p(x_i|y_{1i}, u_i)$, then we only need $p(x_i|y_{1i}, u_i) = p(x_i|y_{1i})$.

Proof:

$$
\begin{aligned}
p(x_i|y_{1i}, u_i) &= p(x_i|y_{1i}, w_2, y_{1i}^-, y_{2i}^-, y_{2i}^+) \\
&= \frac{p(x_i, y_{1i}, w_2, y_{1i}^-, y_{2i}^-, y_{2i}^+)}{p(y_{1i}, w_2, y_{1i}^-, y_{2i}^-, y_{2i}^+)} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i, y_{1i}, y_{1i}^-, y_{1i}^+, y_{2i}^-, y_{2i}^+)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_{1i}, y_{1i}^-, y_{1i}^+, y_{2i}^-, y_{2i}^+)} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i, y_1^n, y_{2i}^-, y_{2i}^+)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_1^n, y_{2i}^-, y_{2i}^+)} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i)p(y_1^n, y_{2i}^-, y_{2i}^+|x_i)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_1^n)p(y_{2i}^-, y_{2i}^+|y_1^n)} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i)p(y_1^n|x_i)p(y_{2i}^-, y_{2i}^+|x_i, y_1^n)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_1^n)p(y_{2i}^-, y_{2i}^+|y_1^n)} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i)p(y_1^n|x_i)p(y_{2i}^-, y_{2i}^+|x_i, y_1^n)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_1^n)p(y_{2i}^-, y_{2i}^+|y_1^n)}. \qquad (3.3)
\end{aligned}
$$

Next we will show $p(y_1^n|x_i) = \prod_{j=1,j\neq i}^{n} p(y_{1j})p(y_{1i}|x_i)$,

$$
\begin{aligned}
p(y_1^n|x_i) &= \sum_{x_i^-,x_i^+} p(y_1^n|x_i, x_i^-, x_i^+)P(x_i^-, x_i^+|x_i) \\
&= \sum_{x_i^-,x_i^+} p(y_1^n|x^n)P(x_i^-, x_i^+) \\
&= \sum_{x_i^-,x_i^+} \prod_{j=1}^{n} p(y_{1j}|x_j) \prod_{j=1,j\neq i}^{n} p(x_j) \\
&= \sum_{x_i^-,x_i^+} \prod_{j=1,j\neq i}^{n} p(y_{1j}, x_j)p(y_{1i}|x_i) \\
&= \prod_{j=1,j\neq i}^{n} p(y_{1j})p(y_{1i}|x_i). \qquad (3.4)
\end{aligned}
$$

The following equations show that $p(y_{2i}^-, y_{2i}^+|x_i, y_1^n) = p(y_{2i}^-, y_{2i}^+|y_{1i}^-, y_{1i}^+)$,

$$\begin{aligned}
p(y_{2i}^-, y_{2i}^+|x_i, y_1^n) &= \sum_{x_i^-, x_i^+} p(y_{2i}^-, y_{2i}^+|x_i, x_i^-, x_i^+, y_1^n) p(x_i^-, x_i^+|x_i, y_1^n) \\
&= \sum_{x_i^-, x_i^+} \sum_{y_{2i}} p(y_{2i}^-, y_{2i}, y_{2i}^+|x_i, x_i^-, x_i^+, y_1^n) \frac{p(x_i^-, x_i^+, x_i|y_1^n)}{p(x_i|y_1^n)} \\
&= \sum_{x_i^-, x_i^+} \sum_{y_{2i}} p(y_2^n|x^n, y_1^n) \frac{\prod_{i=1}^n p(x_i|y_{1i})}{p(x_i|y_{1i})} \\
&= \sum_{x_i^-, x_i^+} \sum_{y_{2i}} p(y_2^n|x^n) \prod_{j=1, j\neq i}^n p(x_j|y_{1j}) \\
&= \sum_{x_i^-, x_i^+} \sum_{y_{2i}} \prod_{i=1}^n p(y_{2i}|x_i) \prod_{j=1, j\neq i}^n p(x_j|y_{1j}) \\
&= \sum_{x_i^-, x_i^+} \prod_{j=1, j\neq i}^n p(y_{2j}|x_j) \prod_{j=1, j\neq i}^n p(x_j|y_{1j}) \\
&= \sum_{x_i^-, x_i^+} p(y_{2i}^-, y_{2i}^+|x_i^-, x_i^+) p(x_i^-, x_i^+|y_{1i}^-, y_{1i}^+) \\
&= \sum_{x_i^-, x_i^+} p(y_{2i}^-, y_{2i}^+|x_i^-, x_i^+, y_{1i}^-, y_{1i}^+) p(x_i^-, x_i^+|y_{1i}^-, y_{1i}^+) \\
&= p(y_{2i}^-, y_{2i}^+|y_{1i}^-, y_{1i}^+). \qquad (3.5)
\end{aligned}$$

Here, we will present $p(y_{2i}^-, y_{2i}^+|y_1^n) = \prod_{j=1, j\neq i}^n p(y_{2i}|y_{1i})$,

$$
\begin{aligned}
p(y_{2i}^-, y_{2i}^+|y_1^n) &= \sum_{x^n, y_{2i}} p(y_2^n, x^n|y_1^n) \\
&= \sum_{x^n, y_{2i}} p(x^n|y_1^n)p(y_2^n|x^n, y_1^n) \\
&= \sum_{x^n, y_{2i}} p(x^n|y_1^n)p(y_2^n|x^n) \\
&= \sum_{x^n, y_{2i}} \prod_{i=1}^n p(x_i|y_{1i})p(y_{2i}|x_i) \\
&= \sum_{x^n, y_{2i}} \prod_{i=1}^n p(x_i|y_{1i})p(y_{2i}|x_i, y_{1i}) \\
&= \sum_{x^n, y_{2i}} \prod_{i=1}^n p(y_{2i}, x_i|y_{1i}) \\
&= \sum_{y_{2i}} \prod_{i=1}^n p(y_{2i}|y_{1i}) \\
&= \prod_{j=1, j\neq i}^n p(y_{2i}|y_{1i}) \\
&= p(y_{2i}^-, y_{2i}^+|y_{1i}^-, y_{1i}^+). \qquad (3.6)
\end{aligned}
$$

Combine Equations (3.3), (3.4), (3.5) and (3.6) together, we will have,

$$
\begin{aligned}
p(x_i|y_{1i},u_i) &= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i)p(y_{1i}|x_i)p(y_{1i}^-, y_{1i}^+)p(y_{2i}^-, y_{2i}^+|y_{1i}^-, y_{1i}^+)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_{1i})p(y_{1i}^-, y_{1i}^+)p(y_{2i}^-, y_{2i}^+|y_{1i}^-, y_{1i}^+))} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i)p(y_{1i}|x_i)p(y_{2i}^-, y_{2i}^+, y_{1i}^-, y_{1i}^+)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_{1i})p(y_{2i}^-, y_{2i}^+, y_{1i}^-, y_{1i}^+))} \\
&= \frac{p(x_i)p(y_{1i}|x_i)p(y_{2i}^-, y_{2i}^+, y_{1i}^-)}{p(y_{1i})p(y_{2i}^-, y_{2i}^+, y_{1i}^-))} \\
&= \frac{p(y_{1i}, x_i)}{p(y_{1i})} \\
&= p(x_i|y_{1i}). \\
\Rightarrow \qquad & p(u_i, x_i|y_{1i}) = p(u_i|y_{1i})p(x_i|y_{1i}),
\end{aligned}
$$

thus we finish the proof of Lemma 2.

