CAUSAL COGNITIVE RADIO: AN INFORMATION THEORETIC PERSPECTIVE

Seyed Hossein Seyedmehdi

A THESIS
SUBMITTED FOR THE DEGREE OF MASTER OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
AT
NATIONAL UNIVERSITY OF SINGAPORE

AUGUST 2008

NATIONAL UNIVERSITY OF SINGAPORE
DEPARTMENT OF
ELECTRICAL AND COMPUTER ENGINEERING

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "**Causal Cognitive Radio: An Information Theoretic Perspective**" by **Seyed Hossein Seyedmehdi** in partial fulfillment of the requirements for the degree of **Master of Engineering**.

Dated: August 2008

Supervisors:

Yong Lian

Yan Xin

Readers:

NATIONAL UNIVERSITY OF SINGAPORE

Date: **August 2008**

Author: **Seyed Hossein Seyedmehdi**

Title: **Causal Cognitive Radio: An Information Theoretic Perspective**

Department: **Electrical and Computer Engineering**

Degree: **M.Eng.** Convocation: **August** Year: **2008**

Permission is herewith granted to National University of Singapore to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

Signature of Author

To my parents,

my sisters,

and my brother.

Table of Contents

List of Tables

List of Figures

Abstract

Nowadays, not only is the current frequency spectrum almost completely allocated, but also the demand for it is daily increasing. According to recently released studies [1], less than 1/5 of the currently licensed frequency spectrum is being efficiently used. This fact have motivated considerable research efforts on improving spectral utilization efficiency. Recently, emerging as a promising technology to achieve this improvement, *cognitive radio* (CR) has been proposed as a new form of cooperative model for wireless communications. One important feature of the CR is that the secondary (cognitive) user is allowed to coexist with the primary (licensed) users. The key idea in CR is that the cognitive user is assumed to be an intelligent user which is capable of sensing and perceiving the environment so that it adapts its way of communication in order to enhance the performance.

In this work, a causal (non-anticipating) cognitive radio (CCR) model is proposed in which the primary and secondary users transmit their messages simultaneously during all the transmission time. In this model, not only is the secondary user a sender with a message to send, but it also acts as a relay which cooperates with the primary user. We refer such a model as interference channel with Causal Unidirectional Cooperation (IC-CUC) or CCR model. An achievable rate region for the IC-CUC is established using a combination of various encoding and decoding methods. Also, the derived achievable rate region in the Gaussian case is demonstrated and compared with the existing results.

Acknowledgements

I would like to thank Professor Yan Xin, my supervisor, for his many suggestions and constant support during this research. I am also thankful to him for his guidance through the period of my chaos and confusion. I also would like to thank Professor Lian Yong for his encouragement and help.

I am also very grateful that I had the chance to join I^{2}R Wireless Communication Laboratory at National University of Singapore. The *A*STAR International Graduate Scholarship*, which was awarded to me for the period 2006–2008, was a crucial support to successful completion of this thesis.

Moreover, I am grateful to my parents for their patience and *love*. Without them this work would never have come into existence (literally).

Finally, I wish to thank the following: Jiang Jinhua, Zhu Yonglan, and Zhang Lan (for their friendship and constructive discussions on the the research topic); *and* Fatemeh, Zahra, and Mohammad (for their dedication and constant support during my education).

NUS, Singapore — Seyed Hossein Seyedmehdi
August 7, 2008

Chapter 1

Introduction

1.1 Background

Wireless communications began to develop in 1888 when H. R. Hertz demonstrated the theory of electromagnetic waves. Transmitting data to further distances with higher transmission rates was later made possible based on the advances in analog modulations and demodulation techniques (such as AM and FM). The main bottleneck at that time was the thermal noise at the electronic circuits (amplifiers, mixers, and filters). At that point, it was widely perceived that the only way to reduce the communication error is to increase the transmitter power while zero error deemed to be practically imposable. In 1948, Shannon opened a new chapter in communication theory when he founded what today is known as *information theory*. In his original paper [38], he showed that for a prescribed transmitter power, one can achieve *reliable* communications by incorporating a proper coding scheme. He also showed that for any rate more than the *channel capacity* there will be an inevitable error independent

of the applied technique[1]. He also obtained this capacity as

$$C = \max_{p(x)} I(X;Y) \tag{1.1.1}$$

for a point-to-point discrete memoryless channel (DMC) with the channel input X and channel output Y. His work proved that there is a code which can achieve the capacity of the channel, but it did not show how to construct such a code.

Since then, there has been a large number of research efforts in this area. One interesting question to be addressed is the achievable limits when the number of users is more than one. Multiple Access Channel (MAC), Broadcast Channel (BC), and Interference Channel (IC) can be pointed out as appealing examples of multiuser channels. The capacity region of the MAC was first obtained by Ahlswede [2]. This channel is of interest because the uplink in mobile communications can be modeled by a MAC. Downlink can, on the other hand, be modeled by a BC. The capacity region of the BC is unknown in general, and it is only known when the channel is degraded. Cover [9] and Gallager [17] investigated the degraded BC and obtained its capacity. The best inner bound on the capacity of this channel was established by Marton [30]. Sato [33] obtained a general form for the outer bound of such a channel. The IC models the communication scenario in which each of users has its own transmitter and receiver. The capacity of this channel is unknown in general. This capacity is only known in some special cases [8, 34]. The best achievable rate region (the inner bound) until today is obtained by Han and Kobayashi (HK) [20] whereas the tightest outer bounds in the Gaussian case so far were given by [16, 37].

[1]Shannon's work can be juxtaposed with the revolutionary work of Einstein (1905) in which he claimed that "nothing can travel faster than light", but he did not propose any solution how to reach the speed of light.

In addition to mentioned channel types, *Relay Channel* (RC) [42] is another appealing type of multi user communication channels. Relay is an extra node that receives the signal from the sender and *cooperate* with the sender via *decode-and-forward*, *compress-and-forward*, or *amplify-and-forward*. Most of the information theoretic results have been obtained by Cover, El Gamal, and Aref [10, 14]. The capacity of the RC is known when the relay is degraded from the primary receiver. This capacity was obtained by applying Block-Markov superposition encoding and list decoding by Cover [10] where he used a binning method to convey the message in two steps. Later, it was shown that *sliding window* decoding [47] achieves the same performance. Unlike the binning method, a window of $k+1$ blocks of codewords for k relays (two blocks in the case of single relay) are used in the sliding window decoder to decode the message of the first codeword of the block. At the sender, the messages are superimposed onto each other (for single relay, each new message is superimposed onto the previous message). Willems [44] proposed a decoding technique know as *backward decoding* for the MAC with feedback. This decoding method was later shown to be capacity achieving for the RC as well [47]. This coding technique was simpler with the cost of delay in decoding at the receiver. Xie and Kumar [48] have shown that the backward decoding can achieve a higher rate in comparison with the sliding window decoding when there are multiple independent sources.

In addition to relaying, another scenario of cooperation can arise when a certain type of *side information* is available at the transmitter, receiver, or both. Particulary, this side information can be the state of the channel. When this Channel Sate Information (CSI) is available at the transmitter (CSIT), there are two different situations depending on how the CSIT has been acquired. In the first one, before sending a

symbol through the channel, encoder is aware of the channel state that this symbol is going to encounter. In other words, encoder has a knowledge about the CSI (only) until the present time. In the second situation, the transmitter knows the complete CSI before sending a codeword X^n. The former is termed as *causal CSIT*, and the latter is termed as *non-causal CSIT*.

Shannon [39], as a pioneer, investigated the capacity of the point-to-point channel when causal CSI is available at the transmitter. By constructing an equivalent channel, which has the same capacity as the original one, he obtained the capacity of such channels. He also showed that knowing only the current channel state would result in the same capacity. Ever since, there has been numerous works examining different types of side information in order to find the performance limits of the underlying communication systems (a comprehensive relevant subject review has recently been given by Keshet et al. [24]). For instance, Shannon model was studied in the multi user configurations by Sigurjonsson and Kim [40]. The non-causal CSI was first studied by Kuznetsov and Tsybakov [28] in which they proposed a primitive coding scheme for this channel. Later, Gel'fand and Pinsker [19] found its capacity as

$$C = \max_{p(u,x|s)} \big(I(U;Y) - I(U;S)\big), \tag{1.1.2}$$

where U is an auxiliary random variable with finite cardinality, and the maximization is subjected to the constraint that $U \to (S, X) \to Y$ forms a Markov chain. This result was extended to the Gaussian channel by Costa [7] which is today known as *Dirty Paper Coding* (DPC). Costa showed that any (White Gaussian) additive interference known at the transmitter can surprisingly be canceled thoroughly at the receiver. As one special case, the encoder can (perfectly) overhear the message of the other user(s) in a causal manner. This encoder which we term it as *Causal Cognitive*

Encoder motivates investigating channels in which the encoder is able to perceive the channel state. Encoders with overhearing capability were first studied by Willems [45] in *multiple access channels with cribbing encoders*. Using the backward decoding technique [44, 50], the capacity region of such channels was obtained. Willems' results were later generalized by Khojastepour et al. [25] to the case in which each encoder receives a noisy version of the other encoder's codeword. Recently, Tuninetti [41] and Cao [5] have studied the CE's in an IC shell.

According to a report recently released by FCC [1], only 15 percent of the currently licensed frequency spectrum is being efficiently used. Besides, the demand for frequency spectrum is rapidly increasing. These two issues have motivated considerable research efforts on improving spectral utilization efficiency. Recently, emerging as a promising technology to achieve this improvement, *Cognitive Radio* (CR) [21] has been proposed as a new form of *cooperative* model for the wireless communications. One important feature of the CR is that the secondary (cognitive) users are allowed to coexist with the primary (licensed) users. The key idea in CR is that the cognitive user is assumed to be a *smart* user which is capable of *sensing* and *perceiving* the environment and *adapting* its way of communication in order to enhance the performance. Most of the previous studies of the CR assume that the secondary user has complete or partial *a priori* (non-causal) knowledge about the message being sent by the primary user [13, 22, 23, 29, 46]. This *genie-aided* CR [13] is also known as the Interference Channel with Degraded Message Set (IC-DMS) [23] or Interference Channel with *Unidirectional Cooperation*. In a general sense, IC-DMS refers to an IC in which primary user transmits its message to the respective receiver, and the cognitive user as a secondary sender has a non-causal knowledge about the message

of the fist sender. This knowledge can be complete as discussed in [13] or partial as studied in [29]. Having the non-causal state information at the cognitive user motivates using DPC to mitigate (or even eliminate in the Gaussian case) the undesired effect of interference at the receivers as well as cooperation with the first sender to facilitate data transferring to the first receiver. This facilitation is feasible because the cognitive user which knows the message of the first sender can allocate a portion of its power to transmit the message of the first user. As mentioned, the information theoretic studies on the CR were initialized by Devroye et al. [13], in which a rate splitting technique and DPC is used to develop an achievable rate region for the IC-DMS. This region is later improved by [22] using a combination of DPC, cooperation, and collaboration. Like [13, 20], [22] also uses the rate splitting technique for encoding the message in the first sender to enlarge the achievable rate region. The results of [22] has been expanded by Marić et al. [29] to a more general case when the cognitive user has a partial knowledge about the message of the primary user. In [23] the capacity of cognitive user without sacrificing the rate of the first user is determined for the Gaussian channel. This capacity is, however, valid only for the case when the communication link between the cognitive sender and primary receiver is weak.

Nevertheless, the primary assumption of *non-causal* knowledge requires that the secondary user has a priori knowledge about the message of the primary user before the message is actually transmitted. This assumption may not be feasible in realistic communication scenarios where the senders and receivers are non-anticipating (causal). The causal CR was first investigated in [13]. Specifically, the paper [13] adopts a two-phase transmission protocol in which the first phase of transmission is

solely allocated for the secondary user to perfectly obtain causal knowledge about the message being sent from the primary user while in the second phase both the primary and secondary users are allowed to transmit their messages simultaneously.

1.2 Contributions and Outline

In this dissertation, we propose a new Causal (*non-anticipating*) Cognitive Radio (CCR) model in which the primary and secondary users transmit their message simultaneously during all the transmission time. In this model, not only is the secondary user a sender that has its own message to send, but it also acts as a relay [10] which cooperates with the primary user. We refer such a model as the *Interference Channel with Causal Unidirectional Cooperation* (IC-CUC). Our model is analogous to the notion of generalized feedback proposed in [41] and interference channels with conferencing [5]. However, in [41, Theorem 1], each sender (performing partially decode and forward) is less capable than receivers in decoding the message of the other sender. We establish an achievable rate region for the IC-CUC by using a combination of rate splitting, block Markov superposition encoding [10, 27, 9], and sliding-window decoding [47]. We also demonstrate the derived achievable rate region in the Gaussian case and compare it with the existing result in [13]. Result of this research has been published in [36].

This dissertation is organized as following:

In Chapter 2, the necessary mathematical tools needed throughout the work is studied. All of the definitions and theorems follow standard information theory text books [11, 12, 18, 31, 49]. In the first part of this chapter, the concept of entropy and mutual information are introduced. Next, the method of type and typical sequences

are discussed. Typical sequences are used in decoder design where the decoder seeks in the set of codewords to find a message whose codeword is jointly typical with the channel output. Then, the differential entropy is defined and it is shown that the signals with normal distribution are entropy maximizers among all input signals with the same variance. Moreover, this chapter elaborates on important results in information theory. As a case in point, the Shannon capacity is discussed and the random coding is shown to be capacity achieving.

The main contribution of this dissertation is included in Chapter 3. As mentioned, a new model for the CR is proposed and its performance is studied. To analyze the performance, we first obtain an inner bound. Then, the inner bound is illustrated in the Gaussian case and it is shown that it outperforms the existing results.

In Chapter 4, the concluding remarks are pointed out. Moreover, the further research is outlined, and some heuristics that can potentially improve the performance are discussed.

In Appendix A, the Fourier Motzkin elimination method for the Theorem 3.2.1 is derived in details.

Chapter 2

Preliminaries

This chapter is devoted to develop mathematical tools needed to analyze multiuser channels. First, the axioms of probability and measure theory are introduced. Then, *entropy* is shown to be an appropriate criteria to measure the amount of information contained in a random variable (RV). Next, the amount of information gain about a RV by knowing another RV is given. In this concept, *Kullback–Libler* distance is stated as a non-metric measure for the distance of two probability functions. Moreover, the method of types as a powerful technique for bounding error probability is discussed. Lastly, these results are extended to continuous random variables and it is shown that Gaussian distribution maximizes the entropy.

2.1 The Axioms of Probability Theory

Throughout this dissertation, capital letters (A), lower case letters (a), and calligraphic letters ($\mathcal{A}$) denote RV's, their sample values, and their alphabets respectively. A similar convention is used for the random vectors and their values. A *Kolmogorov probability space* is shown by a triple $(\Omega, \mathcal{F}, \mu)$ [4]. The first component, Ω, is a nonempty set comprising all possible outcomes. Each element ω of Ω is called an

outcome and Ω is called the *sample space*. The second component, $\mathcal{F}$, is a σ-algebra set including *events* which are subsets of Ω. The third component, μ, is a probability measure on $\mathcal{F}$. Being a σ-algebra set means that $\mathcal{F}$ satisfies

(i) $\Omega \in \mathcal{F}$,

(ii) $\forall A \subset \Omega, A \in \mathcal{F} \rightarrow A^c \in \mathcal{F}$,

(iii) $A, B \in \mathcal{F} \rightarrow A \cup B \in \mathcal{F}$.

Also, the μ-measure $\mu : \mathcal{F} \mapsto [0, 1]$ satisfies

(i) $\mu(A) \geq 0$, $\forall A \in \mathcal{F}$,

(ii) $\forall A, B \in \mathcal{F}, A \cap B = \emptyset \rightarrow \mu(A \cup B) = \mu(A) + \mu(B)$,

(iii) $\mu(\Omega) = 1$.

As an example, consider the experiment of tossing a coin. The possible outcomes of this experience are "heads" (H) and "tails" (T). Then, the set Ω is $\{T, H\}$, and the set of events is $\mathcal{F} = \{\{\}, \{H\}, \{T\}, \{H, T\}\}$. Given that the coin is fair, the μ measure on $\mathcal{F}$ can be written as

$$\mu(\{\}) = 0, \quad \mu(\{H\}) = \tfrac{1}{2}, \quad \mu(\{T\}) = \tfrac{1}{2}, \quad \mu(\{H, T\}) = 1.$$

A RV X in a finite set $\mathcal{X}$ is a mapping $X : \Omega \mapsto \mathcal{X}$ such that $X^{-1}(x) \in \mathcal{F}$ for every $x \in \mathcal{X}$. The probability of an event defined in terms of RV's means μ-measure of the corresponding subset of Ω, e.g.,

$$\Pr\{X \in A\} \triangleq \mu(\{\omega : X(\omega) \in A, \omega \in \Omega\}).$$

The notation of "$X \overset{\text{iid}}{\sim} p(x)$" is used to denote that the RV X is drawn independent and identically distributed (i.i.d.) according to the probability measure $p(\cdot)$ on $\mathcal{X}$.

