Error Exponent for Discrete Memoryless Multiple-Access Channels

by

Ali Nazari

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2011

Doctoral Committee:

Associate Professor Sandeep Pradhan, Co-Chair
Associate Professor Achilleas Anastasopoulos, Co-Chair
Professor David Neuhoff
Associate Professor Erhan Bayraktar
Associate Professor Jussi Keppo

To my parents.

ACKNOWLEDGEMENTS

It is my pleasure to thank the many people who made this thesis possible. I would like to sincerely thank my research advisors Professor Sandeep Pradhan and Professor Achilleas Anastasopoulos for their continuous support and encouragement, for the opportunity they gave me to conduct independent research, and for their exemplary respect. Even when I became frustrated with research, meetings with them always left me with new ideas and renewed optimism. Their high standards in research, creativity and insistence on high-level understanding of a problem are qualities I hope to emulate in my own career.

I would also like to thank my dissertation committee members, Professor David Neuhoff, Professor Erhan Bayraktar, and Professor Jussi Keppo for accepting to be on my dissertation committee. I am grateful to Professor David Neuhoff for his interest in my research and for sharing his deep knowledge. I am indebted to Professor Bayraktar and Professor Keppo for introducing Mathematical Finance to me, which is related to my future career. I feel grateful to Professor Jussi Keppo for accepting to be on my dissertation committee despite the fact that he was going to be out of the United States at the time of my defense.

The supportive work and efforts of the staff members of my department are hereby acknowledged. I wish to thank Becky Turanski, Nancy Goings, Michele Feldkamp, Ann Pace, Karen Liska and Beth Lawson for efficiently and cheerfully helping me deal with myriad administrative matters. I would also like to sincerely thank all of my friends in the Electrical Engineering Department for their invaluable friendship.

I would like to thank all of my Iranian friends, especially Dr. Alireza Tabatabaee and Dr. Javid Moraveji for their invaluable friendship and support since the very first days I came to the United States. I have been lucky to have had wonderful friends Morteza Nick, Sepehr Entezari and Alexandra Tate. I would especially like to thank Alexandra for revising all of my papers over the past couple of years.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES vii

LIST OF TABLES viii

CHAPTERS

1 Introduction 1
  1.1 Dissertation overview 6

2 Background: Error Exponent 9
  2.1 Preliminaries 9
  2.2 Summary of Known Results 13
    2.2.1 Capacity Region for DM-MAC 13
    2.2.2 Known Bounds on the Error Exponents of DM-MAC 15

3 Lower Bounds on the Error Exponent of Multiple-Access Channels . 19
  3.1 Point to Point: Lower Bounds on reliability function 22
    3.1.1 Packing functions 22
    3.1.2 Relation between packing function and probability of error 24
    3.1.3 Packing Lemmas 28
    3.1.4 Error Exponent Bounds 30
  3.2 MAC: Lower Bounds on Error Exponent 34
    3.2.1 Definition of Packing Functions 35
    3.2.2 Relation between probability of error and packing functions 36
    3.2.3 Definition of Information Functions 42
    3.2.4 Packing Lemmas 46
    3.2.5 Error exponent bounds 49
  3.3 Numerical result 52
  3.4 Proof of Theorems 53

3.4.1 Point to Point Proofs 53
3.4.2 MAC Proofs 65

4 Typicality Graphs and Their Properties 89
4.1 Preliminaries 92
4.2 Typicality graphs 94
4.3 Sub-graphs contained in typicality graphs 95
4.3.1 Subgraphs of general degree 95
4.3.2 Nearly complete subgraphs 96
4.3.3 Nearly Empty Subgraphs 98
4.4 Proof of Theorems 100

5 Upper Bounds on the Error Exponent of Multiple-Access Channels . 116
5.1 Preliminaries 118
5.2 Sphere Packing Bound on the Average Error Exponent (Method of Types Technique) 120
5.3 Sphere Packing Bound on the Average Error Exponent (Strong Converse Technique) 123
5.3.1 Point to Point Case 123
5.3.2 MAC Case 124
5.4 A Minimum Distance on the Maximal Error Exponent 126
5.4.1 A Conjectured Tighter Upper Bound 128
5.5 The Maximal Error Exponent vs. The Average Error Exponent 131
5.6 Proof of Theorems 133
5.6.1 Proof of Theorem 5.2.1 133
5.6.2 Proof of Theorem 5.3.1 137
5.6.3 Proof of Fact 5.3.1 140
5.6.4 Proof of Theorem 5.3.2 142
5.6.5 Proof of Theorem 5.3.3 151
5.6.6 Proof of Theorem 5.4.1 154
5.6.7 Proof of Theorem 5.5.1 164

6 Conclusions 166

BIBLIOGRAPHY 169

LIST OF FIGURES

Figure

1.1 Block diagram of communication systems. 2
1.2 Upper and Lower bounds on the error exponent for a DMC. 4
1.3 A schematic of two-user multiple-access channel. 5
2.1 Achievable rates for a fixed channel input distribution P_{UXY}. 14
3.1 Lower bounds on the reliability function for point-to-point channel (random coding $-\cdot$, typical random coding $-$, expurgated $--$. 20

LIST OF TABLES

Table

3.1 Channel Statistics 53
3.2 E_{ex} vs. E_r^{Liu} 54

CHAPTER 1

Introduction

Information theory deals primarily with systems transmitting information or data from one point to another. A rather general block diagram for visualizing the behavior of such systems is given in Figure 1.1. The source output in Figure 1.1 might represent, for example, a voice waveform, the output of a set of sensors or a sequence of binary digits from a magnetic tape. The channel might represent a telephone line, a communication link or a high frequency radio link. The encoder represents any processing of the source output performed prior to transmission. The decoder represents the processing of the channel output with the objective of producing an acceptable replica of the source output at the destination. In coding theory, block codes are one of the two common types of channel codes (the other one being convolutional codes), which enable reliable transmission of digital data over unreliable communication channels subject to channel noise. A block code transforms a message i consisting of a sequence of information symbols over an alphabet $\mathcal{X}$ into a fixed-length sequence $\mathbf{x}_i$ of n encoding symbols, called a codeword. The set of all codewords is called a codebook. For a codebook with M codewords of block length n, the transmission rate, R, is defined as $R = \frac{1}{n} \log M$.

In the early 1940's, it was thought impossible to send information at a positive rate with negligible probability of error. C. E. Shannon surprised the communications theory society by presenting a theory for data transmission over noisy channels and proving that probability of error could be made nearly zero for all transmission rates

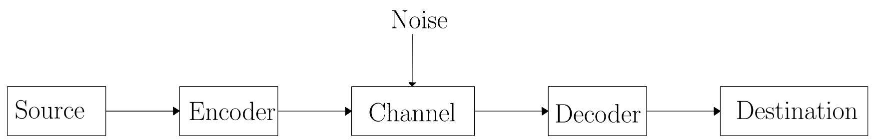

Figure 1.1: Block diagram of communication systems.

below channel capacity. However, Shannon's channel coding theorem is of asymptotic nature; it states that for any transmission rate below the channel capacity, the probability of the error of the channel code can be made arbitrary small as the block length becomes large enough. This theorem does not indicate how large the block length must be in order to achieve a specific error probability. Furthermore, in practical situations, there are limitations on the delay of the communication and the block length of the code cannot be arbitrarily large. Hence, it is important to study how the probability of error drops as a function of block length. A partial answer to this question is provided by examining the error exponent of the channel which is defined as the rate of exponential decay of the probability of error as a function of the block length. It is well-known that the optimum error exponent $E(R)$, at some fixed transmission rate R, (also known as the channel reliability function) gives the decoding error probability exponential rate of decay as a function of block-length for the best sequence of codes.

Error exponents have been meticulously studied for point to point discrete memoryless channels (DMCs) in the literature [1, 17, 22, 23, 25, 45, 46]. Lower and upper bounds on the channel reliability function for the DMC are known. A lower bound, known as the random coding exponent $E_r(R)$, was developed by Fano [23] by upper-bounding the average error probability over an ensemble of codes. This bound is loose at low rates. Later, Gallager [27] considerably reduced the mechanics of developing this bound. Gallager [29] also demonstrated that the random coding bound is the true average error exponent for the random code ensemble. This result illustrates that the weakness of the random coding bound, at low rates, is not due to upper-bounding the ensemble average. Rather, this weakness is due to the fact that the best codes perform much better than the average, especially at low rates. Two upper bounds, known as sphere packing exponent $E_{sp}(R)$ and minimum distance exponent $E_{md}(R)$ were developed by Shannon, Gallager, and Berlekamp [45, 46]. The random coding bound and the sphere packing bound turn out to be equal for code rates greater than a certain value R_{crit}, but are distinctly different at lower rates. Gallager [27] partly closed this gap from below by introducing a technique to purge poor codewords from a random code. This resulted in a new lower bound, the expurgated bound, which is an improvement over the random coding bound at low rates [13, 26, 28]. The expurgated bound, $E_{ex}(R)$, coincides with the minimum distance bound, $E_{md}(R)$, at $R = 0$ [16, pg. 189]. Shannon, Gallager, and Berlekamp [46] further closed this gap from above by combining the minimum distance bound with the sphere packing bound. They proved that a straight line connecting any two points of $E_{sp}(R)$ and $E_{md}(R)$ is an error exponent upper bound. This procedure resulted in a new upper bound, the straight line bound, which is an asymptotic improvement over the sphere packing bound at low rates. Barg and Forney [8] investigated another lower bound for the binary symmetric channel (BSC), called the "typical" random coding bound $E_T(R)$. The authors showed that almost all codes in the standard random coding ensemble exhibit a performance that is as good as the one described by the typical random coding bound. In addition, they showed that the typical error exponent is larger than the random coding exponent and smaller than the expurgated exponent at low rates. Figure 1.2 shows all the upper and lower bounds on the reliability function for a DMC. As we can see in this Figure, the error exponent lies inside the shaded region for all transmission rates below the critical rate.

In information theory, the Hamming distance between two strings of equal length is defined as the number of positions at which the corresponding symbols are different. In the special case of binary codes, extensive study has been devoted not only to bounds on the probability of decoding error but also to bounds on the minimum Hamming distance. The asymptotically best lower bound on the minimum distance was derived by Gilbert [31]. For many years, the asymptotically best upper bound on the minimum distance was the bound first given in an unpublished work by Elias [9], subsequently improved by Welch, et al. [35] and Levenshtein [36]. The best upper

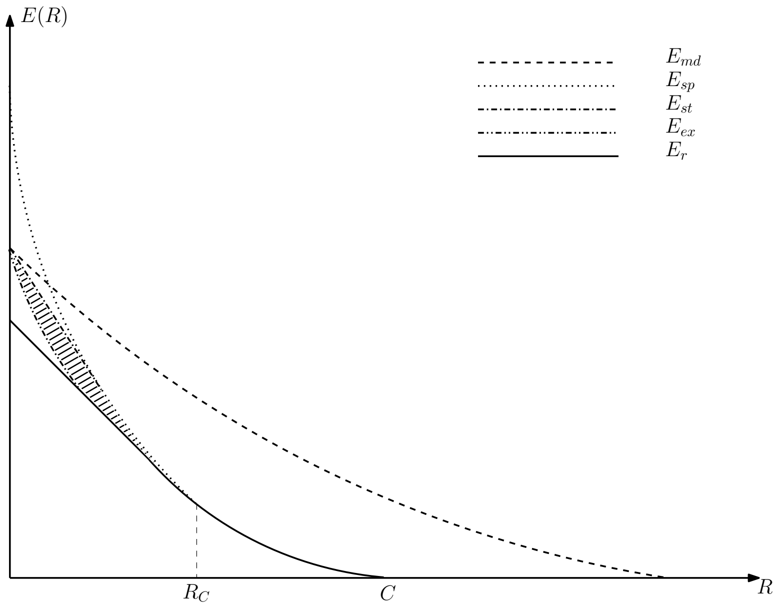

Figure 1.2: Upper and Lower bounds on the error exponent for a DMC.

and lower bounds remain asymptotically different, so the actual asymptotic behavior of the best obtainable minimum Hamming distance remains unanswered.

