VARIABLE PACKET-ERROR CODING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Xiaoqing Fan

December 2017

VARIABLE PACKET-ERROR CODING

Xiaoqing Fan, Ph.D.

Cornell University 2017

Consider a communication scenario in which a source is encoded into N packets, at most T of which may be arbitrarily altered by an omniscient adversary. Unlike prior work in coding theory which seeks to optimize only the worst-case performance of the code, in this work, codes are designed to enable the decoder to reproduce the source subject to a certain distortion constraint when there are no packets errors, subject to a less stringent distortion constraint when there is one error, etc. The topic of this thesis is to find the trade-off between rate and distortion in such communication scenarios.

A code design based on the *Polytope codes* is introduced for the binary source with erasure distortion measure and is also proven to have partial optimality property. Moreover, for the point-to-point scenario ($N = 1$ and $T = 1$), both inner bounds and outer bounds are derived for discrete sources with finite alphabet with general distortion measure. For the binary source with Hamming distortion, these two bounds are proven to be the same.

For a Gaussian source with a mean-square error distortion, it is shown that a natural design based on MDS codes is not order-optimal in the rate as the distortion constraint tends to zero, but a hybrid scheme that involves a form of uncoded transmission is. We derive an outer bound which has a constant gap with the inner bound naturally generated by the codes we design, thus fully characterizing the Rate-Distortion region.

BIOGRAPHICAL SKETCH

Xiaoqing Fan received her B.E. degree in School of Mathematics from Peking University, China, in 2010. She pursued Ph.D in Electrical and Computer Engineering and minored in Computer Science, at Cornell, Ithaca, NY, where she was also awarded the M.S. degree. Her research interest is information-theoretic security. She is also interested in Machine Learning and Digital Communications.

To my loving family,

whose endless support made this achievement possible.

ACKNOWLEDGEMENTS

I have been fortunate to have had a lot of support during my time at Cornell. Firstly, I would like to express my sincerest gratitude to my advisor, Prof. Aaron Wagner. My work has greatly benefited from his continuous support, his unique insights, and his great patience for teaching. His genuine enthusiasm for research, his attention to detail while maintaining an impressively clear view of the broader picture, his commitment to scientific communication, and his outstanding abilities as a lecturer are a source of inspiration to me. In addition to the academic training, Aaron was very patient and supportive during difficult times I had during my PhD. Being his student was a privilege and a pleasure, and I will always be grateful for his mentorship and friendship. I could not have imagined having a better advisor and mentor for my Ph.D study.

I would also like to thank my committee members, Prof. Tang and Prof. Suh. They provided valuable guidance at various stages of my research, and our discussions have significantly improved my work.

I would also like to thank all my colleagues and friends I met in Cornell, Yuguang, Ibrahim, Yucel, Sinem, Nirmal, Omer for the stimulating discussions and for all the fun we have had. I wish them luck!

Finally, and most importantly, I am forever indebted to my mother Ren for supporting me spiritually throughout my life. You have sacrificed so much for me and for my education. Your love, dedication, and support are the constant bright spot. This work is the fruit of your labor first, and mine second. I would also like to thank my beautiful family: my father, Junqing, my husband, Yihan, my mother-in-law and farther-in-law, Meiping and Daqing. All of you have been there to support me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

Consider a communication scenario in which a source sends information to a destination over several nonintersecting paths in a network as depicted in Fig. 1.1. These paths could be used to increase the data rate beyond what would be achievable with a single path, or they could be used to provide redundancy to allow the decoder to recover from errors introduced by the network. It is also possible to simultaneously achieve both goals, subject to a tradeoff between the two, which is the topic of this paper. In particular, it is assumed that some number of paths are subject to adversarial errors, and one shall seek codes that achieve high data rates while still ensuring that the encoder can reconstruct the original message reasonably well in the face of those errors. That is, one is interested in packet-error coding in which the number of packet errors is a variable and a single code simultaneously provides different performance guarantees depending on the number of packet errors. We call this *variable packet-error coding* (VPEC).

In our model, we assume there is an active adversary which is omniscient and omnipotent in every other regard: the adversary knows the source sequence, the encoding and decoding functions, and all sources of randomness in the system. And it can alter the

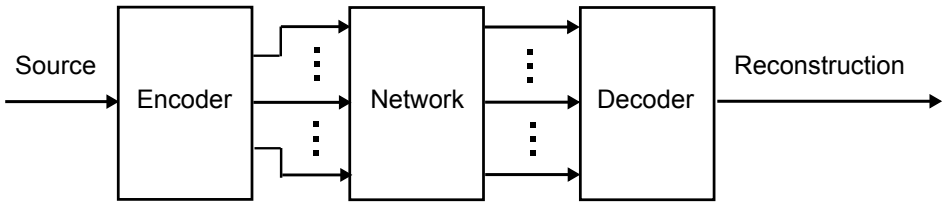

Figure 1.1: Channel Model

packets that it controls in a way that is maximally disruptive to communication. This indicates that secret key is not useful for our model. It is also assumed to have unbounded computational power. Therefore, cryptographic primitives based on computational hardness not useful. However, the adversary can only alter up to some fixed number of paths in the network. This is reasonable since we can transmit packets using a combination of the technologies such as terrestrial channels, aerial relay channels, satellite channels, etc. And it is unlikely to compromise different technologies simultaneously.

While coding for adversarial errors is a classical subject [7,20,41,43], most prior work in coding theory seeks to optimize only the worst-case performance of the code. If the adversary is assumed to cause at most T errors, then the design provides a performance guarantee if there are T errors, but no improved guarantee if the number of errors is lower. In particular, the system does not provide an improved guarantee if there are no errors at all. This is arguably inefficient. Even highly vulnerable systems will spend only a small fraction of time in a state in which an adversary has infiltrated a system and is actively altering messages. Thus, one seeks designs that provide reasonable performance guarantees when an attack occurs without unduly sacrificing the performance of the system when it is not under attack.

In addition, most prior work on the problem of communicating in the presence of adversarial errors, it is worth noting, implicitly assumes source-channel separation [7,43]. This contradicts the fact that source-channel separation is not optimal for this problem in [2]. Therefore, this problem is properly formulated using rate-distortion theory. We assume that a source sequence is encoded into N packets (or messages) at rates $R_1, ..., R_n$, at most T of which may be adversarially altered by the network. In the symmetric case, the rates are the same for all N packets. The decoder receives N packets without knowing which packets were altered or how many have been altered (except that it knows

that the total number of altered packets does not exceed T). The decoder then outputs a reconstruction of the source. Given a *distortion measure* between the source and reproduction, one may seek codes that guarantee a certain level of distortion when there are T errors, a lower level of distortion when there are no errors.

Three models are discussed here. The first model is a symmetric model (each packet has the same rate R) for a binary source under *erasure* distortion measure: the per-letter distortion is zero if the source and reconstruction symbols agree, one if the reconstruction symbol is a special "erasure" symbol, and infinity otherwise. Thus there is an infinite penalty for guessing a source symbol incorrectly, and the decoder should output the erasure symbol for any source symbol about which it is unsure. The erasure distortion measure is reasonable for a wide array of physical sources. For audio and video, it is typically possible to interpolate over unknown samples, pixels, or frames at the receiver. Similarly, humans can often recover a natural language source when some of the characters have been erased [8]. Even executable computer code, which is typically viewed as being unamenable to lossy compression, is suitable to compression under the erasure distortion measure: execution of the program at the decoder could simply pause whenever it reached an erasure and wait for further information, without ever executing incorrect instructions. Focusing on the erasure distortion measure is also a useful simplifying assumption when considering new problems, akin to the way that the binary erasure channel is a good starting point in the study of modern coding theory [33]. Assuming there are no errors in the reconstruction, the distortion of a string is then the fraction of erasures in the reconstruction. A similar model has been discussed in [3]. However, their model only considers the worst-case scenario.

The second model is an asymmetric model (the packets may have different rates) for a Gaussian source under squared error distortion measure. This model is motivated

by the following real world problem. This model can be viewed as adversarial version of the classical multiple description problem [18]. Consider the multi-homing/multi-path communication scenario in which a single media source is encoded into several streams and then transmitted to a destination over a network. One may think of the different streams as being transmitted via N different communication modalities, such as terrestrial links, satellites, aerial communications relays, etc., with the assumption that while an adversary may compromise one such modality, it is unlikely to compromise all, or even most, of them. One may also think of the streams as being communicated via a single modality, such as a terrestrial wireless network, but with N geographically diverse paths. While the adversary may be able to compromise some of the paths in the network, it is unlikely to be able to simultaneously compromise most of them. A solution to this model answers the following question: how to achieve a high-quality connection when there are no errors and a degraded, but still functional, connection when there are T errors?

The third model is a point-to-point (P2P) VPEC model ($N = 1, T = 1$) for general discrete sources (finite alphabet) under general distortion measure. This model can be viewed as an adversarial version of the classical R-D problem with a slight difference. Classical Rate-Distortion (R-D) theory completely solves the following problem [6,36] which can be viewed as a special case of the our VPEC model where $N = 1$ and $T = 0$: given a source distribution and a distortion measure, what is the minimum expected distortion D achievable at a particular rate R? It is well-known that the trade-off can be fully characterized by the Rate-Distortion function. Studying the P2P VPEC model provides valuable insight into solving the general VPEC problem.

1.2 Contributions

1.2.1 Binary Source with Erasure Distortion Measure

In the binary model, a binary source is encoded into N packets (or messages) at a given rate R, at most T of which may be adversarially altered by the network. The decoder receives N packets without knowing which packets were altered or how many have been altered (except that it knows that the total number of altered packets does not exceed T). The decoder then outputs a reconstruction of the source measured by erasure distortion measure.

A code construction which guarantees a certain level of distortion D_T when there are one or more (up to T) errors and a lower level of distortion D_0 when there are no errors is provided. This construction is inspired by the *polytope codes* which are introduced by Kosut, Tong, and Tse [26] in the context of network coding with adversarial nodes. It is achieved by including certain nonlinear check symbols in the packets and by performing all arithmetic operations over the regular integers instead of over a finite field. The construction of polytope codes here departs significantly from that of Kosut, Tong, and Tse. Polytope codes are similar to linear maximum distance separable (MDS) codes but with an added feature: for a certain number of errors, which exceeds the decoding radius of the code, it is possible to always decode some of the codeword symbols even though it is not possible to decode all of them. This is to be contrasted with conventional MDS codes, for which in general none of the coded symbols can be decoded unless they all can. This "partial decodability" property will be crucial in our use of polytope codes. It eliminates the most complicated step of the construction, and it allows for simpler encoding. Nonetheless, it is still called polytope code to emphasize its connection to this earlier work. A partial optimality result for polytope codes is also provided.

1.2.2 Gaussian Source with Squared Error Distortion Measure

In the Gaussian model, it is assumed that a Gaussian source sequence in encoded into N packets at rate $R_1, ..., R_N$, at most T of which may be arbitrarily altered. The decoder receives N packets without knowing which packets were altered or how many have been altered (except that it knows the total number of altered packets does not exceed T). The decoder then outputs a reconstruction of the source which is measured by squared error distortion measure.

At a first glance, we may use successive refinement [13,34] coupled with error correcting and error detecting codes [7,43] to solve this problem. That is, we first use successive refinement to generate a base layer and an enhancement layer. The decoder can achieve distortion D_T given the base layer alone and distortion D_0 given both layers. Then we use error correcting codes (able to correct up to T errors) to encode the base layer and use error detecting codes (able to detect up to T errors) to encode the enhancement layer across N packets. Although this coding scheme is eligible, we show that there is an unbounded gap (as D_0 and D_T go to 0) between its rate region and the outer bound we obtained. Luckily, a slight change of this coding scheme produces a bounded gap. The idea is that we generate a finer base layer and use the same error correcting codes, guaranteeing that the decoder can always correctly recover the base layer. Then we leave the enhancement layer uncoded and send it across the packets. In this case, when the adversary alters the enhancement layer without being detected, we can still achieve distortion D_T.

Similarly, an approximate characterization of the Rate-Distortion region is given. The achievability scheme we provide, naturally generating the inner bound, is based on successive refinement and network error correcting codes. In addition, for $N = 3$ and $T = 1$ with symmetric rate for each packet, motivated by Polytope Codes [24,26], a

better scheme for sufficiently small D_1 and D_0/D_1 is provided. We also derive an outer bound. The key to derive this outer bound is to show that the conditional entropy of the source (or source plus a Gaussian noise) given any $N-2T$ codewords is upper bounded. This is similar to Theorem 4 in [43] which is a generalization of Singleton bound for network error correction codes. Finally, we prove that the gap between the inner and the outer bound is upper bounded by a constant.

1.2.3 P2P VPEC Model

In this P2P VPEC model, given a source distribution and a distortion measure, it is assumed that the source is encoded at rate R and then sent over the channel which may be arbitrarily altered by an adversary. The decoder receives the packet without knowing whether it has been altered or not. The decoder then outputs a reconstruction of the source.

An approximate characterization of the Rate-Distortion region is given here. The achievability scheme we provide, naturally generating the inner bound, is similar to the code design for the classical R-D problem. Next, let $\mathcal{X}^n$ and $\hat{\mathcal{X}}^n$ denote the source alphabet and the reconstruction alphabet, respectively. Suppose we have a code (f, g) for this problem, where f is the encoding function and g is the decoding function. Let $C = \{\hat{x}^n : \exists x^n \in \mathcal{X}^n, g(f(x^n)) = \hat{x}^n\}$ denote the codebook. Since the adversary is omniscient, given any codebook C and a source message x^n, the adversary can alter the packet so that the reconstruction under attack is the codeword in C with largest distortion to x^n. We can view both the encoding function and the adversarial attack as a channel from $C \subseteq \hat{\mathcal{X}}^n$ to $\mathcal{X}^n$, then the problem is closely related to [22], which studies the relations between the image of two channels. The outer bound is then derived from

that. Finally, for the binary source under Hamming distortion measure, the complete R-D region is derived.

1.3 Literature Overview

One closely related work is [3]. Their problem setting is the same: a source is encoded into N packets, any T of which may be altered in an arbitrary way by an adversary. The decoder receives the N packets and, without knowing which packets were altered. However, their goal is to design a code to reconstruct the original source to meet one distortion constraint. In their work, they show that a layered architecture for this problem that separates the lossy compression from the coding for adversarial errors is optimal in the binary-Hamming and quadratic-Gaussian cases yet suboptimal in general. Our model, on the other side, focus on simultaneously satisfying two distortion constraints. And this significantly changes the problem.

Another closely related problem multiple-descriptions (MD) problem [18] in network information theory. The difference is that in the MD problem each message is either received correctly or not received at all; the network does not introduce errors. The MD problem has received considerable attention [4,15,18,31] since it was introduced, including the special case in which the distortion measure is erasure [4]. Allowing the adversary to introduce errors instead of erasures seems to significantly alter the problem, however. In particular, although techniques from coding theory have been successfully applied to the MD problem [31], the polytope codes that shall prove so effective here do not appear to be useful for the MD problem.

The closest constructions to VPEC in the cryptography literature are message authentication codes (MACs) [17, Section 6.1]. MACs assume a computationally-bounded

adversary and a shared key that is available to the transmitter and the receiver but not the adversary. VPEC allows for an adversary that is omniscient and has unbounded computational power.[1] See Ahmed and Wagner [5] for a further discussion of how VPEC relates to cryptographic approaches.

The Binary VPEC problem studied here can be viewed as an instance of a "large-alphabet" channel. In classical studies of channel capacity, the channel law is held fixed and the blocklength is permitted to grow without bound (e.g. [10]). In the case of discrete memoryless channels with finite alphabet, this model well captures the practical regime in which the blocklength is much bigger than the number of channel inputs or outputs. While this model has proven to be very successful, the asymptotic that it considers is not always the right one. For the problem in which a sender sends data over several independent paths in a network, some of which may alter the data adversarially en route, the "blocklength" is naturally viewed as the number of distinct paths, which is generally small, while the "alphabet" is the number of distinct messages that can be sent on one path, which is generally very large. Thus the appropriate model is in some sense dual to the classical one: the blocklength is fixed while the input and output alphabet sizes are permitted to grow without bound, as is done in this paper. Such channels have arisen in network coding [19], although many fundamental Shannon-theoretic questions about them are not well understood. One notable exception is that, as alluded to earlier, source-channel separation is known to be optimal for such channels if the source is Gaussian and the distortion measure is quadratic or if the source is Bernoulli and the distortion measure is Hamming distance but not, in general, if the source is binary and the distortion measure is erasure distortion [4]. Thus we already know that such channels behave differently from conventional ones. channels *packet-error* (or *path-error*) *coding* (PEC).

[1] It is straightforward to show that VPEC with a secret key reduces to the regular multiple-descriptions problem [5].

In addition, the Gaussian VPEC problem is closely related to the MD problem for Gaussian source, which has received considerable research attention and is well studied. [28] shows that for quadratic Gaussian problem with only two descriptions, the achievable region in [15] is tight. However, Zhang and Berger show that the achievable region in [15] is not tight in general [46] and a complete characterization of the rate-distortion (R-D) region has not been found to this date. Recent research attention has shifted to the general N-description problem, partly motivated by the availability of multiple transmission paths in modern communication networks. Though completely characterizing the rate-distortion region of the Gaussian multiple description problem is difficult if not impossible, Tian, Mohajer and Diggavi provide an approximate characterization in [39]. They analyze two achievability schemes: one is based on successive refinement [13,34] coupled with multi-level diversity coding [35,42,44,45], the other one is a generalization of the multilayer coding scheme proposed by Puri, Pradhan and Ramchandran [29,30]. Another special case is analyzed in [4] where distortion measure is erasure.

CHAPTER 2

PROBLEM DESCRIPTION

2.1 Problem Description

Let N be a positive integer and define $[N] = \{1, 2, \ldots, N\}$. Let $\mathcal{X}$ denote the source alphabet. Here, both finite source alphabets and uncountable source alphabets are discussed. The source produces a sequence $X_1, X_2, \ldots, X_n$ i.i.d. $\sim \mathcal{P}(x)$, $x \in \mathcal{X}$.

Let X^n denote[1] the source message in $\mathcal{X}^n$, where $\mathcal{X} = [K]$ is the alphabet for the source. We shall call n the *blocklength* of the source. Given the source sequence X^n, the encoder creates N *packets* (or *messages*, or *codewords*) via the functions

$$f_\ell : \mathcal{X}^n \mapsto \mathcal{X}^{nR_\ell} \quad \ell \in \{1, \ldots, N\}.$$

We call R_ℓ the *rate* of the ℓth packet. The encoder sends the packets

$$(f_1(x^n), f_2(x^n), \ldots, f_N(x^n)),$$

which we will often abbreviate as

$$(C_1, C_2, \ldots, C_N).$$

Let $\hat{\mathcal{X}}$ denote the reconstruction alphabet. The decoder employs a function

$$g : \prod_{\ell=1}^{N} \mathcal{X}^{nR_\ell} \mapsto \hat{\mathcal{X}}^n$$

to reproduce the source given the received packets. The fidelity of the reproduction is measured using a distortion measure

$$d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty].$$

[1] When the length of the vector is particularly important, we indicate it using a superscript.

We extend the single-letter distortion measure $d(\cdot,\cdot)$ to strings in the usual way

$$d(x^n, \hat{x}^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i, \hat{x}_i).$$

We call the tuple $(f_1, \ldots, f_N, g)$ a *code* for the problem. We shall be interested in the distortion of the code under two scenarios, when there are T_0 and T_1 packet errors, denoted by $D_{T_0}(f_1, \ldots, f_N, g)$ and $D_{T_1}(f_1, \ldots, f_N, g)$, respectively. Then, achievable Rate-Distortion (R-D) vectors $(R_1, ..., R_N, T_0, T_1)$ may (or may not) be defined based on $D_{T_0}(f_1, \ldots, f_N, g)$ and $D_{T_1}(f_1, \ldots, f_N, g)$. The topic is to characterize the set of achievable R-D vectors.

Here, we always assume $T_0 = 0$. This is because we are interested in the performance of the codes (in terms of distortion) both when there are no packet errors and when there are T_1 errors. Since T_0 is fixed at zero we shall write T in place of T_1 in the sequel.

2.1.1 Binary Erasure VPEC Problem

This model uses binary source with any arbitrary probability distribution. The distortion measure we use is *erasure distortion measure* [10, p. 338]: for $x \in \mathcal{X}$ and $\hat{x} \in \{\mathcal{X} \cup e\}$,

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } \hat{x} = e \\ \infty & \text{otherwise.} \end{cases} \tag{2.1}$$

The distortion of the code for this model is defined as follows:

$$D_0(f_1, \ldots, f_N, g) := \max_{x^n \in \mathcal{X}^n} d(x^n, g(C_\ell, \ell \in [N])),$$

$$D_{T_1}(f_1, \ldots, f_N, g) := \max_{x^n \in \mathcal{X}^n} \max_{A \subseteq [N]:|A| \leq T_1} \max_{\tilde{C}_A} d(x^n, g(C_{A^c}, \tilde{C}_A)).$$

Here $g(C_{A^c}, \tilde{C}_A)$ denotes the decoder's output when its input is $C_\ell = f_\ell(x^n)$ for all $\ell \in A^c$ and $\tilde{C}_\ell$ for all $\ell \in A$.

The definition for *achievable* R-D vectors is as follows.

Definition 1 *The rate-distortion (R-D) vector* $(R_1, \ldots, R_N, D_0, D_T)$ *is* achievable *if for all* $\epsilon > 0$*, there exists a code* $(f_1, \ldots, f_N, g)$ *for some blocklength n satisfying*

$$\frac{1}{n} \log_K |f_\ell| \leq R_\ell + \epsilon \quad \ell \in [N] \tag{2.2}$$

$$D_0(f_1, \ldots, f_N, g) \leq D_0 + \epsilon \tag{2.3}$$

$$D_T(f_1, \ldots, f_N, g) \leq D_T + \epsilon. \tag{2.4}$$

We shall also focus on characterizing those pairs (R, D) for which

$$(R, R, R, \ldots, R, 0, D)$$

is achievable. In words, we require that all of the packets have rate R, there be lossless reconstruction when there are no packet errors, and there be an erasure distortion of at most D when there are T_1 packet errors.

All of our code constructions shall achieve zero, not merely vanishing, distortion when there are no packet errors, i.e.,

$$\max_{x^n \in \mathcal{X}^n} d(x^n, g(C_\ell, \ell \in [N])) = 0.$$

We call a code that achieves this constraint *feasible*.

2.1.2 Gaussian VPEC Problem

This model uses a Gaussian source. Here, we assume that X_i are i.i.d. Gaussian random variables: $X_i \sim \mathcal{N}(0,1)$; $\mathcal{X} = \mathfrak{R}$.

In this model, the adversary alters up to T of the N packets ($N \geq 2T+1$) and sends the altered ones instead to the decoder, which then attempts to reconstruct X^n from the received packets up to a specified distortion. We assume that the adversary has full knowledge of X^n, all other packets and the decoder's decoding strategy. Moreover, the adversary can observe X^n and then decide which encoders (up to T) to take over. We also assume that the existence and location of the altered packets are unknown to the decoder.

It is natural to define the distortion of the code as follows. The distortion when there are no packet errors at present is:

$$D_0(f_\ell, \ell \in [N], g) \coloneqq E_{X^n}[d(X^n, g(C_\ell, \ell \in [N])].$$

We shall also consider how well the decoder can reproduce the source when up to T of the packets are altered by the adversary:

$$D_T(f_\ell, \ell \in [N], g) \coloneqq E_{X^n}\left[\max_{A\subseteq[N]:|A|=T} \max_{C'_A} d(X^n, g(C_{A^c}, C'_A))\right].$$

Here, $g(C_{A^c}, C'_A)$ denotes the decoder's output when its input is $C_\ell = f_\ell(X^n)$ for all $\ell \in A^c$ and C'_ℓ for all $\ell \in A$. We assume that the adversary can see everything the decoder receivers. Thus, the secret key model does not fit our problem since the adversary also has access the the key. Variable packet-error problem with shared secret key is discussed in "Coding for the Large-Alphabet Adversarial Channel."

Definition 2 *The rate-distortion (R-D) vector* $(R_\ell, \ell \in [N], D_0, D_T)$ *is* achievable *if for any* $\epsilon > 0$*, there exists a code* $(f_\ell, \ell \in [N], g)$ *for some blocklength n with rate at most*

$R_\ell + \epsilon$ *for the* ℓ*th packet such that:*

$$D_0(f_\ell, \ell \in [N], g) \leq D_0 + \epsilon;$$

$$D_T(f_\ell, \ell \in [N], g) \leq D_T + \epsilon.$$

Definition 3 *For any* (D_0, D_T)*, define*

$$R(D_0, D_T) \coloneqq \{(R_\ell, \ell \in [N]) : (R_\ell, \ell \in [N], D_0, D_T) \textit{ is achievable}\}.$$

Notice that by definition, $R(D_0, D_T)$ is a closed set. The goal is to characterize $R(D_0, D_T)$.

2.1.3 P2P VPEC Problem

This model uses a general discrete source with a finite alphabet. The source produces a sequence $X_1, X_2, \ldots, X_n$ i.i.d. $\sim \mathcal{P}(x)$, $x \in \mathcal{X}$. We assume that

$$\mathcal{P}(a) > 0, \quad \forall a \in \mathcal{X}$$

throughout the paper.

The distortion measure we use is the following distortion function:

$$d : \mathcal{X} \times \hat{\mathcal{X}} \mapsto \mathfrak{R}^+,$$

where $\mathfrak{R}^+$ denotes the set of nonnegative real numbers.

For this model, we do not define the distortion of the code D_0, D_T. The definition for *achievable* R-D vectors is as follows.

Definition 4 *The Rate-Distortion (R-D) vector* (R, D_0, D_1) *is achievable if for any* $\epsilon > 0$*, there exists a code* (f, g) *for some blocklength n with rate at most* $R + \epsilon$*, satisfying*

$$\mathcal{P}^n\left(\{x^n : d(x^n, g(f(x^n))) > D_0 + \epsilon\}\right) < \epsilon, \tag{2.5}$$

$$\mathcal{P}^n\left(\left\{x^n : \max_{C'} d(x^n, g(C')) > D_1 + \epsilon\right\}\right) < \epsilon. \tag{2.6}$$

The problem is to characterize the set of achievable R-D vectors.

CHAPTER 3

BINARY ERASURE VPEC

3.1 Main Results for Binary Erasure VPEC Problem

Our main result is the following.

Theorem 1 *Suppose the maximum number of altered packets T satisfies $1 \leq T \leq N$.*

1. *If $0 \leq R < \frac{1}{N-T}$, then there is no finite D for which (R, D) is achievable.*[1]
2. *Let $F(T)$ denote $T + \lfloor \frac{T^2}{4} \rfloor + 1$ and suppose that $N \geq F(T) + 1$. Then for any $\frac{1}{N-T} \leq R \leq \frac{1}{N-2T}$, the rate-distortion pair*

$$\left(R, \frac{F(T)(N-T)(1-(N-2T)R)}{NT}\right)$$

is achievable.

Note that, per the statement of Theorem 1, the resulting scheme can only be applied when $N \geq F(T) + 1$. In particular, the blocklength must grow with the square of the number of errors. This is undesirable; one would prefer to have linear scaling. In Section 3.4, we show that this quadratic scaling cannot be improved by changing the decoder—it is intrinsic to the code itself. Of course, since N represents the number of independent paths in the network between the encoder and the decoder, we are generally interested in small values of N and T, so that the scaling behavior is not paramount.

The performance in part 2) is achieved using polytope codes and should be compared against what can be obtained using conventional MDS codes. Suppose we map $N -$

[1] In a conference version of this result [14], it was incorrectly asserted that feasible codes do not exist if $0 \leq R < \frac{1}{N-T}$. The correct statement is as given here.

$2T$ source symbols to N coded symbols using an $(N, N-2T)$ MDS code (we can, if necessary, group several source symbols together to ensure that the source alphabet is large enough to guarantee the existence of such a code). Let each coded packet consist of exactly one of the coded symbols. The rate per packet is then $R = 1/(N-2T)$, and since the minimum distance of the code is $2T+1$ [37], the decoder can always recover the source sequence exactly, even when there are T errors. Thus this scheme achieves the rate-distortion pair $(1/(N-2T), 0)$.

On the other hand, if we use an $(N, N-T)$ MDS code, then the decoder can reconstruct the source when there are no errors, and since the minimum distance is $T+1$, it can always detect when there are T or fewer errors and output the all-erasure string in response. Hence this code can achieve the rate-distortion pair $(1/(N-T), 1)$. A simple time-sharing argument shows that the line connecting these points

$$\left(R, \frac{N-T}{T} - \frac{(N-T)(N-2T)}{T}R\right)$$

is achievable. This is shown in Fig. 3.1 for $N = 3$ and $T = 1$ and in Fig. 3.2 for $N = 5$ and $T = 2$, along with the achievable rate-distortion pairs from Theorem 1. We see that Theorem 1 does strictly better.

When $N = 3$ and $T = 1$, there is actually a simple design that is not dominated by the above schemes. When $R = \frac{2}{3}$, let the blocklength of the source message be 3 and write the source as (x_1, x_2, x_3). We transmit

$$(x_1, x_2) \quad (x_2, x_3) \quad (x_3, x_1) \tag{3.1}$$

as the three packets. The decoder can check whether the copy of x_i is the same between the two packets in which it appears for each i. If the two packets have the same value of x_i, then this common value must be correct. Since the channel can alter at most one packet, there can be at most two components of (x_1, x_2, x_3) on which there is disagreement. If there is disagreement about two source components, however, then the decoder

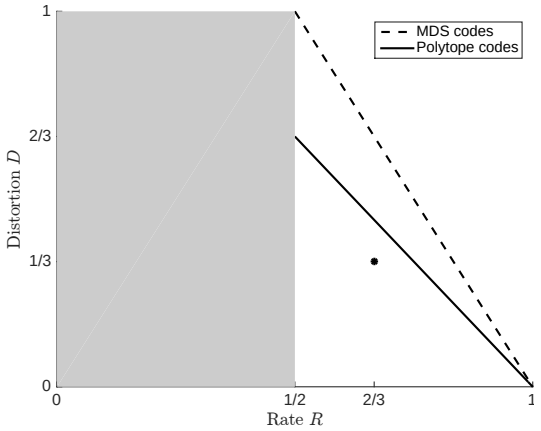

Figure 3.1: Rate-distortion (R-D) tradeoff for $N = 3$ packets and $T = 1$ error. The dashed and solid lines indicate the achievable performance using MDS and polytope codes, respectively. The asterix indicates the rate-distortion performance of the scheme in (3.1). For rates below $1/2$, finite distortion is unachievable for any feasible code. The R-D region for MDS codes is: $\{(R, D)|R \geq \frac{1}{2}, D \geq 2(1 - R)\}$. The R-D region for Polytope codes is: $\{(R, D)|R \geq \frac{1}{2}, D \geq \frac{4}{3}(1 - R)\}$.

can identify which packet was altered, exclude it, and then determine all of the source components from the remaining packets. Thus the maximum number of components about which the decoder can be uncertain is one. It follows that the R-D pair $(2/3, 1/3)$ is achievable. This point lies outside the region achieved by polytope codes, as shown in Fig. 3.1.

Since the rate-distortion pair $(1/(N - 2T), 0)$ is achievable, and the set of achievable pairs is convex, to show part 2) of Theorem 1 it suffices to show that

$$\left(\frac{1}{N - T}, \frac{F(T)}{N}\right)$$

is achievable. In the next section, we will show how polytope codes can be used toward this end.

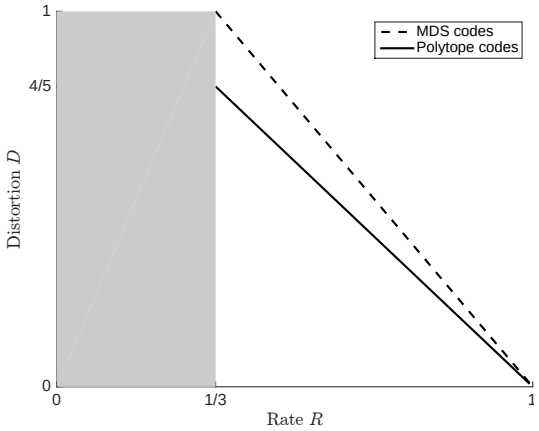

Figure 3.2: Rate-distortion tradeoff for $N = 5$ packets and $T = 2$ errors.

3.2 Polytope Codes

Polytope codes were introduced by Kosut, Tong, and Tse [26] in the context of network coding with adversarial nodes. Polytope codes are akin to linear MDS codes, except that the arithmetic operations are performed over the reals and extra low rate "check" information is included in the transmission. Our construction is somewhat simpler than the one given in [26]. To understand this construction it is helpful to begin with the special case in which there are $N = 3$ packets subject to at most $T = 1$ error.

3.2.1 $N = 3, T = 1$ case

One trivial design for this case is to simply send the true source sequence in all three packets. Since there is at most one error, the decoder can always recover the source sequence by using a majority rule. That is, it can recover the source exactly when there are no errors but also when there is one. As such, this scheme achieves the rate-distortion pair $(1, 0)$. This scheme is unsatisfactory, however, since it is wasteful when there are

no errors.

One may consider using a $(3, 2)$ MDS code instead. For instance, we could choose the blocklength $n = 2$ and encode two source symbols x_1 and x_2 into three packets as

$$x_1 \qquad x_2 \qquad x_1 \oplus x_2, \tag{3.2}$$

where $\oplus$ denotes modulo arithmetic. The decoder can determine whether a single error has been introduced by verifying whether the received packets satisfy the linear relation in (3.2). If so, then there are no errors, and the decoder can reproduce the source exactly. Thus it is feasible. If not, then the decoder knows that one error is present, but it has no way of identifying which packet is in error. Since there is an infinite penalty for guessing a source symbol incorrectly, it must output the all-erasure string, achieving the rate-distortion pair $(1/2, 1)$. The striking thing about this example is that the decoder always receives at least one of the two source symbols correctly; the problem is that it does not know which of the two is correct.

Now suppose that the source is viewed as a pair of vectors of positive integers of length N_0, $x_1^{N_0}$ and $x_2^{N_0}$, and the three transmitted packets consist of

$$y_1^{N_0} = x_1^{N_0} \qquad y_2^{N_0} = x_2^{N_0} \qquad y_3^{N_0} = x_1^{N_0} + x_2^{N_0}, \tag{3.3}$$

where now the addition is performed over the reals. We also send the quantities

$$\langle y_i^{N_0}, y_j^{N_0} \rangle \tag{3.4}$$

for all i and j $(1 \leq i, j \leq 2)$ as part of each packet. As before, the decoder can always detect whether an error has been introduced. If it detects no error, it can output the source sequence correctly. But now if it detects an error, it can always identify at least one of the three packets as correct by the following reasoning. Since the inner products in (3.4) are included in all three packets, they can always be recovered correctly. Further

more, the decoder can correctly compute $\langle y_i^{N_0}, y_3^{N_0} \rangle$ for all i. Let

$$\bar{y}_1^{N_0} \qquad \bar{y}_2^{N_0} \qquad \bar{y}_3^{N_0}, \tag{3.5}$$

denote the vectors in the three received packets, and assume that exactly one of them has been altered.

Now construct a graph with nodes v_1, v_2, and v_3 and an edge (or a self-loop if $i = j$) between v_i and v_j if

$$\langle \bar{y}_i^{N_0}, \bar{y}_j^{N_0} \rangle = \langle y_i^{N_0}, y_j^{N_0} \rangle.$$

We call this the *syndrome graph*.

If v_i does not have a self-loop:

$$\|\bar{y}_i^{N_0}\|^2 \neq \|y_i^{N_0}\|^2,$$

then we know that the ith packet is in error and the other two must be correct. So we shall assume that

$$\|\bar{y}_i^{N_0}\|^2 = \|y_i^{N_0}\|^2,$$

for all i. Under this assumption, if the syndrome graph is fully connected, then for some collection of constants a_{ij} we must have

$$\|\bar{y}_3^{N_0} - \bar{y}_1^{N_0} - \bar{y}_2^{N_0}\|^2 = \sum_{i,j} a_{ij} \langle \bar{y}_i^{N_0}, \bar{y}_j^{N_0} \rangle \tag{3.6}$$

$$= \sum_{i,j} a_{ij} \langle y_i^{N_0}, y_j^{N_0} \rangle \tag{3.7}$$

$$= \|y_3^{N_0} - y_1^{N_0} - y_2^{N_0}\|^2 \tag{3.8}$$

$$= 0. \tag{3.9}$$

Thus

$$\bar{y}_3^{N_0} = \bar{y}_1^{N_0} + \bar{y}_2^{N_0}, \tag{3.10}$$

which contradicts the assumption that one of the these vectors was altered.

Thus the graph must be missing at least one edge. Since only one packet can be received in error, the graph cannot be missing all three edges, however. Thus it must have either one edge or two. If it has exactly one edge, then the vector with no edges must be the one in error, so the other two vectors can be identified as correct. If the graph has two edges, then the vector with two edges must be correct. In the end, then, the decoder can always recover at least one of the transmitted packets correctly. This is of course not the same as recovering one of the source vectors—if the decoder recovers $x_3^{N_0}$ then it cannot reproduce any of the source symbols with certainty. But using a "layering" argument one can transform this code into one for which decoding any of the three transmitted packets correctly allows one to recover some positive fraction of the source symbols correctly (see Section 3.3).

The property that the decoder can always correctly recover a transmitted packet even when the number of errors is outside the decoding radius of the code we call *partial decodability.* Note that to obtain partial decodability in the above construction it is crucial that the arithmetic operations be performed over the reals; (3.10) is not implied by (3.9) under modulo arithmetic. The code is also nonlinear. These two features distinguish the codes described here from conventional MDS codes and network codes [19,20,40]. The antecedent of our code is the polytope code of Kosut, Tong, and Tse, mentioned earlier. The Kosut *et al.* construction requires that $(y_1^{N_0}, y_2^{N_0}, y_3^{N_0})$ have a certain joint empirical distribution. This ensures that the norms and inner products in (3.4) equal certain prespecified values and so they do not need to be transmitted. Encoding for the Kosut *et al.* codes is more complex than for the codes provided here; the former is akin to implementing a constant-composition channel code. The Kosut *et al.* construction also requires selecting the joint empirical distribution of $(y_1^{N_0}, y_2^{N_0}, y_3^{N_0})$ in a particular way that

ensures partial decodability (see [26, Theorem 4]). The present construction eliminates this step.

As with the Kosut *et al.* codes, the partial decodability provided here comes at slight cost in rate compared with conventional MDS codes; one must send the norms and inner products in (3.4) in addition to the vectors, and $x_3^{N_0}$ can take larger values than either $x_1^{N_0}$ or $x_2^{N_0}$ can because the addition in (3.3) is done over the reals. But in the limit of a large source blocklength, this penalty can be made arbitrarily small, and the rate can be made arbitrarily close to $1/2$.

We conclude this subsection with a concrete example of the encoding. Suppose we wish to send a binary source ($K = 2$) with blocklength $n = 2L_0N_0$, where $L_0 = 4$ and $N_0 = 3$. Suppose the source realization is

$$0000 \quad 1111 \quad 0010 \quad 0000 \quad 0100 \quad 0001.$$

We first convert it to a vector in $\{1, \ldots, 2^{L_0}\}^6$,

$$(16, 15, 2, 16, 4, 1),$$

where 0000 is mapped to 2^{L_0} and otherwise we use the usual binary representation. Now define $x_1^3 = (16, 15, 2)$ and $x_2^3 = (16, 4, 1)$. We use the generator matrix:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

and let

$$\begin{bmatrix} y_1^3 \\ y_2^3 \\ y_3^3 \end{bmatrix} = A \begin{bmatrix} x_1^3 \\ x_2^3 \end{bmatrix}.$$

to get $y_1^3 = (16, 15, 2)$, $y_2^3 = (16, 4, 1)$, $y_3^3 = (32, 19, 3)$. Next, compute the inner products and norms

$$F_{11} = \|y_1^3\|^2 = 16^2 + 15^2 + 2^2 = 485;$$
$$F_{22} = \|y_2^3\|^2 = 16^2 + 4^2 + 1^2 = 273;$$
$$F_{12} = \langle y_1^3, y_2^3 \rangle = 16^2 + 15 \times 4 + 2 \times 1 = 318.$$

Thus the three packets are

$$(16, 15, 2, 485, 273, 318)$$
$$(16, 4, 1, 485, 273, 318)$$
$$(32, 19, 3, 485, 273, 318).$$

Each component of each y_i^3 is an element in $\{1, \ldots, 32\}$ and thus requires five bits to describe. Each of F_{11}, F_{22}, F_{12} is an element in $\{3, \ldots, 768\}$ and thus requires ten bits to describe. Each message is therefore 45 bits long.

We next describe how to extend this idea to general N and T. The resulting construction is then used to prove Theorem 1. See [14] for a slightly different decoding algorithm that yields the same performance.

3.2.2 General (N, T): Source

Consider a source message x^n ($x^n \in \mathcal{X}^n$) with length $n = (N - T)N_0L_0$ for some large natural numbers N_0 and L_0. Divide the message into $(N - T)N_0$ subvectors, each having L_0 symbols. We can use an L_0-length vector (each entry taken from $[K]$) to represent K^{L_0} integers $\{1, ..., K^{L_0}\}$; here we use $(0, ..., 0)$ to represent K^{L_0}. Thus, the original source message can also be viewed as an integer vector with length $(N-T)N_0$ whose coordinates are drawn from $\{1, \ldots, K_0^L\}$. Moreover, x^n can be viewed as a concatenation of $N - T$

vectors, each having N_0 entries in $\{1, ..., K^{L_0}\}$. In what follows, we will view the source vector in this way and write

$$x^n = (x_{1,L_0}^{N_0}, ..., x_{N-T,L_0}^{N_0}).$$

3.2.3 Encoding Functions

The encoding is performed with the aid of an eligible generator matrix.

Definition 5 *A is an eligible $(N, N-T)$-generator matrix if its entries are nonnegative integers and*

1. *A is an $N \times (N-T)$ matrix of the following form:*

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ a_{1,1} & a_{1,2} & \cdots & a_{1,N-T} \\ \vdots & \vdots & \vdots & \vdots \\ a_{T,1} & a_{T,2} & \cdots & a_{T,N-T} \end{bmatrix},$$

2. *Every $(N-T) \times (N-T)$ submatrix of A is nonsingular with respect to the field of real numbers.*

The existence of such matrix is guaranteed by the following lemma.

Lemma 1 *For any $T \geq 1$ and $N \geq T$ there exists an eligible $(N, N-T)$-generator matrix of the form*

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ \alpha_1^1 & \alpha_1^2 & \cdots & \alpha_1^{N-T} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_T^1 & \alpha_T^2 & \cdots & \alpha_T^{N-T} \end{bmatrix}, \tag{3.11}$$

where $\alpha_1, \ldots, \alpha_T$ are distinct positive integers. We call such a matrix a V-matrix*, since its lower portion has a Vandermonde structure.*

Proof: One can use the following scheme: pick a sufficiently large integer α, and then pick α_t i.i.d. uniformly at random from $[\alpha]$. For sufficiently large α, the matrix will have the desired property with high probability. See the appendix for a complete argument that does not rely on random selection.

The encoding functions are then as follows:

1. We generate N vectors, $y_1^{N_0} \ldots y_N^{N_0}$ via the linear transformation

$$\begin{bmatrix} y_{1,L_0}^{N_0} \\ \vdots \\ y_{N,L_0}^{N_0} \end{bmatrix} = A \begin{bmatrix} x_{1,L_0}^{N_0} \\ \vdots \\ x_{N-T,L_0}^{N_0} \end{bmatrix},$$

where A is an eligible $(N, N-T)$-generator matrix provided by Lemma 1. In particular, we have

$$y_{i,L_0}^{N_0} = x_{i,L_0}^{N_0}$$

for all $1 \leq i \leq N-T$. Let $\alpha = \max_{i,j} \alpha_{i,j}$. Then $y_{i,L_0}^{N_0}$ is a vector of length N_0, of which the coordinates are positive integers that do not exceed $\alpha(N-T)K^{L_0}$. Thus, each vector can be encoded using $(L_0 + \left\lceil \log_K(\alpha(N-T)) \right\rceil)N_0$ symbols.