Lemma 3 *$U_i \to (Y_{1i}, Y_{2i}) \to X_i$ is a Markov chain, for $U_i = (W_2, Y_{1i}^-, Y_{2i}^-, Y_{2i}^+)$.*

To prove Lemma 3, we need to show $p(u_i, x_i|y_{1i}, y_{2i}) = p(u_i|y_{1i}, y_{2i})p(x_i|y_{1i}, y_{2i})$. Since $p(u_i, x_i|y_{1i}, y_{2i}) = p(u_i|y_{1i}, y_{2i})p(x_i|y_{1i}, y_{2i}, u_i)$, then we only need $p(x_i|y_{1i}, y_{2i}, u_i) = p(x_i|y_{1i}, y_{2i})$.

Proof :

$$
\begin{aligned}
p(x_i|y_{1i}, y_{2i}, u_i) &= p(x_i|y_{1i}, y_{2i}, w_2, y_{1i}^-, y_{2i}^-, y_{2i}^+) \\
&= \frac{p(x_i, y_{1i}, w_2, y_{1i}^-, y_2^n)}{p(y_{1i}, w_2, y_{1i}^-, y_2^n)} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i, y_{1i}, y_{1i}^-, y_{1i}^+, y_2^n)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_{1i}, y_{1i}^-, y_{1i}^+, y_2^n)} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i, y_1^n, y_2^n)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_1^n, y_2^n)} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i)p(y_1^n, y_2^n|x_i)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_1^n)p(y_2^n|y_1^n)} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i)p(y_1^n|x_i)p(y_2^n|x_i, y_1^n)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_1^n)p(y_2^n|y_1^n)}. \qquad (3.7)
\end{aligned}
$$

In the following, $p(y_2^n|x_i,y_1^n) = p(y_{2i}|x_i)\prod_{j=1,j\neq i}^n p(y_{2j}|y_{1j})$ is given,

$$\begin{aligned}
p(y_2^n|x_i,y_1^n) &= \sum_{x_i^-,x_i^+} p(y_2^n|x_i,x_i^-,x_i^+,y_1^n)p(x_i^-,x_i^+|x_i,y_1^n) \\
&= \sum_{x_i^-,x_i^+} p(y_2^n|x^n,y_1^n)\frac{p(x^n|y_1^n)}{p(x_i|y_1^n)} \\
&= \sum_{x_i^-,x_i^+} p(y_2^n|x^n,y_1^n)\frac{\prod_{i=1}^n p(x_i|y_{1i})}{p(x_i|y_{1i})} \\
&= \sum_{x_i^-,x_i^+} p(y_2^n|x^n)\prod_{j=1,j\neq i}^n p(x_j|y_{1j}) \\
&= \sum_{x_i^-,x_i^+} \prod_{i=1}^n p(y_{2i}|x_i)\prod_{j=1,j\neq i}^n p(x_j|y_{1j}) \\
&= p(y_{2i}|x_i)\sum_{x_i^-,x_i^+} \prod_{j=1,j\neq i}^n p(y_{2j}|x_j)\prod_{j=1,j\neq i}^n p(x_j|y_{1j}) \\
&= p(y_{2i}|x_i)\sum_{x_i^-,x_i^+} \prod_{j=1,j\neq i}^n p(y_{2j}|x_j,y_{1j})\prod_{j=1,j\neq i}^n p(x_j|y_{1j}) \\
&= p(y_{2i}|x_i)\sum_{x_i^-,x_i^+} \prod_{j=1,j\neq i}^n p(y_{2j},x_j|y_{1j}) \\
&= p(y_{2i}|x_i)\prod_{j=1,j\neq i}^n p(y_{2j}|y_{1j}). \qquad (3.8)
\end{aligned}$$

Also, we will show $p(y_2^n|y_1^n) = \prod_{i=1}^n p(y_{2i}|y_{1i})$,

$$
\begin{aligned}
p(y_2^n|y_1^n) &= \sum_{x^n} p(y_2^n, x^n|y_1^n) \\
&= \sum_{x^n} p(x^n|y_1^n)p(y_2^n|x^n, y_1^n) \\
&= \sum_{x^n} p(x^n|y_1^n)p(y_2^n|x^n) \\
&= \sum_{x^n} \prod_{i=1}^n p(x_i|y_{1i})p(y_{2i}|x_i) \\
&= \sum_{x^n} \prod_{i=1}^n p(x_i|y_{1i})p(y_{2i}|x_i, y_{1i}) \\
&= \sum_{x^n} \prod_{i=1}^n p(y_{2i}, x_i|y_{1i}). \\
&= \prod_{i=1}^n p(y_{2i}|y_{1i}). \qquad (3.9)
\end{aligned}
$$

Combine Equations (3.4), (3.7), (3.8) and (3.9) together , we will have

$$
\begin{aligned}
p(x_i|y_{1i},y_{2i},u_i) &= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i)p(y_1^n|x_i)p(y_2^n|x_i,y_1^n)}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_1^n)p(y_2^n|y_1^n)} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(x_i)p(y_{1i}|x_i)p(y_{2i}|x_i)\prod_{j=1,j\neq i}^n p(y_{1j})p(y_{2j}|y_{1j})}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} \prod_{j=1}^n p(y_{1i})\prod_{i=1}^n p(y_{2i}|y_{1i})} \\
&= \frac{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} p(y_{1i},y_{2i},x_i)\prod_{j=1,j\neq i}^n p(y_{2j},y_{1j})}{\sum\limits_{y_{1i}^+, f_2(y_1^n)=w_2} \prod_{j=1}^n p(y_{2j},y_{1j})} \\
&= \frac{p(y_{1i},y_{2i},x_i)\prod_{j=1}^{i-1} p(y_{2j},y_{1j})}{\prod_{j=1}^{i} p(y_{2j},y_{1j})} \\
&= \frac{p(y_{1i},y_{2i},x_i)}{p(y_{2i},y_{1i})} \\
&= p(x_i|y_{1i},y_{2i}),
\end{aligned}
$$

$$
\Rightarrow \qquad p(u_i,x_i|y_{1i},y_{2i}) = p(u_i|y_{1i},y_{2i})p(x_i|y_{1i},y_{2i}),
$$

thus finish the proof of Lemma 3.

