UNIVERSITY OF CALIFORNIA, SAN DIEGO

Interactive Schemes in Information Theory and Statistics

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering
(Communication Theory and Systems)

by

Yu Xiang

Committee in charge:

Professor Young-Han Kim, Chair
Professor Ery Arias-Castro
Professor Bruce Driver
Professor Massimo Franceschetti
Professor Alon Orlitsky

2015

The dissertation of Yu Xiang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

DEDICATION

To my family.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I have been extremely fortunate to have Prof. Young-Han Kim as my advisor. I am deeply indebted to Young-Han for his invaluable guidance and support throughout my graduate study. He taught me everything I know about doing research, for which I can't thank him enough. In particular, there are two things I learned from him that have really changed me. First, always strive to bring clarity to complicated problems. Second, attention to detail. I can still vividly remember that he helped me preparing my first ISIT presentation in 2010. We went through the slides over and over again to fix all the details in his hotel room till midnight. In addition, I appreciate his endless patience and the freedom he gave me in all these years.

I would like to thank my committee members Prof. Ery Arias-Castro, Prof. Bruce Driver, Prof. Alon Orlitsky, and Prof. Massimo Franceschetti for agreeing to be on my committee and providing valuable suggestions and feedback. I thank Prof. Alon Orlitsky for his unique interactive teaching style that I feel fortunate to have the opportunity to experience as a student. I am grateful for being able to contribute (a negligible amount) to the ITA workshop. I thank Prof. Massimo Franceschetti for teaching me percolation theory and the "coffee time" puzzle book, which has brought many fun to me and Young-Han's round-robin group meeting. I thank Prof. Bruce Driver for his year-long probability theory course and Prof. Ery Arias-Castro for helpful feedbacks on my interactive hypothesis testing work. I thank Prof. Ofer Shayevitz for many helpful discussions that lead to the variable-length version of my interactive hypothesis testing work. I also thank Prof. Ning Cai for his care and encouragement in all these years.

I would like to thank all my current and former labmates: Fatemeh Arbabjolfaei, Ehsan Ardestanizadeh, Halyun Jeong, and Hwan Joon (Eddy) Kwon for discussing various research topics and their warm friendship. Special thanks to Chiao-Yi Chen and Lele Wang for their support and countless interesting conversations on research and life. I thank Sudeep Kamath and Prof. Himanshu Tyagi for co-organizing the reading group on concentration inequalities, from where I learned a great deal. I have been very fortunate to have many wonderful friends

here at UCSD, in particular, Houdong Hu, Zhengtao Qin, Bo Yang, Jia Guo, Yiran Shen, Dongjin Song, Duo Song, Hua Zhang. It is hard to imagine my graduate life without their support and all the fun times we had. I thank Shengjun Pan for his continual friendship and many helps including sharing this dissertation template.

Finally, my deepest gratitude goes to my family. I thank my parents for their unconditional love and support throughout all my endeavors. I thank my wife Mengke for sharing all the ups and downs with me and making Philadelphia feel like home for me. I thank my late maternal grandmother for giving me endless love and care. I will miss her forever. I am so grateful for having my newborn son Matthew, who has brought so much joy to the whole family. Checking out his new photos has become my daily routine. I thank my mother-in-law and my mother for taking good care of Matthew and my wife during her pregnancy and after her delivery. To them I dedicate this dissertation.

The following co-authored material has been used in the this dissertation. Chapter 3, in part, includes the material in Yu Xiang and Young-Han Kim, "Interactive hypothesis testing with communication constraints," to be submitted for publication in *IEEE Transaction on Information Theory*. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, includes the material in Yu Xiang and Young-Han Kim, "Gaussian channel with noisy feedback and peak energy constraint," *IEEE Transaction on Information Theory,* vol.59, no.8, pp.4746–4756, August 2013. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in part, includes the material in Yu Xiang, Lele Wang, and Young-Han Kim, "Information flooding," *Annual Allerton Conference on Communication, Control, and Computing,* pp. 45–51, Monticello, IL, September 2011. The dissertation author was the primary investigator and author of this paper.

VITA

2008	B.S. in Telecommunication Engineering, Xidian University, China
2010	M.S. in Electrical Engineering (Communication Theory and Systems), University of California, San Diego, United States
2015	Ph.D. in Electrical Engineering (Communication Theory and Systems), University of California, San Diego, United States

PUBLICATIONS

Yu Xiang and Young-Han Kim, "A few meta-theorems in network information theory," to be submitted for publication in *IEEE Transaction on Information Theory.*

Yu Xiang and Young-Han Kim, "Interactive hypothesis testing with communication constraints," to be submitted for publication in *IEEE Transaction on Information Theory.*

Yu Xiang and Young-Han Kim, "A few meta-theorems in network information theory," *IEEE Information Theory Workshop*, pp. 77–81, Hobart, Australia, November 2014.

Yu Xiang and Young-Han Kim, "Gaussian channel with noisy feedback and peak energy constraint," *IEEE Transaction on Information Theory,* vol.59, no.8, pp.4746–4756, August 2013.

Yu Xiang and Young-Han Kim, "Interactive hypothesis testing against independence," *IEEE International Symposium on Information Theory*, pp. 2840–2844, Istanbul, Turkey, July 2013.

Yu Xiang and Young-Han Kim, "Interactive hypothesis testing with communication constraints," *Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL*, pp. 1065–1072, Monticello, IL, October 2012.

Yu Xiang, Lele Wang, and Young-Han Kim, "Information flooding," *Annual Allerton Conference on Communication, Control, and Computing,* pp. 45–51, Monticello, IL, September 2011.

Yu Xiang and Young-Han Kim, "On the AWGN channel with noisy feedback and peak energy constraint," *IEEE International Symposium on Information Theory*, pp. 256–259, Austin, TX, June 2010.

ABSTRACT OF THE DISSERTATION

Interactive Schemes in Information Theory and Statistics

by

Yu Xiang

Doctor of Philosophy in Electrical Engineering
(Communication Theory and Systems)

University of California, San Diego, 2015

Professor Young-Han Kim, Chair

The ever-growing Internet, among other things, provides abundant evidence that we are living in a world of interaction. It is thus of great importance to understand the benefits of interaction in our lives. In this thesis, we investigate the role of interaction in information theory and statistics via three concrete problems: distributed inference, point-to-point channel communication, and communication over networks. First, we consider a classical hypothesis testing problem in a distributed setting, where communication constraints are present. More specifically, two distributed agents, having only partial access to some random data set, are required to perform a hypothesis test regarding the joint data distribution by communicating their observations with each other. The goal is to characterize the

optimal tradeoff between the testing performance and the communication budget. We formulate an interactive version of this problem, where the two agents are allowed to communicate in multiple rounds before making a decision. Interestingly, the testing performance can be strictly improved given the same communication budget. Moreover, we study a sequential version of the interactive hypothesis problem which further improves the testing performance. Second, we investigate the role of interaction in the reliability of communication by studying the optimal coding over the Gaussian channel with noisy Gaussian feedback. While it is well known that the reliability of communication can be strictly improved through noiseless feedback, the theoretical understanding of the benefits of noisy feedback is yet far from being complete. We propose two coding schemes that enable strict improvement of the reliability of communication when the noise power in feedback channel is smaller than a certain threshold. Finally, we study the impact of interaction on relay networks. In particular, we focus on a network communication problem, in which one nodes wishes to broadcast a common message to all the other nodes in the network. Typically, relaying schemes are non-interactive, which can be improved by two-round interaction schemes. We investigate this problem beyond the two-round case and demonstrate via a simple example that infinite rounds of interaction can further improve upon finite rounds interactive schemes.

Chapter 1

Introduction

With the rapid advances of technologies, ranging from wireless communication and embedded systems to the Internet and microelectromechanical systems (MEMS), a vast majority of humans and devices are being connected with each other. This "Internet of things (IoT)", not only enables information being collected at a speed and scale unimaginable before, but more importantly, creates unprecedented opportunities for interaction among all the participants. Considering the fact that billions of people interact through social media like Youtube and Facebook everyday and the number of mobile-connected devices exceeded the world's population, we are indeed living in a world of interaction. It is thus of great importance to explore the impact of interaction so that we can leverage it to facilitate our understanding of the world.

Even though interactive models are more relevant and natural to real world problems than non-interactive models, they have not been the main focus in both information theory and statistics for different reasons. In information theory, the main focus has been traditionally on one-way setups starting from Shannon's seminal work on the point-to-point channel to many classical network information theory problems. The lack of investigation is also partly due to the technical difficulties arising from the interactive models. In statistics, one of the key assumptions of most classical problems is that one has access to all available data, which leaves little room for interaction between multiple parties. However, interaction comes into play for distributed statistical problems, which are becoming more and more

common in the real world.

In this thesis, we investigate the role of interaction in both information theory and statistics through three setups: distributed inference, point-to-point channel communication, and communication over networks.

In Chapter 3, we consider the impact of interaction on distributed hypothesis testing. With advances in technology, a massive amount of data are collected and stored daily. However, the sheer amount of data stored at each data center makes collaborative data processing and analysis across distributed data centers a challenging task. These new challenges motivates us to revisit the classical problem of (one-way) distributed hypothesis testing with communication constraints, where the goal is to characterize the optimal tradeoff between the testing performance (type-II error exponent) and the communication budget (rate). In the hope of improving the testing performance given the same communication budget, we formulate the interactive version of this problem, in which the distributed agents are allowed to communicate with each other in multiple rounds to perform the hypothesis test. For testing whether the observed data at the distributed agents are generated independently or not, we establish a computable characterization of the optimal tradeoff, which generalizes the one-way case. It turns out that, given the communication budget, interaction enables strictly improvement of the testing performance over the one-way case. Moreover, based on the results of our interactive hypothesis testing problem, we show that when the distributed agents are allowed to stop early and make their decisions, the testing performance can be further improved.

In Chapter 4, we study the impact of interaction on the reliability of communication. In particular, we focus on a point-to-point Gaussian channel with feedback. Instead of assuming that the sender has full access to what the receiver observed through a noiseless feedback channel, we consider that the feedback channel is corrupted by Gaussian noise. Unlike the noiseless feedback case, the noise in the feedback channel prevents the sender and the receiver from fully cooperating with each other. The goal is to leverage the noisy feedback channel to improve the reliability of communication. Our main contributions are two coding schemes that

enable strict improvement of the reliability when the noise power in the feedback channel is less than a certain threshold. Both schemes have their own strengths and outperform each other for different ranges of system parameters.

In Chapter 5, we investigate the role of interaction in relay networks. In particular, we study a broadcasting network model, in which one node broadcasts a common message to all the other nodes in the network. Existing relaying schemes including decode–forward, partial decode–forward, hash–forward, compute–forward, and compress–forward are all non-interactive. It is known that two-round interactive relaying can outperform all the non-interactive *–forward schemes. We investigate the benefits of interaction beyond the two-round case. Based on recent results on interactive computing, we demonstrate via a simple example that infinite rounds of interaction can further improve upon the existing finite round schemes.

Chapter 2

Preliminaries

2.1 Notation

We closely follow the notation in [EGK11].

Sets, Scalars, and Vectors

We use lower case letters $x, y, \ldots$ to denote constants and values of random variables. We use $x^j = (x1, x2, \ldots, xj)$ to denote a j-sequence/vector. Sometimes we write $\mathbf{x}, \mathbf{y}, \ldots$ for constant (column) vectors with specified dimension and x_j for the j-th component of $\mathbf{x}$. Let $x(i)$ be a vector indexed by time i and $x_j(i)$ be the j-th component of $x(i)$. The sequence of these vectors will be then written as $x^n = (x(1), x(2), \ldots, x(n))$. Calligraphic letters $\mathcal{X}, \mathcal{Y}, \ldots$ will be used for finite sets, and $|\mathcal{X}|$ denotes the cardinality of the finite set $\mathcal{X}$. The following notation for common sets will be used: $\mathbb{R}^d$ is the d-dimensional real Euclidean space and $\mathbb{C}^d$ is the d-dimensional complex Euclidean space. For a pair of integers $i \leq j$, we define $[i : j] = i, i + 1, \ldots, j$. For a pair of real numbers $b > a$, $[a, b]$ denotes a continuous interval.

Random Variables and Vectors

We use upper case letters $X, Y, \ldots$ to denote random variables. The random variables may take values from finite sets $X, Y, \ldots$ or from the real line $\mathbb{R}$, or from the complex plane $\mathbb{C}$. The probability of the event $\{X \in \mathcal{A}\}$ is denoted by $\mathsf{P}\{X \in \mathcal{A}\}$. In accordance with the notation for constant vectors, we use the notation $X_j = (X_1, X_2, \ldots, X_j)$ to denote a j-sequence/vector of random variables.

The subset of random variables with indices from $\mathcal{J} \subset [1:n]$ is denoted by $X(\mathcal{J}) = (X_j : j \in \mathcal{J})$. Similarly, given k random vectors $(X_1^n, X_2^n, \ldots, X_n^k)$, $X^n(\mathcal{J}) = (X_j^n : j \in \mathcal{J}) = (X_1(\mathcal{J}), X_2(\mathcal{J}), \ldots, X_n(\mathcal{J}))$.

Information Measures

We use $H(X)$ to denote the entropy of a discrete random variable X, and $h(X)$ to denote the differential entropy if X is continuous. The mutual information between two random variables X and Y is denoted by $I(X;Y)$. The relative entropy (KullbackLeibler divergence) between two probability distributions P and Q is denoted by $D(P||Q)$.

In particular, for $X \sim p(x)$ and $\epsilon \in (0,1)$, we define the set of ϵ-typical n-sequences x^n (or the typical set in short) [OR01] as

$$\mathcal{T}_\epsilon^{(n)}(X) = \{x^n : |\#\{i : x_i = x\}/n - p(x)| \leq \epsilon p(x) \text{ for all } x \in \mathcal{X}\}.$$

We say that $X \to Y \to Z$ form a Markov chain if $p(x,y,z) = p(x)p(y|x)p(z|y)$, that is, X and Z are conditionally independent of each other given Y.

2.2 Hypothesis Testing

In this section, we briefly review some basic results on the hypothesis testing problem. In particular, we focus on the simple vs. simple hypothesis testing problem.

Suppose that one observes a sequence of n random variables X^n=(X_1, X_2,..., X_n) that are independently and identically distributed (i.i.d.) according to some unknown distribution, which has two possibilities

$$H_0 : X \sim P(x)$$
$$H_1 : X \sim Q(x),$$

where H_0 denotes the null hypothesis and H_1 denotes the alternative hypothesis. Based on X^n, one can make a decision $\hat{h}(X^n) \in \{0,1\}$, where 0 correspond to H_0 is true and 1 to H_1 is true. Let $\mathcal{A}_n = \{x^n : \hat{h}(x^n) = 0\}$ denote the set of x^n that H_0 is accepted. We call $\mathcal{A}_n$ as acceptance region and $\mathcal{A}_n^c$ as rejection region.

The two hypotheses incur two types of errors as follows. The type-I error, denoted $\alpha_n = \mathsf{P}_0(\mathcal{A}_n^c)$, which is the probability that the H_1 is declared to be true when H_0 is true. Similarly, we have the type-II error, denoted $\beta_n = \mathsf{P}_1(\mathcal{A}_n)$, which is the probability that the H_0 is declared to be true when H_1 is true.

In the Neyman–Pearson framework, one tries to minimize the type-II error given that the type-I error is smaller than some small constant ϵ. The famous Neyman–Pearson lemma says that the optimal test is the likelihood ratio test. Instead of minimizing the type-II error for any finite n, we consider maximizing the type-II error exponent when the n grows to infinity.

Lemma 2.1 (Stein's lemma)**.** *For some $0 < \epsilon < 1$, let*

$$\beta_n^*(\epsilon) := \min_{\mathcal{A}_n : \alpha_n \le \epsilon} \beta_n.$$

Then we have

$$\lim_{n\to\infty} -\frac{1}{n} \log \beta_n^*(\epsilon) = D(P||Q).$$

In the following, we provide an alternative proof of Stein's lemma using the (strong) typical set instead of using the relative entropy typical set as in [CT06]. When it is clear from the context, we will use $\mathcal{T}_\epsilon^{(n)}$ instead of $\mathcal{T}_\epsilon^{(n)}(X)$.

Lemma 2.2. *Let $P(x^n) = \prod_{i=1}^n P(x_i)$. Then for each $x^n \in \mathcal{T}_\epsilon^{(n)}$ and any other distribution Q on $\mathcal{X}$,*

$$D(P||Q) - \delta(\epsilon) \le \frac{1}{n} \log \frac{P(x^n)}{Q(x^n)} \le D(P||Q) + \delta(\epsilon),$$

for some $\delta(\epsilon)$ such that $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Proof. If $x^n \in \mathcal{T}_\epsilon^{(n)}$, we have by definition of $\mathcal{T}_\epsilon^{(n)}$ that

$$(1-\epsilon)P(x) \le \pi(x|x^n) \le (1+\epsilon)P(x).$$

Let $g(x)$ be any function on $\mathcal{X}$, define the following two sets:

$$\mathcal{X}^+ := \{x : g(x) \ge 0\} \text{ and } \mathcal{X}^- := \{x : g(x) < 0\}.$$

Thus we have,

$$\sum_{x\in\mathcal{X}^+} (1-\epsilon)P(x)g(x) \le \sum_{x\in\mathcal{X}^+} \pi(x|x^n)g(x) \le \sum_{x\in\mathcal{X}^+} (1+\epsilon)P(x)g(x) \tag{2.1}$$

and

$$\sum_{x\in\mathcal{X}^-}(1+\epsilon)P(x)g(x) \le \sum_{x\in\mathcal{X}^-}\pi(x|x^n)g(x) \le \sum_{x\in\mathcal{X}^-}(1-\epsilon)P(x)g(x). \tag{2.2}$$

Summing up (2.1) and (2.2), we have

$$\mathsf{E}(g(X)) - \delta(\epsilon) \le \frac{1}{n}\sum_{i=1}^{n} g(x_i) \le \mathsf{E}(g(X)) + \delta(\epsilon),$$

where $\delta(\epsilon) = \epsilon(\sum_{x\in\mathcal{X}^+} P(x)g(x) - \sum_{x\in\mathcal{X}^-} P(x)g(x))$. Choose

$$g(x) = \log(P(x)/Q(x))$$

and since $P(x^n) = \prod_{i=1}^n P(x_i)$ and $Q(x^n) = \prod_{i=1}^n Q(x_i)$, we have

$$D(P||Q) - \delta(\epsilon) \le \frac{1}{n}\log\frac{P(x^n)}{Q(x^n)} \le D(P||Q) + \delta(\epsilon).$$

□

Based on Lemma 2.2, we show the following properties of the $\mathcal{T}_\epsilon^{(n)}$:

- Let $P(x^n) = \prod_{i=1}^n P(x_i)$. Then for each $x^n \in \mathcal{T}_\epsilon^{(n)}$,

$$P(x^n)2^{-n(D(P||Q)+\delta(\epsilon))} \le Q(x^n) \le P(x^n)2^{-n(D(P||Q)-\delta(\epsilon))}.$$

- $P(\mathcal{T}_\epsilon^{(n)}) > 1-\epsilon$ for n sufficiently large.
- $Q(\mathcal{T}_\epsilon^{(n)}) \le 2^{-n(D(P||Q)+\delta(\epsilon))}$.
- $Q(\mathcal{T}_\epsilon^{(n)}) \ge (1-\epsilon)2^{-n(D(P||Q)+\delta(\epsilon))}$ for n sufficiently large.

The first two properties are trivial. The third one follows from

$$\begin{aligned} Q(\mathcal{T}_\epsilon^{(n)}) &= \sum_{x^n\in\mathcal{T}_\epsilon^{(n)}} Q(x^n) \\ &\le \sum_{x^n\in\mathcal{T}_\epsilon^{(n)}} P(x^n)2^{-n(D(P||Q)-\delta(\epsilon))} \\ &= 2^{-n(D(P||Q)-\delta(\epsilon))}P(\mathcal{T}_\epsilon^{(n)}) \\ &\le 2^{-n(D(P||Q)-\delta(\epsilon))}. \end{aligned}$$

The fourth one follows from

$$\begin{aligned} Q(\mathcal{T}_\epsilon^{(n)}) &= \sum_{x^n \in \mathcal{T}_\epsilon^{(n)}} Q(x^n) \\ &\geq \sum_{x^n \in \mathcal{T}_\epsilon^{(n)}} P(x^n) 2^{-n(D(P||Q)-\delta(\epsilon))} \\ &= 2^{-n(D(P||Q)-\delta(\epsilon))} P(\mathcal{T}_\epsilon^{(n)}) \\ &\geq (1-\epsilon) 2^{-n(D(P||Q)-\delta(\epsilon))}, \end{aligned}$$

for n sufficiently large.

Lemma 2.3. *Let $B_n \subset \mathcal{X}^n$ such that $P(B_n) > 1-\epsilon$. Then for any other distribution Q on $\mathcal{X}$ such that $D(P||Q) < \infty$, we have $Q(B_n) > (1-2\epsilon)2^{-n(D(P||Q)+\delta(\epsilon))}$ for n sufficiently large.*

Proof. Since $P(\mathcal{T}_\epsilon^{(n)}) > 1-\epsilon$ for n sufficiently large and $P(B_n) > 1-\epsilon$ by assumption, we have $P(B_n \cap \mathcal{T}_\epsilon^{(n)}) > 1-2\epsilon$ for n sufficiently large. Now

$$\begin{aligned} Q(B_n) &\geq Q(B_n \cap \mathcal{T}_\epsilon^{(n)}) \\ &= \sum_{x^n \in B_n \cap \mathcal{T}_\epsilon^{(n)}} Q(x^n) \\ &\geq \sum_{B_n \cap \mathcal{T}_\epsilon^{(n)}} P(x^n) 2^{-n(D(P||Q)+\delta(\epsilon))} \\ &= 2^{-n(D(P||Q)+\delta(\epsilon))} \sum_{B_n \cap \mathcal{T}_\epsilon^{(n)}} P(x^n) \\ &\geq (1-2\epsilon) 2^{-n(D(P||Q)+\delta(\epsilon))}. \end{aligned}$$

□

Now we prove Stein's lemma using the (strong) typical set instead of the relative entropy typical set.

Proof. For achievability, we choose $\mathcal{A}_n$ as $\mathcal{T}_\epsilon^{(n)}$. As proved before that $P(\mathcal{A}_n^c) < \epsilon$ for n sufficiently large. Also

$$\lim_{n\to\infty} -\frac{1}{n} \log Q(\mathcal{T}_\epsilon^{(n)}) \geq D(P||Q) - \delta(\epsilon).$$

To show the converse, consider any set $B_n \subset \mathcal{X}^n$ such that $P(B_n) > 1-\epsilon$, we have $Q(B_n) > (1-2\epsilon)2^{-n(D(P||Q)+\delta(\epsilon))}$ from lemma 2.3, and therefore

$$\lim_{n\to\infty} -\frac{1}{n}\log Q(B_n) < D(P||Q) + \delta(\epsilon).$$

□

2.3 Technical Lemmas

In this section, we introduce three technical lemmas without proof.

2.3.1 Covering Lemma

Lemma 2.4 ([EGK11]). *Let* $(U, X, \hat{X}) \sim p(u, x, \hat{x})$ *and* $\epsilon' < \epsilon$ *and* $(U^n, X^n) \sim p(u^n, x^n)$ *be* arbitrarily *distributed such that*

$$\lim_{n\to\infty} \mathsf{P}\{(U^n, X^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U, X)\} = 1,$$

and let $\hat{X}^n(m) \sim \prod_{i=1}^n, p_{\hat{X}|U}(\hat{x}_i|u_i), m \in \mathcal{A}$, *where* $|\mathcal{A}| \geq 2^{nR}$, *be conditionally independent of each other and of* X^n *given* U^n. *Then, there exists* $\delta(\epsilon) \to 0$ *as* $\epsilon \to 0$ *such that*

$$\lim_{n\to\infty} \mathsf{P}\{(U^n, X^n, \hat{X}^n(m)) \notin \mathcal{T}_{\epsilon}^{(n)} \text{ for all } m \in \mathcal{A}\} = 0,$$

if $R > I(X; \hat{X}|U) + \delta(\epsilon)$.

2.3.2 Markov Lemma

We present a version of Markov lemma.

Lemma 2.5 ([Tun78] and [EGK11]). *Suppose that* $X \to Y \to Z$ *from a Markov chain. Let* $(x^n, y^n) \in \mathcal{T}_{\epsilon'}^{(n)}(X, Y)$ *and* $Z^n \sim p(z^n|y^n)$, *where the conditional pmf* $p(z^n|y^n)$ *satifies the following conditions:*

$$\lim_{n\to\infty} \mathsf{P}\{(y^n, Z^n) \in \mathcal{T}_{\epsilon'}^{(n)}(Y, Z)\} = 1$$

and for every $z^n \in \mathcal{T}_{\epsilon'}^{(n)}(Z|y^n)$ *and* n *sufficiently large*

$$2^{-n(H(Z|Y)+\delta(\epsilon'))} \leq p(z^n|y^n) \leq 2^{-n(H(Z|Y)-\delta(\epsilon'))}$$

for some $\delta(\epsilon')$ *that tends to zero as* $\epsilon' \to 0$. *Then for some sufficiently small* $\epsilon' < \epsilon$,

$$\lim_{n\to\infty} \mathsf{P}\{(x^n, y^n, Z^n) \in \mathcal{T}_{\epsilon}^{(n)}(X, Y, Z)\} = 1.$$

2.3.3 Blowing-up Lemma

Let $x^n, y^n \in \mathcal{X}^n$ and $d(x^n, y^n)$ be their Hamming distance. For $\mathcal{A} \subseteq \mathcal{X}^n$, define the l-neighborhood of $\mathcal{A}$ as

$$\Gamma_l(\mathcal{A}) = \{x^n : \min_{y^n \in \mathcal{A}} d(x^n, y^n) \leq l\}.$$

Lemma 2.6. *Let* $X^n \sim P_{X^n} = \prod_{i=1}^n P_{X_i}$ *and* $\epsilon_n \to 0$ *as* $n \to \infty$. *There exist* δ_n *and* η_n *(both go to* 0 *as* $n \to \infty$*) such that if* $P_{X^n}(\mathcal{A}) \geq 2^{-n\epsilon_n}$, *then*

$$P_{X^n}(\Gamma_{n\delta_n}(\mathcal{A})) \geq 1 - \eta_n.$$

Chapter 3

Interactive Hypothesis Testing with Communication Constraints

In this chapter, a hypothesis testing problem with communication constraints is studied, in which two nodes separately observe one of two correlated sources and interactly communicate with each other in q rounds to decide between two hypotheses on the joint distribution of the sources. The optimal tradeoff between the communication rates in q rounds interaction and the testing performance is measured by the type II error exponent such that the type I error probability asymptotically vanishes. When testing against independence, that is, the joint distribution of the sources under the alternative hypothesis is the product of the marginal distributions under the null hypothesis, a computable characterization of the optimal tradeoff is obtained. An example is provided that shows that a two-way test strictly outperforms the optimal one-way test and thus that interaction helps for hypothesis testing.

3.1 Introduction

Berger [Ber79], in an inspiring attempt at combining information theory and statistical inference, formulated the problem of hypothesis testing with communication constraints as depicted in Fig. 3.1. Let $(X_1^n, X_2^n) \sim \prod_{i=1}^n p_{X_1,X_2}(x_{1i}, x_{2i})$ be a pair of independent and identically distributed (i.i.d.) n-sequences generated

by a two-component discrete memoryless source (2-DMS) (X_1, X_2). Suppose that there are two hypotheses on the joint distribution of (X_1, X_2), namely,

$$H_0 : (X_1, X_2) \sim p_0(x_1, x_2),$$
$$H_1 : (X_1, X_2) \sim p_1(x_1, x_2).$$

In order to decide which hypothesis is true, nodes 1 and 2 that observe X_1^n and X_2^n, respectively, compress their observed sequences into indices of rates R_1 and R_2, and communicate them over noiseless links to node 3, which then makes a decision $\hat{H} \in \{H_0, H_1\}$ based on the received compression indices. What is the impact of communication constraints on the performance of hypothesis testing? To answer this question, Berger [Ber79] studied the optimal tradeoff between the communication rates and the testing performance that is measured by the exponent of the type II error probability such that the type I error probability is upper bounded by a given $\epsilon < 1$. Despite many natural applications, however, theoretical understanding of this problem is far from complete and a simple characterization of this rate–exponent tradeoff remains open in general.

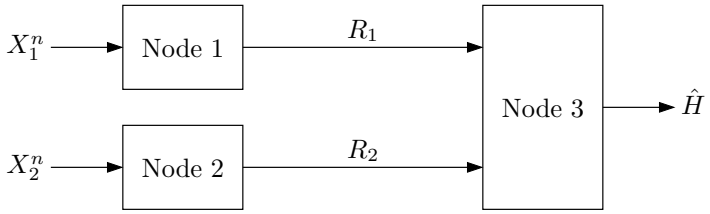


Figure 3.1: Multiterminal hypothesis testing with communication constraints.