The notation of X_i^j, $i \leq j$ is used to show the vector $(X_i, X_{i+1}, \cdots, X_j)$. For brevity, X_1^j is shown by X^j, i.e., the index i is omitted when $i = 1$. A sequence of tuples on $\mathcal{X}^n \times \mathcal{Y}^n \times \cdots \times \mathcal{Z}^n$ is shown by $(x^n, y^n, \cdots, z^n)$ which is by definition equivalent to $\big((x_1, y_1, \cdots, z_1), (x_2, y_2, \cdots, z_2), \cdots, (x_n, y_n, \cdots, z_n)\big)$. The set cardinality is denoted by $|\cdot|$, and the empty set is denoted by $\emptyset$ (clearly, $|\emptyset| = 0$). The events in the most general sense are usually shown by $E_I(\cdot)$ where I pertains to the node in which the event happens. In addition, the compliment of an event $E_I(\cdot)$ is shown by $E_I^c(\cdot)$.

The number of occurrence of a specific symbol, $a \in \mathcal{X}$, in a sequence x^n is shown by the notation $N(a\,;\, x^n)$.

2.2 An Information Measure

We first introduce the concept of *information*[1] of a random event. An event can be outcome of an experiment, received symbol in a communication channel, etc. The term *information* is analogous to the words *surprise* and *uncertainty*, and these three terms usually convey the same concept. Before the event there is an amount of uncertainty. When the event happens, there is an amount of surprise. After the event, there is a gain in the amount of information. Let the RV X represent a sample event with the probability $p(x)$. We use the notation $\mathfrak{I}(p)$ as a measure to determine

[1]There are two widely used information measures in communications theory which are *entropy* and *Fisher information*. The entropy will be introduced in this work, and Fisher information is defined as the variance of the *score* function, i.e., $J(\theta) = E_\theta\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2$. Interestingly, these two information measures are related to each other. It is shown in [11, Section 17.8] that while entropy is related to the volume of typical sets, fisher information is related to the surface area of the typical set with respect to the definition of continuous typical sets stated in Definition 2.5.3.

the amount of information carried by the event outcome x. From what we mentioned above, it can be intuitively understood that the amount of information of an event x is inversely related to the probability of occurrence of that event. The more we expect an event to happen beforehand, the less we get surprised by the occurrence of that event, or in other words, the less information we gain by knowing that event has happened. Moreover, $\mathfrak{I}(p)$ has to be possessed of the following properties

(i) Information of an event is positive, i.e., $\mathfrak{I}(p) \geq 0$.

(ii) The information measure should be additive, i.e., $\mathfrak{I}(p_1 p_2) = \mathfrak{I}(p_1) + \mathfrak{I}(p_2)$.

(iii) $\mathfrak{I}(p)$ is a continuous function on p.

A logarithmic function of the probability distribution satisfies these three properties, i.e., $\mathfrak{I}(p) = \log \frac{1}{p(x)}$. If the base of logarithmic function is 2, the information is measured in bits, and if the base is e, the information is measured in nats[2].

While $\mathfrak{I}(p_i)$ is the amount of information (uncertainty) of the variable x_i, our objective is to know how much information the RV X contains in average. The expected value of the $\mathfrak{I}(p)$ is the desired quantity and is called *entropy* of the random variable X. In brief, the entropy of a RV measures the uncertainty of that RV. In other words, it gives the amount of information required to describe a RV.

Definition 2.2.1. The *entropy* $H(X)$ of a discrete RV X is defined by

$$H(X) = E_p \log \frac{1}{p(X)} = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)}. \tag{2.2.1}$$

However, if the RV X is binary, i.e.,

$$X = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p, \end{cases} \tag{2.2.2}$$

[2]Throughout this dissertation it is assumed that all the logarithmic functions are taken in base 2 unless otherwise is stated.

where $0 \leq p \leq 1$, we use an alternative representation for $H(X)$ as follows

$$H(p) = -\left(p \log p + (1-p) \log(1-p)\right). \tag{2.2.3}$$

Definition 2.2.2. The *joint entropy* $H(X,Y)$ of a pair of discrete random variables (X,Y) with a joint probability distribution $p(x,y)$ is defined as

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y) \log p(x,y). \tag{2.2.4}$$

Definition 2.2.3. The *conditional entropy* $H(Y|X)$ for $(X,Y) \sim p(x,y)$ is defined as

$$\begin{aligned} H(Y|X) &= \sum_{x\in\mathcal{X}} p(x) H(Y|X=x) && (2.2.5)\\ &= -\sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} p(y|x) \log p(y|x) && (2.2.6)\\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y) \log p(y|x) && (2.2.7)\\ &= -E \log p(Y|X). && (2.2.8) \end{aligned}$$

Theorem 2.2.1 (Chain rule)**.** *[11, Theorem 2.2.1, page 17]*

$$H(X,Y) = H(X) + H(Y|X). \tag{2.2.9}$$

The chain rule can be generalized to multiple random variables as following

$$H(X^n) = \sum_{i=1}^{n} H(X_i|X^{i-1}). \tag{2.2.10}$$

For the sake of convention, we sometimes use $H(P)$ or $H(p)$ instead of $H(X)$ if the RV X is drawn according to the probability distribution $p(x)$.

2.3 Distance of Probability Distributions and Mutual Information

Suppose there are two different probability distributions $p(x)$ and $q(x)$ on the RV x. Knowing how much these two probability distributions are *similar* to each other plays

an important role to conceive communication problems. To address this question, a measure to specify the *distance* between two distributions is needed. To define such a measure, we use the previously defined measure of the information of each individual distributions. The difference between the information of $p(x)$ and $q(x)$, i.e., $d_{q||p}(x) = \mathfrak{I}(q(x)) - \mathfrak{I}(p(x))$ will be a criteria showing how *far* two distributions are at the sample point x. We are interested to know how far two distributions are in average, and hence we take the expected value of $d_{q||p}(X)$. Since it does not make sense if the distance is negative, the expected value is taken with respect to $p(x)$ so that this parameter becomes positive.

Definition 2.3.1. The *relative entropy* or *Kullback–Libler (KL) distance*[3] between two probability mass functions $p(x)$ and $q(x)$ is defined as

$$\begin{aligned} D(p||q) &= E_p[\mathfrak{I}(q) - \mathfrak{I}(p)] && (2.3.1)\\ &= E_p \log \frac{1}{q(X)} - E_p \log \frac{1}{p(X)} && (2.3.2)\\ &= \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}. && (2.3.3) \end{aligned}$$

In this definition, we used the conventions that $0 \log \frac{0}{0} = 0$, $0 \log \frac{0}{q} = 0$, and $p \log \frac{p}{0} = \infty$.

To introduce the concept of *mutual information*, consider a communication channel with the channel input symbol x_i and the channel output symbol y_i in i-th channel use. Prior to reception of y_i, *a priori* probability that x_i was sent is $p(x_i)$. After receiving y_i, *a posteriori* probability that x_i was sent is $p(x_i|y_i)$. In other words, there

[3] A function $\rho(x, y)$ is metric if for all x, y,

- $\rho(x, y) \geq 0$,
- $\rho(x, y) = \rho(y, x)$,
- $\rho(x, y) = 0$ if and only if $x = y$,
- $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$.

Consequently, the KL distance is not metric.

is an information gain on what has been sent due to reception of y_i. This information gain is shown by $I(x_i; y_i) = \log(1/p(x_i)) - \log(1/p(x_i|y_i))$ where $-\log p(x_i)$ is the amount of uncertainty (information) in x_i, and $-\log p(x_i|y_i)$ is the amount of uncertainty (information) in x_i after knowing y_i. Therefore, $I(x_i; y_i)$ is the amount of information (in bits) which has been transferred through the channel. As usual, we are interested to know how much we achieve in average.

Definition 2.3.2. The *mutual information* $I(X;Y)$ of two RV's X, Y with joint distribution $p(x, y)$ and marginal distributions $p(x)$ and $p(y)$ is defined as

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && (2.3.4)\\ &= D(p(x,y)||p(x)p(y)). && (2.3.5) \end{aligned}$$

Fig. 2.1 shows the relationship between entropy and mutual information in a Venn diagram. As can be seen, $I(X;Y)$ represents the common part of $H(X)$ and $H(Y)$. From the diagram the following can be inspected

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && (2.3.6)\\ &= H(Y) - H(Y|X) && (2.3.7)\\ &= H(X) + H(Y) - H(X,Y) && (2.3.8)\\ &= I(Y;X). && (2.3.9) \end{aligned}$$

2.4 Typical Sequences

Definition 2.4.1. A sequences $x^n \in \mathcal{X}^n$ is said to be *ϵ-strongly typical* if the sample frequencies are *close* to the true values. More precisely,

$$T_\epsilon^{(n)}(X) = \left\{ x^n \in \mathcal{X}^n : \begin{array}{ll} \left|\frac{1}{n}N(a\,;\,x^n) - P(a)\right| < \frac{\epsilon}{|\mathcal{X}|}, & \text{if } P(a) > 0 \\ N(a\,;\,x^n) = 0, & \text{if } P(a) = 0 \end{array} \right\}. \qquad (2.4.1)$$

In other words, type of any sequence in the typical set does not differ more than $\epsilon/|\mathcal{X}|$ form true probability in any component.

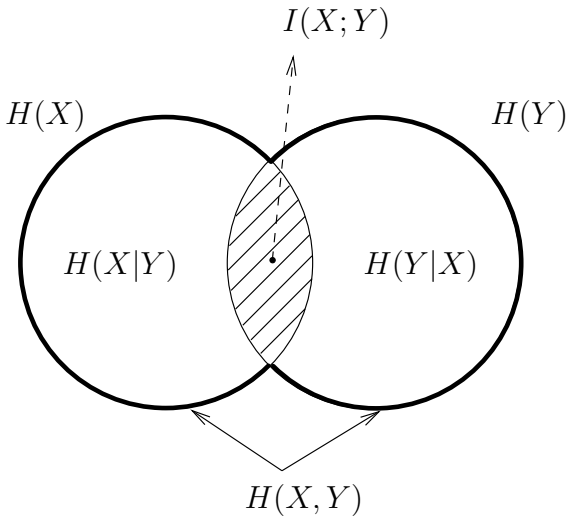

Figure 2.1: Relationship between entropy and mutual information. Entropy of each set is shown by a circle. The intersection of two circles represents the amount of mutual information as shown on the figure.

Definition 2.4.2. A sequence of tuples $(x^n, y^n, \cdots, z^n) \in \mathcal{X}^n \times \mathcal{Y}^n \times \cdots \times \mathcal{Z}^n$ is said to be *ϵ-strongly typical* with respect to distribution $p(x, y, \cdots, z)$ on $\mathcal{X} \times \mathcal{Y} \times \cdots \times \mathcal{Z}$ if

(i) For all $(x, y, \cdots, z) \in \mathcal{X} \times \mathcal{Y} \times \cdots \times \mathcal{Z}$, we have

$$\sum_{\bar{u}} \left| \frac{1}{n} N(\bar{u}\,;\, \bar{\bar{u}}) - p(\bar{u}) \right| < \epsilon, \tag{2.4.2}$$

where $|\bar{\mathcal{U}}| = |\mathcal{X}||\mathcal{Y}| \cdots |\mathcal{Z}|$, $\bar{u} = (x, y, \cdots, z)$, $\bar{\bar{u}} = (x^n, y^n, \cdots, z^n)$, and $N(\bar{u}\,;\, \bar{\bar{u}})$ is the number of occurrence of $(x, y, \cdots, z)$ in the sequence of tuples $(x^n, y^n, \cdots, z^n)$.

(ii) For all $(x, y, \cdots, z) \in \mathcal{X} \times \mathcal{Y} \times \cdots \times \mathcal{Z}$ with $p(x, y, \cdots, z) = 0$, $N(x, y, \cdots, z\,;\, x^n, y^n, \cdots, z^n) = 0$.

The set of sequences $(x^n, y^n, \cdots, z^n) \in \mathcal{X}^n \times \mathcal{Y}^n \times \cdots \times \mathcal{Z}^n$ such that $(x^n, y^n, \cdots, z^n)$ is ϵ-strongly typical is called the *strongly typical* set and is denoted by $T_\epsilon^{(n)}(XY \cdots Z)$ or $T_\epsilon^{(n)}$when the random variables are understood from the context. An alternative definition of typical sets based on Kullback-Leibler distance of empirical probability

distribution of x^n (i.e., $P_{x^n}(x^n)$) and true probability distribution of x (i.e., $Q(x)$) is given in [11, Section 11.2].

Definition 2.4.3. A sequence $y^n \in \mathcal{Y}$ is said to be *ϵ-strongly conditionally typical* with the sequence x^n with respect to the conditional distribution $P_{Y|X}(\cdot|\cdot)$ if

(i) For all $(a, b) \in \mathcal{X} \times \mathcal{Y}$ with $P(b|a) = P_{Y|X}(b|X = a) > 0$

$$\frac{1}{n} |N(a, b\,;\, x^n, y^n) - P(b|a)N(a\,;\, x^n)| \leq \frac{\epsilon}{|\mathcal{Y}| + 1}, \tag{2.4.3}$$

(ii) $N(a, b\,;\, x^n, y^n) = 0$ for all (a, b) such that $V(b|a) = 0$.

The set of all such sequences is called as the *conditionally typical set* and denoted by $T_\epsilon^{(n)}(Y|X^n = x^n)$ or in abbreviated form as $T_\epsilon^{(n)}(Y|x)$.

Lemma 2.4.1 (Asymptotic Equipartition Property (AEP))**.** *Let $X_i \overset{\text{iid}}{\sim} p(x)$. Then $\Pr(T_\epsilon^{(n)}) \to 1$ as $n \to \infty$.*

Theorem 2.4.2 (Probability of jointly typicality)**.** *[11, Lemma 10.6.2, page 327] Let $Y_1, Y_2, \cdots, Y_n \overset{\text{iid}}{\sim} p(y)$. For any $x^n \in T_\epsilon^{(n)}(X)$, the probability that $(x^n, Y^n) \in T_\epsilon^{(n)}(XY)$ is bounded by*

$$\Pr\left((x^n, Y^n) \in T_\epsilon^{(n)}\right) \doteq 2^{-n(I(X;Y) \pm \delta(\epsilon))}, \tag{2.4.4}$$

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

Theorem 2.4.3. *[11, Theorem 15.2.3, page 524] Let $T_\epsilon^{(n)}$ denote the typical set for the probability mass function $p(s_1, s_2, s_3)$, and let $\Pr(\mathbf{S}'_1 = \mathbf{s}_1, \mathbf{S}'_2 = \mathbf{s}_2, \mathbf{S}'_3 = \mathbf{s}_3) = \prod_{i=1}^n p(s_{1i}|s_{3i})p(s_{2i}|s_{3i})p(s_{3i})$, then $\Pr\{(\mathbf{S}'_1, \mathbf{S}'_2, \mathbf{S}'_3) \in T_\epsilon^{(n)}\} \doteq 2^{-n(I(S'_1;S'_2|S'_3) \pm \delta(\epsilon))}$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ and $n \to \infty$.*

2.5 Differential Entropy

In this section, we extend the definition of entropy to the continuous RV.

Definition 2.5.1. The *differential entropy* $h(X)$ of a continuous random variable X with probability density function $f(x)$ is defined as

$$h(X) = \int_{S_X} f(x) \log \frac{1}{f(x)}\, dx \tag{2.5.1}$$

where $S_X = \{x : f(x) > 0\}$ is the support set of X. We show the support set as S when the RV is clear from the context.

Since differential entropy depends only on probability distribution, it is sometimes written as $h(f)$ instead of $h(X)$.

Theorem 2.5.1. *Let* $X_1, X_2, \cdots, X_n \overset{\text{iid}}{\sim} f(x)$. *Then,*

$$-\frac{1}{n}\log f(X_1, X_2, \cdots, X_n) \to E[-\log f(x)] = h(X) \quad \text{in probability.} \tag{2.5.2}$$

Definition 2.5.2. The *volume* $\text{Vol}\,(A)$ of a set $A \subset \mathbb{R}^n$ is defined as

$$\begin{aligned} \text{Vol}\,(A) &= \int_A dx_1\ dx_2\ \cdots\ dx_n & (2.5.3)\\ &= \int_A dx^n. & (2.5.4) \end{aligned}$$

Definition 2.5.3. The typical set $T_\epsilon^{(n)}$ with respect to $f(x)$ for $\epsilon > 0$ and any n is defined as

$$T_\epsilon^{(n)} = \left\{ x^n \in S^n : \left| -\frac{1}{n}\log f(x^n) - h(X) \right| \le \epsilon \right\}, \tag{2.5.5}$$

where $f(x^n) = \prod_{i=1}^n f(x_i)$.