From Lemma 1, we have, for $U_i = (W_2, Y_{1i}^-, Y_{2i}^-, Y_{2i}^+)$,

$$
I(X^n;Y_1^n|W_2,Y_2^n) \geq \sum_{i=1}^n I(X_i;Y_{1i}|U_i,Y_{2i}) \Rightarrow I(X^n;Y_1^n|W_2,Y_2^n)/n \geq \tfrac{1}{n}\sum_{i=1}^n I(X_i;Y_{1i}|U_i,Y_{2i}).
$$

Introducing an timesharing r.v. Q, we rewrite the inequality as

$$
I(X^n;Y_1^n|W_2,Y_2^n)/n \geq \frac{1}{n}\sum_{i=1}^n I(X_i;Y_{1i}|U_i,Y_{2i},Q=i) = I(X_Q;Y_{1Q}|U_Q,Y_{2Q},Q). \quad (3.10)
$$

Q is independent of X_Q, Y_{1Q}, Y_{2Q}. Defining $U^{(n)} = (U_Q, Q), X = X_Q, Y_1 = Y_{1Q}, Y_2 = Y_{2Q}$, we have shown the existence of a r.v. $U^{(n)}$ such that

$$
I(X^n;Y_1^n|W_2,Y_2^n)/n \geq I(X;Y_1|U^{(n)},Y_2). \quad (3.11)
$$

From the definition of $U^{(n)}$, it is clear that $U^{(n)} \to Y_1 \to X$ and $U^{(n)} \to (Y_1, Y_2) \to X$ again form Markov chains. Finally, we show the cardinality range of $U^{(n)}$ can be bounded by $||X||$. $||X||$ is the cardinality of X.

We use the method discussed in [3] and [22] to show the cardinality bound of $U^{(n)}$. The key point in [3] and [22] is the use of two theorems and three lemmas as stated below.

Theorem 11 *Caratheodory's Theorem: Let $\mathcal{X}$ be a subset of $\mathcal{R}^n$, let $\mathcal{C}(\mathcal{X})$ denote the convex hull of $\mathcal{X}$, and let $\bar{y}$ be a point of $\mathcal{C}(\mathcal{X})$; then there exists a set of s points $\bar{X}_1, \cdots, \bar{X}_s$, all belonging to $\mathcal{X}$, with $s \leq n+1$, such that $\bar{y}$ is a point of the simplex whose vertices are $\bar{X}_1, \cdots, \bar{X}_s$.*

Proof: See [9], page 35.

Theorem 12 *(Fenchel-Eggleston): If in the conditions of Caratheodory's theorem it is also assumed that the set $\mathcal{X}$ is the union of at most n connected sets then $s \leq n$.*

Proof: See [9], page 35.

Lemma 4 *(Ahlswede and Körner): Let $\mathcal{P}$ be any subset of $\mathcal{R}^n$, and let $f_j(\bar{P}), j = 1, 2, \cdots, k$ be real valued functions on $\mathcal{P}$. Then to any probability measure μ (on the Borel subset of) $\mathcal{P}$ there exists $k+1$ elements $\bar{P}_i$ of $\mathcal{P}$ and constants $\alpha_i, i = 1, 2, \cdots, k+1, \sum_{i=1}^{k+1} \alpha_i = 1$, such that*

$$\int f_j(\bar{P})\mu(d\bar{P}) = \sum_{i=1}^{k+1} \alpha_i f_j(\bar{P}_i), j = 1, 2, \cdots, k.$$

Proof: See [3], Lemma 3.

Lemma 5 *(Masoud Salehi): Let $\mathcal{P}$ be an subset of $\mathcal{R}^n$ consisting of at most k connected subsets. Let $f_j(\bar{P}), j = 1, 2, \cdots, k$ be real valued continuous functions on $\mathcal{P}$. Then to any probability measure μ (on the Brel subsets of) $\mathcal{P}$ there exists k elements $\bar{P}_i$ of $\mathcal{P}$ and constants $\alpha_i \geq 0, i = a, 2, ..., k, \sum_{i=1}^{k} \alpha_i = 1$, such that*

$$\int f_j(\bar{P})\mu(d\bar{P}) = \sum_{i=1}^{k} \alpha_i f_j(\bar{P}_i), j = 1, 2, ..., k.$$

Proof: See [22], Lemma 2.

Lemma 6 *(Masoud Salehi): Let for $i = 1, 2, ..., n, g_i : \mathcal{R}^m \to \mathcal{R}$ be a set of n positive valued functions, let $T \subseteq \mathcal{R}^m$ be a closed set, and define*

$$\mathcal{E} = \{\bar{X} \in \mathcal{R}^n : \bar{X} \geq \bar{0}, \exists \bar{t} \in T; x_i \leq g_i(\bar{t}), i = 1, 2, ..., n\}.$$

Let $\mathcal{E}' = \mathcal{C}(\mathcal{E})$ denotes the convex hull of $\mathcal{E}$, and let

$$\mathcal{E}'' = \{\bar{X} \in \mathcal{R}^n : \bar{X} \geq \bar{0}, \forall \bar{\lambda} \in \mathcal{R}^n, \bar{\lambda} \geq 0, \bar{\lambda}^t \bar{X} \leq G(\bar{\lambda})\},$$

where

$$G(\bar{\lambda}) = \sup_{\bar{t} \in T} \bar{\lambda}^t [g_1(\bar{t}), g_2(\bar{t}), ..., g_n(\bar{t})]^t.$$

Then $\mathcal{E}' = \mathcal{E}''$.

Proof: See [22], Lemma 3.

Next, we will show the cardinality bound of $U^{(n)}$. For more detailed analysis, please refer to [22]. By lemma 6, we can write

$$\mathcal{V} = \{I(X; Y_1|U^{(n)}, Y_2) \in \mathcal{R} : \forall \lambda \in \mathcal{R}, \lambda \geq 0, \lambda^t I(X; Y_1|U^{(n)}, Y_2) \geq G(\lambda)\}$$

where

$$G(\lambda) = \inf \lambda^t I(X; Y_1|U^{(n)}, Y_2)^t,$$

and infimum is over all r.v. $U^{(n)}$ where $U^{(n)} \to Y_1 \to X$ and $U^{(n)} \to (Y_1, Y_2) \to X$ form Markov chains. We will prove for all λ that $G(\lambda)$ can be achieved by considering those $U^{(n)}$'s with cardinality less than or equal to $||X||$.

Fix $\lambda \geq 0$, and let $\mathcal{P}$ in Lemma 5 be the $||X||$-dimensional probability simplex. Let $\mathcal{X} = 1, 2, ..., ||X||$ be the range of X and interpret

$$\bar{P} = (P(X = 1|U^{(n)} = u), P(X = 2|U^{(n)} = u), ..., P(X = ||X|||U^{(n)} = u))$$

as a point in $\mathcal{P}$. Then each probability on $U^{(n)}$ defines a measure $\mu(d\bar{P})$ on $\mathcal{P}$. Let $P_X^*(.)$ achieve $G(\lambda)$ and let $\mu^*(d\bar{P})$ achieve $P_X(.)$ (and thus $G(\lambda)$). Define

$$f_j(\bar{P}) = P_X(j), j = 1, 2, \cdots, ||X|| - 1$$

$$f_{||X||}(\bar{P}) = \lambda H_{\bar{P}}(X|Y_2) - \lambda H_{\bar{P}}(X|Y_1 Y_2)$$

where $H_{\bar{P}}(X|Y_2)$ and $H_{\bar{P}}(X|Y_1Y_2)$ are the entropies of X given Y_2 and X given Y_1, Y_2 respectively when the distribution of X is $\bar{P}$. Noting that

$$\int f_j(\bar{P})\mu^*(d\bar{P}) = P_X^*(j), j = 1, 2, \cdots, ||X|| - 1$$

and

$$\int f_{||X||}(\bar{P})\mu^*(d\bar{P}) = \lambda H^*(X|Y_2, U^{(n)}) - \lambda H^*(X|Y_1Y_2, U^{(n)}) = \lambda I^*(X; Y_1|U^{(n)}, Y_2).$$

Applying Lemma 5, it is seen that there exists $||X||$ elements $\bar{P}_i \in \mathcal{P}, i = 1, 2, \cdots, ||X||$, and constants $\alpha_i \geq 0, i = 1, 2, \cdots, ||X||, \sum_{i=1}^{||X||} \alpha_i = 1$, such that

$$P_X^*(j) = \sum_{i=1}^{||X||} \alpha_i f_j(\bar{P}_i), j = 1, 2, \cdots, ||X|| - 1$$

$$G(\lambda) = \lambda I^*(X; Y_1|U^{(n)}, Y_2) = \sum_{i=1}^{||X||} \alpha_i f_{||X||}(\bar{P}_i).$$

This means that in order to achieve $G(\lambda)$, it is enough to put positive probability on at most $||X||$ elements of $\mathcal{P}$. This in turn shows it is enough to consinder $U^{(n)}$'s with $||U^{(n)}|| \leq ||X||$.