2. We also transmit $(N-T) + \binom{N-T}{2}$ norms/inner products:

$$F_{ij} = \langle x_{i,L_0}^{N_0}, x_{j,L_0}^{N_0} \rangle, \forall 1 \leq i \leq j \leq N-T$$

in all N packets. This requires that $\lceil (2L_0 + \log_K N_0)[(N-T) + \binom{N-T}{2}] \rceil$ extra symbols to be included in each packet.

3.2.4 General (N, T): Decoding Functions

The decoder receives $\bar{y}_{1,L_0}^{N_0}, ..., \bar{y}_{N,L_0}^{N_0}$ and the norms/inner products between $\{x_1^{N_0}, ..., x_{N-T}^{N_0}\}$. The decoder will identify a subset of the components of $y_1^{N_0}, ..., y_N^{N_0}$ that it is sure have been unaltered.[2] We first note that the norms and inner products can always be recovered without error.

Lemma 2 *The decoder can correctly recover F_{ij} for $i, j \in \{1, ..., N-T\}$ when $N \geq 2T+1$. Since $y_1^{N_0}, ..., y_N^{N_0}$ are linear combinations of $x_1^{N_0}, .., x_{N-T}^{N_0}$. This means that we can correctly recover $F_{ij} = \langle y_i^{N_0}, y_j^{N_0} \rangle$ for $i, j \in \{1, ..., N\}$.*

The proof of this lemma is straightforward and omitted.

Use a graph G with N vertices $V = \{v_1, v_2, ..., v_N\}$ to represent the N received packets. The ith received packet is $\bar{C}_i$, which is composed of the K-symbol representations of $\bar{y}_i^{N_0}$ and $\bar{F}_{j_1 j_2}^{(i)}$ $(1 \leq j_1 \leq j_2 \leq N-T)$. According to Lemma 2, we can correctly recover

[2] Later we will show how to use this identification to prove Theorem 1.

$F_{ij} = \langle y_i^{N_0}, y_j^{N_0} \rangle$. We draw an edge between vertex v_i and vertex v_j ($i \neq j$) iff

$$\langle \bar{y}_i^{N_0}, \bar{y}_j^{N_0} \rangle = F_{ij}.$$

We draw a self-loop on vertex v_i iff

$$\langle \bar{y}_i^{N_0}, \bar{y}_i^{N_0} \rangle = F_{ii}.$$

As in the $N = 3, T = 1$ case, we call this the *syndrome graph.*

The decoder then performs the following operations:

1. Delete all vertices with no loops and their incident edges in the syndrome graph. Let $\hat{G} = (\hat{V}, \hat{\mathcal{E}})$ denote the new graph.
2. Let V' be the set of vertices v_i in $\hat{V}$ such that v_i is contained in a clique of size at least $N - T$ in $\hat{G}$.
3. Let V^* be the set of vertices v_i in V' such that $(v_i, v_j) \in \hat{\mathcal{E}}$ for all v_j in V'.
4. Output the codewords corresponding to the vertices in V^* as correct.

We shall show that the rate of this code can be made arbitrarily close to $1/(N - T)$. We shall then prove that the codewords $\bar{y}_i^{N_0}$ on channels corresponding to the vertices $v_i \in V^*$ are correct.

3.2.5 General (N, T): Coding Rate

Proposition 3 *For any $\epsilon > 0$, there exists natural numbers L_0 and N_0 such that the rate of each packet does not exceed $1/(N - T) + \epsilon$.*

Proof: The rate of each packet is upper bounded by

$$\frac{(L_0 + \left\lceil \log_K(\alpha(N-T)) \right\rceil) N_0}{L_0 N_0 (N-T)} + \frac{\lceil 2L_0 + \log_K N_0 \rceil \left((N-T) + \binom{N-T}{2} \right)}{L_0 N_0 (N-T)}, \tag{3.12}$$

where we recall that $\alpha = \max_{i,j} \alpha_{i,j}$. If we let $N_0 = L_0$ and send both to infinity, the second term tends to zero while the first term tends to $1/(N-T)$.

3.2.6 General (N, T): Partial Decodability of Polytope Codes

We are interested in polytope codes because of the following property.

Theorem 2 *Given T, when $N \geq T + \left\lfloor \frac{T^2}{4} \right\rfloor + 2$, the decoder can identify least $N - T - \left\lfloor \frac{T^2}{4} \right\rfloor - 1$ of the transmitted packets as being received correctly.*

We shall prove Theorem 2 via a sequence of lemmas. The first two establish that the codewords associated with nodes in V^* were received correctly.

Lemma 4 *Suppose the k packets $i_1, \ldots, i_k$ are unaltered, and let i_{k+1} be some other packet for which there exists $l_1, \ldots, l_k$ such that*

$$y_{i_{k+1}}^{N_0} = \sum_{j=1}^{k} l_j y_{i_j}^{N_0}. \tag{3.13}$$

If there is a self-loop on v_{i_k+1} in G, and $(v_{i_{k+1}}, v_{i_j}) \in \mathcal{E}$ for all $j \in \{1, \ldots, k\}$, then the codeword $\bar{y}_{i_{k+1}}^{N_0}$ in packet i_{k+1} is also unaltered.

Proof: We may rewrite (3.13) as

$$\left\| y_{i_{k+1}}^{N_0} - \sum_{j=1}^{k} l_j y_{i_j}^{N_0} \right\|^2 = 0. \tag{3.14}$$

Since there is a self-loop on $v_{i_{k+1}}$,

$$\langle \bar{y}_{i_{k+1}}^{N_0}, \bar{y}_{i_{k+1}}^{N_0} \rangle = \langle y_{i_{k+1}}^{N_0}, y_{i_{k+1}}^{N_0} \rangle.$$

Moreover, since there is an edge $(v_{i_{k+1}}, v_{i_j})$ for all $j \in \{1, \ldots, k\}$,

$$\langle \bar{y}_{i_{k+1}}^{N_0}, \bar{y}_{i_j}^{N_0} \rangle = \langle y_{i_{k+1}}^{N_0}, y_{i_j}^{N_0} \rangle.$$

By expanding the left-hand side of (3.14) in terms of inner products, as in (3.6)-(3.9), we have that

$$0 = \left\| y_{i_{k+1}}^{N_0} - \sum_{j=1}^{k} l_j y_{i_j}^{N_0} \right\|^2 = \left\| \bar{y}_{i_{k+1}}^{N_0} - \sum_{j=1}^{k} l_j y_{i_j}^{N_0} \right\|^2 = \left\| \bar{y}_{i_{k+1}}^{N_0} - y_{i_{k+1}}^{N_0} \right\|^2$$

where we have used the assumption that packets $i_1, \ldots, i_k$ are unaltered, and (3.13). This proves that packet i_{k+1} is unaltered.

Lemma 5 *For any $v_i \in V^*$, we have $\bar{y}_i^{N_0} = y_i^{N_0}$.*

Proof: There must exist $N - T$ packets that are unaltered. Suppose they are packets $i_1, \ldots, i_{N-T}$. Then $v_{i_1}, \ldots, v_{i_{N-T}}$ must form a clique in the syndrome graph $\hat{G}$. From the definition of V^*, for any vertex $v_i \in V^*$, there is a self-loop on v_i and $(v_i, v_{i_j}) \in \mathcal{E}$ for all $j \in \{1, \ldots, N - T\}$. By construction, every $(N - T) \times (N - T)$ submatrix of generator matrix A is nonsingular. This implies that the vector $y_i^{N_0}$ can be represented as a linear combination of the other $N - T$ vectors

$$y_{i_{k+1}}^{N_0} = \sum_{j=1}^{N-T} l_j y_{i_j}^{N_0}$$

for some linear coefficients l_j. By Lemma 4, the codeword $y_i^{N_0}$ in packet i is unaltered.

The final lemma lower bounds the size of V^*. It is a purely graph-theoretic assertion that may have independent uses.

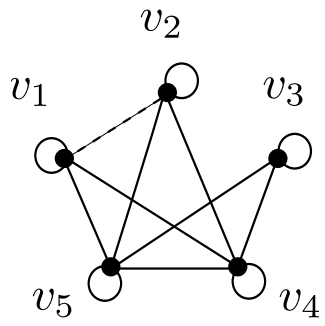

Figure 3.3: An illustration for $\mathcal{E}'$ with $N = 5$, $T = 2$.

Lemma 6 *Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Let Let V' denote the set of nodes that are contained in a clique of size at least $N - T$ and assume that $0 < |V'| \leq N$. Let*

$$V^* = \{v \in V' : (v, \tilde{v}) \in \mathcal{E}\ \forall \tilde{v} \in V'\}$$

denote the set of nodes in V' that are connected to all nodes in V'. Then $|V^*| \geq N - F(T)$, where $F(T) = T + \lfloor \frac{T^2}{4} \rfloor + 1$.

Proof: For any set of edges $\mathcal{E}_0$, let

$$\mathcal{N}(v_i, \mathcal{E}_0) := \{v_j \in V' \backslash \{v_i\} : (v_i, v_j) \notin \mathcal{E}_0\},$$

We construct a set of edges $\mathcal{E}' \supset \mathcal{E}$ as follows. Begin by setting $\mathcal{E}' = \mathcal{E}$. If there is a pair $v_i, v_j \in V'$ such that $(v_i, v_j) \notin \mathcal{E}'$,

$$|\mathcal{N}(v_i, \mathcal{E}')| > 1, \text{ and } |\mathcal{N}(v_j, \mathcal{E}')| > 1,$$

then add (v_i, v_j) to $\mathcal{E}'$. Repeat until there is no such pair v_i, v_j. Fig. 3.3 is an illustration for $N = 5$, $T = 2$ where $\mathcal{V} = \{v_1, ..., v_5\}$ and $\mathcal{E}$ is the set of the solid edges (including 5 self-loops). We can see that $V' = \mathcal{V}$, $V^* = \{v_4, v_5\}$ and $\mathcal{E}'$ is the union of $\mathcal{E}$ and exactly one of the edges (v_1, v_2), (v_2, v_3), and (v_1, v_3).

Note that for the resulting $\mathcal{E}'$, for $v_i \in V'$, $|\mathcal{N}(v_i, \mathcal{E}')| = 0$ if and only if $|\mathcal{N}(v_i, \mathcal{E})| = 0$. Thus

$$V^* = \{v_i \in V' : |\mathcal{N}(v_i, \mathcal{E}')| = 0\}.$$

Moreover, for any pair $(v_i, v_j) \in V'$ with $(v_i, v_j) \notin \mathcal{E}'$, either $|\mathcal{N}(v_i, \mathcal{E}')| = 1$ or $|\mathcal{N}(v_j, \mathcal{E}')| = 1$. For convenience, we write $\mathcal{N}(v_i) := \mathcal{N}(v_i, \mathcal{E}')$ from now on.

Let v_{i_0} be an element of V' maximizing $|\mathcal{N}(v)|$ over v, and let

$$\tilde{l} := |\mathcal{N}(v_{i_0})|.$$

For each element $v_i \in V'$, let $C(v_i)$ be a clique of size exactly $N - T$ containing v_i. [3] Since $\mathcal{E}' \supset \mathcal{E}$, $C(v_i)$ is also a clique on the graph with edges in $\mathcal{E}'$. Let $C_0 = C(v_{i_0})\backslash\{v_{i_0}\}$. Fix $v_{i_1} \in C_0$, and suppose $(v_{i_1}, v_l) \notin \mathcal{E}'$ for $v_l \in V'$. We claim that v_l cannot be in $\mathcal{N}(v_{i_0})$. If it were, then $|\mathcal{N}(v_l)| \geq 2$, in which case $\tilde{l} \geq 2$, which would imply that $|\mathcal{N}(v_{i_0})| \geq 2$. On the other hand, when $v_l \in \mathcal{N}(v_{i_0})$, either $|\mathcal{N}(v_{i_0})| = 1$ or $|\mathcal{N}(v_l)| = 1$ which leads to a contradiction. Moreover, v_l cannot be in $C(v_{i_0})$ by definition. Hence, if $(v_{i_1}, v_l) \neq \mathcal{E}'$, then $v_l \in \mathcal{D}$, where

$$\mathcal{D} := V'\backslash(\mathcal{N}(v_{i_0}) \cup C(v_{i_0})).$$

In particular, if $v_j \in C_0 \cap V'\backslash V^*$, then $(v_j, v_k) \notin \mathcal{E}'$ for some $v_k \in \mathcal{D}$; i.e. $v_j \in \mathcal{N}(v_k)$. Thus

$$\begin{aligned} V'\backslash V^* &\subset (V'\backslash(C_0 \cup V^*)) \cup (V' \cap C_0\backslash V^*) \\ &\subset \{v_{i_0}\} \cup (V'\backslash C(v_{i_0})) \cup \bigcup_{v\in\mathcal{D}} (\mathcal{N}(v) \cap C_0) \\ &\subset \{v_{i_0}\} \cup \mathcal{N}(v_{i_0}) \cup \mathcal{D} \cup \bigcup_{v\in\mathcal{D}} (\mathcal{N}(v) \cap C_0). \end{aligned}$$

Hence,

$$\begin{aligned} |V'| - |V^*| &\leq 1 + |\mathcal{N}(v_{i_0})| + |\mathcal{D}| + \Sigma_{v\in\mathcal{D}}|\mathcal{N}(v)| \\ &\leq (|\mathcal{D}| + 1)(\tilde{l} + 1), \end{aligned} \tag{3.15}$$

[3] There may be several such cliques, in which case $C(v_i)$ can be chosen to be any one of them.

where we have used the fact that $|\mathcal{N}(v)| \leq \tilde{l}$ for all $v \in V'$. Since $\mathcal{N}(v_{i_0}), \mathcal{C}(v_{i_0}) \subset V'$ and $\mathcal{N}(v_{i_0}) \cap \mathcal{C}(v_{i_0}) = \emptyset$,

$$|\mathcal{D}| = |V'| - |\mathcal{N}(v_{i_0})| - |\mathcal{C}(v_{i_0})| = |V'| - \tilde{l} + T - N.$$

Substituting this into (3.15) gives

$$\begin{aligned} |V^*| &\geq |V'| - (|\mathcal{D}| + 1)(\tilde{l} + 1) \\ &= |V'| - (T - \tilde{l} + |V'| - N + 1)(\tilde{l} + 1) \\ &\geq N - (T - \tilde{l} + 1)(\tilde{l} + 1) \\ &\geq N - F(T). \end{aligned}$$

[Proof of Theorem 2]*Proof:* For each $i \in V^*$, we have $\bar{y}_i^{N_0} = y_i^{N_0}$ by Lemma 5 and $|V^*| \geq N - F(T)$ by Lemma 6.

3.3 Proof of Theorem 1

We next show how to use polytope codes to create a code for our original problem. The main difficulty is that, in a polytope code, some of the packets contain only parities, and even if the decoder can determine such packets with certainty, it cannot necessarily recover any of the original source symbols. We circumvent this issue with a layered construction. First we prove the impossibility result in part 1).

3.3.1 Proof of Theorem 1 Part 1)

Suppose the maximum number of altered packets T satisfies $1 \leq T \leq N$. In this subsection, we will prove that if $0 \leq R < \frac{1}{N-T}$, then there is no finite D for which (R, D) is achievable.

Choose $0 < \epsilon < 1/4$ so that

$$1 - (N - T)(R + \epsilon) \geq h(\epsilon) + h(2\epsilon) + 3\epsilon \tag{3.16}$$

where $h(\cdot)$ is the (base-K) binary entropy function

$$h(\theta) = -\theta \log_K \theta - (1 - \theta) \log_K (1 - \theta) \quad \theta \in (0, 1)$$

with $h(\theta) = 0$ if $\theta = 0$ or $\theta = 1$. Suppose there exists a code $(f_1, \ldots, f_N, g)$ with rate at most $R + \epsilon$ satisfying

$$D_0(f_1, \ldots, f_N, g) \leq \epsilon \tag{3.17}$$

$$D_T(f_1, \ldots, f_N, g) \leq D + \epsilon \tag{3.18}$$

for some D and let n denote the length of the source string that it encodes. Consider endowing the space $\mathcal{X}^n$ with an i.i.d. uniform probability distribution. By Fano's and Jensen's inequalities,

$$\begin{aligned}
\frac{1}{n} H(X^n | C_1, \ldots, C_N) &\leq \frac{1}{n} \sum_{i=1}^{n} H(X_i | C_1, \ldots, C_N) \\
&\leq \frac{1}{n} \sum_{i=1}^{n} h(\epsilon_i) + \epsilon_i \\
&\leq h\left(\frac{1}{n} \sum_{i=1}^{n} \epsilon_i \right) + \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \\
&\leq h(\epsilon) + \epsilon,
\end{aligned}$$

where

$$\epsilon_i = \mathbf{Pr}(\hat{X}_i \neq X_i)$$

and

$$\hat{X}^n = g(C_1, \ldots, C_N).$$

Thus

$$\begin{aligned}
&H(X^n|C_{T+1},\dots,C_N)\\
&\geq I(X^n;C_1,\dots,C_T|C_{T+1},\dots,C_N)\\
&= I(X^n;C_1,\dots,C_N) - I(X^n;C_{T+1},\dots,C_N)\\
&= I(X^n;C_1,\dots,C_N) - H(C_{T+1},\dots,C_N)\\
&\geq n - n(h(\epsilon)+\epsilon) - n(N-T)(R+\epsilon)\\
&\geq n(h(2\epsilon)+2\epsilon),
\end{aligned}$$

by (3.16), where all entropy and mutual information quantities are base-K. It follows that there exist $c_{T+1},\dots,c_N$ such that

$$H(X^n|C_{T+1}=c_{T+1},\dots,C_N=c_N) \geq n(h(2\epsilon)+2\epsilon)$$

and by the cardinality bound on entropy, the set of realizations of X^n with positive probability given $C_{T+1}=c_{T+1},\dots,C_N=c_N$, which we shall call $\mathcal{X}^n_{c_{T+1},\dots,c_N}$, satisfies

$$|\mathcal{X}^n_{c_{T+1},\dots,c_N}| \geq K^{nh(2\epsilon)}K^{2\epsilon n}. \tag{3.19}$$

Now any subset of $\mathcal{X}^n$ with Hamming diameter at most $2\epsilon n$ must contain at most

$$\binom{n}{\lfloor 2\epsilon n\rfloor}(K-1)^{\lfloor 2\epsilon n\rfloor} < K^{nh(2\epsilon)}K^{2\epsilon n},$$

sequences, where the inequality follows from, e.g., [10, Example 11.1.3]. Thus there exist x^n and $\tilde{x}^n$ in $\mathcal{X}^n_{c_{T+1},\dots,c_N}$ such that if

$$I_{\neq} = \{i : x_i \neq \tilde{x}_i\}$$

then $|I_{\neq}| \geq 2\epsilon n$. For these two sequences, we must have

$$f_i(x^n) = f_i(\tilde{x}^n) \text{ for all } T+1 \leq i \leq N.$$

Let $\hat{x}^n$ denote the receiver's output when it receives the messages

$$(f_1(x^n),\dots,f_N(x^n)).$$

By (3.17), we must have

$$d(x^n, \hat{x}^n) \leq \epsilon$$

so that if we let

$$I_e = \{i : \hat{x}_i = e\}$$

then $|I_e| \leq \epsilon n$ and $\hat{x}_i = x_i$ for $i \notin I_e$. Now if the true source sequence is $\tilde{x}^n$ and the adversary alters the first T packets so that the decoder receives

$$\begin{aligned}&(f_1(x^n), \ldots, f_T(x^n), f_{T+1}(\tilde{x}^n), \ldots, f_N(\tilde{x}^n))\\ &= (f_1(x^n), \ldots, f_T(x^n), f_{T+1}(x^n), \ldots, f_N(x^n))\end{aligned}$$

then the decoder will output $\hat{x}^n$, so by (3.18),

$$D + \epsilon \geq d(\tilde{x}^n, \hat{x}^n) = \frac{1}{n}\sum_{i=1}^{n} d(\tilde{x}_i, \hat{x}_i). \tag{3.20}$$

Now

$$\begin{aligned}|I_{\neq} \backslash I_e| &\geq |I_{\neq}| - |I_e|\\ &\geq 2\epsilon n - \epsilon n\\ &> 0.\end{aligned}$$

It follows that there exists i in $I_{\neq}\backslash I_e$, for which we must have $d(\tilde{x}_i, \hat{x}_i) = \infty$. It follows from (3.20) that D must be infinite.

3.3.2 Proof of Theorem 1 Part 2)

Suppose the maximum number of altered packets T satisfies $T \geq 1$ and the number of packets N satisfies $N \geq T + \lfloor \frac{T^2}{4} \rfloor + 2$. In this subsection, we will prove: for any $\frac{1}{N-T} \leq R \leq \frac{1}{N-2T}$, the rate-distortion pair

$$\left(R, \frac{F(T)(N-T)(1-(N-2T)R)}{NT}\right)$$

is achievable, where $F(T) = T + \lfloor \frac{T^2}{4} \rfloor + 1$.

As noted earlier it suffices to show that the R-D pair $(\frac{1}{N-T}, \frac{F(T)}{N})$ is achievable. To show this we use a "layered" construction in which we use N polytope codes whose transformation matrices are row rotations of each other. Divide the source into N equal-sized parts. The first part is encoded into packets using a polytope code with transformation matrix

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ \alpha_1^1 & \alpha_1^2 & \cdots & \alpha_1^{N-T} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_T^1 & \alpha_T^2 & \cdots & \alpha_T^{N-T} \end{bmatrix}.$$

The second part is encoded using the transformation matrix

$$A = \begin{bmatrix} \alpha_T^1 & \alpha_T^2 & \cdots & \alpha_T^{N-T} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ \alpha_1^1 & \alpha_1^2 & \cdots & \alpha_1^{N-T} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{T-1}^1 & \alpha_{T-1}^2 & \cdots & \alpha_{T-1}^{N-T} \end{bmatrix},$$

i.e., the first downward row rotation. The other parts of the source are encoded similarly.

The rate of this code can be made arbitrarily close to $1/(N-T)$. At the decoder, we form a syndrome graph in which there is an edge between packets i and j (allowing for

$j = i$) if there is an edge between i and j in the syndrome graphs of all of the layers. For this syndrome graph, delete all nodes without self-loops, along with their edges. The resulting graph must have at least one clique of size at least $N-T$, due to the presence of at least $N-T$ unaltered packets. Thus Lemma 6 implies that there are at least $N-F(T)$ nodes that are connected to all nodes contained in a clique of size at least $N-T$. In particular, these $N-F(T)$ nodes must be connected to an unaltered set of nodes of size $N-T$. By Lemma 4, the codewords in all of these $N-F(T)$ packets were received correctly. For each packet, $N-T$ of its layers correspond to systematic rows of the matrix and T layers correspond to parities. Thus the decoder can reconstruct a fraction

$$\frac{(N-T)(N-F(T))}{N(N-T)} = \frac{N-F(T)}{N}$$

of the source symbols.

3.4 An Impossibility Result

By definition, a polytope code

$$(f_1, \ldots, f_N, g)$$

is characterized by (N, T, A, N_0, L_0), where N is the number of packets, T is the maximum number of packets that can be altered, A is an eligible $(N, N-T)$-generator matrix, and N_0 and L_0 are encoding parameters (see Section 3.2). From Theorem 1, we know that for

$$N \geq F(T) + 1 \text{ and } \frac{1}{N-T} \leq R \leq \frac{1}{N-2T}$$

the R-D pair

$$\left(R, \frac{F(T)(N-T)(1-(N-2T)R)}{NT}\right)$$

is achievable using polytope codes. However, when $N \leq F(T)$, the decoder in Section 3.2.4 no longer works.

This raises the question of whether our design can be improved when $N \leq F(T)$, especially since $F(T)$ grows superlinearly with T. We next show the following impossibility result. When $N = F(T)$, for all sufficiently large N_0 and L_0, our existing polytope code construction lacks the partial decodability property: there exists a set of received packets for which there is no single packet that can be determined to be correct with certainty. Thus, at least as far as partial decodability is concerned, neither the decoder nor the analysis can be improved to relax the $N \geq F(T) + 1$ condition; the code itself would need to change. Recall that, for polytope codes, in order to drive the rate to $1/(N - T)$, we send both N_0 and L_0 to infinity; see (3.12).

To state and prove this result, we use the concept of *possible transmitted codewords*.

Definition 6 *Fix N_0, L_0 and K. Given a set of received codewords $\{\bar{y}_1^{N_0}, ..., \bar{y}_N^{N_0}\}$ and recovered $\{F_{j_1 j_2}\}$ for $j_1, j_2 \in [N]$ (see Lemma 2), if a set of codewords $\{\bar{x}_1^{N_0}, ..., \bar{x}_N^{N_0}\}$ satisfies:*

1. *$F_{j_1 j_2} = \langle \bar{x}_{j_1}^{N_0}, \bar{x}_{j_2}^{N_0} \rangle$, for all $j_1, j_2 \in [N]$;*
2. *The identity $\bar{x}_j^{N_0} = \bar{y}_j^{N_0}$ holds for at least $N - T$ values of j out of $j \in [N]$;*
3. *$\bar{x}_{N-T+i}^{N_0} = \sum_{j=1}^{N-T} a_{i,j} \bar{x}_j^{N_0}$ for all $i \in [T]$.*

then this set of codewords is called a Possible Transmitted Codeword (PTC) *for $\{\bar{y}_1^{N_0}, ..., \bar{y}_N^{N_0}\}$ and $\{F_{j_1 j_2}\}$. Further, let*

$$\text{PTC}(\bar{y}_1^{N_0}, ..., \bar{y}_N^{N_0}, \{F_{j_1 j_2}\}) = \{\{\bar{x}_{1,1}^{N_0}, ..., \bar{x}_{1,N}^{N_0}\}, ..., \{\bar{x}_{M,1}^{N_0}, ..., \bar{x}_{M,N}^{N_0}\}\}$$

denote the set of all possible transmitted codewords for $\{\bar{y}_1^{N_0}, ..., \bar{y}_N^{N_0}\}$ and $\{F_{j_1 j_2}\}$.

Definition 7 *Fix N_0, L_0 and K and then fix a set of received packets $\{\bar{y}_1^{N_0}, ..., \bar{y}_N^{N_0}\}$ and recovered $\{F_{j_1 j_2}\}$ for $j_1, j_2 \in [N]$. We call $\{\bar{y}_1^{N_0}, ..., \bar{y}_N^{N_0}, \{F_{j_1 j_2}\}\}$* totally undecodable

if PTC($\bar{y}_1^{N_0}, ..., \bar{y}_N^{N_0}, \{F_{j_1 j_2}\}$) *has the following property: for any* $i \in [N]$*, there exists* $\{\bar{x}_{i_1,1}^{N_0}, ..., \bar{x}_{i_1,N}^{N_0}\}$ *and* $\{\bar{x}_{i_2,1}^{N_0}, ..., \bar{x}_{i_2,N}^{N_0}\}$ *in* PTC($\bar{y}_1^{N_0}, ..., \bar{y}_N^{N_0}, \{F_{j_1 j_2}\}$) *such that* $\bar{x}_{i_1,i}^{N_0} \neq \bar{x}_{i_2,i}^{N_0}$.

Theorem 3 *Fix* $T > 1$*,* $N = F(T)$ *and let* A *be an* $(N, N - T)$ *V-matrix. Then for all sufficiently large* N_0 *and* L_0 *there exists a set of received packets* $\{\bar{y}_1^{N_0}, ..., \bar{y}_N^{N_0}\}$ *along with* $\{F_{j_1 j_2}\}$ *such that* $\{\bar{y}_1^{N_0}, ..., \bar{y}_N^{N_0}, \{F_{j_1 j_2}\}\}$ *is* totally undecodable.

Proof: See the appendix.

CHAPTER 4

GAUSSIAN VPEC

4.1 Main Results for Gaussian VPEC Problem

Theorem 4 and Theorem 5 give an inner bound and an outer bound, respectively, for $R(D_0, D_T)$. Theorem 6 shows that there is a constant gap between the two bounds. In this paper, we use log to denote the binary logarithm.

Theorem 4 *Given a distortion pair (D_0, D_T), let D_T' satisfy $\sqrt{D_T'} + \sqrt{D_T' - D_0} = \sqrt{D_T}$. If $\{R_\ell\}_{\ell\in[N]}$ satisfies the following condition: $\exists\{R_{a,\ell}\}_{\ell\in[N]}, \{R_{b,\ell}\}_{\ell\in[N]}$ s.t.*

$$\begin{aligned} & R_\ell = R_{a,\ell} + R_{b,\ell}, \quad \forall \ell \in [N]; \\ & \sum_{\ell\in A} R_{a,\ell} > \frac{1}{2}\log\frac{1}{D_T'}, \quad \forall A \subset [N], |A| = N - 2T; \\ & \sum_{\ell=1}^{N} R_{b,\ell} > \frac{1}{2}\log\frac{D_T'}{D_0}; \\ & R_{a,\ell} \geq 0, R_{b,\ell} \geq 0, \quad \forall \ell \in [N], \end{aligned} \tag{4.1}$$

then $(R_\ell, \ell \in [N], D_0, D_T)$ is achievable. Let $R^{ab}(D_0, D_T)$ denote the region defined by Eq.(4.1). Since $R(D_0, D_T)$ is a closed set, we have $cl(R^{ab}(D_0, D_T)) \subseteq R(D_0, D_T)$. Define $R_{\text{in}}(D_0, D_T) \coloneqq cl(R^{ab}(D_0, D_T))$. $R_{\text{in}}(D_0, D_T)$ is an inner bound for $R(D_0, D_T)$.

Theorem 5 *For any $0 < D_0 \leq D_T$, the achievable region $R(D_0, D_T)$ is contained in the following region $R_{\text{out}}(D_0, D_T)$:*

$$\begin{aligned} R_{\text{out}}(D_0, D_T) \coloneqq \{ & (R_\ell, \ell \in [N]) | R_\ell \geq 0, \forall \ell \in [N] \ ; \\ & \sum_{\ell=2T+1}^{N} \tilde{R}_\ell \geq \frac{1}{2}\log\frac{1}{D_T} - \frac{1}{2}; \\ & \sum_{\ell=1}^{N} R_\ell + \frac{2T}{s}\sum_{\ell=2T+s+1}^{N} \tilde{R}_\ell \geq \frac{1}{2}\log\frac{1}{D_0} + \frac{T}{s}\log\frac{1}{D_T} - \frac{2T+s}{s}, \quad \forall 1 \leq s \leq N - 2T \Big\}, \end{aligned} \tag{4.2}$$

where $\{\tilde{R}_1, ..., \tilde{R}_N\}$ *is a permutation of* $\{R_1, ..., R_N\}$ *such that* $\tilde{R}_1 \geq ... \geq \tilde{R}_N$.

Theorem 6 *There is a constant gap between* $R_{\text{in}}(D_0, D_T)$ *and* $R_{\text{out}}(D_0, D_T)$. *That is, for any* $\boldsymbol{R} \in R_{\text{out}}(D_0, D_T)$, *we can find* $\boldsymbol{R}' \in R_{\text{in}}(D_0, D_T)$, *such that:*

$$0 \leq R'_\ell - R_\ell \leq 4, \quad \forall 1 \leq \ell \leq N.$$

Theorem 4 is proven in Section 4.2.3. Theorem 5 is proven in Section 4.4. Theorem 6 is proven in Section 4.5.

4.2 Achievable Schemes for General Cases

4.2.1 Three Useful Lemmas

The first two lemmas are about error correction and error detection, respectively. Consider the communication scenario described in Section 4.1 where the source alphabet $\mathcal{X}$ is $\{0, 1\}$.

Lemma 7 *Fix* $R > 0$. *Let* $\{R_\ell\}_{\ell \in [N]}$ *be a set of non-negative real numbers satisfying the following condition: for any set A such that* $A \subset [N]$, $|A| = N - 2T$,

$$\sum_{\ell \in A} R_\ell > R.$$

Then for any sufficiently large n, there exist encoding functions $\tilde{f}_\ell$:

$$\tilde{f}_\ell : \{0, 1\}^{nR} \mapsto \{0, 1\}^{nR_\ell}, \quad \ell \in [N]$$

and a decoding function $\tilde{g}$:

$$\tilde{g} : \prod_{\ell=1}^{N} \{0, 1\}^{nR_\ell} \mapsto \{0, 1\}^{nR},$$

such that the decoding function can always correct up to T altered packets. That is, for any source message $\boldsymbol{Y} \in \{0,1\}^n$, for any set B such that $B \subset [N]$, $|B| \leq T$ and for any arbitrary $\tilde{f}'_B(\boldsymbol{Y})$,

$$\boldsymbol{Y} \equiv \tilde{g}(\tilde{f}'_B(\boldsymbol{Y}), \tilde{f}_{B^c}(\boldsymbol{Y})).$$

We call $\{\tilde{f}_\ell, \ell \in [N], \tilde{g}\}$ an (N,T)- Error Correcting Code for the Source $\{0,1\}^{nR}$.

Proof: This lemma is an immediate result of Gilbert-Varshamov Bound [16] [7] which shows that there exists an (N,T)-Error Correcting Code for $\{0,1\}^{n_c R}$ for some n_c. See Theorem 2 in [7] for the details.

Lemma 8 *Fix $R > 0$. Let $\{R_\ell\}_{\ell \in [N]}$ be a set of non-negative real numbers satisfying the following condition: for any set B, such that $B \subset [N]$, $|B| = N - T$,*

$$\sum_{\ell \in B} R_\ell > R.$$

Then for sufficiently large n, there exist encoding functions $\tilde{f}_\ell$:

$$\tilde{f}_\ell : \{0,1\}^{nR} \mapsto \{0,1\}^{nR_\ell}, \quad \ell \in [N]$$

and a decoding function $\tilde{g}$:

$$\tilde{g} : \prod_{\ell=1}^{N} \{0,1\}^{nR_\ell} \mapsto \{0,1\}^{nR} \cup \{e\},$$

such that the decoding function can always detect up to T altered packets. That is, when no packet is altered, the decoder correctly outputs the original message:

$$\boldsymbol{Y} \equiv \tilde{g}(\tilde{f}_\ell(\boldsymbol{Y}), \ell \in [N]), \quad \textit{for all } \boldsymbol{Y}.$$

Whenever there is one or more altered packets (at most T), the decoder can detect them: for any $\boldsymbol{Y}$, for any set B such that $B \subset [N]$, $|B| \leq T$ and for any arbitrary $\tilde{f}'_B(\boldsymbol{Y})$,

$$g(\tilde{f}'_B(\boldsymbol{Y}), \tilde{f}_{B^c}(\boldsymbol{Y})) = e.$$

We call $\{\tilde{f}_\ell, \ell \in [N], \tilde{g}\}$ an (N,T)-Error Detecting Code for the Source $\{0,1\}^{nR}$.

Proof: We first assume that all R_ℓ are non-negative integers. The result can be easily generalized to non-negative rational R_ℓ and then to non-negative real R_ℓ. Here, we only show that there exists an (N, T)-Error Correcting Code for the Source $\{0,1\}^{n_d R}$ for some n_d. Again, it is not difficult to generalize it to arbitrary sufficiently large n. Without loss of generality, suppose $R_1 \geq R_2 \geq ... \geq R_N$. Then:

$$\sum_{\ell=T+1}^{N} R_\ell > R.$$

Multiplying a sufficiently large integer n_0 on both sides gives:

$$n_0 \left(\sum_{\ell=T+1}^{N} R_\ell \right) \geq \lceil n_0 R \rceil .$$

Let $k = \lceil n_0 R \rceil$ and $n = n_0 \left(\sum_{\ell=1}^{N} R_\ell \right)$. Let n_1 be a positive integer satisfying $2^{n_1} > n$ and let $q = 2^{n_1}$. For any source message $\in \{0,1\}^{n_1 n_0 R}$, we first convert it to a binary string of length $n_1 \cdot \lceil n_0 R \rceil = n_1 k$ (by appending zeros), treat it as a q-ary string of length k and then use a Reed-Solomon code [32] with message length $k = \lceil n_0 R \rceil$, blocklength $n = n_0 \left(\sum_{\ell=1}^{N} R_\ell \right)$ and alphabet size $q = 2^{n_1}$. The encoding function of the Reed-Solomon code can be viewed as a mapping from $(F_q)^k$ to $(F_q)^n$, where F_q denotes the finite field of size q. Since $n = n_0 \left(\sum_{\ell=1}^{N} R_\ell \right)$, we can divide the codeword of the Reed-Solomon code, a n-length vector, into N subvectors of length $n_0 R_\ell$ ($\ell \in [N]$) respectively, and let these N subvectors form the N packets. Any two codewords of a Reed-Solomon code disagree in at least $n - k + 1 \geq n_0 \left(\sum_{\ell=1}^{T} R_\ell \right) + 1$ positions. Since the adversary can only change up to $n_0 \left(\sum_{\ell=1}^{T} R_\ell \right)$ positions, the decoder is always able to detect up to T altered packets. By doing this, we obtain an (N, T)-Error Detecting Code for $\{0,1\}^{n_d R}$ where $n_d = n_0 n_1$.

The third lemma is about successive refinement coding (see [13] [34]).

Definition 8 *We shall say successive refinement from distortion D_1 to distortion D_2 is*

achievable ($D_1 \geq D_2$) *if there exists a sequence of encoding functions*

$$f_{suc_1,n} : \mathfrak{R}^n \mapsto \{0,1\}^{nR_1}$$

and

$$f_{suc_2,n} : \mathfrak{R}^n \mapsto \{0,1\}^{n(R_2-R_1)},$$

and reconstruction functions:

$$g_{suc_1,n} : \{0,1\}^{nR_1} \mapsto \mathfrak{R}^n$$

and

$$g_{suc_2,n} : \{0,1\}^{nR_1} \times \{0,1\}^{n(R_2-R_1)} \mapsto \mathfrak{R}^n,$$

such that for $\hat{X}_1^n = g_{suc_1,n}(f_{suc_1,n}(X^n))$ *and for* $\hat{X}_2^n = g_{suc_2,n}(f_{suc_1,n}(X^n), f_{suc_2,n}(X^n))$,

$$\limsup_{n\to\infty} Ed(X^n, \hat{X}_1^n) \leq D(R_1),$$

and

$$\limsup_{n\to\infty} Ed(X^n, \hat{X}_2^n) \leq D(R_2),$$

where $D(R)$ *is the Distortion-Rate function.*

Lemma 9 *If* X *is a Gaussian* $\sim \mathcal{N}(0, \sigma^2)$*, then successive refinement from distortion* D_1 *to distortion* D_2*, satisfying the following condition:*

$$\lim_{n\to\infty} \left\{ \sup_{x^n} \|\hat{x}_2^n - \hat{x}_1^n\|_2 \right\} \leq \sqrt{D_1 - D_2}, \tag{4.3}$$

where

$$\hat{x}_1^n = g_{suc_1,n}(f_{suc_1,n}(x^n)),$$

$$\hat{x}_2^n = g_{suc_2,n}(f_{suc_1,n}(x^n), f_{suc_2,n}(x^n)),$$

is achievable for ($\sigma^2 \geq D_1 \geq D_2$).

Proof: If X is a Gaussian with distribution $\mathcal{N}(0, \sigma^2)$, then successive refinement from distortion D_1 to distortion D_2 is achievable for $(\sigma^2 \geq D_1 \geq D_2)$ via the following series of randomized encoding and decoding functions [13].

1. Generate $2^{\frac{n}{2}\log\frac{\sigma^2}{D_1}} = (\sigma^2/D_1)^{n/2}$ vector sequences $\hat{x}^n$ drawn randomly and i.i.d. according to $\mathcal{N}(0, \sigma^2 - D_1)$. Label them

$$\hat{x}^n(1), ..., \hat{x}^n(2^{\frac{n}{2}\log\frac{\sigma^2}{D_1}}).$$

2. Let U be a Gaussian $\sim \mathcal{N}(0, D_1 - D_2)$. Let $\epsilon_n \to 0$ be a sequence of non-negative real numbers such that $P(T^{n,\epsilon_n}(U)) \to 1$. Generate $2^{\frac{n}{2}\log\frac{D_1}{D_2}} = (D_1/D_2)^{n/2}$ vector sequences u^n drawn randomly and i.i.d. according to a uniform distribution over the set $T^{n,\epsilon_n}(U)$. Label them

$$u^n(1), ..., u^n(2^{\frac{n}{2}\log\frac{D_1}{D_2}}).$$

   (Here $T^{n,\epsilon_n}(U)$ refers to the strongly typical set for U. See Def. 15.)

3. For any realization of X^n denoted by x^n, let $f_{suc_1,n}(x^n)$ denote the binary expression of the index i which minimizes $\|\hat{x}^n(i) - x^n\|_2^2$:

$$f_{suc_1,n}(x^n) = Bin(i).$$

   Let $f_{suc_2,n}(x^n)$ denote the binary expression of j which minimizes $\|\hat{x}^n(Bin^{-1}(f_{suc_1,n})) + u^n(j) - x^n\|_2^2$:

$$f_{suc_2,n}(x^n) = Bin(j).$$

4. The corresponding reconstruction functions are:

$$g_{suc_1,n}(Bin(i)) = \hat{x}^n(i),$$

$$g_{suc_2,n}(Bin(i), Bin(j)) = \hat{x}^n(i) + u^n(j).$$

The reconstruction asymptotically achieves distortion D_1 and D_2 at rate $R(D_1) = \frac{1}{2}\log\frac{\sigma^2}{D_1}$ and $R(D_2) = \frac{1}{2}\log\frac{\sigma^2}{D_2}$. Moreover, Eq. (4.3) is also satisfied:

$$\lim_{n\to\infty}\sup_{x^n}\|\hat{x}_2^n - \hat{x}_1^n\|_2 \leq \sup_{u^n\in T^{n,\epsilon_n}(U)}\|u^n\|_2 = \sqrt{D_1 - D_2}.$$

The above coding functions are random because the codebook

$$C = \{x^n(1), ..., x^n(2^{\frac{n}{2}\log\frac{\sigma^2}{D_1}}), u^n(1), ..., u^n(2^{\frac{n}{2}\log\frac{D_1}{D_2}})\}$$

is random. Next, for each n, we pick a specific codebook C which minimize

$$Ed(X^n, \hat{X}_1^n(C)) + Ed(X^n, \hat{X}_2^n(C)),$$

where $\hat{X}_1^n(C)$ and $\hat{X}_2^n(C)$ represent the result of $g_{suc_1,n}$ and $g_{suc_2,n}$ under codebook C respectively. It is not difficult to prove that:

$$\limsup_{n\to\infty} Ed(X^n, \hat{X}_1^n(C)) \leq D(R_1),$$
$$\limsup_{n\to\infty} Ed(X^n, \hat{X}_2^n(C)) \leq D(R_2).$$

After assigning a specific codebook C for each n, we now get a new sequence of deterministic coding functions which also achieves successive refinement from distortion D_1 to distortion D_2.

4.2.2 A Simple Scheme

This simple scheme first uses successive refinement coding and then uses error correcting coding for the coarse description part and error detecting coding for the refinement part.

Encoding and Decoding Scheme

1. Successive Refinement:

   For fixed $0 \le D_0 \le D_T$, let $f_{suc_1,n}$, $f_{suc_2,n}$, $g_{suc_1,n}$ and $g_{suc_2,n}$ be a series of deterministic encoding/ decoding functions that achieve successive refinement from D_T to D_0 given in the proof of Lemma 9. For convenience, let $R'_a = \frac{1}{2}\log\frac{1}{D_T}$ and $R'_b = \frac{1}{2}\log\frac{D_T}{D_0}$ in this section.