Theorem 2.5.2. *[11, Theorem 8.2.2, page 245] The typical set* $T_\epsilon^{(n)}$ *satisfies the following properties*

1. $\Pr\left(T_\epsilon^{(n)}\right) > 1 - \epsilon$ *for* n *sufficiently large.*
2. $\text{Vol}\left(T_\epsilon^{(n)}\right) \le 2^{n(h(X)+\epsilon)}$ *for all* n.
3. $\text{Vol}\left(T_\epsilon^{(n)}\right) \ge (1-\epsilon)2^{n(h(X)-\epsilon)}$ *for* n *sufficiently large.*

This theorem states that for large n, the volume that contains *almost all* of the sequences approaches to 2^{nh} in the first order of exponent. On the other hand, this volume is an n-dimensional volume; and therefore, the corresponding side length $(2^{nh})^{\frac{1}{n}} = 2^h$. In other words, the differential entropy is the logarithm of the side length of the smallest volume that contains almost all of the sequences[4].

As in discrete case, we extend the definition to multiple variables.

[4]While entropy is related to the volume of typical set, Fisher Information is related to the surface of the typical set.

Definition 2.5.4. The *differential entropy* of a set of random variables X^n with probability density function $f(x^n)$ is defined as

$$h(X^n) = -\int_{S^n} f(x^n) \log f(x^n)\, dx^n. \tag{2.5.6}$$

Definition 2.5.5. The conditional entropy of X, Y with joint probability function $f(x, y)$ is defined as

$$h(Y|X) = -\int_{S_X \cup S_Y} f(x, y) \log f(y|x)\, dx\, dy. \tag{2.5.7}$$

Theorem 2.5.3 (Entropy of a multi variable normal distribution)**.** *[11, Theorem 8.4.1, page 249] Let $X_1, X_2, \cdots, X_n$ have a multi variable normal distribution with respective means $\mu_1, \cdots, \mu_n$ and covariance matrix K, i.e., $X^n \sim \mathcal{N}(\underline{\mu}, K)$. Then*

$$\begin{aligned} h(X^n) &= h(\mathcal{N}(\underline{\mu}, K)) && (2.5.8)\\ &= \frac{1}{2}\log|2\pi eK| \qquad \text{bits}, && (2.5.9) \end{aligned}$$

where $|2\pi eK|$ denotes the determinant of $2\pi eK$, and $\underline{\mu}$ is a column vector $\begin{pmatrix}\mu_1 & \mu_2 & \dots & \mu_n\end{pmatrix}^T$.

Proof: The proof of this theorem is given in the mentioned reference based on matrix expansion. We will give an alternative proof based on the properties of matrix operators which is more comprehensive.

Let the vector $\mathbf{x}$ be a column vector $\begin{pmatrix}x_1 & x_2 & \dots & x_n\end{pmatrix}^T$. The probability density function of $f(x^n) : \mathcal{X}^n \mapsto \mathbb{R}$ can be written as

$$f(x^n) = \frac{1}{|2\pi K|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\underline{\mu})^T K^{-1}(\mathbf{x}-\underline{\mu}).} \tag{2.5.10}$$

Then

$$\begin{aligned} h(X^n) &= -E[\ln f(x^n)] \qquad \text{nats} && (2.5.11)\\ &= \frac{1}{2}\ln|2\pi K| + \frac{1}{2}E\left[(\mathbf{x}-\underline{\mu})^T K^{-1}(\mathbf{x}-\underline{\mu})\right] && (2.5.12)\\ &\overset{(a)}{=} \frac{1}{2}\ln|2\pi K| + \frac{1}{2}E\left[\mathrm{tr}\left((\mathbf{x}-\underline{\mu})^T K^{-1}(\mathbf{x}-\underline{\mu})\right)\right] && (2.5.13) \end{aligned}$$

$$\stackrel{(b)}{=} \frac{1}{2}\ln|2\pi K| + \frac{1}{2}E\left[\mathrm{tr}\left((\mathbf{x}-\underline{\mu})(\mathbf{x}-\underline{\mu})^T K^{-1}\right)\right] \tag{2.5.14}$$

$$= \frac{1}{2}\ln|2\pi K| + \frac{1}{2}\mathrm{tr}\left(E\left[(\mathbf{x}-\underline{\mu})(\mathbf{x}-\underline{\mu})^T K^{-1}\right]\right) \tag{2.5.15}$$

$$= \frac{1}{2}\ln|2\pi K| + \frac{1}{2}\mathrm{tr}\left(K\,K^{-1}\right) \tag{2.5.16}$$

$$= \frac{1}{2}\ln|2\pi K| + \frac{1}{2}n \tag{2.5.17}$$

$$= \frac{1}{2}\ln|2\pi eK| \tag{2.5.18}$$

$$= \frac{1}{2}\log|2\pi eK| \qquad \text{bits,} \tag{2.5.19}$$

where

(a) follows form the fact that $(\mathbf{x}-\underline{\mu})^T K^{-1}(\mathbf{x}-\underline{\mu})$ is an scalar and trace of any scalar is equal to that scalar.

(b) can be justified by considering the fact that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for any pair of interchangeable matrices $A_{m\times n}$ and $B_{n\times m}$. ■

Definition 2.5.6. The *relative entropy* between two distributions f and g on x^n is defined as

$$D(f||g) = \int f \log \frac{f}{g}\, dx^n. \tag{2.5.20}$$

Definition 2.5.7. The *mutual information* $I(X;Y)$ between two RV's X and Y with joint density $f(x,y)$ is defined as

$$I(X;Y) = D(f(x,y)||f(x)f(y)) \tag{2.5.21}$$

$$= \int f(x,y)\log\frac{f(x,y)}{f(x)f(y)}\, dx\, dy. \tag{2.5.22}$$

Theorem 2.5.4 (Normal distributions are entropy maximizers)**.** *[11, Theorem 8.6.5, page 254] Let the random vector $X^n \in \mathbb{R}$ have zero mean and covariance $K = E[\mathbf{X}\mathbf{X}^T]$. Then $h(X^n) \le \frac{1}{2}\log|2\pi eK|$, with equality if and only if $\mathbf{X} \sim \mathcal{N}(\underline{0}, K)$.*

2.6 Channel Coding Theorem

In 1948, Shannon published his original paper [38] in which he founded what today is known as *information theory*. In his work, Shannon obtained the capacity of the point-to-point Discrete Memoryless Channel (DMC) using random coding method at the encoder and jointly typical decoder at the receiver. Prior to Shannon's, it was wrongly conceived that the only way to increase the error free data transmission rate is to increase the transmission power. Shannon showed that by coding the message, the zero transmission error can be obtained with any transmission power if the rate by which the data is being transmitted is less than the channel capacity. His pioneering work opened a new chapter in communication theory, and since then, considerable efforts have been made in this area. In this chapter, we present some of the works and well-known results, which will be used in the rest of this work as basic building blocks to establish our results.

A basic single user communication system model is illustrated in Fig. 2.2. As can be seen, this model comprises a message w, an encoder which encodes the message set onto codeword x^n, a DMC, and a decoder that maps the channel outputs onto a message estimate. The channel is shown by $(\mathcal{X}, p(y|x), \mathcal{Y})$ where $\mathcal{X}$ is the set of channel input alphabets, $\mathcal{Y}$ is the set of channel output alphabets, and $p(y|x)$ is the channel transition probability function.

The channel input X^n is subjected to a constraint

$$\frac{1}{n}\sum_{i=1}^{n} E\left\{\phi\left(X_i\right)\right\} \leq \Gamma, \tag{2.6.1}$$

where $\phi : \mathcal{X} \mapsto \{0\} \cup \mathbb{R}^+$ is the *transmission cost function* [32], $\mathbb{R}^+$ is the set of all positive real numbers, and $\Gamma > 0$ is a constant.

Figure 2.2: Basic single user communication channel. The channel is represented by a conditional probability mass function $p(y^n|x^n)$. The encoder maps each message into a codeword x^n. Inversely, the decoder maps the channel output y^n into a message estimate.

Being memoryless implies that $p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$. We further define $(2^{nR}, n)$ code for this channel which consists of the following:

1. A set of messages $\mathcal{W} = \left\{1, 2, \cdots, 2^{nR}\right\}$. Throughout this work, we assume that the message w is uniformly distributed on $\mathcal{W}$.

2. An encoding function that assigns a *codeword* $x^n(w)$ to each message w. The set of all codewords $\left\{x^n(1), x^n(2), \cdots, x^n(2^{nR})\right\}$ is called *codebook* $\mathcal{C}$.

3. A decoding function $g(\cdot)$ that maps the channel output y^n onto the message set $\mathcal{W}$, i.e., $\hat{w} = g(y^n), \hat{w} \in \mathcal{W}$.

The *rate* R of the code is defines as logarithm of the message size divided by the number of channel use. in other words,

$$R = \frac{\log |\mathcal{W}|}{n} \qquad \text{bits per channel use.} \tag{2.6.2}$$

As can bee seen, we do incorporate this definition into the size of the message set and represent[5] the number of messages by 2^{nR}.

[5]Later, we will drive bounds on the rate of different types of channels and these bounds may not result in an integer message set size, i.e., 2^{nR} is not an integer. To address this problem, the number of messages is represented by the *floor* of this quantity, i.e., $\lfloor 2^{nR} \rfloor$ or $2^{n\lfloor R \rfloor}$. For the sake of convenience, we will not use the floor operator $\lfloor \cdot \rfloor$ on the message set size, but as a subtle and obvious assumption, the number of messages are always an integer.

For above mentioned channel, let $\lambda_w = \Pr\{\hat{w} \neq w | w \text{ sent}\}$ be the conditional probability of error given that the transmitted message is w. Then, the average probability of error $P_e^{(n)}$for a code $(2^{nR}, n)$ can be defined as

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w. \tag{2.6.3}$$

A rate R is said to be achievable if there is a code such that $P_e^{(n)} \to 0$ as $n \to \infty$. The supreme of all achievable rates is called the channel capacity.

Theorem 2.6.1 (Shannon [38])**.** *The capacity of a discrete memoryless channel is given by*

$$C = \max_{X \sim p(x)} I(X;Y). \tag{2.6.4}$$

Proof outline: The proof of this theorem is fundamental and seems to be a panacea for various channel types since it involves typical steps being used in other channels. The proof consists of the following steps. The first step is to prove *achievablity* of the capacity. In other words, it should be shown that any rate $R < C$ is achievable, meaning that there exists a sequence of codes $(2^{nR}, n)$ with probability of error $P_e^{(n)} \to 0$. The second step is to prove the *converse*, meaning to justify that for any sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$, the rate R is less than the capacity C. [6]

Proof of achievablity

Codebook generation: A random coding argument is used. Generate 2^{nR} i.i.d. codewords x^n according to $p(x^n) = \prod_{i=1}^{n} p(x_i)$ and label them as $x^n(w)$, where $w \in \mathcal{W} = \left\{1, 2, \cdots, 2^{nR}\right\}$. The generated codebook $\mathcal{C}$ is revealed to both the sender and the receiver before any transmission is being taken place.

[6] Achievablity proof results in an *achievable rate* and the converse part will result in an *upper bound*. The capacity region lies between these two bounds. For any channel, the capacity region is known if these two bounds completely coincide. Otherwise, only the inner and outer bounds to capacity will be known.

Encoding: Suppose that w is the message which is about to be sent. Then the corresponding codeword $x^n(w)$ will be transmitted as the channel input.

Decoding: Let the channel output Y^n be the received sequence. The decoder declares that message $\hat{w}$ was sent if there exists one and only one index $\hat{w} \in \mathcal{W}$ such that $(x^n(\hat{w}), Y^n) \in T_\epsilon^{(n)}$; otherwise, an error is declared.

Probability of error: Assuming that w was sent, the error occurs if $(x^n(w), Y^n) \notin T_\epsilon^{(n)}$, or there is an index $i \neq w$ such that $(x^n(i), Y^n) \in T_\epsilon^{(n)}$. Let the error event $\mathcal{E}$ be the average of $P_e^{(n)}$over all codebooks and all messages w. Then

$$\begin{aligned}
\Pr\{\mathcal{E}\} &= \sum_{\mathcal{C}} p(\mathcal{C}) P_e^{(n)}(\mathcal{C}) && (2.6.5)\\
&= \sum_{\mathcal{C}} p(\mathcal{C}) \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(\mathcal{C}) && (2.6.6)\\
&= 2^{-nR} \sum_{w=1}^{2^{nR}} \sum_{\mathcal{C}} p(\mathcal{C}) \lambda_w(\mathcal{C}) && (2.6.7)\\
&= \sum_{\mathcal{C}} p(\mathcal{C}) \lambda_1(\mathcal{C}) && (2.6.8)\\
&= \Pr(\mathcal{E}|w=1). && (2.6.9)
\end{aligned}$$

Furthermore, we define the event $E(i)$ as

$$E(i) = \left\{(x^n(i), Y^n) \in T_\epsilon^{(n)}\right\}, \qquad i \in \mathcal{W}. \tag{2.6.10}$$

Hence

$$\begin{aligned}
\Pr(\mathcal{E}|w=1) &= \Pr\left\{E^c(1) \bigcup \cup_{k=2}^{2^{nR}} E(k)\right\} && (2.6.11)\\
&\leq P(E^c(1)) + \sum_{k=2}^{2^{nR}} P(E(k)). && (2.6.12)
\end{aligned}$$

According to the Lemma 2.4.1, $P(E^c(1)) \rightarrow 0$ for n sufficiently large. In addition, since $x^n(1)$ and $x^n(k)$, $k \neq 1$ are independent, the channel output Y^n (which is as a result of the channel input $x^n(1)$) and all other codewords $x^n(k)$, $k \neq 1$ are independent as well. Therefore, the probability of $E(k)$, $k \neq 1$ is less than $2^{-n(I(X;Y)-\epsilon)}$ (according to Theorem 2.4.2), and

$$\begin{aligned} \Pr(\mathcal{E}) &\leq \epsilon + \sum_{k=2}^{2^{nR}} 2^{-n(I(X;Y)-\epsilon)} && (2.6.13) \\ &= \epsilon + (2^{nR}-1)2^{-n(I(X;Y)-\epsilon)} && (2.6.14) \\ &\leq \epsilon + 2^{-n(I(X;Y)-R-\epsilon)} && (2.6.15) \\ &\leq \epsilon && (2.6.16) \end{aligned}$$

for n sufficiently large and $R < I(X;Y) - \epsilon$.

Proof of converse

In the converse part it must be shown that any sequence of $(2^{nR}, n)$ code with $P_e^{(n)} \rightarrow 0$ results in $R \leq C$.

The joint p.d.f. of the tuple (W, X^n, Y^n) can be written as

$$(W, X^n, Y^n) \sim p(w, x^n, y^n) = p(w)p(x^n|w)\prod_{i=1}^{n} p(y_i|x_i) \qquad (2.6.17)$$

where $p(w)$ is assumed to be uniform over the message set, i.e., $p(w) = 2^{-nR}$. By Fano's inequality,

$$H(W|\hat{W}) \leq 1 + nRP_e^{(n)} \triangleq n\epsilon_n \qquad (2.6.18)$$

where $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$. Furthermore, since $\hat{W}$ is a function of Y^n, data-processing inequality implies that $H(W|Y^n) \leq H(W|\hat{W})$. We can now write

$$nR \overset{(a)}{=} H(W) \qquad (2.6.19)$$

$$
\begin{aligned}
&= I(W;Y^n) + H(W|Y^n) && (2.6.20)\\
&\overset{(b)}{\leq} I(X^n;Y^n) + n\epsilon_n && (2.6.21)\\
&= H(Y^n) - H(Y^n|X^n) + n\epsilon_n && (2.6.22)\\
&\overset{(c)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) + n\epsilon_n && (2.6.23)\\
&\overset{(d)}{\leq} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) + n\epsilon_n && (2.6.24)\\
&= \sum_{i=1}^{n} I(X_i;Y_i) + n\epsilon_n && (2.6.25)\\
&\leq nC + n\epsilon_n && (2.6.26)
\end{aligned}
$$

where

(a) holds because the messages are uniformly distributed over the message set $\mathcal{W}$,

(b) follows from Fano's inequality,

(c) can be justified considering the fact that the channel is memoryless, and each channel putput y_i is independent form other channel inputs $x_k,\ k \neq i$ given x_i,

(d) follows form the fact that entropy of multiple random variables is less than sum of entropies of those random variables.