Much recent work on communication aspects of information theory has concentrated on network information theory: the theory of simultaneous rates of communication from many senders to many receivers in the presence of interference and noise. Examples of large communication networks include computer networks, satellite networks and phone systems. A complete theory of network information would have wide implications for the design of communication and computer networks. In this thesis, we concentrate on a communication model, in which two transmitters wish to reliably communicate two independent messages to a single receiver. This model is known as a *Multiple-Access Channel*. A schematic is depicted in Figure 1.3. A common example of this channel is a satellite receiver with many independent ground stations as transmitters, or a set of cell phones transmitting to a base station. In this model, the senders must contend not only with the receiver noise but with interference from each other as well.

The capacity region is defined as the closure of the set of all input rates that the network can stably support. The first attempt to calculate capacity regions for

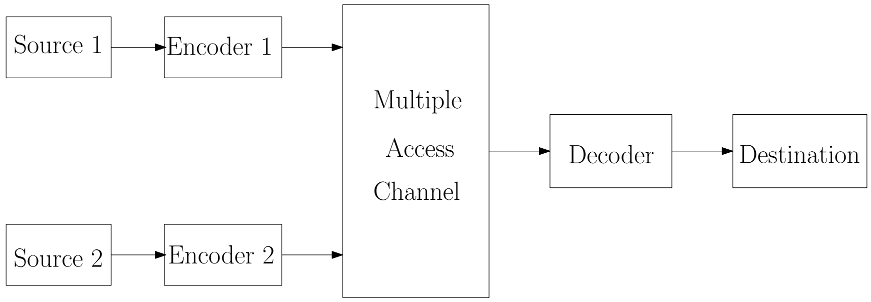

Figure 1.3: A schematic of two-user multiple-access channel.

multi-user systems, systems with more than one transmitter or receiver, were made by Shannon in his fundamental paper [44]. The capacity region for discrete memoryless multiple-access channels was found by Ahlswede in [3] and Liao in [37]. A symmetric characterization of the region was given by Ahlswede, in [4]. In their coding theorem, they proved that for any rate pair in the interior of a certain set $\mathcal{C}_{av}$, and for all sufficiently large block length, there exists a multiuser code with an arbitrary small average probability of error. Conversely, for any rate pair outside of $\mathcal{C}_{av}$, the average probability of error is bounded away from 0. Regarding discrete memoryless multiple-access channels (DM-MACs), stronger versions of Ahlswede and Liao's coding theorem, giving exponential upper and lower bounds for the error probability, were derived by several authors. Slepian and Wolf [47], Dyachkov [20], Gallager [30], Pokorny and Wallmeier [41], and Liu and Hughes [38] studied random coding bounds on the average error exponent. Haroutunian [33] studied a sphere packing bound on the error probability.

Comparing the state of the art in the study of error exponents for DMCs and DM-MACs, we observe that the latter is much less advanced. We believe the main difficulty in the study of error exponents for DM-MACs is due to the fact that error performance in a DM-MAC depends on the pair of codebooks (in the case of a two-user MAC) used by the two transmitters, while at the same time, each transmitter can only control its own codebook. This simple fact has important consequences. For instance, expurgation has not been studied in MAC, because by eliminating some of the "bad" codeword pairs, we may end up with a set of correlated input sequences which is hard to analyze.

1.1 Dissertation overview

This dissertation has four main chapters along with this Introduction chapter and a conclusion statement.

In Chapter 3, we study lower bounds on the average error exponent of DM-MACs. First, we present a unified framework to obtain all known lower bounds (random coding, typical random coding and expurgated bound) on the reliability function of a point-to-point DMC. By using a similar idea for a two-user discrete DM-MAC, three lower bounds on the reliability function are derived. The first one (random coding) is identical to the best known lower bound on the reliability function of DM-MAC [38]. It is shown that the random coding bound is the performance of the average code in the constant composition code ensemble. The second bound (typical random coding) is the typical performance of the constant composition code ensemble. To derive the third bound (expurgated), we eliminate some of the codewords from the codebook with larger rate. This is the first bound of this type that explicitly uses the method of expurgation for MACs. It is shown that the exponent of the typical random coding and the expurgated bounds are greater than or equal to the exponent of the known random coding bounds for all rate pairs. Moreover, an example is given where the exponent of the expurgated bound is strictly larger. These bounds can be universally obtained for all discrete memoryless MACs with given input and output alphabets.

The concept of typicality and typical sequences is central to information theory. It has been used to develop computable performance limits for several communication problems. In Chapter 4, we formally introduce and characterize the typicality graph and investigate some subgraph containment problems. The typicality graphs provide a strong tool in studying a variety of multiuser communication problems. Transmitting correlated information over a MAC, transmitting correlated information over a broadcast channel and communicating over a MAC with feedback, are three problems in which the properties of typicality graphs play a crucial role. The evaluation of performance limits of a multiuser communication problem can be thought of as characterizing certain properties of typicality graphs of random variables associated with the problem. The techniques used to study the typicality graph is applied in Chapter 5 to develop tighter bounds on the error exponents of discrete memoryless multiple-access channels.

In Chapter 5, we study two new upper bounds on the error exponent of a two-user discrete memoryless multiple-access channel. The first bound (sphere packing) is an upper bound on the average error exponent, while the second one (minimum distance) is valid only for the maximal error exponent. To derive the sphere packing bound, first, we revisit the point-to-point case and examine the techniques used for obtaining the sphere bound on the optimum error exponent. By using a similar approach for two-user DM-MACs, we develop a sphere packing bound on the average error exponent of such channels. This bound outperforms the known sphere packing bound derived by Haroutunain [33]. This is the first bound of its type that explicitly imposes independence of the users' input distributions (conditioned on the time-sharing auxiliary variable) and, thus, results in tighter sphere-packing exponents when compared to the tightest known sphere packing exponent in [33]. We also describe a simpler derivation of the Haroutunian's sphere packing bound and we show that we can easily make it tighter by using the properties of the typicality graphs we obtained in Chapter 4. We, furthermore, derive an upper bound (minimum distance) on the maximal error exponent for DM-MACs. To obtain this bound, first, an upper bound on the minimum Bhattacharyya distance between codeword pairs of any multi-user code is derived. For a certain large class of two-user (DM) MACs, an upper bound on the maximal error exponent is derived as a consequence of the upper bound on Bhattacharyya distance. This bound is tighter than the sphere packing bound at low transmission rates. Using a conjecture about the structure of the typicality graph, a tighter minimum distance bound for the maximal error exponent is derived and is shown to be tight at zero rates. Finally, the relationship between average and maximal error probabilities for a two user (DM) MAC is studied. As a result, a method to derive new bounds on the average/maximal error exponent by using known bounds on the maximal/average one is obtained.

CHAPTER 2

Background: Error Exponent

2.1 Preliminaries

For any alphabet $\mathcal{X}$, $\mathcal{P}(\mathcal{X})$ denotes the set of all probability distributions on $\mathcal{X}$. The *type* of a sequence $\mathbf{x} = (x_1, ..., x_n) \in \mathcal{X}^n$ is the distributions $P_{\mathbf{x}}$ on $\mathcal{X}$ defined by

$$P_{\mathbf{x}}(x) \triangleq \frac{1}{n} N(x|\mathbf{x}), \qquad\qquad x \in \mathcal{X}, \tag{2.1}$$

where $N(x|\mathbf{x})$ denotes the number of occurrences of x in $\mathbf{x}$. Let $\mathcal{P}_n(\mathcal{X})$ denote the set of all types in $\mathcal{X}^n$, and define the set of all sequences in $\mathcal{X}^n$ of type P as

$$T_P \triangleq \{\mathbf{x} \in \mathcal{X}^n : P_{\mathbf{x}} = P\}. \tag{2.2}$$

The joint type of a pair $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ is the probability distribution $P_{\mathbf{x},\mathbf{y}}$ on $\mathcal{X} \times \mathcal{Y}$ defined by

$$P_{\mathbf{x},\mathbf{y}}(x, y) \triangleq \frac{1}{n} N(x, y|\mathbf{x}, \mathbf{y}), \qquad\qquad (x, y) \in \mathcal{X} \times \mathcal{Y}, \tag{2.3}$$

where $N(x, y|\mathbf{x}, \mathbf{y})$ is the number of occurrences of (x, y) in $(\mathbf{x}, \mathbf{y})$. The relative entropy or *I-divergence* between probability distributions $P \in \mathcal{P}(\mathcal{X})$ and $Q \in \mathcal{P}(\mathcal{X})$

is defined as

$$D(P||Q) \triangleq \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}. \tag{2.4}$$

Let $\mathcal{W}(\mathcal{Y}|\mathcal{X})$ denote the set of all stochastic matrices with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. Then, given stochastic matrices V, $W \in \mathcal{W}(\mathcal{Y}|\mathcal{X})$, the conditional *I-divergence* is defined by

$$D(V||W|P) \triangleq \sum_{x\in\mathcal{X}} P(x)D\left(V(\cdot|x)||W(\cdot|x)\right). \tag{2.5}$$

Definition 2.1.1. *A discrete memoryless channel (DMC) is defined by a stochastic matrix $W : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$, the input alphabet, and $\mathcal{Y}$, the output alphabet, are finite sets. The channel transition probability for n-sequences is given by*

$$W^n(\mathbf{y}|\mathbf{x}) \triangleq \prod_{i=1}^{n} W(y_i|x_i),$$

where $\mathbf{x} \triangleq (x_1, ..., x_n) \in \mathcal{X}^n$, $\mathbf{y} \triangleq (y_1, ..., y_n) \in \mathcal{Y}^n$. An (n, M) code for a given DMC, W, is a set $C = \{(\mathbf{x}_i, D_i) : 1 \le i \le M\}$ with (a) $\mathbf{x}_i \in \mathcal{X}^n$, $D_i \subset \mathcal{Y}^n$ and (b) $D_i \cap D_{i'} = \varnothing$ for $i \neq i'$. The transmission rate, R, for this code, is defined as $R = \frac{1}{n}\log M$.