2. (N, T)-Error Correcting Code for the Source $\{0,1\}^{nR'_a}$:

   From Lemma 7, if for any $A \subset [N]$, $|A| = N - 2T$, we have

   $$\sum_{\ell \in A} R_{a,\ell} > \frac{1}{2}\log\frac{1}{D_T} = R'_a,$$

   then for sufficiently large n, there exists an (N, T)-Error Correcting Code for the Source $\{0,1\}^{nR'_a}$ denoted by $\{\tilde{f}_{a,\ell}, \ell \in [N], \tilde{g}_a\}$. Notice that:

   $$\tilde{f}_{a,\ell} : \{0,1\}^{nR'_a} \mapsto \{0,1\}^{nR_{a,\ell}}, \quad \ell \in [N];$$
   $$\tilde{g}_a : \prod_{\ell=1}^{N} \{0,1\}^{nR_{a,\ell}} \mapsto \{0,1\}^{nR'_a}.$$

3. (N, T)-Error Detecting Code for the Source $\{0,1\}^{nR'_b}$:

   From Lemma 8, if for any $B \subset [N]$, $|B| = N - T$, we have

   $$\sum_{\ell \in B} R_{b,\ell} > \frac{1}{2}\log\frac{D_T}{D_0} = R'_b,$$

   then for sufficiently large n, there exists an (N, T)-Error Detecting Code for the Source $\{0,1\}^{nR'_b}$ denoted by $\{\tilde{f}_{b,\ell}, \ell \in [N], \tilde{g}_b\}$. Notice that:

   $$\tilde{f}_{b,\ell} : \{0,1\}^{nR'_b} \mapsto \{0,1\}^{nR_{b,\ell}}, \quad \ell \in [N];$$
   $$\tilde{g}_b : \prod_{\ell=1}^{N} \{0,1\}^{nR_{b,\ell}} \mapsto \{0,1\}^{nR'_b} \cup \{e\}.$$

For sufficiently large n, the general encoding functions are:

$$f_{\ell,n}(X^n) = (\tilde{f}_{a,\ell} \circ f_{suc_1,n}(X^n), \tilde{f}_{b,\ell} \circ f_{suc_2,n}(X^n)).$$

Suppose the decoder receives $\{(C_{\ell a}, C_{\ell b})\}_{\ell \in [N]}$. The general decoding function is:

$$g_n((C_{\ell a}, C_{\ell b}), \ell \in [N]) = \begin{cases} g_{suc_2,n}(\tilde{g}_a(C_{\ell a}, \ell \in [N]), \tilde{g}_b(C_{\ell b}, \ell \in [N])), \tilde{g}_b(C_{\ell b}, \ell \in [N]) \neq e \\ g_{suc_1,n}(\tilde{g}_a(C_{\ell a}, \ell \in [N])), \quad g_b(C_{\ell b}, \ell \in [N]) = e. \end{cases}$$

The rate of packet ℓ is $R_\ell = R_{a,\ell} + R_{b,\ell}$. Both distortion constraints are satisfied: for any $\epsilon > 0$ and for all sufficiently large n, we have:

$$D_0(f_\ell, \ell \in [N], g) \leq D_0 + \epsilon, \quad D_T(f_\ell, \ell \in [N], g) \leq D_T + \epsilon.$$

The rate region given by this scheme is stated in the following proposition.

Proposition 10 *Given a distortion pair (D_0, D_T), if $\{R_\ell\}_{\ell \in [N]}$ satisfies the following condition: $\exists \{R_{a,\ell}\}_{\ell \in [N]}$ and $\{R_{b,\ell}\}_{\ell \in [N]}$ s.t.*

$$\begin{aligned} & R_\ell = R_{a,\ell} + R_{b,\ell}, \quad \forall \ell \in [N]; \\ & \sum_{\ell \in A} R_{a,\ell} > \frac{1}{2} \log \frac{1}{D_T}, \quad \forall A \subset [N], |A| = N - 2T; \\ & \sum_{\ell \in B} R_{b,\ell} > \frac{1}{2} \log \frac{D_T}{D_0}, \quad \forall B \subset [N], |B| = N - T; \\ & R_{a,\ell} \geq 0, R_{b,\ell} \geq 0, \quad \forall \ell \in [N], \end{aligned} \tag{4.4}$$

then $(R_\ell, \ell \in [N], D_0, D_T)$ is achievable using the simple scheme in Section 4.2.2.

Comparison

This simple scheme does not give you a good inner bound. Suppose we consider the symmetric case where all packets are required to have the same rate. The simple scheme requires:

$$R \geq \frac{1}{2(N-2T)} \log \frac{1}{D_T} + \frac{1}{2(N-T)} \log \frac{D_T}{D_0},$$

while in Theorem 4, it only requires:

$$R \geq \frac{1}{2(N-2T)} \log \frac{1}{D'_T} + \frac{1}{2N} \log \frac{D'_T}{D_0}.$$

We fix D_T and then let D_0 approach 0. It is obvious that this simple scheme is worse:

$$\begin{aligned}
&\lim_{D_0 \to 0} \left\{ \frac{1}{2(N-2T)} \log \frac{1}{D_T} + \frac{1}{2(N-T)} \log \frac{D_T}{D_0} - \frac{1}{2(N-T)} \log \frac{1}{D_0} \right\} \\
=&\frac{T}{2(N-2T)(N-T)} \log \frac{1}{D_T}; \\
&\lim_{D_0 \to 0} \left\{ \frac{1}{2(N-2T)} \log \frac{1}{D'_T} + \frac{1}{2N} \log \frac{D'_T}{D_0} - \frac{1}{2N} \log \frac{1}{D_0} \right\} \\
=&\frac{T}{N(N-2T)} \log \frac{1}{D'_T}.
\end{aligned}$$

However, this simple scheme can be slightly better than the one achieving the inner bound in Theorem 4. For $0 < D_0 < D_T$, we have $D'_T < D_T$. We can check that for $\boldsymbol{R} =$

$$\frac{1}{2(N-2T)} \log \frac{1}{D_T} (\underbrace{1, ..., 1}_{N}) + \frac{1}{2T} \log \frac{D_T}{D_0} (\underbrace{0, ..., 0}_{N-2T}, \underbrace{1, ..., 1}_{2T}),$$

$(\boldsymbol{R}, D_0, D_T)$ is achievable using this simple scheme, but is not achievable using the one in Theorem 4:

$$\sum_{\ell=1}^{N-2T} R_\ell = \frac{1}{2} \log \frac{1}{D_T} < \frac{1}{2} \log \frac{1}{D'_T}.$$

4.2.3 A Modified Scheme – Proof of Theorem 4

This scheme consists of successive refinement coding (see [13] [34]), (N, T)-Error Correcting Code defined in Section 4.2.1 and Splitting.

Encoding and Decoding Scheme

The encoding scheme consists of the following three parts.

1. Successive Refinement:

   For fixed $0 \leq D_0 \leq D_T$, let D'_T satisfy $\sqrt{D'_T} + \sqrt{D'_T - D_0} = \sqrt{D_T}$. Let $f_{suc_1,n}$, $f_{suc_2,n}$, $g_{suc_1,n}$ and $g_{suc_2,n}$ be a series of encoding/ decoding functions that achieve successive refinement from D'_T to D_0 given in Lemma 9. For convenience, let $R_a = \frac{1}{2}\log\frac{1}{D'_T}$ and $R_b = \frac{1}{2}\log\frac{D'_T}{D_0}$ in this subsection.

2. (N, T)-Error Correcting Code for the Source $\{0,1\}^{nR_a}$:

   From Lemma 7, if for any $A \subset [N]$, $|A| = N - 2T$, we have

   $$\sum_{\ell \in A} R_{a,\ell} > \frac{1}{2}\log\frac{1}{D'_T} = R_a,$$

   then for sufficiently large n, there exists an (N, T)- Error Correcting Code for the Source $\{0,1\}^{nR'_a}$ denoted by $\{\tilde{f}_{a,\ell}, \ell \in [N], \tilde{g}_a\}$. Notice that:

   $$\tilde{f}_{a,\ell} : \{0,1\}^{nR_a} \mapsto \{0,1\}^{nR_{a,\ell}} \quad \ell \in [N];$$
   $$\tilde{g}_a : \prod_{\ell=1}^{N} \{0,1\}^{nR_{a,\ell}} \mapsto \{0,1\}^{nR_a}.$$

3. Splitting:

   $f_{suc_2,n}(X^n) \in \{0,1\}^{nR_b}$ is a binary string of length nR_b. For any $\{R_{b,\ell}\}_{\ell \in [N]}$ such that $\sum_{\ell=1}^{N} R_{b,\ell} > R_b$, after embedding $f_{suc_2,n}(X^n)$ into $\{0,1\}^{n\sum_{\ell=1}^{N} R_{b,\ell}}$, it can be divided into N parts, where the ℓth part can be viewed as a substring of length $nR_{b,\ell}$. The concatenation of these N strings is $f_{suc_2,n}(X^n)$. Let $\tilde{f}_{b,\ell} : \{0,1\}^{nR_b} \mapsto \{0,1\}^{nR_{b,\ell}}$ denote the ℓth splitting function and let $\tilde{g}_b : \prod_{\ell=1}^{N}\{0,1\}^{nR_{b,\ell}} \mapsto \{0,1\}^{nR_b}$ be the concatenation function such that for any $\boldsymbol{Y} \in \{0,1\}^{nR_b}$,

   $$\boldsymbol{Y} \equiv \tilde{g}_b(\tilde{f}_{b,1}(\boldsymbol{Y}), \tilde{f}_{b,2}(\boldsymbol{Y})..., \tilde{f}_{b,N}(\boldsymbol{Y})).$$

The general encoding functions are:

$$f_{\ell,n}(X^n) = (\tilde{f}_{a,\ell} \circ f_{suc_1,n}(X^n), \tilde{f}_{b,\ell} \circ f_{suc_2,n}(X^n)).$$

Suppose the decoder receives $\{(C_{\ell a}, C_{\ell b})\}_{\ell\in[N]}$. The general decoding function is:

$$g_n((C_{\ell a},C_{\ell b}), \ell \in [N]) = g_{suc_2,n}(\tilde{g}_a(C_{\ell a}, \ell \in [N]), \tilde{g}_b(C_{\ell b}, \ell \in [N])).$$

The rate of packet ℓ is $R_\ell = R_{a,\ell} + R_{b,\ell}$.

Distortion of the Coding Scheme

We need to prove that for any $\epsilon > 0$, both distortion constraints can be satisfied for sufficiently large blocklength n. Using the property of successive refinement encoding schemes, it is not difficult to see that for any fixed $\epsilon > 0$, for sufficiently large n,

$$D_0(f_{\ell,n}, \ell \in [N], g_n) \le D_0 + \epsilon.$$

Thus, it is sufficient to show that,

$$\lim_{n\to\infty} D_T(f_{\ell,n}, \ell \in [N], g_n) \le D_T.$$

Since we use (N, T)-Error Correcting Code for the D'_T refinement part, for any $A \subset [N]$, $|A| = T$ and any C'_A,

$$\tilde{g}_a(C_\ell, \ell \in [N]) = \tilde{g}_a(C_{A^c}, C'_A).$$

This indicates that $f_{suc_1,n}(X^n)$ can always be correctly recovered by the decoder. Thus,

$$\begin{aligned}
&\limsup_{n\to\infty} D_T(f_{\ell,n}, \ell \in [N], g_n) \\
&= \lim_{n\to\infty} E_{X^n}\left[\max_{A\subset[N]:|A|=T} \max_{C'_A} d(X^n, g_n(C_{A^c}, C'_A))\right] \\
&= \lim_{n\to\infty} E_{X^n}\left[\max_i \frac{1}{n}\|\hat{x}^n(Bin^{-1}(f_{suc_1,n}(X^n))) + u^n(i) - X^n\|_2^2\right] \\
&\le \lim_{n\to\infty} \frac{1}{n} E_{X^n}[\|\hat{x}^n(Bin^{-1}(f_{suc_1,n}(X^n))) - X^n\|_2^2 + \max_i \|u^n(i)\|_2^2 \\
&\quad + 2\max_i \|u^n(i)\|_2 \|\hat{x}^n(Bin^{-1}(f_{suc_1,n}(X^n))) - X^n\|_2] \\
&\overset{(a)}{\le} D'_T + (D'_T - D_0) + 2\sqrt{D'_T - D_0}\cdot\sqrt{D'_T} \\
&= (\sqrt{D'_T} + \sqrt{D'_T - D_0})^2 = D_T.
\end{aligned}$$

Here, (a) follows from

$$\lim_{n\to\infty}\max_i \frac{1}{n}\|u^n(i)\|_2^2 = D_T' - D_0,$$
$$\lim_{n\to\infty}\max_i \frac{1}{\sqrt{n}}\|u^n(i)\|_2 \le \sqrt{D_T' - D_0},$$

and

$$\left(E_{X^n}[\frac{1}{\sqrt{n}}\|\hat{x}^n(Bin^{-1}(f_{suc_1,n}(X^n))) - X^n\|_2]\right)^2$$
$$\le \frac{1}{n}E_{X^n}[\|\hat{x}^n(Bin^{-1}(f_{suc_1,n}(X^n))) - X^n\|_2^2] \le D_T'.$$

Remark: We can make improvement by time-sharing the enhancement layer. As an example, fix (N, T) and consider a symmetric encoding scheme in Section 4.2.3 with distortion constraints (D_0, D_T), where $R_{a,1} = \ldots = R_{a,N}$, $R_{b,1} = \ldots = R_{b,N}$. Consider N source messages $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N$ ($\boldsymbol{x}^\ell \in \mathcal{X}^n$). Use the first two parts of the encoding scheme in Section 4.2.3 to encode each $\boldsymbol{x}^\ell$. Then, instead of splitting each of the N enhancement layers, we include the ℓth enhancement layer in packet ℓ. At least $N - T$ of the N enhancement layers are received correctly. The improved scheme is able to achieve distortion $\frac{N-T}{N} \cdot D_0 + \frac{T}{N} \cdot D_T < D_T$ facing the adversary. However, this improved scheme produces a much more complicated achievable rate region for non-symmetric cases than the one in Theorem 4.

4.3 Achievable Scheme for $(N, T) = (3, 1)$

In this section, we consider the symmetric adversarial multiple description problem for $(N, T) = (3, 1)$, where all three packets are required to have the same rate: $R_1 = R_2 = R_3 = R$. We consider the case where the distortion D_0, D_1 are very small and the fraction

$\frac{D_0}{D_1}$ is also very small. Notice that there is a natural bijection between $\{0,1\}^{nR}$ and $[2^{nR}]$, in this section, we always consider encoding functions

$$f_\ell : \mathcal{X}^n \mapsto [2^{nR_\ell}], \quad \ell \in [N]$$

and decoding function:

$$g : \prod_{\ell=1}^{N} [2^{nR_\ell}] \mapsto \mathcal{X}^n.$$

We show that there is a better scheme which gives a larger inner bound than the scheme in Section 4.2.3.

Proposition 11 *For any fixed $\epsilon_0 > 0$, there exists a code $(f_i^{\epsilon_0}, i \in [3], g^{\epsilon_0})$ with rate $R_1 = R_2 = R_3 = R^{\epsilon_0}$ and blocklength n_{ϵ_0}, such that:*

$$D_0(f_1^{\epsilon_0}, f_2^{\epsilon_0}, f_3^{\epsilon_0}, g^{\epsilon_0}) \le D_0 + \epsilon,$$
$$D_1(f_1^{\epsilon_0}, f_2^{\epsilon_0}, f_3^{\epsilon_0}, g^{\epsilon_0}) \le D_1 + \epsilon,$$

and

$$\lim_{D_1, D_0/D_1 \to 0} \lim_{\epsilon_0 \to 0} \left(R^{\epsilon_0} - \frac{1}{6} \log \frac{1}{D_0} - \frac{1}{3} \log \frac{1}{D_1} \right) = 1 - \frac{5}{6} \log 3. \tag{4.5}$$

The coding scheme consists of two parts. We first quantize the source message, then inspired by the coding scheme in [15], we design a random coding scheme for the quantized source message and then perform derandomization. [15] provides a coding scheme for two encoders, here we generalize the coding scheme for three encoders. The coding scheme also shares similarity to Polytope codes introduced by Kosut, Tong and Tse [26] in the context of network coding with adversarial nodes. Polytope codes have the "partial decodability" property: the decoder either perfectly reconstruct the source message or decode some of the codeword symbols even though it is not possible to decode all of them. Similarly, our codes either produce a finer reconstruction or produce

a coarser reconstruction when there are adversarial errors. In the following subsections, we use capital letters to denote random variables and lower case letters to denote the realization.

4.3.1 A Useful Random Coding Scheme

Quantization

Let (N_1, N_2, N_3) be a Gaussian vector $\sim \mathcal{N}\left(0, \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{bmatrix}\right)$. The value of σ and ρ will be determined in Section 4.3.2. We first perform a uniform quantization on X and N_i $(i = 1, 2, 3)$:

$$X_{Q_\delta} = \begin{cases} \delta \min(\sqrt{2\ln(1/\delta)}, \lfloor X/\delta \rfloor), & X \geq 0 \\ \delta \max(-\sqrt{2\ln(1/\delta)}, \lceil X/\delta \rceil), & X < 0; \end{cases}$$

$$N_{i,Q_\delta} = \begin{cases} \delta \min(\sqrt{2\ln(1/\delta)}, \lfloor N_i/\delta \rfloor), & N_i \geq 0 \\ \delta \max(-\sqrt{2\ln(1/\delta)}, \lceil N_i/\delta \rceil), & N_i < 0. \end{cases}$$

For each source message x^n, we first perform the above uniform quantization:

$$x^n \to x^n_{Q_\delta}.$$

Random Codebook Generation

Let $\{X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta}\}$ be the summation of two quantized Gaussian random variables:

$$\begin{cases} X_{1,Q_\delta} = X_{Q_\delta} + N_{1,Q_\delta} \\ X_{2,Q_\delta} = X_{Q_\delta} + N_{2,Q_\delta} \\ X_{3,Q_\delta} = X_{Q_\delta} + N_{3,Q_\delta}. \end{cases}$$

Fix a blocklength n, for $i = 1, 2, 3$, generate 2^{nR_i} vector sequences x^n_{i,Q_δ} drawn randomly and i.i.d. according to a uniform distribution over the set $T^{(n,\epsilon)}(X_{i,Q_\delta})$. Assign each codeword an index $w \in [2^{nR_i}]$. Let

$$C := \bigcup_{\ell=1}^{3} \{x^n_{\ell,Q_\delta}(1), ..., x^n_{\ell,Q_\delta}(2^{nR_\ell})\}$$

denote the random codebook. The codebook is revealed to all the encoders and the decoder. The the following sections, we use C to denote the random variable of the codebook and use C_0 to denote a specific realization.

Encoder for Quantized Source Message

Given an $x^n_{Q_\delta}$, find, if possible, a triple (i, j, k) such that

$$(x^n_{Q_\delta}, x^n_{1,Q_\delta}(i), x^n_{2,Q_\delta}(j), x^n_{3,Q_\delta}(k))$$

is in the set $T^{(n,\epsilon)}(X_{Q_\delta}, X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta})$. If no such (i, j, k) exists, simply set $(i, j, k) = (0, 0, 0)$. The index i, j and k form the three transmitted packets.

Decoder for Quantized Source Message

Suppose the decoder receives $\hat{i}$, $\hat{j}$ and $\hat{k}$.

1. If the decoder receives $(0, 0, 0)$, then the decoder announces the zero vector as its reconstruction. Else, go to Step 2.

2. The decoder checks if all the three pairs $(x^n_{1,Q_\delta}(\hat{i}), x^n_{2,Q_\delta}(\hat{j}))$, $(x^n_{2,Q_\delta}(\hat{j}), x^n_{3,Q_\delta}(\hat{k}))$, $(x^n_{1,Q_\delta}(\hat{i}), x^n_{3,Q_\delta}(\hat{k}))$ are ϵ-strongly typical. If so, the decoder announces

$$\frac{1}{(2\rho+1)\sigma^2+3}(x^n_{1,Q_\delta}(\hat{i}) + x^n_{2,Q_\delta}(\hat{j}) + x^n_{3,Q_\delta}(\hat{k}))$$

as its reconstruction. Else, go to Step 3.

3. If at least one of the three is not ϵ-strongly typical, arbitrarily pick one pair. (If there are two pairs that are not ϵ-strongly typical, pick either one.) Without loss of generality, suppose we pick $(x^n_{1,Q_\delta}(\hat{i}), x^n_{2,Q_\delta}(\hat{j}))$. The decoder announces $\frac{1}{1+\sigma^2} x^n_{3,Q_\delta}(\hat{k})$ as its reconstruction.

Expected Error Probability

The above code is random in that the codebook C is random. The distribution of the codebook P_C is completely determined by coefficients

$$\{\sigma, \rho, R_1, R_2, R_3, \delta, \epsilon, n\}.$$

Any fixed $\sigma, \rho, R_1, R_2, R_3, \delta, \epsilon, n$ along with a specific codebook C_0 give a deterministic coding scheme.

For any arbitrary C_0, an error (denoted by $E_0(C_0)$) will occur if one or more of the following events occurs:

1. $E_1(C_0) : x^n_{Q_\delta} \notin T^{(n,\epsilon)}(X_{Q_\delta})$;

2. $E_2(C_0) : \forall (i, j, k) \in [2^{nR_1}] \times [2^{nR_2}] \times [2^{nR_3}]$, we have

$$(x^n_{Q_\delta}, x^n_{1,Q_\delta}(i), x^n_{2,Q_\delta}(j), x^n_{3,Q_\delta}(k)) \notin T^{(n,\epsilon)}(X_{Q_\delta}, X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta}).$$

The probability of encoding error is:

$$P(E_0(C_0)) = P(E_1(C_0) \cup E_2(C_0)) = P(E_1(C_0)) + P(E_2(C_0) \cap E_1(C_0)^c).$$

We are interested in $E_C[P(E_0(C))] = \sum_{C_0} P_C(C_0)P(E_0(C_0))$.

Proposition 12 *Fix $\sigma, \rho, R_1, R_2, R_3, \delta, \epsilon$. If*

$$\begin{aligned}
R_\ell &> I(X_{Q_\delta}; X_{\ell,Q_\delta}), \quad \forall 1 \le \ell \le 3; \\
R_{\ell_1} + R_{\ell_2} &> H(X_{\ell_1,Q_\delta}) + H(X_{\ell_2,Q_\delta}) - H(X_{\ell_1,Q_\delta}, X_{\ell_2,Q_\delta}|X_{Q_\delta}), \quad \forall 1 \le \ell_1 < \ell_2 \le 3; \\
R_1 + R_2 + R_3 &> \sum_{l=1}^{3} H(X_{l,Q_\delta}) - H(X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta}|X_{Q_\delta}),
\end{aligned}$$

then for any $\epsilon_0 > 0$, there exists n'_{ϵ_0}, such that for any $n \ge n'_{\epsilon_0}$,

$$E_C[P(E_0(C))] < n'_{\epsilon_0}.$$

Proof: See Appendix.

We next prove the following proposition.

Proposition 13 *Fix $\sigma, \rho, R_1, R_2, R_3, \delta, \epsilon$. If $R_1 = R_2 = R_3 = R$ and*

$$R > \frac{1}{6} \log \frac{(1+\sigma^2)^3}{\sigma^6(\rho-1)^2(1+2\rho)}, \tag{4.6}$$

then

$$R > \frac{1}{3}\left(3H(X_{1,Q_\delta}) - H(X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta}|X_{Q_\delta})\right) \tag{4.7}$$

holds for any sufficiently small δ. Furthermore, fix any δ such that Inq. (4.7) holds, for any $\epsilon_0 > 0$, there exists n'_{ϵ_0}, such that for any $n \ge n'_{\epsilon_0}$,

$$E_C[P(E_0(C))] < n'_{\epsilon_0}.$$

Proof: Using Theorem 8.4.1 in [10, p. 249], we have

$$\begin{aligned}\lim_{\delta\to 0}\{3H(X_{1,Q_\delta}) - H(X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta}|X_{Q_\delta})\} &= 3h(X_1) - h(X_1, X_2, X_3|X)\\ &= \frac{1}{2}\log\frac{(1+\sigma^2)^3}{\sigma^6(\rho-1)^2(1+2\rho)}.\end{aligned}$$

Thus, the first part of the proposition is correct.

Using the symmetry property of $X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta}$, all the inequalities in Proposition 12 are satisfied if Inq. (4.7) holds. Apply Proposition 12 and it completes the proof.

Distortion Analysis

The above random coding scheme is determined by coefficients σ, ρ, R, δ, ϵ and n. Given a code (f_1, f_2, f_3, g), for any source message X^n, let $\hat{X}^{1,n}$ denote the reconstruction when no packet is altered:

$$\hat{X}^{1,n} := g(C_\ell, \ell \in [3]).$$

Let $\hat{X}^{2,n}$ denote the reconstruction under any one of the most powerful adversarial attacks. We have

$$d(X^n, \hat{X}^{2,n}) \equiv \max_{\ell\in[3]}\max_{C'_{\{\ell\}}} d(X^n, g(C_{\{\ell\}^c}, C'_{\{\ell\}})).$$

Fix $\sigma, \rho, R_1, R_2, R_3, \delta, \epsilon$ and n, then $\hat{X}^{t,n}$ is a random variable depending only on source message X^n and the codebook C. For any arbitrary codebook C_0, let (f_1, f_2, f_3, g) denote the corresponding coding scheme. Then

$$\begin{aligned}D^0(f_1, f_2, f_3, g) &= E_{X^n}[d(X^n, \hat{X}^{1,n}) \mid C = C_0],\\ D^1(f_1, f_2, f_3, g) &= E_{X^n}[d(X^n, \hat{X}^{2,n}) \mid C = C_0].\end{aligned}$$

Recall that we use E_0 to denote the event that there exists an encoding error. For any

arbitrary C_0, we have

$$
\begin{aligned}
&E_{X^n}[d(X^n, \hat{X}^{t,n}) \mid C = C_0] \\
\leq &E_{X^n}[d(X^n, X^n_{Q_\delta})] + 2E_C E_{X^n}[|\langle X^n - X^n_{Q_\delta}, X^n_{Q_\delta} - \hat{X}^{t,n}\rangle| \mid C = C_0] + E_{X^n}[d(X^n_{Q_\delta}, \hat{X}^{t,n}) \mid C = C_0] \\
\leq &E_{X^n}[d(X^n, X^n_{Q_\delta})] + 2E_{X^n}[|\langle X^n - X^n_{Q_\delta}, X^n_{Q_\delta} - \hat{X}^{t,n}\rangle| \mid C = C_0] \\
&+ E_{X^n}[d(X^n_{Q_\delta}, \hat{X}^{t,n}) \mid E_0^c(C), C = C_0] + P(E_0(C_0))E_{X^n}[d(X^n_{Q_\delta}, \hat{X}^{t,n}) \mid E_0(C), C = C_0].
\end{aligned} \tag{4.8}
$$

The following proposition is useful.

Proposition 14 *Fix $\sigma, \rho, R_1, R_2, R_3$. For any $\epsilon_0 > 0$, there exists δ'_{ϵ_0}, ϵ'_{ϵ_0} and n'_{ϵ_0} such that for any $\delta \leq \delta'_{\epsilon_0}$, $\epsilon \leq \epsilon'_{\epsilon_0}$, $n \geq n'_{\epsilon_0}$ and for any arbitrary codebook C_0,*

$$
E_{X^n}[d(X^n_{Q_\delta}, \hat{X}^{1,n}) \mid E_0^c(C), C = C_0] \leq \frac{(2\rho+1)\sigma^2}{(2\rho+1)\sigma^2 + 3} + \epsilon_0; \tag{4.9}
$$

$$
E_{X^n}[d(X^n_{Q_\delta}, \hat{X}^{2,n}) \mid E_0^c(C), C = C_0] \leq \max\left\{\frac{\sigma^2}{1+\sigma^2}, \left(\sqrt{\frac{A}{A+3}} + \frac{\sqrt{8(1-\rho)\sigma^2}}{A+3}\right)^2\right\} + \epsilon_0, \tag{4.10}
$$

where $A = (2\rho+1)\sigma^2$.

Proof: The argument for $\hat{X}^{1,n}$ is straightforward. We next prove the argument for $\hat{X}^{2,n}$. For each x^n which is successfully encoded, the reconstruction under one of the most powerful adversarial attack, $\hat{x}^{2,n}$, is generated in either one of the following two cases.

Case 1: One of the packets is altered and the decoder detects an error.

In this case, according to the decoding function, the decoder will always use an unaltered packet to perform decoding. For sufficiently small δ, ϵ and sufficiently large n, For any arbitrary codebook C_0,

$$
d(x^n_{Q_\delta}, \hat{x}^{2,n}) < \frac{\sigma^2}{1+\sigma^2} + \epsilon_0. \tag{4.11}
$$

Case 2: One of the packets is altered but the decoder does not detect it.

Without loss of generality, we assume that the third packet is altered by the adversary. Suppose the quantized source message is $x^n_{Q_\delta}$. Suppose that $(x^n_{Q_\delta}, x^n_{1,Q_\delta}(i), x^n_{2,Q_\delta}(j), x^n_{3,Q_\delta}(k))$ is ϵ-strongly typical and i, j, k are transmitted. Instead of receiving k, the decoder receives $\tilde{k}$. The error induced in the third packet is:

$$e^n = x^n_{3,Q_\delta}(\tilde{k}) - x^n_{3,Q_\delta}(k) = (x^n_{3,Q_\delta}(\tilde{k}) - x^n_{1,Q_\delta}(i)) + (x^n_{1,Q_\delta}(i) - x^n_{3,Q_\delta}(k)).$$

Since the decoder does not detect an error, we know that both $(x^n_{3,Q_\delta}(\tilde{k}), x^n_{1,Q_\delta}(i))$ and $(x^n_{3,Q_\delta}(k), x^n_{1,Q_\delta}(i))$ are ϵ-strongly typical. Therefore, for any $\epsilon_0' > 0$, the following statement holds. For sufficiently small δ, ϵ and sufficiently large n, for any arbitrary codebook C_0,

$$\begin{aligned}
\frac{1}{n}E[||x^n_{3,Q_\delta}(\tilde{k}) - x^n_{1,Q_\delta}(i)||_2^2] &< 2(1-\rho)\sigma^2 + \epsilon_0', \\
\frac{1}{n}E[||x^n_{3,Q_\delta}(k) - x^n_{1,Q_\delta}(i)||_2^2] &< 2(1-\rho)\sigma^2 + \epsilon_0'.
\end{aligned}$$

Thus, the error is upper bounded by the following:

$$\begin{aligned}
\frac{1}{n}E[||e^n||_2^2] \leq& \frac{2}{n}\left(E[||x^n_{3,Q_\delta}(\tilde{k}) - x^n_{1,Q_\delta}(i)||_2^2] + E[||x^n_{3,Q_\delta}(k) - x^n_{1,Q_\delta}(i)||_2^2]\right) \\
\leq& 8(1-\rho)\sigma^2 + 2\epsilon_0'.
\end{aligned} \tag{4.12}$$

Let $A = (2\rho+1)\sigma^2$. Choosing a sufficiently small ϵ_0', we have

$$d(x^n_{Q_\delta}, \hat{x}^n) < \left(\sqrt{\frac{A}{A+3}} + \frac{\sqrt{\frac{1}{n}E[||e^n||^2]}}{A+3}\right)^2 + \epsilon_0' < \left(\sqrt{\frac{A}{A+3}} + \frac{\sqrt{8(1-\rho)\sigma^2}}{A+3}\right)^2 + \epsilon_0. \tag{4.13}$$

Combine Inq. (4.11) and Inq. (4.13) and it completes the proof.

4.3.2 Proof of Proposition 11

In this section, we show that for any fixed $\epsilon_0 > 0$, there exists a code $(f_i^{\epsilon_0}, i \in [3], g^{\epsilon_0})$ with rate $R_1 = R_2 = R_3 = R^{\epsilon_0}$, blocklength n_{ϵ_0}, coefficients $\sigma_{\epsilon_0}, \rho_{\epsilon_0}, \delta_{\epsilon_0}, \epsilon_{\epsilon_0}$ and codebook C_{ϵ_0} such that:

$$D_0(f_1^{\epsilon_0}, f_2^{\epsilon_0}, f_3^{\epsilon_0}, g^{\epsilon_0}) = E_{X^n}[d(X^n, \hat{X}^{1,n}) \mid C = C_{\epsilon_0}] \leq D_0 + \epsilon,$$
$$D_1(f_1^{\epsilon_0}, f_2^{\epsilon_0}, f_3^{\epsilon_0}, g^{\epsilon_0}) = E_{X^n}[d(X^n, \hat{X}^{2,n}) \mid C = C_{\epsilon_0}] \leq D_1 + \epsilon,$$

and

$$\lim_{D_1, D_0/D_1 \to 0} \lim_{\epsilon_0 \to 0} \left(R^{\epsilon_0} - \frac{1}{6} \log \frac{1}{D_0} - \frac{1}{3} \log \frac{1}{D_1} \right) = 1 - \frac{5}{6} \log 3.$$

To satisfy both the distortion constraints and the rate constraint, we next find appropriate $\sigma_{\epsilon_0}, \rho_{\epsilon_0}, R^{\epsilon_0}, \delta_{\epsilon_0}, \epsilon_{\epsilon_0}, n_{\epsilon_0}$ and a specific codebook C_{ϵ_0} in order. First of all, we choose σ_{ϵ_0} and ρ_{ϵ_0} such that the following inequalities hold for sufficiently small D_1 and D_0/D_1.

$$\frac{(2\rho_{\epsilon_0} + 1)\sigma_{\epsilon_0}^2}{(2\rho_{\epsilon_0} + 1)\sigma_{\epsilon_0}^2 + 3} \leq D_0 + \epsilon_0/2, \tag{4.14}$$
$$\frac{\sigma_{\epsilon_0}^2}{1 + \sigma_{\epsilon_0}^2} \leq D_1 + \epsilon_0/2, \tag{4.15}$$
$$\left(\sqrt{\frac{A_{\epsilon_0}}{A_{\epsilon_0} + 3}} + \frac{\sqrt{8(1 - \rho_{\epsilon_0})\sigma_{\epsilon_0}^2}}{A_{\epsilon_0} + 3} \right)^2 \leq D_1 + \epsilon_0/2. \tag{4.16}$$

Let $A_{\epsilon_0} = (2\rho_{\epsilon_0} + 1)\sigma_{\epsilon_0}^2 = \frac{3(D_0 + \epsilon_0/2)}{1 - (D_0 + \epsilon_0/2)}$ to satisfy Inq. (4.14). Then $A_{\epsilon_0} + 3 = \frac{3}{1 - (D_0 + \epsilon_0/2)}$. To satisfy Inq. (4.16), we let:

$$\sigma_{\epsilon_0}^2 = \frac{9\left(\sqrt{D_1 + \epsilon_0/2} - \sqrt{D_0 + \epsilon_0/2} \right)^2}{8(1 - \rho_{\epsilon_0})(1 - (D_0 + \epsilon_0/2))^2}.$$

We choose $(\sigma_{\epsilon_0}^2, \rho_{\epsilon_0})$ to be the solution to the following equation set:

$$\begin{cases} -0.5 \leq \rho_{\epsilon_0} < 1 \\ \sigma_{\epsilon_0}^2 = \frac{9\left(\sqrt{D_1+\epsilon_0/2}-\sqrt{D_0+\epsilon_0/2}\right)^2}{8(1-\rho_{\epsilon_0})(1-(D_0+\epsilon_0/2))^2} \\ (2\rho_{\epsilon_0}+1)\sigma_{\epsilon_0}^2 = \frac{3(D_0+\epsilon_0/2)}{1-(D_0+\epsilon_0/2)}. \end{cases}$$

This is equivalent with

$$\begin{cases} -0.5 \leq \rho_{\epsilon_0} < 1 \\ \frac{2\rho_{\epsilon_0}+1}{1-\rho_{\epsilon_0}} = \frac{3(D_0+\epsilon_0/2)}{1-(D_0+\epsilon_0/2)} \cdot \frac{8(1-(D_0+\epsilon_0/2))^2}{9\left(\sqrt{D_1+\epsilon_0/2}-\sqrt{D_0+\epsilon_0/2}\right)^2} \\ \sigma_{\epsilon_0}^2 = \frac{9\left(\sqrt{D_1+\epsilon_0/2}-\sqrt{D_0+\epsilon_0/2}\right)^2}{8(1-\rho_{\epsilon_0})(1-(D_0+\epsilon_0/2))^2}. \end{cases}$$

Notice that the function $f(x) = \frac{2x+1}{1-x}$ is an continuous increasing function on $[-0.5, 1)$, and the image of the function is $[0, +\infty)$. Thus, for sufficiently small ϵ_0 and D_0/D_1, there always exists a solution. Furthermore,

$$\lim_{D_0, D_1/D_0 \to 0} \lim_{\epsilon_0 \to 0} \frac{3(D_0+\epsilon_0/2)}{1-(D_0+\epsilon_0/2)} \cdot \frac{8(1-(D_0+\epsilon_0/2))^2}{9\left(\sqrt{D_1+\epsilon_0/2}-\sqrt{D_0+\epsilon_0/2}\right)^2} = 0,$$

indicating that

$$\lim_{D_0/D_1 \to 0} \lim_{\epsilon_0 \to 0} \rho_{\epsilon_0} = -0.5,$$

$$\lim_{D_1, D_0/D_1 \to 0} \lim_{\epsilon_0 \to 0} \sigma_{\epsilon_0} = 0.$$

Inq. (4.14) and Inq. (4.16) are satisfied by construction. Inq. (4.15) can be written as:

$$\sigma_{\epsilon_0}^2 \leq \frac{D_1+\epsilon_0/2}{1-D_1-\epsilon_0/2}.$$

It is not difficult to check that for sufficiently small D_1 and D_0/D_1:

$$\lim_{\epsilon_0 \to 0} \frac{9\left(\sqrt{D_1+\epsilon_0/2}-\sqrt{D_0+\epsilon_0/2}\right)^2}{8(1-\rho_{\epsilon_0})(1-(D_0+\epsilon_0/2))^2} < \frac{D_1}{1-D_1}.$$

This means that the solution to the above equation set also satisfies Inq. (4.14) when ϵ_0 is small.

Secondly, we choose:

$$R^{\epsilon_0} = \frac{1}{6}\log\frac{(1+\sigma_{\epsilon_0}^2)^3}{\sigma_{\epsilon_0}^6(\rho_{\epsilon_0}-1)^2(1+2\rho_{\epsilon_0})} + \epsilon_0.$$

Thirdly, we choose sufficiently small δ_{ϵ_0}, ϵ_{ϵ_0} and sufficiently large $\tilde{n}_{\epsilon_0}$ such that for any $n \geq \tilde{n}_{\epsilon_0}$,

1. Quantization distortion is sufficiently small. For any arbitrary C_0, for $t = 1, 2$

$$E_{X^n}[d(X^n, X^n_{Q_{\delta_{\epsilon_0}}})] + 2E_{X^n}[|\langle X^n - X^n_{Q_{\delta_{\epsilon_0}}}, X^n_{Q_{\delta_{\epsilon_0}}} - \hat{X}^{t,n}\rangle| \mid C = C_0] < \epsilon_0/8. \tag{4.17}$$

2. For any arbitrary C_0, the following two inequalities holds:

$$E_{X^n}[d(X^n_{Q_{\delta_{\epsilon_0}}}, \hat{X}^{1,n}) \mid C = C_0, E_0^c(C)] \leq \frac{(2\rho_{\epsilon_0}+1)\sigma_{\epsilon_0}^2}{(2\rho_{\epsilon_0}+1)\sigma_{\epsilon_0}^2 + 3} + \epsilon_0/8; \tag{4.18}$$

$$E_{X^n}[d(X^n_{Q_{\delta_{\epsilon_0}}}, \hat{X}^{2,n}) \mid C = C_0, E_0^c(C)] \leq \epsilon_0/8 + \max\left\{\frac{\sigma_{\epsilon_0}^2}{1+\sigma_{\epsilon_0}^2}, \left(\sqrt{\frac{A_{\epsilon_0}}{A_{\epsilon_0}+3}} + \frac{\sqrt{8(1-\rho_{\epsilon_0})\sigma_{\epsilon_0}^2}}{A_{\epsilon_0}+3}\right)^2\right\} \tag{4.19}$$

The existence of such δ_{ϵ_0}, ϵ_{ϵ_0} and $\tilde{n}_{\epsilon_0}$ can be easily proven using Proposition 14.

Finally for each $n > \tilde{n}_{\epsilon_0}$, we pick a specific codebook C which minimize $P(E_0(C))$. From Proposition 13, we can find n_{ϵ_0} and the corresponding codebook C_{ϵ_0} such that for $t = 1$ and 2,

$$P(E_0(C_0))E_{X^{\epsilon_0}}[d(X^{n_{\epsilon_0}}_{Q_{\delta_{\epsilon_0}}}, X^{t,n_{\epsilon_0}}) \mid E_0(C), C = C_0] < \epsilon_0/8. \tag{4.20}$$

Combining from Inq. (4.17) to Inq. (4.20), we can easily check that the two rate distor-

tion constraints are satisfied. It remains to verify that Eq. (4.5) holds.

$$\begin{aligned}\lim_{\epsilon_0\to 0} R^{\epsilon_0} &= \lim_{\epsilon_0\to 0}\frac{1}{6}\log\frac{(1+\sigma_{\epsilon_0}^2)^3}{\sigma_{\epsilon_0}^6(\rho_{\epsilon_0}-1)^2(1+2\rho_{\epsilon_0})}+\epsilon_0\\ &= \lim_{\epsilon_0\to 0}\frac{1}{6}\log\frac{(1+\sigma_{\epsilon_0}^2)^3}{\sigma_{\epsilon_0}^6(\rho_{\epsilon_0}-1)^2(1+2\rho_{\epsilon_0})}\\ &= \frac{1}{6}\left(\lim_{\epsilon_0\to 0}\log\frac{(1+\sigma_{\epsilon_0}^2)^3}{(\rho_{\epsilon_0}-1)^2}+\lim_{\epsilon_0\to 0}\log\frac{1}{(1+2\rho_{\epsilon_0})\sigma_{\epsilon_0}^2}+\lim_{\epsilon_0\to 0}\log\frac{1}{\sigma_{\epsilon_0}^4}\right). \end{aligned}\tag{4.21}$$

Recall that $\sigma_{\epsilon_0}^2\to 0, \rho_{\epsilon_0}\to -0.5$ as D_1, $\frac{D_0}{D_1}$ and ϵ_0 approach 0. Moreover,

$$\begin{aligned}&\lim_{\epsilon_0\to 0}(2\rho_{\epsilon_0}+1)\sigma_{\epsilon_0}^2 = \frac{3D_0}{1-D_0},\\ &\lim_{D_1,D_0/D_1\to 0}\lim_{\epsilon_0\to 0}\frac{D_1}{\sigma_{\epsilon_0}^2} = \lim_{D_1,D_0/D_1\to 0}\lim_{\epsilon_0\to 0}\frac{8(1-\rho_{\epsilon_0})(1-(D_0+\epsilon/2))^2D_1}{9\left(\sqrt{D_1+\epsilon_0/2}-\sqrt{D_0+\epsilon_0/2}\right)^2} = \frac{4}{3}.\end{aligned}$$

Thus,

$$\begin{aligned}\lim_{D_1,D_0/D_1\to 0}\lim_{\epsilon_0\to 0}\log\frac{(1+\sigma_{\epsilon_0}^2)^3}{(\rho_{\epsilon_0}-1)^2} &= \log(4/9);\\ \lim_{D_1,D_0/D_1\to 0}\lim_{\epsilon_0\to 0}\left(\log\frac{1}{(1+2\rho_{\epsilon_0})\sigma_{\epsilon_0}^2}-\log\frac{1}{D_0}\right) &= \lim_{D_1,D_0/D_1\to 0}\left(\log\frac{1-D_0}{3D_0}-\log\frac{1}{D_0}\right) = -\log 3;\\ \lim_{D_1,D_0/D_1\to 0}\lim_{\epsilon_0\to 0}\left(\log\frac{1}{\sigma_{\epsilon_0}^4}-2\log\frac{1}{D_1}\right) &= 2\log\frac{4}{3}.\end{aligned}$$

Combining Inq. (4.8), Inq. (4.14)- (4.16) and Inq. (4.17)- (4.21), we can conclude that

$$\begin{aligned}&\lim_{D_1,D_0/D_1\to 0}\lim_{\epsilon_0\to 0}\left(R^{\epsilon_0}-\frac{1}{6}\log\frac{1}{D_0}-\frac{1}{3}\log\frac{1}{D_1}\right)\\ =&\frac{1}{6}(\log(4/9)-\log 3+2\log(4/3)) = 1-\frac{5}{6}\log 3.\end{aligned}$$

4.3.3 Comparison

Consider the scheme in Section 4.2.3. Let $R^*(D_1, D_0)$ denote the corresponding rate. Then:

$$\begin{aligned}
&\lim_{D_1, D_0/D_1 \to 0} \left(R^*(D_1, D_0) - \frac{1}{6} \log \frac{1}{D_0} - \frac{1}{3} \log \frac{1}{D_1} \right) \\
&= \lim_{D_1, D_0/D_1 \to 0} \left(\frac{1}{3} \log \frac{1}{D_1'} - \frac{1}{3} \log \frac{1}{D_1} \right) \\
&= \lim_{D_1 \to 0} \left(\frac{1}{3} \log \frac{4}{D_1} - \frac{1}{3} \log \frac{1}{D_1} \right) = \frac{2}{3}.
\end{aligned}$$

Since $\frac{2}{3} > 1 - \frac{5}{6} \log 3$, the scheme shows a better inner bound than the scheme in 4.2.3 when both D_1 and D_0/D_1 are small. However, this scheme is not better in general since it only works well for sufficiently small D_1 and D_0/D_1.