Thus, there exists a subsequence $\{U^{(n_k)}\} \subseteq \{U^{(n)}\}$ and a r.v. U^*, such that $U^{(n_k)} \to U^*$, as $k \to \infty$ and

$$\lim_{k\to\infty} I(X; Y_1|U^{(n_k)}, Y_2) = I(X; Y_1|U^*, Y_2)$$

and

$$U^* \to Y_1 \to X \text{ and } U^* \to (Y_1, Y_2) \to X \text{ form Markov chains.}$$

From Equation (3.11), we have

$$I(X^n; Y_1^n|W_2, Y_2^n)/n \geq I(X; Y_1|U^{(n)}, Y_2).$$

Thus,

$$I(X^n; Y_1^n | W_2, Y_2^n)/n \geq I(X; Y_1 | U^{(n_k)}, Y_2)$$

$$\lim_{n\to\infty} I(X^n; Y_1^n | W_2, Y_2^n)/n \geq \lim_{k\to\infty} I(X; Y_1 | U^{(n_k)}, Y_2) = I(X; Y_1 | U^*, Y_2).$$

This implies that $I(X; Y_1|U^*, Y_2) = 0$ is a necessary condition for $\lim_{n\to\infty} I(X^n; Y_1^n|W_2, Y_2^n)/n = 0$ to be held. We then obtain a result from Proposition 1.

Proposition 2 *A necessary condition for approaching the maximum rate* $C(\infty) = \sup_{p(x)} I(X; Y_1, Y_2)$ *is* $I(X; Y_1|U^*, Y_2) = 0$.

We will show the condition for $I(X; Y_1|U^*, Y_2) = 0$ as follows.

$$\begin{aligned}
I(X; Y_1|U^*, Y_2) &= \sum_{x,y_1,y_2,u^*} p(x, y_1, y_2, u^*) \log \frac{p(y_1, x|u^*, y_2)}{p(y_1|u^*, y_2)p(x|u^*, y_2)} \\
&= \sum_{x,y_2,u^*} p(y_2, u^*) \sum_{y_1} p(x, y_1|y_2, u^*) \log \frac{p(y_1, x|u^*, y_2)}{p(y_1|u^*, y_2)p(x|u^*, y_2)} \\
&\geq \sum_{x,y_2,u^*} p(y_2, u^*) \sum_{y_1} p(x, y_1|y_2, u^*) \log \frac{\sum_{y_1} p(y_1, x|u^*, y_2)}{\sum_{y_1} p(y_1|u^*, y_2)p(x|u^*, y_2)} \\
&= \sum_{x,y_2,u^*} p(y_2, u^*) \sum_{y_1} p(x, y_1|y_2, u^*) \log \frac{p(x|u^*, y_2)}{p(x|u^*, y_2)} \\
&= 0,
\end{aligned}$$

where "=" holds, iff given x, y_2, u^*, for all y_1, $\frac{p(y_1,x|u^*,y_2)}{p(y_1|u^*,y_2)p(x|u^*,y_2)}$ is a constant.

Equivalently, this implies:

Case 1: given x, y_2, u^* for any $y_1 \neq y_1'$, $p(y_1|u^*, y_2) \neq 0$ and $p(y_1'|u^*, y_2) \neq 0$, we have

$\frac{p(y_1,x|u^*,y_2)}{p(y_1|u^*,y_2)p(x|u^*,y_2)} = \frac{p(y_1',x|u^*,y_2)}{p(y_1'|u^*,y_2)p(x|u^*,y_2)}$. This equation can be further reduced as follows.

$$\begin{aligned}
\frac{p(y_1,x|u^*,y_2)}{p(y_1|u^*,y_2)} &= \frac{p(y_1',x|u^*,y_2)}{p(y_1'|u^*,y_2)} \\
\Leftrightarrow p(x|u^*,y_1,y_2) &= p(x|u^*,y_1',y_2) \\
\Leftrightarrow p(x|y_1,y_2) &= p(x|y_1',y_2) \qquad (3.12) \\
\Leftrightarrow \frac{p(x,y_1,y_2)}{p(y_1,y_2)} &= \frac{p(x,y_1',y_2)}{p(y_1',y_2)} \\
\Leftrightarrow \frac{p(y_1,y_2|x)p(x)}{p(y_2)p(y_1|y_2)} &= \frac{p(y_1',y_2|x)p(x)}{p(y_2)p(y_1'|y_2)} \\
\Leftrightarrow \frac{p(y_1,y_2|x)}{p(y_1|y_2)} &= \frac{p(y_1',y_2|x)}{p(y_1'|y_2)} \\
\Leftrightarrow \frac{p(y_1|x)p(y_2|x)}{p(y_1|y_2)} &= \frac{p(y_1'|x)p(y_2|x)}{p(y_1'|y_2)} \\
\Leftrightarrow \frac{p(y_1|x)}{p(y_1|y_2)} &= \frac{p(y_1'|x)}{p(y_1'|y_2)}.
\end{aligned}$$

Equation (3.12) comes from the fact that $U^* - (Y_1, Y_2) - X$ is Markov chain.

Case 2: given x, y_2, u^*, there exists one y_1 such that $p(y_1|u^*, y_2) = 1$ and for other $y_1' \neq y_1$, $p(y_1'|u^*, y_2) = 0$.