In their celebrated paper [AC86], Ahlswede and Csiszár studied the special case in which the sequence X_2^n is fully available at the destination node, i.e., $R_2 = \infty$. They established single-letter inner and outer bounds on the optimal tradeoff and showed that these bounds are tight for testing *against independence*, i.e., the alternative hypothesis H_1 is $p_1(x_1, x_2) = p_0(x_1)p_0(x_2)$. Later, Han [Han87] and Shimokawa, Han, and Amari [SHA94] provided a new coding scheme that improves upon the Ahlswede and Csiszar inner bound for the general hypothesis testing problem. The Shimokawa–Han–Amari scheme is similar to the Berger–Tung scheme [Tun78], [Ber78] for the distributed lossy source coding problem,

where node 1 and node 2 perform joint typicality encoding followed by binning. A more comprehensive survey on the earlier literature can be found in [HA98]. Several variations of this setup have been studied, including successive refinement hypothesis testing [TC08] and testing against conditional independence [RW12].

This chapter studies an interactive version of hypothesis testing with communication constraints. Two nodes communicate with each other in *q rounds* through noiseless links and one of the nodes is to perform hypothesis testing at the end of interactive communication. For the special case of hypothesis testing *against independence*, we establish a single-letter characterization of the optimal tradeoff between the communication rates and the type II error probability when the type I error probability is arbitrarily small. Part of this chapter has been reported in [XK12] and [XK13b].

The rest of the chapter is organized as follows. In Section II, we review the problem of one-way hypothesis testing with communication constraints. In Section III, we formulate the problem of interactive hypothesis testing with communication constraints and present our main theorem. In Section IV, we compare the interactive hypothesis testing problem with the interactive lossy source coding problem by Kaspi [Kas85].

3.2 One-way Case

As before, let $(X_1^n, X_2^n) \sim \prod_{i=1}^n p_{X_1,X_2}(x_{1i}, x_{2i})$ be a pair of i.i.d. sequences generated by a 2-DMS (X_1, X_2) and consider hypothesis testing against independence

$$\begin{aligned} H_0 &: (X_1, X_2) \sim p_0(x_1, x_2), \\ H_1 &: (X_1, X_2) \sim p_1(x_1, x_2) = p_0(x_1)p_0(x_2). \end{aligned}$$

Here $p_0(x_1)$ and $p_0(x_2)$ are marginal distributions of $p_0(x_1, x_2)$. We consider the special case of the problem depicted in Fig. 3.1, in which $R_2 = \infty$; see Fig. 3.2.

A $(2^{nR_1}, n)$ hypothesis test consists of

- an encoder that assigns an index $m_1(x_1^n) \in [1 : 2^{nR_1}]$ to each sequence $x_1^n \in$

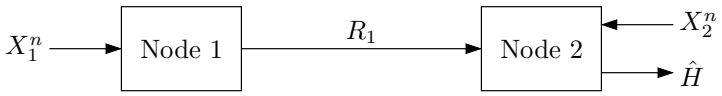


Figure 3.2: One-way hypothesis testing with communication constraint.

$\mathcal{X}_1^n$, and

- a tester that assigns $\hat{h}(m_1, x_2^n) \in \{H_0, H_1\}$ to each $(m_1, x_2^n) \in [1 : 2^{nR_1}] \times \mathcal{X}_2^n$.

The acceptance region is defined as

$$\mathcal{A}_n := \{(m_1, x_2^n) \in [1 : 2^{nR_1}] \times \mathcal{X}_2^n : \hat{h}(m_1, x_2^n) = H_0\}.$$

Then the type I error probability is

$$\mathsf{P}_0(\mathcal{A}_n^c) = \sum_{(x_1^n, x_2^n):(m_1(x_1^n), x_2^n) \in \mathcal{A}_n^c} p_0(x_1^n, x_2^n)$$

and the type II error probability is

$$\mathsf{P}_1(\mathcal{A}_n) = \sum_{(x_1^n, x_2^n):(m_1(x_1^n), x_2^n) \in \mathcal{A}_n} p_1(x_1^n, x_2^n).$$

For $\epsilon \in (0, 1)$, define the optimal type II error probability as

$$\beta_n^*(R_1, \epsilon) := \min \mathsf{P}_1(\mathcal{A}_n),$$

where the minimum is over all $(2^{nR_1}, n)$ tests such that $\mathsf{P}_0(\mathcal{A}_n^c) \leq \epsilon$. Further define the optimal type II error exponent as

$$\theta_1(R_1, \epsilon) := \lim_{n \to \infty} -\frac{1}{n} \log \beta_n^*(R_1, \epsilon).$$

Theorem 3.1 (Ahlswede and Csiszár [AC86])**.** *For every $\epsilon \in (0, 1)$,*

$$\theta_1(R_1, \epsilon) = \max_{p(u_1|x_1):R_1 \geq I(U_1;X_1)} I(U_1; X_2), \tag{3.1}$$

where the cardinality bound for U_1 is $|\mathcal{U}_1| \leq |\mathcal{X}_1| + 1$.

We illustrate the theorem with the following.

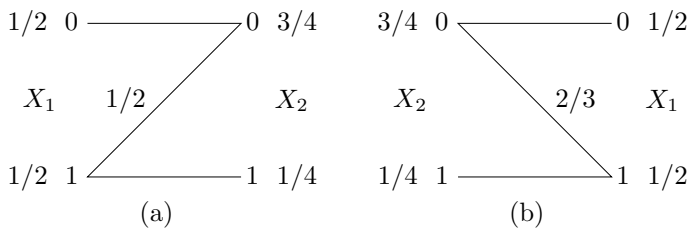


Figure 3.3: (a) Forward Z binary sources and (b) backward Z binary sources.

Example 3.1. *Consider the following* forward *Z binary sources* (X_1, X_2) *depicted in Fig. 4.4(a), where* X_2 *is the output of* X_1 *through a Z channel and*

$$p_{X_1,X_2}(0,0) = 1/2, \quad p_{X_1,X_2}(0,1) = 0,$$
$$p_{X_1,X_2}(1,0) = 1/4, \quad p_{X_1,X_2}(1,1) = 1/4.$$

We now apply Theorem 3.1 and evaluate the optimal type II error exponent in (3.1). *Since* $|\mathcal{U}_1| \leq |\mathcal{X}_1| + 1 = 3$, *we can optimize over all conditional pmfs* $p(u_1|x_1)$ *of the form in Fig. 3.4.*

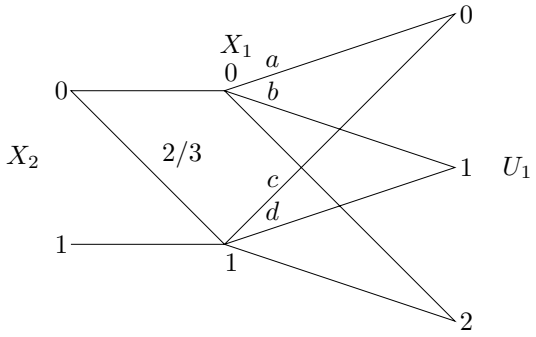


Figure 3.4: Conditional pmf $p(u_1|x_1)$.

Then we have

$$\theta_1(R_1, \epsilon) = \max\left(H_1 - \frac{1}{6}H_2 - \frac{3}{4}H_3\right),$$

where the maximum is over all (a, b, c, d) *such that*

$$R_1 \geq H_1 - \frac{1}{2}H_4 - \frac{1}{2}H_2$$

and H_1 through H_4 are defined as

$$\begin{aligned}
H_1 &:= H\left(\frac{a+c}{2}, \frac{b+d}{2}, \frac{2-a-b-c-d}{2}\right),\\
H_2 &:= H(c, d, 1-c-d),\\
H_3 &:= H\left(\frac{a+2c}{3}, \frac{b+2d}{3}, \frac{3-a-b-2c-2d}{3}\right),\\
H_4 &:= H(a, b, 1-a-b).
\end{aligned}$$

For example, when $R_1 = 1/2$, we have $\theta_1(R_1, \epsilon) \approx 0.1878$. The entire curve of the optimal type II error exponent, denoted by $\theta_1^{\rightarrow}(R_1, \epsilon)$, is plotted in Fig. 4.4(b).

Example 3.2. *Now consider the following* backward *Z binary sources (X_1, X_2) depicted in Fig. 3.3(b), where X_1 is the output of X_2 through an inverted Z channel. Since $|\mathcal{U}_2| \leq 3$, we can again optimize over all conditional pmfs $p(u_2|x_2)$ of the form in Fig. 3.5, which yields*

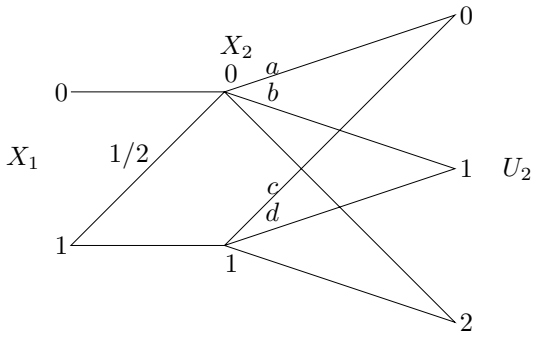


Figure 3.5: Conditional pmf $p(u_2|x_2)$.

$$\theta_1(R_1, \epsilon) = \max\left(H_1 - \frac{1}{2}H_2 - \frac{1}{2}H_3\right),$$

where the maximum is over all (a, b, c, d) such that

$$R_1 \geq H_1 - \frac{3}{4}H_2 - \frac{1}{4}H_4$$

and

$$H_1 := H\left(\frac{3a+c}{4}, \frac{3b+d}{4}, \frac{4-3a-3b-c-d}{4}\right),$$
$$H_2 := H(a, b, 1-a-b),$$
$$H_3 := H\left(\frac{a+c}{2}, \frac{b+d}{2}, \frac{2-a-b-c-d}{2}\right),$$
$$H_4 := H(c, d, 1-c-d).$$

The entire curve of the optimal type II error exponent, denoted by $\theta_1^{\leftarrow}(R_1, \epsilon)$ *this time, is plotted in Fig. 4.4(b). Observe that for every* $R_1 \in (0, 1)$*,*

$$\theta_1^{\leftarrow}(R_1, \epsilon) > \theta_1^{\rightarrow}(R_1, \epsilon). \tag{3.2}$$

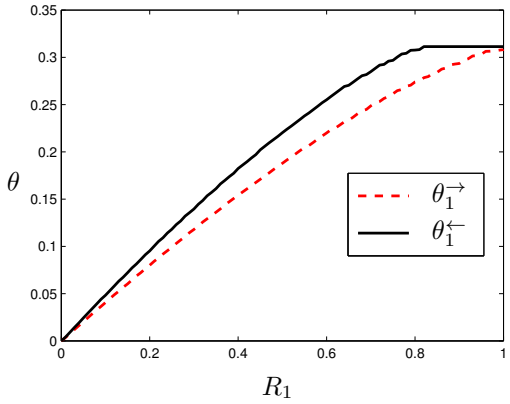


Figure 3.6: The solid black curve corresponds to $\theta_1^{\leftarrow}(R_1, \epsilon)$ and the dotted red curve corresponds to $\theta_1^{\rightarrow}(R_1, \epsilon)$.

3.3 Interactive Case

Suppose now that instead of making an immediate decision based on one round of communication, the two nodes can interactively communicate over a noiseless bidirectional link before one of the nodes performs hypothesis testing. We wish to characterize the optimal tradeoff between the communication rates and the performance of hypothesis testing.

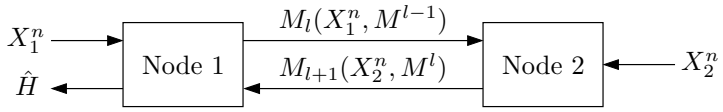


Figure 3.7: Interactive hypothesis testing with communication constraints.

As before, we consider testing against independence. Assume without loss of generality that node 1 sends the first index and that the number of rounds of communication q is even. A $(2^{nR_1}, \ldots, 2^{nR_q}, n)$ hypothesis test consists of

- two encoders, one for each node, where in round $l_j \in \{j, j+2, \ldots, q-2+j\}$, encoder $j \in \{1,2\}$ sends an index $m_{l_j}(x_j^n, m^{l_j-1}) \in [1 : 2^{nr_{l_j}}]$, that is, a function of its sequence and all previously transmitted indices, and
- a tester that assigns $\hat{h}(m^q, x_1^n) \in \{H_0, H_1\}$ to each $(m^q, x_1^n) \in [1 : 2^{nR_1}] \times \cdots \times [1 : 2^{nR_q}] \times \mathcal{X}_1^n$.

The type I and II error probabilities are defined similarly as in the one-way case. In particular, the optimal type II error exponent is

$$\theta_q(R_1, \ldots, R_q, \epsilon) := \lim_{n\to\infty} -\frac{1}{n} \log \beta_n^*(R_1, \ldots, R_q, \epsilon).$$

We establish the optimal tradeoff between the rate constraints and the testing performance by characterizing $\theta_q(R_1, \ldots, R_q, \epsilon)$ in the limit.

Theorem 3.2.

$$\lim_{\epsilon\to 0} \theta_q(R_1, \ldots, R_q, \epsilon) = \max \sum_{l=1}^{q} I(U_l; X_{j_{l+1}} | U^{l-1}), \tag{3.3}$$

where the maximum is over all $\prod_{l=1}^q p(u_l|u^{l-1}, x_{j_l})$ *with* $|\mathcal{U}_l| \le |\mathcal{X}_{j_l}| \cdot \prod_{j=1}^{l-1} |\mathcal{U}_j| + 1$ *such that*

$$R_l \ge I(U_l; X_{j_l} | U^{l-1})$$

for $l \in [1 : q]$ *and* $j_l = 1$ *if* l *is odd and* $j_l = 2$ *if* l *is even.*

Remark 3.1. *By setting* $U_l = \emptyset$ *and* $R_l = 0$ *for* $l = 2, \ldots, q$*, Theorem 3.2 recovers the optimal one-way type II error exponent in Theorem 3.1.*

Remark 3.2. *When $q = 2$ we have the following,*

$$\lim_{\epsilon \to 0} \theta_2(R_1, R_2, \epsilon) = \max\big(I(U_1; X_2) + I(U_2; X_1|U_1)\big),$$

where the maximum is over all $p(u_1|x_1)p(u_2|u_1, x_2)$ with $|\mathcal{U}_1| \leq |\mathcal{X}_1| + 1$ and $|\mathcal{U}_2| \leq |\mathcal{X}_2| \cdot |\mathcal{U}_1| + 1$ such that

$$\begin{aligned} R_1 &\geq I(U_1; X_1), \\ R_2 &\geq I(U_2; X_2|U_1). \end{aligned}$$

Remark 3.3. *Let $\theta_q := \theta_q(R_1, \ldots, R_q, 0+)$ for simplicity, we can express the optimal tradeoff between communication constraints and the type II error exponent by the* rate–exponent region *that consists of all rate–exponent tuples $(R_1, \ldots, R_q, \theta)$ such that*

$$\begin{aligned} R_l &\geq I(U_l; X_{j_l}|U^{l-1}), \quad l \in [1 : q], \\ \theta_q &\leq \sum_{l=1}^{q} I(U_l; X_{j_l}|U^{l-1}) \end{aligned}$$

for some pmfs $\prod_{l=1}^{q} p(u_l|u^{l-1}, x_{j_l})$.

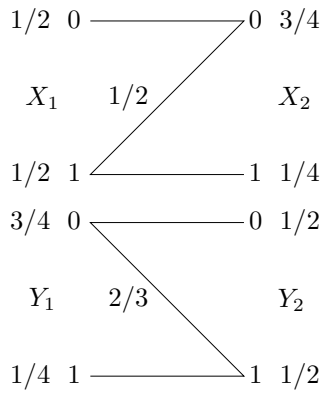


Figure 3.8: Double Z binary sources.

Example 3.3 (Interaction helps)**.** *This example is motivated by [GA10]. We revisit the Z binary sources in Examples 1 and 2. Recall that $\theta_1^{\rightarrow}(R_1, \epsilon)$ and $\theta_1^{\leftarrow}(R_1, \epsilon)$ denote the optimal type II error exponents for the forward and backward Z binary*

sources, respectively. Now consider the following double Z binary sources *as depicted in Fig. 3.8, where* (X_1, X_2) *is independent of* (Y_1, Y_2)*. Let*

$$\theta_2(R, \epsilon) := \max_{R_1, R_2 : R_1 + R_2 = R} \theta_2(R_1, R_2, \epsilon).$$

It can be easily verified that if $R \in (R^*, 2R^*)$*, where* $R^* = \min\{R : \theta^{\leftarrow}(R, \epsilon) = I(X_1; X_2) = 0.3113\}$*, then*

$$\theta_2(R, \epsilon) \geq 2\theta^{\leftarrow}(R/2, \epsilon),$$

while

$$\begin{aligned} \theta_1(R, \epsilon) &= \theta^{\leftarrow}(R^*, \epsilon) + \theta^{\rightarrow}(R - R^*, \epsilon) \\ &\overset{(a)}{<} \theta^{\leftarrow}(R^*, \epsilon) + \theta^{\leftarrow}(R - R^*, \epsilon) \\ &\overset{(b)}{\leq} 2\theta^{\leftarrow}(R/2, \epsilon), \end{aligned}$$

where (a) *follows by* (3.2) *and* (b) *follows by the concavity of* $\theta_1^{\leftarrow}(R)$ *over* $[0, R^*]$ *(see, for example, [AC86, Lemma 1]). For example, when* $R = 3/2$*, we have* $\theta_1(3/2, \epsilon) \approx 0.5548$ *and* $\theta_2(3/2, \epsilon) \geq 0.5934$*. Thus there is strict improvement by using interaction.*

In the following two subsections, we prove Theorem 2 by establishing achievability and the weak converse.

3.3.1 Proof of Achievability

Codebook generation. Fix a conditional pmf $\prod_{l=1}^{q} p(u_l|u^{l-1}, x_{j_l})$ that attains the maximum in (3.3). Let and $p_0(u_l|u^{l-1}) = \sum_{x_{j_l}} p_0(x_{j_l}) p(u_l|u^{l-1}, x_{j_l})$. Randomly and independently generate 2^{nR_l} sequences $u_l^n(m_l|m^{l-1})$, $m_l \in [1 : 2^{nR_l}]$, each according to $\prod_{i=1}^{n} p_0(u_{li}|u_i^{l-1})$. These sequences constitute the codebook $\mathcal{C}$, which is revealed to both nodes.

Encoding for round l. Given a sequence $x_{j_l}^n$, node j_l finds an index m_l such that

$$(u_1^n(m_1), u_2^n(m_2|m_1), \ldots, u_l^n(m_l|m^{l-1}), x_{j_l}^n) \in \mathcal{T}_{\eta_l}^{(n)}.$$

If there is more than one such index, it sends the smallest one among them. If there is no such index, it selects an index from $[1 : 2^{nR_l}]$ uniformly at random.

Testing. Upon receiving m^q, node 1 sets the acceptance region $\mathcal{A}_n$ for H_0 to

$$\mathcal{A}_n = \{(m^q, x_1^n) : (u_1^n(m_1), u_2^n(m_2|m_1), \ldots, u_q^n(m_q|m^{q-1}), x_1^n) \in \mathcal{T}_\eta^{(n)}\},$$

where $\mathcal{T}_\eta^{(n)} = \mathcal{T}_\eta^{(n)}(U^q, X_1)$ is defined with respect to $p_0(x_1, x_2)$, $p(u_l|u^{l-1}, x_{j_l})$ for all l.

Analysis of two types of error. Let M_l denote the chosen indices at node j_l and $\eta_1 < \eta_2 < \cdots < \eta_q < \eta$. Node 1 chooses $\hat{H} \neq H_0$ iff one or more of the following events occur: For $l \in [1 : q]$,

$$\begin{aligned}\mathcal{E}_l = \big\{&(U_1^n(M_1), U_2^n(M_2|M_1), \ldots, U_l^n(M_l|M^{l-1}), X_{j_l}^n) \notin \mathcal{T}_{\eta_l}^{(n)} \\ &\qquad\qquad \text{for all } m_l \in [1 : 2^{nR_l})\big\}, \\ \tilde{\mathcal{E}} = \big\{&(U_1^n(M_1), U_2^n(M_2|M_1), \ldots, U_q^n(M_q|M^{q-1}), X_1^n) \notin \mathcal{T}_\eta^{(n)}\big\}.\end{aligned}$$

For the type I error probability, assume that H_0 is true. Then

$$\alpha_n = \mathsf{P}(\cup_{l=1}^q \mathcal{E}_l \cup \tilde{\mathcal{E}}) \leq \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) + \cdots + \mathsf{P}(\cap_{l=1}^q \mathcal{E}_l^c \cap \tilde{\mathcal{E}}).$$

We now bound each term. By the covering lemma [EGK11, Section 3.7], $\mathsf{P}(\mathcal{E}_1)$ tends to zero as $n \to \infty$ if $R_1 \geq I(U_1; X_1) + \delta(\eta_1)$. Now we bound the second term. Since $\eta_2 > \eta_1$, $\mathcal{E}_1^c = \{(U_1^n(M_1), X_1^n) \in \mathcal{T}_{\eta_1}^{(n)}\}$ and $X_2^n|\{U_1^n(M_1) = u_1^n, X_1^n = x_1^n\} \sim \prod_{i=1}^n p_0(x_{2i}|u_{1i}, x_{1i}) = \prod_{i=1}^n p_0(x_{2i}|x_{1i})$, by the conditional typicality lemma [EGK11, Section 2.5], then $\mathsf{P}\{(U_1^n(M_1), X_1^n, X_2^n) \in \mathcal{T}_{\eta_2}^{(n)}\}$ tends to zero as $n \to \infty$. Therefore, again by the covering lemma, $\mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2)$ tends to zero as $n \to \infty$ if $R_2 \geq I(U_2; X_2|U_1) + \delta(\eta_2)$. Similarly, we have $\mathsf{P}(\mathcal{E}_1^c \cap \cdots \mathcal{E}_{l-1}^c \cap \mathcal{E}_l)$ tends to zero as $n \to \infty$ if $R_l \geq I(U_l; X_{j_l}|U^{l-1}) + \delta(\eta_l)$ for $l \in [1 : q]$.

To bound the last term, we use a version of the Markov lemma [Tun78] in [EGK11, Section 12.1]. Let

$$(x_1^n, x_2^n, u_1^n, \cdots, u_{q-1}^n) \in \mathcal{T}_{\eta_q}^{(n)}$$

and consider

$$\begin{aligned}
&\mathsf{P}\{U_q^n(M_q|M^{q-1}) = u_q^n \,|\, X_1^n = x_1^n, U_{q-1}^n(M_{q-1}|M^{q-2}) = u_{q-1}^n, \ldots, U_1^n(M_1) = u_1^n, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad X_2^n = x_2^n\} \\
&= \mathsf{P}\{U_q^n(M_q|M^{q-1}) = u_q^n \,|\, U_{q-1}^n(M_{q-1}|M^{q-2}) = u_{q-1}^n, \ldots, U_1^n(M_1) = u_1^n, X_2^n = x_2^n\} \\
&= p(u_q^n|u_{q-1}^n, \ldots, u_1^n, x_2^n).
\end{aligned}$$

First note that by the covering lemma,

$$\begin{aligned}
&\mathsf{P}\{U_q^n(M_q|M^{q-1}) \in \mathcal{T}_{\eta_q}^{(n)}(U_q^n|u_{q-1}^n, \ldots, u_1^n, x_2^n) \,| \\
&\qquad\qquad U_{q-1}^n(M_{q-1}|M^{q-2}) = u_{q-1}^n, \ldots, U_1^n(M_1) = u_1^n, X_2^n = x_2^n\}
\end{aligned}$$

tends to one as $n \to \infty$, that is, $p(u_q^n|u_{q-1}^n, \ldots, u_1^n, x_2^n)$ satisfies the first condition in the Markov lemma. For the second condition, the following is proved in the Section 3.6.1.

Lemma 3.1. *For every $u_q^n \in \mathcal{T}_{\eta_q}^{(n)}(U_q^n|u_{q-1}^n, \ldots, u_1^n, x_2^n)$ and n sufficiently large,*

$$p(u_q^n|u_{q-1}^n, \ldots, u_1^n, x_2^n) \doteq 2^{-nH(U_q|U^{q-1}, X_2)}.$$

Hence, by the Markov lemma,

$$\begin{aligned}
&\lim_{n\to\infty} \mathsf{P}\{(x_1^n, x_2^n, u_1^n, \ldots, u_{q-1}^n, U_q^n(M_q|M^{q-1})) \in \mathcal{T}_\eta^{(n)} \,| \\
&\qquad\qquad X_1^n = x_1^n, X_2^n = x_2^n, U_1^n(M_1) = u_1^n, \ldots, U_{q-1}^n(M_{q-1}|M^{q-2}) = u_{q-1}^n\} = 0,
\end{aligned}$$

if $(u_1^n, \ldots, u_{q-1}^n, x_1^n, x_2^n) \in \mathcal{T}_{\eta_q}^{(n)}(U_1, \ldots, U_{q-1}, X_1, X_2)$ and $\eta_q < \eta$ is sufficiently small. Therefore we have,

$$\lim_{n\to\infty} \mathsf{P}(\cap_{l=1}^q \mathcal{E}_l^c \cap \tilde{\mathcal{E}}) = 0.$$

For the type II error probability, assume in this case that H_1 is true. Then

$$\beta_n = \mathsf{P}(\cap_{l=1}^q \mathcal{E}_l^c \cap \tilde{\mathcal{E}}^c) = \mathsf{P}(\mathcal{E}_1^c)\, \mathsf{P}(\mathcal{E}_2^c|\mathcal{E}_1^c) \cdots \mathsf{P}(\mathcal{E}_q^c| \cap_{l=1}^{q-1} \mathcal{E}_l^c)\, \mathsf{P}(\tilde{\mathcal{E}}^c| \cap_{l=1}^q \mathcal{E}_l^c). \tag{3.4}$$

We now bound each factor. By the covering lemma, $\mathsf{P}(\mathcal{E}_1^c)$ tends to one as $n \to \infty$ if $R_1 \geq I(U_1; X_1) + \delta(\eta_1)$. Let

$$\tilde{\mathcal{E}}_l := \{(U_1^n(M_1), \ldots, U_{l-1}^n(M_{l-1}|M^{l-2}), X_{jl}^n) \notin \mathcal{T}_{\eta_l}^{(n)}\}, \text{ for } l \in [2:q].$$

We have for $l \in [2:q]$ that

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_l^c | \cap_{k=1}^{l-1} \mathcal{E}_k^c) &= \mathsf{P}(\mathcal{E}_l^c \cap \tilde{\mathcal{E}}_l^{\,c} | \cap_{k=1}^{l-1} \mathcal{E}_k^c) + \mathsf{P}(\mathcal{E}_l^c \cap \tilde{\mathcal{E}}_l | \cap_{k=1}^{l-1} \mathcal{E}_k^c) \\
&= \mathsf{P}(\mathcal{E}_l^c \cap \tilde{\mathcal{E}}_l^{\,c} | \cap_{k=1}^{l-1} \mathcal{E}_k^c) \\
&= \mathsf{P}(\mathcal{E}_l^c | \tilde{\mathcal{E}}_l^{\,c} \cap \cap_{k=1}^{l-1} \mathcal{E}_k^c) \cdot \mathsf{P}(\tilde{\mathcal{E}}_l^{\,c} | \cap_{k=1}^{l-1} \mathcal{E}_k^c). \qquad (3.5)
\end{aligned}$$

By the covering lemma, the first term in (3.5) tends to one as $n \to \infty$ if $R_l \geq I(U_k; X_{j_l}|U^{l-1}) + \delta(\eta_l)$. To bound the second term, we first make the following observation and the proof can be found in Section 3.6.2.