When message i is transmitted, the conditional probability of error of code C is given by

$$e_i(C, W) \triangleq W^n(D_i^c|\mathbf{x}_i). \tag{2.6}$$

The average probability of error for this code is defined as

$$e(C, W) \triangleq \frac{1}{M}\sum_{i=1}^{M} e_i(C, W), \tag{2.7}$$

and the maximal probability of error is defined as

$$e_m(C,W) \triangleq \max_i W^n(D_i^c|\mathbf{x}_i). \tag{2.8}$$

An (n,M,λ) *code for* $W : \mathcal{X} \to \mathcal{Y}$, *is an* (n,M) *code* C *with* $e_m(C,W) \leq \lambda$. *The average and maximal error exponents, at rate* R, *are defined as:*

$$E_{av}^*(R) \triangleq \limsup_{n\to\infty} \max_{C\in\mathcal{C}} -\frac{1}{n}\log e(C,W), \tag{2.9}$$

$$E_m^*(R) \triangleq \limsup_{n\to\infty} \max_{C\in\mathcal{C}} -\frac{1}{n}\log e_m(C,W), \tag{2.10}$$

where $\mathcal{C}$ *is the set of all codes of length* n *and rate* R. *The typical average error exponent of an ensemble* $\mathcal{C}$, *at rate* R, *is defined as:*

$$E_{av}^T(R) \triangleq \liminf_{\delta\to 0} \limsup_{n\to\infty} \max_{\tilde{\mathcal{C}}\subset\mathcal{C}:\mathbb{P}(\tilde{\mathcal{C}})>1-\delta} \min_{C\in\tilde{\mathcal{C}}} -\frac{1}{n}\log e(C,W), \tag{2.11}$$

where $\mathbb{P}$ *is the uniform distribution over* $\mathcal{C}$. *The typical error exponent is basically the exponent of the average error probability of the worst code belonging to the best high probable collection of the ensemble.*

Definition 2.1.2. *A two-user discrete memoryless multiple-access channel (DM-MAC) is defined by a stochastic matrix* $W : \mathcal{X}\times\mathcal{Y} \to Z$, *where* $\mathcal{X}$, $\mathcal{Y}$, *the input alphabets, and* $\mathcal{Z}$, *the output alphabet, are finite sets. The channel transition probability for n-sequences is given by*

$$W^n(\mathbf{z}|\mathbf{x},\mathbf{y}) \triangleq \prod_{i=1}^n W(z_i|x_i,y_i), \tag{2.12}$$

where $\mathbf{x} \triangleq (x_1,...,x_n) \in \mathcal{X}^n$, $\mathbf{y} \triangleq (y_1,...,y_n) \in \mathcal{Y}^n$, *and* $\mathbf{z} \triangleq (z_1,...,z_n) \in \mathcal{Z}^n$.
An (n,M,N) *multi-user code for a given MAC,* W, *is a set* $C = \{(\mathbf{x}_i,\mathbf{y}_j,D_{ij}) : 1 \leq i \leq M, 1 \leq j \leq N\}$ *with*

- $\mathbf{x}_i \in \mathcal{X}^n$, $\mathbf{y}_j \in \mathcal{Y}^n$, $D_{ij} \subset \mathcal{Z}^n$

- $D_{ij} \cap D_{i'j'} = \varnothing$ *for* $(i,j) \neq (i',j')$.

The transmission rate pair is defined as $(R_X, R_Y) = \left(\frac{1}{n}\log M_X, \frac{1}{n}\log M_Y\right)$. *When message* (i,j) *is transmitted, the conditional probability of error of two-user code* C *is given by*

$$e_{ij}(C,W) \triangleq W^n(D_{ij}^c|\mathbf{x}_i, \mathbf{y}_j). \tag{2.13}$$

The average and maximal probability of error for the two-user code, C*, are defined as*

$$e(C,W) \triangleq \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} e_{ij}(C,W), \tag{2.14}$$

$$e_m(C,W) \triangleq \max_{i,j} e_{ij}(C,W). \tag{2.15}$$

An (n,M,N,λ) *code,* C*, for the DM-MAC,* W*, is an* (n,M,N) *code with*

$$e(C,W) \leq \lambda. \tag{2.16}$$

Finally, the average and maximal error exponents at rate pair (R_X, R_Y)*, are defined as:*

$$E_{av}^*(R_X,R_Y) \triangleq \limsup_{n\to\infty} \max_{C\in\mathcal{C}_M} -\frac{1}{n}\log e(C,W), \tag{2.17}$$

$$E_m^*(R_X,R_Y) \triangleq \limsup_{n\to\infty} \max_{C\in\mathcal{C}_M} -\frac{1}{n}\log e_m(C,W), \tag{2.18}$$

where $\mathcal{C}_M$ *is the set of all codes of length* n *and rate pair* (R_X, R_Y)*. The typical average error exponent of an ensemble* $\mathcal{C}$*, at rate pair* (R_X, R_Y)*, is defined as:*

$$E_{av}^T(R_X,R_Y) \triangleq \liminf_{\delta\to 0} \limsup_{n\to\infty} \max_{\tilde{\mathcal{C}}\subset\mathcal{C}:\mathbb{P}(\tilde{\mathcal{C}})>1-\delta} \min_{C\in\tilde{\mathcal{C}}} -\frac{1}{n}\log e(C,W), \tag{2.19}$$

where $\mathbb{P}$ *is the uniform distribution over* $\mathcal{C}$*.*

2.2 Summary of Known Results

2.2.1 Capacity Region for DM-MAC

The typical results of information theory are of asymptotic character and relate to the existence of codes with certain properties. Theorems asserting the existence of codes are called *direct results* while those asserting non-existence are called *converse results*. A combination of such results giving a complete asymptotic solution is called a *coding theorem*. In particular, a result stating that for rates above capacity, or outside the capacity region, the probability of error, as a function of block length, goes exponentially to 1, is called a *strong converse theorem*.

The capacity region for discrete memoryless multiple-access channels was characterized by Ahlswede [3] and Liao [37]. In their coding theorem, they proved that for any rate pair (R_X, R_Y) in the interior of a certain set $\mathcal{C}_{av}$, and for all sufficiently large blocklength n, there exists a multiuser code with an arbitrary small average probability of error. Conversely, for any rate pair outside of $\mathcal{C}_{av}$, the average probability of error is bounded away from 0. The set $\mathcal{C}_{av}$, called the *capacity region* [47], is defined as

$$\mathcal{C}_{av} \triangleq \bigcup_{P_{UXY} \in \mathcal{B}} \left\{ (R_X, R_Y) \;\middle|\; \begin{array}{c} 0 \leq R_X \leq I(X \wedge Z|Y, U) \\ 0 \leq R_Y \leq I(Y \wedge Z|X, U) \\ 0 \leq R_X + R_Y \leq I(XY \wedge Z|U) \end{array} \right\}, \tag{2.20}$$

and $\mathcal{B}$ is the set of all distributions defined on $\mathcal{U} \times \mathcal{X} \times \mathcal{Y}$ such that (a) $X - U - Y$ form a Markov chain, (b) $U - (X, Y) - Z$ form a Markov chain, (c) $U \in \mathcal{U} = \{1, 2, 3\}$.

In this single-letter characterization, U is called an auxiliary random variable. Unlike the case of point-to-point communication, where the single-letter characterization involves random variables associated with the channel input and the channel output, in many-to-one communication, the single-letter characterization of the capacity region involves, in addition, an auxiliary random variable. This random variable can be interpreted as a source of randomness that all the terminals can share to maximize the transmission rates. The first Markov chain can be interpreted as imposing the condition on the channel input distribution that the two encoders do not communicate with each other while transmitting data. The second Markov chain can be interpreted as imposing the condition that channel does not look at the source of randomness shared among the terminals.

For a given channel input distribution P_{UXY} in $\mathcal{B}$, the rates that are achievable belong to a pentagon. This is depicted in Figure 2.1.

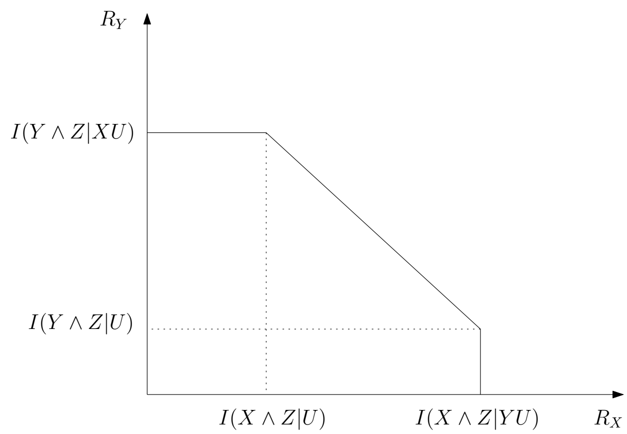

Figure 2.1: Achievable rates for a fixed channel input distribution P_{UXY}.

Since the capacity region uses the average error probability as performance criterion, this type of capacity region is called the *average capacity region.* For maximal error probability, the capacity regions are generally smaller and their determination is a challenging problem. In fact, for a general transmission system, there is a theory of coding for the average error probability and another for the maximal error probability. The drawback of the average error concept is that a small error probability is guaranteed only if both senders use their codewords with equal probabilities. For a DMC, it is unimportant whether we work with average or maximal error. However, for compound channels, the average performance generally does not coincide with the maximal performance. In particular, for discrete memoryless multiple-access channels, Dueck [19] proposed an example in which the maximal error capacity region was strictly smaller than the average error capacity region.

The converse theorems in [3, 37] are weak converse theorems. Dueck [19] proved a strong converse theorem by using the Ahlswede-Gacs-Korner [43] method of "blowing up decoding sets" in conjunction with a new "wringing technique". Later, Ahlswede [5] proved Dueck's result without using the method of "blowing up decoding sets". Ahlswede used his old method to derive upper bounds on the length of maximal error codes in conjunction with a "suitable wringing" technique to derive an upper bound on the length of average error codes. The heart of his approach was the fact that multi-user codes for MAC have subcodes with a certain independence structure.

2.2.2 Known Bounds on the Error Exponents of DM-MAC

The first lower bound on the error exponent of DM-MAC was derived by Slepain and Wolf [47] for a communication situation in which a third information source is jointly encoded by both users of the multiple-access channel. They proved that when the third source is not present, their bound yields an achievable error exponent for the MAC. Their bound does not reflect the possibility of time sharing; hence, it is loose for certain channels. In particular, for some rate pairs interior to the capacity region, their exponent was negative. This problem was remedied by Gallger [30] who presented a tighter random coding bound. All other random coding exponents have been derived by using the method of types. Dyachkov [20] obtained a random coding exponent, improving upon the one of Slepian and Wolf. However, it suffered from a lack of positivity in the interior of the capacity region. Pokorny and Wallmeier [41] derived a random coding bound which could be achieved universally for all MAC's with given input and output alphabets. They observed that the mutual position (mutual type) of the codewords, not the channel itself, plays a crucial role in determining of the probability of decoding error. They used the joint type of the codewords as the measure of distance. The approach used in their proof can be decomposed into a packing lemma and the calculation of the error bound. Pokorny and Wallmeier's packing lemma establishes the existence of codewords with some specified property, i.e., they showed that not too many codeword pairs are at a small distance from a given pair. They used the maximum mutual information decoding rule to bound the average probability of error. Later, Liu and Hughes [38] derived another random coding bound for the average error exponent of DM-MACs. Like Pokorny and Wallmeier's result, their bound is universally achievable, in the sense that neither the choice of codewords nor the choice of decoding rule is dependent on the channel statistics. Their approach was very similar to Pokorny and Wallmeier's approach. The main differences are that their packing lemma incorporated the channel output into all packing inequalities and was proved by using a different random code ensemble which leads to a tighter result. They used the minimum equivocation decoding rule to bound the probability of error. This random coding exponent is greater than or equal to those of previously known bounds. Moreover, they presented examples for which their exponent was strictly larger [38]. In the following, we present their random coding bound:

Fact 2.2.1. *For every finite set $\mathcal{U}$, $P_{XYU} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ satisfying $X - U - Y$, $R_X \geq 0$, $R_Y \geq 0$, $\delta > 0$, and $\mathbf{u} \in T^n_{P_U}$, there exists a multi-user code*

$$C = \{(\mathbf{x}_i, \mathbf{y}_j, D_{ij}) : i = 1, ..., M_X, \;\; j = 1, ..., M_Y\}, \tag{2.21}$$

with $\mathbf{x}_i \in T_{P_{X|U}}(\mathbf{u})$ and $\mathbf{y}_j \in T_{P_{Y|U}}(\mathbf{u})$ for all i and j, $M_X \geq 2^{n(R_X - \delta)}$, and $M_Y \geq 2^{n(R_Y - \delta)}$, such that for every MAC, $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$

$$e(C, W) \leq 2^{-n[E_r^{Liu}(R_X, R_Y, W, P_{XYU}) - \delta]}, \tag{2.22}$$

whenever $n \geq n_1(|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}|, |\mathcal{U}|, \delta)$, where

$$\begin{aligned} &E_r^{Liu}(R_X, R_Y, W, P_{XYU}) \qquad (2.23)\\ &\triangleq \min\left\{E_{rX}^{Liu}(R_X, R_Y, W, P_{XYU}), E_{rX}^{Liu}(R_X, R_Y, W, P_{XYU}), E_{rX}^{Liu}(R_X, R_Y, W, P_{XYU})\right\}, \end{aligned}$$

where E_{rX}^{Liu}, E_{rY}^{Liu}, E_{rXY}^{Liu}, are defined respectively by

$$E_{rX}^{Liu}(R_X, R_Y, W, P_{XYU}) = \min_{V_{UXYZ} \in \mathcal{V}^{Liu}(P_{XYU})} D(V_{Z|XYU}||W|V_{UXY}) + I_V(X \wedge Y|U) + |I_V(X \wedge YZ|U) - R_X|^+, \quad (2.24a)$$

$$E_{rY}^{Liu}(R_X, R_Y, W, P_{XYU}) = \min_{V_{UXYZ} \in \mathcal{V}^{Liu}(P_{XYU})} D(V_{Z|XYU}||W|V_{UXY}) + I_V(X \wedge Y|U) + |I_V(Y \wedge XZ|U) - R_Y|^+, \quad (2.24b)$$

$$E_{rXY}^{Liu}(R_X, R_Y, W, P_{XYU}) = \min_{V_{UXYZ} \in \mathcal{V}^{Liu}(P_{XYU})} D(V_{Z|XYU}||W|V_{UXY}) + I_V(X \wedge Y|U) + |I_V(XY \wedge Z|U) + I_V(X \wedge Y|U) - R_X - R_Y|^+, \quad (2.24c)$$

where $\mathcal{V}^{Liu}(P_{XYU})$ is defined as

$$\mathcal{V}^{Liu}(P_{XYU}) \triangleq \{V_{UXYZ} : V_{XU} = P_{XU}, V_{YU} = P_{YU}\}. \quad (2.25)$$

Liu, and Hughes [38] proved that the average error exponent of MAC, $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, satisfies

$$E_{av}^*(R_X, R_Y) \geq E_r^{Liu}(R_X, R_Y, W), \quad (2.26)$$

where

$$E_r^{Liu}(R_X, R_Y, W) \triangleq \sup_{\mathcal{U}:|\mathcal{U}|=4} \max_{\substack{P_{XYU}: \\ X-U-Y}} E_r^{Liu}(R_X, R_Y, W, P_{XYU}). \quad (2.27)$$

On the other hand, Haroutunian[33] has derived an *upper* bound for the reliability function of MAC W. This result asserts that $E_{av}^*(R_X, R_Y)$ is bounded above by

$$E_{sp}^H(R_X, R_Y, W) \triangleq \max_{P_{XY}} \min_{V_{Z|XY}} D(V_{Z|XY}||W|P_{XY}). \quad (2.28)$$

Here, the maximum is taken over all possible joint distributions on $\mathcal{X} \times \mathcal{Y}$, and the minimum over all channels $V_{Z|XY}$ which satisfy at least one of the following conditions

$$I_V(X \wedge Z|Y) \leq R_X, \tag{2.29a}$$

$$I_V(Y \wedge Z|X) \leq R_Y, \tag{2.29b}$$

$$I_V(XY \wedge Z) \leq R_X + R_Y. \tag{2.29c}$$

This bound tends to be somewhat loose because it does not take into account the separation of the two encoders in the MAC.

CHAPTER 3

Lower Bounds on the Error Exponent of Multiple-Access Channels

In this chapter, we develop two new lower bounds for the reliability function of DM-MACs. These bounds outperform the best known random coding bound derived in [38].

Toward this goal, we first revisit the point-to-point case and look at the techniques that are used for obtaining the lower bounds on the optimum error exponents. The techniques can be broadly classified into three categories. The first is the Gallager technique [29]. Although this yields expressions for the error exponents that are computationally easier to evaluate than others, the expressions themselves are harder to interpret. The second is the Csiszar-Korner technique [16]. This technique gives more intuitive expressions for the error exponents in terms of optimization of an objective function involving information quantities over probability distributions. This approach is more amenable to generalization to multi-user channels. The third is the graph decomposition technique using α-decoding [15]. α-decoding is a class of decoding procedures that includes maximum likelihood decoding and minimum entropy decoding. Although this technique gives a simpler derivation of the exponents, we believe that it is harder to generalize this to multi-user channels. All three classes of techniques give expressions for the random coding and expurgated exponents. The expressions obtained by the three techniques appear in different forms.

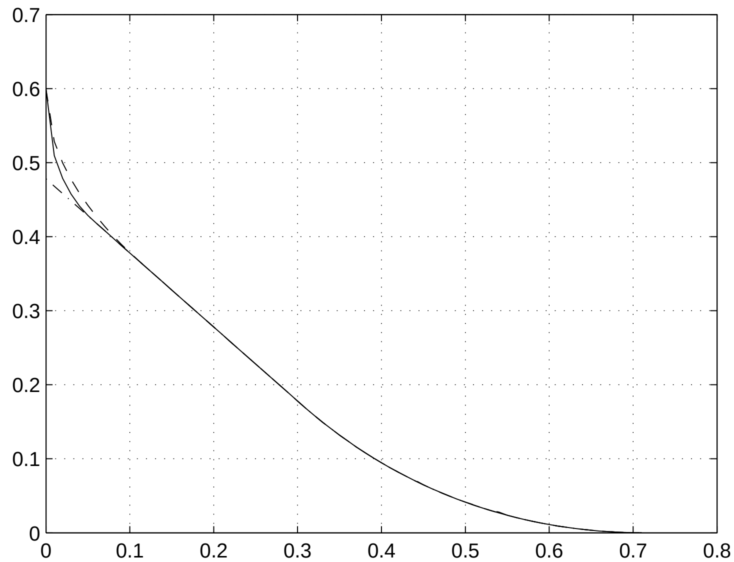

Figure 3.1: Lower bounds on the reliability function for point-to-point channel (random coding $-\cdot$, typical random coding $-$, expurgated $--$.

In developing our main result, we first develop a new simpler technique for deriving the random coding and expurgated exponents for the point-to-point channel using a constant composition code ensemble with α-decoding. We present our results in the format given in [15]. This technique also gives upper bounds on the ensemble averages. As a bonus, we obtain the typical random coding exponent for this channel. This gives an exact characterization (lower and upper bounds that meet) of the error exponent of almost all codes in the ensemble. When specialized to the BSC, this reduces to the typical random coding bound of Barg and Forney [8][1]. Figure 3.1 shows the random coding, the typical random coding, and the expurgated bounds for a BSC with crossover probability $p = 0.05$, which is representative of the general case. All three lower bounds are expressed as minimizations of a single objective function under different constraint sets. The reasons for looking at typical performance are two-fold. The first is that the average error exponent is in general smaller than the typical error exponent at low rates, hence, the latter gives a tighter characterization of the optimum error exponent of the channel. For example, for the BSC, although

[1] Barg and Forney gave only a lower bound in [8].

the average performance of the linear code ensemble is given by the random coding exponent of the Gallager ensemble, the typical performance is given by the expurgated exponent of the Gallager ensemble. In this direction, it was also recently noted in [12] that for the 8-PSK Gaussian channel, the typical performance of the ensemble of group codes over $\mathbb{Z}_8$ equals the expurgated exponent of the Gallager ensemble, whereas the typical performance of the ensemble of binary coset codes (under any mapping) is bounded away from the same. The second is that in some cases, expurgation may not be possible or may not be desirable. For example, (a) in the MAC, the standard expurgation is not possible, and (b) if one is looking at the performance of the best linear code for a channel, then expurgation destroys the linear structure which is not desirable. In the proposed technique, we provide a unified way to derive all the three lower bounds on the optimum error exponents, and upper bounds on the ensemble average and the typical performance. We wish to note that the bounds derived in this chapter are universal in nature. The proposed approach appears to be more amenable to generalization to multi-user channels.

A brief outline of the technique is given as follows. First, for a given constant composition code, we define a pair of packing functions that are independent of the channel. For an arbitrary channel, we relate the probability of error of a code with α-decoding to its packing functions. Packing functions give pair-wise and triple-wise joint-type distributions of the code. This is similar in spirit to the concept of distance distribution of the code. Then, we do random coding and obtain lower and upper bounds on the expected value of the packing functions of the ensemble without interfacing it with the channel. That is, these bounds do not depend on the channel. Finally, using the above relation between the packing function and the probability of error, we get single-letter expressions for the bounds on the optimum error exponents for an arbitrary channel.

Toward extending this technique to MACs, we follow a three-step approach. We start with a constant conditional composition ensemble identical to [38]. Then, we provide a new packing lemma in which the resulting code has better properties in comparison to the packing lemmas in [41] and [38]. This packing lemma is similar to Pokorny's packing lemma, in the sense that the channel conditional distribution does not appear in the inequalities. One of the advantages of our methodology is that it enables us to partially expurgate some of the codewords and end up with a new code with stronger properties. In particular, we do not eliminate pairs of codewords. Rather, we expurgate codewords from only one of the codebooks and analyze the performance of the expurgated code.

Contributions: In summary, the key contributions of the results of this chapter are

- A unified framework to obtain all lower bounds for the error exponent of the DMC.
- An exact characterization of the typical error exponent for the constant composition code ensemble for the DMC.
- Two new tighter lower bounds on the optimum error exponent for the DM-MAC.
- An characterization of the average error exponent of the constant composition code ensemble for the DM-MAC.
- A characterization of the typical error exponent for the constant composition code ensemble for the DM-MAC.

This chapter is organized as follows: Section 3.1 unifies the derivation of all lower bounds on the reliability function for a point-to-point DMC. Our main results for the DM-MAC are introduced in Section 3.2. Some numerical results are presented in Section 3.3. The proofs of some of these results are given in Section 3.4.

3.1 Point to Point: Lower Bounds on reliability function

3.1.1 Packing functions

Consider the class of DMCs with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. In the following, we introduce a unified way to derive all known lower bounds on the reliability function of such a channel. We will follow the random coding approach. First, we choose a constant composition code ensemble. Then, we define a packing function, $\pi : \mathcal{C} \times \mathcal{P}(\mathcal{X} \times \mathcal{X}) \to \mathbb{R}$, on all codebooks in the ensemble. The packing function that we use is the average number of codeword pairs sharing a particular joint type, $V_{X\tilde{X}}$. Specifically, $V_{X\tilde{X}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{X})$, and any code $C = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M\}$, the packing function is defined as:

$$\pi(C, V_{X\tilde{X}}) = \frac{1}{M} \sum_{i=1}^{M} \sum_{j \neq i} 1_{T_{V_{X\tilde{X}}}}(\mathbf{x}_i, \mathbf{x}_j). \tag{3.1}$$

We call this the first order packing function. Using this packing function, we prove three different packing lemmas, each of which shows the existence of a code with some desired properties.