Let us consider the channels which satisfy: given any x, y_2, for any $y_1 \neq y_1'$, $\frac{p(y_1|x)}{p(y_1|y_2)} \neq \frac{p(y_1'|x)}{p(y_1'|y_2)}$. Following the discussions in Case 1 and Case 2 for Proposition 2, it is easy to observe that, for such channels, in order to achieve the maximum rate, Case 2 has to be satisfied. Thus, under the given channel condition, we have

$$I(X; Y_1|U^*, Y_2) = 0 \Rightarrow H(Y_1|U^*, Y_2) = 0. \qquad (3.13)$$

Next, we will show Slepian-Wolf coding is optimal to achieve the maximum rate for this channel model, by verifying the following statement:

$$\tfrac{1}{n} H(Y_1^n|W_2, Y_2^n) \to 0, \text{ as } n \to \infty.$$

Proof :

$$\begin{aligned}
\frac{1}{n}H(Y_1^n|W_2,Y_2^n) &= \frac{1}{n}\sum_{i=1}^{n}H(Y_{1i}|W_2,Y_{1i}^-,Y_{2i}^-,Y_{2i}^+,Y_{2i}) \\
&= \frac{1}{n}\sum_{i=1}^{n}H(Y_{1i}|U_i,Y_{2i}) \\
&= \frac{1}{n}\sum_{i=1}^{n}H(Y_{1i}|U_i,Y_{2i},Q=i) \\
&= H(Y_{1Q}|U_Q,Y_{2Q},Q) \\
&= H(Y_1|U^{(n)},Y_2),
\end{aligned}$$

where $U_i = (W_2, Y_{1i}^-, Y_{2i}^-, Y_{2i}^+)$, Q is timesharing r.v. and $U^{(n)} = (U_Q, Q), Y_1 = Y_{1Q}, Y_2 = Y_{2Q}$. All the definitions are the same as those in Equations (3.2), (3.10) and (3.11). Equivalently, we need to prove

$$H(Y_1|U^{(n)},Y_2) \to 0, \text{ as } n \to \infty.$$

It is clear that for sequence $\{H(Y_1|U^{(n)},Y_2)\}$, $\lim_{n\to\infty}\inf H(Y_1|U^{(n)},Y_2) = 0$. We assume $\lim_{n\to\infty}\sup H(Y_1|U^{(n)},Y_2) = a > 0$. Thus, there exists a subsequence $\{U^{n_k}\} \subseteq \{U^{(n)}\}$, such that $\lim_{k\to\infty} H(Y_1|U^{(n_k)},Y_2) = a > 0$. Furthermore, there exists another subsequence $\{U^{n_{k_l}}\} \subseteq \{U^{(n_k)}\}$ and a r.v. U^{**}, such that $\lim_{l\to\infty} U^{(n_{k_l})} = U^{**}$. Then we have

$$\lim_{l\to\infty} H(Y_1|U^{(n_{k_l})},Y_2) = H(Y_1|U^{**},Y_2) = a > 0. \tag{3.14}$$

However, in order to achieve the maximum rate, it is necessary to ensure:

$$\lim_{n\to\infty} I(X^n;Y_1^n|W_2,Y_2^n)/n \geq \lim_{n\to\infty} I(X;Y_1|U^{(n)},Y_2) = 0$$

$$\Rightarrow$$

$$\lim_{l\to\infty} I(X;Y_1|U^{(n_{k_l})},Y_2) = I(X;Y_1|U^{**},Y_2) = 0.$$

Same reason as for Equation (3.13), we have

$$I(X;Y_1|U^{**},Y_2) = 0 \Rightarrow H(Y_1|U^{**},Y_2) = 0. \tag{3.15}$$

Thus, we obtain a contradiction from Equation (3.14) and Equation (3.15), which implies the assumption "$\lim_{n\to\infty} \sup H(Y_1|U^{(n)}, Y_2) = a > 0$" is invalid. Then we have $\lim_{n\to\infty} \sup H(Y_1|U^{(n)}, Y_2) = 0$. This induces that $\lim_{n\to\infty} H(Y_1|U^{(n)}, Y_2) = 0$.

From previous discussion, we derive the "single-letter criterion" into Theorem 13.

Theorem 13 *(Single-letter Criterion) For the channels satisfying: "given any x, y_2, for any $y_1 \neq y_1'$, $\frac{p(y_1|x)}{p(y_1|y_2)} \neq \frac{p(y_1'|x)}{p(y_1'|y_2)}$", Slepian-Wolf coding is essentially the optimal relay strategy to achieve the maximum rate $C(\infty) = \sup_{p(x)} I(X; Y_1, Y_2)$ and $R_0 = \min_{p(x)} H(Y_1|Y_2)$ is the minimum relay rate.*

3.3 Optimal Relay Rate for the BSC Relay Channel

In this section we will present the optimal relay rate for the BSC relay channel by applying Theorem 13. Two cases will be discussed here: both the source-relay link and the source-destination link are BSC channels with the same error probability; and the source-relay link and the source-destination link are BSC channels with different error probabilities.

Case 1: suppose both the source-relay link and the source-destination link are BSC channels with the same error probability p and the input distribution is $P(x = 0) = \pi, P(x = 1) = 1 - \pi$, $0 < p < \frac{1}{2}, 0 < \pi < 1$. Other related probabilities for the BSC relay channel ($\mathcal{X}$, $p(y_1, y_2|x)$, $\mathcal{Y}_1 \times \mathcal{Y}_2$) are given in the following tables:

Table 3.1: Distributions of X and Y_2 for the BSC relay channel (case 1)

$X \backslash Y_2$	0	1
$P(X)$	π	$1 - \pi$
$P(Y_2)$	$p + \pi - 2\pi p$	$1 - p - \pi + 2\pi p$

Case 2: suppose the source-relay link is BSC channel with error probability p_1, and the source-destination link is BSC channel with error probability p_2 and $p_1 \neq p_2$. The input distribution is $P(x = 0) = \pi, P(x = 1) = 1 - \pi$, $0 < p_1, p_2 < \frac{1}{2}, 0 < \pi < 1$. Other related

Table 3.2: Joint distribution of Y_1 and Y_2, $P(Y_1, Y_2)$ for the BSC relay channel (case 1)

$Y_1 \backslash Y_2$	0	1
0	$\pi(1-p)^2 + (1-\pi)p^2$	$p(1-p)$
1	$p(1-p)$	$\pi p^2 + (1-\pi)(1-p)^2$

Table 3.3: Conditional probabilities of Y_1 given Y_2, $P(Y_1|Y_2)$ for the BSC relay channel (case 1)

$Y_1 \backslash Y_2$	0	1
0	$\frac{\pi(1-p)^2+(1-\pi)p^2}{p+\pi-2\pi p}$	$\frac{p(1-p)}{1-p-\pi(1-2p)}$
1	$\frac{p(1-p)}{p+\pi-2\pi p}$	$\frac{\pi p^2+(1-\pi)(1-p)^2}{1-p-\pi(1-2p)}$

Table 3.4: Values of $\frac{P(Y_1|X)}{P(Y_1|Y_2)}$ for the BSC relay channel (case 1)

$Y_1 \backslash (X, Y_2)$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
0	$\frac{(1-p)(p+\pi-2\pi p)}{\pi(1-p)^2+(1-\pi)p^2}$	$\frac{1-p-\pi(1-2p)}{p}$	$\frac{p(p+\pi-2\pi p)}{\pi(1-p)^2+(1-\pi)p^2}$	$\frac{1-p-\pi(1-2p)}{1-p}$
1	$\frac{p+\pi-2\pi p}{1-p}$	$\frac{p[1-p-\pi(1-2p)]}{\pi p^2+(1-\pi)(1-p)^2}$	$\frac{p+\pi-2\pi p}{p}$	$\frac{(1-p)[1-p-\pi(1-2p)]}{\pi p^2+(1-\pi)(1-p)^2}$

probabilities for the BSC relay channel ($\mathcal{X}$, $p(y_1, y_2|x)$, $\mathcal{Y}_1 \times \mathcal{Y}_2$) are given in the following tables:

Table 3.5: Distributions of X and Y_2 for the BSC relay channel (case 2)

X \ Y_2	0	1
$P(X)$	π	$1-\pi$
$P(Y_2)$	$p_2+\pi-2\pi p_2$	$1-p_2-\pi+2\pi p_2$

Table 3.6: Joint distribution of Y_1 and Y_2, $P(Y_1, Y_2)$ for the BSC relay channel (case 2)