And furthermore, by dividing to n we have

$$R \leq C + \epsilon_n \tag{2.6.27}$$

where $\epsilon_n \to 0$ as $n \to \infty$; and therefore, $R \leq C$. ■

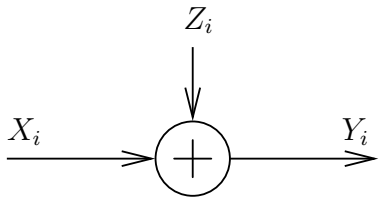

Figure 2.3: Block diagram of a point to point continuous channel with additive white Gaussian noise. The channel has a conditional probability mass function $f(y|x)$.

Continuous alphabet

When the noise is additive Gaussian noise[7] (which is the case in many practical communication channels), the alphabet sets of the channel input and channel output are assumed to be continuous.

Figure 2.3 depicts a point-to-point communication channel with additive white Gaussian noise. As shown, the channel output $Y = X + Z$ where Z is an additive zero mean Gaussian noise with variance N, i.e., $E[Z^2] = N$. The channel input X is a zero mean random variable with power $E[X^2]$ limited to P. The noise and the transmitted codeword are assumed to be independent. The capacity of this channel can be written as

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) && (2.6.28)\\ &= \max_{p(x)} [H(Y) - H(Y|X)] && (2.6.29)\\ &= \max_{p(x)} [H(X+Z) - H(X+Z|X)] && (2.6.30)\\ &= \max_{p(x)} H(X+Z) - H(Z). && (2.6.31) \end{aligned}$$

[7]Noise in wireless communication channels mainly arises at the receiver because of the thermal noise at the amplifiers. This thermal noise can be perfectly modeled by a Gaussian function. There are, however, other sources of noise which are modeled differently. For instance, the noise in optical communication channels can be modeled by a poisson arrival process.

We must find a proper distribution for X that maximizes $H(X+Z)$. Considering Theorem 2.5.4, the distribution of $X+Z$ must be normal so must be the distribution of X because Z is already normal and summation of two normal distribution is normal as well. Hence,

$$\begin{aligned} C &= H(X+Z) - H(Z), \qquad \text{where } X \sim \mathcal{N}(\underline{0}, P) && (2.6.32)\\ &= \frac{1}{2}\log\left|2\pi e E[(X+Z)^2]\right| - \frac{1}{2}\log\left|2\pi e E[Z^2]\right| && (2.6.33)\\ &= \frac{1}{2}\log\left(1+\frac{P}{N}\right). && (2.6.34) \end{aligned}$$

2.7 Multiple Access Channel

Figure 2.4 demonstrates a Discrete Memoryless MAC (DM-MAC). As can be seen, a DM-MAC consists of two[8] encoders, two message sets, a memoryless channel with probability transition matrix $p(y|x_1,\, x_2)$, and one decoder. Encoder t assigns a codeword X_t^n to each message $w_t \in \mathcal{W}_t$, where $t = 1, 2$ is the transmitter index. Decoder maps the channel output Y^n onto message estimations $(\hat{w}_1, \hat{w}_2)$. A $(2^{nR_1}, 2^{nR_2}, n)$ code for the MAC consists of

1. Two message sets $\mathcal{W}_1 = \left\{1, 2, \cdots, 2^{nR_1}\right\}$ and $\mathcal{W}_2 = \left\{1, 2, \cdots, 2^{nR_2}\right\}$. Like the point-to-point case, we assume that the messages (w_1, w_2) are uniformly distributed on $\mathcal{W}_1 \times \mathcal{W}_2$.

2. Two encoding functions that assign codewords $x_1^n(w_1)$ and $x_2^n(w_2)$ to each message pair (w_1, w_2).

[8] In a more general case, a multiple access channel can include multiple senders and single receiver. This case, however, can be easily studied by generalizing the multiple access channel with two senders to that with multiple senders.

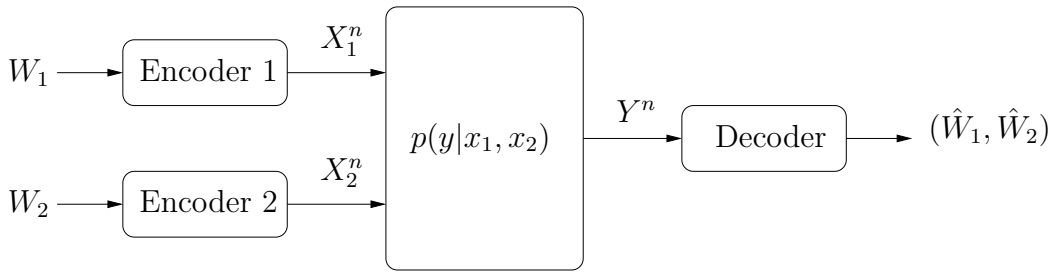

Figure 2.4: A discrete memoryless MAC. Each encoder maps the message onto a codeword and the decoder maps the channel putpuy Y^n onto a pair of message estimate $(\hat{W}_1, \hat{W}_2)$. The channel has a conditional probability mass function $p(y|x_1, x_2)$.

3. A decoding function $g(\cdot)$ that maps each channel output Y^n onto the message set $\mathcal{W}_1 \times \mathcal{W}_2$, i.e., $(\hat{w_1}, \hat{w_2}) = g(y^n), (\hat{w_1}, \hat{w_2}) \in \mathcal{W}_1 \times \mathcal{W}_2$.

For the above mentioned channel, let $\Pr\{(\hat{w_1}, \hat{w_2}) \neq (w_1, w_2) | (w_1, w_2) \text{ sent}\}$ be the conditional probability of error given that the transmitted message is (w_1, w_2). Then, the average probability of error $P_e^{(n)}$for a code $(2^{nR_1}, 2^{nR_2}, n)$ can be defined as

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{(w_1,w_2)\in \mathcal{W}_1 \times \mathcal{W}_2} \Pr\{(\hat{w_1}, \hat{w_2}) \neq (w_1, w_2) | (w_1, w_2) \text{ sent}\}. \quad (2.7.1)$$

The capacity region of MAC was derived in [2].

2.8 Summary

In this chapter, necessary mathematical concepts and theorems were introduced. Firstly, the concept of entropy and information measure have been defined. As discussed, the entropy is a necessary concept to define the distance between probability functions and to define mutual information between two (or more) RVs. Then, the concept of typical sequences and jointly typicality were introduced, and they were

generalized to the continuous RV. Next, the fundamental capacity theorem of a single user DMC was stated and the proof was given. Lastly, the MAC was mentioned as an example of multi-user communication channels.

Chapter 3

Cognitive Radio

As mentioned before, cognitive radio is one of the most promising cooperative models. It was mentioned in Chapter 1.1 that there has been a large number of works studying this model. In this chapter, we introduce the Casual Cognitive Radio mathematical model, and then, study its achievable performance. A practical Gaussian case is illustrated and compared with existing results.

3.1 Mathematical Channel Model

Consider the discrete memoryless IC-CUC as illustrated in Fig. 3.1. The discrete memoryless IC-CUC is denoted by $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2, y|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y})$, where $\mathcal{X}_1$ and $\mathcal{X}_2$ are the finite input alphabets of the primary and secondary users respectively, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are the finite output alphabets of receivers 1 and 2 respectively, and $p(\cdot, \cdot, \cdot|x_1, x_2)$ is a collection of probability distributions on $\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}$ given $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. Following the standard notation adopted in [11], we define a $(2^{nR_1}, 2^{nR_2}, n)$ code for the IC-CUC in the following.

Definition 3.1.1. A $\left(2^{nR_1}, 2^{nR_2}, n\right)$ code for the IC-CUC consists of two message sets $\mathcal{W}_1 = \{1, 2, \ldots, 2^{nR_1}\}$ for the primary user and $\mathcal{W}_2 = \{1, 2, \ldots, 2^{nR_2}\}$ for the

secondary user, an encoding function

$$X_1 : \{1, 2, \ldots, 2^{nR_1}\} \to \mathcal{X}_1^n,$$

a set of functions $\{f_i\}_{i=1}^n$ which will be termed as broadcasting relay functions such that

$$x_{2i} = f_i\left(w_2, Y^{i-1}\right),$$

with $w_2 \in \mathcal{W}_2$, and $Y^{i-1} = (Y_1, Y_2, \ldots, Y_{i-1})$, and two decoding functions

$$g_1 : \mathcal{Y}_1^n \to \mathcal{W}_1, \quad g_2 : \mathcal{Y}_2^n \to \mathcal{W}_2.$$

It should be noted that the broadcasting relay (the secondary user) is non-anticipating. This means that the current output of the secondary user (x_2) depends only on the past received samples as well as the message w_2. The channel is assumed to be memoryless; therefore, for any choice of $p(w_1)$, $p(w_2)$, encoding functions, and broadcasting relay functions, the probability mass function over $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n \times \mathcal{Y}^n$ is given by

$$\begin{aligned} p(w_1, w_2, x_1^n, x_2^n, y_1^n, y_2^n, y^n) &= p(w_1)p(w_2) \\ &\times \prod_{i=1}^{n} p(x_{i1}|w_1)p(x_{2i}|w_2, y^{i-1})p(y_{1i}, y_{2i}, y_i|x_{1i}, x_{2i}). \end{aligned}$$

Note by E_{err}, the event that an error happens in decoder 1 or 2. The average probability of error (happening the event E_{err}) of the code is defined by

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{w_1,w_2} \Pr\left\{ \begin{array}{c|c} g_1(Y_1^n) \neq w_1 & \\ \text{or} & w_1, w_2 \\ g_2(Y_2^n) \neq w_2 & \text{sent} \end{array} \right\}.$$

The probability of error is calculated under the uniform distribution over the codewords $w_1 \in \mathcal{W}_1$ and $w_2 \in \mathcal{W}_2$.

Definition 3.1.2. A rate pair (R_1, R_2) is called achievable for the IC-CUC if there is a sequence of $\left(2^{nR_1}, 2^{nR_2}, n\right)$ codes with $P_e^{(n)} \to 0$. The capacity region of the IC-CUC is the union of set of all achievable rates.

In this dissertation, the primary and secondary users are also referred to as *sender 1* and *sender 2* respectively as depicted in Fig. 3.1.

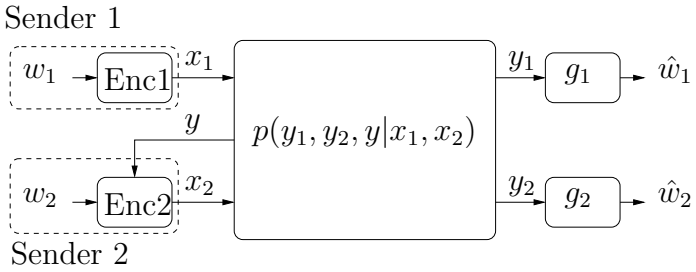

Figure 3.1: Channel model for an IC-CUC, a causal configuration for the CR. x_1 and x_2 are the channel inputs, and y_1, y_2, y are the channel outputs.

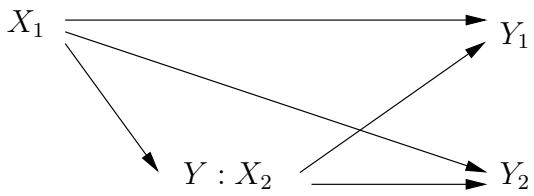

Figure 3.2: A mnemonic channel diagram for the IC-CUC. Solid lines represent the communication channel between a sender and the receiver.

3.2 An Achievable Rate Region

In this section, we establish an achievable rate region for the IC-CUC by using a combination of different coding schemes. The coding scheme needs to take into account the dual roles of the sender 2, which acts as a relay to cooperate with the primary user as well as a sender to transmit its own message to receiver 2. Thus, the sender 2 can be thought as a relay which broadcasts two sets of messages.

In order to demonstrate the achievable region for an IC-CUC, auxiliary random variables Q, U_{11}, U_{12}, U_{21}, U_{22}, X_{11}, X_{12}, X_{21}, and X_{22} are defined over the finite sets $\mathcal{Q}$, $\mathcal{U}_{11}$, $\mathcal{U}_{12}$, $\mathcal{U}_{21}$, $\mathcal{U}_{22}$, $\mathcal{X}_{11}$, $\mathcal{X}_{12}$, $\mathcal{X}_{21}$, and $\mathcal{X}_{22}$ in a random coding argument, where $\mathcal{Q}$ plays the role of a time-sharing random variable [11]. Denote by $\mathcal{P}$ the set of all joint

probability distributions $p(\cdot)$ on Z that can be decomposed as

$$\begin{aligned} p(z) &= p(u_{11}|q)p(u_{12}|q)p(u_{21}|q)p(u_{22}|q) \\ &\quad \times p(x_{11}|u_{11},q)p(x_{12}|u_{12},q)p(x_{21}|u_{21},q) \\ &\quad \times p(x_{22}|u_{22},q)p(x_1|x_{11},x_{12},q) \\ &\quad \times p(x_2|x_{21},x_{22},u_{11},u_{12},q) \\ &\quad \times p(y_1,y_2,y|x_1,x_2), \end{aligned} \tag{3.2.1}$$

where $Z = (Q,\, U_{11},\, U_{12},\, U_{21},\, U_{22},\, X_{11},\, X_{12},\, X_{21},\, X_{22},\, X_1,\, X_2,\, Y_1,\, Y_2,\, Y)$.

Theorem 3.2.1. *For an IC-CUC any non-negative rate pair* $(R_1,\, R_2)$*, where* $R_1 = R_{11} + R_{12}$*,* $R_2 = R_{21} + R_{22}$*, satisfying*

$$R_{11} \le I(X_{11};Y|U_{11}U_{12}X_{12}Q), \tag{3.2.2}$$

$$R_{12} \le I(X_{12};Y|U_{11}U_{12}X_{11}Q), \tag{3.2.3}$$

$$R_{11}+R_{12} \le I(X_{11}X_{12};Y|U_{11}U_{12}Q), \tag{3.2.4}$$

$$\begin{aligned} R_{11} &\le I(X_{11};Y_1|X_{12}X_{21}U_{11}U_{12}U_{21}Q) \\ &\quad + I(U_{11};Y_1|U_{12}U_{21}Q), \end{aligned} \tag{3.2.5}$$

$$\begin{aligned} R_{12} &\le I(X_{12};Y_1|X_{11}X_{21}U_{11}U_{12}U_{21}Q) \\ &\quad + I(U_{12};Y_1|U_{11}U_{21}Q), \end{aligned} \tag{3.2.6}$$

$$\begin{aligned} R_{11}+R_{12} &\le I(X_{11}X_{12};Y_1|X_{21}U_{11}U_{12}U_{21}Q) \\ &\quad + I(U_{11}U_{12};Y_1|U_{21}Q), \end{aligned} \tag{3.2.7}$$

$$\begin{aligned} R_{11}+R_{21} &\le I(X_{11}X_{21};Y_1|X_{12}U_{11}U_{12}U_{21}Q) \\ &\quad + I(U_{11}U_{21};Y_1|U_{12}Q), \end{aligned} \tag{3.2.8}$$

$$\begin{aligned} R_{12}+R_{21} &\le I(X_{12}X_{21};Y_1|X_{11}U_{11}U_{12}U_{21}Q) \\ &\quad + I(U_{12}U_{21};Y_1|U_{11}Q), \end{aligned} \tag{3.2.9}$$

$$\begin{aligned} R_{11}+R_{12}+R_{21} &\le I(X_{11}X_{12}X_{21};Y_1|U_{11}U_{12}U_{21}Q) \\ &\quad + I(U_{11}U_{12}U_{21};Y_1|Q), \end{aligned} \tag{3.2.10}$$

$$\begin{aligned} R_{21} &\le I(X_{21};Y_2|X_{12}X_{22}U_{12}U_{21}U_{22}Q) \\ &\quad + I(U_{21};Y_2|U_{12}U_{22}Q), \end{aligned} \tag{3.2.11}$$

$$\begin{aligned} R_{22} &\le I(X_{22};Y_2|X_{12}X_{21}U_{12}U_{21}U_{22}Q) \\ &\quad + I(U_{22};Y_2|U_{12}U_{21}Q), \end{aligned} \tag{3.2.12}$$

$$
\begin{aligned}
R_{12}+R_{21} &\leq I(X_{12}X_{21};Y_2|X_{22}U_{12}U_{21}U_{22}Q) \\
&\quad +I(U_{12}U_{21};Y_2|U_{22}Q), && (3.2.13)\\
R_{12}+R_{22} &\leq I(X_{12}X_{22};Y_2|X_{21}U_{12}U_{21}U_{22}Q) \\
&\quad +I(U_{12}U_{22};Y_2|U_{21}Q), && (3.2.14)\\
R_{21}+R_{22} &\leq I(X_{21}X_{22};Y_2|X_{12}U_{12}U_{21}U_{22}Q) \\
&\quad +I(U_{21}U_{22};Y_2|U_{12}Q), && (3.2.15)\\
R_{12}+R_{21}+R_{22} &\leq I(X_{12}X_{21}X_{22};Y_2|U_{12}U_{21}U_{22}Q) \\
&\quad +I(U_{12}U_{21}U_{22};Y_2|Q), && (3.2.16)
\end{aligned}
$$

is achievable for some $Z \in \mathcal{P}$.