Lemma 3.2. *If H_1 is true, we have*

$$p_1(u_{l-1}^n, x_{j_l}^n | u_1^n, u_2^n, \ldots, u_{l-2}^n) = p_1(u_{l-1}^n | u_1^n, u_2^n, \ldots, u_{l-2}^n) p_1(x_{j_l}^n | u_1^n, u_2^n, \ldots, u_{l-2}^n).$$

The second term in (3.5) can be upper bounded as follows,

$$\begin{aligned}
\mathsf{P}(\tilde{\mathcal{E}}_l^{\,c} | \cap_{k=1}^{l-1} \mathcal{E}_k^c) &= \sum_{(u_1^n, \ldots, u_{l-1}^n, x_{j_l}) \in \mathcal{T}_{\eta_l}^{(n)}} p_1(u_{l-1}^n | u_1^n, \ldots, u_{l-2}^n) p_1(u_1^n, \ldots, u_{l-2}^n, x_{j_l}^n) \\
&\leq 2^{n(H(U^{l-1}, X_{j_l}) + \delta(\eta_l))} \cdot 2^{-n(H(U_{l-1}|U^{l-2}) - \delta(\eta_{l-1}))} \cdot 2^{-n(H(U^{l-2}, X_{j_l}) - \delta(\eta_{l-1}))} \\
&= 2^{-n(I(U_{l-1}; X_{j_l}|U^{l-2}) - \delta(\eta_l))}.
\end{aligned}$$

The last factor in (3.4) can be upper bounded similarly and we have

$$\mathsf{P}(\tilde{\mathcal{E}}^c | \cap_{l=1}^{q} \mathcal{E}_l^c) \leq 2^{-n(I(U_q; X_1|U^{q-1}) - \delta(\eta_q))}.$$

In summary, the type I error probability averaged over all codebooks is upper bounded by η if $R_l \geq I(U_l; X_{j_l}|U^{l-1})$ for $l \in [1:q]$, while the type II error probability averaged over all codebooks is upper bounded by

$$2^{-n(\sum_{l=1}^{q} I(U_l; X_{j_{l+1}}|U^{l-1}) - \delta(\eta_q))}.$$

Therefore, there exists a codebook such that

$$\theta_q(R_1, \ldots, R_q, \epsilon) \geq \sum_{l=1}^{q} I(U_l; X_{j_{l+1}}|U^{l-1}).$$

This completes the achievability proof.

3.3.2 Proof of the Converse

Consider q is odd and let $j_l = 1$ if l is odd and $j_l = 2$ if l is even. Given a $(2^{nR_1}, \ldots, 2^{nR_q}, n)$ test characterized by the encoding functions m_l, $l = 1, \ldots, q$, and the acceptance region $\mathcal{A}_n$, we have by the data processing inequality for relative entropy that

$$D\big(p_0(x_1^n, m^q)||p_1(x_1^n, m^q)\big) \geq (1-\alpha)\log\frac{1-\alpha}{\beta} + \alpha\log\frac{\alpha}{1-\beta},$$

where $\alpha := \mathsf{P}_0(\mathcal{A}_n^c)$, $\beta := \mathsf{P}_1(\mathcal{A}_n)$,

$$\begin{aligned} p_0(x_1^n, m^q) &:= \sum_{x_2^n} p_0(x_1^n, x_2^n) \prod_{l=1}^{q} p(m_l|m^{l-1}, x_{j_l}^n) \\ &= \left(p_0(x_1^n) \prod_{\substack{l=1 \\ l \text{ odd}}}^{q} p(m_l|m^{l-1}, x_1^n)\right) \cdot \sum_{x_2^n} \left(p_0(x_2^n|x_1^n) \prod_{\substack{l=1 \\ l \text{ even}}}^{q} p(m_l|m^{l-1}, x_2^n)\right), \end{aligned}$$

and

$$\begin{aligned} p_1(x_1^n, m^q) &:= \sum_{x_2^n} p_0(x_1^n) p_0(x_2^n) \prod_{l=1}^{q} p(m_l|m^{l-1}, x_{j_l}^n) \\ &= \left(p_0(x_1^n) \prod_{\substack{l=1 \\ l \text{ odd}}}^{q} p(m_l|m^{l-1}, x_1^n)\right) \cdot \sum_{x_2^n} \left(p_0(x_2^n) \prod_{\substack{l=1 \\ l \text{ even}}}^{q} p(m_l|m^{l-1}, x_2^n)\right). \end{aligned}$$

Let $M_{l_j} = m_{l_j}(x_j^n, M^{l_j-1})$ in round $l_j \in \{j, j+2, \ldots, q-2+j\}$ for $j \in \{1, 2\}$. Then by the definition of $\beta_n^*(R_1, \ldots, R_q, \epsilon)$, we must have

$$\begin{aligned} H(M_l) &\leq nR_l \quad \text{for } l \in [1:q], \\ \alpha &\leq \epsilon, \\ \beta &\leq \beta_n^*(R_1, \ldots, R_q, \epsilon). \end{aligned}$$

Then,

$$\begin{aligned} (1-\alpha)\log\frac{1-\alpha}{\beta} + \alpha\log\frac{\alpha}{1-\beta} &= (1-\alpha)\log\frac{1}{\beta} + \alpha\log\frac{1}{1-\beta} - H(\alpha) \\ &\geq (1-\alpha)\log\frac{1}{\beta} - H(\alpha) \\ &\geq (1-\epsilon)\log\frac{1}{\beta} - H(\alpha). \end{aligned}$$

Thus we have the following multiletter expression upper bound as

$$\lim_{\epsilon\to 0}\theta(R_1,\ldots,R_q,\epsilon)\le\lim_{n\to\infty}\frac{1}{n}D\big(p_0(x_1^n,m^q)||p_1(x_1^n,m^q)\big).$$

It is easy to verify that

$$\prod_{\substack{l=1\\ l\text{ even}}}^{q}p_0(m_l|m^{l-1},x_1^n)=\sum_{x_2^n}\bigg(p_0(x_2^n|x_1^n)\prod_{\substack{l=1\\ l\text{ even}}}^{q}p(m_l|m^{l-1},x_2^n)\bigg),$$
$$\prod_{\substack{l=1\\ l\text{ even}}}^{q}p_1(m_l|m^{l-1},x_1^n)=\sum_{x_2^n}\bigg(p_0(x_2^n)\prod_{\substack{l=1\\ l\text{ even}}}^{q}p(m_l|m^{l-1},x_2^n)\bigg).$$

We now prove that

$$D\big(p_0(x_1^n,m^q)||p_1(x_1^n,m^q)\big)\le\sum_{l=1}^{q}I(M_l;X_{j_{l+1}}^n|M^{l-1}).\tag{3.6}$$

To show this, we expand the relative entropy term in (3.6) as follows.

$$\begin{aligned}
&D\big(p_0(x_1^n,m^q)||p_1(x_1^n,m^q)\big)\\
&=\sum_{x_1^n,m^q}p_0(x_1^n,m^q)\log\frac{p_0(m_q|m^{q-1},x_1^n)p_0(m_{q-2}|m^{q-3},x_1^n)\cdots p_0(m_1|x_1^n)}{p_1(m_q|m^{q-1})p_1(m_{q-2}|m^{q-3})\cdots p_1(m_1)}\\
&=\sum_{x_1^n,m^q}p_0(x_1^n,m^q)\log\bigg(\frac{p_0(m_q|m^{q-1},x_1^n)}{p_0(m_q|m^{q-1})}\frac{p_0(m_q|m^{q-1})}{p_1(m_q|m^{q-1})}\frac{p_0(m_{q-2}|m^{q-3},x_1^n)}{p_0(m_{q-2}|m^{q-3})}\\
&\qquad\qquad\cdot\frac{p_0(m_{q-2}|m^{q-3})}{p_1(m_{q-2}|m^{q-3})}\cdots\frac{p_0(m_1|x_1^n)}{p_0(m_1)}\frac{p_0(m_1)}{p_1(m_1)}\bigg)\\
&=\sum_{\substack{l=1\\ l\text{ even}}}^{q}I(M_l;X_1^n|M^{l-1})+\sum_{\substack{m^l\\ l\text{ even}}}p_0(m^l)\log\frac{p_0(m_l|m^{l-1})}{p_1(m_l|m^{l-1})},
\end{aligned}\tag{3.7}$$

The second term in (3.7) can be upper bounded as

$$\begin{aligned}
&\sum_{m^l}p_0(m^l)\log\frac{p_0(m_l|m^{l-1})}{p_1(m_l|m^{l-1})}\\
&=\sum_{m^{l-2}}p_0(m^{l-2})\sum_{m_l,m_{l-1}}p_0(m_l,m_{l-1}|m^{l-2})\log\frac{p_0(m_l|m^{l-1})}{p_1(m_l|m^{l-1})}\\
&=\sum_{m^{l-2}}p_0(m^{l-2})D\big(p_0(m_l|m^{l-1})p_0(m_{l-1}|m^{l-2})||p_1(m_l|m^{l-1})p_0(m_{l-1}|m^{l-2})\big)\\
&=\sum_{m^{l-2}}p_0(m^{l-2})D\big(p_0(m_{l-1}|m^{l-2})\sum_{x_2^n}p_0(m_l|m^{l-1},x_2^n)p_0(x_2^n|m^{l-1})\big|\big|p_0(m_{l-1}|m^{l-2})
\end{aligned}$$

$$\cdot \sum_{x_2^n} p_0(m_l|m^{l-1}, x_2^n) p_0(x_2^n|m^{l-2})\big)$$

$$\leq \sum_{m^{l-2}} p_0(m^{l-2}) D\big(p_0(m_{l-1}|m^{l-2}) p_0(m_l|m^{l-1}, x_2^n) p_0(x_2^n|m^{l-1}) \big\| p_0(m_{l-1}|m^{l-2})$$

$$\cdot\, p_0(m_l|m^{l-1}, x_2^n) p_0(x_2^n|m^{l-2})\big)$$

$$= I(M_{l-1}; X_2^n|M^{l-2}). \tag{3.8}$$

Thus we have established (3.6). To complete the proof, we single-letterize the upper bound in (3.6) as

$$\begin{aligned} I(M_l; X_1^n|M^{l-1}) &= \sum_{i=1}^n I(M_l; X_{1i}|M^{l-1}, X_1^{i-1}) \\ &= \sum_{i=1}^n I(M_l; X_{1i}|M^{l-1}, X_1^{i-1}, X_2^{i-1}) \\ &\quad + I(M_l; X_2^{i-1}|M^{l-1}, X_1^{i-1}) - I(M_l; X_2^{i-1}|M^{l-1}, X_1^i) \\ &\overset{(a)}{\leq} \sum_{i=1}^n I(M_l; X_{1i}|M^{l-1}, X_1^{i-1}, X_2^{i-1}), \end{aligned}$$

where (a) follows from

$$\begin{aligned} &I(M_l; X_2^{i-1}|M^{l-1}, X_1^{i-1}) - I(M_l; X_2^{i-1}|M^{l-1}, X_1^i) \\ &= H(X_2^{i-1}|M^{l-1}, X_1^{i-1}) - H(X_2^{i-1}|M^l, X_1^{i-1}) \\ &\quad - H(X_2^{i-1}|M^{l-1}, X_1^i) + H(X_2^{i-1}|M^l, X_1^i) \\ &= I(X_2^{i-1}; X_{1i}|M^{l-1}, X_1^{i-1}) - I(X_2^{i-1}; X_{1i}|M^l, X_1^{i-1}) \\ &\leq 0. \end{aligned}$$

To bound the rate constraints, consider for l even,

$$\begin{aligned} nR_l &\geq H(M_l) \\ &\geq I(M_l; X_1^n, X_2^n|M^{l-1}) \\ &= \sum_{i=1}^n I(M_l; X_{1i}, X_{2i}|M^{l-1}, X_1^{i-1}, X_2^{i-1}) \\ &\geq \sum_{i=1}^n I(M_l; X_{2i}|M^{l-1}, X_1^{i-1}, X_2^{i-1}). \end{aligned}$$

When $l > 1$ is odd, the rate constraints and the terms in (3.6) can be bounded similarly. The case of $l = 1$ needs to be considered separately and it can be easily verified that

$$nR_1 \geq \sum_{i=1}^{n} I(M_1, X_1^{i-1}, X_2^{i-1}; X_{2i})$$
$$I(M_1; X_1^n) \leq \sum_{i=1}^{n} I(M_1, X_2^{i-1}, X_1^{i-1}; X_{1i}).$$

Identify $U_{1i} = (M_1, X_1^{i-1}, X_2^{i-1})$ and $U_{li} = M_l$ for $l \geq 2$. Define the time-sharing random variable Q to be uniformly distributed over $[1 : n]$ and independent of (M^q, X_1^n, X_2^n), and let $U_l = (Q, U_{lQ})$, $X_1 = X_{1Q}$, and $X_2 = X_{2Q}$. Clearly, $U_l \to (U^{l-1}, X_{j_l}) \to X_{j_{l+1}}$ form Markov chains. Finally, the cardinality bounds on U_l follow the standard technique, in particular, the one used in the 2-round interactive lossy source coding problem [Kas85]. This completes the converse proof.

3.4 An Equivalent Characterization of the Optimal Rate–exponent Tradeoff

In this section, we give an alternative characterization of the rate-exponent region, which reveals an interesting connection between the interactive hypothesis testing problem and the interactive lossy source coding problem described in Section 3.4.1.

First denote the rate–exponent region in remark 3.3 as $\mathcal{R}_1$, which consists of rate–exponent tuples $(R_1, \ldots, R_q, \theta_q)$ such that

$$R_l \geq I(U_l; X_{j_l}|U^{l-1}), \quad l \in [1 : q],$$
$$\theta_q \leq \sum_{l=1}^{q} I(U_l; X_{j_l}|U^{l-1})$$

for some pmfs $\prod_{l=1}^{q} p(u_l|u^{l-1}, x_{j_l})$, where $j_l = 1$ if l is odd and $j_l = 2$ if l is even. Define a rate–exponent region $\mathcal{R}_2$ that consists of rate–exponent triples

$(R_1, \ldots, R_q, \theta_q)$ such that

$$\begin{aligned} \theta_q &\le \sum_{l=1}^{q} I(U_l; X_{j_l}|U^{l-1}), \\ R_l &\ge I(U_l; X_{j_l}|U^{l-1}) - I(U_l; X_{j_{l+1}}|U^{l-1}), \quad l \in [1:q], \\ \sum_{l=1}^{q} R_l - \theta_q &\ge \sum_{l=1}^{q} \big(I(U_l; X_{j_l}|U^{l-1}) - I(U_l; X_{j_l}|U^{l-1})\big) \end{aligned}$$

for some $\prod_{l=1}^{q} p(u_l|u^{l-1}, x_{j_l})$, where $j_l = 1$ if l is odd and $j_l = 2$ if l is even. We can show the following, the proof of which can be found in Section 3.6.3.

Proposition 3.1. *The two regions are equivalent, i.e.,*

$$\mathcal{R}_1 = \mathcal{R}_2.$$

3.4.1 Relationship to Interactive Lossy Compression

In this section, we compare the two-round interactive hypothesis testing problem with the two-round interactive lossy source coding problem. Consider the interactive lossy source coding problem depicted in Fig. 3.9. Here two nodes interactively communicate with each other so that each node can reconstruct the source observed by the other node with prescribed distortions. Kaspi [Kas85] established the optimal tradeoff between communication constraints and the distortion pair (D_1, D_2). (See also Ma and Ishwar [MI11] for an ingenious example demonstrating that interactive lossy compression can strictly outperforms one-way lossy compression.)

The optimal tradeoff between communication and distortion is characterized by the rate–distortion region, the formal definition of which can be found in [Kas85] or [EGK11, Section 20.3].

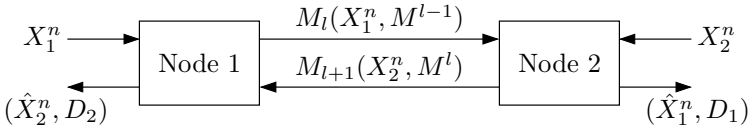


Figure 3.9: Interactive lossy compression.

Theorem 3.3 (Kaspi [Kas85])**.** *The two-round rate–distortion region is the set of all rate pairs* (R_1, R_2) *such that*

$$R_1 \geq I(X_1; U^q | X_2),$$
$$R_2 \geq I(X_2; U^q | X_1)$$

for some conditional pmf $\prod_{l=1}^{q} p(u_l|u^{l-1}, x_{j_l})$ *with* $|\mathcal{U}_l| \leq |\mathcal{X}_{j_l}| \cdot (\prod_{j=1}^{l} |\mathcal{U}_j|) + 1$ *and functions* $\hat{x}_1(u^q, x_2)$ *and* $\hat{x}_1(u^q, x_1)$ *that satisfy* $\mathsf{E}(d_j(X_j, \hat{X}_j)) \leq D_j$*,* $j = 1, 2$*, where* $j_l = 1$ *if* l *is odd and* $j_l = 2$ *if* l *is even.*

Achievability is established by performing Wyner–Ziv coding [WZ76] in each round, i.e., joint typicality encoding followed by binning. By contrast, the scheme we used for the interactive hypothesis testing problem is joint typicality encoding in each round (without binning). It turns out, however, that this distinction between binning and no binning is not fundamental. In fact, by using Wyner–Ziv coding in the interactive hypothesis testing problem, we can establish that $\mathcal{R}_2$ characterizes the tradeoff between communication constraints and the testing performance. The proof for the two-round case ($q = 2$) can be found in Section 3.4.2 and the general case follows straightforwardly. Therefore, the coding scheme for q-round interactive lossy source coding leads to an essentially identical scheme for q-round interactive hypothesis testing. It is refreshing to note that the same scheme is optimal for both problems.

Remark 3.4. *For the one-way case, Shimokawa, Han, and Amari [SHA94] showed that, by using binning, the the testing performance can be strictly improved given the same rate. The necessity of binning is also investigated by Rahman and Wagner [RW12] in a slightly different setup, where the test is testing against conditional independence (the distributed nodes share a common random variable) and they showed that binning is necessary to achieve the optimal rate–exponent tradeoff.*

3.4.2 Proof of Achievability

Codebook generation. Fix the conditional pmf $p(u_1|x)$ and let $p(u_1) = \sum_x p(x)p(u_1|x)$. Randomly and independently generate $2^{n\tilde{R}_1}$ sequences $u_1^n(l_1)$, $l_1 \in$

$[1 : 2^{n\tilde{R}_1}]$, each according to $\prod_{i=1}^{n} p_{U_1}(u_{1i})$. Partition the set of indices $l_1 \in [1 : 2^{n\tilde{R}_1}]$ into equal-size subsets referred to as bins $B(m_1) = [(m_1 - 1)2^{n(\tilde{R}_1 - R_1)} + 1 : m_1 2^{n(\tilde{R}_1 - R_1)}]$, $m_1 \in [1 : 2^{nR_1}]$. Fix the conditional pmf $p(u_2|u_1, y)$ and let $p(u_2) = \sum_{u_1, y} p(u_1, y)p(u_2|u_1, y)$. Randomly and independently generate $2^{n\tilde{R}_2}$ sequences $u_2^n(l_2|l_1)$, $l_2 \in [1 : 2^{n\tilde{R}_2}]$, each according to $\prod_{i=1}^{n} p_{U_2|U_1}(u_{2i}|u_{1i})$. Partition the set of indices $l_2 \in [1 : 2^{n\tilde{R}_2}]$ into equal-size subsets referred to as bins $B(m_2) = [(m_2 - 1)2^{n(\tilde{R}_2 - R_2)} + 1 : m_2 2^{n(\tilde{R}_2 - R_2)}]$, $m_2 \in [1 : 2^{nR_2}]$. The codebook $\mathcal{C}$ is revealed to both the encoder and decoder.

Encoding. We use joint typicality encoding. Given a sequence x^n, find an index l_1 such that $(x^n, u_1^n(l_1)) \in \mathcal{T}_\epsilon^{(n)}$. If there is more than one such index, it sends the smallest one among them. If there is no such index, it selects an index from $[1 : 2^{n\tilde{R}_1}]$ uniformly at random. Node 1 sends the bin index m_1 such that $l_1 \in B(m_1)$. Given a sequence y^n, find an index l_2 such that $(y^n, u_1^n(l_1), u_2^n(l_2)) \in \mathcal{T}_\epsilon^{(n)}$. If there is more than one such index, it sends the smallest one among them. If there is no such index, it selects an index from $[1 : 2^{n\tilde{R}_2}]$ uniformly at random. Node 2 sends the bin index m_2 such that $l_2 \in B(m_2)$.

Decoding. Let $\epsilon > \epsilon'$. The decoder first compute the $\mathcal{T}_\epsilon^{(n)}(U_1, Y)$ with respect to $P_{X,Y}(x, y)$ and $p(u_1|x)$ and $\mathcal{T}_\epsilon^{(n)}(U_2, X, Y)$ with respect to $P_{X,Y}(x, y)$, $p(u_1|x)$ and $p(u_2|u_1, y)$. Then set $\hat{H} = H_0$ if there is a unique $\hat{l}_2 \in B(m_2)$ such that $(u_1^n(\hat{l}_1), u_2^n(\hat{l}_2), x^n) \in \mathcal{T}_\epsilon^{(n)}$, otherwise set $\hat{H} = H_1$.

Analysis of two types of error. Let (L_1, L_2, M_1, M_2) denote the chosen indices. The decoder choose $\hat{H} \neq H_0$ iff one or more of the following events occur:

$$\begin{aligned}
\mathcal{E}_1 &= \{(U_1^n(l_1), X^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } l_1 \in [1 : 2^{n\tilde{R}_1})\},\\
\mathcal{E}_2 &= \{\nexists\, l_1 \in B(M_1) \text{ s.t. } (U_1^n(l_1), Y^n) \in \mathcal{T}_\epsilon^{(n)}\},\\
\mathcal{E}_3 &= \{(U_2^n(l), U_1^n(l_1), Y^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } l_2 \in [1 : 2^{n\tilde{R}_2})\},\\
\mathcal{E}_4 &= \{\nexists\, l_2 \in B(M_2) \text{ s.t. } (U_2^n(l_2), U_1^n(L_1), X^n) \in \mathcal{T}_\epsilon^{(n)}\}.
\end{aligned}$$

For the type I error, in this case, H_0 is true. Following similar steps from

the proof of the Wyner-Ziv coding scheme, we have $\alpha_n \to 0$ if

$$\begin{aligned} \tilde{R}_1 &\geq I(X;U_1) + \delta(\epsilon'), \\ \tilde{R}_1 - R_1 &\leq I(Y;U_1) - \delta(\epsilon), \\ \tilde{R}_2 &\geq I(U_2;Y|U_1) + \delta(\epsilon'), \\ \tilde{R}_2 - R_2 &\leq I(U_2;X|U_1) - \delta(\epsilon) \end{aligned}$$

For the type II error, in this case, H_1 is true.

$$\begin{aligned} \beta_n &= \mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3^c \cap \mathcal{E}_4^c) \\ &= \mathsf{P}(\mathcal{E}_1^c)\,\mathsf{P}(\mathcal{E}_2^c|\mathcal{E}_1^c)\,\mathsf{P}(\mathcal{E}_3^c|\mathcal{E}_2^c \cap \mathcal{E}_1^c)\,\mathsf{P}(\mathcal{E}_4^c|\mathcal{E}_3^c \cap \mathcal{E}_2^c \cap \mathcal{E}_1^c). \end{aligned}$$

By the covering lemma, $\mathsf{P}(\mathcal{E}_1^c)$ tends to one as $n \to \infty$ if $\tilde{R}_1 \geq I(U_1;X) + \delta(\epsilon')$. Let

$$\mathcal{E}'(l) = \{(U_1^n(l), Y^n) \in \mathcal{T}_\epsilon^{(n)}\},$$

we have

$$\begin{aligned} \mathsf{P}(\mathcal{E}_2^c|\mathcal{E}_1^c) &\leq \sum_{l_1 \in B(M_1)} \mathsf{P}(\mathcal{E}'(l_1) \cap \bigcap_{k \neq l_1} \mathcal{E}'(k)^c) \\ &= \sum_{l_1 \in B(M_1)} \mathsf{P}(\mathcal{E}'(l_1)) \prod_{k \neq l_1} (1 - \mathsf{P}(\mathcal{E}'(k))) \\ &\leq 2^{n(\tilde{R}_1 - R_1)} 2^{-n(I(U_1;Y) - \delta(\epsilon))} (1 - 2^{-n(I(U_1;Y) + \delta(\epsilon))})^{2^{n(\tilde{R}_1 - R_1)}} \\ &\overset{(a)}{\leq} 2^{-n(I(U_1;Y) + R_1 - \tilde{R}_1 - \delta(\epsilon))} e^{-2^{n(\tilde{R}_1 - R_1 - I(U_1;Y) - \delta(\epsilon))}} \\ &\overset{(b)}{\leq} 2^{-n(I(U_1;Y) + R_1 - \tilde{R}_1 - \delta'(\epsilon))}, \end{aligned}$$

where (a) follows by the joint typical lemma and the inequality $(1-x)^k \leq e^{-kx}$, and (b) follows since $\tilde{R}_1 - R_1 - I(U_1;Y) < 0$. By the covering lemma, $\mathsf{P}(\mathcal{E}_3^c|\mathcal{E}_1^c, \mathcal{E}_2^c) = \mathsf{P}(\mathcal{E}_3^c)$ tends to one as $n \to \infty$ if $\tilde{R}_2 \geq I(U_2;Y|U_1) + \delta(\epsilon')$. Let

$$\mathcal{E}''(l) = \{(U_2^n(l), U_1^n(L_1), X^n) \in \mathcal{T}_\epsilon^{(n)}\},$$

we have

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_4^c|\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3^c) &\leq \sum_{l_2 \in B(M_2)} \mathsf{P}(\mathcal{E}''(l_2) \cap \bigcap_{k \neq l_2} \mathcal{E}''(k)^c) \\
&= \sum_{l_2 \in B(M_2)} \mathsf{P}(\mathcal{E}''(l_2)) \prod_{k \neq l_2} (1 - \mathsf{P}(\mathcal{E}''(k))) \\
&\overset{(a)}{\leq} 2^{n(\tilde{R}_2 - R_2)} 2^{-n(I(U_2;X|U_1) - \delta(\epsilon))} \\
&\qquad \cdot (1 - 2^{-n(I(U_2;X|U_1)+\delta(\epsilon))})^{2^{n(\tilde{R}_2 - R_2)}} \\
&\overset{(b)}{\leq} 2^{-n(I(U_2;X|U_1)+R_2-\tilde{R}_2-\delta(\epsilon))} e^{-2^{n(\tilde{R}_2 - R_2 - I(U_2;X|U_1) - \delta(\epsilon))}} \\
&\overset{(c)}{\leq} 2^{-n(I(U_2;X|U_1)+R_2-\tilde{R}_2-\delta'(\epsilon))},
\end{aligned}$$

where (a) follows since $U_2 \to U_1 \to X$ forms a Markov chain when H_1 is true, (b) follows by the joint typical lemma and the inequality $(1-x)^k \leq e^{-kx}$, and (c) follows since $\tilde{R}_2 - R_2 < I(U_2; X|U_1)$. Thus the following type-II error is achievable,

$$2^{-n(I(U_2;X|U_1)+R_2-\tilde{R}_2-\delta'(\epsilon))} 2^{-n(I(U_1;Y)+R_1-\tilde{R}_1-\delta'(\epsilon))}.$$

Therefore, we have the following conditions

$$\begin{aligned}
I(U_1;X) \leq \tilde{R}_1 &\leq I(U_1;Y) + R_1, \\
I(U_2;Y|U_1) \leq \tilde{R}_2 &\leq I(U_2;X|U_1) + R_2, \\
R_1 + R_2 \leq \tilde{R}_1 + \tilde{R}_2 &\leq I(U_1;Y) + I(U_2;X|U_1) - \theta_2 + R_1 + R_2,
\end{aligned}$$

Eliminating $\tilde{R}_1$ and $\tilde{R}_2$ by the Fourier-Motzkin procedure yields the following characterization

$$\begin{aligned}
\mathcal{R}_2 = \bigcup_{p(u_1|x),p(u_2|u_1,y)} \{(R_1, R_2, \theta_2) : \\
\theta_2 &\leq I(U_1;Y) + I(U_2;X|U_1), \\
R_1 &\geq I(U_1;X) - I(U_1,Y), \\
R_2 &\geq I(U_2;Y|U_1) - I(U_2;X|U_1), \\
I(U_1;X) + I(U_2;Y|U_1) &\leq I(U_1;Y) + I(U_2;X|U_1) - \theta_2 + R_1 + R_2\}.
\end{aligned}$$

3.5 Discussions

In this section, we first present one interesting variant of the interactive hypothesis testing problem and then a few open questions.

3.5.1 Variable-length Setting

In this section, we discuss a variant of the interactive hypothesis testing problem. We still focus on the testing against independence case. A $(2^{nR_0}, 2^{nR_1}, n)$ test consists of two testers, one for each node. In each round, tester can either declare that H_0 or H_1 is true and terminate the communication or continue the communication.

More precisely, in round $l_j \in \{j, j+2, j+4, \ldots\}$, tester $j \in \{1, 2\}$ assigns $\hat{h}(x_j^n, m^{l_j}) \in \{H_0, H_1, C\}$ to its sequence and all previously transmitted indices,

- if $\hat{h}(x_j^n, m^{l_j}) = C$, communication continues and tester assigns $m_{l_j}(x_j^n, m^{l_j-1}) \in [1 : 2^{nr_{l_j}}]$,
- if $\hat{h}(x_j^n, m^{l_j}) \in \{H_0, H_1\}$, communication terminates.