Remark: If we apply the time-sharing scheme to the enhancement layer as described in Section 4.2.3 and let $\tilde{R}^*(D_1, D_0)$ denote the corresponding rate. Then we still have:

$$\lim_{D_1, D_0/D_1 \to 0} \left(\tilde{R}^*(D_1, D_0) - \frac{1}{6} \log \frac{1}{D_0} - \frac{1}{3} \log \frac{1}{D_1} \right) = \frac{2}{3} - \frac{1}{3} \log 3 > 1 - \frac{5}{6} \log 3.$$

4.4 Outer Bound – Proof of Theorem 5

In order to prove Theorem 5, it is sufficient to prove that for any $(R_\ell, \ell \in [N]) \in R(D_0, D_T)$ for which $\{R_\ell\}_{\ell \in [N]}$ is non-increasing, we have:

$$\sum_{\ell=2T+1}^{N} R_\ell \geq \frac{1}{2} \log \frac{1}{D_T} - \frac{1}{2}, \tag{4.22}$$

and $\forall 1 \leq s \leq N - 2T$,

$$\sum_{\ell=1}^{N} R_\ell + \frac{2T}{s} \sum_{\ell=2T+s+1}^{N} R_\ell \geq \frac{1}{2} \log \frac{1}{D_0} + \frac{T}{s} \log \frac{1}{D_T} - \frac{2T+s}{s}. \tag{4.23}$$

We assume that an empty sum is zero. This means, when $s = N-2T$, the left hand side of Inq. (4.23) is $\sum_{\ell=1}^{N} R_\ell$. Without loss of generality, we assume $\{R_\ell\}_{\ell\in[N]}$ is non-increasing for the following subsections in Section 4.4.

4.4.1 Important Lemmas

Let $(f_\ell, \ell \in [N], g)$ be a code for the problem such that:

$$D_0(f_\ell, \ell \in [N], g) \leq D_0;$$

$$D_T(f_\ell, \ell \in [N], g) \leq D_T.$$

Let C_ℓ denote the ℓth packet. Then the following two lemmas hold.

Lemma 15 *The following inequality holds:*

$$\frac{1}{n} h(X^n | C_\ell, \ell \in [N]) \leq \frac{1}{2} \log(2\pi e D_0).$$

Proof: This lemma is an immediate result of the Rate-Distortion Theory and the Rate-Distortion function for Gaussian source [10, Theorem 10.2.1, 10.3.2].

Lemma 16 *Let $W \sim \mathcal{N}(0, \sigma^2)$ be a Gaussian random variable independent of X and $\{C_\ell\}_{\ell\in[N]}$. For any $A \subset [N]$, $|A| = N - 2T$, the following inequality holds:*

$$\frac{1}{n} h(X^n + W^n | C_A) \leq \frac{1}{2} \log(2\pi e(2D_T + 2\sigma^2)). \tag{4.24}$$

Proof: See appendix for the proof.

Lemma 17 *(Entropy Power Inequality) If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent random n-vectors with densities, then*

$$2^{\frac{2}{n}h(\boldsymbol{X}+\boldsymbol{Y})} \geq 2^{\frac{2}{n}h(\boldsymbol{X})} + 2^{\frac{2}{n}h(\boldsymbol{Y})}.$$

Lemma 18 *X^n, $\tilde{X}^n$ and Y^n are random n-vectors. $\tilde{X}^n$ and Y^n are independent Gaussian vectors. X^n and Y^n are independent. If $h(X^n + Y^n) - h(X^n) = h(\tilde{X}^n + Y^n) - h(\tilde{X}^n)$, then $h(X^n) \geq h(\tilde{X}^n)$.*

Proof: Using Lemma 17, we have:

$$\begin{aligned}
&2^{\frac{2}{n}h(\tilde{X}^n)} + 2^{\frac{2}{n}h(Y^n)} = 2^{\frac{2}{n}h(\tilde{X}^n+Y^n)} \\
&= 2^{\frac{2}{n}h(X^n+Y^n)} \cdot 2^{\frac{2}{n}(h(\tilde{X}^n)-h(X^n))} \\
&\geq (2^{\frac{2}{n}h(X^n)} + 2^{\frac{2}{n}h(Y^n)}) \cdot 2^{\frac{2}{n}(h(\tilde{X}^n)-h(X^n))} \\
&= 2^{\frac{2}{n}h(\tilde{X}^n)} + 2^{\frac{2}{n}h(Y^n)} \cdot 2^{\frac{2}{n}(h(\tilde{X}^n)-h(X^n))}.
\end{aligned}$$

Hence, $1 \geq 2^{\frac{2}{n}(h(\tilde{X}^n)-h(X^n))}$, indicating that $h(X^n) \geq h(\tilde{X}^n)$.

The following two lemmas are the keys to prove Inq. (4.23).

Lemma 19 *Let X be a Gaussian random variable with variance σ_x^2. Let Z be an arbitrary discrete random variable. Let $Y_a = X + W_a$ where W_a is a Gaussian random variable independent of both X and Z, with variance σ_a^2. Let $D_a = \sigma_a^2$. If Z, X^n and Y_a^n form a Markov chain in that order ($Z \to X^n \to Y_a^n$) and*

$$\frac{1}{n}h(X^n|Z) \leq \frac{1}{2}\log(2\pi e\tilde{D}),$$

then we have:

$$I(Z;X^n) - I(Z;Y_a^n) \geq \frac{n}{2}\log\frac{\sigma_x^2(\tilde{D}+D_a)}{\tilde{D}(\sigma_x^2+D_a)}.$$

Proof: See appendix for the proof.

Lemma 20 *Let X denote the source, $X \sim \mathcal{N}(0,1)$. Let Z be an arbitrary discrete random variable. Let $Y_b = X + W_b$, $Y_a = X + W_a + W_b$ where W_a and W_b are mutually independent Gaussian random variables and are both independent of X and Z, with variance*

σ_a^2 and σ_b^2 respectively. Let $D_a = \sigma_a^2 + \sigma_b^2$, $D_b = \sigma_b^2$. If $\frac{1}{n}h(X^n + W_b^n|Z) \leq \frac{1}{2}\log(2\pi e\tilde{D})$, then we have:

$$I(Z; Y_b^n) - I(Z; Y_a^n) \geq \frac{n}{2}\log\frac{(1+D_b)(\tilde{D}+D_a-D_b)}{\tilde{D}(1+D_a)}.$$

Proof: Treat Y_b as X in Lemma 19. Y_b is a Gaussian random variable with variance $1 + D_b$. Moreover, Z, Y_b^n and Y_a^n form a Markov chain in that order by construction. Lemma 20 can be viewed as a corollary of Lemma 19.

4.4.2 Proof of Inq. (4.22)

For any $\boldsymbol{R} \in R(D_0, D_T)$ and any $\epsilon > 0$, there exists a code $(f_\ell, \ell \in [N], g)$ such that:

$$D_0(f_\ell, \ell \in [N], g) \leq D_0 + \epsilon,$$

$$D_T(f_\ell, \ell \in [N], g) \leq D_T + \epsilon,$$

and

$$n(R_\ell + \epsilon) \geq H(C_\ell),$$

where n is the blocklength of the code. Using Lemma 16 and setting $\sigma = 0$, for any $A \subset [N]$, $|A| = N - 2T$, we have

$$\frac{1}{n}h(X^n|C_A) \leq \frac{1}{2}\log(4\pi e(D_T + \epsilon)). \tag{4.25}$$

This indicates that

$$\begin{aligned}
&\sum_{\ell\in A}(R_\ell + \epsilon) \geq \frac{1}{n}\sum_{\ell\in A}H(C_\ell) \\
&\geq \frac{1}{n}H(C_A) \geq \frac{1}{n}I(X^n; C_A) \\
&= \frac{1}{n}h(X^n) - \frac{1}{n}h(X^n|C_A) = \frac{1}{2}\log(2\pi e) - \frac{1}{n}h(X^n|C_A) \\
&\geq \frac{1}{2}\log(2\pi e) - \frac{1}{2}\log(4\pi e(D_T + \epsilon)) = \frac{1}{2}\log\frac{1}{D_T+\epsilon} - \frac{1}{2}.
\end{aligned}$$

Let $A = \{2T + 1, ..., N\}$ and $\epsilon \to 0$ and it completes the proof.

4.4.3 Proof of Inq. (4.23)

The proof is similar to the proof of Theorem 1 in [44]. We first quote a few definitions and results from [44]. Let $\boldsymbol{v}$ be a vector in $\{0,1\}^N$ and denote the ith component of $\boldsymbol{v}$ by $\boldsymbol{v}_i$. Let $G_{\boldsymbol{v}} \coloneqq \{i : \boldsymbol{v}_i = 1\}$. Let $|\boldsymbol{v}|$ be the Hamming weight of $\boldsymbol{v}$. For $1 \leq \alpha \leq N$, define:

$$\Omega_N^\alpha = \{\boldsymbol{v} \in \{0,1\}^N, \quad |\boldsymbol{v}| = \alpha\}.$$

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vectors in $\Re^N$. Define $\boldsymbol{u} \geq \boldsymbol{v}$ iff $u_i \geq v_i$ for $1 \leq i \leq N$. For any $\boldsymbol{A} = (A_1, ..., A_N) \geq 0$, a mapping $c_\alpha : \Omega_N^\alpha \to \Re^+$ satisfying the following properties:

$$c_\alpha(\boldsymbol{v}) \geq 0, \quad \forall \boldsymbol{v} \in \Omega_N^\alpha,$$

and

$$\sum_{\boldsymbol{v} \in \Omega_N^\alpha} c_\alpha(\boldsymbol{v}) \boldsymbol{v} \leq \boldsymbol{A}$$

is called an *α-resolution* for $\boldsymbol{A}$ and it will be abbreviated as $\{c_\alpha(\boldsymbol{v})\}$ or simply as c_α when there is no ambiguity.

Define a function $f_\alpha : (\Re^+)^N \to \Re^+$ for $1 \leq \alpha \leq N$ by

$$f_\alpha(\boldsymbol{A}) = \max \sum_{\boldsymbol{v} \in \Omega_N^\alpha} c_\alpha(\boldsymbol{v}),$$

where the maximum is taken over all α-resolutions of $\boldsymbol{A}$. If $\{c_\alpha(\boldsymbol{v})\}$ achieves $f_\alpha(\boldsymbol{A})$, then it is called an *optimal α-resolution* for $\boldsymbol{A}$ or simply *α-optimal* for $\boldsymbol{A}$.

Definition 9 *For $2 \leq \alpha \leq N$, let c_α and $c_{\alpha-1}$ be an α-optimal resolution and $(\alpha-1)$-optimal resolution for $\boldsymbol{A}$ respectively. Then $c_{\alpha-1}$ covers c_α, denoted by $c_{\alpha-1} \succ c_\alpha$, if*

$$\sum_{\boldsymbol{u} \in \Omega_N^{\alpha-1}} c_{\alpha-1}(\boldsymbol{u}) H(C_\ell, \ell \in G_{\boldsymbol{u}}) \geq \sum_{\boldsymbol{v} \in \Omega_N^\alpha} c_\alpha(\boldsymbol{v}) H(C_\ell, \ell \in G_{\boldsymbol{v}})$$

for any N jointly distributed discrete random variables C_1,...,C_N.

The following lemma is Theorem 3 in [44] and is crucial for us to derive the outer bound.

Lemma 21 *For any $\boldsymbol{A} \geq 0$, there exist c_α for $1 \leq \alpha \leq N$, where c_α is α-optimal for $\boldsymbol{A}$ and $c_1 \succ c_2 \succ ... \succ c_N$.*

We consider the following distribution:

$$Y_\alpha = X + \sum_{i=\alpha}^{N-1} W_i, \quad \alpha \in [N],$$

where $W_i \sim \mathcal{N}(0, \sigma_i^2)$ are mutually independent and independent of X and $\{C_\ell\}_{\ell \in [N]}$. Let Z_α denote $\sum_{i=\alpha}^{N-1} W_i$. Let d_α denote the variance of Z_α. Here, $Y_N = X$ and $d_N = 0$. For any $s \in [N - 2T]$, let

$$\boldsymbol{A}_s = (\underbrace{1, ..., 1}_{2T+s}, \underbrace{1 + 2T/s, ..., 1 + 2T/s}_{N-2T-s}).$$

Let $c_1, c_2, ..., c_N$ be a set of α-resolution for $\boldsymbol{A}_s$ as defined in Lemma 21 (c_α is α-optimal and $c_1 \succ ... \succ c_N$). For any $\boldsymbol{R} \in R(D_0, D_T)$ and any $\epsilon > 0$, there exists a code $(f_\ell, \ell \in [N], g)$ such that:

$$D_0(f_\ell, \ell \in [N], g) \leq D_0 + \epsilon,$$

$$D_T(f_\ell, \ell \in [N], g) \leq D_T + \epsilon,$$

and

$$n(R_\ell + \epsilon) > H(C_\ell), \quad \forall \ell \in [N],$$

where n is the blocklength of the code. Let $\boldsymbol{\epsilon} = (\epsilon, ..., \epsilon) \in (\Re^+)^N$, then:

$$\begin{aligned}
n\boldsymbol{A}_s(\boldsymbol{R}+\boldsymbol{\epsilon})^T &\geq n\left(\sum_{\boldsymbol{v}\in\Omega_N^1} c_1(\boldsymbol{v})\boldsymbol{v}\right)(\boldsymbol{R}+\boldsymbol{\epsilon})^T \\
&= n\sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=1} c_1(\boldsymbol{v})\left(\sum_{\ell\in G_{\boldsymbol{v}}}(R_\ell+\epsilon)\right) \\
&\geq \sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=1} c_1(\boldsymbol{v})H(C_\ell, \ell\in G_{\boldsymbol{v}}) \\
&= \sum_{\alpha=1}^{N-1}\left(\sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})H(C_\ell,\ell\in G_{\boldsymbol{v}}) - \sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=\alpha+1} c_{\alpha+1}(\boldsymbol{v})H(C_\ell,\ell\in G_{\boldsymbol{v}})\right) \\
&\quad + \sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=N} c_N(\boldsymbol{v})H(C_\ell,\ell\in G_{\boldsymbol{v}}),
\end{aligned}$$

where the last equation is obtained by adding and subtracting the same terms. Using the conditional version of the covering property of the given sequence $c_1, .., c_N$, we have:

$$\sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})H(C_\ell,\ell\in G_{\boldsymbol{v}}|Y_\alpha^n) - \sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=\alpha+1} c_{\alpha+1}(\boldsymbol{v})H(C_\ell,\ell\in G_{\boldsymbol{v}}|Y_\alpha^n) \geq 0.$$

Thus, we can get Inq. (4.26).

$$\begin{aligned}
&n\boldsymbol{A}_s(\boldsymbol{R}+\boldsymbol{\epsilon})^T \\
&\geq \sum_{\alpha=1}^{N-1}\left(\sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})H(C_\ell,\ell\in G_{\boldsymbol{v}}) - \sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=\alpha+1} c_{\alpha+1}(\boldsymbol{v})H(C_\ell,\ell\in G_{\boldsymbol{v}})\right) \\
&\quad + \sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=N} c_N(\boldsymbol{v})H(C_\ell,\ell\in G_{\boldsymbol{v}}) \\
&\quad - \sum_{\alpha=1}^{N-1}\left(\sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})H(C_\ell,\ell\in G_{\boldsymbol{v}}|Y_\alpha^n) - \sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=\alpha+1} c_{\alpha+1}(\boldsymbol{v})H(C_\ell,\ell\in G_{\boldsymbol{v}}|Y_\alpha^n)\right) \\
&= \sum_{\alpha=2}^{N-1}\sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})(I(C_\ell,\ell\in G_{\boldsymbol{v}};Y_\alpha^n) - I(C_\ell,\ell\in G_{\boldsymbol{v}};Y_{\alpha-1}^n)) \\
&\quad + \sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=1} c_1(\boldsymbol{v})I(C_\ell,\ell\in G_{\boldsymbol{v}};Y_1^n) + \sum_{G_{\boldsymbol{v}}:|\boldsymbol{v}|=N} c_N(\boldsymbol{v})(H(C_\ell,\ell\in G_{\boldsymbol{v}}) - I(C_\ell,\ell\in G_{\boldsymbol{v}};Y_{N-1}^n)) \\
&\geq \sum_{\alpha=1}^{N} f_\alpha(\boldsymbol{A}_s)\min_{\boldsymbol{v}\in\Omega_N^\alpha}\{I(C_\ell,\ell\in G_{\boldsymbol{v}};Y_\alpha^n) - I(C_\ell,\ell\in G_{\boldsymbol{v}};Y_{\alpha-1}^n)\}. \qquad (4.26)
\end{aligned}$$

For Inq. (4.26), we let Y_0^n denote a constant random variable. We let

$$d_\alpha = \begin{cases} \infty, & 0 \le \alpha \le N - 2T - 1 \\ D_T, & N - 2T \le \alpha \le N - 1 \\ 0, & \alpha = N. \end{cases}$$

Then for $\alpha \neq N - 2T$ or N, for any $\boldsymbol{v} \in \Omega_N^\alpha$,

$$I(C_\ell, \ell \in G_{\boldsymbol{v}}; Y_\alpha^n) - I(C_\ell, \ell \in G_{\boldsymbol{v}}; Y_{\alpha-1}^n) = 0.$$

For $\alpha = N$, $\Omega_N^N = \{(\underbrace{1, ..., 1}_{N})\}$. Lemma 15 gives

$$\frac{1}{n} h(X^n | C_\ell, \ell \in [N]) \le \frac{1}{2} \log(2\pi e(D_0 + \epsilon)).$$

Substituting into Lemma 19,

$$\begin{aligned} I(C_\ell, \ell \in [N]; X^n) - I(C_\ell, \ell \in [N]; Y_{N-1}^n) \ge & \frac{n}{2} \log \frac{(D_0 + D_T + \epsilon)}{(D_0 + \epsilon)(1 + D_T)} \\ > & \frac{n}{2} \log \frac{D_T}{(D_0 + \epsilon)(1 + D_T)}. \end{aligned}$$

For $\alpha = N - 2T$, from Lemma 16 we know that for any $\boldsymbol{v} \in \Omega_N^{N-2T}$,

$$\frac{1}{n} h(Y_{N-2T}^n | C_\ell, \ell \in G_{\boldsymbol{v}}) \le \frac{1}{2} \log(2\pi e(2D_T + 2\epsilon + 2d_{N-2T})) = \frac{1}{2} \log(2\pi e \cdot (4D_T + 2\epsilon)).$$

Substituting into Lemma 20,

$$I(C_\ell, \ell \in G_{\boldsymbol{v}}; Y_\alpha^n) - I(C_\ell, \ell \in G_{\boldsymbol{v}}; Y_{\alpha-1}^n) \ge \frac{n}{2} \log \frac{1 + D_T}{4D_T + 2\epsilon}.$$

For $\alpha = N - 2T$, it is not difficult to verify that for $\boldsymbol{v} \in \Omega_N^{N-2T}$,

$$\tilde{c}_{N-2T}(\boldsymbol{v}) = \begin{cases} \frac{(s+2T)/s}{\binom{s+2T}{s}}, & \boldsymbol{v}_{N-2T-s+1} = ... = \boldsymbol{v}_N = 1; \\ 0, & \text{otherwise.} \end{cases}$$

is an $N - 2T$-resolution of $\boldsymbol{A}_s$. By definition,

$$f_{N-2T}(\boldsymbol{A}_s) \geq \sum_{\boldsymbol{v} \in \Omega_N^{N-2T}} \tilde{c}_{N-2T}(\boldsymbol{v}) = \frac{s+2T}{s}.$$

Moreover, it is not difficult to compute that

$$f_N(\boldsymbol{A}_s) = 1.$$

Substituting these into Inq. (4.26), we have:

$$\begin{aligned}
&\boldsymbol{A}_s(\boldsymbol{R}+\boldsymbol{\epsilon})^T \\
&\geq \frac{s+2T}{2s} \log \frac{1+D_T}{4D_T+2\epsilon} + \frac{1}{2} \log \frac{D_T}{1+D_T} + \frac{1}{2} \log \frac{1}{D_0+\epsilon} \\
&= \frac{1}{2} \log \frac{1}{D_0+\epsilon} + \frac{T}{s} \log \frac{1+D_T}{4D_T+2\epsilon} + \frac{1}{2} \log \frac{D_T}{4D_T+2\epsilon} \\
&> \frac{1}{2} \log \frac{1}{D_0+\epsilon} + \frac{T}{s} \log \frac{1}{4D_T+2\epsilon} + \frac{1}{2} \log \frac{D_T}{4D_T+2\epsilon}.
\end{aligned}$$

Let $\epsilon \to 0$ and it completes the proof:

$$\boldsymbol{A}_s \boldsymbol{R}^T \geq \frac{1}{2} \log \frac{1}{D_0} + \frac{T}{s} \log \frac{1}{D_T} - \frac{2T+s}{s}.$$

4.5 Relationship Between Outer and Inner Bounds

We now have both an inner and an outer bound for $R(D_0, D_T)$:

$$R_{\text{in}}(D_0, D_T) \subseteq R(D_0, D_T) \subseteq R_{\text{out}}(D_0, D_T).$$

The problem is whether the gap between $R_{\text{in}}(D_0, D_T)$ and $R_{\text{out}}(D_0, D_T)$ is small. Theorem 6 states that the gap is bounded above by 4. We next prove this theorem.

4.5.1 Another Expression for $R_{\text{in}}(D_0, D_T)$

$R_{\text{in}}(D_0, D_T)$ is characterized by $\{R_{a,\ell}\}$ and $\{R_{b,\ell}\}$. The next lemma gives an inner bound of $R(D_0, D_T)$ with a different expression.

Definition 10 *Let $W, w_1, ..., w_K$ be $K+1$ non-negative real numbers such that $\sum_{i=1}^{K} w_i \geq W$. Call $\{a_i\}_{i\in[K]}$ eligible for $(W, w_i, i \in [K])$ if it satisfies:*

$$\begin{cases} 0 \leq a_i \leq w_i, & \forall i \in [K] \\ \sum_{i=1}^{K} a_i = W \end{cases}.$$

Define

$$\mathcal{G}(K, W, w_i, i \in [K]) \coloneqq \min_{\text{eligible}\{a_i\}_{i\in[K]}} \max \{a_i\}_{i\in[K]}.$$

A solution $\{a_i\}_{i\in[K]}$ achieving the min max *is called an optimal solution for $(W, w_i, i \in [K])$.*

Lemma 22 *Let $R'_{\text{in}}(D_0, D_T)$ denote the set of $(R_1, ..., R_N)$ satisfying:*

$$\begin{aligned} &\sum_{\ell=2T+1}^{N} \tilde{R}_\ell \geq \frac{1}{2}\log\frac{1}{D'_T}; \\ &\sum_{\ell=1}^{N} \tilde{R}_\ell \geq \frac{1}{2}\log\frac{1}{D_0} + \\ &\qquad 2T \cdot \mathcal{G}(N-2T, \frac{1}{2}\log\frac{1}{D'_T}, \tilde{R}_{2T+1}, ..., \tilde{R}_N); \\ &R_\ell \geq 0, \quad \forall \ell \in [N], \end{aligned}$$

where $\{\tilde{R}_1, ..., \tilde{R}_N\}$ is a permutation of $\{R_1, ..., R_N\}$ such that $\tilde{R}_1 \geq ... \geq \tilde{R}_N$. Then

$$R'_{\text{in}}(D_0, D_T) = R_{\text{in}}(D_0, D_T) = cl(R^{ab}(D_0, D_T)).$$

Proof: See Appendix for the proof.

We next solve $\mathcal{G}(N-2T, \frac{1}{2}\log\frac{1}{D'_T}, \tilde{R}_{2T+1}, ..., \tilde{R}_N)$. $\forall (R_\ell, \ell \in [N]) \in R'_{\text{in}}(D_0, D_T)$, we have:

$$\sum_{\ell=2T+1}^{N} \tilde{R}_\ell \geq \frac{1}{2}\log\frac{1}{D'_T}.$$

Thus the solution set of the following

$$\begin{cases} \sum_{i=1}^{N-2T} a_i = \frac{1}{2}\log\frac{1}{D'_T} \\ a_i \leq \tilde{R}_{2T+i}, \quad \forall i \in [N-2T] \end{cases}$$

is non-empty. It is not difficult to check that an optimal solution $\{a_i\}_{i\in[K]}$ for $(\frac{1}{2}\log\frac{1}{D'_T}, \tilde{R}_{2T+1}, ..., \tilde{R}_N)$ has the following property: there exists an integer s ($1 \leq s \leq N-2T$), such that:

$$a_i = \begin{cases} a, & 1 \leq i \leq s; \\ \tilde{R}_{2T+i}, & s+1 \leq i \leq N-2T. \end{cases}$$

Then,

$$\begin{aligned} &\mathcal{G}(N-2T, \frac{1}{2}\log\frac{1}{D'_T}, \tilde{R}_{2T+1}, ..., \tilde{R}_N) \\ =&\begin{cases} \frac{1}{s}\left(\frac{1}{2}\log\frac{1}{D'_T} - \sum_{\ell=2T+1+s}^{N}\tilde{R}_\ell\right), \\ \qquad\qquad 1 \leq s \leq N-2T-1; \\ \frac{1}{2(N-2T)}\log\frac{1}{D'_T}, \quad s = N-2T. \end{cases} \end{aligned} \tag{4.27}$$

4.5.2 Proof of Theorem 6

Define

$$R^{\text{Order}}_{\text{out}}(D_0, D_T) := \{(R_\ell, \ell \in [N]) | (R_\ell, \ell \in [N]) \in R_{\text{out}}(D_0, D_T), R_1 \geq R_2 \geq ... \geq R_N\};$$

$$R^{\text{Order}}_{\text{in}}(D_0, D_T) := \{(R_\ell, \ell \in [N]) | (R_\ell, \ell \in [N]) \in R_{\text{in}}(D_0, D_T), R_1 \geq R_2 \geq ... \geq R_N\}.$$

Then $R_{\text{out}}^{\text{Order}}(D_0, D_T)$ denotes the set of $(R_1, ..., R_N)$ satisfying:

$$\begin{aligned}
&R_1 \geq R_2 \cdots \geq R_N \geq 0;\\
&\sum_{\ell=2T+1}^{N} R_\ell \geq \frac{1}{2}\log\frac{1}{D_T} - \frac{1}{2};\\
&\sum_{\ell=1}^{N} R_\ell + \frac{2T}{s}\sum_{\ell=2T+s+1}^{N} R_\ell \geq \frac{1}{2}\log\frac{1}{D_0} + \frac{T}{s}\log\frac{1}{D_T} - \frac{2T+s}{s}, \quad \forall 1 \leq s \leq N-2T.
\end{aligned}$$

By Lemma 22, $R_{\text{in}}^{\text{Order}}(D_0, D_T)$ denotes the set of $(R_1, ..., R_N)$ satisfying:

$$\begin{aligned}
&R_1 \geq R_2 \cdots \geq R_N \geq 0;\\
&\sum_{\ell=2T+1}^{N} R_\ell \geq \frac{1}{2}\log\frac{1}{D'_T};\\
&\sum_{\ell=1}^{N} R_\ell \geq \frac{1}{2}\log\frac{1}{D_0} + 2T\cdot\mathcal{G}(N-2T, \frac{1}{2}\log\frac{1}{D'_T}, R_{2T+1}, ..., R_N),
\end{aligned}$$

where D'_T satisfies $\sqrt{D'_T} + \sqrt{D'_T - D_0} = \sqrt{D_T}$. It is sufficient to prove that $R_{\text{out}}^{\text{Order}}(D_0, D_T) \subseteq R_{\text{in}}^{\text{Order}}(D_0, D_T) + 4$. We will show that

$$\forall \boldsymbol{R} = (R_1, ..., R_N) \in R_{\text{out}}^{\text{Order}}(D_0, D_T),$$

we can find $\boldsymbol{R}' = (R'_1, ..., R'_N) \in R_{\text{in}}^{\text{Order}}(D_0, D_T)$ s.t. $\boldsymbol{R}' \geq \boldsymbol{R}$ and

$$R'_\ell - R_\ell \leq 4, \quad \forall \ell \in [N].$$

Step 1: If $\sum_{\ell=2T+1}^{N} R_\ell < \frac{1}{2}\log\frac{1}{D'_T}$, let

$$\lambda_1 = \frac{1}{N-2T}\left(\frac{1}{2}\log\frac{1}{D'_T} - \sum_{\ell=2T+1}^{N} R_\ell\right)$$

and let $\boldsymbol{R}'' = (R''_1, ..., R''_N)$, where

$$R''_\ell = \begin{cases} R_\ell + \lambda_1, & 2T+1 \leq \ell \leq N;\\ \max\{R_\ell, R_{2T+1} + \lambda_1\}, & 1 \leq \ell \leq 2T. \end{cases}$$

Else, let $\boldsymbol{R}'' = \boldsymbol{R}$.

Step 2: If $\boldsymbol{R}''$ is in $R_{\text{in}}^{\text{Order}}(D_0, D_T)$, let $\boldsymbol{R}' = \boldsymbol{R}''$. Else, suppose that for some $s_0 \in [N - 2T]$,

$$\mathcal{G}(N-2T, \frac{1}{2}\log\frac{1}{D_T'}, R_{2T+1}'', ..., R_N'') = \frac{1}{s_0}\left(\frac{1}{2}\log\frac{1}{D_T'} - \sum_{\ell=2T+1+s_0}^{N} R_\ell''\right).$$

Let λ_2 satisfy:

$$\sum_{\ell=1}^{2T}(R_\ell'' + \lambda_2) + \sum_{\ell=2T+1}^{2T+s_0} R_\ell'' + \frac{s_0+2T}{s_0}\sum_{\ell=2T+s_0+1}^{N} R_\ell'' = \frac{1}{2}\log\frac{1}{D_0} + \frac{T}{s_0}\log\frac{1}{D_T'}. \tag{4.28}$$

Then let $\boldsymbol{R}' = (R_1', ..., R_N')$ where

$$R_\ell' = \begin{cases} R_\ell'' + \lambda_2, & 1 \le \ell \le 2T; \\ R_\ell'', & 2T+1 \le \ell \le N. \end{cases}$$

Then $\boldsymbol{R}'$ satisfies: $\sum_{\ell=1}^{N} R_\ell' = \frac{1}{2}\log\frac{1}{D_0} + 2T \cdot \mathcal{G}(N-2T, \frac{1}{2}\log\frac{1}{D_T'}, R_{2T+1}', ..., R_N')$. We obtain $\boldsymbol{R}' \in R_{\text{in}}^{\text{order}}(D_0, D_T)$.

We next bound λ_1 and λ_2. Since $4D_T' - 2D_0 = 2D_T' + 2(D_T' - D_0) \ge (\sqrt{D_T'} + \sqrt{D_T' - D_0})^2 = D_T$, we have:

$$D_T' \ge \frac{1}{4}(D_T + 2D_0).$$

Thus,

$$\begin{aligned}
\lambda_1 &= \frac{1}{N-2T}\left(\frac{1}{2}\log\frac{1}{D_T'} - \sum_{\ell=2T+1}^{N} R_\ell\right) \\
&\le \frac{1}{N-2T}\left(\frac{1}{2}\log\frac{1}{D_T'} - \frac{1}{2}\log\frac{1}{D_T} + \frac{1}{2}\right) \\
&\le \frac{1}{N-2T}\left(\frac{1}{2}\log\frac{4}{D_T} - \frac{1}{2}\log\frac{1}{D_T} + \frac{1}{2}\right) \\
&= \frac{3}{2(N-2T)} \le \frac{3}{2}.
\end{aligned}$$

When $\boldsymbol{R}'' \in R_{\text{in}}^{\text{order}}$, $\lambda_2 = 0$. Otherwise, since $\boldsymbol{R} \in R_{\text{out}}^{\text{Order}}(D_0, D_T)$, we have

$$\sum_{\ell=1}^{2T+s_0} R_\ell + \frac{s_0+2T}{s_0}\sum_{\ell=2T+s_0+1}^{N} R_\ell \ge \frac{1}{2}\log\frac{1}{D_0} + \frac{T}{s_0}\log\frac{1}{D_T} - \frac{2T+s_0}{s_0}.$$

Since $\boldsymbol{R}'' \geq \boldsymbol{R}$ by construction, the above inequality gives:

$$\sum_{\ell=1}^{2T+s_0} R''_\ell + \frac{s_0+2T}{s_0} \sum_{\ell=2T+s_0+1}^{N} R''_\ell \geq \frac{1}{2}\log\frac{1}{D_0} + \frac{T}{s_0}\log\frac{1}{D_T} - \frac{2T+s_0}{s_0}. \tag{4.29}$$

Combining Eq. (4.28) and Inq. (4.29), we have,

$$\begin{aligned}
\lambda_2 &= \frac{1}{2T}\left(\frac{1}{2}\log\frac{1}{D_0} + \frac{T}{s_0}\log\frac{1}{D'_T} - \sum_{\ell=1}^{2T+s_0} R''_\ell - \frac{s_0+2T}{s_0}\sum_{\ell=2T+s_0+1}^{N} R''_\ell\right) \\
&\leq \frac{1}{2T}\left(\frac{T}{s_0}\log\frac{4}{D_T+2D_0} - \frac{T}{s_0}\log\frac{1}{D_T} + \frac{2T+s_0}{s_0}\right) \\
&\leq \frac{4T+s_0}{2s_0T} \leq \frac{5}{2}.
\end{aligned}$$

Thus, for any $\ell \in [N]$,

$$0 \leq R'_\ell - R_\ell \leq \lambda_1 + \lambda_2 \leq 4.$$

CHAPTER 5

P2P VPEC

5.1 Main Results for P2P VPEC Problem

5.1.1 Achievability

We use $X \overset{d}{=} Y$ to denote that X and Y have the same probability distribution. For any arbitrary random variables X, Y and Z, we use the notation $X \to Y \to Z$ to denote that X, Y and Z form a Markov chain in that order.

Theorem 7 *If there exists a triple of variables* $(U, Y, X) \in \mathcal{U} \times \hat{\mathcal{X}} \times \mathcal{X}$ *($\mathcal{U}$ is finite) with joint probability distribution* P_{UYX} *satisfying the following conditions:*

1. $U \to Y \to X$;
2. $P_X = \mathcal{P}$;
3. $E[d(X, Y)] \leq D_0$;
4. $R \geq I(X; Y)$;
5. *For any random variable* $Z \in \mathcal{X}$ *satisfying:*

$$Z \overset{d}{=} X;$$

$$I(U; X) \geq I(U; Z);$$

$$I(X; Y) \geq I(U; Z) + I(Z; Y|U),$$

we have $E[d(Z, Y)] \leq D_1$

then (R, D_0, D_1) *is achievable.*

Theorem 8 *If there exists a pair of random variables $(Y, X) \in \hat{\mathcal{X}} \times \mathcal{X}$ with joint probability distribution P_{YX} satisfying the following four conditions:*

1. $P_X = \mathcal{P}$;
2. $E[d(X, Y)] \leq D_0$;
3. $R \geq I(X; Y)$;
4. *For any random variable $Z \in \mathcal{X}$ satisfying:*

$$Z \stackrel{d}{=} X;$$

$$I(X; Y) \geq I(Z; Y),$$

we have $E[d(Z, Y)] \leq D_1$,

then (R, D_0, D_1) is achievable.

Remark: Theorem 8 can be viewed as a special case of Theorem 7 with $U \equiv 1$.

5.1.2 Optimality

We adopt the notations in [22] with slight modifications. We assume that all the random variables in this section have finite range. Let $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, and $Z \in \mathcal{Z}$ be jointly distributed random variables. Denote by $\mathcal{P}$, $\mathcal{Q}$ and $\mathcal{R}$ their respective distributions, and set

$$V(a|b) := P(X = a|Y = b); V'(a|b) := P(Y = a|X = b);$$

$$W(c|b) := P(Z = c|Y = b); W'(c|b) := P(Y = c|Z = b).$$

Let $\{(X_i, Y_i, Z_i)\}_{i=1}^{\infty}$ be i.i.d. random variables such that the distribution of (X_i, Y_i, Z_i) coincides with that of (X, Y, Z). Let $\tilde{X}^n$, $\tilde{Y}^n$ and $\tilde{Z}^n$ be random variables taking values in $\mathcal{X}^n$, $\mathcal{Y}^n$ and $\mathcal{Z}^n$, respectively, and let $U \in \mathcal{U}$ be an arbitrary discrete random variable.

Definition 11 *We write* $(U, \tilde{Y}^n, \tilde{X}^n, \tilde{Z}^n) \in \mathscr{P}^n_{V,W}$ *if* $(U, \tilde{Y}^n, \tilde{X}^n, \tilde{Z}^n)$ *is a quadruple of random variables such that:*

$$U \to \tilde{Y}^n \to (\tilde{X}^n, \tilde{Z}^n);$$

$$P_{\tilde{X}^n|\tilde{Y}^n} = V^n;$$

$$P_{\tilde{Z}^n|\tilde{Y}^n} = W^n.$$

We also write $(\tilde{Y}^n, \tilde{X}^n, \tilde{Z}^n) \in \mathscr{P}^n_{V,W}$ *for a triple of random variables* $(\tilde{Y}^n, \tilde{X}^n, \tilde{Z}^n)$ *such that*

$$P_{\tilde{X}^n|\tilde{Y}^n} = V^n;$$

$$P_{\tilde{Z}^n|\tilde{Y}^n} = W^n.$$

We omit the superscript n *if* $n = 1$. *We may omit the subscript* V, W *when there is no ambiguity.*

Definition 12 *We write* $(U, \tilde{Y}^n, \tilde{X}^n, \tilde{Z}^n) \in \mathscr{P}^n_{V,W}(Q)$ *if* $(U, \tilde{Y}^n, \tilde{X}^n, \tilde{Z}^n) \in \mathscr{P}^n_{V,W}$ *and* $P_{\tilde{Y}^n} = Q^n$. *We also write* $(\tilde{Y}^n, \tilde{X}^n, \tilde{Z}^n) \in \mathscr{P}^n_{V,W}(Q)$ *if* $(\tilde{Y}^n, \tilde{X}^n, \tilde{Z}^n) \in \mathscr{P}^n_{V,W}$ *and* $P_{\tilde{Y}^n} = Q^n$. *Again, we omit the superscript* n *if* $n = 1$. *We may omit the subscript* V, W *when there is no ambiguity.*

For our communication scenario, we always let $\mathcal{X} = \mathcal{Z}$, $\mathcal{Y} = \hat{\mathcal{X}}$. Fix the distribution for X (denoted by $\mathcal{P}$). Recall that we always use V to denote the stochastic matrix describing the conditional conditional distribution of X given Y and use V' to denote the stochastic matrix describing the conditional conditional distribution of Y given X. W

and W' are defined analogously. For convenience, we write

$$d(V,\mathcal{Q}) = d(\mathcal{P},V') = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathcal{Q}(y)V(x|y)d(x,y)$$
$$d(W,\mathcal{Q}) = d(\mathcal{R},W') = \sum_{z\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathcal{Q}(y)W(z|y)d(z,y).$$

We also let $I(\mathcal{P},V') = I(V,\mathcal{Q})$ denote the mutual information between the random variables (X,Y) on $\mathcal{X}\times\mathcal{Y}$ with probability mass function (pmf) $P_{XY}(x,y) = \mathcal{P}(x)V'(y|x) = \mathcal{Q}(y)V(x|y)$. $I(\mathcal{R},W')$ and $I(W,\mathcal{Q})$ are defined analogously. Moreover, we use $\mathcal{Q}\cdot V$ to denote a pmf on $\mathcal{X}$ defined by

$$\mathcal{Q}\cdot V(x) = \sum_{y\in\mathcal{Y}} \mathcal{Q}(y)V(x|y).$$

Definition 13 *For any distributions $\mathcal{P}$, $\mathcal{Q}$ and distortions D_0, D_1, define*

$$\mathcal{S}(\mathcal{P},\mathcal{Q},D_0) = \{V : \mathcal{Q}\cdot V = \mathcal{P}; d(V,\mathcal{Q}) \le D_0\}, \tag{5.1}$$

and

$$\begin{aligned}\mathcal{S}(\mathcal{P},\mathcal{Q},D_0,D_1) = \{V : V \in \mathcal{S}(\mathcal{P},\mathcal{Q},D_0);\\ \forall W,\ s.t.\ 1)\mathcal{Q}\cdot W = \mathcal{P},\quad 2)I(U;Z) \le I(U;X),\\ \text{for all } (U,Y,X,Z) \in \mathscr{P}_{V,W}(\mathcal{Q}), |\mathcal{U}| \le |\mathcal{Y}| + 2,\\ \text{we have } d(W,\mathcal{Q}) \le D_1\}.\end{aligned} \tag{5.2}$$

where $|\cdot|$ denotes the cardinality of a set and $F_{U|Y}$ ranges over all stochastic matrices from $\mathcal{Y}$ to $\mathcal{U}$.

Definition 14 *For any distributions $\mathcal{P}$, $\mathcal{Q}$ and distortions D_0, D_1, define*

$$R(\mathcal{P},\mathcal{Q},D_0,D_1) = \inf_{V\in\mathcal{S}(\mathcal{P},\mathcal{Q},D_0,D_1)} I(V,\mathcal{Q}). \tag{5.3}$$

When $\mathscr{S}(\mathcal{P}, \mathcal{Q}, D_0, D_1) = \emptyset$ *, set* $R(\mathcal{P}, \mathcal{Q}, D_0, D_1) = +\infty$*. Let*

$$\mathscr{G} = \{(\mathcal{P}, D_0, D_1, R) : \exists \mathcal{Q}, V \in \mathscr{S}(\mathcal{P}, \mathcal{Q}, D_0, D_1), s.t.\ R \geq I(V, \mathcal{Q})\},$$

and let $\bar{\mathscr{G}} = cl(\mathscr{G})$ *denote the closure of* $\mathscr{G}$*. Define*

$$R(\mathcal{P}, D_0, D_1) = \min_{R:(\mathcal{P}, D_0, D_1, R) \in \bar{\mathscr{G}}} R. \tag{5.4}$$

The following theorem gives an outer bound of the achievable region.

Theorem 9 *Given a distribution* $\mathcal{P}$ *on* $\mathcal{X}$*, if* (R, D_0, D_1) *is achievable, then:*

$$R \geq R(\mathcal{P}, D_0, D_1).$$

5.2 Notations

Before we start to prove the main results, we first introduce some notations that is useful in the proof.

We use capital letters to denote random variables and we use lower case letters to denote their realizations. We use log to denote the binary logarithm. We also use $H(\mathcal{P})$ to denote the entropy of the random variable X with distribution $\mathcal{P}$ and use $H_2(p)$ to denote the entropy of a Bernoulli(p) random variable.

For any distribution $\mathcal{Q}$ and conditional distribution V, let $H(V|\mathcal{Q})$ denote the conditional entropy $H(X|Y)$ of random variables X and Y such that Y has distribution $\mathcal{Q}$ and V describes the conditional distribution of X given Y.

We adopt the definitions of strong typicality of sequences [10, Ch. 10] [23].