Proof Outline: The key idea in Theorem 3.2.1 is to use regular encoding at the senders and sliding-window decoding [47] at the receivers. The rate of each sender is split into two parts R_{t1} and R_{t2} such that $R_t = R_{t1} + R_{t2}$, $t = 1, 2$. Therefore, sender t has message indices $w_{t1} \in \{1, 2, \ldots, 2^{nR_{t1}}\}$, and $w_{t2} \in \{1, 2, \ldots, 2^{nR_{t2}}\}$, $t = 1, 2$ to send. By splitting the rate, receiver 1 (or 2) can also decode a part of the message of sender 2 (or 1) instead of treating it entirely as noise. During the encoding phase, sender 1 sends $x_1^n(w_{11,b}, w_{12,b}|w_{11,b-1}, w_{12,b-1})$ in block b. Before sending the block b, sender 2 decodes messages of sender 1 in the previous block, i.e., $w_{11,b-1}, w_{12,b-1}$. Then, sender 2 superimposes its messages, i.e., $w_{21,b}, w_{22,b}$, onto $w_{11,b-1}, w_{12,b-1}$ and its own messages in the previous block and sends $x_2^n(w_{21,b}, w_{22,b}|w_{11,b-1}, w_{12,b-1}, w_{21,b-1}, w_{22,b-1})$ in the block b. After receiving block b, receiver 1 decodes the message tipple $(\hat{w}_{11,b-1}, \hat{w}_{12,b-1}, \hat{w}_{21,b-1})$, and receiver 2 decodes the message tipple $(\hat{\hat{w}}_{12,b-1}, \hat{\hat{w}}_{21,b-1}, \hat{\hat{w}}_{22,b-1})$.

Note that in [41, Theorem 3], the messages at sender 2 is not split, and instead is treated entirely as noise at its non-pairing receiver. This encoding approach may potentially reduce the achievable rate region.

Proof of Theorem 3.2.1

As mentioned before, in order to represent the achievable region for the IC-CUC in Theorem 3.2.1, auxiliary random variables Q, U_{tr}, and X_{tr} are defined over the finite sets $\mathcal{Q}$, $\mathcal{U}_{tr}$, and $\mathcal{X}_{tr}$ in a random coding argument where $t, r = 1, 2$. Therefore, the family $p(q, u_{11}, u_{12}, u_{21}, u_{22}, x_{11}, x_{12}, x_{21}, x_{22}, x_1, x_2, y_1, y_2, y)$ can be written as (3.2.1). A regular block Markov superposition coding argument similar to [47] is used. For notational simplicity, the time-sharing random variable Q is omitted throughout the proof, but it can be easily substituted back using a standard time-sharing argument (see [11], [10] for details).

Codebook generation:

- Generate $2^{nR_{tr}}$ i.i.d. codewords $u_{tr}^n(w'_{tr})$, $w'_{tr} \in \mathcal{W}_{tr} = \{1, \ldots, 2^{nR_{tr}}\}$, according to $\prod_{i=1}^n p(u_{tri})$ where $t, r = 1, 2$.

- For each codeword $u_{tr}^n(w'_{tr})$, generate $2^{nR_{tr}}$ i.i.d. codewords $x_{tr}^n(w_{tr}, w'_{tr})$, $w_{tr} \in \mathcal{W}_{tr}$, according to
  $\prod_{i=1}^n p(x_{tri}|u_{tri}(w'_{tr}))$ where $t, r = 1, 2$.

- For each message tuple $(w_{11}, w_{12}, w'_{11}, w'_{12})$, generate an i.i.d. codeword $x_1^n(w_{11}, w_{12}, w'_{11}, w'_{12})$ according to $\prod_{i=1}^n p(x_{1i}|x_{11i}(w_{11}, w'_{11}), x_{12i}(w_{12}, w'_{12}))$.

- For each message tuple $(w'_{11}, w'_{12}, w'_{21}, w'_{22}, w_{21}, w_{22})$, generate an i.i.d. codeword $x_2^n(w_{21}, w_{22}, w'_{11}, w'_{12}, w'_{21}, w'_{22})$ according to
  $\prod_{i=1}^n p(x_{2i}|x_{21i}(w_{21}, w'_{21}), x_{22i}(w_{22}, w'_{22}), u_{11i}(w'_{11}), u_{12i}(w'_{12}))$.

Therefore, the codebook $\mathcal{C}_0$ is generated and revealed to all senders and receivers. Repeating above process, another random codebook $\mathcal{C}_1$ similar to $\mathcal{C}_0$ is generated.

These codebooks are used alternatively as follows: In block b the coodebook $\mathcal{C}_{b \mod 2}$ is used. Hence, codewords in two consecutive blocks are independent.

Encoding: In block b, sender 1 sends $x_1^n(w_{11,b}, w_{12,b}, w_{11,b-1}, w_{12,b-1})$ in order to transmit the message pair $(w_{11,b}, w_{12,b})$ where $w_{tr,b}$ is the message being transmitted in block b and $w_{tr,b-1}$ is the message being transmitted in block $b-1$ for $r, t = 1, 2$. At the beginning of block b, sender 2 has the estimation $\bar{\hat{w}}_{11,b-1}, \bar{\hat{w}}_{12,b-1}$, of the messages of sender 1 in the previous block, i.e., $w_{11,b-1}, w_{12,b-1}$ (see the decoding part). In the b^{th} block, sender 2 sends $x_2^n(w_{21,b}, w_{22,b}, w_{21,b-1}, w_{22,b-1}, \bar{\hat{}}w_{11,b-1}, \bar{\hat{}}w_{12,b-1})$ in order to transmit message pair $(w_{21,b}, w_{22,b})$. Note that the rate of senders 1 and 2 are defined as $R_1 = R_{11} + R_{12}$ and $R_2 = R_{21} + R_{22}$ respectively.

Note that in block b, sender 2 knows U_{1r}, $r = 1, 2$ part of the message being transmitted by sender 1. Therefore, sender 2 can use dirty paper coding [7] to mitigate the interference effect caused by sender 1 at receiver 2. This encoding scheme has been used in [5] for interference channels with conferencing. However, whether it can outperform the encoding scheme adopted in this paper is not clear and is currently under investigation.

Decoding: At the relay, we apply the regular encoding sliding-window decoding [27, 47], which achieves the same rate as the irregular encoding successive decoding [10] for the single relay channel. At the end of block b, decoding happens at the sender 2, receiver 1, and receiver 2 simultaneously.

Sender 2 declares that $(\bar{\hat{w}}_{11,b}, \bar{\hat{w}}_{12,b})$ was sent if there is a unique message pair $(\bar{\hat{w}}_{11,b}, \bar{\hat{w}}_{12,b}) \in \mathcal{W}_{11} \times \mathcal{W}_{12}$ such that

$$\left(x_{11}^n(\bar{\hat{w}}_{11,b}, w_{11,b-1}), x_{12}^n(\bar{\hat{w}}_{12,b}, w_{12,b-1}), u_{11}^n(w_{11,b-1}), u_{12}^n(w_{12,b-1}), Y_b^n\right) \in T_\epsilon^{(n)}, \quad (3.2.17)$$

if such a pair exists and is unique; otherwise, an error is declared.

Receiver 1 declares that message triple $(\hat{w}_{11,b-1}, \hat{w}_{12,b-1}, \hat{w}_{21,b-1}) \in \mathcal{W}_{11} \times \mathcal{W}_{12} \times \mathcal{W}_{21}$ is sent such that in both blocks b and $b-1$

$$\big(u_{11}^n(\hat{w}_{11,b-1}), u_{12}^n(\hat{w}_{12,b-1}), u_{21}^n(\hat{w}_{21,b-1}), Y_{1,b}^n\big) \in T_\epsilon^{(n)}, \tag{3.2.18}$$

$$\begin{aligned}\big(x_{11}^n(\hat{w}_{11,b-1}, w_{11,b-2}), x_{12}^n(\hat{w}_{12,b-1}, w_{12,b-2}),\\ x_{21}^n(\hat{w}_{21,b-1}, w_{21,b-2}), u_{11}^n(w_{11,b-2}),\\ u_{12}^n(w_{12,b-2}), u_{21}^n(w_{21,b-2}), Y_{1,b-1}^n\big) \in T_\epsilon^{(n)},\end{aligned} \tag{3.2.19}$$

if such a message triple exists and is unique; otherwise, an error is declared.

Receiver 2 declares that message triple $(\hat{\hat{w}}_{12,b-1}, \hat{\hat{w}}_{21,b-1}, \hat{\hat{w}}_{22,b-1}) \in \mathcal{W}_{12} \times \mathcal{W}_{21} \times \mathcal{W}_{22}$ is sent such that in both blocks b and $b-1$

$$\big(u_{12}^n(\hat{\hat{w}}_{12,b-1}), u_{21}^n(\hat{\hat{w}}_{21,b-1}), u_{22}^n(\hat{\hat{w}}_{22,b-1}), Y_{2,b}^n\big) \in T_\epsilon^{(n)}, \tag{3.2.20}$$

$$\begin{aligned}\big(x_{12}^n(\hat{\hat{w}}_{12,b-1}, w_{12,b-2}), x_{21}^n(\hat{\hat{w}}_{21,b-1}, w_{21,b-2}),\\ x_{22}^n(\hat{\hat{w}}_{22,b-1}, w_{22,b-2}), u_{12}^n(w_{12,b-2}),\\ u_{21}^n(w_{21,b-2}), u_{22}^n(w_{22,b-2}), Y_{2,b-1}^n\big) \in T_\epsilon^{(n)},\end{aligned} \tag{3.2.21}$$

if such a message triple exists and is unique; otherwise, an error is declared. Table 3.1 summarizes the encoding and decoding process for Theorem 3.2.1. As can be seen, during the first block, receivers 1 and 2 do not decode any message. In other words, the actual rate of senders are $\frac{b-1}{b} R_t$, $t = 1, 2$. This rate, however, approaches R_t as $b \to \infty$.

Analysis of Probability of Error: To obtain the probability of error for decoding in block b, we assume that no error has been made in decoding the previous

$b-1$ blocks. On the other hand, the codewords are independently and uniformly generated. In addition, the codebook in the block b is independent of that in block $b-1$. Therefore, without loss of generality, it can be assumed that in blocks $b-1$ and b, the messages $w_{tr,b-1}=1$, and $w_{tr,b}=1$ were sent for $r,t=1,2$. Moreover, we state the following definition and lemma as they will be frequently used in the proof.

Considering (3.2.17) – (3.2.21), we define events $E(\cdot)$, $E_{1u}(\cdot)$, $E_{1x}(\cdot)$, $E_{2u}(\cdot)$, and $E_{2x}(\cdot)$ as (3.2.22) – (3.2.26) respectively.

$$E(ijkl):=\{(x_{11}^n(i,k),x_{12}^n(j,l),u_{11}^n(k),u_{12}^n(l),y^n)\in T_\epsilon^{(n)}\}, \tag{3.2.22}$$

$$E_{1u}(ijk):=\{(u_{11}^n(i),u_{12}^n(j),u_{21}^n(k),y_1^n)\in T_\epsilon^{(n)}\}, \tag{3.2.23}$$

$$\begin{aligned}E_{1x}(ijklmn):=\{&(x_{11}^n(i,l),x_{12}^n(j,m),x_{21}^n(k,n),u_{11}^n(l),\\ &u_{12}^n(m),u_{21}^n(n),y_1^n)\in T_\epsilon^{(n)}\},\end{aligned} \tag{3.2.24}$$

$$E_{2u}(ijk):=\{(u_{12}^n(i),u_{21}^n(j),u_{22}^n(k),y_2^n)\in T_\epsilon^{(n)}\}, \tag{3.2.25}$$

$$\begin{aligned}E_{2x}(ijklmn):=\{&(x_{12}^n(i,l),x_{21}^n(j,m),x_{22}^n(k,n),u_{12}^n(l),\\ &u_{21}^n(m),u_{22}^n(n),y_2^n)\in T_\epsilon^{(n)}\}.\end{aligned} \tag{3.2.26}$$

Let the event that an error occurs at the sender 2 in block b be $E_{e,b}(Y)$. Therefore, at the sender 2, an error in decoding $(w_{11,b},w_{12,b})$ occurs with the probability

$$\begin{aligned}&\Pr(E_{e,b}(Y))\\ &=\Pr(E_{e,b}(Y)|(w_{11,b},w_{12,b},w_{11,b-1},w_{12,b-1})=(1,1,1,1))\\ &=\Pr\Big(E^c(1111)\bigcup\cup_{k\neq 1}E(k111)\bigcup\cup_{l\neq 1}E(1l11)\\ &\qquad\bigcup\cup_{\substack{k\neq 1\\ l\neq 1}}E(kl11)\Big|1111\Big),\end{aligned} \tag{3.2.27}$$

where $E^c(\cdot)$ indicates the complement of the event $E(\cdot)$, and $\Pr(\cdot)$ is the probability

measure.

For a randomly i.i.d. generated codebook, the probability of the events $E(k111)$, $E(1l11)$, and $E(kl11)$ for $k, l \neq 1$ are the same as those for $E(2111)$, $E(1211)$, and $E(2211)$ respectively given that $(w_{11,b}, w_{12,b}, w_{11,b-1}, w_{12,b-1}) = (1,1,1,1)$ was sent. On the other hand, according to joint asymptotic equipartition property (AEP) [11, Theorem 15.2.1], $\Pr(E^c(1111)|1111)$ approaches to zero when $n \to \infty$. We further apply the union bound to (3.2.27) and we will have

$$\begin{aligned}
&\Pr(E_{e,b}(Y)) \\
&\leq \ (2^{nR_{11}} - 1)P(E(2111)|1111) + \\
&\quad (2^{nR_{12}} - 1)P(E(1211)|1111) + \\
&\quad (2^{nR_{11}} - 1)(2^{nR_{12}} - 1)P(E(2211)|1111). \qquad (3.2.28)
\end{aligned}$$

Moreover, by letting $s_1 = \{x_{11}^n\}$, $s_2 = \{y^n\}$, and $s_3 = \{u_{11}^n, u_{12}^n, x_{12}^n\}$, it directly follows from Theorem 2.4.3 that

$$\Pr(E(2111)|1111) \doteq 2^{-nI(X_{11};Y|U_{11}U_{12}X_{12})\pm 6\epsilon)}.$$

Note that given the message tuple $(w_{11,b}, w_{12,b}, w_{11,b-1}, w_{12,b-1}) = (1,1,1,1)$ was sent , the probability mass function $P(x_{11}^n(2,1), x_{12}^n(1,1), u_{11}^n(1), u_{12}^n(1), y^n)$ can be decomposed as $P(x_{11}^n(2,1)|x_{12}^n(1,1), u_{11}^n(1), u_{12}^n(1)) \times P(y^n|x_{12}^n(1,1), u_{11}^n(1), u_{12}^n(1)) \times P(x_{12}^n(1,1), u_{11}^n(1), u_{12}^n(1))$.

Using the same approach, by adopting $s_1 = \{x_{12}^n\}$, $s_2 = \{y^n\}$, and $s_3 = \{u_{11}^n, u_{12}^n, x_{11}^n\}$, we have

$$\Pr(E(1211)|1111) \doteq 2^{-n(I(X_{12};Y|U_{11}U_{12}X_{11})\pm 6\epsilon)},$$

and further,

$$\Pr(E(2211)|1111) \doteq 2^{-n(I(X_{11}X_{12};Y|U_{11}U_{12})\pm 6\epsilon)},$$

as $n \to \infty$. By substituting these quantities in (3.2.28), we have $\Pr(E_{e,b}(Y)) \to 0$ when n is sufficiently large and (3.2.2) – (3.2.4) hold.