Y_1 \ Y_2	0	1
0	$\pi(1-p_1)(1-p_2)+(1-\pi)p_1p_2$	$\pi(1-p_1)p_2+(1-\pi)p_1(1-p_2)$
1	$\pi p_1(1-p_2)+(1-\pi)(1-p_1)p_2$	$\pi p_1p_2+(1-\pi)(1-p_1)(1-p_2)$

Table 3.7: Conditional probabilities of Y_1 given Y_2, $P(Y_1|Y_2)$ for the BSC relay channel (case 2)

Y_1 \ Y_2	0	1
0	$\frac{\pi(1-p_1)(1-p_2)+(1-\pi)p_1p_2}{p_2+\pi-2\pi p_2}$	$\frac{\pi(1-p_1)p_2+(1-\pi)p_1(1-p_2)}{1-p_2-\pi+2\pi p_2}$
1	$\frac{\pi p_1(1-p_2)+(1-\pi)(1-p_1)p_2}{p_2+\pi-2\pi p_2}$	$\frac{\pi p_1p_2+(1-\pi)(1-p)_1(1-p)_2}{1-p_2-\pi+2\pi p_2}$

Table 3.8: Values of $\frac{P(Y_1|X)}{P(Y_1|Y_2)}$ for the BSC relay channel (case 2)

Y_1 \ (X, Y_2)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
0	$\frac{(1-p_1)(p_2+\pi-2\pi p_2)}{\pi-\pi(p_1+p_2)+p_1p_2}$	$\frac{(1-p_1)(1-p_2-\pi+2\pi p_2)}{p_1-p_1p_2-\pi(p_1-p_2)}$	$\frac{p_1(p_2+\pi-2\pi p_2)}{\pi-\pi(p_1+p_2)+p_1p_2}$	$\frac{p_1(1-p_2-\pi+2\pi p_2)}{p_1-p_1p_2-\pi(p_1-p_2)}$
1	$\frac{p_1(p_2+\pi-2\pi p_2)}{\pi(p_1-p_2)+p_2-p_1p_2}$	$\frac{p_1(1-p_2-\pi+2\pi p_2)}{1-\pi-(1-\pi)(p_1+p_2)+p_1p_2}$	$\frac{(1-p_1)(p_2+\pi-2\pi p_2)}{\pi(p_1-p_2)+p_2-p_1p_2}$	$\frac{(1-p_1)(1-p_2-\pi+2\pi p_2)}{1-\pi-(1-\pi)(p_1+p_2)+p_1p_2}$

From the results in Table 3.4 and Table 3.8, we observe: for the BSC relay channel ($\mathcal{X}$, $p(y_1, y_2|x)$, $\mathcal{Y}_1 \times \mathcal{Y}_2$), given any x, y_2 and for any $y_1 \neq y_1^{'}$, $\frac{p(y_1|x)}{p(y_1|y_2)} \neq \frac{p(y_1^{'}|x)}{p(y_1^{'}|y_2)}$ holds. Apply Theorem 13, we obtain the following proposition for the BSC relay channel.

Proposition 3 *For the BSC relay channel, Slepian-Wolf coding is essentially the optimal relay strategy to achieve the maximum rate* $C(\infty) = \max_{p(x)} I(X; Y_1, Y_2)$ *and the minimum relay rate* R_0 *is* $\min_{p(x)} H(Y_1|Y_2)$.

Furthermore, we calculate the value of the maximum rate by referring to the values in Table 3.7:

$$
\begin{aligned}
I(X; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X) \\
&= H(Y_1, Y_2) - H(Y_1|X) - H(Y_2|X) \\
&= H(Y_1, Y_2) - H(p_1) - H(p_2) \\
&\leq \max_{p(x)} H(Y_1, Y_2) - H(p_1) - H(p_2)
\end{aligned}
$$

We define

$$H(p_1, p_2, \cdots, p_n) = -p_1 \log p_1 - p_2 \log p_2 - \cdots - p_n \log p_n, \sum_{i=1}^{n} p_i = 1.$$

From Table 3.6, we have,

$$
\begin{aligned}
H(Y_1, Y_2) = \; & H(\pi(1-p_1)(1-p_2) + (1-\pi)p_1p_2, \pi(1-p_1)p_2 + (1-\pi)p_1(1-p_2), \\
& \pi p_1(1-p_2) + (1-\pi)(1-p_1)p_2, \pi p_1p_2 + (1-\pi)(1-p_1)(1-p_2)).
\end{aligned}
$$

$H(Y_1, Y_2)$ can be regarded as a function of $\pi, 0 < \pi < 1$. Let $f(\pi) = H(Y_1, Y_2)$, then take the derivative of π, we have:

When $0 < \pi \leq \frac{1}{2}$, $f'(\pi) \geq 0$ and when $\frac{1}{2} \leq \pi < 1$, $f'(\pi) \leq 0$. This implies $\max_{p(x)} H(Y_1, Y_2) = \max_\pi f(\pi)$ is achieved by $\pi = \frac{1}{2}$. Thus,

$$I(X; Y_1, Y_2) \leq H(\frac{1}{2}(1 - p_1 - p_2) + p_1 p_2, p_1 - p_1 p_2 + \frac{1}{2}(p_2 - p_1),$$
$$\frac{1}{2}(p_1 - p_2) + p_2 - p_1 p_2, p_1 p_2 + \frac{1}{2}(1 - p_1 - p_2)) - H(p_1) - H(p_2),$$

where equality is held when the input distribution is uniform. Hence the maximum rate $\sup_{p(x)} I(X; Y_1, Y_2)$ of a the BSC relay channel is

$$\begin{aligned} C(\infty) &= \sup_{p(x)} I(X; Y_1, Y_2) \\ &= H(\frac{1}{2}(1 - p_1 - p_2) + p_1 p_2, p_1 - p_1 p_2 + \frac{1}{2}(p_2 - p_1), \\ &\quad \frac{1}{2}(p_1 - p_2) + p_2 - p_1 p_2, p_1 p_2 + \frac{1}{2}(1 - p_1 - p_2)) - H(p_1) - H(p_2), \end{aligned}$$

and,
$C(\infty) = H(\frac{1}{2}(1-p)^2 + \frac{1}{2}p^2, \frac{1}{2}p^2 + \frac{1}{2}(1-p)^2, p(1-p), p(1-p)) - 2H(p)$, if $p_1 = p_2 = p$.

3.4 Optimal Relay Rate for the BEC Relay Channel

In this section we will provide the optimal relay rate for the BEC relay channel by applying Theorem 13. Two cases will be discussed here: both the source-relay link and the source-destination link are BEC channels with the same erasure probability; and the source-relay link and the source-destination link are BEC channels with different erasure probabilities.