The stopping time τ_n is defined as the first time that H_0 or H_1 is declared, i.e.,

$$\tau_n := \min\Big\{ l : \ \hat{h}(\, x_{l_o}^n,\, m^l) \in \{H_0, H_1\} \Big\},$$

where $l_o := (l \bmod 2) + 1$. We focus on the tests that terminate in $T < \infty$ rounds under both hypotheses, i.e.,

$$\mathsf{P}_0(\tau_n \le T) = 1 \ \text{ and } \ \mathsf{P}_1(\tau_n \le T) = 1, \quad \text{for all } n,$$

where P_0 and P_1 denote the probability measure under H_0 and H_1, respectively. The expected sum rate constraints under both hypotheses are

$$\lim_{n\to\infty} \mathsf{E}_0\left[\sum_{l=1}^{\tau_n} r_l\right] \le R_0 \text{ when } H_0 \text{ is true },$$

$$\lim_{n\to\infty} \mathsf{E}_1\left[\sum_{l=1}^{\tau_n} r_l\right] \le R_1 \text{ when } H_1 \text{ is true },$$

where E_0 and E_1 denote the expectation under H_0 and H_1, respectively. Let $\mathcal{A}_{k,n}(l)$, $k \in \{0,1\}$, be defined as

$$\mathcal{A}_{k,n}(l) := \left\{ (x_{l_o}^n, m^l) : \ \hat{h}(\, x_{l_o}^n \,, m^l) = H_k \right\}, \quad k \in \{0,1\}$$

and $\mathcal{A}_n(l) := \mathcal{A}_{0,n}(l) \cup \mathcal{A}_{1,n}(l)$. Then the type I error probability is

$$\mathsf{P}_0(\mathcal{A}_{1,n}) = \sum_{(x_1^n, x_2^n):(x_1^n, x_2^n, \{m\}) \in \mathcal{A}_{1,n}} p_0(x_1^n, x_2^n),$$

where $\{m\} := \{m_1, m_2, \ldots\}$ and $\{(x_1^n, x_2^n) : (x_1^n, x_2^n, \{m\}) \in \mathcal{A}_{1,n}\}$ is defined as the set of (x_1^n, x_2^n) such that

$$\bigvee_{l \geq 1} \left\{ \bigwedge_{l' < l} \left\{ (x_{l'_o}^n, m^{l'}) \in \mathcal{A}_n(l')^c \right\} \bigwedge \left\{ (x_{l_o}^n, m^l) \in \mathcal{A}_{1,n}(l) \right\} \right\}.$$

The type II error probability is defined similarly as

$$\mathsf{P}_1(\mathcal{A}_{0,n}) = \sum_{(x_1^n, x_2^n):(x_1^n, x_2^n, \{m\}) \in \mathcal{A}_{0,n}} p_1(x_1^n, x_2^n).$$

A tuple $(R_0(\theta), R_1(\theta), \theta)$ is said to be achievable if there exists a $(2^{nR_0(\theta)}, 2^{nR_1(\theta)}, n)$ test such that

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} -\frac{1}{n} \log \left(\min \mathsf{P}_1(\mathcal{A}_{0,n}) \right) \geq \theta,$$

where the minimum is over all tests such that $\mathsf{P}_0(\mathcal{A}_{1,n}) \leq \epsilon$. The rate exponent region is defined as the closure of all achievable tuples.

Consider the fixed-length setting for some error exponent $\theta > 0$ with q rounds communication and denote the sum rate as $R_{\text{sum},q}(\theta)$, then the theorem 2 can be rewritten as follows.

Corollary 3.1.

$$R_{sum,q}(\theta) = \min \sum_{l=1}^{q} I(U_l; X_{jl} | U^{l-1}),$$

where the minimum is over all $\prod_{l=1}^{q} p(u_l | x_{jl}, u^{l-1})$*, s.t.*

$$\sum_{l=1}^{q} I(U_l; X_{jl} | U^{l-1}) \geq \theta.$$

From the formulation of our variable-length setting, we are bound to have $R_0(\theta) \leq R_{\text{sum},q}(\theta)$ and $R_1(\theta) \leq R_{\text{sum},q}(\theta)$. The benefit of the variable-length setting can be reflected by the following result.

Theorem 3.4. *For any fixed error exponent $\theta > 0$,*

$$R_0(\theta) \leq R_{sum,q}(\theta) \quad \textit{and} \quad 0 = R_1(\theta) < R_{sum,q}(\theta).$$

Proof. Fix a fixed-length q-round test $\mathscr{T}$ for θ, i.e., fix an acceptance region $\mathcal{A}'_n$ such that

$$\mathsf{P}_0(\mathcal{A}'^c_n) \to 0 \quad \text{and} \quad -\frac{1}{n}\log \mathsf{P}_1(\mathcal{A}'_n) \to \theta.$$

Fix a $p(u_1|x_1)$ such that $I(U_1; X_1) \leq \epsilon_0$ for ϵ_0 arbitrarily small. Generate $u_1^n(m_1)$, $m_1 \in [1 : 2^{nr_1}]$, each according to $\prod_{i=1}^n p_{U_1}(u_{1i})$.

Testing for round 1. Encoder 1 looks for $u_1^n(m_1)$ such that $(u_1^n(m_1), x_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}$. If there exists one, it sends m_1 to node 2. Otherwise, set $m_1 = 0$

Testing for round 2. Let $\epsilon > \epsilon'$. Upon receiving $u_1^n(m_1)$, if $(u_1^n(m_1), x_2^n) \notin \mathcal{T}_{\epsilon}^{(n)}$, it declares H_1 is true. Otherwise, send 1-bit notification m_2 to node 1. Thus

$$\mathcal{A}_{0,n}(1) = \emptyset \quad \text{and} \quad \mathcal{A}_{1,n}(1) = \{(x_2^n, m_1) : (u_1^n(m_1), x_2^n) \notin \mathcal{T}_{\epsilon}^{(n)}\}.$$

Testing for round 3 until round (q+2). If it receives the 1-bit notification m_2, it applies the fixed test $\mathscr{T}$ in the following q rounds. Thus

$$\mathcal{A}_{0,n}(q+2) = \mathcal{A}'_n \quad \text{and} \quad \mathcal{A}_{1,n}(q+2) = \mathcal{A}'^c_n.$$

Analysis. Let

$$\mathcal{E}_1 := \{(U_1^n(m_1), X_1^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } m_1 \in [1 : 2^{nr_1}]\},$$
$$\mathcal{E}_2 := \{(U_1^n(m_1), X_1^n, X_2^n) \notin \mathcal{T}_{\epsilon}^{(n)}\}.$$

If H_0 is true, the sum rate is

$$\lim_{n\to\infty} \mathsf{E}\left[\sum_{l=1}^{\tau_n} r_l\right] = r_1 + R_{\text{sum},q}(\theta) \cdot \lim_{n\to\infty} \mathsf{P}\{(U_1^n(m_1), X_2^n) \in \mathcal{T}_{\epsilon}^{(n)}\}.$$

By the covering lemma [EGK11, Section 3.7], we have $\mathsf{P}(\mathcal{E}_1)$ tends to zero as $n \to \infty$, if $r_1 \geq I(U_1; X_1) + \delta(\epsilon')$. Since $\epsilon > \epsilon'$, by the conditional typicality lemma [EGK11, Section 2.5], we have that $\mathsf{P}(\mathcal{E}_2^c)$ tends to one as $n \to \infty$ and thus $\mathsf{P}\{(U_1^n(m_1), X_2^n) \in \mathcal{T}_\epsilon^{(n)}\}$ tends to one as $n \to \infty$. Thus $R_0(\theta) = r_1 + R_{\text{sum},q}(\theta)$. Type I error is

$$\begin{aligned}
\mathsf{P}_0(\mathcal{A}_{1,n}) &= \mathsf{P}_0\Big(\mathcal{A}_{1,n}(1) \cup \Big(\mathcal{A}_{1,n}(1)^c \cap \mathcal{A}_{1,n}(q+2)\Big)\Big) \\
&\leq \mathsf{P}_0(\mathcal{A}_{1,n}(1)) + \mathsf{P}_0\Big(\mathcal{A}_{1,n}(1)^c \cap \mathcal{A}_{1,n}(q+2)\Big) \\
&\leq \mathsf{P}_0(\mathcal{A}_{1,n}(1)) + \mathsf{P}_0(\mathcal{A}_{1,n}(q+2)) = \mathsf{P}\{(U_1^n(m_1), X_2^n) \notin \mathcal{T}_\epsilon^{(n)}\} + \mathsf{P}_0(\mathcal{A}_n'^c).
\end{aligned}$$

Thus by the assumption of the $\mathcal{T}$ and the covering lemma, we have that $\mathsf{P}_0(\mathcal{A}_{1,n})$ tends to zero if $r_1 \geq I(U_1; X_1) + \delta(\epsilon')$.

If H_1 is true, the sum rate is

$$r_1 + R_{\text{sum},q}(\theta) \cdot \lim_{n\to\infty} \mathsf{P}\{(U_1^n(m_1), X_2^n) \in \mathcal{T}_\epsilon^{(n)}\} \overset{(a)}{=} r_1,$$

where (a) follows since $\mathsf{P}\{(U_1^n(m_1), X_2^n) \in \mathcal{T}_\epsilon^{(n)}\} \leq 2^{-nI(U_1;X_2)}$ by the jointly typicality lemma [EGK11, Section 2.5]. Type II error is $\mathsf{P}_1(\mathcal{A}_{0,n}) = \mathsf{P}_1(\mathcal{A}_{0,n}(q+2)) = \mathsf{P}_1(\mathcal{A}_n')$. Since ϵ_0 can be arbitrarily small, thus we have shown that $(R_{\text{sum},q}(\theta), 0, \theta)$ is achievable in the variable-length setting. □

Remark 3.5. *It is easy to see that the above result holds for general hypothesis testing problem.*

Proposition 3.2. *For two hypotheses on the joint distribution of* (X_1, X_2)*:*

$$H_0 : p_0(x_1, x_2) \quad \text{and} \quad H_1 : p_1(x_1, x_2),$$

and any fixed error exponent $\theta > 0$*, we have* $R_0(\theta) \leq R_{sum,q}(\theta)$ *and* $0 = R_1(\theta) < R_{sum,q}(\theta)$*.*

It is still unknown whether there is a strict separation between the fixed-length scheme and the variable-length scheme.

3.5.2 Guassian Source

In this section, we discuss the Gaussian source case. It is unknown whether in this case, the interaction is strictly helpful or not. Let $X = Y + Z$, where $Y \sim \mathrm{N}(0, P)$ and $Z \sim \mathrm{N}(0, N)$ is independent of Y. For the one-round case, the rate-exponent function is

$$R(\theta) = \min_{p(u|y):I(U;X)\geq\theta} I(U;Y).$$

It is easy to see that $R(\theta)$ is equivalent to the following function

$$r(\alpha) = \max_{p(u|y):h(X|U)\leq\alpha} h(Y|U),$$

where $\alpha = h(X) - \theta$ and $r(\alpha) = h(Y) - R(\theta)$. Applying the entropy power inequality, we have

$$2^{2h(Y|U)} \leq 2^{2h(X|U)} - 2^{2h(Z|U)} \leq 2^{2\alpha} - 2^{2h(Z)} = 2^{2\alpha} - 2\pi eN.$$

Thus we have

$$h(Y|U) \leq \frac{1}{2}\log(2\pi e(P+N)2^{-2\theta} - 2\pi eN)$$

and

$$R(\theta) \geq \frac{1}{2}\log\frac{P}{(P+N)2^{-2\theta} - N}. \tag{3.9}$$

Now we show that the lower bound of $R(\theta)$ in (3.9) can be achieved. Let $U = Y + V$ with $V \sim \mathrm{N}(0, Q)$. It is easy to obtain the following

$$I(X;U) = \frac{1}{2}\log\frac{(P+N)(P+Q)}{PQ+PN+NQ}, \tag{3.10}$$

$$I(Y;U) = \frac{1}{2}\log\left(1 + \frac{P}{Q}\right). \tag{3.11}$$

With $I(Y;U) = R$, we have

$$Q = \frac{P}{2^{2R} - 1}. \tag{3.12}$$

Plugging (3.12) into the right hand side of (3.10), we have

$$R(\theta) \leq \frac{1}{2}\log\frac{P}{(P+N)2^{-2\theta} - N}.$$

For the two-round case, the rate-exponent function is

$$R(\theta) = \max I(U_1; X) + I(U_2; Y|U_1),$$

where the maximum is over all $p(u_1|x)$ and $p(u_2|u_1, y)$ such that

$$I(U_1; Y) + I(U_2; X|U_1) \geq \theta.$$

It is equivalent to

$$r(\alpha) = \max h(X|U_1) - h(Y|U_1) + h(Y|U_1, U_2),$$

where the maximum is over all $p(u_1|x)$ and $p(u_2|u_1, y)$ such that

$$h(Y|U_1) - h(X|U_1) + h(X|U_1, U_2) \leq \alpha$$

with $\alpha = h(Y) - \theta$ and $r(\alpha) = h(X) - R(\theta)$. However, it is no known how to provide a closed form expression for the two-round case.

As a comparison, in the following, we consider the interactive lossy source coding problem as a comparison with the interactive hypothesis testing problem. More specifically, we consider the same quadratic Gaussian source described at the beginning of this section.

Let $X = Y + Z$, where $Y \sim \mathrm{N}(0, P)$ and $Z \sim \mathrm{N}(0, N)$ is independent of Y. For the one-way case, the rate distortion function is

$$R_1(D) = \min_{\hat{y}(u,x), p(u|y): \mathsf{E}[d(Y,\hat{Y})] \leq D} I(Y; U|X).$$

Similar to the converse proof of the quadratic Gaussian lossy source coding problem, we have

$$\begin{aligned} I(Y; U|X) &= h(Y|X) - h(Y|U, X) \\ &= \frac{1}{2}\log(2\pi e \operatorname{Var}(Y|X)) - h(Y - \hat{Y}|U, X, \hat{Y}) \\ &\geq \frac{1}{2}\log(2\pi e \operatorname{Var}(Y|X)) - \frac{1}{2}\log(2\pi e D) \\ &= \frac{1}{2}\log\frac{\operatorname{Var}(Y|X)}{D}, \end{aligned} \tag{3.13}$$

where (3.13) follows since $\hat{Y}$ is a function of (U, X). Thus we have

$$R_1(D) \geq \frac{1}{2} \log \frac{\mathrm{Var}(Y|X)}{D}.$$

The achievability can be shown as in [EGK11, Example 11.2] by choosing $U = X + V$, where $V \sim \mathrm{N}(0, Q)$ in independent of (X, Y) and $Q = \mathrm{Var}(Y|X)D/(\mathrm{Var}(Y|X) - D)$. For the two-round case, the rate distortion function is

$$R_2(D) = \min_{p(u_1|x), p(u_2|u_1,y), \hat{y}(x,u_1,u_2): \mathsf{E}[d(Y,\hat{Y})] \leq D} I(X; U_1|Y) + I(Y; U_2|U_1, X).$$

Now we show that $R_2(D) \geq R_1(D)$, which implies that $R_2(D) = R_1(D)$.

$$\begin{aligned}
& I(X; U_1|Y) + I(Y; U_2|U_1, X) \\
&= h(X|Y) - h(X|U_1, Y) + h(Y|U_1, X) - h(Y|U_1, U_2, X) \\
&\geq h(Y|X) - h(Y - \hat{Y}|U_1, U_2, X) \qquad (3.14) \\
&\geq h(Y|X) - h(Y - \hat{Y}) \\
&\geq R_1(D),
\end{aligned}$$

where (3.14) follows since

$$I(X; U_1|Y) \geq I(Y; U_1|X) = 0$$

implies

$$h(X|Y) - h(X|U_1, Y) + h(Y|U_1, X) \geq h(Y|X)$$

and $\hat{Y}$ is a function of (U_1, U_2, X).

3.5.3 Strong Converse

The strong converse of the interactive hypothesis problem is still open. The strong converse proof of the one-way case is due to Ahlswede and Csiszár [AC86]. In this section, we provide a streamlined proof of their result. We need to introduce the following notation:

- P_{x^n}: type of a sequence x^n, i.e.,

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

- $\mathcal{P}_n$: the set of all possible types of $x^n \in \mathcal{X}^n$.
- $\mathcal{T}_P^n$: the set of sequences of type P, i.e.,

$$\mathcal{T}_P^n := \{x^n : P_{x^n} = P\}.$$

- (X, η)-essentail type: given a random variable X and $\eta > 0$, we call $P \in \mathcal{P}_n$ an (X, η)-essentail type if

$$\max_x |P(x) - P_X(x)| \leq \eta.$$

- $\mathcal{T}_\eta^n(X)$: (X, η)-typical sequences, i.e.,

$$\mathcal{T}_\eta^n(X) := \bigcup_{(X,\eta)\text{-essentail type } P} \mathcal{T}_P^n.$$

Theorem 3.5 ([AC86]). *For any $0 < \lambda < \epsilon < 1$, $\alpha > 0$, and $R' \geq R + \alpha$,*

$$\theta(R, \epsilon) \leq \theta(R', \lambda) + \alpha.$$

Note that from the theorem, we have

$$\begin{aligned}\theta(R, \epsilon) &\leq \lim_{\lambda\to 0} \lim_{\alpha\to 0} \big(\theta(R', \lambda) + \alpha\big) \\ &= \lim_{\lambda\to 0} \theta(R, \lambda) \\ &\leq \max_{p(u|x):R\geq I(U;X)} I(U; Y).\end{aligned}$$

Combined with the achievability proof, we have

$$\theta(R, \epsilon) = \max_{p(u|x):R\geq I(U;X)} I(U; Y), \quad \forall \epsilon \in (0, 1).$$

Outline of the proof of Theorem 3.5:

- Based on the type I error constraint and rate constraint, construct sets $E \subseteq \{x^n : m(x^n) = i_0\} \subseteq \mathcal{X}^n$ and $F \in \mathcal{Y}^n$ such that for any $x^n \in E$,

$$\begin{aligned}&\mathsf{P}_0(Y^n \in F | X^n = x^n) \geq \delta > 0 \\ &\mathsf{P}_1(Y^n \in F | X^n = x^n) \leq \beta_n(R, \epsilon) 2^{n\delta}.\end{aligned}$$

- Blow up E to $C := \Gamma^k E \cap \mathcal{T}_\eta^n(X)$ since the size of E is not big enough to cover the typical sequences

$$\{x^n \in \mathcal{T}_\eta^n(X) : m(x^n) = i_0\}.$$

- Blow up F to $D := \Gamma^{k+l} F$ so that $\mathsf{P}_0(Y^n \in \Gamma^{k+l} F | X^n = x^n) \approx 1$.

- Construct a test from C and D to construct a $(2^{nR'}, n)$ test with $R' \geq R$ to upper bound $\theta(R, \epsilon)$.

Proof of Theorem 3.5. • Step 1: construct sets $E \in \mathcal{X}^n$ and $F \in \mathcal{Y}^n$.

Consider any $m : \mathcal{X}^n \to [1 : 2^{nR}]$ and set $A \subseteq [1 : 2^{nR}] \times \mathcal{Y}^n$ such that

$$\mathsf{P}_0(A) \geq 1 - \epsilon, \tag{3.15}$$

$$\mathsf{P}_1(A) = \beta_n(R, \epsilon). \tag{3.16}$$

The acceptance region can be written as

$$A = \bigcup_{i=1}^{2^{nR}} i \times G_i, \quad G_i \subset \mathcal{Y}^n \text{ and } i = 1, 2, \ldots, 2^{nR}.$$

Then (3.15) and (3.16) imply

$$\mathsf{P}_0(Y^n \in G_{m(X^n)}) \geq 1 - \epsilon,$$
$$\mathsf{P}_1(Y^n \in G_{m(X^n)}) = \beta_n(R, \epsilon).$$

Fix δ and η which will be decided later. For $x^n \in \mathcal{T}_\eta^n(X)$, let

$$s(x^n) := \mathsf{P}_0(Y^n \notin G_{m(X^n)} | X^n = x^n),$$
$$t(x^n) := \mathsf{P}_1(Y^n \in G_{m(X^n)} | X^n = x^n),$$

and

$$B := \left\{x^n \in \mathcal{T}_\eta^n(X) : s(x^n) \leq 1 - \delta, t(x^n) \leq \beta_n(R, \epsilon) 2^{n\delta}\right\}.$$

Since

$$\sum_{x^n \in \mathcal{T}_\eta^n(X)} \mathsf{P}_0(X^n = x^n) s(x^n) = \mathsf{P}_0(X^n \in \mathcal{T}_\eta^n(X), Y^n \notin G_{m(X^n)}) \leq \epsilon,$$

$$\sum_{x^n \in \mathcal{T}_\eta^n(X)} \mathsf{P}_1(X^n = x^n) t(x^n) = \mathsf{P}_1(X^n \in \mathcal{T}_\eta^n(X), Y^n \notin G_{m(X^n)}) \leq \beta_n(R, \epsilon).$$

By the Markov inequality and the union bound, for n sufficiently large , we have

$$\begin{aligned}\mathsf{P}_0(X^n \in B) &\geq \mathsf{P}_0(X^n \in \mathcal{T}_\eta^n(X)) - \frac{\epsilon}{1-\delta} - 2^{-n\delta} \\ &\overset{(a)}{\geq} \frac{1-\epsilon}{2},\end{aligned}$$

where (a) follows by choosing $\delta \in (0, (1-\epsilon)/2)$.

Let

$$i_0 = \arg\min_i \mathsf{P}_0(X^n \in B, m(X^n) = i),$$

and choose

$$\begin{aligned}E &:= B \cap m^{-1}(i_0), \\ F &:= G_{i_0}.\end{aligned}$$

The sets E and F have the following properties, for every $x^n \in E$,

$$\mathsf{P}_0(X^n \in E) \geq \frac{1-\epsilon}{2^{nR+1}}, \tag{3.17}$$

$$\mathsf{P}_0(Y^n \in F | X^n = x^n) = 1 - s(x^n) \geq \delta, \tag{3.18}$$

$$\mathsf{P}_1(Y^n \in F | X^n = x^n) = t(x^n) \leq \beta_n(R, \epsilon) 2^{n\delta}. \tag{3.19}$$

- Step 2: blow up E and F to get C and D.

  Now we blow up E and F as follows,

  $$\begin{aligned}C &:= \Gamma^k E \cap \mathcal{T}_\eta^n(X), \\ D &:= \Gamma^{k+l} F,\end{aligned}$$

  where k and l will be specified later.

  Note that (3.17) implies that there exists a (X, η)-essential type $\bar{P} \in \mathcal{P}_n$ such that

  $$|E \cap \mathcal{T}_{\bar{P}}^n| \geq \frac{1-\epsilon}{2^{nR+1}} |\mathcal{T}_{\bar{P}}|.$$

  Let $P \in \mathcal{P}_n$ be any other (X, η)-essentail type. Then

  $$\max_x |P(x) - \bar{P}(x)| \leq 2\eta.$$

Then η and k are chosen such that for every (X,η)-essential $P \in \mathcal{P}_n$,

$$|C \cap \mathcal{T}_P^n| = |\Gamma^k E \cap \mathcal{T}_P^n| \geq \frac{1}{2^{nR}} 2^{(n(H(X)-2\delta))}.$$

For any $\bar{x}^n \in C$ and $x^n \in E$,

$$\begin{aligned} \mathsf{P}_0(Y^n \in \Gamma^k F | X^n = \bar{x}^n) &\geq |\mathcal{Y}|^{-2k} \, \mathsf{P}_0(Y^n \in F | X^n = x^n) \\ &\overset{(a)}{\geq} |\mathcal{Y}|^{-2k} \delta, \end{aligned}$$

where (a) follows from (3.18).

Then by the blowing up lemma, there exists l such that for any $\bar{x}^n \in C$,

$$\mathsf{P}_0(Y^n \in \Gamma^{k+l} F | X^n = \bar{x}^n) \geq 1 - \frac{\lambda}{2}.$$

Finally, for any $\bar{x}^n \in C$,

$$\begin{aligned} \mathsf{P}_1(Y^n \in D | X^n = \bar{x}^n) &= \mathsf{P}_1(Y^n \in \Gamma^{k+l} F | X^n = \bar{x}^n) \\ &\leq \mathsf{P}_1(Y^n \in F | X^n = x^n) 2^{2n\delta} \\ &\overset{(a)}{\leq} \beta_n(R,\epsilon) 2^{4n\delta}, \end{aligned}$$

where (a) follows from (3.19).

Thus, the sets C and D satisfies the following properties, for any $x^n \in C$,

$$|C \cap \mathcal{T}_P^n| \geq \frac{1}{2^{nR}} 2^{(n(H(X)-2\delta))}, \quad \forall (X,\eta)\text{-essential type } P \in \mathcal{P}_n, \tag{3.20}$$

$$\mathsf{P}_0(Y^n \in D | X^n = x^n) \geq 1 - \frac{\lambda}{2}, \tag{3.21}$$

$$\mathsf{P}_1(Y^n \in D | X^n = x^n) \leq \beta_n(R,\epsilon) 2^{4n\delta}. \tag{3.22}$$

- Step 3: construct a $(2^{nR'}, n)$ test based on C and D, where $R' \geq R + 3\delta$.

  Because of (3.20), by the covering lemma, for any (X,η)-essential type $P \in \mathcal{P}_n$, there exists permutations $\pi_{1,P}, \ldots, \pi_{N,P}$ such that

$$\mathcal{T}_P^n \subset \bigcup_{i=1}^{N} \pi_{i,P}(C), \quad N \leq 2^{n(R'-\delta/2)}.$$

Let $\pi_1, \ldots, \pi_M$ be all permutations selected as P runs over all the (X, η)-essential types. Then

$$\mathcal{T}_\eta^n(X) \subset \bigcup_{i=1}^{M} \pi_i(C), \quad M \le (n+1)^{|\mathcal{X}|} 2^{n(R'-\delta/2)}.$$

Define an encoder $m' : \mathcal{X}^n \to [1 : M]$ as

$$m'(x^n) = \begin{cases} 0, & \text{if } x^n \notin \mathcal{T}_\eta^n(X), \\ \text{smallest } i \text{ with } x^n \in \pi_i(C), & \text{if } x^n \in \mathcal{T}_\eta^n(X). \end{cases}$$

Note that the rate constraint is satisfied, we want to find $A' \in [1 : M] \times \mathcal{Y}^n$ such that

$$\mathsf{P}_0(A') \ge 1 - \lambda, \tag{3.23}$$

$$\mathsf{P}_1(A') \le \beta_n(R, n) 2^{4n\delta} 2^{n\delta}. \tag{3.24}$$

We now show that

$$A' := \bigcup_{i=1}^{M} \{i\} \times \pi_i(D)$$

satisfies both (3.23) and (3.24).

For (3.23),

$$\begin{aligned} \mathsf{P}_0(A') &= \mathsf{P}_0\big((m(X^n), Y^n) \in A'\big) \\ &= \sum_{i=1}^{M} \mathsf{P}_0\big(Y^n \in \pi_{m(X^n)}(D), m(X^n) = i\big) \\ &= \sum_{i=1}^{M} \sum_{x^n} \mathsf{P}_0\big(Y^n \in \pi_i(D), m(X^n) = i \big| X^n = x^n\big)\, \mathsf{P}_0(X^n = x^n) \\ &= \sum_{i=1}^{M} \sum_{x^n \in m^{-1}(i)} \mathsf{P}_0\big(Y^n \in \pi_i(D) \big| X^n = x^n\big)\, \mathsf{P}_0(X^n = x^n). \end{aligned}$$

Since (X^n, Y^n) is a pair of i.i.d. sequence,

$$\mathsf{P}_0\big(Y^n \in \pi_i(D) \big| X^n = x^n\big) = \mathsf{P}_0\big(Y^n \in D \big| X^n = \pi_i^{-1}(x^n)\big),$$

and by the definition of m', we have $\pi_i^{-1}(x^n) \in C$. Thus

$$\begin{aligned}\mathsf{P}_0(A') &\geq \mathsf{P}_0\bigg(X^n \in \bigcup_{i=1}^{M} m'^{-1}(i)\bigg)(1-\lambda/2)\\ &= \mathsf{P}_0\big(X^n \in \mathcal{T}_\eta^n(X)\big)(1-\lambda/2)\\ &\geq 1-\lambda.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\mathsf{P}_1(A') &= \sum_{i=1}^{M}\sum_{x^n\in m^{-1}(i)} \mathsf{P}_1\big(Y^n \in D\big|X^n = \pi_i^{-1}(x^n)\big)\,\mathsf{P}_1(X^n = x^n)\\ &\leq \mathsf{P}_1\big(X^n \in \mathcal{T}_\eta^n(X)\big)\beta_n(R,\epsilon)2^{4n\delta}\\ &\leq \beta_n(R,\epsilon)2^{4n\delta}2^{n\delta}.\end{aligned}$$

Therefore

$$\beta_n(R',\lambda) \leq \beta_n(R,\epsilon)2^{5n\delta}$$

for n sufficiently large and $R' > R + 3\delta$.