In the first packing lemma, tight upper and lower bounds on the expectation of the packing function over the ensemble are derived. By using this packing lemma, upper and lower bounds on the expectation of the average probability of error over the ensemble are derived. These bounds meet for all transmission rates below the critical rate[2]. In the second packing lemma, by using the expectation and the variance of the packing function, we prove that for almost all codes in the constant composition code ensemble, the bounds in the first packing lemma are still valid. By using this tight bound on the performance of almost every code in the ensemble, we provide a tighter bound on the error exponent which we call the "typical" random coding bound. As we see later in the chapter, the typical random coding bound is indeed the typical performance of the constant composition code ensemble. In the third packing lemma, we use one of the typical codes and eliminate some of its "bad" codewords. The resulting code satisfies some stronger constraints in addition to all the previous properties. By using this packing lemma and an efficient decoding rule, we re-derive the well-known expurgated bound.

[2]This is essentially a re-derivation of the upper and lower bounds on the average probability of error obtained by Gallager in a different form. The present results are for constant composition codes.

To provide upper bounds on the average error exponents, such as those given below in Fact 3.1.1 and Theorem 3.1.1, for every $V_{X\tilde{X}\hat{X}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{X} \times \mathcal{X})$, we define a second packing function $\lambda : \mathcal{C} \times \mathcal{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{X}) \to \mathbb{R}$ on all codes in the constant composition code ensemble as follows:

$$\lambda(C, V_{X\tilde{X}\hat{X}}) \triangleq \frac{1}{M} \sum_{i=1}^{M} \sum_{j \neq i} \sum_{k \neq i,j} 1_{T_{V_{X\tilde{X}\hat{X}}}}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k). \tag{3.2}$$

We call this the second order packing function. As it is clear from the definition, this quantity is the average number of codeword triplets sharing a common joint distribution in code C.

3.1.2 Relation between packing function and probability of error

First, we consider the decoding rule at the receiver, and secondly we relate the average probability of error to the packing function.

Decoding Rule: In our derivation, error probability bounds using maximum- likelihood and minimum-entropy decoding rules will be obtained in a unified way. The reason is that both can be given in terms of a real-valued function on the set of distributions on $\mathcal{X} \times \mathcal{Y}$. This type of decoding rule was introduced in [15] as the $\alpha -$ *decoding* rule. For a given real-valued function α, a given code C, and for a received sequence $\mathbf{y} \in \mathcal{Y}^n$, the $\alpha -$ *decoder* accepts the codeword $\hat{\mathbf{x}} \in C$ for which the joint type of $\hat{\mathbf{x}}$ and $\mathbf{y}$ minimizes the function α, i.e., the decoder accepts $\hat{\mathbf{x}}$ if

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in C} \alpha(P \cdot V_{\mathbf{y}|\mathbf{x}}), \tag{3.3}$$

where P is the fixed composition of the codebook, and $V_{\mathbf{y}|\mathbf{x}}$ is the conditional type of $\mathbf{y}$ given $\mathbf{x}$. It was shown in [15] that for fixed composition codes, maximum-likelihood and minimum-entropy are special cases of this decoding rule. In particular, for maximum-likelihood decoding,

$$\alpha(P \cdot V) = D(V||W|P) + H(V|P), \tag{3.4}$$

and for minimum entropy decoding,

$$\alpha(P \cdot V) = H(V|P), \tag{3.5}$$

where V is the conditional type of $\mathbf{y}$ given $\mathbf{x}$.

Relation between probability of error and packing function: Next, for a given channel, we derive an upper bound and a lower bound on the average probability of error of an arbitrary constant composition code in terms of its first order and second order packing functions. The rest of the chapter is built on this crucial derivation. Consider the following argument about the average probability of error of a code C used on a channel W.

$$\begin{aligned} e(C,W) &= \frac{1}{M}\sum_{i=1}^{M} W^n(D_i^c|\mathbf{x}_i) \\ &= \frac{1}{M}\sum_{i=1}^{M} W^n\left(\left\{\mathbf{y} : \alpha(P \cdot V_{\mathbf{y}|\mathbf{x}_i}) \geq \alpha(P \cdot V_{\mathbf{y}|\mathbf{x}_j}) \text{ for some } j \neq i\right\}|\mathbf{x}_i\right) \\ &= \sum_{V_{X\tilde{X}Y} \in \mathcal{P}_n^r} \left(2^{-n[D(V_{Y|X}||W|P) + H_V(Y|X)]} \left[\frac{1}{M}\sum_{i=1}^{M} A_i\left(V_{X\tilde{X}Y}, C\right)\right]\right), \end{aligned} \tag{3.6}$$

where $\mathcal{P}_n^r$ and $A_i\left(V_{X\tilde{X}Y}, C\right)$ are defined as follows

$$\mathcal{P}_n^r \triangleq \left\{V_{X\tilde{X}Y} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{X} \times \mathcal{Y}) : V_X = V_{\tilde{X}} = P\ ,\ \alpha(P \cdot V_{Y|\tilde{X}}) \leq \alpha(P, V_{Y|X})\right\}, \tag{3.7}$$

$$A_i\left(V_{X\tilde{X}Y}, C\right) \triangleq \left|\left\{\mathbf{y} : (\mathbf{x}_i, \mathbf{x}_j, \mathbf{y}) \in T_{V_{X\tilde{X}Y}} \text{ for some } j \neq i\right\}\right|. \tag{3.8}$$

From the inclusion-exclusion principle, it follows that $A_i(V_{X\tilde{X}Y}, C)$ satisfies

$$B_i(V_{X\tilde{X}Y}, C) - C_i(V_{X\tilde{X}Y}, C) \leq A_i(V_{X\tilde{X}Y}, C) \leq B_i(V_{X\tilde{X}Y}, C), \tag{3.9}$$

where

$$B_i(V_{X\tilde{X}Y}, C) \triangleq \sum_{j \neq i} 1_{T_{V_{X\tilde{X}}}}(\mathbf{x}_i, \mathbf{x}_j) \left| \left\{ \mathbf{y} : \mathbf{y} \in T_{V_{Y|X\tilde{X}}}(\mathbf{x}_i, \mathbf{x}_j) \right\} \right|, \tag{3.10}$$

$$C_i(V_{X\tilde{X}Y}, C) \triangleq \sum_{j \neq i} \sum_{k \neq i,j} 1_{T_{V_{X\tilde{X}}}}(\mathbf{x}_i, \mathbf{x}_j) 1_{T_{V_{X\tilde{X}}}}(\mathbf{x}_i, \mathbf{x}_k) \left| \left\{ \mathbf{y} : \mathbf{y} \in T_{V_{Y|X\tilde{X}}}(\mathbf{x}_i, \mathbf{x}_j) \cap T_{V_{Y|X\tilde{X}}}(\mathbf{x}_i, \mathbf{x}_k) \right\} \right|. \tag{3.11}$$

Next, we provide an upper bound on the second term on the right hand side of (3.6) as follows.

$$\frac{1}{M} \sum_{i=1}^{M} A_i \left(V_{X\tilde{X}Y}, C \right) \leq \frac{1}{M} \sum_{i=1}^{M} B_i \left(V_{X\tilde{X}Y}, C \right) \tag{3.12a}$$

$$= \frac{1}{M} \sum_{i=1}^{M} \sum_{j \neq i} 1_{T_{V_{X\tilde{X}}}}(\mathbf{x}_i, \mathbf{x}_j) \left| \left\{ \mathbf{y} : \mathbf{y} \in T_{V_{Y|X\tilde{X}}}(\mathbf{x}_i, \mathbf{x}_j) \right\} \right| \tag{3.12b}$$

$$\leq \frac{1}{M} \sum_{i=1}^{M} \sum_{j \neq i} 1_{T_{V_{X\tilde{X}}}}(\mathbf{x}_i, \mathbf{x}_j) 2^{nH(Y|X\tilde{X})} \tag{3.12c}$$

$$= \pi(C, V_{X\tilde{X}}) 2^{nH(Y|X\tilde{X})} \tag{3.12d}$$

On the other hand

$$\left\{ \mathbf{y} : (\mathbf{x}_i, \mathbf{x}_j, \mathbf{y}) \in T_{V_{X\tilde{X}Y}} \text{ for some } j \neq i \right\} \subset T_{V_{Y|X}}(\mathbf{x}_i), \tag{3.13}$$

so we can conclude that

$$\frac{1}{M} \sum_{i=1}^{M} A_i(V_{X\tilde{X}Y}, C) \leq 2^{nH_V(Y|X)}. \tag{3.14}$$

Combining the above with (3.6), we have an upper bound on the probability of error in terms of the first order packing function as follows.

$$e(C,W) \leq \sum_{V_{X\tilde{X}Y} \in \mathcal{P}_n^r} 2^{-n[D(V_{Y|X}||W|P)]} \min\left\{2^{-nI_V(\tilde{X}\wedge Y|X)}\pi(C,V_{X\tilde{X}}),1\right\} \tag{3.15}$$

Next, we consider the lower bound. For that, we provide a lower bound on B_i and upper bound on C_i as follows.

$$\begin{aligned}\frac{1}{M}\sum_{i=1}^{M} B_i(V_{X\tilde{X}Y},C) &= \frac{1}{M}\sum_{i=1}^{M}\sum_{j\neq i} 1_{T_{V_{X\tilde{X}}}}(\mathbf{x}_i,\mathbf{x}_j)\left|\{\mathbf{y}:\mathbf{y}\in T_{V_{Y|X\tilde{X}}}(\mathbf{x}_i,\mathbf{x}_j)\}\right| \\ &\geq \pi(C,V_{X\tilde{X}})2^{n[H(Y|X\tilde{X})-\delta]},\end{aligned} \tag{3.16}$$

and

$$\begin{aligned}&\frac{1}{M}\sum_{i=1}^{M} C_i(V_{X\tilde{X}Y},C) = \\ &\frac{1}{M}\sum_{i=1}^{M}\sum_{j\neq i}\sum_{k\neq i,j} 1_{T_{V_{X\tilde{X}}}}(\mathbf{x}_i,\mathbf{x}_j)1_{T_{V_{X\tilde{X}}}}(\mathbf{x}_i,\mathbf{x}_k)\left|\left\{\mathbf{y}:\mathbf{y}\in T_{V_{Y|X\tilde{X}}}(\mathbf{x}_i,\mathbf{x}_j)\cap T_{V_{Y|X\tilde{X}}}(\mathbf{x}_i,\mathbf{x}_k)\right\}\right| \\ &= \sum_{\substack{V_{X\tilde{X}\hat{X}Y}: \\ V_{X\hat{X}Y}=V_{X\tilde{X}Y}}} \frac{1}{M}\sum_{i=1}^{M}\sum_{j\neq i}\sum_{k\neq i,j} 1_{T_{V_{X\tilde{X}\hat{X}}}}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_k)\left|\left\{\mathbf{y}:\mathbf{y}\in T_{V_{Y|X\tilde{X}\hat{X}}}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_k)\right\}\right|.\end{aligned} \tag{3.17}$$

By using $2^{nH(Y|X\tilde{X}\hat{X})}$ as an upper bound on the size of $\left\{\mathbf{y}:\mathbf{y}\in T_{V_{Y|X\tilde{X}\hat{X}}}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_k)\right\}$, it can be shown that (3.17) can be upper bounded by

$$\leq \sum_{\substack{V_{X\tilde{X}\hat{X}Y}: \\ V_{X\hat{X}Y}=V_{X\tilde{X}Y}}} 2^{nH(Y|X\tilde{X}\hat{X})}\lambda(C,V_{X\tilde{X}\hat{X}}) \tag{3.18}$$

Combining (3.6), (3.16), and (3.18) we have the following lower bound on the average probability of error.

$$e(C,W) \geq \sum_{V_{X\tilde{X}Y}\in\mathcal{P}_n^r} 2^{-n[D(V_{Y|X}||W|P)+I_V(\tilde{X}\wedge Y|X)+\delta]}$$
$$\left|\pi(C,V_{X\tilde{X}}) - \sum_{\substack{V_{X\tilde{X}\hat{X}Y}:\\ V_{X\hat{X}Y}=V_{X\tilde{X}Y}}} 2^{-n[I_V(\hat{X}\wedge Y|X\tilde{X})]}\lambda(C,V_{X\tilde{X}\hat{X}})\right|^+ \quad (3.19)$$

Observe that these upper and lower bounds apply for every code C. We have accomplished the task of relating the average probability of error to the two packing functions. The key results of this subsection are given by (3.15) and (3.19). Next, we use the packing lemmas to derive the bounds on the error exponents.