Definition 15 *A sequence $x^n \in \mathcal{X}^n$ is said to be* δ-strongly typical *with respect to a distribution $P_X(x)$ on $\mathcal{X}$ if:*

1. *For all $a \in \mathcal{X}$ with $P_X(a) > 0$, we have*

$$\left|\frac{1}{n}N(a|x^n) - P_X(a)\right| < \frac{\delta}{|\mathcal{X}|}.$$

2. *For all $a \in \mathcal{X}$ with $P_X(a) = 0$, $N(a|x^n) = 0$.*

Here $N(a|x^n)$ is the number of occurrences of the symbol a in the sequence x^n. The set of sequences $x^n \in \mathcal{X}^n$ such that x^n is δ-strongly typical is called the strongly typical set *and is denoted $T^{(n,\delta)}(P_X)$.*

Let $\{r_n\}_{n=1}^{\infty}$ be a sequence of natural numbers such that $r_n \cdot n^{-1} \to 0$, and $r_n \cdot n^{-1/2} \to \infty$ as $n \to \infty$. Define

$$T^n(P_X) := T^{(n,|\mathcal{X}|r_n/n)}(P_X).$$

We fix such a sequence throughout the rest of the paper.

The *type* of $x^n \in \mathcal{X}^n$ is a pmf P_{x^n} on $\mathcal{X}$ where $P_{x^n}(a)$ is the relative frequency of the symbol a ($a \in \mathcal{X}$) in x^n. For convenience, we use Q_{y^n} to denote the type of $y^n \in \mathcal{Y}^n$ and use R_{z^n} to denote the type of $z^n \in \mathcal{Z}^n$. For any distribution $\mathcal{P}$ on $\mathcal{X}$, we use $\bar{T}^n(\mathcal{P})$ to denote the set of sequences of type $\mathcal{P}$ in $\mathcal{X}^n$. A distribution $\mathcal{P}$ on $\mathcal{X}$ is called a *type of sequences* in $\mathcal{X}^n$ if $\bar{T}^n(\mathcal{P}) \neq \emptyset$.

We also adopt Definition 2.4, 2.9 and 2.11 in [11].

Definition 16 *We say that $x^n \in \mathcal{X}^n$ has* conditional type *V given $y^n \in \mathcal{Y}^n$ if*

$$N(a, b|x^n, y^n) = N(b|y^n)V(a|b).$$

*For any given y^n and stochastic matrix $V : X \to Y$, the set of sequences $x^n \in \mathcal{X}^n$ having conditional type V given y^n will be called the V-*shell *of y^n, denoted by $T_V(y^n)$.*

REMARK: The conditional type of x^n given y^n is not uniquely determined if some $b \in \mathcal{Y}$ do not occur in y^n. Still, the set $T_V(y^n)$ containing x^n is unique. Moreover, if V_1 and V_2 are both conditional type of x^n given y^n, then $H(V_1|Q_{y^n}) = H(V_2|Q_{y^n})$. For any $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, we use $F_{x^n|y^n} : \mathcal{Y} \to \mathcal{X}$ to denote any one of the stochastic matrix such that

$$x^n \in T_{F_{x^n|y^n}}(y^n).$$

The choice of $F_{x^n|y^n}$ will *not* affect $H(F_{x^n|y^n}|Q_{y^n})$

Definition 17 *For a stochastic matrix $V : \mathcal{Y} \to \mathcal{X}$, a sequence $x^n \in \mathcal{X}^n$ is V-*typical *under the condition $y^n \in \mathcal{Y}^n$ with constant δ if*

$$\left|\frac{1}{n}N(a,b|x^n,y^n) - \frac{1}{n}N(b|y^n)V(a|b)\right| \le \delta, \quad \forall a \in \mathcal{X}, b \in \mathcal{Y},$$

in addition, $N(a,b|x^n,y^n) = 0$ whenever $V(a|b) = 0$. The set of such sequences x^n will be denoted as $T_{[V]_\delta}(y^n)$.

Definition 18 *Recall the sequence $\{r_n\}_{n=1}^\infty$ defined in Definition 15. Define $T_{[V]}(y^n) := T_{[V]_{r_n/n}}(y^n)$. For any set $\mathcal{B} \subseteq \mathcal{Y}^n$, define $T_{[V]}(\mathcal{B}) = \bigcup_{y^n \in \mathcal{B}} T_{[V]}(y^n)$.*

It is useful to recall three distance measures for pmf's on $\mathcal{X}$. The *Kullback-Leibler (K-L) divergence* between the pmf's P_1 and P_2 on $\mathcal{X}$ is defined as:

$$D(P_1\|P_2) = \sum_{a\in\mathcal{X}} P_1(a)\log\frac{P_1(a)}{P_2(a)}.$$

The *variational distance* between P_1 and P_2 is defined as:

$$\|P_1 - P_2\| = \sum_{a\in\mathcal{X}} |P_1(a) - P_2(a)|.$$

The *Kolmogorov distance* between P_1 and P_2 is defined as:

$$||P_1 - P_2||_\infty = \max_{a \in \mathcal{X}} |P_1(a) - P_2(a)|.$$

Both the variational distance and the Kolmogorov distance satisfy the Triangle Inequality.

We also adopt the following notation in [23] which is similar to K-L divergence. For random variables X_1 and X_2 ($X_i \in \mathcal{X}$) with distributions $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively, define

$$H_{\mathcal{P}_1}(\mathcal{P}_2) = H_{X_1}(X_2) := \sum_{x \in \mathcal{X}} \mathcal{P}_2(x) \log \frac{\mathcal{P}_1(x)}{\mathcal{P}_2(x)}$$

For any random variables X, Y and a distribution $\bar{\mathcal{P}}$ on $\mathcal{X}$, define

$$H_{\bar{\mathcal{P}}}(X|Y) := \sum_{y \in \mathcal{Y}} P_Y(y) \cdot \sum_{x \in \mathcal{X}} P_{X|Y}(x|y) \log \frac{\bar{\mathcal{P}}(x)}{P_{X|Y}(x|y)}.$$

Finally, we introduce some topology prerequisites. For any finite set $\mathcal{X}$, let $\mathcal{S}$ be the set of distributions on $\mathcal{X}$. $\mathcal{S}$ is convex. We can obtain a metric space $(\mathcal{S}, d)$ by supplying $\mathcal{S}$ with a distance function d as follows (see Definition 7.1.1 in [27]):

$$d(\mathcal{P}_1, \mathcal{P}_2) = ||\mathcal{P}_1 - \mathcal{P}_2||_\infty.$$

For any function $f : \mathcal{S} \to \Re$, we always assume that the underlying metric spaces are $(\mathcal{S}, d)$ and the Euclidean space. f is continuous at $Q \in \mathcal{S}$, if and only if for any $\epsilon > 0$, there exists a $\delta > 0$ such that for any Q' with $||Q' - Q||_\infty \leq \delta$, we have: $|f(Q') - f(Q)| \leq \epsilon$. The function f is continuous on $\mathcal{S}$ if it is continuous at each point of $\mathcal{S}$. (see Definition 7.1.6 in [27]).

Similarly, for any finite sets $\mathcal{X}$ and $\mathcal{Y}$, let $\mathcal{S}$ be the set of all stochastic matrices from $\mathcal{Y}$ to $\mathcal{X}$. For any two stochastic matrices $V, W : \mathcal{Y} \to \mathcal{X}$, we let:

$$d(V, W) = ||V - W||_\infty := \max_{a \in \mathcal{X}, b \in \mathcal{Y}} |V(a|b) - W(a|b)|.$$

Then, by supplying $\mathcal{S}$ with distance d, we can obtain a metric space $(\mathcal{S}, d)$. Similarly, we always assume the underlying metric space is $(\mathcal{S}, d)$.

5.3 Proof of Theorem 7

Consider the following coding scheme for any distribution P_{UYZ} without the restriction $U \to Y \to X$.

1. *Generation of codebook:* Generate $2^{\lfloor nR_1 \rfloor}$ sequences

$$\left\{U^n(1), ..., U^n(2^{\lfloor nR_1 \rfloor})\right\},$$

drawn i.i.d. according to P'_{U^n} defined as follows:

$$P'_{U^n}(u^n) = \begin{cases} 0, & u^n \notin T^{(n,\delta/2)}(P_U) \\ P_U^n(u^n) \cdot \frac{1}{P_U^n(T^{(n,\delta/2)}(P_U))}, & \text{otherise.} \end{cases}$$

Here $P_U^n(u^n) = \prod_{i=1}^n P_U(u_i)$. Then for each $U^n(i)$, define

$$\mathcal{T}(U^n(i)) = \left\{y^n : (U^n(i), y^n) \in T^{(n,\delta)}(P_{UY})\right\}.$$

Generate $2^{\lfloor nR_2 \rfloor}$ sequences

$$\left\{Y^n(i,1), ..., Y^n(i, 2^{\lfloor nR_2 \rfloor})\right\},$$

drawn i.i.d. according to the conditional distribution $P'_{Y^n|U^n}(\cdot|U^n = U^n(i))$ defined as follows:

$$P'_{Y^n|U^n}(y^n|U^n = U^n(i)) = \begin{cases} 0, & y^n \notin \mathcal{T}(U^n(i)) \\ P_{Y^n|U^n}(y^n|U^n(i)) \cdot \frac{1}{P_{Y^n|U^n}(\mathcal{T}(U^n(i))|U^n(i))}, & \text{othewise.} \end{cases}$$

Here $P_{Y^n|U^n}(y^n|u^n) = \prod_{j=1}^n P_{Y|U}(y_j|u_j)$. Let

$$C(n) = \bigcup_{\substack{i \in 2^{\lfloor nR_1 \rfloor}, \\ j \in 2^{\lfloor nR_2 \rfloor}}} \{(U^n(i), Y^n(i,j))\}$$

denote the random codebook. (If $\mathcal{T}(U^n(i))$ is an empty set, then we let $Y^n(i,j) = \vec{0}$ for all j.)

2. *Encoding:* Encode X^n by $(i(X^n), j(X^n)) = (i, j)$ if there exists a pair such that

$$(U^n(i), Y^n(i, j), X^n) \in T^{(n,\delta)}(P_{UYX}).$$

If there is more than one such pair, send the pair with the least i; if there are more than one such pair with the least i, send the pair with the least j. If there is no such pair, let $i = j = 0$. We can use an integer in $\left[0, 2^{\lfloor nR_1 \rfloor + \lfloor nR_2 \rfloor}\right]$ to represent (i, j). For convenience, we write (i, j) to denote its *integer representation*.

3. *Decoding:* If $(i, j) = (0, 0)$, reproduce the n-sequence with all a's: $(a, ..., a)$, where a is some arbitrary letter of the reconstruction alphabet. Else, reproduce $Y^n(i, j)$.

It is worth noting that for any $u^n \in T^{(n,\delta/2)}(P_U)$ and $y^n \in \mathcal{T}(u^n)$, we always have:

$$P_U^n(u^n) \leq P'_{U^n}(u^n) \tag{5.5}$$

$$P_{Y^n|U^n}(y^n|u^n) \leq P'_{Y^n|U^n}(y^n|u^n). \tag{5.6}$$

We next calculate the distortion. Fix a distribution P_{UYX} $(U \to Y \to X)$ and rates R_1, R_2. Any arbitrary coefficients n, δ along with a codebook C gives a specific coding scheme. For any C, n, δ, let

$$E_1(n, \delta) = \left\{x^n : \forall u^n \in \mathcal{U}^n, y^n \in \mathcal{Y}^n, (u^n, y^n, x^n) \notin T^{(n,\delta)}(P_{UYX})\right\};$$

$$E_2(C, n, \delta) = \{x^n : x^n \notin E_1(n, \delta);$$

$$\forall (i, j) \in \left[2^{\lfloor nR_1 \rfloor}\right] \times \left[2^{\lfloor nR_2 \rfloor}\right], (u^n(i), y^n(i, j), x^n) \notin T^{(n,\delta)}(P_{UYX})\}.$$

Let $E_0(C, n, \delta) = E_1(n, \delta) \cup E_2(C, n, \delta)$ denote the encoding error event.

The following lemma is useful and is proved in Appendix.

Lemma 23 *Fix a distribution P_{UYX}. Let $\mathcal{P}$ denote P_X. Let $R_1, R_2 \in \mathfrak{R}^+$ be the rates in the above coding scheme.*

1. *If* $R_1 > I(U;X)$, $R_1 + R_2 > I(U;X) + I(X;Y|U)$, *then for all sufficiently small* $\delta > 0$,

$$\lim_{n\to\infty} E_{C(n)}[\mathcal{P}^n(E_0(C(n), n, \delta))] = 0.$$

2. *Let* $C > 1$ *be an arbitrary fixed constant. For any* (n, δ), *we let* $\delta^* = C\delta$. *For any distribution* P^*_{UYX} *satisfying* $P^*_{UY} = P_{UY}$ *and* $P^*_X = P_X$, *let* (U^*, Y^*, X^*) *be a triple of random variables with distribution* P^*_{UYX}. *Let*

$$\mathcal{T}^{(n,\delta^*)}(P^*_{UYX}) = \Bigg\{(u^n, y^n, x^n) : \|P_{u^n,y^n,x^n} - P^*_{UYX}\|_\infty \le \frac{\delta^*}{|\mathcal{U}||\mathcal{Y}||\mathcal{X}|};$$
$$N(a, b|u^n, y^n) = 0, \forall (a, b) \in \mathcal{U} \times \mathcal{Y}, P^*_{UY}(a, b) = 0\}.$$

Define:

$$E^*_1(n, \delta) = \{x^n : x^n \notin T^{(n,\delta^*)}(P_X)\};$$
$$E^*_2(C, n, \delta, P^*_{UYX}) = \Big\{x^n : x^n \notin E^*_1(n, \delta);$$
$$\forall (i, j) \in \left[2^{\lfloor nR_1 \rfloor}\right] \times \left[2^{\lfloor nR_2 \rfloor}\right], (u^n(i), y^n(i, j), x^n) \notin \mathcal{T}^{(n,\delta^*)}(P^*_{UYX})\Big\}.$$

and let

$$E^*_0(C, n, \delta, P^*_{UYX}) = E^*_1(n, \delta) \cup E^*_2(C, n, \delta, P^*_{UYX}).$$

Observe that

$$E^{*c}_0(C, n, \delta, P^*_{UYX}) = \Big\{x^n : \exists (i, j) \in \left[2^{\lfloor nR_1 \rfloor}\right] \times \left[2^{\lfloor nR_2 \rfloor}\right]$$
$$s.t.\ (u^n(i), y^n(i, j), x^n) \in \mathcal{T}^{(n,\delta^*)}(P^*_{UYX})\}.$$

Then, there exists a $\xi(\delta, P^*_{UYX})$, *such that for any* $\delta > 0$, *if*

$$\limsup_{n\to\infty} E_{C(n)}[\mathcal{P}^n(E^{*c}_0(C(n), n, \delta, P^*_{UYZ}))] > 0,$$

then we have: $R_1 \ge I(U^*;X^*) - \xi(\delta, P^*_{UYZ})$, $R_1 + R_2 \ge I(U^*;X^*) + I(X^*;Y^*|U^*) - \xi(\delta, P^*_{UYZ})$. *Moreover, if* $\{(\delta_k, P^{(k)}_{UYX})\}$ *is a sequence satisfying:*

$$P^{(k)}_{UY} = P_{UY}, \quad P^{(k)}_X = P_X,$$

and

$$\lim_{k\to\infty} \delta_k = 0, \quad \lim_{k\to\infty} P^{(k)}_{UYX} = P^*_{UYX},$$

then

$$\lim_{k\to\infty} \xi(\delta_k, P^{(k)}_{UYX}) = 0.$$

From now on, we set $R_1 = I(U;X) + \delta$, $R_2 = I(X;Y|U) + \delta$. Then, R_1 and R_2 are determined by distribution P_{UYX} and δ. Thus, for a fixed distribution P_{UYX}, any arbitrary coefficients n, δ along with a codebook C gives a specific coding scheme. For any source message X^n, let $\hat{X}^n(C, n, \delta)$ denote the reconstruction of the decoder under that coding scheme; let $\tilde{X}^n(C) = g(\tilde{i}(X^n), \tilde{j}(X^n))$ denote the reconstruction under the most powerful adversarial attack:

$$d(x^n, \tilde{x}^n(C)) \equiv \max_{(i,j)\in(0,0)\cup\left[2^{\lfloor nR_1 \rfloor}\right]\times\left[2^{\lfloor nR_2 \rfloor}\right]} d(x^n, g(i,j)),$$

where g denotes the corresponding decoding function. We next prove Theorem 7 by showing that for any $\epsilon > 0$, we can use P_{UYX} with $U \to Y \to X$ to construct a code with coefficients $(n_\epsilon, \delta_\epsilon)$ and a specific codebook C_ϵ such that both the rate constraint and Inq. (2.5), (2.6) are satisfied.

For any specific codebook C, let

$$A_0(C, n, \delta) = \{x^n : d(x^n, \hat{x}^n(C, n, \delta)) > D_0 + \epsilon\};$$

$$A_1(C) = \{x^n : d(x^n, \tilde{x}^n(C)) > D_1 + \epsilon\}.$$

(Here $A_1(C)$ is only determined by C.) Then we have

$$\lim_{\delta\to 0}\lim_{n\to\infty} E_{C(n)}[\mathcal{P}^n(A_0(C(n), n, \delta))] = 0; \tag{5.7}$$

$$\lim_{\delta\to 0}\lim_{n\to\infty} E_{C(n)}[\mathcal{P}^n(A_1(C(n)))] = 0. \tag{5.8}$$

Eq. (5.7) follows from Lemma 23 Part 1). Eq. (5.8) is proved in Appendix. Now, for each (n, δ), we let $C^{(n,\delta)}$ be the codebook that minimize

$$\mathcal{P}^n(A_0(C, n, \delta)) + \mathcal{P}^n(A_1(C)).$$

By Eq. (5.7) and Eq. (5.8), we have:

$$\lim_{\delta \to 0} \lim_{n \to \infty} E_{C(n)} \left[\mathcal{P}^n(A_0(C(n, \delta), n, \delta)) + \mathcal{P}^n(A_1(C(n, \delta))) \right] = 0.$$

Thus, we can find $n_\epsilon, \delta_\epsilon$ satisfying $2\delta_\epsilon + 1/n_\epsilon < \epsilon$ and a codebook $C_\epsilon = C^{(n_\epsilon, \delta_\epsilon)}$ such that:

$$\mathcal{P}^{n_\epsilon}(A_0(C^{(n_\epsilon,\delta_\epsilon)}, n_\epsilon, \delta_\epsilon)) < \epsilon;$$

$$\mathcal{P}^{n_\epsilon}(A_1(C^{(n_\epsilon,\delta_\epsilon)})) < \epsilon,$$

and

$$R_{n_\epsilon} = \frac{1}{n_\epsilon} \left(\lfloor nR_1 \rfloor + \lfloor nR_2 \rfloor + 1 \right) \leq (I(U; X) + \delta_\epsilon) + (I(X; Y|U) + \delta_\epsilon) + \frac{1}{n_\epsilon} < I(X; Y) + \epsilon.$$

5.4 Proof of Theorem 9

5.4.1 Blowing-up Lemma

The following lemma (Blowing-up Lemma) is useful and is proved in [11, Chapter. 5].

Lemma 24 *Let $\mathcal{X}$ be a finite set. Let $\mathcal{P}$ denote a probability distribution on $\mathcal{X}$. Let $d_H : \mathcal{X} \times \mathcal{X} \mapsto \Re^+$ denote the Hamming distance. We extend $d_H(\cdot, \cdot)$ to strings in the following way:*

$$d_H(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i).$$

For any set $A \in \mathcal{X}^n$, let $\Gamma^d(A) := \{x^n \in \mathcal{X}^n : \exists y^n \in A, d_H(x^n, y^n) < d\}$. Let ϵ_n be an arbitrary positive sequence where $\lim_{n\to\infty} \epsilon_n = 0$. Then, there exists two positive

sequences $\{\delta_n\}$, $\{\eta_n\}$ *where* $\lim_{n\to\infty}\delta_n = \lim_{n\to\infty}\eta_n = 0$, *s.t. if* $A_n \in \mathcal{X}^n$ *and* $\mathcal{P}^n(A_n) \geq 2^{-n\epsilon_n}$, *then*

$$\mathcal{P}^n(\Gamma^{\delta_n}(A_n)) \geq 1 - \eta_n.$$

5.4.2 An equivalent expression of $\mathcal{S}(\mathcal{P}, \mathcal{Q}, D_0, D_1)$

We again adopt the definitions in [22,23] with slight modifications.

Definition 19 *For* $\mathcal{B} \in \mathcal{Y}^n$ $(n = 1, 2, \ldots)$, *a probability distribution* $\mathcal{Q}$ *on* $\mathcal{Y}$, $0 < \eta < 1$, *the minimum size of the* η*-image of* $\mathcal{B}$ *via channel* V *will be defined as:*

$$G_{V,\mathcal{Q}}(\mathcal{B}, \eta) = \min\left\{\mathcal{P}^n(\mathcal{A}) : \mathcal{A} \subseteq \mathcal{X}^n, V^n(\mathcal{A}|y^n) \geq \eta, \quad \forall y^n \in \mathcal{B}\right\},$$

where $\mathcal{P} = \mathcal{Q} \cdot V$. $G_{W,\mathcal{Q}}(\mathcal{B}, \eta)$ *is defined analogously via channel* W.

Definition 20 *Channel* V *is less noisy than channel* W *under distribution* $\mathcal{Q}$ *if for every* $0 < \eta < 1$,

$$\liminf_{n\to\infty} \frac{1}{n} \min_{\mathcal{B}\subseteq T^n(\mathcal{Q})} \left[\log G_{W,\mathcal{Q}}(\mathcal{B}, \eta) - \log G_{V,\mathcal{Q}}(\mathcal{B}, \eta)\right] \geq 0.$$

We write $W \overset{\mathcal{Q}}{\ll} V$ *if* V *is less noisy than* W *under* $\mathcal{Q}$.

Definition 21 *For any distribution* $\mathcal{Q}$ *on* $\mathcal{Y}$ *and stochastic matrices* $V, W : \mathcal{Y} \to \mathcal{X}$, *let* $\mathcal{P} = \mathcal{Q} \cdot V$ *and* $\mathcal{R} = \mathcal{Q} \cdot W$. *Define*

$$T_{V,W,\mathcal{Q}}(\mathcal{Q}) := \inf\{H_{\mathcal{R}}(\tilde{Z}|\tilde{U}) - H_{\mathcal{P}}(\tilde{X}|\tilde{U}) : (\tilde{U}, \tilde{Y}, \tilde{X}, \tilde{Z}) \in \mathscr{P}_{V,W}(\mathcal{Q})\}.$$

Lemma 25 *Fix stochastic matrices* $V, W : \mathcal{Y} \to \mathcal{X}$ *and a distribution* $\mathcal{Q}$ *on* $\mathcal{Y}$. *The function* $T_{V,W,\mathcal{Q}}(\tilde{\mathcal{Q}})$ *is a continuous function of* $\tilde{\mathcal{Q}}$.

Proof: See the appendix.

Proposition 26 *Fix a distribution Q on $\mathcal{Y}$. Let $V, W : \mathcal{Y} \to X$ be two stochastic matrices. The following three statements are equivalent:*

1. $W \overset{Q}{\ll} V$;
2. $I(U;Z) \le I(U;X)$ *for any* $(U,Y,X,Z) \in \mathscr{P}_{V,W}(Q)$ *with* $|\mathcal{U}| \le |\mathcal{Y}| + 2$;
3. $I(U;Z) \le I(U;X)$ *for any* $(U,Y,X,Z) \in \mathscr{P}_{V,W}(Q)$.

Proof: See the appendix.

Via Proposition 26, we now have another expression of $\mathscr{S}(\mathcal{P}, Q, D_0, D_1)$:

$$\begin{aligned}\mathscr{S}(\mathcal{P}, Q, D_0, D_1) = \{V : Q \cdot V = \mathcal{P}; d(V,Q) \le D_0;\\ \forall W, \text{ s.t. } Q \cdot W = \mathcal{P}, W \overset{Q}{\ll} V,\\ \text{we have } d(W,Q) \le D_1\}.\end{aligned}$$

5.4.3 Proof of Theorem 9

In this section, we fix the distribution of the source: $P_X = \mathcal{P}$. The following proof is similar to the proof of Proposition 1 in [21].

For any ϵ, let $\{(f_n, g_n)\}$ be a sequence of codes with rate $\le R + \epsilon$ such that Inq. (2.5) and (2.6) hold for all n sufficiently large. We can assume that $f_n : \mathcal{X}^n \to [M_n]$ is a surjection and $g_n : [M_n] \to \hat{\mathcal{X}}^n$ is an injection. The reason is as follows. If f_n is not a surjection, let $\mathcal{A}$ denote the image of $\mathcal{X}^n$ under f_n and let $f_n' : \mathcal{X}^n \to \mathcal{A}$ be a surjection s.t. $f_n' = f_n$. Let $g_n' : \mathcal{A} \to \hat{\mathcal{X}}^n$ be the restriction of g_n to $\mathcal{A}$. Then (f_n', g_n') is a code

with strictly less rate compared to (f_n, g_n). Furthermore, Inq. (2.5) and (2.6) hold for (f'_n, g'_n) as long as they hold for (f_n, g_n). If g_n is not an injection, then let $f'_n(x^n)$ equals the least index i such that $g_n(i) = g_n(f_n(x^n))$. Let $\mathcal{A}$ denote the image of $\mathcal{X}^n$ under f'_n. Let $g'_n : \mathcal{A} \to \hat{\mathcal{X}}^n$ be the restriction of g_n to $\mathcal{A}$. g'_n is an injection. Again, (f'_n, g'_n) is a code with strictly less rate compared to (f_n, g_n). Inq. (2.5) and (2.6) hold for (f'_n, g'_n) as long as they hold for (f_n, g_n). For any code $\{(f_n, g_n)\}$, let

$$C_{f_n,g_n} = \{y^n(i) : y^n(i) = g_n(i), i = 1, \ldots, M_n\}$$

denote the set of the codewords. Here, $M_n \leq 2^{n(R+\epsilon)}$ and $\{y^n(i)\}$ are pairwise different.

For any $\epsilon > 0$, let

$$A_{i,n} = \{x^n \in \mathcal{X}^n : g_n(f_n(x^n)) = y^n(i)\},$$
$$B_{i,n}(\epsilon) = \{x^n \in A_{i,n} : d(x^n, y^n(i)) \leq D_0 + \epsilon\}.$$

It is straightforward that $\{A_{i,n}\}_{i=1}^{M_n}$ is a partition of $\mathcal{X}^n$. Moreover, since Inq. (1) holds for all large n, we have

$$P\left(\bigcup_{i=1}^{M_n} B_{i,n}(\epsilon)\right) \geq 1 - \epsilon.$$

For any $\epsilon > 0$, we also define

$$\mathscr{F}_\epsilon(\mathcal{P},\mathcal{Q}, D_0, D_1) = \{V : V \in \mathscr{S}(\mathcal{P}, \mathcal{Q}, D_0 + \epsilon),$$
$$\exists W, \text{ s.t. } \mathcal{Q} \cdot W = \mathcal{P}, W \overset{Q}{\ll} V, d(W, \mathcal{Q}) \geq D_1 + 2\epsilon\},$$

and

$$\mathscr{F}^c_\epsilon(\mathcal{P}, \mathcal{Q}, D_0, D_1) = \mathscr{S}(\mathcal{P}, \mathcal{Q}, D_0 + \epsilon) \backslash \mathscr{F}_\epsilon(\mathcal{P}, \mathcal{Q}, D_0, D_1).$$

It is obvious that $\mathscr{F}_\epsilon(\mathcal{P}, \mathcal{Q}, D_0, D_1)$ and $\mathscr{F}^c_\epsilon(\mathcal{P}, \mathcal{Q}, D_0, D_1)$ are disjoint, and

$$\mathscr{F}_\epsilon(\mathcal{P}, \mathcal{Q}, D_0, D_1) \cup \mathscr{F}^c_\epsilon(\mathcal{P}, \mathcal{Q}, D_0, D_1) = \mathscr{S}(\mathcal{P}, \mathcal{Q}, D_0 + \epsilon), \tag{5.9}$$
$$\mathscr{F}^c_\epsilon(\mathcal{P}, \mathcal{Q}, D_0, D_1) \subseteq \mathscr{S}(\mathcal{P}, \mathcal{Q}, D_0 + \epsilon, D_1 + 2\epsilon). \tag{5.10}$$

For any $\delta > 0$ and $\zeta > 0$, it holds for all n sufficiently large (depending on ζ) that $\mathcal{P}^n(T^n([\mathcal{P}]_\delta)) \geq 1 - \zeta$, where

$$T^n([\mathcal{P}]_\delta) = \bigcup_{\substack{\tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|<\delta,\\ \exists x^n\in\mathcal{X}^n, P_{x^n}=\tilde{\mathcal{P}}}} \bar{T}^n(\tilde{\mathcal{P}}).$$

Thus, similar to the proof of the main theorem in [21], for all large n (depending on ϵ, δ, ζ)

$$\begin{aligned} 1 - \epsilon - \zeta &\leq \sum_{i=1}^{M_n} \mathcal{P}^n\left(T^n([\mathcal{P}]_\delta) \cap B_{i,n}(\epsilon)\right) \\ &= \sum_{i=1}^{M_n} \sum_{\substack{\tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|<\delta,\\ \exists x^n\in\mathcal{X}^n, P_{x^n}=\tilde{\mathcal{P}}}} \mathcal{P}^n(\bar{T}^n(\tilde{\mathcal{P}}) \cap B_{i,n}(\epsilon)). \end{aligned} \tag{5.11}$$

A suitable choice of ϵ, δ, ζ will be given later.

Recall that $T_V(y^n(i))$ denote the V-shell of $y^n(i)$. Then for $i = 1, \ldots, M_n$,

$$\bar{T}^n(\tilde{\mathcal{P}}) \cap B_{i,n}(\epsilon) = \bigcup_{V:V\in\mathscr{S}(\tilde{\mathcal{P}},Q_{y^n(i)},D_0+\epsilon)} \left(T_V(y^n(i)) \cap B_{i,n}(\epsilon)\right). \tag{5.12}$$

Let

$$S_1(n,\epsilon,\delta) = \sum_{i=1}^{M_n} \sum_{\substack{\tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\leq\delta,\\ \exists x^n\in\mathcal{X}^n, P_{x^n}=\tilde{\mathcal{P}};\\ V\in\mathcal{F}_\epsilon^c(\tilde{\mathcal{P}},Q_{y^n(i)},D_0,D_1)}} \mathcal{P}^n(T_V(y^n(i)) \cap B_{i,n}(\epsilon)),$$

$$S_2(n,\epsilon,\delta) = \sum_{i=1}^{M_n} \sum_{\substack{\tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\leq\delta,\\ \exists x^n\in\mathcal{X}^n, P_{x^n}=\tilde{\mathcal{P}};\\ V\in\mathcal{F}_\epsilon(\tilde{\mathcal{P}},Q_{y^n(i)},D_0,D_1)}} \mathcal{P}^n(T_V(y^n(i)) \cap B_{i,n}(\epsilon)).$$

Combining Inq. (5.11) and Eq. (5.12), we know that

$$S_1(n,\epsilon,\delta) + S_2(n,\epsilon,\delta) \geq 1 - \epsilon - \zeta \tag{5.13}$$

for all n sufficiently large.

Lemma 2.5 in [11] gives:

$$\frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} 2^{nH(V|Q_{y^n})} \leq |T_V(y^n)| \leq 2^{nH(V|Q_{y^n})}. \tag{5.14}$$

Therefore,

$$\begin{aligned}
&S_1(n,\epsilon,\delta)\\
=&\sum_{i=1}^{M_n} \sum_{\substack{\tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\leq\delta,\\ \exists x^n\in\mathcal{X}^n, P_{x^n}=\tilde{\mathcal{P}};\\ V\in\mathcal{F}_\epsilon^c(\tilde{\mathcal{P}},Q_{y^n(i)},D_0,D_1)}} \left\{|T_V(y^n(i))\cap B_{i,n}(\epsilon))|\times 2^{-n\left(H(\tilde{\mathcal{P}})+D(\tilde{\mathcal{P}}\|\mathcal{P})\right)}\right\}\\
\leq&\sum_{i=1}^{M_n} \sum_{\substack{\tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\leq\delta,\\ \exists x^n\in\mathcal{X}^n, P_{x^n}=\tilde{\mathcal{P}};\\ V\in\mathcal{F}_\epsilon^c(\tilde{\mathcal{P}},Q_{y^n(i)},D_0,D_1)}} \left\{|T_V(y^n(i))|\times 2^{-n\left(H(\tilde{\mathcal{P}})+D(\tilde{\mathcal{P}}\|\mathcal{P})\right)}\right\}\\
\leq&\sum_{i=1}^{M_n} \sum_{\substack{\tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\leq\delta,\\ \exists x^n\in\mathcal{X}^n, P_{x^n}=\tilde{\mathcal{P}}}} \max_{V\in \mathcal{F}_\epsilon^c(\tilde{\mathcal{P}},Q_{y^n(i)},D_0,D_1)} \left\{2^{n\left(H(\tilde{\mathcal{P}})-I(V,Q_{y^n(i)})\right)}\times(n+1)^{|\mathcal{X}||\mathcal{Y}|}2^{-nH(\tilde{\mathcal{P}})}\right\}\\
\leq&(n+1)^{|\mathcal{X}|(|\mathcal{Y}|+1)}2^{n(R+\epsilon)}\cdot 2^{-n\min_{(\tilde{\mathcal{P}},i):\|\tilde{\mathcal{P}}-\mathcal{P}\|\leq\delta} R(\tilde{\mathcal{P}},Q_{y^n(i)},D_0+\epsilon,D_1+2\epsilon)}.
\end{aligned}$$

The second inequality uses Inq. (5.14), $D(\tilde{\mathcal{P}}\|\mathcal{P})\geq 0$ and the fact that there are at most $(n+1)^{|\mathcal{X}||\mathcal{Y}|}$ different *conditional types* (Lemma 2.2 in [11]). The last inequality uses the fact that there are at most $(n+1)^{|\mathcal{X}|}$ *types of sequences*, $M_n \leq 2^{n(R+\epsilon)}$ and Inq. (5.10). Although we have M_n codewords, $Q_{y^n(i)}$ can only take values in $\tilde{M}_n$ pmf's: $\{\tilde{Q}_{j,n}\}_{j=1}^{\tilde{M}_n}$, where $\tilde{M}_n < (n+1)^{|\mathcal{Y}|}$. Therefore,

$$S_1(n,\epsilon,\delta) \leq (n+1)^{|\mathcal{X}|(|\mathcal{Y}|+1)}\cdot 2^{n\left(R+\epsilon-\min_{(\tilde{\mathcal{P}},j):\|\tilde{\mathcal{P}}-\mathcal{P}\|\leq\delta} R(\tilde{\mathcal{P}},\tilde{Q}_{j,n},D_0+\epsilon,D_1+2\epsilon)\right)}. \tag{5.15}$$

Let $\mathscr{B}_{j,n} = \{y^n(i) : Q_{y^n(i)} = \tilde{Q}_{j,n}\}$ and let $T_V(\mathscr{B}_{j,n}) = \bigcup_{y^n(i)\in\mathscr{B}_{j,n}} T_V(y^n(i))$. Conse-

quently,

$$
\begin{aligned}
&S_2(n,\epsilon,\delta)\\
\le&(n+1)^{|\mathcal{X}||\mathcal{Y}|}\times\sum_{\substack{\tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\le\delta,\\ \exists x^n\in\mathcal{X}^n,P_{x^n}=\tilde{\mathcal{P}}}}\sum_{j=1}^{\tilde{M}_n}\max_{V\in\mathcal{F}_\epsilon(\tilde{\mathcal{P}},\tilde{\mathcal{Q}}_{j,n},D_0,D_1)}\mathcal{P}^n(T_V(\mathscr{B}_{j,n}))\\
\le&(n+1)^{(|\mathcal{X}|+1)|\mathcal{Y}|}\times\sum_{\substack{\tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\le\delta,\\ \exists x^n\in\mathcal{X}^n,P_{x^n}=\tilde{\mathcal{P}}}}\max_{\substack{(V,j):V\in\\ \mathcal{F}_\epsilon(\tilde{\mathcal{P}},\tilde{\mathcal{Q}}_{j,n},D_0,D_1)}}\mathcal{P}^n(T_V(\mathscr{B}_{j,n}))\\
\le&(n+1)^{(|\mathcal{X}|+1)|(\mathcal{Y}|+1)}\max_{\substack{(V,j,\tilde{\mathcal{P}}):V\in\\ \mathcal{F}_\epsilon(\tilde{\mathcal{P}},\tilde{\mathcal{Q}}_{j,n},D_0,D_1);\\ \tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\le\delta,\\ \exists x^n\in\mathcal{X}^n,P_{x^n}=\tilde{\mathcal{P}}}}\mathcal{P}^n(T_V(\mathscr{B}_{j,n})).
\end{aligned}
\tag{5.16}
$$

Here, the first inequality follows from the facts that $\{T_V(y^n(i))\cap B_{i,n}(\epsilon)\}_{i=1}^{M_n}$ are pairwise disjoint and

$$T_V(y^n(i))\cap B_{i,n}(\epsilon)\subseteq T_V(y^n(i))\subseteq T_V(\mathscr{B}_{j,n})$$

holds for any (i,j) such that $y^n(i)\in\mathscr{B}_{j,n}$.

Suppose $(V_n,j_n,\tilde{\mathcal{P}}_n)$ achieves the maximum $\mathcal{P}^n(T_V(\mathscr{B}_{j,n}))$ for each n (here $\tilde{\mathcal{P}}_n=\tilde{\mathcal{Q}}_{j_n,n}\cdot V_n$). Without loss of generality, suppose

$$\lim_{n\to\infty}(V_n,\tilde{\mathcal{Q}}_{j_n,n},\tilde{\mathcal{P}}_n)=(\tilde{V}^*,\tilde{\mathcal{Q}}^*,\tilde{\mathcal{P}}^*).$$

This shows that $S_2(n,\epsilon,\delta)$ is upper bounded by a polynomial of n times

$$\max_{\substack{(V,j,\tilde{\mathcal{P}}):V\in\\ \mathcal{F}_\epsilon(\tilde{\mathcal{P}},\tilde{\mathcal{Q}}_{j,n},D_0,D_1);\\ \tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\le\delta,\\ \exists x^n\in\mathcal{X}^n,P_{x^n}=\tilde{\mathcal{P}}}}\mathcal{P}^n(T_V(\mathscr{B}_{j,n}))$$

That is, there always exists (V_n,j_n), such that

$$S_2(n,\epsilon,\delta)\le(n+1)^{(|\mathcal{X}|+1)|(\mathcal{Y}|+1)}\mathcal{P}^n(T_{V_n}(\mathscr{B}_{j_n,n})).\tag{5.17}$$

We can further assume that for all large n, for any y^n with $Q_{y^n}=\tilde{\mathcal{Q}}_{j_n,n}$, $y^n\in T^n(\tilde{\mathcal{Q}}^*)$. The reason is as follows. If $\{\tilde{\mathcal{Q}}_{j_n,n}\}$ do not have the property, then let $\{\tilde{\mathcal{Q}}_{j'_n,n}\}$ be a sequence

of *types of sequences* on $\mathcal{Y}$ such that for all large n,

$$y^n \in T^n(\tilde{Q}^*), \quad \forall y^n \text{ s.t. } Q_{y^n} = \tilde{Q}_{j'_n,n}.$$

Here $\{\tilde{Q}_{j'_n,n}\}$ also converges to $\tilde{Q}^*$. For any codeword $y^n(i) \notin \mathscr{B}_{j_n,n}$, let $y'^n(i) = y^n(i)$. For any codeword $y^n(i) \in \mathscr{B}_{j_n,n}$, let $y'^n(i) \in \mathcal{Y}^n$ be a sequence with *type of sequences* $\tilde{Q}_{j'_n,n}$ which minimizes $d_H(y^n(i), y'^n(i))$. Since $\lim_{n\to\infty} \|\tilde{Q}_{j_n,n} - \tilde{Q}_{j'_n,n}\|_\infty = 0$, we have

$$\lim_{n\to\infty} d_H(y^n(i), y'^n(i)) = 0. \tag{5.18}$$

Now we can define another code (f'_n, g'_n) where $f'_n = f_n$ and $g'_n(i) = y'^n(i)$. By Eq. (5.18), for any n, there exists ϵ'_n, satisfying $\lim_{n\to\infty} \epsilon'_n = 0$, such that for any $y^n(i) \in \mathscr{B}_{j_n,n}$, $x^n \in \mathcal{X}^n$

$$\left|d(x^n, y^n(i)) - d(x^n, y'^n(i))\right| < \epsilon'_n.$$

We know that $\{(f'_n, g'_n)\}$ is a sequence of code with exactly the same rate ($< R + \epsilon + \epsilon'_n$) as $\{(f_n, g_n)\}$. Moreover, for all n sufficiently large,

$$\mathcal{P}^n\left(\{x^n : d(x^n, g(f'_n(x^n))) > D_0 + \epsilon + \epsilon'_n\}\right) < \epsilon + \epsilon'_n,$$
$$\mathcal{P}^n\left(\left\{x^n : \max_{C'} d(x^n, g'_n(C')) > D_1 + \epsilon + \epsilon'_n\right\}\right) < \epsilon + \epsilon'_n.$$

For the new code (f'_n, g'_n) and $\epsilon' \geq \epsilon + \epsilon'_n$, we can similarly define $S'_1(n, \epsilon', \delta)$ and $S'_2(n, \epsilon', \delta)$:

$$S'_1(n, \epsilon', \delta) = \sum_{i=1}^{M_n} \sum_{\substack{\tilde{\mathcal{P}}: \|\tilde{\mathcal{P}} - \mathcal{P}\| \leq \delta, \\ \exists x^n \in \mathcal{X}^n, P_{x^n} = \tilde{\mathcal{P}}; \\ V \in \mathcal{F}^c_{\epsilon'}(\tilde{\mathcal{P}}, Q_{y^n(i)}, D_0, D_1)}} \mathcal{P}^n(T_V(y'^n(i)) \cap B_{i,n}(\epsilon')),$$

$$S'_2(n, \epsilon', \delta) = \sum_{i=1}^{M_n} \sum_{\substack{\tilde{\mathcal{P}}: \|\tilde{\mathcal{P}} - \mathcal{P}\| \leq \delta, \\ \exists x^n \in \mathcal{X}^n, P_{x^n} = \tilde{\mathcal{P}}; \\ V \in \mathcal{F}_{\epsilon'}(\tilde{\mathcal{P}}, Q_{y^n(i)}, D_0, D_1)}} \mathcal{P}^n(T_V(y'^n(i)) \cap B_{i,n}(\epsilon')).$$

It is not difficult to verify that there exists $\epsilon'_0 > 0$ such that for any ϵ' with $0 < \epsilon' - \epsilon < \epsilon'_0$, we have:

$$S'_2(n, \epsilon', \delta) \leq 2S_2(n, \epsilon, \delta).$$

Therefore, if we consider codes $\{f_n', g_n'\}$ for all large n (for these n, we always have $\epsilon_n' + \epsilon < \epsilon_0'$), we can still prove that $S_2'(n, \epsilon', \delta)$ is upper bounded by a polynomial of n times

$$\max_{\substack{(V,j,\tilde{\mathcal{P}}):V\in \\ \mathcal{F}_{\epsilon'}(\tilde{\mathcal{P}},\tilde{\mathcal{Q}}_{j,n},D_0,D_1); \\ \tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\le\delta, \\ \exists x^n\in\mathcal{X}^n, P_{x^n}=\tilde{\mathcal{P}}}} \mathcal{P}^n(T_V(\mathscr{B}_{j,n}))$$

as follows:

$$\begin{aligned} S_2'(n, \epsilon', \delta) \le & 2S_2(n, \epsilon, \delta) \\ \le & 2(n+1)^{(|\mathcal{X}|+1)|(\mathcal{Y}|+1)} \mathcal{P}^n(T_{V_n}(\mathscr{B}_{j_n,n})) \\ \le & 2(n+1)^{(|\mathcal{X}|+1)|(\mathcal{Y}|+1)} \max_{\substack{(V,j,\tilde{\mathcal{P}}):V\in \\ \mathcal{F}_{\epsilon'}(\tilde{\mathcal{P}},\tilde{\mathcal{Q}}_{j,n},D_0,D_1); \\ \tilde{\mathcal{P}}:\|\tilde{\mathcal{P}}-\mathcal{P}\|\le\delta, \\ \exists x^n\in\mathcal{X}^n, P_{x^n}=\tilde{\mathcal{P}}}} \mathcal{P}^n(T_V(\mathscr{B}_{j,n})). \end{aligned}$$

This indicates that our assumption is eligible.