□

3.6 Technical Proofs

3.6.1 Proof of Lemma 3.1

To simplify the notation, let $\mathcal{T}_{\eta_q}^{(n)} := \mathcal{T}_{\eta_q}^{(n)}(U_q^n|u_{q-1}^n,\ldots,u_1^n,x_2^n)$ and $U_l^n := U_l^n(M_l|M^{l-1})$. Then for every $u_q^n \in \mathcal{T}_{\eta_q}^{(n)}$,

$$\begin{aligned}&\mathsf{P}\{U_q^n = u_q^n \,|\, U_{q-1}^n = u_{q-1}^n,\ldots,U_1^n = u_1^n, X_2^n = x_2^n\}\\ &= \mathsf{P}\{U_q^n = u_q^n, U_q^n \in \mathcal{T}_{\eta_q}^{(n)} \,|\, U_{q-1}^n(M_{q-1}|M^{q-2}) = u_{q-1}^n,\ldots,U_1^n = u_1^n, X_2^n = x_2^n\}\\ &= \mathsf{P}\{U_q^n \in \mathcal{T}_{\eta_q}^{(n)} \,|\, U_{q-1}^n = u_{q-1}^n,\ldots,U_1^n = u_1^n, X_2^n = x_2^n\}\\ &\quad\cdot \mathsf{P}\{U_q^n = u_q^n \,|\, U_{q-1}^n = u_{q-1}^n,\ldots,U_1^n = u_1^n, X_2^n = x_2^n, U_q^n \in \mathcal{T}_{\eta_q}^{(n)}\}\\ &\leq \mathsf{P}\{U_q^n = u_q^n \,|\, U_{q-1}^n = u_{q-1}^n,\ldots,U_1^n = u_1^n, X_2^n = x_2^n, U_q^n \in \mathcal{T}_{\eta_q}^{(n)}\}\\ &= \sum_{m^q} \mathsf{P}\{U_q^n = u_q^n, M^q = m^q \,|\, U_{q-1}^n = u_{q-1}^n,\ldots,U_1^n = u_1^n,\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad X_2^n = x_2^n, U_q^n \in \mathcal{T}_{\eta_q}^{(n)}\} \\
&= \sum_{m^q} \mathsf{P}\{M^q = m^q \,|\, U_{q-1}^n = u_{q-1}^n, \ldots, U_1^n = u_1^n, X_2^n = x_2^n, U_q^n \in \mathcal{T}_{\eta_q}^{(n)}\} \\
&\quad \cdot \mathsf{P}\{U_q^n = u_q^n \,|\, U_{q-1}^n = u_{q-1}^n, \ldots, U_1^n = u_1^n, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad X_2^n = x_2^n, U_q^n \in \mathcal{T}_{\eta_q}^{(n)}, M^q = m^q\} \\
&\overset{(a)}{=} \sum_{m^q} \mathsf{P}\{M^q = m^q \,|\, U_{q-1}^n = u_{q-1}^n, \ldots, U_1^n = u_1^n, X_2^n = x_2^n, U_q^n \in \mathcal{T}_{\eta_q}^{(n)}\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdot \mathsf{P}\{U_q^n = u_q^n \,|\, U_q^n \in \mathcal{T}_{\eta_q}^{(n)}\} \\
&\overset{(b)}{\leq} \sum_{m^q} \mathsf{P}\{M^q = m^q \,|\, U_{q-1}^n = u_{q-1}^n, \ldots, U_1^n = u_1^n, X_2^n = x_2^n, U_q^n \in \mathcal{T}_{\eta_q}^{(n)}\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdot 2^{-n(H(U_q|U^{q-1},X_2)-\delta(\eta_q))} \\
&= 2^{-n(H(U_q|U^{q-1},X_2)-\delta(\eta_q))},
\end{aligned}
$$

where (a) follows since $U_q^n(m_q|m^{q-1})$ is independent of X_2^n and $U_2^n(m_q'|m^{q-1})$ for $m_q' \neq m_q$ and is conditionally independent of M^q given (X_2^n, $U_1(m_1)$,..., $U_{q-1}(m_{q-1}|m^{q-2})$) and the indicator variables of the event $U_q^n(m_q|m^{q-1}) \in \mathcal{T}_{\eta_q}^{(n)}$, $m_q \in [1 : 2^{nR_q}]$, which implies that the event $\{U_q^n(m_q|m^{q-1}) = u_q^n\}$ is conditionally independent of $\{X_2^n, U_1(m_1), \ldots, U_{q-1}(m_{q-1}|m^{q-2}), M^q = m^q\}$ given $U_q^n(m_q|m^{q-1})$ $\in \mathcal{T}_{\eta_q}^{(n)}$. Step (b) follows from the properties of typical sequences. Similarly, for every for every $u_q^n \in \mathcal{T}_{\eta_q}^{(n)}$ and n sufficiently large,

$$
\begin{aligned}
&\mathsf{P}\{U_q^n(M_q|M^{q-1}) = u_q^n \,|\, U_{q-1}^n(M_{q-1}|M^{q-2}) = u_{q-1}^n, \ldots, U_1(M_1)^n = u_1^n, X_2^n = x_2^n\} \\
&\geq (1-\eta_q) 2^{-n(H(U_q|U^{q-1},X_2)+\delta(\eta_q))}.
\end{aligned}
$$

This completes the proof of Lemma 1.

3.6.2 Proof of Lemma 3.2

We have

$$
\begin{aligned}
&p_1(u_{l-1}^n, x_{j_l}^n | u_1^n, u_2^n, \ldots, u_{l-2}^n) \\
&= \sum_{x_{j_{l-1}}^n} p_1(u_{l-1}^n, x_{j_l}^n, x_{j_{l-1}}^n | u_1^n, u_2^n, \ldots, u_{l-2}^n)
\end{aligned}
$$

$$\begin{aligned}
&\overset{(a)}{=} \sum_{x^n_{j_{l-1}}} p_1(x^n_{j_l}, x^n_{j_{l-1}}|u^n_1, u^n_2, \ldots, u^n_{l-2}) p_1(u^n_{l-1}|u^n_1, u^n_2, \ldots, u^n_{l-2}, x^n_{j_{l-1}}) \\
&\overset{(b)}{=} \sum_{x^n_{j_{l-1}}} p_1(x^n_{j_l}|x^n_{j_{l-1}}, u^n_1, u^n_2, \ldots, u^n_{l-2}) p_1(x^n_{j_{l-1}}|u^n_1, u^n_2, \ldots, u^n_{l-2}) \\
&\qquad\qquad \cdot p_1(u^n_{l-1}|u^n_1, u^n_2, \ldots, u^n_{l-2}, x^n_{j_{l-1}}) \\
&= \sum_{x^n_{j_{l-1}}} p_1(x^n_{j_l}|u^n_1, u^n_2, \ldots, u^n_{l-2}) p_1(u^n_{l-1}, x^n_{j_{l-1}}|u^n_1, u^n_2, \ldots, u^n_{l-2}) \\
&= p_1(x^n_{j_l}|u^n_1, u^n_2, \ldots, u^n_{l-2}) p_1(u^n_{l-1}|u^n_1, u^n_2, \ldots, u^n_{l-2}),
\end{aligned}$$

where (a) follows since U^n_{l-1} is conditionally independent of $X^n_{j_l}$ given

$$(U^n_1, U^n_2, \ldots, U^n_{l-2}, X^n_{j_{l-1}})$$

and (b) follows since $X^n_{j_{l-1}}$ and $X^n_{j_l}$ are independent under H_1.

3.6.3 Proof of Proposition 3.1

In this section, we focus on the two-round case and the q-round case follows straightforwardly. Define the two regions as

$$\begin{aligned}
\mathcal{R}_1 = &\bigcup_{p(u_1|x), p(u_2|u_1,y)} \{(R_1, R_2, \theta) : R_1 \geq I_1, R_2 \geq I_2, \theta \leq I_3 + I_4\}, \\
\mathcal{R}_2 = &\bigcup_{p(u_2|x), p(u_2|u_1,y)} \{(R_1, R_2, \theta) : R_1 \geq I_1 - I_3, R_2 \geq I_2 - I_4, \\
&\qquad\qquad \theta \leq I_3 + I_4, R_1 + R_2 - \theta \geq I_1 + I_2 - I_3 - I_4\},
\end{aligned}$$

where for fixed $p(u_1|x)$ and $p(u_2|u_1, y)$, let

$$\begin{aligned}
I_1 &:= I(U_1; X), \\
I_2 &:= I(U_2; Y|U_1), \\
I_3 &:= I(U_1; Y), \\
I_4 &:= I(U_2; X|U_1).
\end{aligned}$$

We now show that $\mathcal{R}_1 = \mathcal{R}_2$ as follows. The corner points of $\mathcal{R}_1$ are

$$c_1 := (I_1, I_2, 0) \text{ and } c_2 := (I_1, I_2, I_3 + I_4).$$

The corner points of $\mathcal{R}_2$ are

$$d_1 := (I_1 - I_3, I_2 - I_4, 0) \text{ and } d_2 := (I_1, I_2, I_3 + I_4).$$

It is easy to see that $\mathcal{R}_1 \subseteq \mathcal{R}_2$. Since $d_1 \in \mathcal{R}_1$ by choosing $U = \emptyset$ and $d_2 \in \mathcal{R}_1$ by choosing the fixed $p(u_1|x)$ and $p(u_2|u_1, y)$, thus $\mathcal{R}_2 \subseteq \mathcal{R}_1$.

Chapter 3, in part, includes the material in Yu Xiang and Young-Han Kim, "Interactive hypothesis testing with communication constraints," *Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL*, pp. 1065–1072, Monticello, IL, October 2012, and Yu Xiang and Young-Han Kim, "Interactive hypothesis testing against independence," *IEEE International Symposium on Information Theory*, pp. 2840–2844, Istanbul, Turkey, July 2013. The dissertation author was the primary investigator and author of this paper.

Chapter 4

Gaussian Channel with Noisy Feedback

In this chapter, the optimal coding over the additive white Gaussian noise channel under the *peak* energy constraint is studied when there is noisy feedback over an orthogonal additive white Gaussian noise channel. As shown by Pinsker, under the peak energy constraint, the best error exponent for communicating an M-ary message, $M \geq 3$, with noise-free feedback is strictly larger than the one without feedback. In this chapter, we extend Pinsker's result and show that if the noise power in the feedback link is sufficiently small, the best error exponent for communicating an M-ary message can be strictly larger than the one without feedback. The proof involves two feedback coding schemes. One is motivated by a two-stage noisy feedback coding scheme of Burnashev and Yamamoto for binary symmetric channels, while the other is a linear noisy feedback coding scheme that extends Pinsker's noise-free feedback coding scheme. When the feedback noise power α is sufficiently small, the linear coding scheme outperforms the two-stage (nonlinear) coding scheme, and is asymptotically optimal as α tends to zero. By contrast, when α is relatively larger, the two-stage coding scheme performs better.

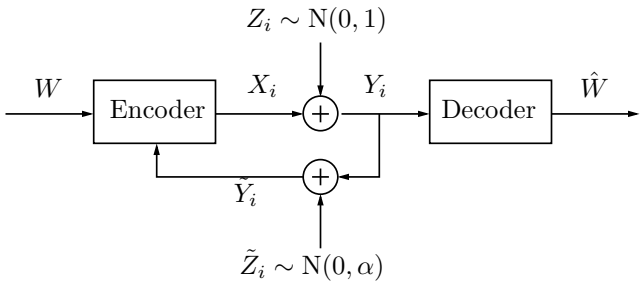


Figure 4.1: Gaussian channel with noisy feedback.

4.1 Introduction

We consider a communication problem for an additive white Gaussian noise (AWGN) *forward* channel with feedback over an orthogonal additive white Gaussian noise *backward* channel as depicted in Fig. 1. Suppose that the sender wishes to communicate a message $W \in [1:M] := \{1, 2, \ldots, M\}$ over the (forward) additive white Gaussian noise channel

$$Y_i = X_i + Z_i,$$

where X_i, Y_i, and Z_i respectively denote the channel input, channel output, and additive Gaussian noise. The sender has a causal access to a noisy version $\tilde{Y}_i$ of Y_i over the feedback (backward) additive white Gaussian noise channel

$$\tilde{Y}_i = Y_i + \tilde{Z}_i,$$

where $\tilde{Z}_i$ is the Gaussian noise in the backward link. We assume that the forward noise process $\{Z_i\}_{i=1}^{\infty}$ and the backward noise process $\{\tilde{Z}_i\}_{i=1}^{\infty}$ are independent of each other, and respectively white Gaussian $\mathrm{N}(0,1)$ and $\mathrm{N}(0,\alpha)$.

We define an (M,n) code with the encoding functions $x_i(w, \tilde{y}^{i-1})$, $i \in [1:n]$, and the decoding function $\hat{w}(y^n)$. We assume a *peak energy constraint*

$$\mathsf{P}\left\{\sum_{i=1}^{n} x_i^2(w, \tilde{Y}^{i-1}) \le nP\right\} = 1 \quad \text{for all } w. \tag{4.1}$$

The probability of error of the code is defined as

$$\begin{aligned} P_e^{(n)} &= \mathsf{P}\{W \neq \hat{W}(Y^n)\} \\ &= \frac{1}{M}\sum_{w=1}^{M} \mathsf{P}\{W \neq \hat{W}(Y^n)|W = w\}, \end{aligned}$$

where W is distributed uniformly over $\{1, 2, \ldots, M\}$ and is independent of $(Z^n, \tilde{Z}^n)$.

As is well known, the capacity of the channel (the supremum of $(\log M)/n$ such that there exists a sequence of (M, n) codes with $\lim_{n\to\infty} P_e^{(n)} \to 0$) stays the same with or without feedback. Hence, our main focus is the reliability of communication, which is captured by the error exponent

$$\lim_{n\to\infty} -\frac{1}{n} \ln P_e^{(n)}$$

of the given code. The error exponent is sensitive to the presence of noise in the feedback link. Schalkwijk and Kailath showed in their celebrated work [SK66] that *noise-free* feedback can improve the error exponent dramatically under the *expected energy constraint*

$$\sum_{i=1}^{n} \mathsf{E}[x_i^2(w, \tilde{Y}^{i-1})] \leq nP \quad \text{for all } w, \tag{4.2}$$

(in fact, $P_e^{(n)}$ decays much faster than exponentially in n). Kim, Lapidoth, and Weissman [KLW07] studied the optimal error exponent under the expected energy constraint and noisy feedback, and showed that the error exponent is inversely proportional to α for small α.

Another important factor that affects the error exponent is the energy constraint on the channel inputs—the peak energy constraint in (4.1) vs. the expected energy constraint in (4.2). Wyner [Wyn68] showed that the error probability of the Schalkwijk–Kailath coding scheme [SK66] degrades to an exponential form under the peak energy constraint. In fact, Shepp, Wolf, Wyner, and Ziv [SWWZ69] showed that for the binary-message case ($M = 2$), the best error exponent under the peak energy constraint is achieved by simple nonfeedback antipodal signaling, regardless of the presence of feedback. This negative result might lead to an impression that under the peak energy constraint, even noise-free feedback does not

improve the reliability of communication. Pinsker [Pin68] proved the contrary by showing that the best error exponent for sending an M-ary message does not depend on M and, hence can be strictly larger than the best error exponent without feedback for $M \geq 3$.

In this chapter, we show that noisy feedback can improve the reliability of communication under the peak energy constraint, provided that the feedback noise power α is sufficiently small. Let

$$E_M(\alpha) := \limsup_{n\to\infty} -\frac{1}{n} \ln P_e^*(M,n),$$

where $P_e^*(M,n)$ denotes the best error probability over all (M,n) codes for the AWGN channel with the noisy feedback. Thus, $E_M(\infty)$ denotes the best error exponent for communicating an M-ary message over the AWGN channel *without* feedback. Shannon [Sha59] showed that

$$E_M(\infty) = \frac{M}{4(M-1)}P. \tag{4.3}$$

This follows by first upper bounding the error exponent with the sphere packing bound and then achieving this upper bound by using a regular simplex code on the sphere of radius $\sqrt{nP}$, that is, each codeword $x^n(w)$ satisfies $\sum_{i=1}^n x_i^2(w) = nP$ and is at the same Euclidean distance from every other codeword. In particular, for $M = 3$,

$$\begin{aligned}
x^n(1) &= \sqrt{nP} \cdot (\quad 0, \quad\quad 1, 0, \ldots, 0),\\
x^n(2) &= \sqrt{nP} \cdot (-1/2, -\sqrt{3}/2, 0, \ldots, 0),\\
x^n(3) &= \sqrt{nP} \cdot (\quad 1/2, -\sqrt{3}/2, 0, \ldots, 0),
\end{aligned}$$

and

$$E_3(\infty) = \frac{3}{8}P.$$

At the other extreme, $E_M(0)$ denotes the best error exponent for communicating an M-ary message over the AWGN channel with *noise-free* feedback. Pinsker [Pin68] showed that

$$E_M(0) \equiv \frac{P}{2}$$

for all M. In particular,

$$E_3(0) = \frac{P}{2}.$$

Clearly, $E_M(\alpha)$ is decreasing in α and

$$E_M(\infty) \le E_M(\alpha) \le E_M(0)$$

for every α and M.

Is $E_M(\alpha)$ strictly larger than $E_M(\infty)$ (i.e., is noisy feedback better than no feedback)? Does $E_M(\alpha)$ tend to $E_M(0)$ as $\alpha \to 0$ (i.e., does the performance degrade gracefully with small noise in the feedback link)? What is the optimal feedback coding scheme that achieves $E_M(\alpha)$? To answer these questions, we establish the following results.

Theorem 4.1. *For* $0 \le s \le 1$,

$$E_M(\alpha^*(s)) \ge \frac{P}{2}\left(1 - \frac{3(M-2)}{M(s^2-2s+4)+3(M-2)}\right),$$

where

$$\alpha^*(s) = \frac{3s^2}{4(s^2-2s+4)}.$$

By comparing the lower bound with (4.3) and identifying the critical point $\alpha = \alpha^*(1) = 1/4$, we obtain the following.

Corollary 4.1.

$$E_M(\alpha) > E_M(\infty) \quad \textit{for } \alpha < \frac{1}{4}.$$

Thus, if the noise power in the feedback link is sufficiently small, then the noisy feedback improves the reliability of communication even under the peak energy constraint. The proof of Theorem 1 is motivated by recent results of Burnashev and Yamamoto in a series of papers [BY08a], [BY08b], where they considered a communication model with a forward BSC(p) and a backward BSC(αp), and showed that when α is sufficiently small, the best error exponent is strictly larger than the one without feedback.

The lower bound in Theorem 1 shows that $\liminf_{\alpha\to 0} E_M(\alpha) \ge 2PM/(7M-6)$, which is strictly less than $E_M(0) = P/2$. To obtain a better asymptotic behavior for $\alpha \to 0$, we establish the following.

Theorem 4.2.

$$E_M(\alpha) \geq \frac{P}{2} \frac{1}{1+\alpha+4(\lfloor M/2 \rfloor)^2\alpha+4(\lfloor M/2 \rfloor)\sqrt{\alpha(1+\alpha)}}$$
$$\geq \frac{P}{2} \frac{1}{(\sqrt{\alpha}M+\sqrt{1+\alpha})^2}.$$

This theorem leads to the following.

Corollary 4.2.

$$\lim_{\alpha \to 0} E_M(\alpha) = E_M(0).$$

Thus, the lower bound in Theorem 2 is tight for $\alpha \to 0$. The proof of Theorem 2 extends Pinsker's linear noise-free feedback coding scheme [Pin68] to the noisy case.

Fig. 4.2 compares the two bounds for the $M = 3$ case. The linear noisy feedback coding scheme performs better when α is sufficiently small, while the two-stage noisy feedback coding scheme performs better when α is relatively larger.

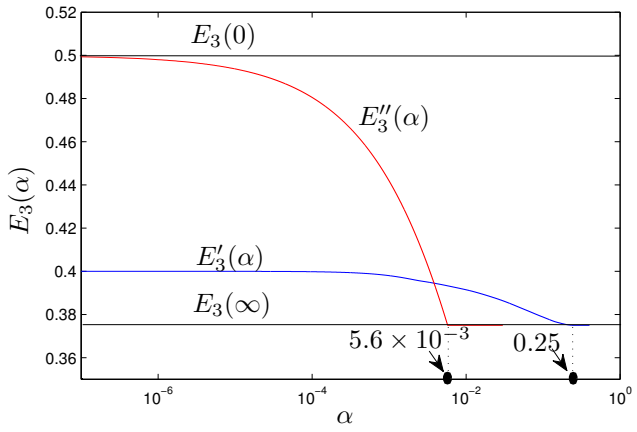


Figure 4.2: Comparison of the two noisy feedback coding scheme for $M = 3$.

The rest of the chapter is organized as follows. In Section II, we study a two-stage noisy feedback coding scheme motivated by recent results of Burnashev and Yamamoto and establish Theorem 1. In Section III, we extend Pinsker's noise-free linear feedback coding scheme to the noisy feedback case and establish Theorem 2. Section IV concludes this chapter.

4.2 Two-stage Noisy Feedback Scheme

4.2.1 Background

It is instructive to first consider a two-stage noise-free feedback coding scheme for $M = 3$. This two-stage scheme has been studied by Schalkwijk and Barron [SB71] and Yamamoto and Itoh [YI79] for a general M.

Encoding. Fix some $\lambda \in (0, 1)$. For simplicity of notation, assume throughout that λn is an integer. To send message $w \in [1 : 3]$, during the transmission time interval $[1 : \lambda n]$ (namely, stage 1), the encoder uses the simplex signaling:

$$x^{\lambda n}(w) = \begin{cases} \sqrt{\lambda nP} \cdot (\quad 0, \quad\quad 1, 0, \ldots, 0) & \text{for } w = 1, \\ \sqrt{\lambda nP} \cdot (-1/2, -\sqrt{3}/2, 0, \ldots, 0) & \text{for } w = 2, \\ \sqrt{\lambda nP} \cdot (\ 1/2, -\sqrt{3}/2, 0, \ldots, 0) & \text{for } w = 3. \end{cases} \tag{4.4}$$

Based on the feedback $y^{\lambda n}$, the encoder then chooses the two most probable message estimates $\hat{w}_1$ and $\hat{w}_2$, where

$$p(\hat{w}_1|y^{\lambda n}) \geq p(\hat{w}_2|y^{\lambda n}) \geq p(\hat{w}_3|y^{\lambda n}) \tag{4.5}$$

and in case of a tie the one with the smaller index is chosen. Since the channel is Gaussian and W is uniform, (4.5) can be written as

$$||x^{\lambda n}(\hat{w}_1) - y^{\lambda n}|| \leq ||x^{\lambda n}(\hat{w}_2) - y^{\lambda n}|| \leq ||x^{\lambda n}(\hat{w}_3) - y^{\lambda n}||,$$

where $||\cdot||$ denotes the Euclidean distance. During the transmission time interval $[\lambda n + 1 : n]$ (stage 2), the encoder uses antipodal signaling for w if $w \in \{\hat{w}_1, \hat{w}_2\}$ and transmits all-zero sequence otherwise:

$$\begin{aligned} &x^n_{\lambda n+1}(w) \\ &= \begin{cases} \sqrt{(1-\lambda)nP} \cdot (1, 0, 0, \ldots, 0) & \text{if } w = \min\{\hat{w}_1, \hat{w}_2\}, \\ \sqrt{(1-\lambda)nP} \cdot (-1, 0, 0, \ldots, 0) & \text{if } w = \max\{\hat{w}_1, \hat{w}_2\}, \\ (0, 0, 0, \ldots, 0) & \text{otherwise.} \end{cases} \end{aligned}$$

Decoding. At the end of stage 1, the decoder chooses the two most probable message estimates $\hat{w}_1$ and $\hat{w}_2$ based on $Y^{\lambda n}$ as the encoder does. At the end of stage 2, the decoder declares that $\hat{w}$ is sent if

$$\begin{aligned}\hat{w} &= \arg \min_{w \in \{\hat{w}_1, \hat{w}_2\}} ||x^n(w) - y^n|| \\ &= \arg \min_{w \in \{\hat{w}_1, \hat{w}_2\}} \big(||x^{\lambda n}(w) - y^{\lambda n}||^2 + ||x^n_{\lambda n+1}(w) - y^n_{\lambda n+1}||^2 \big)^{1/2}.\end{aligned}$$

Analysis of the probability of error. Let $\hat{W}_1$ and $\hat{W}_2$ denote the two most probable message estimates at the end of stage 1. The decoder makes an error if and only if one of the following events occurs:

$$\begin{aligned}\mathcal{E}_1 &= \big\{ W \neq \hat{W}_1 \text{ and } W \neq \hat{W}_2 \big\}, \\ \mathcal{E}_2 &= \big\{ W \in \{\hat{W}_1, \hat{W}_2\} \text{ and } \hat{W} \neq W \big\}.\end{aligned}$$

Thus, the probability of error is

$$P_e^{(n)} = \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2).$$

By symmetry, we assume without loss of generality that $W = 1$ is sent. For brevity, we do not explicitly condition on the event $\{W = 1\}$ in probability expressions in the following, whenever it is clear from the context. Referring to Fig. 4.3, let

$$\begin{aligned}A_{23} = \big\{ y^{\lambda n} \colon & ||x^{\lambda n}(1) - y^{\lambda n}|| \geq ||x^{\lambda n}(2) - y^{\lambda n}|| \\ & \text{and } ||x^{\lambda n}(1) - y^{\lambda n}|| \geq ||x^{\lambda n}(3) - y^{\lambda n}|| \big\},\end{aligned}$$

we have

$$\begin{aligned}\mathsf{P}(\mathcal{E}_1) &= \mathsf{P}\{Y^{\lambda n} \in A_{23}\} \\ &\leq Q(d_1) \\ &\overset{(a)}{\leq} \frac{1}{2} \exp\left(- \frac{\lambda n P}{2} \right),\end{aligned}$$

where (a) follows since $Q(x) \leq (1/2)\exp(-x^2/2)$ for $x \geq 0$ (see [WJ65, Problem 2.26]).

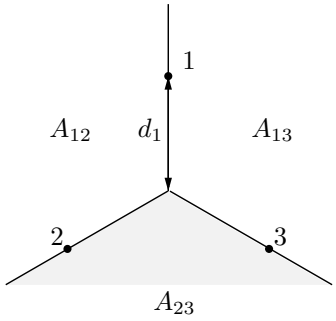


Figure 4.3: The error event $\mathcal{E}_1$ when $W = 1$. Here $d_1 = \sqrt{\lambda nP}$ and 1, 2, and 3 denote $x^{\lambda n}(1)$, $x^{\lambda n}(2)$, and $x^{\lambda n}(3)$, respectively.

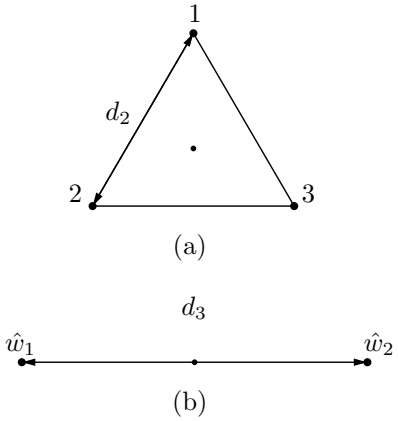


Figure 4.4: The error event $\mathcal{E}_2$. Here $d_2 = \sqrt{3\lambda nP}$ and $d_3 = \sqrt{4(1-\lambda)nP}$.