3.1.3 Packing Lemmas

Lemma 3.1.1. ***(Random Coding Packing Lemma)*** *Fix $R>0$, $\delta>0$, a sufficient large n and any type P of sequences in $\mathcal{X}^n$ satisfying $H(P)>R$. For any $V_{X\tilde{X}}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{X})$, the expectation of the first order packing function over the constant composition code ensemble is bounded by*

$$2^{n(R-I_V(X\wedge\tilde{X})-\delta)} \leq \mathbb{E}\left(\pi(X^M,V_{X\tilde{X}})\right) \leq 2^{n(R-I_V(X\wedge\tilde{X})+\delta)}, \quad (3.20)$$

where $X^M \triangleq (X_1,X_2,...,X_M)\subset T_P$ are independent and X_is are uniformly distributed on T_P, and $2^{n(R-\delta)}\leq M\leq 2^{nR}$. Moreover, the following inequality holds for the second order packing function:

$$\mathbb{E}\left(\lambda(X^M,V_{X\tilde{X}\hat{X}})\right) \leq 2^{n[2R-I_V(X\wedge\tilde{X})-I_V(\hat{X}\wedge X\tilde{X})+4\delta]} \qquad \textit{for all } V_{X\tilde{X}\hat{X}}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{X}\times\mathcal{X}). \quad (3.21)$$

Proof. The proof follows directly from the fact that two words drawn independently from T_P have a joint type $V_{X\tilde{X}}$ with probability close to $2^{-nI(X\wedge\hat{X})}$. The details are provided in Section 3.4.1. □

Lemma 3.1.2. ***(Typical Random Code Packing Lemma)*** *Fix $R>0$, $\delta>0$, a sufficient large n and any type P of sequences in $\mathcal{X}^n$ satisfying $H(P)>R$. Almost* *every code, C^t, with $2^{n(R-\delta)} \le M \le 2^{nR}$ codewords, in the constant composition code ensemble satisfies the following inequalities*

$$2^{n[R-I_V(X\wedge\tilde{X})-2\delta]} \le \pi(C^t, V_{X\tilde{X}}) \le 2^{n[R-I_V(X\wedge\tilde{X})+2\delta]} \qquad \text{for all } V_{X\tilde{X}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X}), \tag{3.22}$$

and

$$\lambda(C^t, V_{X\tilde{X}\hat{X}}) \le 2^{n[2R-I_V(X\wedge\tilde{X})-I_V(\hat{X}\wedge X\tilde{X})+4\delta]} \qquad \text{for all } V_{X\tilde{X}\hat{X}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X}\times\mathcal{X}). \tag{3.23}$$

Proof. The proof is provided in Section 3.4.1. In the proof, we evaluate the variance of the packing function and use Chebyshev's inequality to show that with high probability the packing function is close to its expected value. □

Lemma 3.1.3. ***(Expurgated Packing Lemma)*** *For every sufficiently large n, every $R > 0$, $\delta > 0$ and every type P of sequences in $\mathcal{X}^n$ satisfying $H(P) > R$, there exists a set of codewords $C^{ex} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{M^*}\} \subset T_P$ with $M^* \ge \frac{2^{n(R-\delta)}}{2}$, such that for any $V_{X\tilde{X}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X})$,*

$$\pi(C^{ex}, V_{X\tilde{X}}) \le 2^{n(R-I_V(X\wedge\tilde{X})+2\delta)}, \tag{3.24}$$

and for every sequence $\mathbf{x}_i \in C^{ex}$,

$$|T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C^{ex}| \le 2^{n(R-I_V(X\wedge\tilde{X})+2\delta)}. \tag{3.25}$$

Proof. The proof is provided in Section 3.4.1. The basic idea of the proof is simple. From Lemma 3.1.1, we know that for every $V_{X\tilde{X}}$, there exists a code whose packing function is upper bounded by a number that is close to $2^{n(R-I_V(X\wedge\tilde{X}))}$. Since the packing function is an average over all codewords in the code, we infer that for at least half of the codewords, the corresponding property (3.25) is satisfied. In Section 3.4.1, we show that there exists a single code that works for every joint type. □

3.1.4 Error Exponent Bounds

Now, we obtain the bounds on the error exponents using the results from the previous three subsections. We present three lower bounds and two upper bounds. The lower bounds are the random coding exponent, typical random coding exponent and expurgated exponent. All the three lower bounds are expressed as minimization of the same objective function under different constraint sets. Similar structure is manifested in the case of upper bounds. For completeness, we first rederive the well-known result of random coding exponent.

Fact 3.1.1. ***(Random Coding Bound)*** *For every type P of sequences in $\mathcal{X}^n$ and $0 \leq R \leq H(P)$, $\delta > 0$, every DMC, $W : \mathcal{X} \to \mathcal{Y}$, and $2^{n(R-\delta)} \leq M \leq 2^{nR}$, the expectation of the average error probability over the constant composition code ensemble with M codewords of type P, can be bounded by*

$$2^{-n[E_{rL}(R,P,W)+3\delta]} \leq \bar{P}_e \leq 2^{-n[E_r(R,P,W)-2\delta]}, \tag{3.26}$$

whenever $n \geq n_1(|\mathcal{X}|, |\mathcal{Y}|, \delta)$, where

$$\begin{aligned} E_r(R,P,W) &\triangleq \min_{V_{X\tilde{X}Y} \in \mathcal{P}^r} D(V_{Y|X}||W|P) + |I_V(\tilde{X} \wedge XY) - R|^+, & (3.27)\\ E_{rL}(R,P,W) &\triangleq \min_{\substack{V_{X\tilde{X}Y} \in \mathcal{P}^r: \\ I_V(\tilde{X} \wedge XY) \geq R}} D(V_{Y|X}||W|P) + I_V(\tilde{X} \wedge XY) - R, & (3.28) \end{aligned}$$

and

$$\mathcal{P}^r \triangleq \left\{ V_{X\tilde{X}Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{Y}) : V_X = V_{\tilde{X}} = P \ , \ \alpha(P, V_{Y|\tilde{X}}) \leq \alpha(P, V_{Y|X}) \right\}. \tag{3.29}$$

In particular, there exists a set of codewords $C^r = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M\} \subset T_P$, with $M \geq 2^{n(R-\delta)}$, such that for every DMC, $W : \mathcal{X} \to \mathcal{Y}$,

$$e(C^r, W) \leq 2^{-n[E_r(R,P,W)-3\delta]}. \tag{3.30}$$

Proof. The proof is straightforward and is outlined in Section 3.4.1. □

It is well known that for $R \geq R_{crit}$, the random coding error exponent is equal to the sphere packing error exponent, and as a result the random coding bound is a tight bound. In addition, the following is true.

Corollary 3.1.1. *For any* $R \leq R_{crit}$,

$$\max_{P \in \mathcal{P}(\mathcal{X})} E_{rL}(R, P, W) = \max_{P \in \mathcal{P}(\mathcal{X})} E_r(R, P, W). \tag{3.31}$$

Proof. The proof is provided in the Section 3.4.1. □

Next, we have an exact characterization of the typical performance of the constant composition code ensemble.

Theorem 3.1.1. ***(Typical random Coding Bound)*** *For every type P of sequences in $\mathcal{X}^n$, $\delta > 0$, and every transmission rate satisfying $0 \leq R \leq H(P)$, almost all codes, $C^t = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M\}$ with $\mathbf{x}_i \in T_P$ for all i, $M \geq 2^{n(R-\delta)}$, satisfy*

$$2^{-n[E_{TL}(R,P,W)+4\delta]} \leq e(C^t, W) \leq 2^{-n[E_T(R,P,W)-3\delta]}, \tag{3.32}$$

for every DMC, $W : \mathcal{X} \to \mathcal{Y}$, whenever $n \geq n_1(|\mathcal{X}|, |\mathcal{Y}|, \delta)$. Here,

$$E_T(R, P, W) \triangleq \min_{V_{X\tilde{X}Y} \in \mathcal{P}^t} D(V_{Y|X}||W|P) + |I_V(\tilde{X} \wedge XY) - R|^+, \tag{3.33}$$

$$E_{TL}(R, P, W) \triangleq \min_{\substack{V_{X\tilde{X}Y} \in \mathcal{P}^t: \\ I_V(\tilde{X} \wedge XY) \geq R}} D(V_{Y|X}||W|P) + I_V(\tilde{X} \wedge XY) - R, \tag{3.34}$$

where

$$\mathcal{P}^t \triangleq \big\{V_{X\tilde{X}Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{Y}) : \; V_X = V_{\tilde{X}} = P, \; I_V(X \wedge \tilde{X}) \leq 2R, \\ \alpha(P, V_{Y|\tilde{X}}) \leq \alpha(P, V_{Y|X})\big\}. \tag{3.35}$$

Proof. The proof is provided in Section 3.4.1. □

In Theorem 3.1.1, we proved the existence of a high probability (almost 1) collection of codes such that every code in this collection satisfies (3.32). This provides a lower bound on the typical average error exponent for the constant composition code ensemble as defined in equation (2.11). In the following, we show that the typical performance of the best high-probability collection cannot be better than that given in Theorem 3.1.1.

Corollary 3.1.2. *For every type P of sequences in $\mathcal{X}^n$ and every transmission rate satisfying $0 \leq R \leq H(P)$,*

$$E_T(R, P, W) \leq E_{av}^T(R, P) \leq E_{TL}(R, P, W), \tag{3.36}$$

where $E_{av}^T(R, P)$ is the typical average error exponent of the constant composition (P) code ensemble.

Proof. The proof is provided in the Section 3.4.1. □

Clearly, since the random coding bound is tight for $R \geq R_{crit}$, the same is true for the typical random coding bound. For $R \leq R_{crit}$ we have the following result.

Corollary 3.1.3. *For any $R \leq R_{crit}$,*

$$\max_{P \in \mathcal{P}(\mathcal{X})} E_{TL}(R, P, W) = \max_{P \in \mathcal{P}(\mathcal{X})} E_T(R, P, W). \tag{3.37}$$

Proof. The proof is very similar to that of Corollary 3.1.1 and is omitted. □

It can be seen that the typical random coding bound is the true error exponent for almost all codes, with M codewords, in the constant composition code ensemble. A similar lower bound on the typical random coding bound was derived by Barg and Forney [8] for the binary symmetric channel. Although the approach used here is completely different from the one in [8], in the following corollary we show that these two bounds coincide for binary symmetric channels.