Denote by $E_{e,b}(Y_1)$the event that an error occurs at the receiver 1 in block b. At the receiver 1, an error in decoding $(\hat{w}_{11,b-1}, \hat{w}_{12,b-1}, \hat{w}_{21,b-1})$ occurs with the probability

$$\begin{aligned}
&\Pr(E_{e,b}(Y_1)) \\
&= \Pr\big(E_{e,b}(Y_1)|(w_{11,b-1}, w_{12,b-1}, w_{21,b-1}, w_{11,b-2}, \\
&\qquad\qquad w_{12,b-2}, w_{21,b-2}) = (1,1,1,1,1,1)\big) \\
&= \Pr\Big((E^c_{1u}(111) \cup E^c_{1x}(111111)) \\
&\quad \bigcup \cup_{k\neq 1}(E_{1u}(k11) \cap E_{1x}(k11111)) \\
&\quad \bigcup \cup_{l\neq 1}(E_{1u}(1l1) \cap E_{1x}(1l1111)) \\
&\quad \bigcup \cup_{\substack{k\neq 1\\ l\neq 1}}(E_{1u}(kl1) \cap E_{1x}(kl1111)) \\
&\quad \bigcup \cup_{\substack{k\neq 1\\ m\neq 1}}(E_{1u}(k1m) \cap E_{1x}(k1m111)) \\
&\quad \bigcup \cup_{\substack{l\neq 1\\ m\neq 1}}(E_{1u}(1lm) \cap E_{1x}(1lm111)) \\
&\quad \bigcup \cup_{\substack{k\neq 1\\ l\neq 1\\ m\neq 1}}(E_{1u}(klm) \cap E_{1x}(klm111))\big|111111\Big),
\end{aligned} \tag{3.2.29}$$

where $E_{1u}(\cdot) \cap E_{1x}(\cdot)$ means that both events $E_{1u}(\cdot)$ *and* $E_{1x}(\cdot)$ happen simultaneously. Note that the codebooks in two consecutive blocks are independent; hence, the probability of $E_{1u}(\cdot) \cap E_{1x}(\cdot)$ can be written as a product of probabilities of $E_{1u}(\cdot)$ and $E_{1x}(\cdot)$. By applying the union bound to (3.2.29) and using the same argument as that used to compute (3.2.27), we have

$$\begin{aligned}
&\Pr(E_{e,b}(Y_1)) \leq \\
&(2^{nR_{11}} - 1)\Pr(E_{1u}(211)|111)\Pr(E_{1x}(211111)|111111) + \\
&(2^{nR_{12}} - 1)\Pr(E_{1u}(121)|111)\Pr(E_{1x}(121111)|111111) +
\end{aligned}$$

$$
\begin{aligned}
&(2^{nR_{11}}-1)(2^{nR_{12}}-1)\Pr(E_{1u}(221)|111)\Pr(E_{1x}(221111)|111111)+\\
&(2^{nR_{11}}-1)(2^{nR_{21}}-1)\Pr(E_{1u}(212)|111)\Pr(E_{1x}(212111)|111111)+\\
&(2^{nR_{12}}-1)(2^{nR_{21}}-1)\Pr(E_{1u}(122)|111)\Pr(E_{1x}(122111)|111111)+\\
&(2^{nR_{11}}-1)(2^{nR_{12}}-1)(2^{nR_{21}}-1)\times\\
&\qquad\Pr(E_{1u}(222)|111)\Pr(E_{1x}(222111)|111111),
\end{aligned}
\tag{3.2.30}
$$

where $\Pr(E_{1u}(\cdot)|111)$ denotes the probability of $E_{1u}(\cdot)$ given that massage triple $(w_{11,b-1}, w_{12,b-1}, w_{21,b-1}) = (1,1,1)$ was sent, and $\Pr(E_{1x}(\cdot)|111111)$ denotes the probability of event $E_{1x}(\cdot)$ given that massage tuple $(w_{11,b-1}, w_{12,b-1}, w_{21,b-1}, w_{11,b-2}, w_{12,b-2}, w_{21,b-2}) = (1,1,1,1,1,1)$ was sent.

Let $s_1 = \{u_{11}^n\}$, $s_2 = \{y_1^n\}$, and $s_3 = \{u_{12}^n, u_{21}^n\}$. By applying Theorem 2.4.3, we have

$$
\Pr(E_{1u}(211)|111) \doteq 2^{-n(I(U_{11};Y_1|U_{12}U_{21})\pm 6\epsilon)}. \tag{3.2.31}
$$

Using a similar argument for each of $\Pr(E_{1u}(\cdot)|111)$ in (3.2.30), it can be written that

$$
\Pr(E_{1u}(121)|111) \doteq 2^{-n(I(U_{12};Y_1|U_{11}U_{21})\pm 6\epsilon)}, \tag{3.2.32}
$$

$$
\Pr(E_{1u}(221)|111) \doteq 2^{-n(I(U_{11}U_{12};Y_1|U_{21})\pm 6\epsilon)}, \tag{3.2.33}
$$

$$
\Pr(E_{1u}(212)|111) \doteq 2^{-n(I(U_{11}U_{21};Y_1|U_{12})\pm 6\epsilon)}, \tag{3.2.34}
$$

$$
\Pr(E_{1u}(122)|111) \doteq 2^{-n(I(U_{12}U_{21};Y_1|U_{11})\pm 6\epsilon)}. \tag{3.2.35}
$$

The probability of $E_{1u}(222)$ given message triple $(w_{11,b-1}, w_{12,b-1}, w_{21,b-1}) = (1,1,1)$ was sent can be computed as

$$
\begin{aligned}
&\Pr(E_{1u}(222)|111)\\
&= \sum_{(u_{11}^n, u_{12}^n, u_{21}^n, y_1^n)\in T_\epsilon^{(n)}} p(u_{11}^n, u_{12}^n, u_{21}^n, y_1^n)
\end{aligned}
$$

$$\begin{aligned}
&\overset{(a)}{=} |T_\epsilon^{(n)}|p(u_{11}^n)p(u_{12}^n)p(u_{21}^n)p(y_1^n) \\
&\overset{(b)}{\doteq} 2^{n(H(U_{11}U_{12}U_{21}Y_1)\pm 2\epsilon)}2^{-n(H(U_{11})\pm\epsilon)}2^{-n(H(U_{12})\pm\epsilon)} \\
&\quad \times 2^{-n(H(U_{21})\pm\epsilon)}2^{-n(H(Y_1)\pm\epsilon)} \\
&\overset{(c)}{\doteq} 2^{n(H(U_{11}U_{12}U_{21}Y_1)\pm 2\epsilon)}2^{-n(H(U_{11}U_{12}U_{21})\pm 3\epsilon)} \\
&\quad \times 2^{-n(H(Y_1)\pm\epsilon)} \\
&\doteq 2^{-n(I(U_{11}U_{12}U_{21};Y_1)\pm 6\epsilon)}, \qquad (3.2.36)
\end{aligned}$$

where

(a) follows from the fact that codewords are generated randomly, identically, and independently.

(b) follows from [11, Theorem 15.2.1],

(c) can be justified by considering [11, Theorem 15.2.1] and the fact that U_{11}, U_{12}, and U_{21} are generated independently.

To calculate $\Pr(E_{1x}(211111))$ given $(w_{11,b-1}, w_{12,b-1}, w_{21,b-1}, w_{11,b-2}, w_{12,b-2}, w_{21,b-2}) = (1,1,1,1,1,1)$, let $s_1 = \{x_{11}^n\}$, $s_2 = \{y_1^n\}$, and $s_3 = \{x_{12}^n, x_{21}^n, u_{11}^n, u_{12}^n, u_{21}^n\}$, and then, apply Theorem 2.4.3. Therefore, $\Pr(E_{1x}(211111)|111111)$ can be written as

$$\Pr(E_{1x}(211111)|111111) \doteq 2^{-n(I(X_{11};Y_1|X_{12}X_{21}U_{11}U_{12}U_{21})\pm 6\epsilon)}. \qquad (3.2.37)$$

Using a similar argument, the remaining probability terms in (3.2.30) can be computed as follows

$$\Pr(E_{1x}(121111)|111111) \doteq 2^{-n(I(X_{12};Y_1|X_{11}X_{21}U_{11}U_{12}U_{21})\pm 6\epsilon)}, \qquad (3.2.38)$$

$$\Pr(E_{1x}(221111)|111111) \doteq 2^{-n(I(X_{11}X_{12};Y_1|X_{21}U_{11}U_{12}U_{21})\pm 6\epsilon)}, \qquad (3.2.39)$$

Table 3.1: Summary of encoding and decoding processes for Theorem 3.2.1.

Block	1	2	$\cdots$	$b-1$	b
X_1	$x_1^n(w_{11,1}, w_{12,1}, \emptyset, \emptyset)$	$x_1^n(w_{11,2}, w_{12,2}, w_{11,1}, w_{12,1})$	$\cdots$	$x_1^n(w_{11,b-1}, w_{12,b-1}, w_{11,b-2}, w_{12,b-2})$	$x_1^n(w_{11,b}, w_{12,b}, w_{11,b-1}, w_{12,b-1})$
X_2	$x_2^n(w_{21,1}, w_{22,1}, \emptyset, \emptyset, \emptyset, \emptyset)$	$x_2^n(w_{21,2}, w_{22,2}, w_{11,1}, w_{12,1}, w_{21,1}, w_{22,1})$	$\cdots$	$x_2^n(w_{21,b-1}, w_{22,b-1}, w_{11,b-2}, w_{12,b-2}, w_{21,b-2}, w_{22,b-2})$	$x_2^n(w_{21,b}, w_{22,b}, w_{11,b-1}, w_{12,b-1}, w_{21,b-1}, w_{22,b-1})$
Y	$\bar{\hat{w}}_{11,1}, \bar{\hat{w}}_{12,1}$	$\bar{\hat{w}}_{11,2}, \bar{\hat{w}}_{12,2}$	$\cdots$	$\bar{\hat{w}}_{11,b-1}, \bar{\hat{w}}_{12,b-1}$	$\bar{\hat{w}}_{11,b}, \bar{\hat{w}}_{12,b}$
Y_1	$\emptyset, \emptyset, \emptyset$	$\hat{w}_{11,1}, \hat{w}_{12,1}, \hat{w}_{21,1}$	$\cdots$	$\hat{w}_{11,b-2}, \hat{w}_{12,b-2}, \hat{w}_{21,b-2}$	$\hat{w}_{11,b-1}, \hat{w}_{12,b-1}, \hat{w}_{21,b-1}$
Y_2	$\emptyset, \emptyset, \emptyset$	$\hat{\hat{w}}_{12,1}, \hat{\hat{w}}_{21,1}, \hat{\hat{w}}_{22,1}$	$\cdots$	$\hat{\hat{w}}_{12,b-2}, \hat{\hat{w}}_{21,b-2}, \hat{\hat{w}}_{22,b-2}$	$\hat{\hat{w}}_{12,b-1}, \hat{\hat{w}}_{21,b-1}, \hat{\hat{w}}_{22,b-1}$

$$\Pr(E_{1x}(212111)|111111) \doteq 2^{-n(I(X_{11}X_{21};Y_1|X_{12}U_{11}U_{12}U_{21})\pm 6\epsilon)}, \tag{3.2.40}$$

$$\Pr(E_{1x}(122111)|111111) \doteq 2^{-n(I(X_{12}X_{21};Y_1|X_{11}U_{11}U_{12}U_{21})\pm 6\epsilon)}, \tag{3.2.41}$$

$$\Pr(E_{1x}(222111)|111111) \doteq 2^{-n(I(X_{11}X_{12}X_{21};Y_1|U_{11}U_{12}U_{21})\pm 6\epsilon)}. \tag{3.2.42}$$

By substituting these probabilities into (3.2.30), we have $\Pr(E_{e,b}(Y_1)) \to 0$ when n is sufficiently large and (3.2.5) – (3.2.10) hold.

We further define the event that an error occurs in the block b at receiver 2 by $E_{e,b}(Y_2)$. In a similar manner, we can obtain the probability of error in decoding $(\hat{\hat{w}}_{12,b-1}, \hat{\hat{w}}_{21,b-1}, \hat{\hat{w}}_{22,b-1})$ at receiver 2, i.e., $\Pr(E_{e,b}(Y_2))$. The procedure for calculating the $\Pr(E_{e,b}(Y_2))$ can be imitated from that for $\Pr(E_{e,b}(Y_1))$ (i.e., (3.2.29) – (3.2.35), and (3.2.37) – (3.2.42)) by slight changes in indices. By doing so, it can be seen that $\Pr(E_{e,b}(Y_2)) \to 0$ when $n \to \infty$ and (3.2.11) – (3.2.16) hold.

Finally, it must be shown that any propensity for a catastrophic error propagation through out the blocks is excluded. We do this step using the same approach as

that in [10, 47]. Denote the union of the $E_{e,b}(Y)$, $E_{e,b}(Y_1)$, and $E_{e,b}(Y_2)$by F_b, i.e., $F_b = E_{e,b}(Y) \cup E_{e,b}(Y_1) \cup E_{e,b}(Y_2)$. In other words, F_b is the event of an error (false decoding) in block b at the receivers 1, 2, or sender 2. Also, denote the complement of this event (no error occurs in the block b) by F_b^c. Now, the probability of error can be written as

$$\begin{aligned} P_e^{(n)} &= \Pr\left(\cup_{i=1}^{b} F_i\right) \\ &= \Pr\left(\cup_{i=1}^{b}\left(F_i - \cup_{j=1}^{i-1} F_j\right)\right) \\ &= \sum_{i=1}^{b} \Pr\left(F_i \cap F_1^c \cap F_2^c \cdots \cap F_{i-1}^c\right) \\ &\leq \sum_{i=1}^{b} \Pr\left(F_i | F_1^c \cap F_2^c \cdots \cap F_{i-1}^c\right). \end{aligned} \tag{3.2.43}$$

Thus, if $\Pr\left(F_i | F_1^c \cap F_2^c \cdots \cap F_{i-1}^c\right) \to 0$, $i = 1, \ldots, b$, then $P_e^{(n)} \to 0$. ■

As mentioned before, sender 2 in Theorem 3.2.1 acts as a DF relay. Therefore, as a subtle assumption, the channel output Y_1 should be a degraded form of Y for this coding scheme; otherwise, the rate of sender 1 is unnecessarily limited by the sender 2.

An alternative approach for decoding at receivers can be backward decoding [50]. Although it rises to large delay, the backward decoding can give a simplified description for the achievable rate region. It has also been shown that the backward decoding outperforms sliding-window and successive decoding in the channels involving multiple access [48, 45], or in other words, channels with multi independent sources. This potential improvement of backward decoding over sliding widow decoding in our scenario (where the sources are not thoroughly independent) is currently under investigation.

As can be seen, the rate region, $\{(R_1, R_2)\}$, in Theorem 3.2.1 is given in an

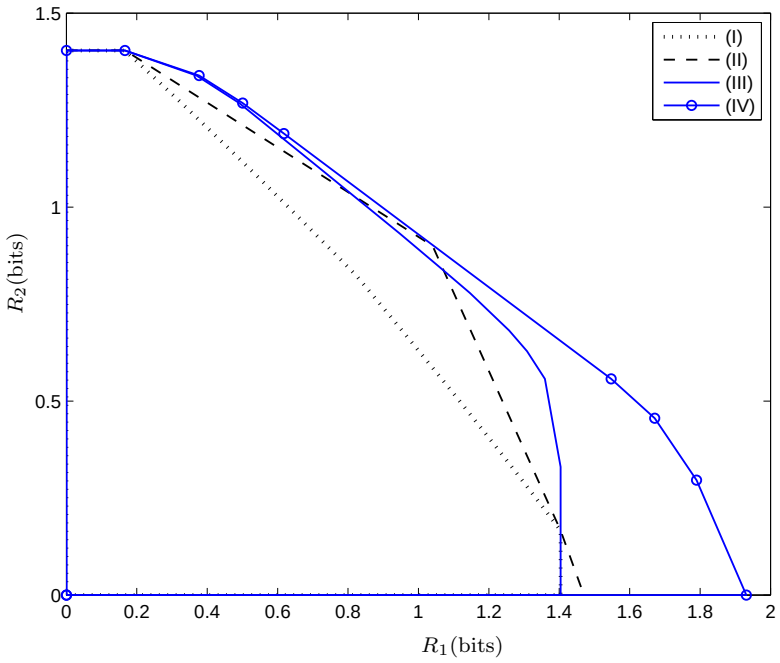

Figure 3.3: Comparison between achievable rate regions in the Gaussian CCR (GIC-CUC): (I) IC (HK rate region), (II) CCR in [13] when $c^2 = 10$, (III) Theorem 3.2.1 when $c = 1$, (IV) Theorem 3.2.1 when $c^2 = 10$. Channel parameters are $P_1 = 6$, $P_2 = 6$, $N = 1$, $N_1 = 1$, $N_2 = 1$, $a = 0.55$, and $b = 0.55$.

implicit form. We further apply the well-known Fourier-Motzkin elimination to obtain an explicit rate region. Define the right hand sides of (3.2.5) – (3.2.16), (3.2.2) – (3.2.4) as $c_1, c_2, \ldots, c_{15}$ respectively, and define $a_1 = \min\{c_1, c_{15}\}$, $a_2 = \min\{c_2, c_{16}\}$, $a_4 = \min\{c_3, c_{17}\}$, $a_i = c_i$ $(i = 4, 5, 6, \ldots, 15)$. The derivation procedure is shown in Appendix A and only the final result is presented here. The explicit rate region is given as

$$\begin{aligned} R_1 \leq \min\{&a_1 + a_5,\ a_1 + a_9,\ a_2 + a_4,\ a_6,\ a_4 + a_5,\ a_4 + a_9, \\ &a_1 + a_{10},\ a_1 + a_2,\ a_3\}, \end{aligned} \tag{3.2.44}$$