Since $V_n \in \mathcal{F}_\epsilon(\tilde{\mathcal{P}}_n, \tilde{\mathcal{Q}}_{j_n,n}, D_0, D_1)$, we can find W_n such that:

$$\tilde{\mathcal{Q}}_{j_n,n} \cdot W_n = \tilde{\mathcal{Q}}_{j_n,n} \cdot V_n = \tilde{P}_n; \tag{5.19}$$

$$W_n \overset{\tilde{\mathcal{Q}}_{j_n,n}}{\ll} V_n; \tag{5.20}$$

$$d(W_n, \tilde{\mathcal{Q}}_{j_n,n}) \ge D_1 + 2\epsilon. \tag{5.21}$$

Note that $V_n, W_n, j_n, \tilde{\mathcal{P}}_n$ all depend on (ϵ, δ) and (f_n, g_n). We omit (ϵ, δ) when there is no ambiguity. Again, we can suppose that $\lim_{n\to\infty} W_n = \tilde{W}^*$. Note that

$$\lim_{n\to\infty} \tilde{P}_n = \tilde{P}^*, \quad \lim_{\delta\to 0} \tilde{P}^*(\epsilon, \delta) = \mathcal{P}. \tag{5.22}$$

In addition, we also assume that δ are sufficiently small such that $\forall \tilde{P}$ s.t. $\|\tilde{P} - P\| \le \delta$, $\forall a \in \mathcal{X}$, $\tilde{P}(a) > 0$.

We first prove that for any $\epsilon > 0$, there exists a $\delta < \epsilon$ such that:

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathcal{P}^n(T_{V_n}(\mathscr{B}_{j_n,n})) < 0. \tag{5.23}$$

In order to prove Inq. (5.23), we first prove that for any $\epsilon > 0$, there exists a $\beta_\epsilon > 0$ (depending on ϵ), such that for any positive sequence $\{\delta_n\}$,

$$\frac{1}{n}\log \mathcal{P}^n(T_{[W_n(\epsilon,\delta_n)]}(\mathscr{B}_{j_n(\epsilon,\delta_n),n})) \leq -\beta_\epsilon. \tag{5.24}$$

We will prove Inq. (5.24) by contradiction. We assume that

$$\mathcal{P}^n(T_{[W_n(\epsilon,\delta_n)]}(\mathscr{B}_{j_n(\epsilon,\delta_n),n})) = 2^{-n\epsilon_n},$$

where $\lim_{n\to\infty}\epsilon_n = 0$. According to Lemma 24, there exist two positive sequences $\{\delta'_n\}$, $\{\eta'_n\}$ where

$$\lim_{n\to\infty}\delta'_n = \lim_{n\to\infty}\eta'_n = 0,$$

and

$$\mathcal{P}^n(\Gamma^{\delta'_n}(T_{[W_n(\epsilon,\delta_n)]}(\mathscr{B}_{j_n(\epsilon,\delta_n),n}))) \geq 1-\eta'_n.$$

It is obvious that for all n sufficiently large $1-\eta'_n > \epsilon$. For any $x^n \in \Gamma^{\delta'_n}(T_{[W_n(\epsilon,\delta_n)]}(\mathscr{B}_{j_n(\epsilon,\delta_n),n}))$, we can find $\tilde{x}^n \in T_{[W_n(\epsilon,\delta_n)]}(\mathscr{B}_{j_n(\epsilon,\delta_n),n})$ such that $d_H(x^n,\tilde{x}^n) \leq \delta'_n$. For sufficiently large n, $x^n_j \neq \hat{x}^n_j$ for at most $n\delta'_n$ coordinates j. Let $D_{\max} = \max_{x\in\mathcal{X},y\in\mathcal{Y}} d(x,y)$. Therefore, for any $y^n \in \hat{\mathcal{X}}^n$,

$$d(x^n,y^n) - d(\tilde{x}^n,y^n) \geq -\delta'_n D_{\max}.$$

Using $\lim_{n\to\infty} r_n/n = 0$, for all large n, we can find a codeword $y^n \in \mathscr{B}_{j_n(\epsilon,\delta_n),n}$, such that $\tilde{x}^n \in T_{[W_n(\epsilon,\delta_n)]}(y^n)$ and

$$d(\tilde{x}^n,y^n) - d(W_n(\epsilon,\delta_n), Q_{j_n(\epsilon,\delta_n),n}) \geq -\frac{\epsilon}{2}.$$

So, $\forall x^n \in \Gamma^{\delta'_n}(T_{[W_n(\epsilon,\delta_n)]}(\mathscr{B}_{j_n(\epsilon,\delta_n),n}))$, by Inq. (5.21), when $\delta'_n \leq \frac{\epsilon}{2D_{\max}}$, we can find a $\tilde{x}^n \in T_{[W_n(\epsilon,\delta_n)]}(\mathscr{B}_{j_n(\epsilon,\delta_n),n})$ and a codeword $y^n \in \mathscr{B}_{j_n(\epsilon,\delta_n),n}$ s.t.

$$\begin{aligned}
&d(x^n,y^n)\\
=&d(W_n(\epsilon,\delta_n), Q_{j_n(\epsilon,\delta_n),n}) + (d(x^n,y^n) - d(\tilde{x}^n,y^n)) + (d(\tilde{x}^n,y^n) - d(W_n(\epsilon,\delta_n), Q_{j_n(\epsilon,\delta_n),n}))\\
\geq&(D_1+2\epsilon) - \epsilon/2 - \epsilon/2\\
=&D_1+\epsilon.
\end{aligned}$$

This means Inq. (2.6) cannot holds for n sufficiently large. Contradiction.

Next, we will prove that for any $\gamma, \epsilon > 0$, when δ is sufficiently small, for all large n (depending on γ, δ, ϵ),

$$\frac{1}{n} \log \frac{\mathcal{P}^n(T_{V_n}(\mathscr{B}_{j_n,n}))}{\mathcal{P}^n(T_{[W_n]}(\mathscr{B}_{j_n,n}))} < \gamma. \tag{5.25}$$

The following lemma is useful in proving Eq. (5.25). Note that $W_n \to \tilde{W}^*$ and $V_n \to \tilde{V}^*$. Therefore, for any $\gamma, \epsilon > 0$, when δ is sufficiently small, for all large n,

$$\frac{1}{n} \log \frac{\mathcal{P}_n^n(T_{[\tilde{W}^*]}(\mathscr{B}_{j_n,n}))}{\mathcal{P}^n(T_{[W_n]}(\mathscr{B}_{j_n,n}))} < \gamma/4, \tag{5.26}$$

and

$$\frac{1}{n} \log \frac{\mathcal{P}_n^n(T_{V_n}(\mathscr{B}_{j_n,n}))}{\mathcal{P}^n(T_{[\tilde{V}^*]}(\mathscr{B}_{j_n,n}))} < \gamma/4. \tag{5.27}$$

By Lemma 2.12 in [11] and the fact that $W_n \to \tilde{W}^*$, for all large n, $\forall y^n \in \mathscr{B}_{j_n,n}$, we have

$$\tilde{W}^{*n}(T_{[\tilde{W}^*]}(\mathscr{B}_{j_n,n})|y^n) \geq \tilde{W}^{*n}(T_{[\tilde{W}^*]}(y^n)|y^n) \geq 1/3.$$

Recall the assumption that for any y^n with $Q_{y^n} = \tilde{Q}_{j_n,n}$, $y^n \in T^n(\tilde{Q}^*)$, we have, $\mathscr{B}_{j_n,n} \subseteq T^n(\tilde{Q}^*)$. Therefore, by Definition 19, we know that for arbitrary fixed δ and ϵ, for all large n (depending on δ, ϵ):

$$\frac{1}{n} \log \frac{G_{\tilde{W}^*,\tilde{Q}^*}(\mathscr{B}_{j_n,n}, 1/3)}{\mathcal{P}_n^n(T_{[\tilde{W}^*]}(\mathscr{B}_{j_n,n}))} < 0. \tag{5.28}$$

The following lemma is proved in Appendix.

Lemma 27 *Let $\{Q_n\}$ be a sequence of distributions on $\mathcal{Y}$ such that*

$$\lim_{n\to\infty} Q_n = Q.$$

Let $\{V_n\}$ $(V_n : \mathcal{Y} \to \mathcal{X})$ *and* $\{W_n\}$ $(W_n : \mathcal{Y} \to \mathcal{Z})$ *be sequences of stochastic matrices such that*

$$W_n \overset{Q_n}{\ll} V_n, \quad \lim_{n\to\infty} V_n = V, \quad \lim_{n\to\infty} W_n = W.$$

Then $W \overset{Q}{\ll} V$.

Via Lemma 27, we know that $\tilde{W}^* \overset{\tilde{Q}^*}{\ll} \tilde{V}^*$. Recall the assumption that for any y^n with $Q_{y^n} = \tilde{Q}_{j_n,n}$, $y^n \in T^n(\tilde{Q}^*)$, we have, $\mathscr{B}_{j_n,n} \subseteq T^n(\tilde{Q}^*)$. Thus for all large n,

$$\frac{1}{n} \log \frac{G_{\tilde{V}^*,\tilde{Q}^*}(\mathscr{B}_{j_n,n}, 1/3)}{G_{\tilde{W}^*,\tilde{Q}^*}(\mathscr{B}_{j_n,n}, 1/3)} < \gamma/4. \tag{5.29}$$

We can also prove that for all large n (depending on γ, δ, ϵ),

$$\frac{1}{n} \log \frac{\mathcal{P}^n(T_{[\tilde{V}^*]}(\mathscr{B}_{j_n,n}))}{G_{\tilde{V}^*,\tilde{Q}^*}(\mathscr{B}_{j_n,n}, 1/3)} < \gamma/4. \tag{5.30}$$

Suppose we can pick at most $\hat{M}_n$ sequences from $\mathscr{B}_{j_n,n}$ and $\hat{M}_n$ pairwise disjoint sets $\subseteq T_{[\tilde{V}^*]}(\mathscr{B}_{j_n,n})$:

$$\{y_1^n, \ldots, y_{\hat{M}_n}^n; \mathcal{A}_1, \ldots, \mathcal{A}_{\hat{M}_n}\}$$

such that for any $i \in [\hat{M}_n]$,

$$\tilde{V}^{*n}(\mathcal{A}_i | y_i^n) > \frac{1}{3}; \quad \mathcal{A}_i \subseteq T_{[V_n]}(\mathscr{B}_{j_n,n}).$$

Here, $\hat{M}_n$ denotes the maximum number of the sequences that we can pick for each n. We can use the following procedure to find an eligible set $\left\{\{y_i^n\}; \{\mathcal{A}_i\}\right\}$:

1. Start with an empty set $\mathcal{G}$.
2. Pick a sequence $y^n \in \mathscr{B}_{j_n,n}$ which has not been processed yet. Suppose at this moment, there are t elements in $\mathcal{G}$:

$$\mathcal{G} = \{y_1^n, \ldots, y_t^n; \mathcal{A}_1, \ldots, \mathcal{A}_t\}.$$

If

$$\tilde{V}^{*n}(T_{[\tilde{V}^*]}(y^n)\setminus \bigcup_{\ell\in[t]} \mathcal{A}_\ell \mid y^n) > \frac{1}{3},$$

then let $y^n_{t+1} = y^n$ and

$$\mathcal{A}_{t+1} = T_{[\tilde{V}^*]}(y^n)\setminus \bigcup_{\ell\in[t]} \mathcal{A}_\ell$$

and add $(y_{t+1}, \mathcal{A}_{t+1})$ to $\mathcal{G}$. Repeat until all sequences in $\mathscr{B}_{j_n,n}$ has been processed.

Again, use Lemma 2.12 in [11], we know that when δ is small, for all large n (depending on γ, δ, ϵ),

$$\frac{1}{n}\log \hat{M}_n \geq \frac{1}{n}\log \mathcal{P}^n(T_{[\tilde{V}^*]}(\mathscr{B}_{j_n,n})) + I(\tilde{V}^*, \tilde{Q}^*) - \gamma/8.$$

For all large n, $\mathscr{B}_{j_n,n} \subseteq T^n(\tilde{Q}^*)$. Applying Theorem 1 Part 2) in [23], we know that for all large n

$$\begin{aligned}
&\frac{1}{n}\log G_{\tilde{V}^*,\tilde{Q}^*}(\mathscr{B}_{j_n,n}, 1/3)\\
\geq&\frac{1}{n}\log G_{\tilde{V}^*,\tilde{Q}^*}(\{y^n_1,\ldots,y^n_{\hat{M}_n}\}, 1/3)\\
\geq&\left(\frac{1}{n}\log \hat{M}_n - I(\tilde{V}^*, \tilde{Q}^*) - \gamma/8\right) - \gamma/8\\
\geq&\frac{1}{n}\log \tilde{\mathcal{P}}^{*n}(T_{[V_n]}(\mathscr{B}_{j_n,n})) - \gamma/4.
\end{aligned}$$

Combining Inq. (5.26) to Inq. (5.30) gives Inq. (5.25).

For any $\epsilon > 0$, let $\gamma_\epsilon = \frac{1}{2}\beta_\epsilon$ where β_ϵ is given in Inq. (5.24). Applying Inq. (5.25), we know that there exists $\delta < \epsilon$ such that

$$\limsup_{n\to\infty}\frac{1}{n}\log\frac{\mathcal{P}^n(T_{V_n}(\mathscr{B}_{j_n,n}))}{\mathcal{P}^n(T_{[W_n]}(\mathscr{B}_{j_n,n}))} \leq \gamma_\epsilon.$$

Then fix (ϵ,δ) and combine Inq. (5.24) and Inq. (5.25), we have:

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\log \mathcal{P}^n(T_{V_n(\epsilon,\delta)}(\mathscr{B}_{j_n(\epsilon,\delta),n}))\\
\leq&-\beta_\epsilon + \limsup_{n\to\infty}\frac{1}{n}\log\frac{\mathcal{P}^n(T_{V_n(\epsilon,\delta)}(\mathscr{B}_{j_n(\epsilon,\delta),n}))}{\mathcal{P}^n(T_{[W_n(\epsilon,\delta)]}(\mathscr{B}_{j_n(\epsilon,\delta),n}))}\\
\leq&-\frac{1}{2}\beta_\epsilon < 0.
\end{aligned}$$

It completes the proof of Inq. (5.23).

By Inq. (5.17), when ϵ, δ are such that Inq. (5.23) holds,

$$\limsup_{n\to\infty} S_2(n, \epsilon, \delta) = 0.$$

Let $\zeta = 1/4$, then when $0 < \epsilon < 1/4$, by Inq. (5.13),

$$\liminf_{n\to\infty} S_1(n, \epsilon, \delta) \geq 1 - \epsilon - \zeta > \frac{1}{2}.$$

By Inq. (5.15), we know:

$$\begin{aligned} R &\geq \min_{(\tilde{\mathcal{P}}, j):\|\tilde{\mathcal{P}}-\mathcal{P}\|\leq\delta} R(\tilde{\mathcal{P}}, \tilde{Q}_{j,n}, D_0 + \epsilon, D_1 + 2\epsilon) - \epsilon \\ &\geq \min_{(\tilde{\mathcal{P}}):\|\tilde{\mathcal{P}}-\mathcal{P}\|\leq\delta} R(\tilde{\mathcal{P}}, D_0 + \epsilon, D_1 + 2\epsilon) - \epsilon. \end{aligned}$$

So by letting $\epsilon \to 0$ and using the definition of $R(\mathcal{P}, D_0, D_1)$, we have

$$R \geq R(\mathcal{P}, D_0, D_1).$$

5.5 Binary Source with Hamming Distortion

5.5.1 Characterization of the Achievable Region

In this section, we consider the communication scenario for a Bernoulli(p) source with Hamming distortion: $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1\}$, $P_X(x = 1) = p$ and $P_X(x = 0) = 1 - p$. Recall that $H_2(p)$ denotes the entropy of a Bernoulli(p) random variable. Without loss of generality, we may assume that $p \leq \frac{1}{2}$. Using the Rate-Distortion Theory for binary source in [10, Theorem 10.3.1], if $R < \max\{H_2(p) - H_2(D_0), 0\}$, then for any $D_1 \in \mathfrak{R}^+$, (R, D_0, D_1) is not achievable. Otherwise, $(R, D_0, D_1) = (R, D_0, 1)$ is always achievable. When $R \geq$

$\max\{H_2(p) - H_2(D_0), 0\}$, we let

$$D_1^*(R, D_0) = \inf_{D_1}\{D_1 : (R, D_0, D_1) \text{ is achievable}\}.$$

The following proposition characterizes the achievable region.

Proposition 28 *For (R, D_0) with $R \geq \max\{H_2(p) - H_2(D_0), 0\}$, $D_1^*(R, D_0)$ satisfies:*

$$D_1^*(R, D_0) \geq 2p - D_0.$$

Moreover, for $D_0 < p$, let $\tilde{Y}$ be a random variable in $\{0, 1\}$ with conditional distribution:

$$P(\tilde{Y} = 0|X = 1) = \frac{D_0}{p}; \quad P(\tilde{Y} = 1|X = 0) = 0.$$

Let V' denote the stochastic matrix describing the conditional distribution of of Y given X. If $R \geq I(\mathcal{P}, V')$, then $D_1^*(R, D_0) = 2p - D_0$.

Proof: We first prove that $D_1^*(R, D_0) \geq 2p - D_0$. Let $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$. Let $\mathcal{P}$ denote the Bernoulli(p) distribution for the source. It is equivalent to show that for any random variable Y with distribution $\mathcal{Q}$ on $\{0, 1\}$ and for any $D_1 < 2p - D_0$, $\mathscr{S}(\mathcal{P}, \mathcal{Q}, D_0, D_1) = \emptyset$. Suppose Y is jointly distributed with X w.r.t. V such that $\mathcal{Q} \cdot V = \mathcal{P}$ and $d(V, \mathcal{Q}) \leq D_0$. Let Z be a random variable jointly distributed with X, Y and satisfies: $Y \to X \to Z$. The stochastic matrix T describing the conditional distribution of Z given X is as follows:

$$\begin{aligned} T(0|1) &= 1, & T(1|1) &= 0; \\ T(1|0) &= \frac{p}{1-p}, & T(0|0) &= \frac{1-2p}{1-p}. \end{aligned}$$

Then the stochastic matrix describing the conditional distribution of of Z given Y is $W = V \cdot T$. Here, W satisfies: $\mathcal{Q} \cdot W = (\mathcal{Q} \cdot V) \cdot T = \mathcal{P} \cdot T = \mathcal{P}$. By the Remark of Proposition 2 in [22], we know that $W \overset{\mathcal{Q}}{\ll} V$. We can also prove that $d(W, \mathcal{Q}) \geq 2p - D_0$.

Define $f : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \mapsto \Re$ as $f(X,Y,Z) = d_H(X,Y) + d_H(Y,Z)$. By the definition of $d_H(\cdot,\cdot)$, $f(X,Y,Z) \geq \mathbf{1}_{\{X \neq Z\}}$, where $\mathbf{1}_{\{\cdot\}}$ is the indication function. Thus,

$$\begin{aligned} & E_{XY}[d_H(X,Y)] + E_{YZ}[d_H(Z,Y)] \\ =& E_{XYZ}[f(X,Y,X)] \geq E_{XYZ}[\mathbf{1}_{\{X \neq Z\}}] \\ =& E_{XZ}[\mathbf{1}_{\{X \neq Z\}}] = 2p. \end{aligned}$$

Combining $E_{XY}[d_H(X,Y)] = d(V,Q) \leq D_0$, we have:

$$d(W,Q) = E_{YZ}[d_H(Z,Y)] \geq 2p - D_0.$$

So, when $D_1 < 2p - D_0$, $V \notin \mathscr{S}(\mathcal{P}, Q, D_0, D_1)$.

For $D_0 < p$ and $\tilde{Y}$, let $P_{\tilde{Y}X}$ denote the joint probability distribution. By computation, we know that $P(\tilde{Y} = 1) = p - D_0$ and $P(\tilde{Y} = 0) = 1 - p + D_0$. It is not difficult to verify that for any $Z \overset{d}{=} X$, $E[d_H(Z,Y)] \leq 2p - D_0$. Applying Theorem 8, we know that $(R, D_0, 2p - D_0)$ is achievable.

APPENDIX A

PROOFS

A.1 Proofs for Binary VPEC

A.1.1 Proof of Lemma 1

proof: We find the required $\alpha_1, \ldots, \alpha_T$ by induction. Clearly there exists a positive integer α_1 such that

$$A_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ \alpha_1^1 & \alpha_1^2 & \cdots & \alpha_1^{N-T} \end{bmatrix},$$

is such that every $(N-T) \times (N-T)$ submatrix is nonsingular. Indeed, taking $\alpha_1 = 1$ suffices. Now suppose we have positive integers $\alpha_1, \ldots \alpha_{t-1}$ such that every $(N-T) \times (N-T)$ submatrix of

$$A_{t-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ \alpha_1^1 & \alpha_1^2 & \cdots & \alpha_1^{N-T} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{t-1}^1 & \alpha_{t-1}^2 & \cdots & \alpha_{t-1}^{N-T} \end{bmatrix}$$

is nonsingular. Consider the matrix

$$A_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ \alpha_1^1 & \alpha_1^2 & \cdots & \alpha_1^{N-T} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_t^1 & \alpha_t^2 & \cdots & \alpha_t^{N-T} \end{bmatrix},$$

viewed as a function of the variable α_t. For any given $(N-T)\times(N-T)$ submatrix of A_t of the form

$$\begin{bmatrix} & \tilde{A} & \\ \alpha_t^1 & \alpha_t^2 & \cdots & \alpha_t^{N-T} \end{bmatrix}, \tag{A.1}$$

there must exist a natural number α_t such that this particular $(N-T)\times(N-T)$ matrix is nonsingular, by the following reasoning. The rows of $\tilde{A}$ are linearly independent by the induction hypothesis. Let $[v_1 \;\; v_2 \;\; \cdots v_{N-T}]$ be a nonzero row vector such that

$$\begin{bmatrix} & \tilde{A} & \\ v_1 & v_2 & \cdots & v_{N-T} \end{bmatrix}, \tag{A.2}$$

is full rank. Then let $[\tilde{v}_1 \;\; \tilde{v}_2 \;\; \cdots \tilde{v}_{N-T}]$ denote the component of $[v_1 \;\; v_2 \;\; \cdots v_{N-T}]$ that is orthogonal to the row space of $\tilde{A}$ and note that $[\tilde{v}_1 \;\; \tilde{v}_2 \;\; \cdots \tilde{v}_{N-T}]$ must be nonzero. Then we can find a natural number α_t so that

$$\sum_{i=1}^{N-T} \tilde{v}_i \alpha_t^i \neq 0.$$

This follows from the fact that the left-hand side is a nonzero $(N-T)$-degree polynomial in α_t, so that there must be a positive integer that is not a root. We conclude that the determinant of the $(N-T)\times(N-T)$ matrix in (A.1), which is evidently an $(N-T)$-degree polynomial in α_t, is not identically zero.

Next we show that there is one choice of α_t that ensures that every $(N-T)\times(N-T)$ submatrix of A_t is nonsingular. The determinant of any given $(N-T)\times(N-T)$ submatrix is a nonzero $(N-T)$-degree polynominal in α_t, as noted earlier. Thus it has at most $(N-T)$ roots according to fundamental theorem of algebra. Thus all of the submatrices together have at most $\binom{N-T+t-1}{N-T-1}(N-T)$ roots. Since this is finite, there must exist a natural number α_t that is not a root of any of these polynomials.

A.1.2 Proof of Lemma 29

Lemma 29 *For any integer $m > 1$ and $\Lambda \in \mathbb{Z}^{(m-1)\times m}$, there exists a non-zero vector $x^m \in \mathbb{Z}^m$ such that $\Lambda x^m = 0$. Furthermore, if* $\text{rank}(\Lambda') = m-1$ *for all $(m-1)$-by-$(m-1)$ submatrices Λ' of Λ, then any such an x^m must be in $(\mathbb{Z}\backslash\{0\})^m$.*

Proof: Let $\lambda_1^m, ..., \lambda_{m-1}^m$ denote the rows of Λ. Using the Gram-Schmidt procedure, we may assume that $\lambda_1^m, ..., \lambda_{m-1}^m$ are orthogonal. Since $\lambda_1^m, ..., \lambda_{m-1}^m$ cannot span $\mathbb{R}^m$ but $\mathbb{N}^m$ does, there must exist a vector $\lambda^m \in \mathbb{N}^m$ that is not in the span of $\lambda_1, ..., \lambda_{m-1}$. Then the vector:

$$\lambda^m - \sum_{i=1}^{m-1} \frac{(\lambda_i^m)^T \lambda^m}{(\lambda_i^m)^T \lambda_i^m} \lambda_i^m,$$

where the sum excludes those i for which λ_i^m is the zero vector, is in $\mathbb{Q}^m$ and is orthogonal to $\lambda_1^m, ..., \lambda_{m-1}^m$. Multiplying λ^m by the least common denominator gives a non-zero integer solution to $\Lambda x^m = 0$.

When $\text{rank}(\Lambda') = m-1$ for all Λ', we prove that all the entries of x^m must be non-zero

by contradiction. Without loss of generality, suppose that $x_1 = 0$. Then

$$\begin{bmatrix} \Lambda_2 & \cdots & \Lambda_m \end{bmatrix} \begin{bmatrix} x_2 \\ \vdots \\ x_m \end{bmatrix} = 0,$$

where Λ_2 through Λ_m are the second through last columns of Λ. Now $\begin{bmatrix} \Lambda_2 & \cdots & \Lambda_m \end{bmatrix}$ is a non-singular matrix by hypothesis. The above linear system then has a unique solution, namely the zero vector. This implies that x^m is the zero vector, which is a contradiction.

A.1.3 Proof of Theorem 3

Proof: We begin by showing the conclusion for some N_0 and for all sufficiently large L_0.

Write the V-matrix as:

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ a_{1,1} & a_{1,2} & \cdots & a_{1,N-T} \\ \vdots & \vdots & \vdots & \vdots \\ a_{T,1} & a_{T,2} & \cdots & a_{T,N-T} \end{bmatrix}.$$

Observe that $\left\lfloor \frac{T}{2} \right\rfloor \left\lceil \frac{T}{2} \right\rceil = N - T - 1$. For $i \in \{0, ..., \left\lceil \frac{T}{2} \right\rceil - 1\}$, let μ_i denote a length-$\left\lfloor \frac{T}{2} \right\rfloor$ integer vector in the right null-space of the $(\left\lfloor \frac{T}{2} \right\rfloor - 1)$-by $\left\lfloor \frac{T}{2} \right\rfloor$ matrix

$$\begin{bmatrix} a_{1,i\lfloor \frac{T}{2} \rfloor + 1} & \cdots & a_{1,(i+1)\lfloor \frac{T}{2} \rfloor} \\ \vdots & \ddots & \vdots \\ a_{\lfloor \frac{T}{2} \rfloor - 1, i\lfloor \frac{T}{2} \rfloor + 1} & \cdots & a_{\lfloor \frac{T}{2} \rfloor - 1,(i+1)\lfloor \frac{T}{2} \rfloor} \end{bmatrix}. \tag{A.3}$$

Such a vector exists by Lemma 29 in the appendix (if $T = 2$, then set $\mu_0 = 1$). Since A is a V-matrix, all $(\lfloor \frac{T}{2} \rfloor - 1)$-by-$(\lfloor \frac{T}{2} \rfloor - 1)$ submatrices of the matrix in (A.3) have rank $\left\lfloor \frac{T}{2} \right\rfloor - 1$ (see Lemma 30 in the appendix). Let $\mu_{i,j}$ refer to the jth entry of the column vector μ_i. Then $\mu_{i,j}$ is non-zero for all i and j by Lemma 29. For $i \in \{0, ..., \left\lceil \frac{T}{2} \right\rceil - 1\}$, let $\nu_i \in \mathbb{N}^{\lfloor \frac{T}{2} \rfloor}$ be chosen so that the components of $\nu_i + \mu_i$ are all positive, then let

$$c_i = \begin{bmatrix} \nu_i & \nu_i + \mu_i \end{bmatrix}$$

be an $\mathbb{N}^{\lfloor \frac{T}{2} \rfloor \times 2}$ matrix. Let $\mu_{\lceil \frac{T}{2} \rceil} \in \mathbb{Z}$ be a natural number whose value will be chosen later, and let $\nu_{\lceil \frac{T}{2} \rceil} = 1$. Let

$$c_{\lceil \frac{T}{2} \rceil} = \begin{bmatrix} \nu_{\lceil \frac{T}{2} \rceil} & \nu_{\lceil \frac{T}{2} \rceil} + \mu_{\lceil \frac{T}{2} \rceil} \end{bmatrix}$$

be an $\mathbb{N}^{1 \times 2}$ matrix.

From c_i define the matrices

$$c_i^+ = \begin{bmatrix} \nu_i + \frac{\mu_i}{2} & \nu_i + \frac{\mu_i}{2} \end{bmatrix},$$

and

$$c_i^- = \begin{bmatrix} -\frac{\mu_i}{2} & \frac{\mu_i}{2} \end{bmatrix}.$$

Now let H denote an L-by-L Hadamard matrix for some L satisfying

$$L \geq \left\lceil \frac{T}{2} \right\rceil + 1,$$

which exists by Sylvester's construction [38]. Each element of H is -1 or 1, and the rows are orthogonal. We use H to construct an $(N - T)$-by-$2L$ matrix X according to (A.4).

$$X = \begin{bmatrix} c_0^+ + c_0^- H_{1,1} & c_0^+ + c_0^- H_{1,2} & \cdots & c_0^+ + c_0^- H_{1,L} \\ c_1^+ + c_1^- H_{2,1} & c_1^+ + c_1^- H_{2,2} & \cdots & c_1^+ + c_1^- H_{2,L} \\ \vdots & \vdots & \ddots & \vdots \\ c_{\lceil \frac{T}{2} \rceil - 1}^+ + c_{\lceil \frac{T}{2} \rceil - 1}^- H_{\lceil \frac{T}{2} \rceil, 1} & c_{\lceil \frac{T}{2} \rceil - 1}^+ + c_{\lceil \frac{T}{2} \rceil - 1}^- H_{\lceil \frac{T}{2} \rceil, 2} & \cdots & c_{\lceil \frac{T}{2} \rceil - 1}^+ + c_{\lceil \frac{T}{2} \rceil - 1}^- H_{\lceil \frac{T}{2} \rceil, L} \\ c_{\lceil \frac{T}{2} \rceil}^+ + c_{\lceil \frac{T}{2} \rceil}^- H_{\lceil \frac{T}{2} \rceil + 1, 1} & c_{\lceil \frac{T}{2} \rceil}^+ + c_{\lceil \frac{T}{2} \rceil}^- H_{\lceil \frac{T}{2} \rceil + 1, 2} & \cdots & c_{\lceil \frac{T}{2} \rceil}^+ + c_{\lceil \frac{T}{2} \rceil}^- H_{\lceil \frac{T}{2} \rceil + 1, L} \end{bmatrix} \tag{A.4}$$

Note that for any $i \in \{0, ..., \left\lceil \frac{T}{2} \right\rceil\}$, if $H_{i+1,j} = 1$,

$$c_i^+ + c_i^- H_{i+1,j} = \begin{bmatrix} \nu_i & \nu_i + \mu_i \end{bmatrix},$$

and if $H_{i+1,j} = -1$,

$$c_i^+ + c_i^- H_{i+1,j} = \begin{bmatrix} \nu_i + \mu_i & \nu_i \end{bmatrix}.$$

Evidently, the rows of X can be divided into $\left\lceil \frac{T}{2} \right\rceil + 1$ blocks, the first $\left\lceil \frac{T}{2} \right\rceil$ blocks consisting of $\left\lfloor \frac{T}{2} \right\rfloor$ rows and the last block consisting of a single row. For $i \in \{0, ..., \left\lceil \frac{T}{2} \right\rceil\}$, we define a modified version of X, X_i, obtained by replacing the ith row block in X with

$$\begin{bmatrix} c_i^+ + c_i^-(-H_{i+1,1}) & \cdots & c_i^+ + c_i^-(-H_{i+1,L}) \end{bmatrix}.$$

Note that this has the effect of replacing $\begin{bmatrix} \nu_i & \nu_i + \mu_i \end{bmatrix}$ with $\begin{bmatrix} \nu_i + \mu_i & \nu_i \end{bmatrix}$ and vice versa. We view X and the various X_i as different source realizations with blocklength $(N - T)N_0K_0$ where $N_0 = 2L$ and K_0 is any integer satisfying

$$\log_K K_0 \geq \max_{i,j} \mu_{i,j} + \nu_{i,j}.$$

Since H is Hadamard, the inner product between any two rows of X must equal the inner product between the corresponding rows of X_i for all i. Thus, all of these source realizations will result in the same norms and inner products being sent as part of the polytope code. Let $\{F_{j_1 j_2}\}$ denote these norms and inner products.

Next we construct codewords from these source realizations. Let

$$\bar{X} = AX$$

and for $i \in \{0, ..., \left\lceil \frac{T}{2} \right\rceil\}$, let

$$\bar{X}_i = AX_i.$$

Observe that since μ_i is in the null space of the matrix in (A.3), rows

$$\left\{ N - T + 1, ..., N - T + \left\lfloor \frac{T}{2} \right\rfloor - 1 \right\}$$

of $\bar{X}$ and $\bar{X}_i$ will be the same for all $i \in \{0, ..., \left\lceil \frac{T}{2} \right\rceil - 1\}$.

Finally, construct a set of received packets as follows. Packets 1 through $N - T$ are the first $N - T$ rows of $\bar{X}$, respectively. Packets

$$\left\{ N - T + 1, ..., N - T + \left\lfloor \frac{T}{2} \right\rfloor - 1 \right\}$$

are set to be rows $\{N - T + 1, ..., N - T + \left\lfloor \frac{T}{2} \right\rfloor - 1\}$ of any of the $\bar{X}_i$, $i \in \{0, ..., \left\lceil \frac{T}{2} \right\rceil - 1\}$ (recall that these rows coincide across $\bar{X}$ and these $\bar{X}_i$). For

$$i \in \left\{ N - T + \left\lfloor \frac{T}{2} \right\rfloor, ..., N \right\},$$

received packet i is set to the corresponding row of $\bar{X}_{i-(N-T+\lfloor \frac{T}{2} \rfloor)}$. Define the matrix $\bar{Y}$ to be the set of received packets, one per row, starting with the first.

Now the number of packets that differ between $\bar{Y}$ and $\bar{X}_i$ is at most

$$\left\lfloor \frac{T}{2} \right\rfloor + \left\lceil \frac{T}{2} \right\rceil = T$$

if $i \in \{0, ..., \left\lceil \frac{T}{2} \right\rceil - 1\}$. Likewise, codeword $\bar{X}_{\lceil \frac{T}{2} \rceil}$ differs from $\bar{Y}$ in at most

$$1 + \left(\left\lfloor \frac{T}{2} \right\rfloor - 1 \right) + \left\lceil \frac{T}{2} \right\rceil = T.$$

Thus, $\bar{X}_i$, $i \in \{0, ..., \left\lceil \frac{T}{2} \right\rceil\}$ is in PTC($\bar{Y}, \{F_{j_1 j_2}\}$). For each $i \in \{1, ..., N-T\}$, there exists i_1 and i_2 s.t. row i in $\bar{X}_{i_1}$ and $\bar{X}_{i_2}$ disagree. Moreover, we can pick $\mu_{\lceil \frac{T}{2} \rceil}$ such that for each $i \in \{N-T+1, ..., N\}$, row i in $\bar{X}_1$ and $\bar{X}_{\lceil \frac{T}{2} \rceil}$ disagree. This is because for each $i \in \{N-T+1, ..., N\}$, there is at most one value for $\mu_{\lceil \frac{T}{2} \rceil}$ such that row i in $\bar{X}_1$ and $\bar{X}_{\lceil \frac{T}{2} \rceil}$ are the same. Thus the set of integers for which $\mu_{\lceil \frac{T}{2} \rceil}$ does not satisfy the desired condition has at most T elements, and we can choose $\mu_{\lceil \frac{T}{2} \rceil}$ to be any positive integer not in this set.

This establishes the conclusion for $N_0 = 2L$ and all sufficiently large K_0. One can accommodate larger values of N_0 by prepending a vector of ones to each of the X_i source realizations.

A.1.4 Lemma 30 and its Proof

Lemma 30 *Let $\alpha_1, \ldots \alpha_m$ be distinct natural numbers. Then for any integer $k \geq 0$, every m-by-m submatrix of*

$$M = \begin{bmatrix} \alpha_1^k & \alpha_1^{k+1} & \cdots & \alpha_1^{k+m} \\ \alpha_2^k & \alpha_2^{k+1} & \cdots & \alpha_2^{k+m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_m^k & \alpha_m^{k+1} & \cdots & \alpha_m^{k+m} \end{bmatrix}.$$

is nonsingular.

Proof: Let $a = [a_0 \;\; a_1 \;\; \cdots \;\; a_m]^T$ be such that $Ma = 0$ and $a_i = 0$ for some i. It suffices

to show that a must be the zero vector. Now a is in the nullspace of

$$M = \begin{bmatrix} 1 & \alpha_1^1 & \cdots & \alpha_1^m \\ 1 & \alpha_2^1 & \cdots & \alpha_2^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m^1 & \cdots & \alpha_m^m \end{bmatrix}.$$

Consider the polynomial

$$P(x) = \sum_{i=0}^{m} a_i x^i.$$

Evidently P is a degree-m polynomial with roots $\alpha_1, \ldots, \alpha_m$. There is a unique nonzero degree-m polynomial with these roots, however, namely,

$$P'(x) = \prod_{i=0}^{m} (x - \alpha_i) = \sum_{i=0}^{m} a_i' x^i.$$

Since all of the α_i are positive, all of the a_i' must be nonzero. It follows that $P(\cdot) \neq P'(\cdot)$ and so $P(\cdot)$ must be the all-zero polynomial.