On the other hand, $\mathsf{P}(\mathcal{E}_2)$ is determined by the distance between the simplex signaling in stage 1 and the distance between the antipodal signaling in stage 2 (see Fig. 4.4). In particular,

$$||X^n(\hat{W}_1) - X^n(\hat{W}_2)|| = \sqrt{d_2^2 + d_3^2} = \sqrt{(4-\lambda)nP}.$$

Thus,

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_2) &= Q\bigg(\frac{||X^n(\hat{W}_1) - X^n(\hat{W}_2)||}{2}\bigg) \\
&= Q\bigg(\sqrt{\Big(1-\frac{\lambda}{4}\Big)nP}\bigg) \\
&\leq \frac{1}{2}\exp\bigg(-\frac{1}{2}\Big(1-\frac{\lambda}{4}\Big)nP\bigg).
\end{aligned}$$

Therefore, the error exponent of the two-stage feedback coding scheme is lower bounded as

$$\begin{aligned} E_3'(0) &= \limsup_{n\to\infty} -\frac{1}{n} \ln P_e^{(n)} \\ &= \limsup_{n\to\infty} -\frac{1}{n} \max\{\ln \mathsf{P}(\mathcal{E}_1), \ln \mathsf{P}(\mathcal{E}_2)\} \\ &\geq \min\left\{ \frac{\lambda P}{2}, \frac{P}{2}\left(1 - \frac{\lambda}{4}\right) \right\}. \end{aligned}$$

Now let $\lambda = 4/5$. Then it can be readily verified that both terms in the minimum are the same and we have

$$E_3(0) \geq E_3'(0) \geq \frac{2P}{5}.$$

Remark 4.1. *Since $E_3(0) = P/2$, this two-stage noise-free feedback coding scheme is strictly suboptimal.*

Remark 4.2. *We need only three transmissions: two for stage 1 and one for stage 2. Thus λ actually divides only the total energy nP, not the block length n.*

4.2.2 Coding Scheme and Performance Analysis

Based on the two-stage noise-free feedback coding scheme in the previous subsection and a new idea of *signal protection* introduced by Burnashev and Yamamoto [BY08a], [BY08b], we present a two-stage noisy feedback coding scheme for $M = 3$. The coding scheme for an arbitrary M is given in the Appendix.

In the two-stage noise-free feedback coding scheme, the encoder and decoder agree on the same set of message estimates $\hat{w}_1$ and $\hat{w}_2$ at the end of stage 1. When there is noise in the feedback link, however, this coordination is not always possible. To solve this problem, we assign a signal protection region B_w, $w \in [1:3]$, to each signal $x^{\lambda n}(w)$ as depicted in Fig. 4.5. Let $x^{\lambda n}$ and $y^{\lambda n}$ denote the transmitted and received signals, respectively, and $\tilde{y}^{\lambda n}$ denote the feedback sequence at the encoder. Let $d' = ||x^{\lambda n}(1) - x^{\lambda n}(2)|| = \sqrt{3\lambda n P}$ and the signal protection region

B_w for $x^{\lambda n}(w)$, $w \in [1:3]$, is defined as

$$\begin{aligned} B_w = \big\{ y^{\lambda n} \colon & ||x^{\lambda n}(w) - y^{\lambda n}|| \le ||x^{\lambda n}(w') - y^{\lambda n}|| \\ & \qquad \text{for } w' \ne w, \\ & \big| ||x^{\lambda n}(w') - y^{\lambda n}|| - ||x^{\lambda n}(w'') - y^{\lambda n}|| \big| \le td' \\ & \qquad \text{for } w', w'' \ne w \big\} \end{aligned} \tag{4.6}$$

which means that message w is the most probable and the other messages w' and w'' are of approximately equal posterior probabilities. Here $t \in [0, (\sqrt{3}-1)/2]$ is a fixed parameter which will be optimized later in the analysis.

Encoding. In stage 1, the encoder uses the same simplex signaling as in the noise-free feedback case (see (4.4)). Then based on the noisy feedback $\tilde{y}^{\lambda n}$, the encoder chooses $\tilde{w}_1$ and $\tilde{w}_2$ such that

$$||x^{\lambda n}(\hat{w}_1) - \tilde{y}^{\lambda n}|| \le ||x^{\lambda n}(\hat{w}_2) - \tilde{y}^{\lambda n}|| \le ||x^{\lambda n}(\hat{w}_3) - \tilde{y}^{\lambda n}||,$$

In stage 2, the encoder uses antipodal signaling for w if $w \in \{\tilde{w}_1, \tilde{w}_2\}$ and transmits all-zero sequence otherwise.

Decoding. The decoder makes a decision immediately at the end of stage 1 if the received signal lies in one of the signal protection regions, i.e., $y^{\lambda n} \in B_w$ for $w \in [1:3]$. Otherwise, it chooses the two most probable message estimates $\hat{w}_1$ and $\hat{w}_2$ and wait for the transmission in stage 2. At the end of stage 2, the decoder declares that $\hat{w}$ is sent if

$$\begin{aligned} \hat{w} &= \arg \min_{w \in \{\hat{w}_1, \hat{w}_2\}} ||x^n(w) - y^n|| \\ &= \arg \min_{w \in \{\hat{w}_1, \hat{w}_2\}} \big(||x^{\lambda n}(w) - y^{\lambda n}||^2 + ||x^n_{\lambda n+1}(w) - y^n_{\lambda n+1}||^2 \big)^{1/2}. \end{aligned}$$

Remark 4.3. *The signal protection region corresponds to the case in which the two least probable messages are of approximately equal posterior probabilities, i.e.,* $||x^{\lambda n}(w) - y^{\lambda n}|| \ll ||x^{\lambda n}(w') - y^{\lambda n}|| \approx ||x^{\lambda n}(w'') - y^{\lambda n}||$.

Analysis of the probability of error. Let $(\tilde{W}_1, \tilde{W}_2)$ and $(\hat{W}_1, \hat{W}_2)$ denote the pairs of the two most probable message estimates at the encoder and the decoder,

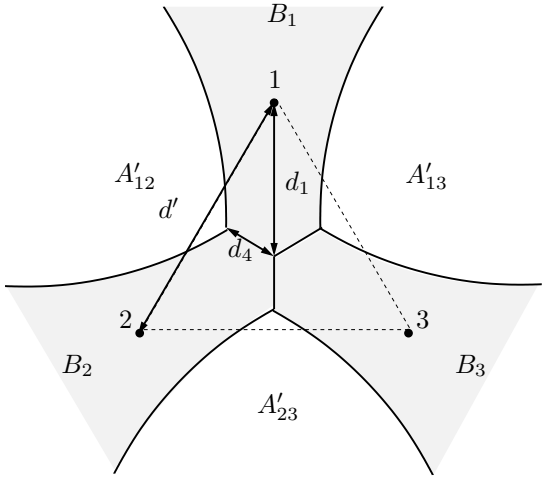


Figure 4.5: Signal protection regions. The shaded areas B_w for $w = 1, 2, 3$ are the signal protection regions for $x^{\lambda n}(1)$, $x^{\lambda n}(2)$, and $x^{\lambda n}(3)$, respectively. Here $d_4 = sd_1/2 = (s/2)\sqrt{\lambda nP}$ for some parameter $s = s(t) \in [0, 1]$ to be optimized later.

respectively. As before, we assume that $W = 1$ is sent. Referring to Fig. 4.5, let

$$A'_{ww'} = A_{ww'} \backslash (\cup_{w''} B_{w''}), \quad w, w' \in [1 : 3]$$

where

$$\begin{aligned} A_{ww'} = \{y^{\lambda n} \colon \max\{||y^{\lambda n} - x^{\lambda n}(w)||, ||y^{\lambda n} - x^{\lambda n}(w')||\} \\ \leq ||y^{\lambda n} - x^{\lambda n}(w'')||, \ w'' \neq w, w'\}. \end{aligned}$$

The decoder makes an error only if one or more of the following events occur:

- decoding error at the end of stage 1

$$\mathcal{E}_1 = \left\{Y^{\lambda n} \in B_2 \cup B_3 \cup A'_{23}\right\},$$

- miscoordination due to the feedback noise

$$\begin{aligned} \tilde{\mathcal{E}}_{12} &= \left\{Y^{\lambda n} \in A'_{12}, \tilde{Y}^{\lambda n} \in A_{13} \cup A_{23}\right\}, \\ \tilde{\mathcal{E}}_{13} &= \left\{Y^{\lambda n} \in A'_{13}, \tilde{Y}^{\lambda n} \in A_{12} \cup A_{23}\right\}, \end{aligned}$$

- decoding error at the end of stage 2

$$\mathcal{E}_2 = \big\{W \in \{\hat{W}_1, \hat{W}_2\} = \{\tilde{W}_1, \tilde{W}_2\} \text{ and } \hat{W} \neq W\big\}.$$

Thus, the probability of error is upper bounded as

$$\begin{aligned} P_e^{(n)} &\leq \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\tilde{\mathcal{E}}_{12}) + \mathsf{P}(\tilde{\mathcal{E}}_{13}) + \mathsf{P}(\mathcal{E}_2) \\ &= \mathsf{P}(\mathcal{E}_1) + 2\,\mathsf{P}(\tilde{\mathcal{E}}_{12}) + \mathsf{P}(\mathcal{E}_2). \end{aligned}$$

To simplify the analysis, we introduce a new parameter $s \in [0,1]$ such that $d_4 = sd_1/2 = (s/2)\sqrt{\lambda nP}$. It can be easily checked that $s \in [0,1]$ corresponds to $t \in [0, (\sqrt{3}-1)/2]$ and that this constraint guarantees that

$$d_5 = \min_{y^{\lambda n} \in A'_{23} \cup B_2 \cup B_3} ||x^{\lambda n}(1) - y^{\lambda n}|| \quad \text{(see Fig. 4.6(a))}.$$

Hence, for the first term

$$\begin{aligned} \mathsf{P}(\mathcal{E}_1) &= \mathsf{P}\left\{Y^{\lambda n} \in A'_{23} \cup B_2 \cup B_3\right\} \\ &\leq 2Q(d_5) \\ &\leq \exp\bigg(-\frac{\lambda nP}{8}(s^2 - 2s + 4)\bigg). \end{aligned} \tag{4.7}$$

The second term $\mathsf{P}(\tilde{\mathcal{E}}_{12})$ can be upper bounded (see Fig. 4.6(b)) as

$$\begin{aligned} \mathsf{P}(\tilde{\mathcal{E}}_{12}) &= \mathsf{P}\left\{Y^{\lambda n} \in A'_{12}, \tilde{Y}^{\lambda n} \in A_{13} \cup A_{23}\right\} \\ &\leq \mathsf{P}\left\{\tilde{Y}^{\lambda n} \in A_{13} \cup A_{23} | Y^{\lambda n} \in A'_{12}\right\} \\ &\leq 2Q\bigg(\frac{d_6}{\sqrt{\alpha}}\bigg) \\ &\leq \exp\bigg(-\frac{3s^2\lambda nP}{32\alpha}\bigg). \end{aligned} \tag{4.8}$$

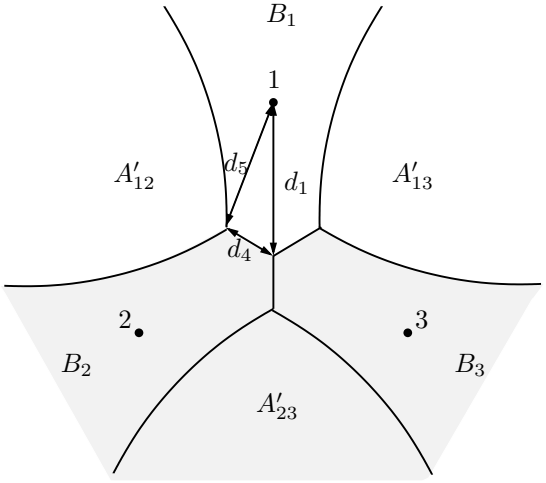
Finally, the third term $\mathsf{P}(\mathcal{E}_2)$ can be upper bounded in the exactly same manner as in the noise-free feedback case:

$$\mathsf{P}(\mathcal{E}_2) \leq \frac{1}{2}\exp\bigg(-\frac{1}{2}\bigg(1 - \frac{\lambda}{4}\bigg)nP\bigg).$$

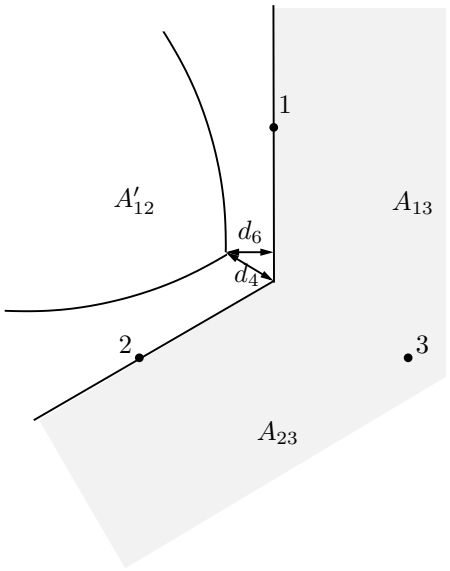
Therefore, the error exponent of the two-stage noisy feedback coding scheme is lower bounded as

$$E'_3(\alpha) = \limsup_{n\to\infty} -\frac{1}{n}\ln P_e^{(n)}$$

$$\geq \limsup_{n\to\infty} -\frac{1}{n}\max\{\ln \mathsf{P}(\mathcal{E}_1), \ln \mathsf{P}(\tilde{\mathcal{E}}_{12}), \ln \mathsf{P}(\mathcal{E}_2)\}$$
$$\geq \min\left\{ \frac{\lambda P}{8}(s^2-2s+4),\ \frac{3\lambda s^2 P}{32\alpha},\ \frac{P}{2}\left(1-\frac{1}{4}\lambda\right)\right\}.$$



(a)



(b)

Figure 4.6: (a) The error event $\mathcal{E}_1$ when $W = 1$. Since $0 \leq s \leq 1$, we have $d_5 = \sqrt{d_1^2 + d_4^2 - d_1 d_4} = \sqrt{(\lambda nP/4)(s^2 - 2s + 4)}$. (b) The error event $\tilde{\mathcal{E}}_{12}$ when $W = 1$ and $\{\tilde{W}_1, \tilde{W}_2\} = \{1, 3\}$. Here $d_6 = (\sqrt{3}/2)d_4 = s\sqrt{(3\lambda nP/16)}$.

Now let

$$\alpha = \alpha^*(s) = \frac{3s^2}{4(s^2 - 2s + 4)}$$

and

$$\lambda = \lambda^*(s) = \frac{4}{s^2 - 2s + 5}.$$

Then it can be readily verified that all the three terms in the minimum are the same and we have

$$E_3'(\alpha^*(s)) \geq \frac{P}{2}\frac{s^2 - 2s + 4}{s^2 - 2s + 5} =: \phi(s). \tag{4.9}$$

Note that if $s < 1$,

$$\phi(s) > \frac{3}{8}P = E_3(\infty),$$

and $\alpha^*(s)$ is monotonically increasing over $s \in [0, 1]$. Thus

$$E_3(\alpha) > E_3(\infty) \quad \text{for } \alpha < \alpha^*(1) = \frac{1}{4}.$$

This completes the proof of Theorem 1 for the $M = 3$ case.

Remark 4.4. *It can be easily checked that the lower bound in* (4.9) *is tight and characterizes the exact error exponent $E_3'(\alpha)$ of the two-stage noisy feedback coding scheme.*

4.3 Linear Noisy Feedback Coding Scheme

4.3.1 Background

It is instructive to revisit (a slightly simplified version of) the linear noise-free feedback coding scheme by Pinsker [Pin68], which shows that $E_M(0) \geq E_2(\infty) = P/2$ for all $M \geq 2$. This lower bound is tight since $E_2(0) = E_2(\infty)$ [SWWZ69] and $E_M(0)$ is nonincreasing in M.

Encoding. To send message $w \in [1 : M]$, the encoder transmits

$$X_1(w) = \begin{cases} \frac{L+1-w}{L}\sqrt{P} & \text{if } M = 2L + 1, \\ \frac{L+1/2-w}{L}\sqrt{P} & \text{if } M = 2L. \end{cases} \tag{4.10}$$

Because of the feedback Y_1, the encoder can learn the noise $Z_1 = Y_1 - X_1$. Subsequently it transmits

$$X_i = (1+\delta)Z_{i-1}, \quad i \in [2:\eta],$$

and $X_i = 0$ afterwards, where $\delta > 0$ will be optimized later and the random time $\eta = \eta(w, Z^n)$ is the largest $k \le \bar{n} = \sqrt{n}$ such that

$$\sum_{i=1}^{k} X_i^2 \le nP.$$

Decoding. Upon receiving Y^n, the decoder estimates X_1 by

$$\hat{X}_1 = \sum_{i=1}^{\bar{n}} (-1)^{i-1} \frac{Y_i}{(1+\delta)^{i-1}}$$

and declares that $\hat{w}$ is sent if

$$\hat{w} = \arg \min_{w \in [1:M]} |X_1(w) - \hat{X}_1|.$$

Remark 4.5. *It can be easily checked that each time $i \in [2:\eta]$, the encoder transmits the error*

$$\sum_{j=1}^{i-1} (-1)^{j-1} \frac{Y_j}{(1+\delta)^{j-1}} - X_1 = (-1)^{i-2} \frac{Z_{i-1}}{(1+\delta)^{i-2}}$$

in the decoder's current estimate of the initial transmission (up to scaling). Thus, Pinsker's coding scheme is another instance of iterative refinement used in the Schalkwijk-Kailath coding scheme [SK66] for the Gaussian channel and the Horstein coding scheme [Hor63] for the binary symmetric channel.

Analysis of the probability of error. For simplicity of notation, assume throughout that $\bar{n} = \sqrt{n}$ is an integer. We use ϵ_n to denote a generic sequence of nonnegative numbers that tends to zero as $n \to \infty$. When there are multiple such functions $\epsilon_n^{(1)}, \epsilon_n^{(2)}, \cdots, \epsilon_n^{(k)}$, we denote them all by ϵ_n with the understanding that $\epsilon_n = \max\{\epsilon_n^{(1)}, \epsilon_n^{(2)}, \cdots, \epsilon_n^{(k)}\}$. It is easy to see that decoding error occurs only if $|X_1(w) - \hat{X}_1| > \sqrt{P}/(2L)$. The probability of error is thus upper bounded as

$$P_{\mathrm{e}}^{(n)} = \mathsf{P}\{W \neq \hat{W}\} \le \mathsf{P}\left\{|X_1 - \hat{X}_1| > \frac{\sqrt{P}}{2L}\right\}.$$

The key idea in the analysis is to introduce a "virtual" transmission

$$X_i' = \begin{cases} X_1 & \text{if } i = 1, \\ (1+\delta)Z_{i-1} & \text{if } i \in [2:\bar{n}], \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

Let

$$Y_i' = X_i' + Z_i \tag{4.12}$$

and define the estimate $\hat{X}_1'$ of X_1' as

$$\hat{X}_1' = \sum_{i=1}^{\bar{n}} (-1)^{i-1} \frac{Y_i'}{(1+\delta)^{i-1}}. \tag{4.13}$$

Then, it can be easily shown that

$$\hat{X}_1' = X_1 + (-1)^{\bar{n}-1} \frac{Z_{\bar{n}}}{(1+\delta)^{\bar{n}-1}}.$$

Thus we have

$$\begin{aligned} &\mathsf{P}\left\{|X_1 - \hat{X}_1| > \frac{\sqrt{P}}{2L}\right\} \\ &\le \mathsf{P}\left\{|X_1 - \hat{X}_1'| + |\hat{X}_1' - \hat{X}_1| > \frac{\sqrt{P}}{2L}\right\} \\ &\le \mathsf{P}\left\{|X_1 - \hat{X}_1'| > \frac{\sqrt{P}}{2L}\right\} + \mathsf{P}\{|\hat{X}_1' - \hat{X}_1| > 0\} \\ &=: P_1 + P_2. \end{aligned}$$

Now we upper bound the two terms. For the first term, we have

$$\begin{aligned} P_1 &= \mathsf{P}\left\{\left|\frac{Z_{\bar{n}}}{(1+\delta)^{\bar{n}-1}}\right| > \frac{\sqrt{P}}{2L}\right\} \\ &= 2Q\left(\frac{\sqrt{P}(1+\delta)^{\bar{n}-1}}{2L}\right) \\ &\le \exp\left(-\frac{P(1+\delta)^{2(\bar{n}-1)}}{8L^2}\right). \end{aligned}$$

For the second term, note that $X_i = X_i'$ for all $i \in [1:n]$ if and only if $\sum_{i=1}^{\bar{n}} X_i^2 \leq nP$ (i.e., $\bar{n} = \eta$), and thus that $\hat{X}_1' \neq \hat{X}_1$ only if $\sum_{i=1}^{\bar{n}} X_i^2 > nP$. Therefore,

$$\begin{aligned} P_2 &\leq \mathsf{P}\bigg\{\sum_{i=1}^{\bar{n}} X_i^2 > nP\bigg\} \\ &\overset{(a)}{\leq} \mathsf{P}\bigg\{\sum_{i=2}^{\bar{n}} (1+\delta)^2 Z_{i-1}^2 > (n-1)P\bigg\} \\ &= \mathsf{P}\bigg\{\chi_{\bar{n}-1}^2 > \frac{(n-1)P}{(1+\delta)^2}\bigg\}, \end{aligned}$$

where (a) follows since $X_1^2 \leq P$ (recall (4.10)) and $\chi_{\bar{n}-1}^2$ denotes a chi-square random variable with $\bar{n}-1$ degrees of freedom. By upper bounding the tail probability of the chi-square random variable [IL06] as

$$\mathsf{P}\{\chi_k^2 > x\} \leq \exp\bigg(-\frac{x}{2} + \frac{k}{2}\log\frac{ex}{k}\bigg) \quad \text{for any } k \geq 1 \text{ and } x \geq k, \tag{4.14}$$

we have

$$\begin{aligned} P_2 &\leq \mathsf{P}\bigg\{\chi_{\bar{n}-1}^2 > \frac{(n-1)P}{(1+\delta)^2}\bigg\} \\ &\leq \exp\bigg(-\frac{1}{2}\frac{(n-1)P}{(1+\delta)^2} + \frac{\bar{n}-1}{2}\log\frac{e(n-1)P}{(\bar{n}-1)(1+\delta)^2}\bigg) \\ &\leq \exp\bigg(-\frac{1}{2}\frac{(n-1)P}{(1+\delta)^2} + \frac{\bar{n}-1}{2}\log\frac{e(n-1)P}{(\bar{n}-1)}\bigg) \\ &\leq \exp\bigg(-\frac{1}{2}\frac{nP}{(1+\delta)^2} + n\epsilon_n\bigg), \end{aligned}$$

where ϵ_n tends to zero as $n \to \infty$. Therefore, the error exponent of the linear feedback coding scheme is lower bounded as

$$\begin{aligned} E_M''(0) &\geq \limsup_{n\to\infty} -\frac{1}{n}\ln P_e^{(n)} \\ &= \limsup_{n\to\infty} -\frac{1}{n}\max\{\ln P_1, \ln P_2\} \\ &\geq \limsup_{n\to\infty} \min\bigg\{\frac{P(1+\delta)^{2(\bar{n}-1)}}{8nL^2},\ \frac{P}{2(1+\delta)^2}\bigg\}. \end{aligned}$$

for any $\delta > 0$. Now let

$$\delta = \delta(n) = \frac{\ln(4nL^2)}{2\bar{n}},$$

which tends to zero as $n \to \infty$. Then the limits of both terms in the minimum are the same. Therefore,

$$E_M''(0) \geq \limsup_{n\to\infty} \frac{P}{2(1+\delta(n))^2} = \frac{P}{2},$$

which completes the proof of achievability.

4.3.2 Coding Scheme and Performance Analysis

Now we formally describe and analyze a linear noisy feedback coding scheme based on Pinsker's noise-free feedback coding scheme.

Encoding. Fix some $\lambda \in (0,1)$. To send message $w \in [1:M]$, the encoder transmits

$$X_1(w) = \begin{cases} \frac{L+1-w}{L}\sqrt{\lambda n P} & \text{if } M = 2L+1, \\ \frac{L+1/2-w}{L}\sqrt{\lambda n P} & \text{if } M = 2L. \end{cases} \tag{4.15}$$

Because of the noisy feedback $\tilde{Y}_1$, the encoder can learn $Z_1 + \tilde{Z}_1 = \tilde{Y}_1 - X_1$. Subsequently it transmits

$$X_i = (1+\delta)(Z_{i-1} + \tilde{Z}_{i-1}), \quad i \in [2:\eta],$$

where $\delta > 0$ will be optimized later and the random time $\eta = \eta(w, Z^n, \tilde{Z}^n)$ is the largest $k \leq \bar{n} = \sqrt{n}$ such that

$$\sum_{i=1}^{k} X_i^2 \leq nP.$$

Decoding. Upon receiving Y^n, the decoder estimates X_1 by

$$\hat{X}_1 = \sum_{i=1}^{\bar{n}} (-1)^{i-1} \frac{Y_i}{(1+\delta)^{i-1}}$$

and declares that $\hat{w}$ is sent if

$$\hat{w} = \arg\min_{w\in[1:M]} |X_1(w) - \hat{X}_1|.$$

Remark 4.6. *The main difference between this noisy feedback coding scheme and Pinsker's noise-free feedback coding scheme in the previous subsection is that we let the power of the initial transmission grow linearly with the block length n (exploiting the peak energy constraint in* (4.1)*) and thus that the initial transmission contains much more information about the message than in Pinsker's scheme. This makes the coding scheme more robust to combat the noise in the feedback link.*

Analysis of the probability of error. As before we assume that $\bar{n} = \sqrt{n}$ is an integer. Let

$$X_i' = \begin{cases} X_1 & \text{if } i = 1, \\ (1+\delta)(Z_{i-1} + \tilde{Z}_{i-1}) & \text{if } i \in [2 : \bar{n}], \\ 0 & \text{otherwise,} \end{cases} \tag{4.16}$$

and let Y_i' and $\hat{X}_1'$ be defined as in (4.12) and (4.13). Then, it can be easily shown that

$$\hat{X}_1' = X_1 + (-1)^{\bar{n}-1} \frac{Z_{\bar{n}}}{(1+\delta)^{\bar{n}-1}} + \sum_{i=1}^{\bar{n}-1} (-1)^i \frac{\tilde{Z}_i}{(1+\delta)^{i-1}}.$$

Thus we have

$$\begin{aligned} P_{\mathrm{e}}^{(n)} &= \mathsf{P}\{W \neq \hat{W}\} \\ &\le \mathsf{P}\left\{ |X_1 - \hat{X}_1| > \frac{\sqrt{\lambda n P}}{2L} \right\} \\ &\le \mathsf{P}\left\{ |X_1 - \hat{X}_1'| + |\hat{X}_1' - \hat{X}_1| > \frac{\sqrt{\lambda n P}}{2L} \right\} \\ &\le \mathsf{P}\left\{ |X_1 - \hat{X}_1'| > \frac{\sqrt{\lambda n P}}{2L} \right\} + \mathsf{P}\{|\hat{X}_1' - \hat{X}_1| > 0\} \\ &=: P_1 + P_2. \end{aligned}$$

Now we upper bound the two terms. For the first term we have

$$\begin{aligned} P_1 &= \mathsf{P}\left\{ \left| (-1)^{\bar{n}-1} \frac{Z_{\bar{n}}}{(1+\delta)^{\bar{n}-1}} + \sum_{i=1}^{\bar{n}-1} (-1)^i \frac{\tilde{Z}_i}{(1+\delta)^{i-1}} \right| > \frac{\sqrt{\lambda n P}}{2L} \right\} \\ &= 2Q\left(\frac{\sqrt{\lambda n P / N}}{2L} \right) \\ &\le \exp\left(-\frac{\lambda n P}{8 L^2 N} \right), \end{aligned}$$

where

$$\begin{aligned} N &= \sum_{i=1}^{\bar{n}-1} \frac{\alpha}{(1+\delta)^{2(i-1)}} + \frac{1}{(1+\delta)^{2(\bar{n}-1)}} \\ &= \frac{\alpha\left(1 - \frac{1}{(1+\delta)^{2(\bar{n}-2)}}\right)}{1 - \frac{1}{(1+\delta)^2}} + \frac{1}{(1+\delta)^{2(\bar{n}-1)}} \\ &\le \frac{\alpha(1+\delta)^2}{(1+\delta)^2 - 1} + \epsilon_n, \end{aligned}$$

where ϵ_n tends to zero as $n \to \infty$. Thus

$$P_1 \le \exp\left(- \frac{\lambda n P}{8L^2} \left(\frac{\alpha(1+\delta)^2}{(1+\delta)^2 - 1} + \epsilon_n \right)^{-1} \right). \tag{4.17}$$

For the second term, we have

$$\begin{aligned} P_2 &\le \mathsf{P}\left\{ \sum_{i=1}^{\bar{n}} X_i^2 > nP \right\} \\ &\overset{(a)}{\le} \mathsf{P}\left\{ \sum_{i=2}^{\bar{n}} (1+\delta)^2 (Z_{i-1} + \tilde{Z}_{i-1})^2 > (1-\lambda) nP \right\} \\ &= \mathsf{P}\left\{ \chi^2_{\bar{n}-1} > \frac{(1-\lambda)nP}{(1+\delta)^2(1+\alpha)} \right\}, \end{aligned}$$

where (a) follows since $X_1 \le \lambda n P$ (recall (4.15)). By (4.14), we have

$$\begin{aligned} P_2 &\le \mathsf{P}\left\{ \chi^2_{\bar{n}-1} > \frac{(1-\lambda)nP}{(1+\delta)^2(1+\alpha)} \right\} \\ &\le \exp\left(- \frac{1}{2} \frac{(1-\lambda)nP}{(1+\delta)^2(1+\alpha)} \right. \\ &\qquad \left. + \frac{\bar{n}-1}{2} \log \frac{e(1-\lambda)nP}{(\bar{n}-1)(1+\delta)^2(1+\alpha)} \right) \\ &\le \exp\left(- \frac{1}{2} \frac{(1-\lambda)nP}{(1+\delta)^2(1+\alpha)} + n\epsilon_n \right), \end{aligned} \tag{4.18}$$

where ϵ_n tends to zero as $n \to \infty$. Therefore, the error exponent of the linear noisy feedback coding scheme is lower bounded as

$$\begin{aligned} E''_M(\alpha) &= \limsup_{n\to\infty} -\frac{1}{n} \ln P_e^{(n)} \\ &= \limsup_{n\to\infty} -\frac{1}{n} \max\{\ln P_1, \ln P_2\} \\ &\ge \min\left\{ \frac{\lambda P}{8L^2\alpha} \frac{(1+\delta)^2 - 1}{(1+\delta)^2}, \ \frac{(1-\lambda)P}{2(1+\delta)^2(1+\alpha)} \right\}. \end{aligned}$$

Now let

$$\delta = \delta(\alpha) = \left(1 + \sqrt{\frac{4L^2\alpha}{1+\alpha}}\right)^{1/2} - 1$$

and

$$\lambda = \lambda(\alpha) = \left(1 + \sqrt{\frac{1+\alpha}{4L^2\alpha}}\right)^{-1}.$$

Then it can be readily verified that both terms in the minimum are the same and we have

$$E''_M(\alpha) \geq \frac{P}{2} \frac{1}{1 + \alpha + 4(\lfloor M/2 \rfloor)^2\alpha + 4(\lfloor M/2 \rfloor)\sqrt{\alpha(1+\alpha)}},$$

which completes the proof of Theorem 2.