Case 1: suppose both the source-relay link and the source-destination link are BEC channels with the same erasure probability p, and e is the erasure symbol. The input distribution is $P(x = 0) = \pi, P(x = 1) = 1 - \pi$, $0 < p < \frac{1}{2}, 0 < \pi < 1$. Other related probabilities for the BEC relay channel ($\mathcal{X}$, $p(y_1, y_2|x)$, $\mathcal{Y}_1 \times \mathcal{Y}_2$) are given in the following tables:

Table 3.9: Distributions of X and Y_2 for the BEC relay channel (case 1)

X \ Y_2	0	e	1
$P(X)$	π	–	$1-\pi$
$P(Y_2)$	$\pi(1-p)$	p	$(1-\pi)(1-p)$

Table 3.10: Joint distribution of Y_1 and Y_2, $P(Y_1, Y_2)$ for the BEC relay channel (case 1)

Y_1 \ Y_2	0	e	1
0	$\pi(1-p)^2$	$\pi p(1-p)$	0
e	$\pi p(1-p)$	p^2	$(1-\pi)p(1-p)$
1	0	$(1-\pi)p(1-p)$	$(1-\pi)(1-p)^2$

Table 3.11: Conditional probabilities of Y_1 given Y_2, $P(Y_1|Y_2)$ for the BEC relay channel (case 1)

Y_1 \ Y_2	0	e	1
0	$1-p$	$\pi(1-p)$	0
e	p	p	p
1	0	$(1-\pi)(1-p)$	$1-p$

Case 2: suppose the source-relay link is BEC channel with erasure probability p_1 and the source-destination link is BEC channel with erasure probability p_2 and $p_1 \neq p_2$. e is the erasure symbol and the input distribution is $P(x=0) = \pi, P(x=1) = 1-\pi$, $0 < p_1, p_2 < \frac{1}{2}, 0 < \pi < 1$. Other related probabilities for the BEC relay channel ($\mathcal{X}$, $p(y_1, y_2|x)$, $\mathcal{Y}_1 \times \mathcal{Y}_2$) are given in the following tables:

From column 2 and column 7 in Table 3.12 and Table 3.16, we observe: given $y_2 = 0, x = 0$ or $y_2 = 1, x = 1$, both $y_1 = 0$ and $y_1 = 1$ satisfy $\frac{p(y_1|x)}{p(y_1|y_2)} = \frac{p(y_1'|x)}{p(y_1|y_2)}$. It seems that the condition in Theorem 13 is not held. However, for BEC channel, when $y_2 = 0$ or $y_2 = 1$,

Table 3.12: Values of $\frac{P(Y_1|X)}{P(Y_1|Y_2)}$ for the BEC relay channel (case 1)

Y_1 \ (X, Y_2)	$(0,0)$	$(0,e)$	$(0,1)$	$(1,0)$	$(1,e)$	$(1,1)$
0	1	$\frac{1}{\pi}$	–	–	0	–
e	1	1	–	–	1	1
1	–	0	–	–	$\frac{1}{1-\pi}$	1

Table 3.13: Distributions of X and Y_2 for the BEC relay channel (case 2)

X \ Y_2	0	e	1
$P(X)$	π	–	$1-\pi$
$P(Y_2)$	$\pi(1-p_2)$	p_2	$(1-\pi)(1-p_2)$

Table 3.14: Joint distribution of Y_1 and Y_2, $P(Y_1, Y_2)$ for the BEC relay channel (case 2)

Y_1 \ Y_2	0	e	1
0	$\pi(1-p_1)(1-p_2)$	$\pi(1-p_1)p_2$	0
e	$\pi p_1(1-p_2)$	p_1p_2	$(1-\pi)p_1(1-p_2)$
1	0	$(1-\pi)(1-p_1)p_2$	$(1-\pi)(1-p_1)(1-p_2)$

we know exactly what x is. In this case y_1 has no contribution to the entire communication system. Thus, we only consider the case that y_2 cannot decode the message, *i.e.* $y_2 = e$. Then the relay can aid the transmission. From the results in column 3 and column 6 in Table 3.12, given $y_2 = e$ and for any $y_1 \neq y_1'$, $\frac{p(y_1|x)}{p(y_1|y_2)} \neq \frac{p(y_1'|x)}{p(y_1'|y_2)}$ holds. Applying Theorem 13, we have the following proposition:

Proposition 4 *For the BEC relay channel, Slepian-Wolf coding is essentially the optimal relay strategy to achieve the maximum rate* $C(\infty) = \max_{p(x)} I(X; Y_1, Y_2)$ *and the minimum*

Table 3.15: Conditional probabilities of Y_1 given Y_2, $P(Y_1|Y_2)$ for the BEC relay channel (case 2)

Y_1 \ Y_2	0	e	1
0	$1-p_1$	$\pi(1-p_1)$	0
e	p_1	p_1	p_1
1	0	$(1-\pi)(1-p_1)$	$1-p_1$

Table 3.16: Values of $\frac{P(Y_1|X)}{P(Y_1|Y_2)}$ for the BEC relay channel (case 2)

Y_1 \ (X,Y_2)	$(0,0)$	$(0,e)$	$(0,1)$	$(1,0)$	$(1,e)$	$(1,1)$
0	1	$\frac{1}{\pi}$	$-$	$-$	0	$-$
e	1	1	$-$	$-$	1	1
1	$-$	0	$-$	$-$	$\frac{1}{1-\pi}$	1

relay rate R_0 is $\min_{p(x)} H(Y_1|Y_2)$.

Furthermore, we calculate the value of the maximum rate by referring to the values in Table 3.15:

$$\begin{aligned} I(X;Y_1,Y_2) &= H(Y_1,Y_2) - H(Y_1,Y_2|X) \\ &= H(Y_1,Y_2) - H(Y_1|X) - H(Y_2|X) \\ &= H(Y_1,Y_2) - H(p_1) - H(p_2). \end{aligned}$$

From Table 3.14, we have,

$$\begin{aligned}
& H(Y_1, Y_2) \\
= \; & H(\pi(1-p_1)(1-p_2), \pi(1-p_1)p_2, \pi p_1(1-p_2), p_1 p_2, \\
& (1-\pi)p_1(1-p_2), (1-\pi)(1-p_1)p_2, (1-\pi)(1-p_1)(1-p_2)) \\
= \; & (1-p_1p_2)H(\pi) + H(p_1) + H(p_2),
\end{aligned}$$

thus,

$$\begin{aligned}
I(X;Y_1,Y_2) \;\; & \leq \;\; \max_{p(x)} H(Y_1,Y_2) - (H(p_1)+H(p_2)) \qquad (3.16) \\
& = \;\; \max_{\pi}(1-p_1p_2)H(\pi) + H(p_1) + H(p_2) - (H(p_1)+H(p_2)) \\
& = \;\; \max_{\pi}(1-p_1p_2)H(\pi) \\
& = \;\; 1 - p_1p_2,
\end{aligned}$$

where the equality at (3.16) is held when the input of channel is uniform, *i.e.*, $\pi = \frac{1}{2}$. Hence, the maximum rate $\sup_{p(x)} I(X;Y_1,Y_2)$ of a the BEC relay channel is

$$C(\infty) = \sup_{p(x)} I(X;Y_1,Y_2) = (1-p_1p_2)$$
and,
$$C(\infty) = (1-p^2) \text{ if } \; p_1 = p_2 = p.$$

Remark: All the proofs and results are obtained basing on this assumption that $X_1, X_2, ..., X_n$ are i.i.d.. This assumption can be replaced by reducing the original open problem into a smaller problem that the coding scheme at the source is random coding according to any distribution $p(x)$, and we consider the best relay coding strategy to achieve the maximum rate $\sup_{p(x)} I(X;Y_1,Y_2)$.