A.2 Proofs for Gaussian VPEC

A.2.1 Proof of Proposition 12

It is equivalent to prove the following. Fix $\sigma, \rho, R_1, R_2, R_3, \delta, \epsilon$. If

$$R_\ell > I(X_{Q_\delta}; X_{\ell,Q_\delta}), \quad \forall 1 \leq \ell \leq 3;$$

$$R_{\ell_1} + R_{\ell_2} > H(X_{\ell_1,Q_\delta}) + H(X_{\ell_2,Q_\delta}) H(X_{\ell_1,Q_\delta}, X_{\ell_2,Q_\delta} \mid X_{Q_\delta}), \quad \forall 1 \leq \ell_1 < \ell_2 \leq 3;$$

$$R_1 + R_2 + R_3 > \sum_{l=1}^{3} H(X_{l,Q_\delta}) - H(X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta} | X_{Q_\delta}),$$

then

$$\lim_{n\to\infty} E_C[P(E_0(C))] = \lim_{n\to\infty} E_C[P(E_1(C))] + \lim_{n\to\infty} E_C[P(E_2(C) \cap E_1(C)^c)] = 0.$$

$\lim_{n\to\infty} E_C[P(E_1(C)] = 0$ is straightforward. It remains to prove that

$$\lim_{n\to\infty} E_C[P(E_2(C) \cap E_1(C)^c)] = 0.$$

The following proof is similar to the proof in [12]. For each $x^n_{Q_\delta} \in T^{(n,\epsilon)}(X^n_{Q_\delta})$, define

$$\begin{aligned} C_{x^n_{Q_\delta}} :=&\{(i,j,k) \in [2^{nR_1}] \times [2^{nR_2}] \times [2^{nR_3}] \mid (x^n_{Q_\delta}, x^n_{1,Q_\delta}(i), \\ & x^n_{2,Q_\delta}(j), x^n_{3,Q_\delta}(k)) \in T^{(n,\epsilon)}(X_{Q_\delta}, X^n_{1,Q_\delta}, X^n_{2,Q_\delta}, X^n_{3,Q_\delta})\}. \end{aligned}$$

Here, $C_{x^n_{Q_\delta}}$ and $|C_{x^n_{Q_\delta}}|$ can be viewed as a function of the random codebook C. Then, by changing the order of expectation, we have

$$\begin{aligned} E_C[P(E_2(C) \cap E_1(C)^c)] =& E_{X^n_{Q_\delta}}\left[P\{|C_{X^n_{Q_\delta}}| = 0\} \cdot \mathbf{1}_{\{X^n_{Q_\delta} \in T^{(n,\epsilon)}(X_{Q_\delta})\}}\right] \\ \leq & \max_{x^n_{Q_\delta} \in T^{(n,\epsilon)}(X_{Q_\delta})} P\{|C_{x^n_{Q_\delta}}| = 0\}. \end{aligned}$$

For any $x^n_{Q_\delta} \in T^{(n,\epsilon)}(X_{Q_\delta})$ and $0 < \alpha < 1$, we have

$$P\{|C_{x^n_{Q_\delta}}| = 0\} \leq P\{|C_{x^n_{Q_\delta}}| - E|C_{x^n_{Q_\delta}}| \leq -\alpha E|C_{x^n_{Q_\delta}}|\} \leq \mathrm{var}|C_{x^n_{Q_\delta}}|/(\alpha E|C_{x^n_{Q_\delta}}|)^2.$$

Write $|C_{x^n_{Q_\delta}}|$ as the sum of indicator functions:

$$|C_{x^n_{Q_\delta}}| = \sum_{k=1}^{2^{nR_3}} \sum_{j=1}^{2^{nR_2}} \sum_{i=1}^{2^{nR_1}} \mathbf{1}_{\{(x^n_{Q_\delta}, x^n_{1,Q_\delta}(i), x^n_{2,Q_\delta}(j), x^n_{3,Q_\delta}(k)) \in T^{(\epsilon,n)}(X_{Q_\delta}, X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta})\}}.$$

For convenience, for $\ell = 0, 1, 2$, we define

$$S_\ell = \{(i,j,k,i',j',k') \in \mathbb{N}^6 | \ell \text{ of the three inequalities } i = i', j = j', k = k' \text{ holds.}\},$$

where $\mathbb{N}$ denotes the set of natural numbers. We can compute $E\left[|C_{x^n_{Q_\delta}}|^2\right]$ and $\left(E|C_{x^n_{Q_\delta}}|\right)^2$

(see the next page).

$$
\begin{aligned}
E\left[|C_{x^n_{Q_\delta}}|^2\right] &= \sum_{k=1}^{2^{nR_3}}\sum_{j=1}^{2^{nR_2}}\sum_{i=1}^{2^{nR_1}} E\mathbf{1}_{\{(x^n_{Q_\delta},x^n_{1,Q_\delta}(i),x^n_{2,Q_\delta}(j),x^n_{3,Q_\delta}(k))\in T^{(\epsilon,n)}(X_{Q_\delta},X_{1,Q_\delta},X_{2,Q_\delta},X_{3,Q_\delta})\}} \\
&\quad + \sum_{\ell=1}^{2} \sum_{\substack{(i,j,k),(i',j',k')\in \\ [2^{nR_1}]\times[2^{nR_2}]\times[2^{nR_3}] \\ (i,j,k,i',j',k')\in S_\ell}} E\left(\mathbf{1}_{\{(x^n_{Q_\delta},x^n_{1,Q_\delta}(i),x^n_{2,Q_\delta}(j),x^n_{3,Q_\delta}(k))\in T^{(\epsilon,n)}(X_{Q_\delta},X_{1,Q_\delta},X_{2,Q_\delta},X_{3,Q_\delta})\}}\right. \\
&\quad \left. \cdot\ \mathbf{1}_{\{(x^n_{Q_\delta},x^n_{1,Q_\delta}(i'),x^n_{2,Q_\delta}(j'),x^n_{3,Q_\delta}(k'))\in T^{(\epsilon,n)}(X_{Q_\delta},X_{1,Q_\delta},X_{2,Q_\delta},X_{3,Q_\delta})\}}\right) \\
&\quad + \sum_{\substack{(i,j,k),(i',j',k')\in \\ [2^{nR_1}]\times[2^{nR_2}]\times[2^{nR_3}] \\ (i,j,k,i',j',k')\in S_0}} \left(E\mathbf{1}_{\{(x^n_{Q_\delta},x^n_{1,Q_\delta}(i),x^n_{2,Q_\delta}(j),x^n_{3,Q_\delta}(k))\in T^{(\epsilon,n)}(X_{Q_\delta},X_{1,Q_\delta},X_{2,Q_\delta},X_{3,Q_\delta})\}}\right) \\
&\quad \cdot \left(E\mathbf{1}_{\{(\mathbf{x^n_{Q_\delta}},\mathbf{x^n_{1,Q_\delta}}(\mathbf{i'}),\mathbf{x^n_{2,Q_\delta}}(\mathbf{j'}),\mathbf{x^n_{3,Q_\delta}}(\mathbf{k'}))\in \mathbf{T}^{(\epsilon,\mathbf{n})}(\mathbf{X_{Q_\delta}},\mathbf{X_{1,Q_\delta}},\mathbf{X_{2,Q_\delta}},\mathbf{X_{3,Q_\delta}})\}}\right). \\
\left(E|C_{x^n_{Q_\delta}}|\right)^2 &= \left(\sum_{k=1}^{2^{nR_3}}\sum_{j=1}^{2^{nR_2}}\sum_{i=1}^{2^{nR_1}} E\mathbf{1}_{\{(x^n_{Q_\delta},x^n_{1,Q_\delta}(i),x^n_{2,Q_\delta}(j),x^n_{3,Q_\delta}(k))\in T^{(\epsilon,n)}(X_{Q_\delta},X_{1,Q_\delta},X_{2,Q_\delta},X_{3,Q_\delta})\}}\right)^2 \\
&> \sum_{\substack{(i,j,k),(i',j',k')\in \\ [2^{nR_1}]\times[2^{nR_2}]\times[2^{nR_3}] \\ (i,j,k,i',j',k')\in S_0}} \left(E\mathbf{1}_{\{(x^n_{Q_\delta},x^n_{1,Q_\delta}(i),x^n_{2,Q_\delta}(j),x^n_{3,Q_\delta}(k))\in T^{(\epsilon,n)}(X_{Q_\delta},X_{1,Q_\delta},X_{2,Q_\delta},X_{3,Q_\delta})\}}\right) \\
&\quad \cdot \left(E\mathbf{1}_{\{(\mathbf{x^n_{Q_\delta}},\mathbf{x^n_{1,Q_\delta}}(\mathbf{i'}),\mathbf{x^n_{2,Q_\delta}}(\mathbf{j'}),\mathbf{x^n_{3,Q_\delta}}(\mathbf{k'}))\in \mathbf{T}^{(\epsilon,\mathbf{n})}(\mathbf{X_{Q_\delta}},\mathbf{X_{1,Q_\delta}},\mathbf{X_{2,Q_\delta}},\mathbf{X_{3,Q_\delta}})\}}\right).
\end{aligned}
$$

For $(i, j, k, i', j', k') = (i, j, k, i, j, k') \in S_2$, if

$$R_1 + R_2 > H(X_{1,Q_\delta}) + H(X_{2,Q_\delta}) - H(X_{1,Q_\delta}, X_{2,Q_\delta}|X_{Q_\delta}),$$

then we get Inq. A.5.

$$
\begin{aligned}
&\lim_{n\to\infty}\left\{\frac{1}{\left(E|C_{x^n_{Q_\delta}}|\right)^2}\cdot \sum_{\substack{(i,j,k),(i',j',k')\in \\ [2^{nR_1}]\times[2^{nR_2}]\times[2^{nR_3}] \\ i=i',j=j',k\neq k'}} E\left(\mathbf{1}_{\{(x^n_{Q_\delta},x^n_{1,Q_\delta}(i),x^n_{2,Q_\delta}(j),x^n_{3,Q_\delta}(k))\in T^{(\epsilon,n)}(X_{Q_\delta},X_{1,Q_\delta},X_{2,Q_\delta},X_{3,Q_\delta})\}}\right.\right. \\
&\left.\left. \cdot\ \mathbf{1}_{\{(x^n_{Q_\delta},x^n_{1,Q_\delta}(i'),x^n_{2,Q_\delta}(j'),x^n_{3,Q_\delta}(k'))\in T^{(\epsilon,n)}(X_{Q_\delta},X_{1,Q_\delta},X_{2,Q_\delta},X_{3,Q_\delta})\}}\right)\right\} \\
&\le \lim_{n\to\infty} 2^{-n(R_1+R_2-H(X_{1,Q_\delta})-H(X_{2,Q_\delta})+H(X_{1,Q_\delta},X_{2,Q_\delta}|X_{Q_\delta})-o(1))} = 0. \qquad \text{(A.5)}
\end{aligned}
$$

For $(i,j,k,i',j',k') = (i,j,k,i,j',k') \in S_1$, if $R_1 > I(X_{Q_\delta}; X_{1,Q_\delta})$, then we get Inq. A.6.

$$\lim_{n\to\infty} \left\{ \frac{1}{\left(E|C_{x^n_{Q_\delta}}|\right)^2} \cdot \sum_{\substack{(i,j,k),(i',j',k')\in \\ [2^{nR_1}]\times[2^{nR_2}]\times[2^{nR_3}] \\ i=i', j\neq j', k\neq k'}} E\left(\mathbf{1}_{\{(x^n_{Q_\delta}, x^n_{1,Q_\delta}(i), x^n_{2,Q_\delta}(j), x^n_{3,Q_\delta}(k))\in T^{(\epsilon,n)}(X_{Q_\delta}, X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta})\}} \right.\right.$$

$$\left.\left. \cdot\ \mathbf{1}_{\{(x^n_{Q_\delta}, x^n_{1,Q_\delta}(i'), x^n_{2,Q_\delta}(j'), x^n_{3,Q_\delta}(k'))\in T^{(\epsilon,n)}(X_{Q_\delta}, X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta})\}} \right) \right\}$$

$$\leq \lim_{n\to\infty} 2^{-n(R_1 - I(X_{Q_\delta}; X_{1,Q_\delta}) - o(1))} = 0. \tag{A.6}$$

Thus, when

$$R_\ell > I(X_{Q_\delta}; X_{i,Q_\delta}), \qquad \forall \ell \in [3];$$

$$R_{\ell_1} + R_{\ell_2} > H(X_{\ell_1,Q_\delta}) + H(X_{\ell_2,Q_\delta}) \quad - H(X_{\ell_1,Q_\delta}, X_{\ell_2,Q_\delta}|X_{Q_\delta}), \qquad \forall 1 \leq \ell_1 < \ell_2 \leq 3,$$

we have

$$\begin{aligned} &\lim_{n\to\infty} P\{|C_{x^n_{Q_\delta}}| = 0\} \leq \mathrm{var}|C_{x^n_{Q_\delta}}|/(\alpha E|C_{x^n_{Q_\delta}}|)^2 \\ &\leq \frac{1}{\alpha^2} \cdot 2^{-n(R_1+R_2+R_3+H(X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta}|X_{Q_\delta})+o(1))} \cdot 2^{n\sum_{l=1}^3 H(X_{l,Q_\delta})} \end{aligned} \tag{A.7}$$

If $R_1 + R_2 + R_3 > \sum_{l=1}^3 H(X_{l,Q_\delta}) - H(X_{1,Q_\delta}, X_{2,Q_\delta}, X_{3,Q_\delta}|X_{Q_\delta})$ also holds, then for any $x^n_{Q_\delta}$,

$$\lim_{n\to\infty} P\{|C_{x^n_{Q_\delta}}| = 0\} = 0,$$

and it completes the proof:

$$0 \leq \lim_{n\to\infty} E_C[P(E_2(C) \cap E_1(C)^c)] \leq \lim_{n\to\infty} \left\{ \max_{x^n_{Q_\delta} \in T^{(n,\epsilon)}(X_{Q_\delta})} P\{|C_{x^n_{Q_\delta}}| = 0\} \right\} = 0.$$

A.2.2 Proof of Lemma 16

Let H_1 and H_2 be two sets $\subset [N]$ such that $[N] = H_1 \cup H_2$, $|H_1| = |H_2| = N - T$. Let $A = H_1 \cap H_2$. For any $D > 0$, define $Q_D :=$

$$\left\{ (X^n, W^n) : \max_{i=1,2} \max_{C_{H_i^c}} \frac{1}{n} \sum_{\ell=1}^n d(X_\ell + W_\ell, \hat{X}_\ell) \leq D \right\}.$$

Here X_ℓ denote the ℓth component of X^n. For every codeword c_A, let

$$Q_{D,c_A} = \{X^n + W^n | (X^n, W^n) \in Q_D, f_A(X^n) = c_A\}.$$

We have:

$$D_T \geq E\left[\max_{H\subset[N],|H|=N-T} \max_{C_H^c} \frac{1}{n}\sum_{\ell=1}^{n} d(X_\ell, \hat{X}_\ell)\right].$$

We also have $\frac{1}{n}\|X^n + W^n - \hat{X}^n\|_2^2 \leq \frac{2}{n}\|X^n - \hat{X}^n\|_2^2 + \frac{2}{n}\|W^n\|_2^2$, i.e.,

$$d(X^n + W^n, \hat{X}^n) \leq 2d(X^n, \hat{X}^n) + \frac{2}{n}\|W^n\|_2^2.$$

Combining the above two, we have:

$$\begin{aligned}
&2D_T + 2\sigma^2 \\
\geq\ & E\left[\max_{H\subset[N],|H|=N-T} \max_{C_H^c} \frac{1}{n}\sum_{\ell=1}^{n} d(X_\ell + W_\ell, \hat{X}_\ell)\right] \\
\geq\ & E\left[\max_{i=1,2} \max_{C_{H_i}^c} \frac{1}{n}\sum_{\ell=1}^{n} d(X_\ell + W_\ell, \hat{X}_\ell)\right].
\end{aligned}$$

Define $\tilde{D}(X^n, W^n) \coloneqq \max_{i=1,2} \max_{C_{H_i}^c} \frac{1}{n}\sum_{\ell=1}^{n} d(X_\ell + W_\ell, \hat{X}_\ell)$. Hence, we have:

$$2D_T + 2\sigma^2 \geq E_{X^n,W^n}[\tilde{D}(X^n, W^n)].$$

For convenience, we now use $\tilde{D}$ as an abbreviation of $\tilde{D}(X^n, W^n)$. Fix $\Delta > 0$ and let $\tilde{D}_\Delta = \Delta\left\lceil \frac{\tilde{D}}{\Delta} \right\rceil$ be a quantized version of $\tilde{D}$. Note that $\tilde{D}_\Delta$ is also a function of X^n and W^n. It is obvious that $\tilde{D}_\Delta \leq \tilde{D} + \Delta$. Therefore,

$$E(\tilde{D}_\Delta) \leq E(\tilde{D} + \Delta) \leq 2D_T + 2\sigma^2 + \Delta.$$

We then have:

$$\begin{aligned}
&\frac{1}{n}h(X^n + W^n | C_A) \\
=&\frac{1}{n}h(X^n + W^n | C_A, \tilde{D}_\Delta) + \frac{1}{n}I(X^n + W^n; \tilde{D}_\Delta, C_A) && \text{(A.8)} \\
\leq&\frac{1}{n}h(X^n + W^n | C_A, \tilde{D}_\Delta) + \frac{1}{n}H(\tilde{D}_\Delta). && \text{(A.9)}
\end{aligned}$$

Note that $\tilde{D} \leq \tilde{D}_\Delta$, so $X^n + W^n \in Q_{\tilde{D}_\Delta, C_A}$. Therefore, by the uniform bound on entropy,

$$h(X^n + W^n | C_A, \tilde{D}_\Delta) \leq E\left[\log \text{Vol}(cl(Q_{\tilde{D}_\Delta, C_A})\right]. \tag{A.10}$$

Since $\tilde{D}_\Delta$ is quantized, it can be shown using a maximum entropy distribution result [10, Theorem 12.1.1] that

$$H(\tilde{D}_\Delta) \leq \frac{E(\tilde{D}_\Delta)}{\Delta} h\left(\frac{\Delta}{E(\tilde{D}_\Delta)}\right), \tag{A.11}$$

where $h(q) = -q \log q - (1-q)\log(1-q)$ is the binary entropy function. Substituting Inq. (A.10) and Inq. (A.11) into Inq. (A.9), we have

$$\frac{1}{n} h(X^n + W^n | C_A) \leq \frac{1}{n} E\left[\log \text{Vol}(cl(Q_{\tilde{D}_\Delta, C_A})\right] + \frac{1}{n} \frac{E(\tilde{D}_\Delta)}{\Delta} h\left(\frac{\Delta}{E(\tilde{D}_\Delta)}\right). \tag{A.12}$$

Let X^n, W^n, X'^n and W'^n satisfy: $X^n + W^n, X'^n + W'^n \in Q_{D'|c_A}$. Suppose that the decoder receives the following set of codewords

$$(c_A, c_{H_1 \setminus H_2} = f_{H_1 \setminus H_2}(X^n), c_{H_2 \setminus H_1} = f_{H_2 \setminus H_1}(X'^n)).$$

Let $\hat{X}^n$ denote the resulting reconstruction. Then,

$$d(X^n + W^n, \hat{X}^n) \leq D' \Rightarrow \|X^n + W^n - \hat{X}^n\|_2 \leq \sqrt{nD'},$$
$$d(X'^n + W'^n, \hat{X}^n) \leq D' \Rightarrow \|X'^n + W'^n - \hat{X}^n\|_2 \leq \sqrt{nD'}.$$

By the triangle inequality,

$$\|X^n + W^n - X'^n - W'^n\|_2 \leq 2\sqrt{nD'}.$$

Q_{D', c_A} has diameter at most $2\sqrt{nD'}$. We know that the volume of any subset of $\mathfrak{R}^n$ is no more than that of the n-ball with the same diameter (see Lemma 1 in [25]). The following result gives the volume of an n-ball [9, Chapter 21.2.C].

Lemma 31 *The n-dimensional volume of a Euclidean ball of radius R in n-dimensional Euclidean space is:*

$$V_n(R) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)} R^n,$$

where Γ *is Leonhard Euler's Gamma Function.*

Thus, using Lemma 31, we can get Eq. (18) in [9, p. 9]:

$$\mathrm{Vol}(cl(Q_{D'|C_A})) \leq (2\pi e D')^{\frac{n}{2}},$$

indicating that

$$\mathrm{Vol}(cl(Q_{\tilde{D}_\Delta, C_A})) \leq (2\pi e \tilde{D}_\Delta)^{\frac{n}{2}}.$$

Since $E(\tilde{D}_\Delta) \leq 2D_T + 2\sigma^2 + \Delta$, we have

$$H(\tilde{D}_\Delta) \leq \frac{E(\tilde{D}_\Delta)}{\Delta} h\left(\frac{\Delta}{E(\tilde{D}_\Delta)}\right) \leq \frac{2D_T + 2\sigma^2 + \Delta}{\Delta} h\left(\frac{\Delta}{2D_T + 2\sigma^2 + \Delta}\right).$$

Substituting these into Inq. (A.12),

$$\begin{aligned}
&\frac{1}{n} h(X^n + W^n | C_A) \\
&\leq \frac{1}{n} E\left[\log(2\pi e \tilde{D}_\Delta)^{\frac{n}{2}}\right] + \frac{1}{n} \frac{2D_T + 2\sigma^2 + \Delta}{\Delta} \cdot h\left(\frac{\Delta}{2D_T + 2\sigma^2 + \Delta}\right) \\
&\leq \frac{1}{2} \log\left(2\pi e (2D_T + 2\sigma^2 + \Delta)\right) + \frac{1}{n} \frac{2D_T + 2\sigma^2 + \Delta}{\Delta} \cdot h\left(\frac{\Delta}{2D_T + \sigma^2 + \Delta}\right).
\end{aligned}$$

Now, replicate the code for m times, let $C_\ell^{(m)}$ denote the ℓth packet for m-replication. We have:

$$\begin{aligned}
&\frac{1}{n} h(X^n + W^n | C_A) = \frac{1}{nm} h(X^{mn} + W^{mn} | C_A^{(m)}) \\
&\leq \frac{1}{2} \log\left(2\pi e (2D_T + 2\sigma^2 + \Delta)\right) + \frac{1}{nm} \frac{2D_T + 2\sigma^2 + \Delta}{\Delta} \cdot h\left(\frac{\Delta}{2D_T + \sigma^2 + \Delta}\right).
\end{aligned}$$

Let $m \to \infty$, $\Delta \to 0$ and it completes the proof.

A.2.3 Proof of Lemma 19

$$I(Z;X^n) - I(Z;Y_a^n) \geq \frac{n}{2}\log\frac{\sigma_x^2(\tilde{D}+D_a)}{\tilde{D}(\sigma_x^2+D_a)}$$
$$\Longleftrightarrow h(Y_a^n|Z) - h(X^n|Z) \geq \frac{n}{2}\log\frac{\tilde{D}+D_a}{\tilde{D}}.$$

Let $\bar{X}_Z$ be a random variable depending on the value of Z. For each realization z, $\bar{X}_{Z=z}$ is a zero-mean Gaussian random variable with variance chosen to satisfy the following equation:

$$h(X^n + W_a^n|Z=z) - h(X^n|Z=z) = h(\bar{X}_Z^n + W_a^n|Z=z) - h(\bar{X}_Z^n|Z=z).$$

Using Lemma 18, we have:

$$h(\bar{X}_Z|Z=z) \leq h(X^n|Z=z),$$

indicating that

$$h(\bar{X}_Z|Z) \leq h(X^n|Z) \leq \frac{n}{2}\log(2\pi e\tilde{D}). \tag{A.13}$$

Hence,

$$\begin{aligned}
&h(Y_a^n|Z) - h(X^n|Z)\\
&= \sum_z P(Z=z)\cdot\Big(h(Y_a^n|Z=z) - h(X^n|Z=z)\Big)\\
&= \sum_z P(Z=z)\cdot(h(\bar{X}_Z^n + W_a^n|Z=z) - h(\bar{X}_Z^n|Z=z))\\
&= \sum_z P(Z=z)\cdot\left(\frac{n}{2}\log\left(2^{\frac{2}{n}h(\bar{X}_Z^n|Z=z)} + 2^{\frac{2}{n}h(W_a^n)}\right) - h(\bar{X}_Z^n|Z=z)\right)\\
&\geq \frac{n}{2}\log\left(2^{\frac{2}{n}\sum_z P(Z=z)\cdot h(\bar{X}_Z^n|Z=z)} + 2^{\frac{2}{n}h(W_a^n)}\right) - \sum_z P(Z=z)\cdot h(\bar{X}_Z^n|Z=z) \qquad \text{(A.14)}\\
&= \frac{n}{2}\log\left(2^{\frac{2}{n}h(\bar{X}_Z^n|Z)} + 2^{\frac{2}{n}h(W_a^n)}\right) - h(\bar{X}_Z^n|Z)\\
&\geq \frac{n}{2}\log(2\pi e(\tilde{D}+D_a)) - \frac{n}{2}\log(2\pi e\tilde{D}) = \frac{n}{2}\log\frac{\tilde{D}+D_a}{\tilde{D}}. \qquad \text{(A.15)}
\end{aligned}$$

Inq. (A.14) follows from Jensen's Inequality and the fact that the following function

$$f_1(x) = \frac{n}{2}\log(2^{\frac{2}{n}x} + C)$$

is convex for any constant $C \geq 0$. Inq. (A.15) follows from Inq. (A.13) and the fact that the following function

$$f_2(x) = \frac{n}{2}\log\left(2^{\frac{2}{n}x} + C\right) - x$$

is non-increasing for any constant $C \geq 0$.

A.2.4 Proof of Lemma 22

We first prove that $R'_{\text{in}}(D_0, D_T) \subseteq R_{\text{in}}(D_0, D_T)$. It is equivalent to prove for any $\boldsymbol{R} = (R_1, ..., R_N) \in R'_{\text{in}}(D_0, D_T)$, we can construct $\{R_{a,\ell}\}$ and $\{R_{b,\ell}\}$ which satisfy:

$$\begin{aligned}
&R_\ell = R_{a,\ell} + R_{b,\ell}, \quad \forall \ell \in [N];\\
&\sum_{\ell \in A} R_{a,\ell} \geq \frac{1}{2}\log\frac{1}{D'_T}, \quad \forall A \subset [N], |A| = N - 2T;\\
&\sum_{\ell=1}^{N} R_{b,\ell} \geq \frac{1}{2}\log\frac{D'_T}{D_0};\\
&R_{a,\ell} \geq 0, R_{b,\ell} \geq 0, \quad \forall \ell \in [N].
\end{aligned}$$

Without loss of generality, we assume that $R_1 \geq \cdots \geq R_N$. Let $\{R_{a,\ell}\}_{2T+1 \leq \ell \leq N}$ be an *optimal solution* for $(\frac{1}{2}\log\frac{1}{D'_T}, R_{2T+1}, ..., R_N)$. This solution exists because $\sum_{\ell=2T+1}^{N} R_\ell \geq \frac{1}{2}\log\frac{1}{D'_T}$. Then we have: $\sum_{\ell=2T+1}^{N} R_{a,\ell} = \frac{1}{2}\log\frac{1}{D'_T}$. Without loss of generality, we can assume that $R_{a,2T+1} \geq \cdots \geq R_{a,N}$. Let $R_{b,\ell} = R_\ell - R_{a,\ell} \geq 0$ for $2T + 1 \leq \ell \leq N$. Finally, for $1 \leq \ell \leq 2T$, let

$$R_{a,\ell} = R_{a,2T+1}; \quad R_{b,\ell} = R_\ell - R_{a,\ell} \geq 0.$$

Then for all $A \subset [N]$, $|A| = N - 2T$,

$$\sum_{\ell \in A} R_{a,\ell} \geq \sum_{\ell=2T+1}^{N} R_{a,\ell} = \frac{1}{2}\log\frac{1}{D'_T}.$$

Furthermore,

$$\begin{aligned}
&\sum_{\ell=1}^{N} R_{b,\ell} = \sum_{\ell=1}^{N} (R_\ell - R_{a,\ell}) \\
&= \sum_{\ell=1}^{N} R_\ell - \sum_{\ell=1}^{2T} R_{a,\ell} - \sum_{\ell=2T+1}^{N} R_{a,\ell} \\
&\geq \frac{1}{2}\log\frac{1}{D_0} + 2T \cdot \mathcal{G}(N-2T, \frac{1}{2}\log\frac{1}{D'_T}, R_{2T+1}, ..., R_N) \\
&\quad - 2T \cdot \mathcal{G}(N-2T, \frac{1}{2}\log\frac{1}{D'_T}, R_{2T+1}, ..., R_N) - \frac{1}{2}\log\frac{1}{D'_T} \\
&= \frac{1}{2}\log\frac{D'_T}{D_0}.
\end{aligned}$$

We next prove that $R_{\text{in}}(D_0, D_T) \subseteq R'_{\text{in}}(D_0, D_T)$. For any $\boldsymbol{R} = (R_1, ..., R_N) \in R_{\text{in}}(D_0, D_T)$, there exist $\{R_{a,\ell}\}$ and $\{R_{b,\ell}\}$ such that:

$$R_\ell = R_{a,\ell} + R_{b,\ell}, \quad \forall \ell \in [N]; \tag{A.16}$$

$$\sum_{\ell \in A} R_{a,\ell} \geq \frac{1}{2}\log\frac{1}{D'_T}, \quad \forall A \subset [N], |A| = N - 2T; \tag{A.17}$$

$$\sum_{\ell=1}^{N} R_{b,\ell} \geq \frac{1}{2}\log\frac{D'_T}{D_0}; \tag{A.18}$$

$$R_{a,\ell} \geq 0, R_{b,\ell} \geq 0, \quad \forall \ell \in [N]. \tag{A.19}$$

We next show that $\boldsymbol{R} \in R'_{\text{in}}(D_0, D_T)$. Without loss of generality, we can assume $R_1 \geq \cdots \geq R_N$. We can pick $\{R_{a,\ell}\}_{\ell \in [N]}$ such that:

$$R_{a,1} \geq \cdots \geq R_{a,N}.$$

This is because if there exist $R_{a,i}$ and $R_{a,j}$ such that $R_{a,i} < R_{a,j}$ and $i < j$, we can let $R'_{a,i} = R_{a,j}$, $R'_{a,j} = R_{a,i}$ and let $R'_{b,i} = R_i - R'_{a,i}$, $R'_{b,j} = R_j - R'_{a,j}$. The other $R_{a,\ell}$ and $R_{b,\ell}$ remain the same. The new $\{R_{a,\ell}\}$ and $\{R_{b,\ell}\}$ also satisfy Eq. (A.16). Since $R_{a,2T+1}, ..., R_{a,N}$ satisfy:

$$\begin{cases} 0 \leq R_{a,2T+\ell} \leq R_{2T+\ell}, & \forall \ell \in [2N-T] \\ \sum_{\ell=2T+1}^{N} R_{a,\ell} \geq \frac{1}{2}\log\frac{1}{D'_T} \end{cases}.$$

We have:

$$R_{a,2T+1} \geq \mathcal{G}(N-2T, \frac{1}{2}\log\frac{1}{D'_T}, R_{2T+1}, ..., R_N).$$

Therefore,

$$\begin{aligned}
&\sum_{\ell=1}^{N} R_\ell = \sum_{\ell=1}^{N}(R_{a,\ell} + R_{b,\ell}) \\
&= \sum_{\ell=2T+1}^{N} R_{a,\ell} + \sum_{\ell=1}^{N} R_{b,\ell} + \sum_{\ell=1}^{2T} R_{a,\ell} \\
&\geq \frac{1}{2}\log\frac{1}{D'_T} + \frac{1}{2}\log\frac{D'_T}{D_0} + 2TR_{a,2T+1} \\
&\geq \frac{1}{2}\log\frac{1}{D_0} + 2T\cdot\mathcal{G}(N-2T, \frac{1}{2}\log\frac{1}{D'_T}, R_{2T+1}, ..., R_N).
\end{aligned}$$

Moreover,

$$\sum_{\ell=2T+1}^{N} R_\ell = \sum_{\ell=2T+1}^{N}(R_{a,\ell} + R_{b,\ell}) \geq \sum_{\ell=2T+1}^{N} R_{a,\ell} \geq \frac{1}{2}\log\frac{1}{D'_T}.$$

The above two inequalities indicate that $\boldsymbol{R} \in R'_{\text{in}}(D_0, D_T)$.

A.3 Proofs for P2P VPEC

A.3.1 Proof of Lemma 23

Keep in mind that $P_X = \mathcal{P}$ throughout this proof. Given (n, δ), recall that $\{U^n(i)\}$ are drawn i.i.d. according to the same distribution; for any fixed $U(i)$, $\left\{Y(i,1), ..., Y(i, 2^{nR_2})\right\}$ are drawn i.i.d. according to the same conditional distribution. Define

$$\mathcal{A}(n,\delta) := \{(u^n, x^n) : u^n \in T^{(n,\delta/2)}(P_U); \quad \exists y^n, \text{ s.t. } (u^n, y^n, x^n) \in T^{(n,\delta)}(P_{UYX})\}.$$

For $x^n \in E_1^c(n,\delta)$ define

$$p_1(n,\delta,x^n) := P((U^n(i), x^n) \in \mathcal{A}(n,\delta)),$$

and for $(u^n, x^n) \in \mathcal{A}(n, \delta)$, define

$$p_2(n, \delta, x^n, u^n) := P((U^n(i), Y^n(i, j), x^n) \in T^{(n,\delta)}(P_{UYX}) | U^n(i) = u^n).$$

Since $\{U^n(i)\}$ and $\left\{Y(i, 1), ..., Y(i, 2^{nR_2})\right\}$ drawn i.i.d., $p_1(n, \delta, x^n)$ and $p_2(n, \delta, x^n, u^n)$ are well-defined. In addition, they are always non-negative.

Similarly, given distribution P^*_{UYX} satisfying $P^*_{UY} = P_{UY}$ and $P^*_X = P_X$, for $x^n \in T^{(n,\delta^*)}(P^*_X)$, define

$$p_3(n, \delta, x^n, P^*_{UYX}) := P((U^n(i), x^n) \in T^{(n,\delta^*)}(P^*_{UX})),$$

for $(u^n, x^n) \in T^{(n,\delta^*)}(P^*_{UYX})$, define

$$p_4(n, \delta, x^n, u^n, P^*_{UYX}) := P((U^n(i), Y^n(i, j), x^n) \in \mathcal{T}^{(n,\delta^*)}(P^*_{UYX}) | U^n(i) = u^n).$$

Let $N_1(n, \delta, x^n)$ denote the number of indices $i \in \left[2^{\lfloor nR_1 \rfloor}\right]$, such that $(u^n(i), x^n) \in \mathcal{A}(n, \delta)$. $N_1(n, \delta, x^n)$ is a random variable. Let

$$B(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k}$$

denote the pmf of a Binomial distribution. It is not difficult to verify that the distribution of $N_1(n, \delta, x^n)$ is

$$\begin{aligned}
&P(N_1(n, \delta, x^n) = k) \\
=&\binom{2^{\lfloor nR_1 \rfloor}}{k} p_1(n, \delta, x^n)^k (1 - p_1(n, \delta, x^n))^{2^{\lfloor nR_1 \rfloor} - k} \\
=&B(k; 2^{\lfloor nR_1 \rfloor}, p_1(n, \delta, x^n)).
\end{aligned}$$

Similarly, let $N_2(n, \delta, x^n, P^*_{UYX})$ denote the number of indices $i \in \left[2^{\lfloor nR_1 \rfloor}\right]$, such that $(u^n(i), x^n) \in T^{(n,\delta^*)}(P^*_{UX})$. It is not difficult to verify that the distribution of

$N_2(n, \delta, x^n, P^*_{UYX})$ is

$$\begin{aligned}
&P(N_2(n, \delta, x^n, P^*_{UYX}) = k) \\
=&\binom{2^{\lfloor nR_1 \rfloor}}{k} p_3(n, \delta, x^n, P^*_{UYX})^k (1 - p_3(n, \delta, x^n, P^*_{UYX}))^{2^{\lfloor nR_1 \rfloor} - k} \\
=&B(k; 2^{\lfloor nR_1 \rfloor}, p_3(n, \delta, x^n, P^*_{UYX})).
\end{aligned}$$

Part 1) For each (n, δ), let

$$\begin{aligned}
\underline{p}_1(n, \delta) &= \min_{x^n \in E_1^c(n,\delta)} p_1(n, \delta, x^n), \\
\underline{p}_2(n, \delta) &= \min_{(u^n, x^n) \in \mathcal{A}(n,\delta)} p_2(n, \delta, x^n, u^n).
\end{aligned}$$

For any $\delta > 0$, we first observe that

$$\lim_{n \to \infty} E_{C(n)}[\mathcal{P}^n(E_1(C(n), n, \delta))] = 0.$$

Since $\{U^n(i)\}$ are drawn i.i.d., for any $x^n \in E_1^c(n, \delta)$, we have

$$P(x^n \in E_2(C(n), n, \delta) | N_1(n, \delta, x^n) = k) \leq (1 - \underline{p}_2(n, \delta))^{k \cdot 2^{\lfloor nR_2 \rfloor}}.$$

Therefore, for any $x^n \in E_1^c(n, \delta)$, we have

$$\begin{aligned}
P(x^n \in E_2(C(n), n, \delta)) = &\sum_{k=0}^{2^{\lfloor nR_1 \rfloor}} \{P(N_1(n, \delta, x^n) = k) \cdot \\
&P(x^n \in E_2(C(n), n, \delta) | N_1(n, \delta, x^n) = k)\} \\
\leq &\sum_{k=0}^{2^{\lfloor nR_1 \rfloor}} B(k; 2^{\lfloor nR_1 \rfloor}, p_1(n, \delta, x^n))(1 - \underline{p}_2(n, \delta))^{k \cdot 2^{\lfloor nR_2 \rfloor}}.
\end{aligned}$$

So $E_{C(n)}[\mathcal{P}^n(E_2(C(n), n, \delta))]$ is bounded as follows:

$$\begin{aligned}
&E_{C(n)}[\mathcal{P}^n(E_2(C(n), n, \delta))] \\
&= \sum_{x^n \in E_1^c(n,\delta)} \mathcal{P}^n(x^n) P(x^n \in E_2(C(n), n, \delta)) \\
&\leq \sum_{x^n \in E_1^c(n,\delta)} \mathcal{P}^n(x^n) \cdot \sum_{k=0}^{2^{\lfloor nR_1 \rfloor}} \{(1 - \underline{p}_2(n, \delta))^{k \cdot 2^{\lfloor nR_2 \rfloor}} \cdot B(k; 2^{\lfloor nR_1 \rfloor}, p_1(n, \delta, x^n))\}.
\end{aligned}$$

Using binomial expansion, we know that for any $a \in \mathfrak{R}$

$$\sum_{k=0}^{n} B(k; n, p) a^k = ((1 - p) + pa)^n = (1 - p(1 - a))^n.$$

Thus,

$$E_{C(n)}[\mathcal{P}^n(E_2(C(n), n, \delta))] \leq \left(1 - \underline{p}_1(n, \delta)\left(1 - (1 - \underline{p}_2(n, \delta))^{2^{\lfloor nR_2 \rfloor}}\right)\right)^{2^{\lfloor nR_1 \rfloor}}. \tag{A.20}$$

For any $x^n \in E_1^c(n, \delta)$, there exists at least one string $u^n \in \mathcal{U}^n$ such that $(u^n, x^n) \in \mathcal{A}(n, \delta)$. Let u'^n be one such string, then

$$p_1(n, \delta, x^n) = P'_{U^n}(\{u^n : (u^n, x^n) \in \mathcal{A}(n, \delta)\}) \geq P'_{U^n}(T_{F_{u'^n|x^n}}(x^n)).$$

For any u^n such that $(u^n, x^n) \in \mathcal{A}(n, \delta)$, we must have $u^n \in T^{(n,\delta)}(P_U)$. Thus, by Inq. (5.5),

$$p_1(n, \delta, x^n) \geq P'_{U^n}(T_{F_{u'^n|x^n}}(x^n)) \geq P^n_U(T_{F_{u'^n|x^n}}(x^n)).$$

by definition. For any u^n with type of sequences P_{u^n}, we have $P^n_U(u^n) =$

$2^{-n(H(P_{u^n})+D(P_{u^n}\|P_U))}$. So, combining Lemma 2.5 in [11],

$$\begin{aligned}
\underline{p}_1(n,\delta) &= \min_{x^n \in E_1^c(n,\delta)} p_1(n,\delta,x^n) \\
&\geq \min_{(u^n,x^n)\in\mathcal{A}(n,\delta)} P_U^n(T_{F_{u^n|x^n}}(x^n)) \\
&= \min_{(u^n,x^n)\in\mathcal{A}(n,\delta)} |T_{F_{u^n|x^n}}(x^n)| \cdot 2^{-n(H(P_{u^n})+D(P_{u^n}\|P_U))} \\
&\geq (n+1)^{-|\mathcal{X}\|\mathcal{U}|} \cdot \min_{(u^n,x^n)\in\mathcal{A}(n,\delta)} 2^{n(H(F_{u^n|x^n}|P_{x^n})-H(P_{u^n})-D(P_{u^n}\|P_U))}.
\end{aligned}$$

Therefore,

$$\lim_{\delta\to 0}\lim_{n\to\infty}\frac{1}{n}\log \underline{p}_1(n,\delta) \geq -I(U;X),$$

Similarly, using Inq. (5.6),

$$\underline{p}_2(n,\delta) \geq \min_{(u^n,y^n,x^n)\in T^{(n,\delta)}(P_{UYX})} P_Y^n(T_{F_{y^n|(u^n,x^n)}}(u^n,x^n)),$$

and

$$\lim_{\delta\to 0}\lim_{n\to\infty}\frac{1}{n}\log \underline{p}_2(n,\delta) \geq -I(X;Y|U).$$

If $R_1 > I(U;X)$, $R_1+R_2 > I(U;X)+I(X;Y|U)$, then when δ is sufficiently small, for all n sufficiently large (depending on δ), we have

$$\begin{aligned}
-\frac{1}{n}\log \underline{p}_1(n,\delta) &\leq \frac{1}{2}(R_1 + I(U;X)) < R_1, \\
-\frac{1}{n}\log(\underline{p}_1(n,\delta)\cdot p_2(n,\delta)) &\leq \frac{1}{2}(R_1+R_2+I(U;X)+I(X;Y|U)) < R_1+R_2.
\end{aligned}$$

Thus, for all small δ,

$$\lim_{n\to\infty}\left\{\underline{p}_1(n,\delta)\cdot 2^{\lfloor nR_1\rfloor}\right\} = \infty, \tag{A.21}$$

$$\lim_{n\to\infty}\left\{\underline{p}_1(n,\delta)\underline{p}_2(n,\delta)\cdot 2^{\lfloor nR_1\rfloor+\lfloor nR_2\rfloor}\right\} = \infty. \tag{A.22}$$

Using Inq. (A.20) and the inequality $(1-x)^n \le e^{-nx}$ for $x \ge 0$, we know that for all small δ,

$$\begin{aligned}
&\lim_{n\to\infty} E_{C(n)}[\mathcal{P}^n(E_2(C(n),n,\delta))] \\
&\le \limsup_{n\to\infty} \exp\left(-\underline{p}_1(n,\delta)2^{\lfloor nR_1\rfloor}\cdot\left(1-(1-\underline{p}_2(n,\delta))^{2^{\lfloor nR_2\rfloor}}\right)\right) \\
&\le \limsup_{n\to\infty} \exp\left(-\underline{p}_1(n,\delta)2^{\lfloor nR_1\rfloor}\cdot\left(1-\exp\left(-\underline{p}_2(n,\delta)2^{\lfloor nR_2\rfloor}\right)\right)\right).
\end{aligned}$$

If there exists a constant c_1 such that $\underline{p}_2(n,\delta)2^{\lfloor nR_2\rfloor} \ge c_1$, then

$$\begin{aligned}
&\liminf_{n\to\infty}\left\{\underline{p}_1(n,\delta))2^{\lfloor nR_1\rfloor}\cdot\left(1-\exp\left(-\underline{p}_2(n,\delta)2^{\lfloor nR_2\rfloor}\right)\right)\right\} \\
&\ge \liminf_{n\to\infty} \underline{p}_1(n,\delta))2^{\lfloor nR_1\rfloor}\cdot(1-\exp(-c_1)) = \infty.
\end{aligned}$$

Else, $\lim_{n\to\infty}\underline{p}_2(n,\delta)2^{\lfloor nR_2\rfloor} = 0$. Note that $\lim_{x\to 0}\frac{1-e^{-x}}{x} = 1$. Therefore

$$\lim_{n\to\infty}\frac{1-\exp\left(-\underline{p}_2(n,\delta)2^{\lfloor nR_2\rfloor}\right)}{\underline{p}_2(n,\delta)2^{\lfloor nR_2\rfloor}} = 1,$$

and

$$\begin{aligned}
&\liminf_{n\to\infty}\left\{\underline{p}_1(n,\delta))2^{\lfloor nR_1\rfloor}\cdot\left(1-\exp\left(-\underline{p}_2(n,\delta)2^{\lfloor nR_2\rfloor}\right)\right)\right\} \\
&= \liminf_{n\to\infty}\underline{p}_1(n,\delta))\underline{p}_2(n,\delta)2^{\lfloor nR_1\rfloor}2^{\lfloor nR_2\rfloor}\cdot\lim_{n\to\infty}\frac{1-\exp\left(-\underline{p}_2(n,\delta)2^{\lfloor nR_2\rfloor}\right)}{\underline{p}_2(n,\delta)2^{\lfloor nR_2\rfloor}} = \infty.
\end{aligned}$$

In both cases, we all have:

$$\lim_{n\to\infty} E_{C(n)}[\mathcal{P}^n(E_2(C(n),n,\delta))] = 0.$$

Part 2) We first observe that

$$\limsup_{n\to\infty} E_{C(n)}[\mathcal{P}^n(E_0^{*c}(C(n),n,\delta,P^*_{UYZ}))] > 0$$

implies

$$\begin{aligned}
&\liminf_{n\to\infty} E_{C(n)}[\mathcal{P}^n(E_1^*(C(n),n,\delta,P^*_{UYZ}))] + \liminf_{n\to\infty} E_{C(n)}[\mathcal{P}^n(E_2^*(C(n),n,\delta,P^*_{UYZ}))] \\
&\le \liminf_{n\to\infty} E_{C(n)}[\mathcal{P}^n(E_0^*(C(n),n,\delta,P^*_{UYZ}))] < 1.
\end{aligned}$$

Recall that (U^*, Y^*, X^*) has distribution P^*_{UYX}. Keep in mind that in the proof of Part 2), we always have $P^*_{UY} = P_{UY}$, $P^*_X = P_X = \mathcal{P}$, and $\delta^* = C\delta$. Let

$$\bar{p}_3(n, \delta, P^*_{UYX}) = \max_{x^n \in T^{(n,\delta^*)}(P^*_X)} p_3(n, \delta, x^n, P^*_{UYX}),$$

$$\bar{p}_4(n, \delta, P^*_{UYX}) = \max_{(u^n, x^n) \in T^{(n,\delta^*)}(P^*_{UX})} p_4(n, \delta, x^n, u^n, P^*_{UYX}).$$

We can compute that

$$E_{C(n)}[\mathcal{P}^n(E^*_1(C(n), n, \delta, P^*_{UYX}))] = 1 - \mathcal{P}^n(T^{(n,\delta^*)}(P^*_X)).$$

Since $\{U^n(i)\}$ are drawn i.i.d., for any $x^n \in T^{(n,\delta^*)}(P^*_X)$, we have

$$P(x^n \in E^*_2(C(n), n, \delta, , P^*_{UYX}) | N_2(n, \delta, x^n, P^*_{UYX}) = k) \geq (1 - \bar{p}_4(n, \delta, P^*_{UYX}))^{k \cdot 2^{\lfloor nR_2 \rfloor}}.$$

Therefore, for any $x^n \in T^{(n,\delta^*)}(P_X)$, we have

$$\begin{aligned}
&P(x^n \in E^*_2(C(n), n, \delta, P^*_{UYX})) \\
&= \sum_{k=0}^{2^{\lfloor nR_1 \rfloor}} \{P(N(n, \delta, x^n, P^*_{UYX}) = k) \cdot P(x^n \in E^*_2(C(n), n, \delta, P^*_{UYX}) | N(n, \delta, x^n, P^*_{UYX}) = k)\} \\
&\geq \sum_{k=0}^{2^{\lfloor nR_1 \rfloor}} \{B(k; 2^{\lfloor nR_1 \rfloor}, p_3(n, \delta, x^n, P^*_{UYX})) \cdot (1 - \bar{p}_4(n, \delta, P^*_{UYX}))^{k \cdot 2^{\lfloor nR_2 \rfloor}}\}.
\end{aligned}$$

So $E_{C(n)}[\mathcal{P}^n(E^*_2(C(n), n, \delta, P^*_{UYX}))]$ is bounded as follows:

$$\begin{aligned}
&E_{C(n)}[\mathcal{P}^n(E^*_2(C(n), n, \delta, P^*_{UYX}))] \\
&= \sum_{\substack{x^n \in \\ T^{(n,\delta^*)}(P^*_X)}} \mathcal{P}^n(x^n) P(x^n \in E^*_2(C(n), n, \delta, P^*_{UYX})) \\
&\geq \sum_{\substack{x^n \in \\ T^{(n,\delta^*)}(P^*_X)}} \mathcal{P}^n(x^n) \cdot \sum_{k=0}^{2^{\lfloor nR_1 \rfloor}} \{(1 - \bar{p}_4(n, \delta, P^*_{UYX}))^{k \cdot 2^{\lfloor nR_2 \rfloor}} \cdot B(k; 2^{\lfloor nR_1 \rfloor}, p_3(n, \delta, x^n, P^*_{UYX}))\}
\end{aligned}$$

Similarly, using binomial expansion, we know that

$$\begin{aligned}
&E_{C(n)}[\mathcal{P}^n(E_2^*(C(n), n, \delta, P^*_{UYX}))] \\
\geq &\mathcal{P}^n(T^{(n,\delta^*)}(P^*_X)) \cdot \left(1 - \bar{p}_3(n, \delta, P^*_{UYX})\left(1 - (1 - \bar{p}_4(n, \delta, P^*_{UYX}))^{2^{\lfloor nR_2 \rfloor}}\right)\right)^{2^{\lfloor nR_1 \rfloor}}
\end{aligned} \tag{A.23}$$

Using Lemma 2.2 in [11], for any n, there are at most $(n+1)^{|\mathcal{U}||\mathcal{X}|}$ different *types of sequences* in $(\mathcal{U} \times \mathcal{X})^n$. Again, combining Lemma 2.5 in [11] and using the fact that $P^*_U = P_U$ along with the definition of P'_{U^n}, we know that there exists a constant α_1 for all P^*_{UYX}, such that for all large n:

$$\begin{aligned}
&\bar{p}_3(n, \delta, P^*_{UYX}) \\
\leq &(n+1)^{|\mathcal{X}||\mathcal{U}|} \max_{(u^n, x^n) \in T^{(n,\delta^*)}(P^*_{UX})} P'_{U^n}(T_{F_{u^n|x^n}}(x^n)) \\
\leq &\alpha_1 (n+1)^{|\mathcal{X}||\mathcal{U}|} \max_{(u^n, x^n) \in T^{(n,\delta^*)}(P^*_{UX})} P^n_U(T_{F_{u^n|x^n}}(x^n)) \\
= &\alpha_1 (n+1)^{|\mathcal{X}||\mathcal{U}|} \cdot \max_{(u^n, x^n) \in T^{(n,\delta^*)}(P^*_{UX})} \left\{|T_{F_{u^n|x^n}}(x^n)| \cdot 2^{-n(H(P_{u^n}) + D(P_{u^n} \| P^*_U))}\right\} \\
\leq &\alpha_1 (n+1)^{|\mathcal{X}||\mathcal{U}|} \cdot \max_{(u^n, x^n) \in T^{(n,\delta^*)}(P^*_{UX})} 2^{nH(F_{u^n|x^n}|P_{x^n}) - H(P_{u^n}) - D(P_{u^n} \| P^*_U))} \\
\leq &\alpha_1 (n+1)^{|\mathcal{X}||\mathcal{U}|} \cdot \max_{(u^n, x^n) \in T^{(n,\delta^*)}(P^*_{UX})} 2^{nH(F_{u^n|x^n}|P_{x^n}) - H(P_{u^n}))}.
\end{aligned}$$

Let $\Gamma_1(\delta, P^*_{UYX}) = \{\tilde{P}_{UX} : \|\tilde{P}_{UX} - P^*_{UX}\| \leq \delta^*\}$. Let $\Gamma_2(\delta, P^*_{UYX}) = \{\tilde{P}_U : \|\tilde{P}_U - P^*_U\| \leq \delta^*\}$. Combining the above analysis, for any distribution P^*_{UYX}, it is not difficult to verify that:

$$\lim_{n \to \infty} \frac{1}{n} \log \bar{p}_3(n, \delta, P^*_{UYX}) \leq \sup_{\tilde{P}_{UX} \in \Gamma_1(\delta, P^*_{UYX})} H_{\tilde{P}_{UX}}(\tilde{U}|\tilde{X}) - \inf_{\tilde{P}_U \in \Gamma_2(\delta, P^*_{UYX})} H(\tilde{P}_U).$$

(Here $H_{\tilde{P}_{UX}}(\tilde{U}|\tilde{X})$ denote the conditional distribution of $\tilde{U}$ given $\tilde{X}$ where $(\tilde{U}, \tilde{X})$ has distribution $\tilde{P}_{UY}$). $\Gamma_1(\delta, P^*_{UYX})$ and $\Gamma_2(\delta, P^*_{UYX})$ are compact sets. $H_{\tilde{P}_{UX}}(\tilde{U}|\tilde{X})$ and $H(\tilde{P}_U)$ and are continuous function of $\tilde{P}_{UX}$ and $\tilde{P}_U$, respectively. Thus, it is eligible to substitute

max and min for sup and inf, respectively:

$$\limsup_{n\to\infty}\frac{1}{n}\log\bar{p}_3(n,\delta)\le\max_{\tilde{P}_{UX}\in\Gamma_1(\delta,P^*_{UYX})}H_{\tilde{P}_{UX}}(\tilde{U}|\tilde{X})-\min_{\tilde{P}_U:\in\Gamma_2(\delta,P^*_{UYX})}H(\tilde{P}_U)$$

$$:=\xi_1(\delta,P^*_{UYX}). \tag{A.24}$$

Similarly, using the facts that $P^*_{UY}=P_{UY}$ and $P^*_X=P_X$, we have:

$$\limsup_{n\to\infty}\frac{1}{n}\log\bar{p}_4(n,\delta)$$

$$\le\max_{\tilde{P}_{UYX}\in\Gamma_3(\delta,P^*_{UYX})}H_{\tilde{P}_{UYX}}(\tilde{Y}|\tilde{U},\tilde{X})-\min_{\tilde{P}_{UY}\in\Gamma_4(\delta,P^*_{UYX})}H_{\tilde{P}_{UY}}(\tilde{Y}|\tilde{U}) \tag{A.25}$$

$$:=\xi_2(\delta,P^*_{UYX}), \tag{A.26}$$

where

$$\Gamma_3(\delta,P^*_{UYX})=\{\tilde{P}_{UYX}:\|\tilde{P}_{UYX}-P^*_{UYX}\|\le\delta^*\},$$

$$\Gamma_4(\delta,P^*_{UYX})=\{\tilde{P}_{UY}:\|\tilde{P}_{UY}-P^*_{UY}\|\le\delta^*\}.$$

(Here, $H_{\tilde{P}_{UYX}}(\tilde{Y}|\tilde{U},\tilde{X})$ and $H_{\tilde{P}_{UY}}(\tilde{Y}|\tilde{U})$ are defined analogously as $H_{\tilde{P}_{UX}}(\tilde{U}|\tilde{X})$.)