Corollary 3.1.4. *For a binary symmetric channel with crossover probability p, and* *for* $0 \leq R \leq R_{crit}$

$$\max_{P \in \mathcal{P}(\mathcal{X})} E_T(R, P, W) = E_{TRC}(R), \tag{3.38}$$

where E_{TRC} *is the lower bound for the error exponent of a typical random code in [8].*

Finally, we re-derive the well-known expurgated error exponent in a rather straightforward way.

Fact 3.1.2. ***(Expurgated Bound)*** *For every type* P *of sequences in* $\mathcal{X}^n$ *and* $0 \leq R \leq H(P)$*,* $\delta > 0$*, there exists a set of codewords* $C^{ex} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{M^*}\} \subset T_P$ *with* $M^* \geq \frac{2^{n(R-\delta)}}{2}$*, such that for every DMC,* $W : \mathcal{X} \to \mathcal{Y}$*,*

$$e(C^{ex}, W) \leq 2^{-n[E_{ex}(R,P,W) - 3\delta]} \tag{3.39}$$

whenever $n \geq n_1(|\mathcal{X}|, |\mathcal{Y}|, \delta)$*, where*

$$E_{ex}(R, P, W) \triangleq \min_{V_{X\tilde{X}Y} \in \mathcal{P}^{ex}} D(V_{Y|X}||W|P) + |I_V(\tilde{X} \wedge XY) - R|^+ \tag{3.40}$$

where

$$\begin{aligned} \mathcal{P}^{ex} \triangleq \quad & \big\{V_{X\tilde{X}Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{Y}) : \; V_X = V_{\tilde{X}} = P, \quad I_V(X \wedge \tilde{X}) \leq R, \\ & \qquad \alpha(P, V_{Y|\tilde{X}}) \leq \alpha(P, V_{Y|X})\big\} \end{aligned} \tag{3.41}$$

Proof. The proof is provided in Section 3.4.1. □

Note that none of the mentioned three bounds have their "traditional format" as found in [16], [28], but rather the format introduced in [15] by Csiszar and Korner. It was shown in [15] that the new random coding bound is equivalent to the original one for maximum likelihood and minimum entropy decoding rule. Furthermore, the new format for the expurgated bound is equivalent to the traditional one for maximum likelihood-decoding and it results in a bound that is the maximum of the traditional expurgated and random coding bounds.

3.2 MAC: Lower Bounds on Error Exponent

Consider a DM-MAC, W, with input alphabets $\mathcal{X}$ and $\mathcal{Y}$, and output alphabet $\mathcal{Z}$. In this section, we present three achievable lower bounds on the reliability function (upper bound on the average error probability of the best code) for this channel. The method we are using is very similar to the point-to-point case. Again, the goal is first proving the existence of a good code and then analyzing its performance. The first step is choosing the ensemble. The ensemble, $\mathcal{C}$, we are using is similar to the ensemble in [38]. For a fixed distribution, $P_U P_{X|U} P_{Y|U}$, the codewords of each code in the ensemble are chosen from $T_{P_{X|U}}(\mathbf{u})$ and $T_{P_{Y|U}}(\mathbf{u})$ for some sequence $\mathbf{u} \in T_{P_U}$. Intuitively, we expect that the codewords in a "good" code must be far from each other. In accordance with the ideas of Csiszar and Korner [16], we use conditional types to quantify this statement. We select a prescribed number of sequences in $\mathcal{X}^n$ and $\mathcal{Y}^n$ so that the shells around each pair have small intersections with the shells around other sequences. In general, two types of packing lemmas have been studied in the literature based on whether the shells are defined on the channel input space or channel output space. The packing lemma in [41] belongs to the first type, and the one in [38] belongs to the second type. All the inequalities in the first type depend only on the channel input sequences. However, in the second type, the lemma incorporates the channel output into the packing inequalities. In this chapter, we use the first type. In the following, we follow a four step procedure to arrive at the error exponent bounds. In step one, we define first-order and second-order packing functions. These functions are independent of the channel statistics. Next, in step two, for any constant composition code and any DM-MAC, we provide upper and lower bounds on the probability of decoding error in terms of these packing functions. In step three, by using a random coding argument on the constant composition code ensemble, we show the existence of codes whose packing functions satisfy certain conditions. Finally, in step four, by connecting the results in step two and three, we provide lower and upper bounds on the error exponents. Our results include a new tighter lower bound on the error exponent for DM-MAC using a new partial expurgation method for multi-user codes. We also give a tight characterization of the typical performance of the constant composition code ensemble. Both the expurgated bound as well as the typical bound outperform the random coding bound of [38], which is derived as special case of our methodology.

3.2.1 Definition of Packing Functions

Let $C_X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{M_X}\}$ and $C_Y = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{M_Y}\}$ be constant composition codebooks with $\mathbf{x}_i \in T_{P_{X|U}}(\mathbf{u})$ and $\mathbf{y}_j \in T_{P_{Y|U}}(\mathbf{u})$, for some $\mathbf{u} \in T_{P_U}$. In the following, for a two-user code $C = C_X \times C_Y$, we define the following quantities that we will use later in this section.

Definition 3.2.1. *Fix a finite set* $\mathcal{U}$, *and a joint type* $V_{UXY\tilde{X}\tilde{Y}} \in \mathcal{P}_n(\mathcal{U} \times (\mathcal{X} \times \mathcal{Y})^2)$. *For code* C, *the first-order packing functions are defined as follows:*

$$N_U(C, V_{UXY}) \triangleq \frac{1}{M_X M_Y} \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} 1_{T_{V_{UXY}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j), \tag{3.42a}$$

$$N_X(C, V_{UXY\tilde{X}}) \triangleq \frac{1}{M_X M_Y} \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} \sum_{k \neq i} 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k), \tag{3.42b}$$

$$N_Y(C, V_{UXY\tilde{Y}}) \triangleq \frac{1}{M_X M_Y} \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} \sum_{l \neq j} 1_{T_{V_{UXY\tilde{Y}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{y}_l), \tag{3.42c}$$

$$N_{XY}(C, V_{UXY\tilde{X}\tilde{Y}}) \triangleq \frac{1}{M_X M_Y} \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} \sum_{k \neq i} \sum_{l \neq j} 1_{T_{V_{UXY\tilde{X}\tilde{Y}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k, \mathbf{y}_l). \tag{3.42d}$$

Moreover, for any $V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}} \in \mathcal{P}_n\left(\mathcal{U} \times (\mathcal{X} \times \mathcal{Y})^3\right)$, *we define a set of second-*

order packing functions as follows:

$$\Lambda_X(C, V_{UXY\tilde{X}\hat{X}}) \triangleq \frac{1}{M_X M_Y} \sum_{i,j} \sum_{k \neq i} \sum_{k' \neq i,k} 1_{T_{V_{UXY\tilde{X}\hat{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k, \mathbf{x}_{k'}), \tag{3.43a}$$

$$\Lambda_Y(C, V_{UXY\tilde{Y}\hat{Y}}) \triangleq \frac{1}{M_X M_Y} \sum_{i,j} \sum_{l \neq j} \sum_{l' \neq j,l} 1_{T_{V_{UXY\tilde{Y}\hat{Y}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{y}_l, \mathbf{y}_{l'}), \tag{3.43b}$$

$$\Lambda_{XY}(C, V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}) \triangleq \frac{1}{M_X M_Y} \sum_{i,j} \sum_{\substack{k \neq i \\ l \neq j}} \sum_{\substack{k' \neq i,k \\ l' \neq j,l}} 1_{T_{V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k, \mathbf{y}_l, \mathbf{x}_{k'}, \mathbf{y}_{l'}). \tag{3.43c}$$

The second-order packing functions are used to prove the tightness of the results of Theorem 3.2.1 and Theorem 3.2.2. Next, we will obtain upper and lower bounds on the probability of decoding error for an arbitrary two-user code that depend on its packing functions defined above.

3.2.2 Relation between probability of error and packing functions

Consider the multiuser code C as defined above, and a function $\alpha : \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}) \to \mathbb{R}$. Taking into account the given $\mathbf{u}$, α-decoding yields the decoding sets

$$D_{ij} = \left\{\mathbf{z} : \alpha(P_{\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{z}}) \leq \alpha(P_{\mathbf{u},\mathbf{x}_k,\mathbf{y}_l,\mathbf{z}}) \text{ for all } (k,l) \neq (i,j)\right\}. \tag{3.44}$$

The average error probability of this multiuser code on DM-MAC, W, can be written as

$$\begin{aligned} e(C, W) &\triangleq \frac{1}{M_X M_Y} \sum_{i,j} W^n(D_{ij}^c | \mathbf{x}_i, \mathbf{y}_j) \\ &= \frac{1}{M_X M_Y} \sum_{i,j} W^n(\bigcup_{k \neq i} D_{kj} | \mathbf{x}_i, \mathbf{y}_j) + \frac{1}{M_X M_Y} \sum_{i,j} W^n(\bigcup_{l \neq j} D_{il} | \mathbf{x}_i, \mathbf{y}_j) \\ &\quad + \frac{1}{M_X M_Y} \sum_{i,j} W^n(\bigcup_{\substack{k \neq i \\ l \neq j}} D_{kl} | \mathbf{x}_i, \mathbf{y}_j). \end{aligned} \tag{3.45}$$

The first term on the right side of (3.45) can be written as

$$\begin{aligned}
&\frac{1}{M_X M_Y}\sum_{i,j} W^n(\bigcup_{k\neq i} D_{kj}|\mathbf{x}_i,\mathbf{y}_j)\\
&=\frac{1}{M_X M_Y}\sum_{i,j} W^n\Big(\left\{\mathbf{z}:\alpha(P_{\mathbf{u},\mathbf{x}_k,\mathbf{y}_j,\mathbf{z}})\le\alpha(P_{\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{z}}),\text{for some } k\neq i\right\}|\mathbf{u},\mathbf{x}_i,\mathbf{y}_j\Big)\\
&=\frac{1}{M_X M_Y}\sum_{i,j}\sum_{\substack{\mathbf{z}:\\ \alpha(P_{\mathbf{u},\mathbf{x}_k,\mathbf{y}_j,\mathbf{z}})\le\alpha(P_{\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{z}})\\ \text{for some } k\neq i}} W^n\left(\mathbf{z}|\mathbf{u},\mathbf{x}_i,\mathbf{y}_j\right)\\
&=\frac{1}{M_X M_Y}\sum_{i,j}\sum_{\substack{V_{UXY\tilde{X}Z}\\ \in\mathcal{V}^r_{X,n}}}\sum_{\substack{\mathbf{z}:\\ \alpha(P_{\mathbf{u},\mathbf{x}_k,\mathbf{y}_j,\mathbf{z}})\le\alpha(P_{\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{z}})\\ \text{for some } k\neq i}} 1_{T_{V_{UXY\tilde{X}Z}}}(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{x}_k,\mathbf{z})W^n\left(\mathbf{z}|\mathbf{u},\mathbf{x}_i,\mathbf{y}_j\right)
\end{aligned} \tag{3.46}$$

$$\begin{aligned}
=\sum_{V_{UXY\tilde{X}Z}\in\mathcal{V}^r_{X,n}} & 2^{-n[D(V_{Z|XYU}||W|V_{XYU})+H_V(Z|XYU)]}\cdot\\
&\Big[\frac{1}{M_X M_Y}\sum_{i,j} 1_{T_{V_{UXY}}}(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j)\cdot A^X_{i,j}\left(V_{UXY\tilde{X}Z},C\right)\Big],
\end{aligned} \tag{3.47}$$

where

$$\begin{aligned}
&A^X_{i,j}\left(V_{UXY\tilde{X}Z},C\right)\triangleq\left|\{\mathbf{z}:(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{x}_k,\mathbf{z})\in T_{V_{UXY\tilde{X}Z}}\text{for some } k\neq i\}\right|\\
&\mathcal{V}^r_{X,n}\triangleq\{V_{UXY\tilde{X}Z}:\alpha(V_{UXYZ})\ge\alpha(V_{U\tilde{X}YZ}),V_{UX}=V_{U\tilde{X}}=P_{UX},V_{UY}=P_{UY}\}\,.
\end{aligned} \tag{3.48}$$