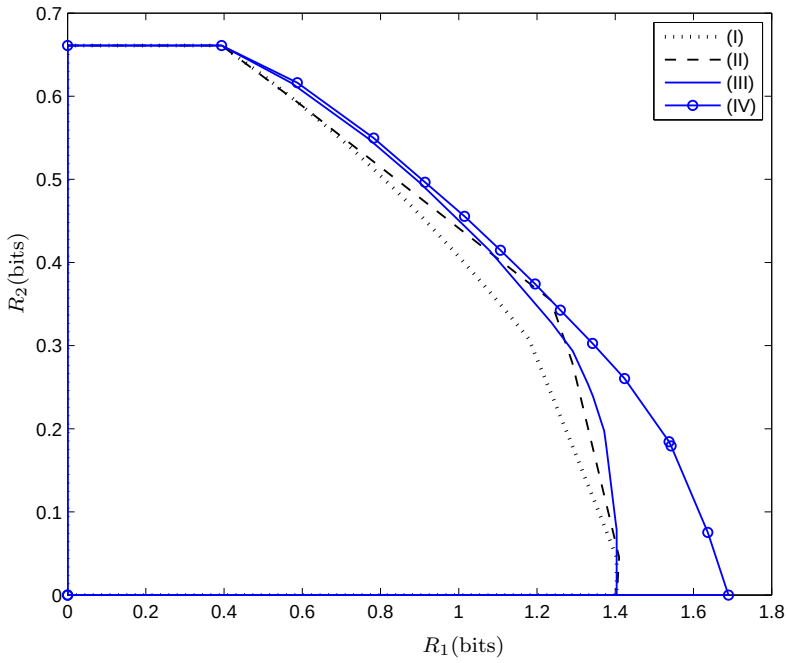

Figure 3.4: Comparison between achievable rate regions in the Gaussian CCR (GIC-CUC): (I) IC (HK rate region), (II) CCR in [13] when $c^2 = 10$, (III) Theorem 3.2.1 when $c = 1$, (IV) Theorem 3.2.1 when $c^2 = 10$. Channel parameters are $P_1 = 6$, $P_2 = 1.5$, $N = 1$, $N_1 = 1$, $N_2 = 1$, $a = 0.55$, and $b = 0.55$.

$$\begin{aligned} R_1 + R_2 \;\le\; \min\{&a_4 + a_{12},\; a_1 + a_5 + a_8,\; a_1 + a_9 + a_8,\; a_2 + \\ &a_4 + a_8,\; a_6 + a_8,\; a_1 + a_5 + a_{10},\; a_1 + a_9 + a_{10}, \\ &a_2 + a_4 + a_{10},\; a_6 + a_{10},\; a_1 + a_4 + a_{10}, \\ &a_1 + a_5 + a_{10},\; a_1 + a_7 + a_{10},\; a_1 + a_9 + a_{10}, \\ &a_4 + a_{10},\; \tfrac{1}{2}(a_1 + a_4 + a_{10} + a_{12}),\; a_1 + a_{12}\}, \end{aligned} \tag{3.2.45}$$

$$\begin{aligned} R_2 \;\le\; \min\{&a_4 + a_8,\; a_5 + a_8,\; a_7 + a_8,\; a_8 + a_9,\; a_4 + a_{10}, \\ &a_5 + a_{10},\; a_7 + a_{10},\; a_9 + a_{10},\; a_{11},\; a_{12}\}, \end{aligned} \tag{3.2.46}$$

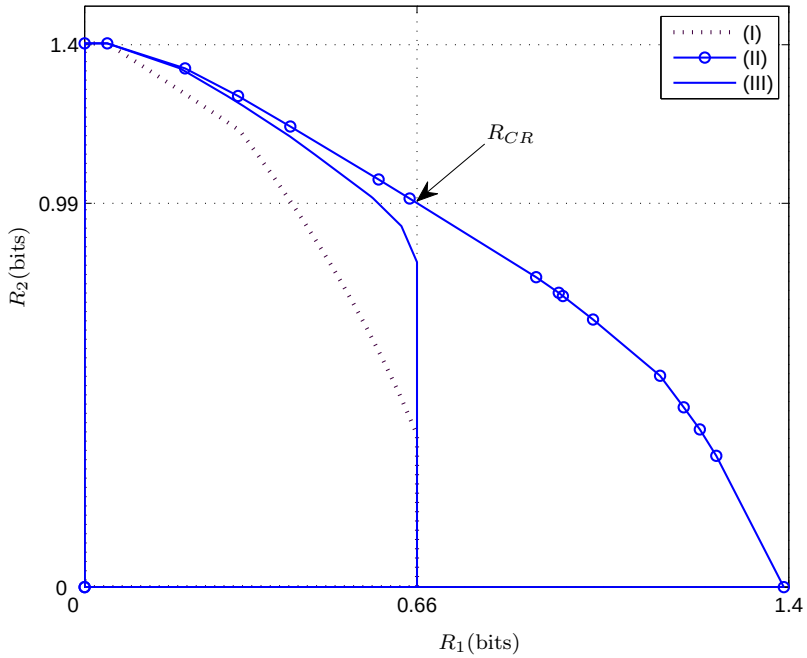

Figure 3.5: Achievable rate region for (I) HK region, (II) GIC-CUC when $c^2 = 1$, and (III) GIC-CUC when $c^2 = 10$. Other channel parameters are $P_1 = 1.5$, $P_2 = 6$, $N = 1$, $N_1 = 1$, $N_2 = 1$, $a = 0.55$, and $b = 0.55$. The point on the boundary of the regions marked by R_{CR}demonstrates the achievable rate by sender 1 as if no interference is caused by sender 2.

$$3R_1 + 2R_2 \leq \min\{2(a_1 + a_{10}) + a_4 + a_5,\ 2(a_1 + a_{10}) + a_4 + a_9\}, \tag{3.2.47}$$

$$\begin{aligned} R_1 + 2R_2 \leq \min\{&a_4 + a_5 + 2a_8,\ a_4 + a_9 + 2a_8,\ a_4 + a_5 + 2a_{10},\\ &a_4 + a_9 + 2a_{10},\ a_4 + a_8 + a_{12},\ a_4 + a_{10} + a_{12}\}, \end{aligned} \tag{3.2.48}$$

$$\begin{aligned} 2R_1 + R_2 \leq \min\{&2a_1 + a_5 + a_{10},\ 2a_1 + a_9 + a_{10},\ a_1 + a_6 + a_{10},\\ &a_1 + a_2 + a_4 + a_{10}\ \}. \end{aligned} \tag{3.2.49}$$

Further, the region in Theorem 3.2.1 is illustrated for the Gaussian case in Figs. 3.3,

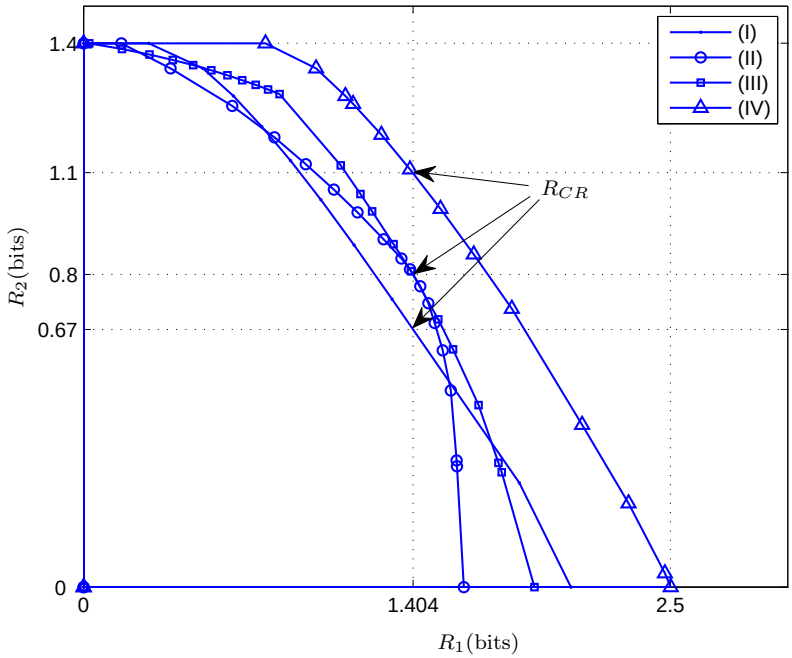

Figure 3.6: Achievable rate region for GIC-CUC for channel parameters $P_1 = 6$, $P_2 = 6$, $N = 1$, $N_1 = 1$, $N_2 = 1$, $c^2 = 10$, (I) $a = 0.74$, and $b = 0.74$, (II) $a = 0.2$, and $b = 0.54$, (III) $a = 0.54$, and $b = 0.2$, (IV) $a = 1.5$, and $b = 1.5$. The point on the boundary of the regions marked by R_{CR} denotes the achievable rate by sender 1 as if no interference is caused by sender 2.

3.4, 3.5, and 3.6.

3.3 The Gaussian IC-CUC

In this section, the achievable rate region in Theorem 3.2.1 is demonstrated for the Gaussian IC-CUC (GIC-CUC). Without loss of the information-theoretic optimality, the GIC-CUC can be converted into the GIC-CUC in the standard form through invertible transformations. We thus only focus on the GIC-CUC in the standard

form represented as follows:

$$Y_1 = X_1 + a\,X_2 + Z_1, \tag{3.3.1}$$

$$Y_2 = b\,X_1 + X_2 + Z_2, \tag{3.3.2}$$

$$Y = c\,X_1 + Z, \tag{3.3.3}$$

where Z_1, Z_2, and Z are additive white Gaussian noises whose powers are N_1, N_2, and N respectively. X_1 and X_2 are the inputs of Gaussian IC-CUC with respective maximum transmit powers P_1 and P_2. To compute the achievable rate region in Theorem 3.2.1, all random variables in (3.2.1) should be mapped into gaussian random variables. In order to map the generated codebook into Gaussian random variables, normal kernels U_{11}, U_{12}, U_{21}, U_{22}, X'_{11} ,X'_{12}, X'_{21}, and X'_{22} with zero means and unit variances are defined. Using these kernels for $0 \leq \alpha_t, \beta_t, \gamma_t, \lambda, \xi \leq 1$, and $\bar{\alpha}_t = 1 - \alpha_t$, $\bar{\beta}_t = 1 - \beta_t$, $\bar{\gamma}_t = 1 - \gamma_t$, $\bar{\lambda} = 1 - \lambda$, $\bar{\xi} = 1 - \xi$, where $t = 1, 2$, the codebook can be mapped as

$$X_{11} = \sqrt{\gamma_1 \alpha_1 P_1}\, X'_{11} + \sqrt{\gamma_1 \bar{\alpha}_1 P_1}\, U_{11}, \tag{3.3.4}$$

$$X_{12} = \sqrt{\bar{\gamma}_1 \beta_1 P_1}\, X'_{12} + \sqrt{\bar{\gamma}_1 \bar{\beta}_1 P_1}\, U_{12}, \tag{3.3.5}$$

$$X_1 = X_{11} + X_{12}, \tag{3.3.6}$$

$$X_{21} = \sqrt{\lambda \gamma_2 \alpha_2 P_2}\, X'_{21} + \sqrt{\lambda \gamma_2 \bar{\alpha}_2 P_2}\, U_{21}, \tag{3.3.7}$$

$$X_{22} = \sqrt{\lambda \bar{\gamma}_2 \beta_2 P_2}\, X'_{22} + \sqrt{\lambda \bar{\gamma}_2 \bar{\beta}_2 P_2}\, U_{22}, \tag{3.3.8}$$

$$X_2 = X_{21} + X_{22} + \sqrt{\bar{\lambda} \xi P_2}\, U_{11} + \sqrt{\bar{\lambda} \bar{\xi} P_2}\, U_{12}. \tag{3.3.9}$$

By this mapping, the mutual information terms in (3.2.2) – (3.2.16) can be computed.

In a general sense, the interference channel (IC), when there is no cooperation between either senders or receivers, can be thought as a special case of the CCR

(see [13, Protocol 3]). The HK region is the best known achievable rate region for this channel [20]. In addition, [13] proposes three other protocols for the CCR and derives the convex hull of all those four protocols as an achievable rate region for CCR.

Figs. 3.3 and 3.4 illustrate and compare rate regions of HK region for IC, CCR in [13], and Theorem 3.2.1. As demonstrated in Figs. 3.3 and 3.4, when there is a low noise channel between sender 1 and CR (for example $c^2/N = 10$ in this case), both regions in [13] and Theorem 3.2.1 outperforms the HK rate region. On the other hand, when this channel is noisy (for example $c^2/N = 1$), the the rate region for CCR in [13] almost reduces to HK rate region, while the rate region of Theorem 3.2.1 includes that of HK. As shown in Figs. 3.3 and 3.4, the rate region in Theorem 3.2.1outperforms that of the CCR in [13].

Fig. 3.5 shows the achievable rate region for the GIC-CUC when the sender 2 has more power than sender 1. As can be seen, the sender 2's cooperation can significantly increase the rate of sender 1. Fig. 3.6 shows the achievable rate region for different channel gains. As expected, the weaker interference results in better performance.

There is indeed an interesting point in this region (R_{CR}) in which sender 1 can achieve its point to point capacity, i.e, $0.5\log(1+P_1/N_1)$, as if there is no interference caused by the sender 2. This point can be thought as a rate pair in which the cognitive user does not degrade the performance of the primary user while having a non-zero rate. In other words, the point may be considered as a performance criteria for the cognitive radio. Therefore, the goal for cognitive users is to achieve maximum rate while letting the primary user to have the point to point capacity.

3.4 Summary

In this chapter, we proposed a coding scheme for the IC-CUC and derived an achievable rate region for this channel. It was also shown that the derived achievable rate outperforms that of CCR in [13] when the receiver 1 is a degraded form that of the sender 2. This improvement becomes more pronounced as Y_1 becomes more degraded form Y (for example when $c^2/N \gg 1$ in the Gaussian channel).

Chapter 4

Conclusion

In this dissertation, we mainly studied the CR in an information theoretic perspective. For this purpose, the necessary mathematical tools were defined and introduced in Chapter 2. Firstly, the concept of entropy as an information measure has been defined. The entropy as a quintessential concept was used to define the distance of probability distributions and mutual information between two (or more) RVs. Next, the fundamental capacity theorem of a single user DMC was stated and the proof was given according to what Shannon [38] had arrived at. As mentioned the proving steps in this theorem tends to be a panacea for almost all other channels. In other words, to achieve the capacity region, we first obtain an inner bound on capacity, and then, an outer bound. If the inner and outer bounds completely overlap with each other, the capacity is obtained. As it was shown (and also Shannon [38] showed), random coding and jointly typical decoder achieve the capacity as the codeword length approaches to infinity.[1]

[1]It is, however, interesting that such an inefficient decoder can achieve the capacity. There are basically two different types of decoders. The first one is Maximum Likelihood (ML) decoder. In ML decoder, the decoder calculates the probability of sent symbol given the received symbol. Then the symbol with highest probability is declared as the transmitted symbol. The second type of decoders is the Jointly Typical decider. This decoder compares that empirical probability of received codeword and each of the possible transmitted codewords. Then, it declares the transmitted codeword if this

Chapter 3 elaborates on the main contribution of this dissertation. In this chapter, the new causal concurrent model for the CR was investigated. In this model, the cognitive user has the dual role of relaying the message of the first user as well as broadcasting its own message. Moreover, the nature of the channel is of the interference channel. Therefore, a combination of coding methods is used to establish the achievable rate region. First of all, the rates are split into two parts so that each receiver can decode a part of other sender's message instead of treating it entirely as noise. Secondly, since there is one cooperative node (which is the CR), two blocks of messages are superimposed onto each other. In other words, the CR decodes the message of current block and will forward it in the next block. The performance of the overall coding and decoding scheme is shown to outperform that of existing results. This performance improves when the channel between the primary and the secondary users has a high capacity. Also, it was shown that by implementing the proposed coding scheme, the primary user can have its point to point capacity as if there is no interference by the second user while the second user can have a nonzero rate. Nonetheless, the performance of our coding scheme suffers severe degradation when the channel between sender 1 and sender 2 is more noisy than the one between sender 1 and receiver 1. This scenario happens when two senders are geometrically far apart from each other.