4.4 Discussion

When α is very small, the linear feedback coding scheme (which is optimal for noise-free feedback) outperforms the two-stage (nonlinear) feedback coding scheme. When α is relatively large, however, linear feedback coding scheme amplifies the feedback noise, while the two-stage scheme achieves a more robust performance via signal protection. While this dichotomy agrees with the usual engineering intuition, it would be aesthetically more pleasing if a single feedback coding scheme performs uniformly better over all ranges of α, and the search for such a coding scheme invites further investigation. We finally note that $\alpha^* = 1/4$ is the threshold for all M in the two-stage noisy feedback coding scheme (see the Appendix). In both schemes, the error exponents are strictly larger than those for the no feedback case only when α is sufficiently small. Thus it is natural to ask whether the noisy feedback is useful for all α or there exists a fundamental threshold beyond which noisy feedback becomes useless.

Following Yamamoto and Burnashev's work [BY10] on noisy feedback communication over the binary symmetric channel at positive rates, we can extend our result to a positive rate, i.e., $M = e^{nR}$ with $R > 0$. Let $E(R;\alpha)$ denote the maximum error exponent, namely, the reliability function. Although the $E(R;\infty)$

is not known for all $R \in [0, C]$ (see, e.g., [ABL00]), Shannon [Sha59] showed that

$$E(0+;\infty) := \lim_{R\to 0} E(R;\infty) = \frac{P}{4}.$$

We can easily adapt the analysis of our two-stage noisy feedback coding scheme in the Appendix to show that

$$\lim_{\alpha\to 0} E(0+;\alpha) = \frac{2}{7}P > E(0+;\infty).$$

Moreover, we have the following.

Proposition 4.1.

$$E(R;\alpha) > E(R;\infty) \quad \textit{for } R < \frac{P}{24} \textit{ and } \alpha < \alpha(s),$$

where $s \in [0, 1]$ *is the root of* $(s-1)^2 = 24R/P$.

Thus, the best error exponent can be strictly larger than the one without feedback if the rate and the feedback noise power are sufficiently small.

Finally, we note that our discussion has been limited to the peak energy constraint (4.1). In some practical systems, however, it would be more relevant to consider peak power constraints

$$\mathsf{P}\big\{x_i^2(w, \tilde{Y}^{i-1}) \le P\big\} = 1 \quad \text{for all } w \text{ and } i,$$

or

$$\mathsf{E}[x_i^2(w, \tilde{Y}^{i-1})] \le P \quad \text{for all } w \text{ and } i.$$

It remains to be seen whether noisy feedback still improves the reliability under these more stringent conditions.

4.5 Technical Proofs

4.5.1 Proof of Theorem 1 for the General Case

Encoding. In stage 1, the encoder uses the simplex signaling for an M-ary message:

$$x^{\lambda n}(w) = A\bigg(e_w - \frac{1}{M}\sum_{w=1}^{M} e_w\bigg) \quad \text{for } w \in [1:M],$$

where $A = \sqrt{M\lambda nP/(M-1)}$ and

$$e_w = (\underbrace{0, \cdots, 0}_{w-1}, 1, 0, \cdots, 0).$$

Then based on the noisy feedback $\tilde{y}^{\lambda n}$, the encoder chooses the two most probable message estimates $\tilde{w}_1$ and $\tilde{w}_2$ among M candidates. In stage 2, the encoder uses antipodal signaling for w if $w \in \{\tilde{w}_1, \tilde{w}_2\}$ and transmits all-zero sequence otherwise.

Decoding. The signal protection region for the M-ary message is defined as in (4.6) (with w, w', $w''\in [1:M]$). The decoder makes a decision immediately at the end of stage 1 if the received signal $y^{\lambda n}$ lies in one of the signal protection regions. Otherwise, it chooses the two most probable message estimates $\hat{w}_1$ and $\hat{w}_2$, and wait for the transmission in stage 2. At the end of stage 2, the decoder declares that $\hat{w}$ is sent if

$$\hat{w} = \arg \min_{w \in \{\hat{w}_1, \hat{w}_2\}} \left(||x^{\lambda n}(w) - y^{\lambda n}||^2 + ||x_{\lambda n+1}^n(w) - y_{\lambda n+1}^n||^2 \right)^{1/2}.$$

Analysis of the probability of error. Let $(\tilde{W}_1, \tilde{W}_2)$ and $(\hat{W}_1, \hat{W}_2)$ denote the pairs of the two most probable message estimates at the encoder and the decoder, respectively. The decoder makes an error only if one or more of the following events occur:

- decoding error at the end of stage 1

$$\mathcal{E}_1 = \left\{ Y^{\lambda n} \in \cup_{w \neq 1} B_w \cup (\cup_{w, w' \neq 1} A'_{ww'}) \right\},$$

- miscoordination due to the feedback noise

$$\tilde{\mathcal{E}}_{1w} = \left\{ Y^{\lambda n} \in A'_{1w} \text{ and } \tilde{Y}^{\lambda n} \in \cup_{\{w', w''\} \neq \{1, w\}} A_{w'w''} \right\},$$

- decoding error at the end of stage 2

$$\mathcal{E}_2 = \left\{ W \in \{\hat{W}_1, \hat{W}_2\} = \{\tilde{W}_1, \tilde{W}_2\} \text{ and } \hat{W} \neq W \right\}.$$

Thus, the probability of error is upper bounded as

$$P_e^{(n)} \leq \mathsf{P}(\mathcal{E}_1) + M\, \mathsf{P}(\tilde{\mathcal{E}}_{1w}) + \mathsf{P}(\mathcal{E}_2).$$

As before, we assume that $W = 1$ was sent. For the first term, by the union of events bound,

$$\begin{aligned}\mathsf{P}(\mathcal{E}_1) &= \mathsf{P}\left\{Y^{\lambda n} \in \cup_{w\neq 1} B_w \cup (\cup_{w,w'\neq 1} A'_{ww'})\right\} \\ &\leq M^2\, \mathsf{P}\left\{Y^{\lambda n} \in B_2 \cup A'_{23}\right\}.\end{aligned}$$

For $\mathsf{P}(\tilde{\mathcal{E}}_{1w})$, again by the union of events bound,

$$\begin{aligned}\mathsf{P}(\tilde{\mathcal{E}}_{1w}) &= \mathsf{P}\left\{Y^{\lambda n} \in A'_{1w} \text{ and } \tilde{Y}^{\lambda n} \in \cup_{\{w',w''\}\neq\{1,w\}} A_{w'w''}\right\} \\ &\leq M^2\, \mathsf{P}\left\{Y^{\lambda n} \in A'_{1w} \text{ and } \tilde{Y}^{\lambda n} \in A_{w'w''}\right\}.\end{aligned}$$

We use d'_j, $j \in [1:6]$, to denote the distances corresponding to d_j in the $M = 3$ case (see Fig. 4.6). It can be easily checked that $d'_j = d_j\sqrt{3(M-1)/(2M)}$. Thus by replacing d_5 by d'_5 in (4.7) and d_6 by d'_6 in (4.8), we have

$$\begin{aligned}\mathsf{P}(\mathcal{E}_1) &\leq M^2 Q(d'_5) \\ &\leq \frac{M^2}{2}\exp\left(-\frac{M}{12(M-1)}\lambda n P(s^2-2s+4)\right)\end{aligned}$$

and

$$\mathsf{P}(\tilde{\mathcal{E}}_{12}) \leq M^2 Q\left(\frac{d'_6}{\sqrt{\alpha}}\right) \leq \frac{M^2}{2}\exp\left(-\frac{s^2 M}{16(M-1)\alpha}\lambda n P\right).$$

The third term $\mathsf{P}(\mathcal{E}_2)$ can be upper bounded in the same manner as for the $M = 3$ case,

$$\begin{aligned}\mathsf{P}(\mathcal{E}_2) &= Q\left(-\sqrt{\left(1-\frac{M-2}{2(M-1)}\lambda\right)nP}\right) \\ &\leq \frac{1}{2}\exp\left(-\frac{nP}{2}\left(1-\frac{M-2}{2(M-1)}\lambda\right)\right).\end{aligned}$$

Therefore,

$$\begin{aligned}E'_M(\alpha) &= \limsup_{n\to\infty} -\frac{1}{n}\ln P_e^{(n)} \\ &\geq \limsup_{n\to\infty} -\frac{1}{n}\max\{\ln \mathsf{P}(\mathcal{E}_1), \ln\big(M\,\mathsf{P}(\tilde{\mathcal{E}}_{12})\big), \ln \mathsf{P}(\mathcal{E}_2)\} \\ &\geq \min\Bigg\{\frac{\lambda MP}{12(M-1)}(s^2-2s+4),\ \frac{s^2\lambda MP}{16(M-1)\alpha}, \\ &\qquad\qquad \frac{P}{2}\left(1-\frac{M-2}{2(M-1)}\lambda\right)\Bigg\}.\end{aligned}$$

Now let

$$\alpha = \alpha^*(s) = \frac{3s^2}{4(s^2 - 2s + 4)}$$

and

$$\lambda = \lambda^*(s) = \left(\frac{M}{6(M-1)}(s^2 - 2s + 4) + \frac{M-2}{2(M-1)} \right)^{-1}.$$

Then it can be readily verified that all the three terms in the minimum are the same and we have

$$\begin{aligned} E'_M(\alpha^*(s)) &\geq \frac{P}{2}\left(1 - \frac{3(M-2)}{M(s^2-2s+4)+3(M-2)}\right) \\ &=: \phi(s). \end{aligned}$$

Note that if $s < 1$,

$$\phi(s) > \frac{M}{4(M-1)}P = E_M(\infty),$$

and $\alpha^*(s)$ is monotonically increasing over $s \in [0, 1]$. Thus

$$E'_M(\alpha) > E_M(\infty) \quad \text{for } \alpha < \alpha^*(1) = \frac{1}{4}.$$

This completes the proof of Theorem 1 for the general case.

Remark 4.7. *Note that $E'_M(\alpha)$ is decreasing in M, while $\alpha^*(s)$ is still independent of M.*

4.5.2 Proof of Proposition 4.1

Following the analysis in the Appendix from the chapter and replacing M with e^{nR}, the error exponent of the two-stage feedback coding scheme is lower bounded as

$$E'(R;\alpha) \geq \min\left\{ -2R + \frac{\lambda P}{12}(s^2 - 2s + 4),\ -3R + \frac{s^2\lambda P}{16\alpha},\ \frac{P}{2}\left(1 - \frac{\lambda}{2}\right) \right\}.$$

Now let

$$\alpha = \alpha^*(s, R) = \frac{3s^2(P + 4R)}{4((s^2 - 2s + 4)P + 6(s^2 - 2s + 5)R)}$$

and

$$\lambda = \lambda^*(s, R) = \frac{6(P + 4R)}{P(s^2 - 2s + 7)}.$$

Then it can be readily verified that all the three terms in the minimum are the same and we have

$$E(R';\alpha^*(s,R)) \geq \frac{s^2-2s+4}{2(s^2-2s+7)}P - \frac{6R}{s^2-2s+7}.$$

Thus

$$E'(0+;0) = \frac{2}{7}P > \frac{1}{4}P = E(0+;\infty).$$

Moreover, we have

$$E(R';\alpha) > E(0+;\infty) \geq E(R;\infty) \text{ for } \frac{(s-1)^2}{R} > \frac{24}{P}.$$

Note that $\alpha^*(s,R)$ is monotonically increasing over $s \in [0,1]$ for fixed R. We have,

$$E(R';\alpha) > E(R;\infty) \text{ for } R < \frac{P}{24} \text{ and } \alpha < \alpha(s_0),$$

where $s_0 \in [0,1]$ is the root of $(s-1)^2 = 24R/P$.

Chapter 4, in part, includes the material in Yu Xiang and Young-Han Kim, "On the AWGN channel with noisy feedback and peak energy constraint," *IEEE International Symposium on Information Theory*, pp. 256–259, Austin, TX, June 2010, and Yu Xiang and Young-Han Kim, "Gaussian channel with noisy feedback and peak energy constraint," *IEEE Transaction on Information Theory,* vol.59, no.8, pp.4746–4756, August 2013. The dissertation author was the primary investigator and author of this paper.

Chapter 5

Interactive Relaying over Networks

In this chapter, we studies the problem of broadcasting a common message over a relay network as the canonical platform to investigate the utilities and limitations of traditional relay coding schemes. For a few special classes of networks, such as the 3-node relay channel and the 4-node diamond network, the decode–forward coding scheme by Cover and El Gamal, and its generalization to networks by Xie and Kumar, and Kramer, Gastpar, and Gupta achieve the cutset bound, establishing the capacity. When the network has cycles, however, decode–forward is suboptimal in general and is outperformed by partial decode–forward, compress–forward, or more generally, interactive relaying built upon these $*$–forward coding schemes. In particular, it is demonstrated via a simple example that a coding scheme based on interactive computing by Orlitsky and Roche, and its infinite-round generalization by Ma and Ishwar can strictly outperform existing noninteractive or finite-round interactive coding schemes. Roughly speaking, when the network is to be flooded with information, it is more efficient for the relays to spray tiny droplets of the information back and forth than to splash a huge amount at a time.

5.1 Introduction

Consider the discrete memoryless network (DMN) model $(\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_N, p(y^N|x^N), \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_N)$ that consists of N sender-receiver alphabet pairs $(\mathcal{X}_k, \mathcal{Y}_k)$, $k \in [1:N] := \{1, 2 \ldots, N\}$, and a collection of channel conditional pmfs (probability mass functions) $p(y^N|x^N) := p(y_1, y_2, \ldots, y_N|x_1, x_2, \ldots, x_N)$. Suppose that source node 1 wishes to communicate a common message M to the rest of the network, as depicted in Figure 5.1. Compared to the unicast (one node wishes to recover M) or multicast (some nodes wishes to recover M), this problem is relatively simpler since every node in the network has the symmetric goal of recovering the same message.

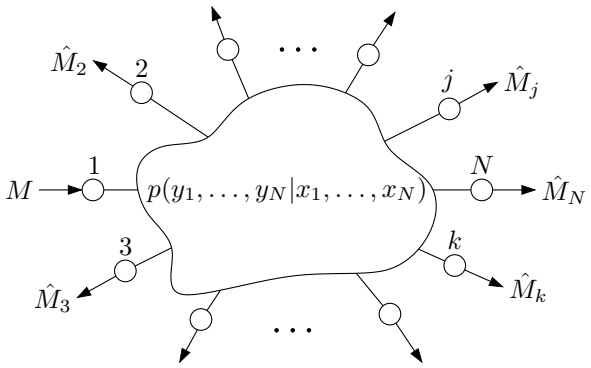


Figure 5.1: Common message broadcasting over a noisy network.

When the nodes in the network cannot adapt their transmissions based on its received sequence (that is, no relaying or feedback is allowed), then the problem reduces to common message communication over a broadcast channel and the capacity is

$$C_{\mathrm{BC}} = \max_{p(x_1), x_2, \ldots, x_N} \min_{k \in [2:N]} I(X_1; Y_k).$$

Now suppose that each node in the network can adapt its transmission based on the received sequence (that is, relaying is allowed) and thus help other nodes recover the message as well. Despite its relative simplicity, this problem still captures the essential richness of relaying over networks. This chapter attempts to demonstrate the inherent complexity in relaying by studying the information flow

questions on broadcasting:

- What is the capacity?
- What is the optimal relaying coding scheme that achieves the capacity?

We are now ready to define the common-message broadcasting problem formally. A $(2^{nR}, n)$ broadcast code for the DMN $p(y^N|x^N)$ consists of

- a message set $[1 : 2^{nR}]$,
- a source encoder that assigns a symbol $x_{1i}(m, y_1^{i-1})$ to each message $m \in [1 : 2^{nR}]$ and received sequence y_1^{i-1} for $i \in [1 : n]$,
- a set of relay encoders, where encoder $k \in [2 : N]$ assigns a symbol $x_{ki}(y_k^{i-1})$ to every received sequence y_k^{i-1} for $i \in [1 : n]$, and
- a set of decoders, where decoder $k \in [2 : N]$ assigns $\hat{m}_k$ to each y_k^n.

We assume that the message M is uniformly distributed over the message set. The average probability of error is defined as $P_e^{(n)} = \mathsf{P}\{\hat{M}_k \neq M \text{ for some } k \in [2 : N]\}$. A rate R is said to be achievable if there exists a sequence of $(2^{nR}, n)$ broadcast codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$. The broadcast capacity of the DMN is the supremum of all achievable rates.

El Gamal [EG81] established the cutset upper bound on the capacity:

$$C \leq \max_{p(x^N)} \min_{k\in[2:N]} \min_{\mathcal{S}:\, 1\in\mathcal{S},\, k\in\mathcal{S}^c} I(X(\mathcal{S}); Y(\mathcal{S}^c) \mid X(\mathcal{S}^c)). \tag{5.1}$$

Xie and Kumar [XK05] and Kramer, Gastpar, and Gupta [KGG05] generalized the decode–forward coding scheme by Cover and El Gamal [CEG79] and established the network decode–forward lower bound on the capacity:

$$C \geq \max_{p(x^N)} \min_{k\in[1:N-1]} I(X^k; Y_{k+1}|X_{k+1}^N). \tag{5.2}$$

These two bounds coincide and establish the broadcast capacity when the network is degraded, i.e.,

$$p(y_{k+2}^N|x^N, y^{k+1}) = p(y_{k+2}^N|x_{k+1}^N, y_{k+1}) \tag{5.3}$$

for $k \in [1 : N-2]$ (up to relabeling of nodes).

In the following section, we discuss two other special cases—3-node relay channels and layered networks—for which the decode–forward lower bound is tight. Decode–forward, however, is suboptimal for general networks. We demonstrate gradually through simple examples that optimal relaying can be more sophisticated than simple decode–forward and require partial decode–forward, compress–forward, or interactive relaying built upon these $*$–forward coding schemes. Our discussion will culminate with the binary broadcast relay channel example for which not only interactive communication between relays strictly outperforms the existing noninteractive coding schemes, but also the number of communication rounds needs to go to infinity to fully enjoy the benefit of interaction.

5.2 Formulation and Existing Schemes

5.2.1 Decode–Forward

It is already mentioned that the decode–forward coding scheme is optimal when the network is degraded; see (5.3). Another case in which decode–forward is natural is when the network is *acyclic*, i.e.,

$$p(y^N|x^N) = \prod_{k=1}^{N} p(y_k|x^k, y^{k-1})$$

(up to relabeling of nodes). For this case, node k does not receive any signal from its downstream (nodes $j \in [k+1 : N]$). Thus it is natural to decode its received signal at once and forward the recovered message downstream. In the following, we revisit a few special classes of acyclic networks for which this decode–forward coding scheme is optimal.

We first consider the *relay channel* $p(y_2, y_3|x_1, x_2)$ [vdM71, CEG79] depicted in Figure 5.2. It is well known that *decode–forward* is optimal and the capacity is

$$C = \max_{p(x_1,x_2)} \min\big\{I(X_1; Y_2|X_2),\, I(X_1, X_2; Y_3)\big\}.$$

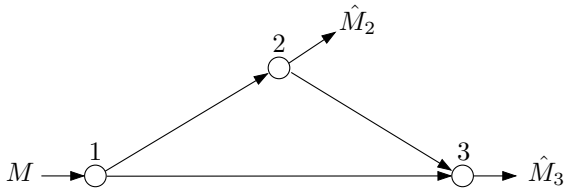


Figure 5.2: Relay channel.

Consider the diamond network $p(y_2, y_3|x_1)p(y_4|x_2, x_3)$ [SG00] depicted in Figure 5.3. Again, decode–forward is optimal and the capacity is

$$C = \max_{p(x_1)p(x_2,x_3)} \min\big\{I(X_1;Y_2), I(X_1,Y_3),\ I(X_2,X_3;Y_4)\big\}.$$

To prove the converse, simplify the cutset bound in (5.1) as

$$\begin{aligned} C &\le \max_{p(x^3)} \min\big\{I(X_1,X_3;Y_2|X_2),\ I(X_1,X_2;Y_3|X_3), \\ &\qquad\qquad I(X_1,X_2,X_3;Y_4)\big\} \\ &\overset{(a)}{\le} \max_{p(x^3)} \min\big\{I(X_1;Y_2),\ I(X_1;Y_3),\ I(X_2,X_3;Y_4)\big\} \\ &\overset{(b)}{=} \max_{p(x_1)p(x_2,x_3)} \min\big\{I(X_1;Y_2),\ I(X_1;Y_3),\ I(X_2,X_3;Y_4)\big\}, \end{aligned}$$

where (a) follows since $(X_2, X_3) \to X_1 \to Y_2$, $(X_2, X_3) \to X_1 \to Y_3$, and $X_1 \to (X_2, X_3) \to Y_4$, respectively, form Markov chains, and (b) follows since the mutual information terms $I(X_1;Y_2), I(X_1;Y_3)$, and $I(X_2, X_3;Y_4)$ depend on the channel input pmf $p(x_1, x_2, x_3)$ only through the marginals $p(x_1)$ and $p(x_2, x_3)$. The achiev-

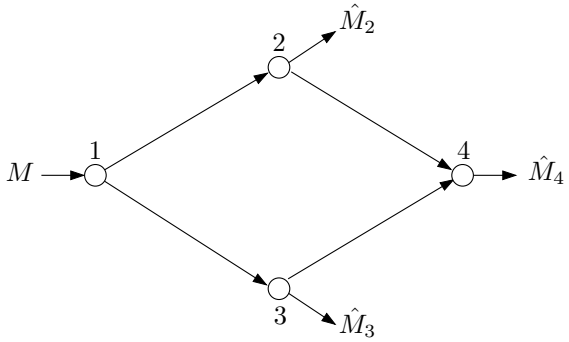


Figure 5.3: Diamond network.

ability follows by simplifying the decode-forward lower bound in (5.2) as

$$\begin{aligned} C &\geq \max_{p(x^3)} \min\big\{ I(X_1; Y_2|X_2, X_3),\ I(X_1, X_2; Y_3|X_3), \\ &\qquad\qquad I(X_1, X_2, X_3; Y_4)\big\} \\ &\overset{(a)}{\geq} \max_{p(x_1)p(x_2,x_3)} \min\{ I(X_1; Y_2|X_2, X_3), I(X_1; Y_3|X_3), \\ &\qquad\qquad I(X_2, X_3; Y_4)\} \\ &\overset{(b)}{\geq} \max_{p(x_1)p(x_2,x_3)} \min\big\{ I(X_1; Y_2),\ I(X_1; Y_3),\ I(X_2, X_3; Y_4)\big\}, \end{aligned}$$

where (a) follows since the maximum is over a smaller set and (b) follows since X_1 is independent of (X_2, X_3).

This result can be easily generalized to *layered network*

$$p(y^N|x^N) = \prod_{l=1}^{\lambda} p(y(\mathcal{L}_l)|x(\mathcal{L}_{l-1}))$$

depicted in Figure 5.4, where the layers of nodes $\mathcal{L}_0 = \{1\}$ and $\mathcal{L}_j$, $j \in [1 : \lambda]$ partition the network, i.e.,

$$\mathcal{L}_0 \uplus \mathcal{L}_1 \uplus \cdots \uplus \mathcal{L}_\lambda = [1 : N].$$

The capacity of the layered network is

$$C = \max_{\prod_{l=1}^{\lambda} p(x(\mathcal{L}_{l-1}))} \min_{l \in [1:\lambda]} \min_{j \in \mathcal{L}_l} I(X(\mathcal{L}_{l-1}); Y_j).$$

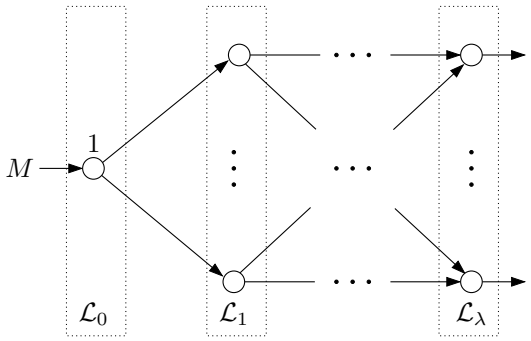


Figure 5.4: Layered network.

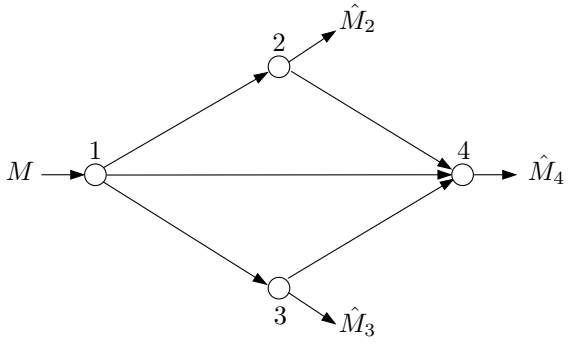


Figure 5.5: Diamond network with direct link.

Now we consider the variant of the diamond network depicted in Figure 5.5, which is defined as

$$p(y_2, y_3, y_4|x_1, x_2, x_3) = p(y_2, y_3|x_1)p(y_4|x_1, x_2, x_3).$$

For this case, the cutset bound simplifies to

$$C \leq \max_{p(x^3)} \min\big\{ I(X_1; Y_2|X_2),\ I(X_1; Y_3|X_3),\\ I(X_1, X_2, X_3; Y_4),\ I(X_1; Y_2, Y_3|X_2, X_3)\big\},$$

while the decode–forward lower bound simplifies to

$$C \geq \max_{p(x^3)} \min\big\{ I(X_1; Y_2|X_2, X_3),\ I(X_1; Y_3|X_3),\\ I(X_1, X_2, X_3; Y_4)\big\}.$$

Thus, it is not known whether decode–forward is optimal for acyclic networks in general, even though it seems to be the only reasonable coding scheme when there is no cycle in the information flow.

5.2.2 Partial Decode–Forward

In general, when the network has cycles, it is more advantageous to recover only part of the message at the beginning and recover the rest with the help of other nodes. This idea is best explained by a 3-node *cyclic graphical network* example depicted in Figure 5.6. Here the network is modeled by a

weighted directed cyclic graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} = \{1, 2, 3\}$ is the set of nodes, $\mathcal{E} = \{(1, 2), (1, 3), (2, 3), (3, 2)\}$ is the set of edges, each of which models an orthogonal communication link that can carry 1 bit per transmission. Note that the corresponding conditional pmf $p(y^3|x^3)$ is given by $X_1 = (X_{12}, X_{13})$, $Y_2 = (X_{12}, X_3)$ and $Y_3 = (X_{13}, X_2)$, where X_{12}, X_{13}, X_2, and X_3 are binary.

It can be easily verified that the cutset bound simplifies to $C \leq 2$ and the decode–forward lower bound simplifies to $C \geq 1$. But by simply *routing* one bit along the path $1 \to 2 \to 3$ and another bit along the path $1 \to 3 \to 2$, we can easily achieve 2 bits per transmission.