For any $x^n\in T^{(n,\delta)}(P^*_X)$, we must have

$$\|P_{x^n}-P^*_X\|\le\delta^*.$$

By definition, $\xi_1(\delta,P^*_{UYX})\ge -I(X^*;U^*)$ and $\xi_2(\delta,P^*_{UYX})\ge -I(X^*;Y^*|U^*)$. Let

$$\xi(\delta,P^*_{UYX})$$

$$:=\max\left\{\xi_1(\delta,P^*_{UYX})+I(X^*;U^*),\xi_1(\delta,P^*_{UYX})+\xi_2(\delta,P^*_{UYX})+I(X^*;U^*)+I(X^*;Y^*|U^*)\right\}.$$

It is obvious that

$$\lim_{\delta\to0}\xi_1(\delta,P^*_{UYX})=-I(X^*;U^*),$$

$$\lim_{\delta\to0}\xi_2(\delta,P^*_{UYX})=-I(X^*;Y^*|U^*).$$

For any other distribution P^{**}_{UYX}, suppose $\tilde{P}'_{UX}$ and $\tilde{P}''_{U}$ achieve max and min, respectively in the expression of $\xi_1(\delta, P^{**}_{UYX})$:

$$\xi_1(\delta, P^{**}_{UYX}) = H_{\tilde{P}'_{UX}}(\tilde{U}'|\tilde{X}') - H(\tilde{P}''_{U}).$$

(Here $\tilde{P}'_{U}$ may not equal $\tilde{P}''_{U}$.) Since $\|\tilde{P}'_{UX} - P^{**}_{UX}\| \leq \delta$, by Triangle Inequality,

$$\|\tilde{P}'_{UX} - P^{*}_{UX}\| \leq \delta + \|P^{**}_{UX} - P^{*}_{UX}\|.$$

Since $\|\tilde{P}''_{U} - P^{*}_{U}\| \leq \delta$, by Triangle Inequality,

$$\|\tilde{P}''_{U} - P^{*}_{U}\| \leq \delta + \|P^{**}_{U} - P^{*}_{U}\|.$$

It is not difficult to verify that $\|P^{**}_{UX} - P^{*}_{UX}\| \geq \|P^{**}_{U} - P^{*}_{U}\|$. Therefore, for any δ and P^{*}_{UYX},

$$\tilde{P}'_{UX} \in \Gamma_1(\delta + \|P^{**}_{UX} - P^{*}_{UX}\|, P^{*}_{UYX}),$$
$$\tilde{P}''_{U} \in \Gamma_2(\delta + \|P^{**}_{UX} - P^{*}_{UX}\|, P^{*}_{UYX}).$$

We have

$$\xi_1(\delta, P^{**}_{UYX}) \leq \xi_1(\delta + \|P^{**}_{UX} - P^{*}_{UX}\|, P^{*}_{UYX}).$$

Similarly, for any δ and P^{*}_{UYX}, we have

$$\xi_2(\delta, P^{**}_{UYX}) \leq \xi_2(\delta + \|P^{**}_{UYX} - P^{*}_{UYX}\|, P^{*}_{UYX}).$$

Thus, if $\{(\delta_k, P^{(k)}_{UYZ})\}$ is a sequence satisfying:

$$\lim_{k\to\infty} \delta_k = 0, \quad \lim_{k\to\infty} P^{(k)}_{UYZ} = P^{*}_{UYZ},$$

then

$$\begin{aligned}
&\limsup_{k\to\infty} \xi_1(\delta_k, P^{(k)}_{UX}) \\
\leq &\limsup_{k\to\infty} \xi_1(\delta_k + \|P^{(k)}_{UX} - P^{*}_{UX}\|, P^{*}_{UYX}) \\
= &- I(X;U),
\end{aligned}$$

and

$$\begin{aligned}
&\limsup_{k\to\infty} \xi_2(\delta_k, P^{(k)}_{UYX}) \\
\leq &\limsup_{k\to\infty} \xi_2(\delta_k + \|P^{(k)}_{UYX} - P^*_{UYX}\|_\infty, P^*_{UYX}) \\
= &- I(X;Y|U).
\end{aligned}$$

This gives $\lim_{k\to\infty} \xi(\delta_k, P^{(k)}_{UYX}) = 0$.

A.3.2 Proof of Inq. (5.8)

To prove Inq. (5.8), it is equivalent to prove that

$$\limsup_{\delta\to 0}\{\limsup_{n\to\infty} E_{C(n)}[\mathcal{P}^n(A_1(C(n)))]\} = 0.$$

Notice that for any $\delta > 0$,

$$\lim_{n\to\infty} E_{C(n)}[\mathcal{P}^n(E_1^c(C(n), n, \delta))] = 1.$$

Thus, it is equivalent to prove that

$$\begin{aligned}
&\limsup_{\delta\to 0}\limsup_{n\to\infty} E_{C(n)}[\mathcal{P}^n(A_1(C(n)) \cap E_1^c(C(n), n, \delta))] \\
&= 0. \qquad \text{(A.27)}
\end{aligned}$$

We next prove Eq. (A.27) by contradiction. If Eq. (A.27) does not hold, then there exists a positive $\alpha \in (0, 1]$, a positive sequence $\{\delta_k\}$ ($\lim_{k\to\infty} \delta_k = 0$) and a sequence $\{n_{k,\ell}\}$ for each k s.t.

$$E_{C(n_{k,\ell},\delta_k)}[\mathcal{P}^n(A_1(C(n_{k,\ell}, \delta_k)) \cap E_1^c(C(n_{k,\ell}, \delta_k), n_{k,\ell}, \delta_k))] > \alpha, \quad \forall k, \ell \in \mathbb{Z}^+. \qquad \text{(A.28)}$$

Consider the metric space $(\mathcal{S}, \|\cdot\|_\infty)$, where $\mathcal{S}$ is the set of all probability distributions on $\mathcal{U} \times \mathcal{Y} \times \mathcal{X}$. For any $\delta > 0$, let

$$\begin{aligned}\mathcal{D}_\delta := \{P'_{UYX} : &\|P_X - P'_X\|_\infty \leq \frac{\delta}{|\mathcal{X}|}, \\ &\|P_{UY} - P'_{UY}\|_\infty \leq \frac{\delta}{|\mathcal{U}||\mathcal{Y}|}, \\ &P'_{UY}(a,b) = 0, \quad \forall (a,b) \in \mathcal{U} \times \mathcal{Y}, \text{ s.t. } P_{UY}(a,b) = 0\}.\end{aligned}$$

$\forall P'_{UYX} \in \mathcal{D}_\delta$, let $C = |\mathcal{X}| + \frac{|\mathcal{U}||\mathcal{Y}||\mathcal{X}|}{\min_{a\in\mathcal{X}} P_X(a)}$ and let

$$\Gamma(P'_{UYX}) = \{P''_{UYX} : \|P''_{UYX} - P'_{UYX}\|_\infty < C\delta\}.$$

Then, $\bigcup_{P'_{UYX}: P'_{UY}=P_{UY}, P'_X=P_X} \Gamma(P'_{UYX})$ is an open cover of $\mathcal{D}_\delta$. For any $P''_{UYX} \in \mathcal{D}_\delta$, let $F : X \to X$ be a stochastic matrix such that $P_X = P_{UY} \cdot F''_{X|UY} \cdot F$, where $F''_{X|UY}$ is the stochastic matrix corresponding to P''_{UYX}. In addition, $\forall a \in \mathcal{X}$, if $P'_X(a) \leq P_X(a)$, then $F(a|a) = 1$, $F(b|a) = 0, \forall b \neq a$; else $F(a|a) = \frac{P_X(a)}{P'_X(a)} \geq \frac{P_X(a)}{P_X(a)+\delta/|\mathcal{X}|}$. In conclusion, F has the following property: $\forall a, b \in \mathcal{X}$, if $a = b$, then $0 \leq 1 - F(a|a) \leq \frac{\delta/|\mathcal{X}|}{\delta/|\mathcal{X}|+P_X(a)} \leq \frac{\delta}{P_X(a)|\mathcal{X}|}$; if $a \neq b$, then $0 \leq F(a|b) \leq \frac{\delta}{P_X(a)|\mathcal{X}|}$. Let P'_{UYX} be a distribution such that $P'_{UY} = P_{UY}$ and

$$P'_{X|UY}(x|u,y) = \sum_{x'\in\mathcal{X}} P''_{X|UY}(x'|u,y) F(x|x').$$

By construction, $P'_{UY} = P_{UY}$, $P'_X = P_X$. Moreover, for any $(u, y, x) \in \mathcal{U} \times \mathcal{Y} \times \mathcal{X}$,

$$\begin{aligned}&|P'_{UYX}(u,y,x) - P''_{UYX}(u,y,x)| \\ =&|P_{UY}(u,y)P'_{X|UY}(x|u,y) - P''_{UY}(u,y)P''_{X|UY}(x|u,y)| \\ \leq&P_{UY}(u,y)\sum_{x'\in\mathcal{X}} P''_{X|UY}(x'|u,y)\left|F(x|x') - \mathbf{1}_{\mathbf{x}=\mathbf{x}'}\right| + \left|(P_{UY}(u,y) - P''_{UY}(u,y))P''_{X|UY}(x|u,y)\right| \\ \leq&\frac{\delta}{\min_{a\in\mathcal{X}} P_X(a)} + \frac{\delta}{|\mathcal{U}||\mathcal{Y}|} = \frac{C\delta}{|\mathcal{U}||\mathcal{Y}||\mathcal{X}|}.\end{aligned}$$

$\mathcal{D}_\delta$ is closed and bounded. By the Heine-Borel Covering Theorem (Theorem 9.7.7 in [27]), $\mathcal{D}_\delta$ is a compact set, that is, every open cover of it has a finite sub-cover. So,

we can find a finite sub-cover:

$$\mathcal{D}_\delta \subseteq \bigcup_{t=1}^{N_\delta} \Gamma(P_{UYX}^{(\delta,t)}), \tag{A.29}$$

where $\forall t \in [N_\delta]$, $P_{UY}^{(\delta,t)} = P_{UY}$ and $P_X^{(\delta,t)} = P_X$. For any $x^n \in A_1(C) \cap E_1^c(C,n,\delta)$, the *type* of $(u^n(\tilde{i}(x^n)), y^n(\tilde{i}(x^n), \tilde{j}(x^n)), x^n)$ satisfies:

$$P_{u^n(\tilde{i}(x^n)),y^n(\tilde{i}(x^n),\tilde{j}(x^n)),x^n} \in \mathcal{D}_\delta.$$

By Inq. (A.29), there exists $t \in [N_\delta]$ s.t.

$$P_{u^n(\tilde{i}(x^n)),y^n(\tilde{i}(x^n),\tilde{j}(x^n)),x^n} \in \Gamma(P_{UYX}^{(\delta,t)}),$$

indicating that $(u^n(\tilde{i}(x^n)), y^n(\tilde{i}(x^n), \tilde{j}(x^n)), x^n)$ belongs to $\mathcal{T}^{(n,C\delta)}(P_{UYX}^{(\delta,t)})$. For convenience, let $S(C,n,\delta,t)$ denote the following set:

$$\begin{aligned} S(C,n,\delta,t) = \Big\{ & x^n : x^n \in A_1(C) \cap E_1^c(C,n,\delta), \\ & (u^n(\tilde{i}(x^n)), y^n(\tilde{i}(x^n), \tilde{j}(x^n)), x^n) \in \mathcal{T}^{(n,C\delta)}(P_{UYX}^{(\delta,t)}) \Big\}. \end{aligned}$$

Here, for fixed (C,n,δ), $\{S(C,n,\delta,t)\}_{t\in[N_\delta]}$ do not need to be pairwise disjoint. Using the assumption Inq. (A.28), we know that for any δ_k, there exists a probability distribution $P_{UYX}^{(\delta_k,t_k)}$ ($t_k \in [M_{\delta_k}]$) and a subsequence $\{n_{k,\tilde{\ell}}\}$ of $\{n_{k,\ell}\}$, such that:

$$E_{C(n_{k,\tilde{\ell}})}[\mathcal{P}^n(S(C(n_{k,\tilde{\ell}}), n_{k,\tilde{\ell}}, \delta_k, t_k))] \geq \frac{\alpha}{N_{\delta_k}}.$$

Now, we have a sequence of probability distribution $\{P_{UYX}^{(\delta_k,t_k)}\}_{k=1}^\infty$. We can always find a subsequence of it such that it converges to P_{UYX}^*. Without loss of generality, we can assume that

$$\lim_{k\to\infty} P_{UYX}^{(\delta_k,t_k)} = P_{UYX}^*. \tag{A.30}$$

Let Z be a random variable such that $P_{UYZ} = P_{UYX}^*$. Then, by construction

$$Ed(Y,Z) \geq D_1 + \epsilon; \quad Z \overset{d}{=} X.$$

For each k, apply Lemma 23 Part 2) with constant $C = |\mathcal{X}| + \frac{|\mathcal{U}||\mathcal{Y}||\mathcal{X}|}{\min_{a\in\mathcal{X}} P_X(a)}$, $R_1 = I(U;X) + \delta_k$, $R_2 = I(X;Y|U) + \delta_k$ and distribution $P_{UYX}^{(\delta_k,t_k)}$. Note that for any C and n,

$$S(C, n, \delta_k, t_k) \subseteq E_0^*(C, n, \delta_k, P_{UYX}^{(\delta_k,t_k)})$$

Then we have:

$$I(U;X) + \delta_k \geq I_{P_{UYX}^{(\delta_k,t_k)}}(U;X) - \xi(\delta_k, P_{UYX}^{(\delta_k,t_k)}); \tag{A.31}$$

$$I(X;Y) + 2\delta_k \geq I_{P_{UYX}^{(\delta_k,t_k)}}(U;X) + I_{P_{UYX}^{(\delta_k,t_k)}}(X;Y|U) - \xi(\delta_k, P_{UYX}^{(\delta_k,t_k)}). \tag{A.32}$$

By Eq. (A.30) and $\lim_{k\to\infty} \delta_k = 0$, we have:

$$\lim_{k\to\infty} \xi(\delta_k, P_{UYX}^{(\delta_k,t_k)}) = 0.$$

Since Inq. (A.31) and Inq. (A.32) hold for any k, let $k \to \infty$ and we have

$$I(U;X) \geq I(U;Z),$$

$$I(X;Y) \geq I(U;Z) + I(Z;Y|U).$$

This contradicts the last condition in Theorem 7.

A.3.3 Proof of Lemma 25

This proof is similar to the proof of Lemma 3 in [23]. Let

$$\mathscr{A} := \{(\tilde{\boldsymbol{Q}}, H_{\mathcal{R}}(\tilde{Z}|\tilde{U}) - H_{\mathcal{P}}(\tilde{X}|\tilde{U})) : (\tilde{U}, \tilde{Y}, \tilde{X}, \tilde{Z}) \in \mathscr{P}_{V,W}(\tilde{\boldsymbol{Q}})\}$$

Let $(\tilde{U}_i, \tilde{Y}_i, \tilde{X}_i, \tilde{Z}_i) \in \mathscr{P}_{V,W}(\tilde{\boldsymbol{Q}}_i)$, $i = 1, 2$ be two arbitrary quadruple of random variables and choose any $\alpha \in [0, 1]$. Define an random variable T such that T is independent from $(\tilde{U}_i, \tilde{Y}_i, \tilde{X}_i, \tilde{Z}_i)$, $i = 1, 2$ and takes the values 1 and 2 with probabilities α and $1 - \alpha$, respectively. Let $\tilde{U}_0 = (\tilde{U}_T, \tilde{T})$ and $\tilde{Y}_0 = \tilde{Y}_T$, $\tilde{X}_0 = \tilde{X}_T$ and $\tilde{Z}_0 = \tilde{Z}_T$. Then:

$$\tilde{U}_0 \to \tilde{Y}_0 \to (\tilde{X}_0, \tilde{Z}_0),$$

$$(\tilde{Y}_0, \tilde{X}_0, \tilde{Z}_0) \in \mathscr{P}_{V,W}.$$

The distribution of $\tilde{Y}_0$ is $\tilde{\mathbf{Q}}_0 = \alpha\tilde{\mathbf{Q}}_1 + (1-\alpha)\tilde{\mathbf{Q}}_2$. Thus, $(\tilde{U}_0, \tilde{Y}_0, \tilde{X}_0, \tilde{Z}_0) \in \mathscr{P}_{V,W}(\tilde{\mathbf{Q}}_0)$. By the independency of T, for $t = 1, 2$,

$$P(\tilde{X}_T = x|T = t, \tilde{U}_T = u) = P(\tilde{X}_t = x|T = t, \tilde{U}_t = u) = P(\tilde{X}_t = x|\tilde{U}_t = u).$$

Therefore,

$$\begin{aligned}
&H_{\mathcal{P}}(\tilde{X}_0|\tilde{U}_0) \\
=&\sum_{u_0} P(\tilde{U}_0 = u_0) \sum_x P_{\tilde{X}_0|\tilde{U}_0}(x|u_0) \cdot \log \frac{\mathcal{P}(x)}{P_{\tilde{X}_0|\tilde{U}_0}(x|u_0)} \\
=&\sum_{u\in\mathcal{U}} P(T=1, \tilde{U}_T = u) \sum_x \left\{ P(\tilde{X}_T = x|T=1, \tilde{U}_T = u) \cdot \log \frac{\mathcal{P}(x)}{P(\tilde{X}_T = x|T=1, \tilde{U}_T = u)} \right\} + \\
&\sum_{u\in\mathcal{U}} P(T=2, \tilde{U}_T = u) \sum_x \left\{ P(\tilde{X}_T = x|T=2, \tilde{U}_T = u) \cdot \log \frac{\mathcal{P}(x)}{P(\tilde{X}_T = x|T=2, \tilde{U}_T = u)} \right\} \\
=&\sum_{u\in\mathcal{U}} \alpha P_{\tilde{U}_1}(u) \sum_x \left\{ P_{\tilde{X}_1|\tilde{U}_1}(x|u) \cdot \log \frac{\mathcal{P}(x)}{P_{\tilde{X}_1|\tilde{U}_1}(x|u)} \right\} + \\
&\sum_{u\in\mathcal{U}} (1-\alpha) P_{\tilde{U}_2}(u) \sum_x \left\{ P_{\tilde{X}_2|\tilde{U}_1}(x|u) \cdot \log \frac{\mathcal{P}(x)}{P_{\tilde{X}_2|\tilde{U}_2}(x|u)} \right\} \\
=&\alpha H_{\mathcal{P}}(\tilde{X}_1|\tilde{U}_1) + (1-\alpha) H_{\mathcal{P}}(\tilde{X}_2|\tilde{U}_2).
\end{aligned}$$

Similarly,

$$H_{\mathcal{R}}(\tilde{Z}_0|\tilde{U}_0) = \alpha H_{\mathcal{R}}(\tilde{Z}_1|\tilde{U}_1) + (1-\alpha) H_{\mathcal{R}}(\tilde{Z}_2|\tilde{U}_2).$$

This proves that the plane region $\mathscr{A}$ is convex, indicating that $cl(\mathscr{A})$ is also convex. (The closure of a convex set is also convex.) Note that $T_{V,W,Q}(\tilde{\mathbf{Q}})$ is the lower boundary of $cl(\mathscr{A})$. Therefore, for any distributions $\tilde{\mathbf{Q}}_1$ and $\tilde{\mathbf{Q}}_2$ and any $\alpha \in [0, 1]$, we have

$$T_{V,W,Q}(\alpha\tilde{\mathbf{Q}}_1) + T_{V,W,Q}((1-\alpha)\tilde{\mathbf{Q}}_2) \geq T_{V,W,Q}(\tilde{\mathbf{Q}}_0),$$

where

$$\tilde{\mathbf{Q}}_0 = \alpha\tilde{\mathbf{Q}}_1 + (1-\alpha)\tilde{\mathbf{Q}}_2.$$

So, $T_{V,W,Q}(\tilde{\mathbf{Q}})$ is a convex function of $\tilde{\mathbf{Q}}$, and is thus a continuous function of $\tilde{\mathbf{Q}}$ (convexity implies continuity).

A.3.4 Proof of Proposition 26

The proof is similar to the proof of Proposition 2 in [22].

(1) $\Rightarrow$ (2). Choose a quadruple $(U, Y, X, Z) \in \mathscr{P}_{V,W}(Q)$, a $\delta > 0$ and an $\epsilon > 0$. Let P_U denote the marginal distribution of U and let P_{UY} denote the distribution of U and Y. For $n = 1, 2, \ldots$, write $\mathscr{B}_n = \{y^n : (u^n, y^n) \in T^{n,\epsilon}(P_{UY})\}$, where $u^n \in T^{n,\epsilon}(P_U)$ is some fixed sequence. By Definition 20, for any $0 < \eta < 1$ and any $\epsilon > 0$, when n is large enough,

$$\frac{1}{n} \log G_{V,Q}(\mathscr{B}_n, \eta) \leq \frac{1}{n} \log G_{W,Q}(\mathscr{B}_n, \eta) + \delta.$$

Since for n large enough and ϵ small enough, we have by Lemma 1 in [23],

$$\left| \frac{1}{n} \log G_{V,Q}(\mathscr{B}_n, \eta) - I(X;U) \right| \leq \delta,$$
$$\left| \frac{1}{n} \log G_{W,Q}(\mathscr{B}_n, \eta) - I(Z;U) \right| \leq \delta.$$

This implies that $I(U;Z) \leq I(U;X) + 3\delta$.

(2) $\Rightarrow$ (3). We use Lemma 3 in [1] to prove it.

Lemma 32 (Lemma 3 [1]) *Let $\mathcal{S}_N$ be the set of all probability N-vectors $\vec{p} = (p_1, ..., p_N)$ and let $f_j(\vec{p}), j = 1, ..., K$ be continuous function on $\mathcal{S}_N$. Then, for any probability measure μ on (the Borel subsets of) $\mathcal{S}_N$, there exist $K+1$ elements $\vec{p}_i$ of $\mathcal{S}_N$ and constants $\alpha_i \geq 0$, $i = 1, ..., K+1$ with $\sum_{i=1}^{K+1} \alpha_i = 1$, such that*

$$\int f_j(\vec{p}) d\mu = \sum_{i=1}^{K+1} \alpha_i f_j(\vec{p}_i), \quad j = 1, \ldots, K.$$

The proof is similar to the proof in [1]. Without loss of generality, we suppose that $\mathcal{Y} = [N]$ and $|\mathcal{Y}| = N$. We choose $\mathcal{S}_N$ as the set of all probability distributions on $\mathcal{Y}$. We can interpret $P_{Y|U}(\cdot|u)$ as an element of $\mathcal{S}_N$ and $\{P_U(u)\}_{u \in \mathcal{U}}$ as a Borel measure on $\mathcal{S}_N$.

Consider the following $N+1$ continuous functions on $\mathcal{S}_N$. For $\vec{p} = (p(1), ..., p(N)) \in \mathcal{S}_N$, set

1. $f_j(\vec{p}) = p(j)$, $j = 1, ..., N-1$;
2. $f_N(\vec{p}) = H(X) + \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} V(x|y)p(y))\cdot$
$\log\Big(\log \sum_{y\in\mathcal{Y}} V(x|y)p(y)\Big)$;
3. $f_{N+1}(\vec{p}) = H(Z) + \sum_{z\in\mathcal{Z}} \sum_{y\in\mathcal{Y}} W(z|y)p(y))\cdot$
$\log\Big(\log \sum_{y\in\mathcal{Y}} W(z|y)p(y)\Big)$.

Let $(U, Y, X, Z) \in \mathscr{P}_{V,W}(Q)$. Then for $j = 1, ..., N-1$

$$\sum_{u\in\mathcal{U}} f_j(P_{Y|U}(\cdot|u))P_U(u) = Q(j).$$

In addition,

$$\sum_{u\in\mathcal{U}} f_N(P_{Y|U}(\cdot|u))P_U(u) = I(U;X),$$
$$\sum_{u\in\mathcal{U}} f_{N+1}(P_{Y|U}(\cdot|u))P_U(u) = I(U;Z).$$

Lemma 32 implies that there exists a U^* with $|\mathcal{U}^*| \le |\mathcal{Y}|+2$ such that $I(U;X) = I(U^*;X)$ and $I(U;Z) = I(U^*;Z)$.

(3) $\Rightarrow$ (1). For any $0 < \eta < 1, \delta > 0$, we first choose an ϵ_δ such that:

$$H_2(\epsilon_\delta) + \epsilon_\delta|\mathcal{Y}| = \delta/8.$$

For any n, let $\mathscr{B}'_n \subseteq T^n(Q)$ achieve

$$\min_{\mathscr{B}\subseteq T^n(Q)} \frac{1}{n}[\log G_{W,Q}(\mathscr{B},\eta) - \log G_{V,Q}(\mathscr{B},\eta)].$$

By Theorem 1 in [23], we can select from $\mathscr{B}'_n$ a maximal ϵ_δ- code $\mathscr{B}_n$ for the channel V such that for all n sufficiently large ($n > n_0(\epsilon_\delta, \delta/16, \eta)$ where n_0 is given in Theorem 1

in [23])

$$\frac{1}{n}\log G_{V,Q}(\mathscr{B}_n,\eta) \geq |\mathscr{B}_n| - I(V,Q) - \delta/16 \tag{A.33}$$

$$\geq \frac{1}{n}\log G_{V,Q}(\mathscr{B}'_n,\eta) - \delta/8. \tag{A.34}$$

Here Inq. (A.33) follows from the converse part of the theorem and Inq. (A.34) follows from the direct part of the theorem.

Since $\mathscr{B}_n \subseteq \mathscr{B}'_n$,

$$\frac{1}{n}\log G_{W,Q}(\mathscr{B}_n,\eta) \leq \frac{1}{n}\log G_{W,Q}(\mathscr{B}'_n,\eta). \tag{A.35}$$

Let $\mathcal{P} = Q \cdot V$ and $\mathcal{R} = Q \cdot W$. By Lemma 5 in [23], for all large n (depending on η, δ),

$$\frac{1}{n}\log G_{V,Q}(\mathscr{B}_n,\eta) \leq \frac{1}{n}\left(H_{\mathcal{P}^n}(\hat{X}^n) + H(\hat{Y}^n|\hat{X}^n)\right) + \delta/8, \tag{A.36}$$

$$\frac{1}{n}\log G_{W,Q}(\mathscr{B}_n,\eta) \geq \frac{1}{n}H_{\mathcal{R}^n}(\hat{Z}^n) - \delta/8, \tag{A.37}$$

where the random variables $(\hat{Y}^n, \hat{X}^n, \hat{Z}^n) \in \mathscr{P}^n_{V,W}$ are defined by

$$Pr(\hat{Y}^n = y^n) = \begin{cases} Q^n(y^n)/Q^n(\mathscr{B}_n), & \text{if } y^n \in \mathscr{B}_n \\ 0, & \text{otherwise.} \end{cases}$$

Since $\mathscr{B}_n$ is an ϵ_δ code for the channel V, from Fano's inequality, we have:

$$\frac{1}{n}H(\hat{Y}^n|\hat{X}^n) \leq \frac{1}{n}H_2(\epsilon_\delta) + \epsilon_\delta|\mathcal{Y}| \leq \delta/8. \tag{A.38}$$

By Inq. (4.11) – (4.13) in [23], we know that there exists a distribution $\tilde{Q}_n$ on $\mathcal{Y}$ satisfying

$$|\tilde{Q}_n(y) - Q(y)| \leq \frac{r_n}{n}, \quad \forall y \in \mathcal{Y},$$

and a quadruple of random variables $(\tilde{U}_n, \tilde{Y}_n, \tilde{X}_n, \tilde{Z}_n) \in \mathscr{P}_{V,W}(\tilde{Q}_n)$ such that:

$$\frac{1}{n}H_{\mathcal{R}^n}(\hat{Z}^n) - \frac{1}{n}H_{\mathcal{P}^n}(\hat{X}^n) = H_{\mathcal{R}}(\tilde{Z}_n|\tilde{U}_n) - H_{\mathcal{P}}(\tilde{X}_n|\tilde{U}_n). \tag{A.39}$$

Combining Inq. (A.34) to (A.39), we know that for all large n (depending on η, δ)

$$\begin{aligned}
&\frac{1}{n}\log G_{W,Q}(\mathscr{B}'_n,\eta)-\frac{1}{n}\log G_{V,Q}(\mathscr{B}'_n,\eta)\\
\geq&\frac{1}{n}\log G_{W,Q}(\mathscr{B}_n,\eta)-\left(\frac{1}{n}\log G_{V,Q}(\mathscr{B}_n,\eta)+\delta/8\right)\\
\geq&\left(\frac{1}{n}H_{\mathcal{R}^n}(\hat{Z}^n)-\delta/8\right)-\left(\frac{1}{n}\left(H_{\mathcal{P}^n}(\hat{X}^n)+H(\hat{Y}^n|\hat{X}^n)\right)+\delta/8\right)-\delta/8\\
\geq&\frac{1}{n}H_{\mathcal{R}^n}(\hat{Z}^n)-\frac{1}{n}H_{\mathcal{P}^n}(\hat{X}^n)-\delta/2\\
=&H_{\mathcal{R}}(\tilde{Z}_n|\tilde{U}_n)-H_{\mathcal{P}}(\tilde{X}_n|\tilde{U}_n)-\delta/2\\
\geq&T_{V,W,Q}(\tilde{Q}_n)-\delta/2.
\end{aligned}$$

The third statement of the proposition implies that $T_{V,W,Q}(Q)\geq 0$. Since $\|\tilde{Q}_n-Q\|_\infty\leq\frac{r_n}{n}$, combining Lemma 25, for all n sufficiently large,

$$T_{V,W,Q}(\tilde{Q}_n)\geq T_{V,W,Q}(Q)-\delta/2\geq-\delta/2.$$

Thus, for any $0<\eta<1$ and any $\delta>0$, for all large n (depending on η,δ),

$$\min_{\mathscr{B}\subseteq T^n(Q)}\frac{1}{n}[\log G_{W,Q}(\mathscr{B},\eta)-\log G_{V,Q}(\mathscr{B},\eta)]\geq-\delta.$$

This completes the proof.

A.3.5 Proof of Lemma 27

By Proposition 26, it is equivalent to prove that $I(U;Z)\leq I(U;X)$ for any $(U,Y,X,Z)\in\mathscr{P}_{V,W}(Q)$. For any $(U,Y,X,Z)\in\mathscr{P}_{V,W}(Q)$, let P_{UYXZ} denote its distribution. Let $F_{U|Y}$ denote the stochastic matrix describing the conditional distribution of U given Y. $F_{Y|U}$ is defined analogously. Then, let (U_n,Y_n,X_n,Z_n) be a quadruple of random variables with joint distribution $P_{U_nY_nX_nZ_n}=P_{U_nY_n}P_{X_n|Y_n}P_{Z_n|Y_n}$. Here, $P_{U_nY_n}=P_{Y_n}P_{U_n|Y_n}$ where $P_{Y_n}=Q_n$ and $P_{U_n|Y_n}=P_{U|Y}$. $P_{X_n|Y_n}$ and $P_{Z_n|Y_n}$ are described by V_n and W_n, respectively. Since

$V_n \to V$, $W_n \to W$ and $Q_n \to Q$, we have:

$$\lim_{n\to\infty} P_{U_nY_nX_n} = P_{UYX}; \quad \lim_{n\to\infty} P_{U_nY_nZ_n} = P_{UYZ}.$$

It is easy to check that $(U_n, Y_n, X_n, Z_n) \in \mathscr{P}_{V_n,W_n}(Q_n)$. Since $W_n \overset{Q_n}{\ll} V_n$, via Proposition 26, $I(U_n; Z_n) \leq I(U_n; X_n)$. Note that mutual information $I(X; Y)$ is a continuous function of the distribution of (X, Y). So

$$I(U; Z) = \lim_{n\to\infty} I(U_n; Z_n) \leq \lim_{n\to\infty} I(U_n; X_n) = I(U; X).$$

BIBLIOGRAPHY

[1] R. Ahlswede and J. Korner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, 21(6):629–637, 1975.

[2] E. Ahmed and A. B. Wagner. Lossy source coding with byzantine adversaries. In *Proc. Information Theory Workshop*, pages 462–466. IEEE, 2011.

[3] E. Ahmed and A. B. Wagner. Lossy source coding with Byzantine adversaries. In *Proc. IEEE ITW*, pages 462–466, oct 2011.

[4] E. Ahmed and A. B. Wagner. Erasure multiple descriptions. *IEEE Trans. Inf. Theory*, 58(3):1328–1344, 2012.

[5] E. Ahmed and A. B. Wagner. Coding for the large-alphabet adversarial channel. *IEEE Trans. Inf. Theory*, to appear.

[6] T. Berger. Rate distortion theory: A mathematical basis for data compression. 1971.

[7] N. Cai and R. W. Yeung. Network error correction, ii: Lower bounds. *Communications in Information & Systems*, 6(1):37–54, 2006.

[8] P. A. Chou and Z. Miao. Rate-distortion optimized streaming of packetized media. *IEEE Trans. Multimedia*, 8(2):390–404, 2006.

[9] John Horton Conway and Neil James Alexander Sloane. *Sphere packings, lattices and groups*, volume 290. Springer Science & Business Media, 2013.

[10] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[11] I. Csiszar and J. Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[12] A. El Gamal and T. M. Cover. Information theory of multiple descriptions. In *Shannon Theory Workshop, Mt. Kisco, NY*, 1979.

[13] W. H. Equitz and T. M. Cover. Successive refinement of information. *IEEE Transactions on Information Theory*, 37(2):269–275, 1991.

[14] X. Fan, A. B. Wagner, and E. Ahmed. Polytope codes for large-alphabet channels. In *Proc. Allerton Conference on Communication, Control, and Computing*, pages 948–955, 2013.

[15] A. E. Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Inf. Theory*, 28(6):851–857, 1982.

[16] E. N. Gilbert. A comparison of signalling alphabets. *Bell System Technical Journal*, 31(3):504–522, 1952.

[17] O. Goldreich. *Foundations of cryptography: volume 2, basic applications*. Cambridge university press, 2009.

[18] V. K. Goyal. Multiple description coding: Compression meets the network. *Signal Processing Magazine, IEEE*, 18(5):74–93, 2001.

[19] T. Ho and D. Lun. *Network coding: an introduction*, volume 6. Cambridge University Press Cambridge, 2008.

[20] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Médard. Resilient network coding in the presence of Byzantine adversaries. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pages 616–624. IEEE, 2007.

[21] A. Kanlis, S. Khudanpur, and P. Narayan. Typicality of a good rate-distortion code. *Problems of Information Transmission*, 32(1):96–103, 1996.

[22] J. Körner and K. Marton. Comparison of two noisy channels. *Topics in information theory*, (16), 1977.

[23] J. Korner and K. Marton. Images of a set via two channels and their role in multi-user communication. *IEEE Transactions on Information Theory*, 23(6):751–761, 1977.

[24] O. Kosut. Polytope codes for distributed storage in the presence of an active omniscient adversary. In *Proc. International Symposium on Information Theory*, pages 897–901. IEEE, 2013.

[25] Oliver Kosut and Lang Tong. The quadratic gaussian ceo problem with byzantine agents. In *Proc. International Symposium on Information Theory*, pages 1145–1149. IEEE, 2009.

[26] Oliver Kosut, Lang Tong, and NC David. Polytope codes against adversaries in networks. *IEEE Transactions on Information Theory*, 60(6):3308–3344, 2014.

[27] P. A. Loeb. *Real Analysis*. Springer International Publishing, 2016.

[28] L. Ozarow. On a source-coding problem with two channels and three receivers. *The Bell System Technical Journal*, 59(10):1909–1921, 1980.

[29] S. S. Pradhan, R. Puri, and K. Ramchandran. n-channel symmetric multiple descriptions-part i: (n, k) source-channel erasure codes. *IEEE Transactions on Information Theory*, 50(1):47–61, 2004.

[30] R. Puri, S. S. Pradhan, and K. Ramchandran. n-channel symmetric multiple descriptions-part ii: An achievable rate-distortion region. *IEEE Transactions on Information Theory*, 51(4):1377–1392, 2005.

[31] R. Puri and K. Ramchandran. Multiple description source coding using forward error correction codes. In *Signals, Systems, and Computers, 1999. Conference Record of the Thirty-Third Asilomar Conference on*, volume 1, pages 342–346. IEEE, 1999.

[32] I. S Reed and G. Solomon. Polynomial codes over certain finite fields. *Journal of the society for industrial and applied mathematics*, 8(2):300–304, 1960.

[33] T. Richardson and R. Urbanke. *Modern coding theory*. Cambridge University Press, 2008.

[34] B. Rimoldi. Successive refinement of information: Characterization of the achievable rates. *IEEE Transactions on Information Theory*, 40(1):253–259, 1994.

[35] J. R. Roche, R. W. Yeung, and K. P. Hau. Symmetrical multilevel diversity coding. *IEEE Transactions on Information Theory*, 43(3):1059–1064, 1997.

[36] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record*, 4(142-163):1, 1959.

[37] R. Singleton. Maximum distance-nary codes. *IEEE Trans. Inf. Theory*, 10(2):116–118, 1964.

[38] J. J. Sylvester. Lx. thoughts on inverse orthogonal matrices, simultaneous signsuccessions, and tessellated pavements in two or more colours, with applications to

newton's rule, ornamental tile-work, and the theory of numbers. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 34(232):461–475, 1867.

[39] C. Tian, S. Mohajer, and S. N. Diggavi. Approximating the gaussian multiple description rate region under symmetric distortion constraints. *IEEE Transactions on Information Theory*, 55(8):3869–3891, 2009.

[40] Stephen B. Wicker. *Error Control Systems for Digital Communication and Storage*. Prentice Hall, 1995.

[41] H. Yao, D. Silva, S. Jaggi, and M. Langberg. Network codes resilient to jamming and eavesdropping. *IEEE/ACM Transactions on Networking (TON)*, 22(6):1978–1987, 2014.

[42] R. W. Yeung. Multilevel diversity coding with distortion. *IEEE Transactions on Information Theory*, 41(2):412–422, 1995.

[43] R. W. Yeung and N Cai. Network error correction, i: Basic concepts and upper bounds. *Communications in Information & Systems*, 6(1):19–35, 2006.

[44] R. W. Yeung and Z. Zhang. On symmetrical multilevel diversity coding. *IEEE Transactions on Information Theory*, 45(2):609–621, 1999.

[45] R. Zamir and R. W. Yeung. Multilevel diversity coding via successive refinement. In *Proc. International Symposium on Information Theory*, page 265. IEEE, 1997.

[46] Z. Zhang and T. Berger. New results in binary multiple descriptions. *IEEE Transactions on Information Theory*, 33(4):502–521, 1987.