As a future work, the case when the receiver 1 is not degraded from sender 2 can be investigated. In this case, the compress and forward method [6, 10] should be applied to the relay part of the model. Therefore, the CR is not obligated to decode the message of the first sender thoroughly, and consequently, the rate of the first

empirical probability is *close* to the actual probability.

sender will not be limited by the CR. In other words, the CR will not be a bottle neck for the rate of the first user. Another interesting question to be addressed is the capacity of the CR when the primary user holds the point to point capacity. In addition, there seems to be some other techniques that can potentially enlarge the achievable rate region. For instance, the known message at the CR can be treated as a known state whose interference effect can be mitigated by incorporating Gel'fand Pinsker binning method [19]. In other words, given that the $U_{1r,b}^n$ is known for the second sender, DPC can be used to mitigate the interference effect on the second sender.

As a cooperator, the CR can act selfishly meaning that it consumes all of its power to transmit its own message. On the other hand, the CR can sacrifice and allocate significant portion of its power to amplify the received state. In other words, it can help to resolve the uncertainly about the first sender's message at the primary receiver. The optimum tradeoff between these two situations is interesting to be investigated as a future work. On one hand, the CR can *amplify* the message of the first user as a known state information [26]. On the other hand, the cognitive radio might need to mask the message of the first user rather than conveying it. State masking was studied in the work of Merhav and Shamai [32]. Masking of the state is important to address the privacy issue. In other words, it is important to know how much can be learned about the state at the second receiver form the channel output.

The cooperation discussed in this dissertation is unidirectional meaning that one of the senders receive the message of the other one. This cooperation can also be bidirectional. The bidirectional cooperation has been investigates under different names [5, 41]. The overall rate region of an interference channel with bidirectional

cooperation is the union of rate regions of that with cooperation from sender one to the sender two and vise versa. The ultimate goal in interference channel with perfect bidirectional cooperation is to reach the capacity of the MIMO broadcast channel. This capacity has been obtained by Weingarten et al. [43] for the Gaussian channel.

Appendix A

Fourier Motzkin Elimination for Theorem 3.2.1

As shown in Chapter 3, the rate region, $\{(R_1, R_2)\}$, in Theorem 3.2.1 is given in an implicit form. To obtain the rate region in an explicit form, the Fourier Motzkin elimination method [35, pp. 155–157] has to be applied.

The rate region in Theorem 3.2.1 is rewritten here again as

$$\begin{aligned}
R_{11} &\leq a_1, \\
R_{12} &\leq \min\{a_2, a_8\}, \\
R_{21} &\leq \min\{a_3, a_9\}, \\
R_{11} + R_{12} &\leq a_4, \\
R_{11} + R_{21} &\leq a_5, \\
R_{12} + R_{21} &\leq \min\{a_6, a_{11}\}, \\
R_{11} + R_{12} + R_{21} &\leq a_7, \\
R_{22} &\leq a_{10}, \\
R_{12} + R_{22} &\leq a_{12}, \\
R_{21} + R_{22} &\leq a_{13},
\end{aligned}$$

$$R_{12} + R_{21} + R_{22} \leq a_{14}.$$

Firstly, we replace R_{11}, R_{12}, R_{21}, R_{22} with S_1, T_1, S_2, T_2 respectively, and add the nonnegativity conditions on the rates. Hence,

$$\begin{aligned}
S_1 &\leq a_1, \\
T_1 &\leq \min\{a_2, a_8\}, \\
T_2 &\leq \min\{a_3, a_9\}, \\
R_1 &\leq a_4, \\
S_1 + T_2 &\leq a_5, \\
T_1 + T_2 &\leq \min\{a_6, a_{11}\}, \\
R_1 + T_2 &\leq a_7, \\
S_2 &\leq a_{10}, \\
T_1 + S_2 &\leq a_{12}, \\
R_2 &\leq a_{13}, \\
T_1 + R_2 &\leq a_{14}, \\
-S_1 &\leq 0, \\
-T_1 &\leq 0, \\
-S_2 &\leq 0, \\
-T_2 &\leq 0.
\end{aligned}$$

Then, S_1 and S_2 will be substituted with their respective values $R_1 - T_1$ and $R_2 - T_2$. Hence,

$$R_1 - T_1 \leq a_1,$$

$$
\begin{aligned}
T_1 &\leq \min\{a_2, a_8\},\\
T_2 &\leq \min\{a_3, a_9\},\\
R_1 &\leq a_4,\\
R_1 - T_1 + T_2 &\leq a_5,\\
T_1 + T_2 &\leq \min\{a_6, a_{11}\},\\
R_1 + T_2 &\leq a_7,\\
R_2 - T_2 &\leq a_{10},\\
T_1 + R_2 - T_2 &\leq a_{12},\\
R_2 &\leq a_{13},\\
T_1 + R_2 &\leq a_{14},\\
-T_1 &\leq 0,\\
-T_2 &\leq 0,\\
-R_1 + T_1 &\leq 0,\\
-R_2 + T_2 &\leq 0.
\end{aligned}
$$

Suppose the objective is now to eliminate the variable T_1. To do so, we cluster these equations in three groups. The first group contains all in inequalities in which sign of T_1 is positive. The second group contains all inequalities in which T_1 has a negative sign, and the rest will be in the third group. In other words,

$$
\begin{aligned}
T_1 &\leq \min\{a_2, a_8\},\\
T_1 + T_2 &\leq \min\{a_6, a_{11}\},\\
T_1 + R_2 - T_2 &\leq a_{12},\\
T_1 + R_2 &\leq a_{14},
\end{aligned}
$$

$$
\begin{aligned}
T_1 - R_1 &\leq 0,\\
\hline
-T_1 &\leq 0,\\
-T_1 + R_1 &\leq a_1,\\
-T_1 + T_2 + R_1 &\leq a_5,\\
\hline
T_2 &\leq \min\{a_3, a_9\},\\
R_1 &\leq a_4,\\
R_1 + T_2 &\leq a_7,\\
R_2 - T_2 &\leq a_{10},\\
R_2 &\leq a_{13},\\
-T_2 &\leq 0,\\
-R_2 + T_2 &\leq 0.
\end{aligned}
$$

Then, all the inequalities from the first group are added with all inequalities in the second group, and as a result, there will be $5 \times 3 = 15$ new inequalities. These inequalities in addition to the ones from the third group result in total $15 + 7 = 22$ inequalities. The inequalities with a similar left hand side can be embedded together and it can be written

$$
\begin{aligned}
T_2 &\leq \min\{a_3, a_5, a_6, a_9, a_{11}\},\\
R_1 + T_2 &\leq \min\{a_1 + a_6, a_1 + a_{11}, a_2 + a_5, a_5 + a_8, a_7\},\\
R_1 + 2T_2 &\leq \min\{a_5 + a_6, a_5 + a_{11}\},\\
R_1 + R_2 + T_2 &\leq \min\{a_5 + a_{14}\},\\
-R_2 + T_2 &\leq 0,\\
R_2 - T_2 &\leq \min\{a_{10}, a_{11}\},
\end{aligned}
$$

$$
\begin{aligned}
R_1 + R_2 - T_2 &\leq a_1 + a_{12}, \\
R_1 &\leq \min\{a_1 + a_2, a_1 + a_8, a_4\}, \\
R_2 &\leq \min\{a_{13}, a_{14}\}, \\
R_1 + R_2 &\leq \min\{a_1 + a_{14}, a_5 + a_{12}\}.
\end{aligned}
$$

In the next step, the variable T_2 has to be eliminated. To do so, the inequality set will be grouped in a similar manner, and we will have

$$
\begin{aligned}
T_2 &\leq \min\{a_3, a_5, a_6, a_9, a_{11}\}, \\
T_2 + R_1 &\leq \min\{a_1 + a_6, a_1 + a_{11}, a_2 + a_5, a_5 + a_8, a_7\}, \\
T_2 + \frac{1}{2}R_1 &\leq \frac{1}{2}\min\{a_5 + a_6, a_5 + a_{11}\}, \\
T_2 + R_1 + R_2 &\leq \min\{a_5 + a_{14}\}, \\
T_2 - R_2 &\leq 0, \\
\hline
-T_2 + R_2 &\leq \min\{a_{10}, a_{11}\}, \\
-T_2 + R_1 + R_2 &\leq a_1 + a_{12}, \\
\hline
R_1 &\leq \min\{a_1 + a_2, a_1 + a_8, a_4\}, \\
R_2 &\leq \min\{a_{13}, a_{14}\}, \\
R_1 + R_2 &\leq \min\{a_1 + a_{14}, a_5 + a_{12}\}.
\end{aligned}
$$

After eliminating T_2 and embedding the appropriate inequalities, we will have

$$
\begin{aligned}
R_1 &\leq \min\{a_1 + a_5,\ a_1 + a_9,\ a_2 + a_4,\ a_6,\ a_4 + a_5,\ a_4 + a_9, \\
&\qquad a_1 + a_{10},\ a_1 + a_2,\ a_3\}, \qquad \text{(A.0.1)} \\
R_1 + R_2 &\leq \min\{a_4 + a_{12},\ a_1 + a_5 + a_8,\ a_1 + a_9 + a_8,\ a_2 + \\
&\qquad a_4 + a_8,\ a_6 + a_8,\ a_1 + a_5 + a_{10},\ a_1 + a_9 + a_{10},
\end{aligned}
$$

$$a_2 + a_4 + a_{10},\ a_6 + a_{10},\ a_1 + a_4 + a_{10},$$

$$a_1 + a_5 + a_{10},\ a_1 + a_7 + a_{10},\ a_1 + a_9 + a_{10},$$

$$a_4 + a_{10},\ \tfrac{1}{2}(a_1 + a_4 + a_{10} + a_{12}),\ a_1 + a_{12}\}, \tag{A.0.2}$$

$$R_2 \leq \min\{a_4 + a_8,\ a_5 + a_8,\ a_7 + a_8,\ a_8 + a_9,\ a_4 + a_{10},$$

$$a_5 + a_{10},\ a_7 + a_{10},\ a_9 + a_{10},\ a_{11},\ a_{12}\}, \tag{A.0.3}$$

$$3R_1 + 2R_2 \leq \min\{2(a_1 + a_{10}) + a_4 + a_5,\ 2(a_1 + a_{10}) + a_4 + a_9\}, \tag{A.0.4}$$

$$R_1 + 2R_2 \leq \min\{a_4 + a_5 + 2a_8,\ a_4 + a_9 + 2a_8,\ a_4 + a_5 + 2a_{10},$$

$$a_4 + a_9 + 2a_{10},\ a_4 + a_8 + a_{12},\ a_4 + a_{10} + a_{12}\}, \tag{A.0.5}$$

$$2R_1 + R_2 \leq \min\{2a_1 + a_5 + a_{10},\ 2a_1 + a_9 + a_{10},\ a_1 + a_6 + a_{10},$$

$$a_1 + a_2 + a_4 + a_{10}\ \}, \tag{A.0.6}$$

and that is the final step since we have the explicit rate region including only R_1 and R_2.

Bibliography

[1] "Federal Communications Commission, Cognitive Radio Technologies Proceeding (CRTP)." [Online]. Available: http://www.fcc.gov/oet/cognitiveradio/

[2] R. Ahlswede, "Multiway communication channels," in *Proceeding of International Symposium on Information Theory (ISIT)*, (Tsahkadsor), Arminian S.S.R. Prague, 1971, pp. 23–52.

[3] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory*, vol. 19, pp. 197–207, 1973.

[4] P. Billingsley, *Probability and Measure*, 3rd ed. Haboken, New Jersey: John Wiley & Sons, 1995.

[5] Y. Cao and B. Chen, "An achievable rate region for interference channels with conferencing," in *Proceeding of International Symposium on Information Theory (ISIT)*, Nice, France, Jun. 2007, pp. 1251–1255.

[6] H. F. Chong, M. Motani, and H. K. Garg, "Generalized backward decoding strategies for the relay channel," *IEEE Transactions on Information Theory*, vol. 53, no. 1, pp. 394–401, Jan. 2007.

[7] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[8] M. H. M. Costa and A. A. El Gamal, "The capacity region of the discrete memoryless interference channel with strong interference (corresp.)," *IEEE Transactions on Information Theory*, vol. 33, pp. 710–711, Sep. 1987.

[9] T. M. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 2–14, Jan. 1972.

[10] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, pp. 572–584, Sep. 1979.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Haboken, New Jersey: John Wiley & Sons, 2006.

[12] I. Csiszár and J. Kőrner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

[13] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.

[14] A. A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Transactions on Information Theory*, vol. 28, p. 536, May 1982.

[15] A. A. El Gamal and E. C. van der Meulen, "A proof of marton's coding theorem for the discrete memoryless broadcast channel (corresp.)," *IEEE Transactions on Information Theory*, vol. 27, pp. 120–122, Jan. 1981.

[16] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," 2007. [Online]. Available: http://arxiv.org/abs/cs/0702045v2

[17] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredaci Informaccii*, vol. 10, no. 3, pp. 3–14, 1974.

[18] ——, *Information Theory and Reliable Communication.* Haboken, New Jersey: John Wiley & Sons, 1968.

[19] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[20] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Transactions on Information Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.

[21] S. Haykin, "Cognitive radio: Brain–empowered wireless communications," *IEEE Transactions on Information Theory*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[22] J. Jiang and Y. Xin, "On the achievable rate regions for interference channels with degraded message sets," *IEEE Transactions on Information Theory*, (to appear).

[23] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," 2006. [Online]. Available: http://arxiv.org/abs/cs/0604107

[24] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information: subject review," *accepted to Foundations and Trends in Communications and Information Theory*, Jan. 2007.

[25] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "Improved achievable rates for user cooperation and relay channels," in *Proceeding of International Symposium on Information Theory (ISIT)*, Chicago, USA, Jun./Jul. 2004, p. 4.

[26] Y. H. Kim, A. Sutivong, and T. M. Cover, "State amplification," 2007. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0703005

[27] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 572–584, Sep. 2005.

[28] N. V. Kuznetsov and B. S. Tsybakov, "Coding in memories with defective cells," *Problemy peredachi informatsii*, vol. 10, no. 2, pp. 52–60, 1974.

[29] I. Marić, A. Goldsmith, G. Kramer, and S. Shamai, "On the capacity of interference channels with a partially-cognitive transmitter," 2007. [Online]. Available: http://systems.stanford.edu/ ivanam/publications.html

[30] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. 25, pp. 306–311, May 1979.

[31] R. J. McEliece, *The Theory of Information and Coding : a mathematical framework for communication.* Addison-Wesley, 1977.

[32] N. Merhav and S. Shamai, "Information rates subjected to state masking," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 2254–2261, Jun. 2007.

[33] H. Sato, "An outer bound to the capacity region of broadcast channels," *IEEE Transactions on Information Theory*, vol. 24, pp. 374–377, May 1978.

[34] ——, "The capacity of the gaussian interference channel under strong interference (corresp.)," *IEEE Transactions on Information Theory*, vol. 27, pp. 786–788, Nov. 1981.

[35] A. Schrijver, *Theory of Linear and Integer Programming.* Haboken, New Jersey: John Wiley & Sons, 1986.

[36] S. H. Seyedmehdi, Y. Xin, and Y. Lian, "An achievable rate region for the causal cognitive radio," in *Proc. 45th Ann. Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, USA, Sep. 2007.

[37] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "On the capacity of mimo interference channels," 2008. [Online]. Available: http://arxiv.org/abs/cs/08071543v1

[38] C. E. Shannon, "A mathematical theory of communicaion," *Bell Sys. Tech. Journal*, no. 27, pp. 379–423, 623–656, 1948.

[39] ——, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, pp. 289–293, Oct. 1958.

[40] S. Sigurjonsson and Y. H. Kim, "On multiple user channels with state information at the transmitters," in *Proceeding of International Symposium on Information Theory (ISIT)*, Adelaide, Australia, Sep. 2005, pp. 72–76.

[41] D. Tuninetti, "On interference channel with generalized feedback (ifc-gf)," in *Proceeding of International Symposium on Information Theory (ISIT)*, Nice, France, Jun. 2007, pp. 2861–2865.

[42] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[43] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[44] F. M. J. Willems and E. C. van der Meulen, "Partial feedback for the discrete memoryless multiple access channel," *IEEE Transactions on Information Theory*, vol. 29, no. 2, pp. 287–290, Mar. 1983.

[45] ——, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Transactions on Information Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[46] W. Wu, S. Vishwanath, and A. Arapostathis, "On the capacity of interference channel with degraded message sets," 2006. [Online]. Available: http://www.arxiv.org/abs/cs/0605071v1

[47] L. L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.

[48] ——, "Multi-source, multi-destination, multi-relay wireless networks," *To appear in IEEE Trans. on Inform. Theory, Special Issue on Models, Theory and Codes for Relaying and Cooperation in Communication Networks*, Oct. 2007.

[49] R. W. Yeung, *A First Cours in Information Theory.* Kluwer Academic, 2002.

[50] C. M. Zeng, F. Kuhlmann, and A. Buzo, "Achievability proof of some multiuser channel coding theorems using backward decoding," *IEEE Transactions on Information Theory*, vol. 35, no. 6, pp. 1160–1165, Nov. 1989.