Chapter 4

Conclusions and Future Work

4.1 Conclusions

In this thesis, we investigated Cover's open problem on the capacity of the relay channel. Our goal was to find the smallest relay rate such that the maximum message rate can be achieved. Under the assumption: $X_1, X_2, ..., X_n$ are i.i.d., our work gave partial solutions to our original problem in two special cases: the BSC relay channel and the BEC relay channel.

Our main results are:

1. A "single-letter criterion" is given as shown below.

   Single-letter Criterion:

   For the channels satisfying: "given any x, y_2, for any $y_1 \neq y_1'$, $\frac{p(y_1|x)}{p(y_1|y_2)} \neq \frac{p(y_1'|x)}{p(y_1'|y_2)}$", Slepian-Wolf coding is essentially the optimal relay strategy to achieve the maximum rate $C(\infty) = \sup_{p(x)} I(X; Y_1, Y_2)$ and $R_0 = \min_{p(x)} H(Y_1|Y_2)$ is the minimum relay rate.

2. For both the BSC relay channel and the BEC relay channel, we proved that Slepian-Wolf coding is essentially the optimal relay strategy to achieve the maximum rate and $R_0 = H(Y_1|Y_2)$ is the minimum relay rate to achieve the maximum rate.

3. The maximum rates $\sup_{p(x)} I(X; Y_1, Y_2)$ for the BSC relay channel and the BEC relay channel were given respectively.

BSC relay channel:

$$\begin{aligned}C(\infty) &= \sup_{p(x)} I(X;Y_1,Y_2)\\ &= H(\frac{1}{2}(1-p_1-p_2)+p_1p_2, p_1-p_1p_2+\frac{1}{2}(p_2-p_1),\\ &\quad \frac{1}{2}(p_1-p_2)+p_2-p_1p_2, p_1p_2+\frac{1}{2}(1-p_1-p_2))-H(p_1)-H(p_2),\end{aligned}$$

and,

$$C(\infty) = H(\frac{1}{2}(1-p)^2+\frac{1}{2}p^2, \frac{1}{2}p^2+\frac{1}{2}(1-p)^2, p(1-p), p(1-p)) - 2H(p), if\ p_1 = p_2 = p.$$

BEC relay channel:

$$C(\infty) = (1-p_1p_2), and, C(\infty) = (1-p^2), if\ p_1 = p_2 = p.$$

4.2 Future Work

There are much to be done in this area of research. The followings are a few possible extensions.

1. If lossy compression is applied at the relay, we can potentially find a relay rate smaller than $H(Y_1|Y_2)$. For future work, we can try to figure out the specific condition under which a smaller relay rate can be obtained, and what the exact value of the smaller relay rate is.

2. We conjecture that if the "sing-letter criterion" is not satisfied, then Slepian-Wolf coding is not optimal. For future work, we can try to verify this conjecture.

References

[1] S. Kulkarni A. Reznik and S. Verdu. Capacity and optimal resource allocation in the degraded gaussian relay channel with multiple relays. *40th Allerton conf. Commun., Contr. Comput.*, Oct. 2002. 1, 14

[2] R. Ahlswede. Multi-way communication channels. *Proc. 2nd Int. Symp. Inform. Theory, : Armenian S.S.R*, pages 23–52, 1971. 1

[3] R.F. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE, Trans.Inform.Thoery*, IT-21:629–632, Nov. 1975. 32

[4] T.M. Cover. Broadcast channels. *IEEE, Trans. Inform. Theory*, IT-18:2–14, 1972. 1

[5] T.M. Cover. A proof of the data compression theorem of slepian and wolf for ergodic sources. *IEEE, Trans. Inform.Theory*, 21:226–228, Mar. 1975. 3, 16

[6] T.M. Cover and A.E. Gamal. Capacity theorems for the relay channel. *IEEE, Trans.Inform. Theory*, 25:572–584, 1979. 1, 14

[7] T.M. Cover and B. Gopinath. *Open Problems in Communication and Computation.* Spring-Verlag Inc., N.Y., 1987. iii, 1

[8] T.M. Cover and J. Thomas. *Elements of Information Theory.* Wiley, N.Y., 1991. 5, 14, 15, 16

[9] H.G. Eggleston. *Convexity.* Cambridge University Press, N.Y., 1963. 32

[10] M. Gastpar G. Kramer and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE, Trans. Inf. Theory*, IT-51:3037–3063, 2005. 1

[11] R.G. Gallager. *Information theory and Reliable Communication*. Wiley, N.Y., 1968. 5, 10

[12] A.E. Gamal and M. Aref. The capacity of the semideterministic relay channel. *IEEE, Trans. Inform. Theory*, IT-28:536, May 1982. 1

[13] A.E. Gamal and Y.H. Kim. Lecture notes on network information theory. *Stanford University and UCSD*, 2009. 14

[14] M. Gastpar and M. Vetterli. On the capacity of wireless networks: the relay case. *IEEE INFO.COM*, 3:1577–1586, 2002. 1, 14

[15] P. Gupta and P.R. Kumar. Towards and information theory of large networks: An achievable rate region. *IEEE, ISIT*, 49:1877–1894, Aug. 2003. 1

[16] D.-K. He J. Chen and E.-H. Yang. On the codebook-level duality between slepian-wolf coding and channel coding. *Information Theory and Applications Workshop*, pages 84–93, 2007. 3, 16

[17] Y.-H. Kim. Coding techniques for primitive relay channels. *Proc. Forty-Fifth Annual Allerton conf. Commun., Contr. Comput.*, 2007. 1

[18] G. Kramer. Capacity results for the discrete memoryless network. *IEEE, Trans. Inform. Theory*, 49:4–21, Jan. 2003. 1

[19] J. Kusuma. Slepian-wolf coding and related problems, 2001. `http://www.mit.edu/~6.454/www_fall_2001/kusuma/summary.ps`. 3, 16

[20] E.C. Van Der Meulen. Three-terminal communication channels. *Adv. Appl. Prob.*, 3:120–154, 1971. 1

[21] M.B. Pursley. *Introduction to Digital Communications*. Pearson Prentice Hall, 2005.

[22] M. Salehi. Cardinality bounds on auxiliary variables in multiple-user theory via the method of ahlswede and körner. *Stanford University*, 1978. 32, 33

[23] B. Schein and R. Gallager. The gaussian parallel relay network. *ISIT*, page 22, Jun. 2000. 1, 14

[24] C.E. Shannon. A mathematical theory of communication. *Bell Sys. Tech*, 27:379–423, 623–656, 1948. 13

[25] D. Slepian and J.K. Wolf. Noiseless coding of correlated information sources. *IEEE, Trans. Inform.Theory*, 19:47–480, Jul. 1973. 3, 16

[26] X. Wu and L.-L. Xie. On the optimal compressions in the compress-and-forward relay schemes, 2011. `https://ece.uwaterloo.ca/~llxie/pdf_files/1009.5959v3.pdf`. 1, 14

[27] L.-L. Xie. An improvement of cover/el gamal's compress-and-forward relay scheme, Aug. 2009. `http://arxiv.org/abs/0908.0163`. 1, 14

[28] Z. Zhang. Partial converse for a relay channel. *IEEE, Trans. on Inform. Theory*, IT-34:1106–1110, Sep. 1988. 1