This observation can be readily generalized to any graphical networks, for which the capacity is achieved by routing as in the unicast case [FF56, EFS56]. Note that unlike the multicast case, network coding [ACLY00] is unnecessary for broadcasting. When the network suffers noise, the partial decode–forward coding scheme by Cover and El Gamal [CEG79] and its extension to networks by Aref [Are80] provide a means of splitting the message into independent parts and forwarding them along multiple paths.

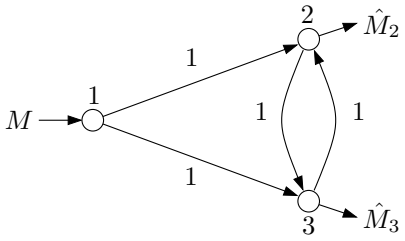


Figure 5.6: Cyclic graphical network.

5.2.3 Compress–Forward

In some cases, decoding is actually impossible at the beginning and more sophisticated coding schemes are necessary. To illustrate the depth of this problem, throughout the rest of the chapter we focus on a simple 3-node cyclic network model depicted in Figure 5.7, which is commonly referred to as the *broadcast relay channel*. Here the message is sent over a broadcast channel $p(y_2, y_3|x_1)$. In

addition, nodes 2 and 3 are connected via two noiseless links of rates R_2 and R_3, respectively, that are orthogonal to the main broadcast channel. Let $C(R)$ be the broadcast capacity as a function of the sum $R = R_2 + R_3$ of the link capacities between nodes 2 and 3.

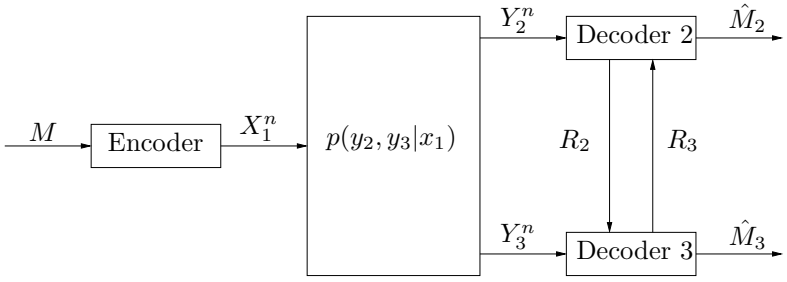


Figure 5.7: Broadcast relay channel.

To be more specific, we consider the Gaussian broadcast relay channel depicted in Figure 5.8. The channel outputs corresponding to the input X_1 are

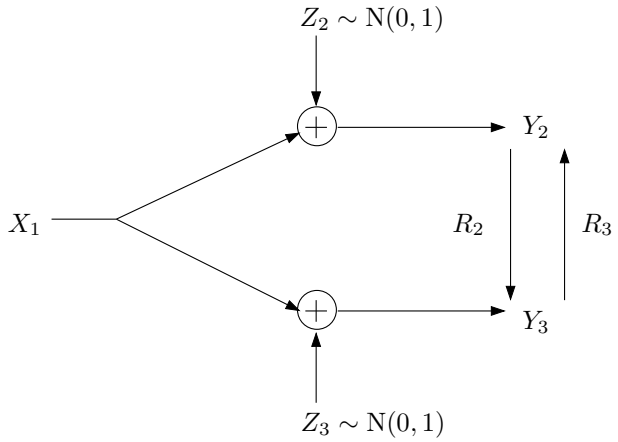


Figure 5.8: Gaussian broadcast relay channel.

$$\begin{aligned} Y_2 &= X_1 + Z_2, \\ Y_3 &= X_1 + Z_3, \end{aligned} \tag{5.4}$$

where Z_2 are Z_3 are jointly Gaussian with zero mean, equal variance $\mathsf{E}(Z_2^2) = \mathsf{E}(Z_3^2) = 1$, and correlation coefficient $\rho = \mathsf{E}(Z_2 Z_3)$. Note that the capacity without

the two noiseless links between the two receivers is

$$C(0) = \frac{1}{2}\log(1+P).$$

In the following, we focus on the case of $\rho = 0$.

By the cutset bound, the capacity is upper bounded as

$$C(R) \le C(0) + \frac{R}{2}, \tag{5.5}$$

where the optimal $R_2 = R_3 = R/2$ by symmetry. In comparison, since both receivers are symmetric, i.e., $F_{Y_2|X_1}(y|x) = F_{Y_3|X_1}(y|x)$, one recovers exactly what the other can recover about the message. Thus, any decoding-based relaying scheme (decode–forward, partial decode–forward, or compute–forward [NG07]) cannot achieve more than $C(0)$, which tends to zero as $P \to 0$.

Now we consider the compress–forward coding scheme for the relay channel by Cover and El Gamal [CEG79], which can be readily extended to the current setup. It can be easily shown [Kim07] that the corresponding lower bound (with the optimal rate splitting $R_2 = R_3 = R/2$) simplifies to

$$C(R) \ge \max_{F(x_1)F(\hat{y}_2|y_2)F(\hat{y}_3|y_3)} \min\{I_1,\, I_2,\, I_3,\, I_4\}, \tag{5.6}$$

where

$$\begin{aligned}
I_1 &= I(X_1; Y_2, \hat{Y}_3),\\
I_2 &= I(X_1; \hat{Y}_2, Y_3),\\
I_3 &= I(X_1; Y_2) - I(Y_3; \hat{Y}_3|X_1, Y_2) + R/2,\\
I_4 &= I(X_1; Y_3) - I(Y_2; \hat{Y}_2|X_1, Y_3) + R/2.
\end{aligned}$$

Evaluated with the Gaussian input distribution and test channels, this lower bound simplifies to

$$C(R) \ge \frac{1}{2}\log\left(1 + \frac{2P(P+1)(2^R-1) + P(2P+1)}{(P+1)(2^R-1) + (2P+1)}\right),$$

which is strictly larger than $C(0)$ for every $R > 0$. Thus, compress–forward strictly improves upon decoding-based relaying schemes. Note that when $\rho = -1$, the corresponding compress–forward lower bound coincides with the cutset bound in (5.5). This lower bound can be also achieved by the hash–forward coding scheme [CK07, Kim08].

5.3 Interactive Relaying

In the relay coding schemes we have discussed so far—(partial) decode–forward, compress–forward, compute–forward, hash–forward, each node summarizes its received signal and forwards it to other nodes. It turns out, however, that interactive cooperation between nodes can achieve higher rates, as demonstrated by Draper, Frey, and Kschschang [DFK03] for the broadcast relay channel consisting of two binary erasure channels.

In this section, we adapt their interactive relaying scheme to the Gaussian broadcast relay channel in (5.4) studied in the previous section. Suppose that node 2 first uses compress–forward to help node 3 recover the message and node 3 then uses decode–forward to help node 2 recover the message. It can be easily shown that this "compress–forward-followed-by-decode–forward" coding scheme yields the following lower bound on the capacity:

$$C(R) \geq \max_{F(x_1)F(\hat{y}_2|y_2)} \min\{I_2,\, I_3'\}, \tag{5.7}$$

where

$$\begin{aligned} I_2 &= I(X_1; \hat{Y}_2, Y_3), \\ I_3' &= I(X_1; Y_2) - I(Y_2; \hat{Y}_2|Y_3) + R. \end{aligned}$$

By symmetry, it can be shown that this lower bound strictly improves upon the compress–forward lower bound in (5.6). Thus, two-round interactive relaying is sometimes better than noninteractive relaying. When evaluated with the Gaussian input distribution and test channels, the lower bound in (5.7) simplifies to

$$\begin{aligned} C(R) \geq \max_{\sigma^2} \min\Bigg\{ & \frac{1}{2}\log\left(1 + \frac{2P + P\sigma^2}{1+\sigma^2}\right), \\ & R + \frac{1}{2}\log(1+P) - \frac{1}{2}\log\left(1 + \frac{2P+1}{(P+1)\sigma^2}\right)\Bigg\}. \end{aligned}$$

Figure 5.9 compares the cutset bound and the (partial) decode–forward, compress–forward, and compress–forward decode–forward lower bounds.

As shown in the previous section, interactive relaying can outperform noninteractive $*$–forward coding schemes. It is natural to ask the following:

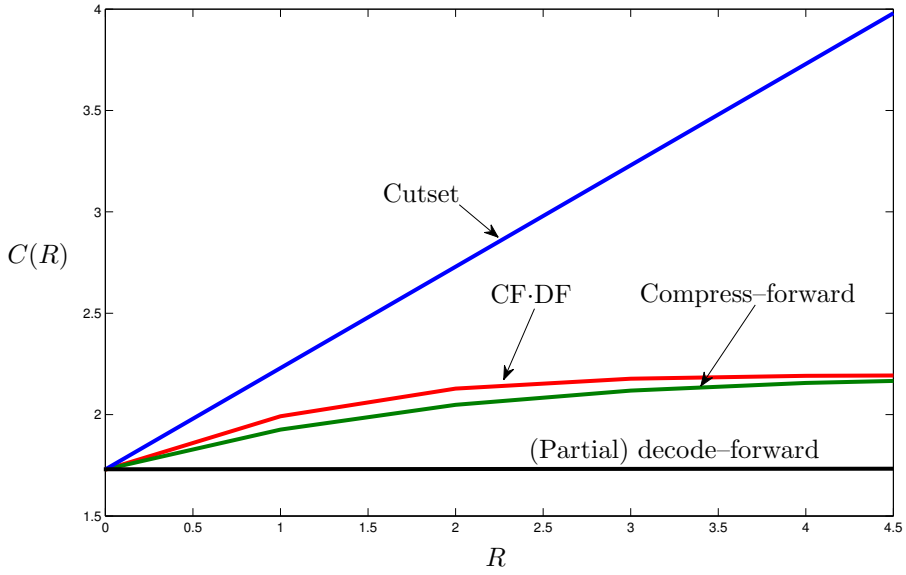


Figure 5.9: Comparison of the capacity bounds for the Gaussian broadcast relay channel when $P = 10$.

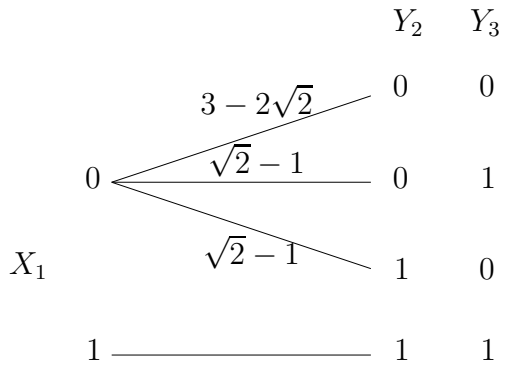
- Would more than two rounds of interactive relaying further outperform two rounds of interactive relaying?
- If so, how many rounds would be necessary?

In this section, we study a simple binary broadcast relay channel that consists of two correlated Z channels as depicted in Figure 5.10, and show that infinite rounds of interactive relaying can strictly outperform known finite-round coding schemes.

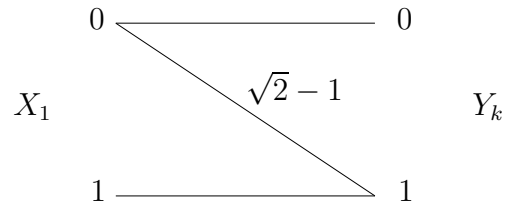
As before, we focus on the capacity $C(R)$ as a function of the sum-rate R of communication between two receivers. In particular, we will focus on the optimal rate of interaction

$$R^* = \min\{R\colon C(R) = 1\}.$$

It is easy to see that $C(0) = 0.3941$, which is the capacity of the Z channel, while $C(R) = 1$ for $R \geq 2$, which is the capacity of the DMC from X_1 to (Y_2, Y_3). In other words, $R^* \leq 2$. Note further that $X_1 = Y_2 \cdot Y_3$ and that when $X \sim \text{Bern}(1/2)$, Y_2 and Y_3 are independent and identically distributed $\text{Bern}(1/\sqrt{2})$.



(a) $p(y_2, y_3|x_1)$



(b) $p(y_k|x_1)$, $k = 2, 3$

Figure 5.10: Two correlated Z channels.

We now compare the existing bounds on the capacity. First, the cutset bound simplifies (under the optimal choice $R_2 = R_3 = R/2$) to

$$\begin{aligned} C(R) &\le \max_{p(x_1)} \min\{I(X_1; Y_2) + R/2,\, I(X_1; Y_2, Y_3)\} \\ &= \max_{\alpha \in [0:1]} \min\{H((2-\sqrt{2})\alpha) - \alpha H(\sqrt{2}-1) + R/2,\, H(\alpha)\}. \end{aligned}$$

In particular, $C(R) < 1$ for $R < 1.2338$; in other words, $R^* \ge 1.2338$.

Since the channel is symmetric as in the Gaussian case, decoding-based coding schemes are useless and the (partial) decode–forward lower bound simplifies as

$$C(R) \ge C(0) = 0.3941.$$

The capacity $C(R)$ lies between two simple bounds as plotted in Figure 5.11.

While the compress–forward lower bound in (5.6) can be evaluated only

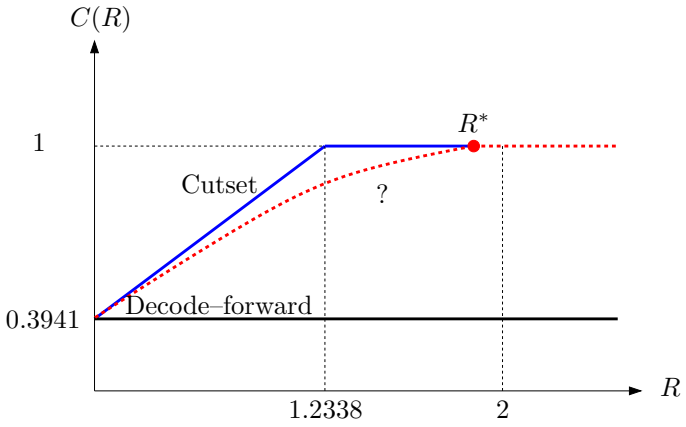


Figure 5.11: Optimal $C(R)$ curve.

numerically, one extreme point can be calculated analytically. Let R^*_{CF} be the minimum R such that the compress–forward lower bound $C_{\mathrm{CF}}(R) = 1$. Then, the inverse problem of finding R^*_{CF} is equivalent to finding the minimum sum-rate of noninteractive communication between nodes 2 and 3 so that each of the nodes can losslessly compute $X_1 = Y_2 \cdot Y_3 \sim \mathrm{Bern}(1/2)$. Thus, we can apply Orlitsky and Roche's result on coding for computing [OR01] and conclude that

$$\begin{aligned} R^*_{\mathrm{CF}} &= H_{\mathcal{G}}(Y_2|Y_3) + H_{\mathcal{G}}(Y_3|Y_2) \\ &= H(Y_2) + H(Y_3) \\ &= 2H\left(\frac{1}{\sqrt{2}}\right) \\ &= 1.7449, \end{aligned}$$

where $H_{\mathcal{G}}(Y_2|Y_3)$ and $H_{\mathcal{G}}(Y_3|Y_2)$ denote the conditional graph entropies that characterize the minimum rates to compute X_1 at node 3 and node 2, respectively. In other words, $C(R) = 1$ for $R \geq 1.7449$. Note that noninteractive extensions of compress–forward including hash–forward [CK07, Kim08], noisy network coding [LKEGC11], and hybrid coding [KLM11] do not perform better than compress–forward.

Compress–forward can be improved instead by making communication between nodes 2 and 3 interactive. Suppose that node 2 first uses the regular compress–forward to help node 3 recover the message and then node 3 uses a

modified version of compress–forward that incorporates the signal from node 2 as side information to help node 2 recover the message; see Kaspi [Kas85] for the origin of this idea in two-way lossy source coding. While this modification does not improve upon the noninteractive compress–forward lower bound in (5.6) for the Gaussian case (since the quadratic Gaussian rate–distortion function is the same with or without side information at the encoder [WZ76]), it provides strict improvements in general, for example, in the current setup of the binary broadcast relay channel. As before, the inverse problem of finding the corresponding minimum sum-rate $R^*_{\mathrm{CF}^2}$ of this coding scheme can be recast into the problem of finding the minimum sum-rate for computing X_1 at nodes 2 and 3 via two-round communication. Following the results on interactive coding for computing by Orlitsky and Roche [OR01], and Ma and Ishwar [MI08], it can be shown that

$$\begin{aligned} R^*_{\mathrm{CF}^2} &= H(Y_2) + H(X_1|Y_3) \\ &= H\left(\frac{1}{\sqrt{2}}\right) + \frac{1}{\sqrt{2}} H\left(\frac{1}{\sqrt{2}}\right) \\ &= 1.4893. \end{aligned}$$

In other words, $C(R) = 1$ for $R \geq 1.4893$.

As for the Gaussian case in Section 5.3, we can adapt the coding scheme by Draper, Frey, and Kschischan [DFK03], in which compress–forward is followed by decode–forward. This interactive relaying scheme in general yields a tighter lower bound than the two-round interactive compress–forward lower bound, since it is more efficient to use full knowledge of the message (decode–forward) for the second-round communication. At the extreme point of the 1-bit message, however, there is no gain since computing X_1 is equivalent to decoding the message itself. Hence, compress–forward followed by decode–forward yields the same upper bound on the minimum sum-rate R^* as

$$\begin{aligned} R^* &\leq 1 - I(X_1; Y_2) + H_{\mathcal{G}}(Y_2|Y_3) \\ &= H(Y_2) + H(X_1|Y_3) \\ &= 1.4893. \end{aligned}$$

Now we further generalize the idea of interactive relaying to q-round interactive compress–forward. Again at the extreme point of the 1-bit message, the inverse problem of finding the minimum sum-rate $R^*_{\mathrm{CF}^q}$ is equivalent to q-round interactive coding for computing, in which nodes 2 and 3 exchange messages in q rounds of communication to losslessly recover X_1. While the exact characterization of this minimum sum-rate for q-round computing seems to be intractable, using ingenious techniques Ma and Ishwar [MI08], [MI09] characterized its limiting behavior as

$$\begin{aligned}\lim_{q\to\infty} R^*_{\mathrm{CF}^q} &= (1+p)H(p) + p\log(pe^{1-p})\,\big|_{p=1/\sqrt{2}} \\ &= 1.4346.\end{aligned}$$

They further showed that for the natural coding scheme that achieves this limiting behavior, the corresponding sum-rate R_{CF^q} is strictly larger than $R^*_{\mathrm{CF}^\infty} = \lim_{q\to\infty} R^*_{\mathrm{CF}^q}$. Thus, $C(R) = 1$ for $R \geq 1.4346$, and among all *known* relay coding schemes this can be achieved only by infinite rounds of interactive relaying! Therefore, roughly speaking, when the network is to be flooded with information, it is more efficient for the relays to spray tiny droplets of the information back and forth than to splash a huge amount at a time.

Chapter 5, in part, includes the material in Yu Xiang, Lele Wang, and Young-Han Kim, "Information flooding," *Annual Allerton Conference on Communication, Control, and Computing,* pp. 45–51, Monticello, IL, September 2011. The dissertation author was the primary investigator and author of this paper.

Bibliography

[ABL00] Alexei Ashikhmin, Alexander Barg, and Simon Litsyn. A new upper bound on the reliability function of the gaussian channel. *IEEE Trans. Inf. Theory*, 46(6):1945–1961, September 2000.

[AC86] Rudolph Ahlswede and Imre Csiszár. Hypothesis testing with communication constraints. *IEEE Trans. Inf. Theory*, 32(4):533–542, 1986.

[ACLY00] Rudolf Ahlswede, Ning Cai, Shuo-Yen Robert Li, and Raymond W. Yeung. Network information flow. *IEEE Trans. Inf. Theory*, 46(4):1204–1216, 2000.

[Are80] Mohammad Reza Aref. *Information flow in relay networks*. Ph.D. thesis, Stanford University, Stanford, CA, October 1980.

[Ber78] Toby Berger. Multiterminal source coding. In Giuseppe Longo, editor, *The Information Theory Approach to Communications*. Springer-Verlag, New York, 1978.

[Ber79] Toby Berger. Decentralized estimation and decision theory. In *Proc. IEEE Inf. Theory Workshop*, Mt. Kisco, NY, September 1979.

[BY08a] Marat Burnashev and Hirosuke Yamamoto. On BSC, noisy feedback and three messages. In *Proc. IEEE Internat. Symp. Inf. Theory*, pages 886–889, Toronto, Canada, July 2008.

[BY08b] Marat Burnashev and Hirosuke Yamamoto. On zero-rate error exponent for BSC with noisy feedback. *Problems of Information Transmission*, 44(3):33–49, 2008.

[BY10] Marat Burnashev and Hirosuke Yamamoto. On reliability function of BSC with noisy feedback. *Problems of Information Transmission*, 46(2):2–23, 2010.

[CEG79] Thomas M. Cover and Abbas El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inf. Theory*, 25(5):572–584, September 1979.

[CK07] Thomas M. Cover and Young-Han Kim. Capacity of a class of deterministic relay channels. In *Proc. IEEE Internat. Symp. Inf. Theory*, pages 591–595, Nice, France, June 2007.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, New York, second edition, 2006.

[DFK03] S. C. Draper, B. J. Frey, and F. R. Kschischang. Interactive decoding of a broadcast message. In *Proc. 41st Ann. Allerton Conf. Commun. Control Comput.*, Monticello, IL, October 2003.

[EFS56] Peter Elias, Amiel Feinstein, and Claude E. Shannon. A note on the maximum flow through a network. *IRE Trans. Inf. Theory*, 2(4):117–119, December 1956.

[EG81] Abbas El Gamal. On information flow in relay networks. In *Proc. IEEE National Telecom Conf.*, volume 2, pages D4.1.1–D4.1.4. November 1981.

[EGK11] Abbas El Gamal and Young-Han Kim. *Network Information Theory*. Cambridge University Press, Cambridge, 2011.

[FF56] L. R. Ford, Jr. and D. R. Fulkerson. Maximal flow through a network. *Canad. J. Math.*, 8(3):399–404, 1956.

[GA10] A. A. Gohari and Venkat Anantharam. Information-theoretic key agreement of multiple terminals—I: Source model. *IEEE Trans. Inf. Theory*, 56(8):3973–3996, August 2010.

[HA98] Te Sun Han and Shun Amari. Statistical inference under multiterminal data compression. *IEEE Trans. Inf. Theory*, 44(6):2300–2324, October 1998.

[Han87] Te Sun Han. Hypothesis testing with multiterminal data compression. *IEEE Trans. Inf. Theory*, 33(6):759–772, 1987.

[Hor63] Michael Horstein. Sequential transmission using noiseless feedback. *IEEE Trans. Inf. Theory*, 9(3):136–143, July 1963.

[IL06] Tadeusz Inglot and Teresa Ledwina. Asymptotic optimality of new adaptive test in regression model. *Ann. Inst. H. Poincare Probab. Statist.*, 42(5):579–590, 2006.

[Kas85] Amiram H. Kaspi. Two-way source coding with a fidelity criterion. *IEEE Trans. Inf. Theory*, 31(6):735–740, 1985.

[KGG05] Gerhard Kramer, Michael Gastpar, and Piyush Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Trans. Inf. Theory*, 51(9):3037–3063, September 2005.

[Kim07] Young-Han Kim. Coding techniques for primitive relay channels. In *Proc. 45th Ann. Allerton Conf. Commun. Control Comput.*, Monticello, IL, September 2007.

[Kim08] Young-Han Kim. Capacity of a class of deterministic relay channels. *IEEE Trans. Inf. Theory*, 54(3):1328–1329, March 2008.

[KLM11] Young-Han Kim, Sung Hoon Lim, and Paolo Minero. Relaying via hybrid coding. In *Proc. IEEE Internat. Symp. Inf. Theory*, pages 1946–1950, St. Petersburg, Russia, August 2011.

[KLW07] Young-Han Kim, Amos Lapidoth, and Tsachy Weissman. The Gaussian channel with noisy feedback. In *Proc. IEEE Internat. Symp. Inf. Theory*, pages 1416–1420, Nice, France, June 2007.

[LKEGC11] Sung Hoon Lim, Young-Han Kim, Abbas El Gamal, and Sae-Young Chung. Noisy network coding. *IEEE Trans. Inf. Theory*, 57(5):3132–3152, May 2011.

[MI08] Nan Ma and Prakash Ishwar. Two-terminal distributed source coding with alternating messages for function computation. In *Proc. IEEE Internat. Symp. Inf. Theory*, pages 51–55, Toronto, Canada, July 2008.

[MI09] Nan Ma and Prakas Ishwar. Infinite-message distributed source coding for two-terminal interactive computing. In *Proc. 47th Ann. Allerton Conf. Commun. Control Comput.*, pages 1510–1517, Monticello, IL, September 2009.

[MI11] Nan Ma and Prakash Ishwar. Some results on distributed source coding for interactive function computation. *IEEE Trans. Inf. Theory*, 57(9):6180–6195, September 2011.

[NG07] Bobak Nazer and Michael Gastpar. Computation over multiple-access channels. *IEEE Trans. Inf. Theory*, 53(10):3498–3516, October 2007.

[OR01] Alon Orlitsky and James R. Roche. Coding for computing. *IEEE Trans. Inf. Theory*, 47(3):903–917, 2001.

[Pin68] M. S. Pinsker. The probability of error in block transmission in a memoryless Gaussian channel with feedback. *Probl. Inf. Transm.*, 4(4):3–19, 1968.

[RW12] Md. Saifur Rahman and Aaron Wagner. On the optimality of binning for distributed hypothesis testing. *IEEE Trans. Inf. Theory*, 58(10):6282–6303, October 2012.

[SB71] Schalkwijk and M. Barron. Sequential signalling under a peak power constraint. *IEEE Trans. Inf. Theory*, 17(3):278–282, May 1971.

[SG00] B. Schein and R. G. Gallager. The Gaussian parallel relay channel. In *Proc. IEEE Internat. Symp. Inf. Theory*, page 22, Sorrento, Italy, June 2000.

[Sha59] C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *Bell System Tech. J.*, 38:611–656, 1959.

[SHA94] H. Shimokawa, Te Sun Han, and Shun Amari. Error bound of hypothesis testing with data compression. In *Proc. IEEE Internat. Symp. Inf. Theory*, page 29, June 1994.

[SK66] J. Pieter M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback—I: No bandwidth constraint. *IEEE Trans. Inf. Theory*, 12(2):172–182, April 1966.

[SWWZ69] Lawrence A. Shepp, Jack K. Wolf, Aaron D. Wyner, and Jacob Ziv. Binary communication over the Gaussian channel using feedback with a peak energy constraint. *IEEE Trans. Inf. Theory*, 15(4):476–478, 1969.

[TC08] Chao Tian and Jun Chen. Successive refinement for hypothesis testing and lossless one-helper problem. *IEEE Trans. Inf. Theory*, 54(10):4666–4681, October 2008.

[Tun78] S.-Y. Tung. *Multiterminal Source Coding*. Ph.D. thesis, Cornell University, Ithaca, NY, 1978.

[vdM71] Edward C. van der Meulen. Three-terminal communication channels. *Adv. Appl. Probab.*, 3(1):120–154, 1971.

[WJ65] John M. Wozencraft and Irwin M. Jacobs. *Principles of Communication Engineering*. Wiley, New York, 1965.

[Wyn68] A. D. Wyner. On the Schalkwijk–Kailath coding scheme with a peak energy constraint. *IEEE Trans. Inf. Theory*, 14(1):129–134, January 1968.

[WZ76] Aaron D. Wyner and Jacob Ziv. The rate–distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory*, 22(1):1–10, 1976.

[XK05] Liang-Liang Xie and P. R. Kumar. An achievable rate for the multiple-level relay channel. *IEEE Trans. Inf. Theory*, 51(4):1348–1358, 2005.

[XK12] Yu Xiang and Young-Han Kim. Interactive hypothesis testing with communication constraints. In *Proc. 50th Ann. Allerton Conf. Commun. Control Comput.*, pages 1065–1072, Monticello, IL, October 2012.

[XK13a] Yu Xiang and Young-Han Kim. Gaussian channel with noisy feedback and peak energy constraint. *IEEE Trans. Inf. Theory*, 59(8):4746–4756, August 2013.

[XK13b] Yu Xiang and Young-Han Kim. Interactive hypothesis testing against independence. In *Proc. IEEE Internat. Symp. Inf. Theory*, pages 2840–2844, Istanbul, Turkey, July 2013.

[XK15] Yu Xiang and Young-Han Kim. Interactive hypothesis testing with communication constraints. To be submitted to *IEEE Trans. Inf. Theory*, 2015.

[XWK11] Yu Xiang, Lele Wang, and Young-Han Kim. Information flooding. In *Proc. 49th Ann. Allerton Conf. Commun. Control Comput.*, pages 45–51, Monticello, IL, September 2011.

[YI79] H. Yamamoto and K. Itoh. Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback. *IEEE Trans. Inf. Theory*, 25(6):729–733, November 1979.