Note that $\mathcal{V}^r_{X,n}$ is a set of types of resolution n, therefore, we use a subscript n to define it. Similarly, the second and third term term on the right side of (3.45) can be written as follows:

$$
\begin{aligned}
&\frac{1}{M_X M_Y}\sum_{i,j} W^n(\bigcup_{l\neq j} D_{il}|\mathbf{x}_i,\mathbf{y}_j)\\
&= \sum_{V_{UXY\tilde{Y}Z}\in\mathcal{V}^r_{Y,n}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU})+H_V(Z|XYU)]}\\
&\qquad\qquad\qquad \cdot\Big[\frac{1}{M_X M_Y}\sum_{i,j} 1_{T_{V_{UXY}}}(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j).A^Y_{i,j}\left(V_{UXY\tilde{Y}Z},C\right)\Big], \quad (3.49)
\end{aligned}
$$

where

$$
\begin{aligned}
&A^Y_{i,j}\left(V_{UXY\tilde{Y}Z},C\right) \triangleq \left|\{\mathbf{z}:(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{y}_l,\mathbf{z})\in T_{V_{UXY\tilde{Y}Z}}\text{for some } l\neq j\}\right|\\
&\mathcal{V}^r_{Y,n} \triangleq \left\{V_{UXY\tilde{Y}Z}:\alpha(V_{UXYZ})\geq\alpha(V_{UX\tilde{Y}Z}), V_{UX}=P_{UX}, V_{UY}=V_{U\tilde{Y}}=P_{UY}\right\}, \quad (3.50)
\end{aligned}
$$

and,

$$
\begin{aligned}
&\frac{1}{M_X M_Y}\sum_{i,j} W^n(\bigcup_{\substack{k\neq i\\ l\neq j}} D_{kl}|\mathbf{x}_i,\mathbf{y}_j)\\
&= \sum_{V_{UXY\tilde{X}\tilde{Y}Z}\in\mathcal{V}^r_{XY,n}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU})+H_V(Z|XYU)]}\\
&\qquad\qquad\qquad \cdot\Big[\frac{1}{M_X M_Y}\sum_{i,j} 1_{T_{V_{UXY}}}(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j)\cdot A^{XY}_{i,j}\left(V_{UXY\tilde{X}\tilde{Y}Z},C\right)\Big],
\end{aligned}
\tag{3.51}
$$

where

$$
\begin{aligned}
&A^{XY}_{i,j}\left(V_{UXY\tilde{X}\tilde{Y}Z},C\right) \triangleq \left|\{\mathbf{z}:(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{x}_k,\mathbf{y}_l,\mathbf{z})\in T_{V_{UXY\tilde{X}\tilde{Y}Z}}\text{for some } k\neq i, l\neq j\}\right|\\
&\mathcal{V}^r_{XY,n} \triangleq \left\{V_{UXY\tilde{X}\tilde{Y}Z}\;\middle|\; \begin{array}{c}\alpha(V_{UXYZ})\geq\alpha(V_{U\tilde{X}\tilde{Y}Z})\\ V_{UX}=V_{U\tilde{X}}=P_{UX},\; V_{UY}=V_{U\tilde{Y}}=P_{UY}\end{array}\right\}.
\end{aligned}
\tag{3.52}
$$

Clearly, $A_{i,j}^X\left(V_{UXY\tilde{X}Z}\right)$ satisfies

$$B_{i,j}^X\left(V_{UXY\tilde{X}Z}, C\right) - C_{i,j}^X\left(V_{UXY\tilde{X}Z}, C\right) \leq A_{i,j}^X\left(V_{UXY\tilde{X}Z}, C\right) \leq B_{i,j}^X\left(V_{UXY\tilde{X}Z}, C\right), \tag{3.53}$$

where

$$B_{i,j}^X\left(V_{UXY\tilde{X}Z}, C\right) \triangleq \sum_{k \neq i} 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k) \cdot \left|\{\mathbf{z} : \mathbf{z} \in T_{V_{Z|UXY\tilde{X}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k\}\right|, \tag{3.54}$$

$$\begin{aligned} C_{i,j}^X\left(V_{UXY\tilde{X}Z}, C\right) \triangleq \sum_{k \neq i} \sum_{k' \neq k,i} & 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k) 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_{k'}) \\ & \cdot \left|\{\mathbf{z} : \mathbf{z} \in T_{V_{Z|UXY\tilde{X}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k) \cap T_{V_{Z|UXY\tilde{X}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_{k'})\}\right|. \end{aligned} \tag{3.55}$$

Having related the probability of error and the function $B_{i,j}^X$, $B_{i,j}^Y$ and $B_{i,j}^{XY}$, our next task is to provide a simple upper bound on these functions. This is done as follows.

$$\begin{aligned} & \frac{1}{M_X M_Y} \sum_{i,j} 1_{T_{V_{UXY}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j) B_{i,j}^X\left(V_{UXY\tilde{X}Z}, C\right) \\ & \quad = \frac{1}{M_X M_Y} \sum_{i,j} \sum_{k \neq i} 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k) \left|\left\{\mathbf{z} : \mathbf{z} \in T_{V_{Z|UXY\tilde{X}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k)\right\}\right| \\ & \quad \leq 2^{nH(Z|UXY\tilde{X})} \frac{1}{M_X M_Y} \sum_{i,j} \sum_{k \neq i} 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k) \\ & \quad = 2^{nH(Z|UXY\tilde{X})} N_X(C, V_{UXY\tilde{X}}) \end{aligned} \tag{3.56}$$

Similarly, we can provide upper bounds for $B_{i,j}^Y$ and $B_{i,j}^{XY}$. Moreover, we can also provide trivial upper bounds on $A(\cdot)$ functions as was done in the point-to-point case.

$$A_{i,j}^X(V_{UXY\tilde{X}Z}, C) \leq 2^{nH_V(Z|XYU)}.$$

The same bound applies to A^Y and A^{XY}. Collecting all these results, we provide the following upper bound on the probability of error.

$$
\begin{aligned}
e(C,W) \leq & \sum_{\substack{V_{UXY\tilde{X}Z} \\ \in \mathcal{V}^r_{X,n}}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU})]} \min\left\{2^{-nI_V(\tilde{X}\wedge Z|XYU)} N_X(C, V_{UXY\tilde{X}}), 1\right\} \\
& + \sum_{\substack{V_{UXY\tilde{Y}Z} \\ \in \mathcal{V}^r_{Y,n}}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU})]} \min\left\{2^{-nI_V(\tilde{Y}\wedge Z|XYU)} N_Y(C, V_{UXY\tilde{Y}}), 1\right\} \\
& + \sum_{\substack{V_{UXY\tilde{X}\tilde{Y}Z} \\ \in \mathcal{V}^r_{XY,n}}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU})]} \min\left\{2^{-nI_V(\tilde{X}\tilde{Y}\wedge Z|XYU)} N_{XY}(C, V_{UXY\tilde{X}\tilde{Y}}), 1\right\}
\end{aligned}
\tag{3.57}
$$

Next, we consider lower bounds on $B(\cdot)$ functions and upper bounds on $C(\cdot)$ functions. One can use a similar argument to show the following

$$
\frac{1}{M_X M_Y} \sum_{i,j} 1_{T_{V_{UXY}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j) B^X_{i,j}(V_{UXY\tilde{X}Z}, C) \geq 2^{n[H(Z|UXY\tilde{X})-\delta]} N_X(C, V_{UXY\tilde{X}}).
$$

Similar lower bounds can be obtained for B^Y and B^{XY}. Moreover, we have the following arguments for bounding from above the function C^X.

$$
\begin{aligned}
& \frac{1}{M_X M_Y} \sum_{i,j} 1_{T_{V_{UXY}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j) \cdot C^X_{i,j}(V_{UXY\tilde{X}Z}) \\
& = \frac{1}{M_X M_Y} \sum_{i,j} 1_{T_{V_{UXY}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j) \sum_{k\neq i} \sum_{k'\neq k,i} 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k) 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_{k'}) \\
& \qquad \cdot \left|\left\{\mathbf{z} : \mathbf{z} \in T_{V_{Z|UXY\tilde{X}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k) \cap T_{V_{Z|UXY\tilde{X}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_{k'})\right\}\right|
\end{aligned}
$$

$$= \frac{1}{M_X M_Y} \sum_{i,j} \sum_{\substack{V_{UXY\tilde{X}\hat{X}Z}: \\ V_{UXY\hat{X}Z} = V_{UXY\tilde{X}Z}}} \sum_{k \neq i} \sum_{k' \neq k,i} 1_{T_{V_{UXY\tilde{X}\hat{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k, \mathbf{x}_{k'})$$

$$\cdot \left| \left\{ \mathbf{z} : \mathbf{z} \in T_{V_{Z|UXY\tilde{X}\hat{X}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k, \mathbf{x}_{k'}) \right\} \right|$$

$$\leq \sum_{\substack{V_{UXY\tilde{X}\hat{X}Z}: \\ V_{UXY\hat{X}Z} = V_{UXY\tilde{X}Z}}} 2^{nH(Z|UXY\tilde{X}\hat{X})} \frac{1}{M_X M_Y} \sum_{i,j} \sum_{k \neq i} \sum_{k' \neq k,i} 1_{T_{V_{UXY\tilde{X}\hat{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k, \mathbf{x}_{k'})$$

$$= \sum_{\substack{V_{UXY\tilde{X}\hat{X}Z}: \\ V_{UXY\hat{X}Z} = V_{UXY\tilde{X}Z}}} 2^{nH(Z|UXY\tilde{X}\hat{X})} \Lambda_X(C, V_{UXY\tilde{X}\hat{X}}). \tag{3.58}$$

Similar relation can be obtained that relate C^Y and Λ_Y, C^{XY} and Λ_{XY}. Combining the lower bounds on $B(\cdot)$-functions and upper bounds on $C(\cdot)$-functions, we have the following lower bound on the probability of decoding error.

$$e(C, W)$$

$$\geq \sum_{\substack{V_{UXY\tilde{X}Z} \\ \in \mathcal{V}^r_{X,n}}} 2^{-n[D(V_{Z|XYU}||W|V) + I(\tilde{X} \wedge Z|XYU)]} \left| N_X - \sum_{\substack{V_{UXY\tilde{X}\hat{X}Z}: \\ V_{UXY\hat{X}Z} = V_{UXY\tilde{X}Z}}} 2^{nI(\hat{X} \wedge Z|UXY\tilde{X})} \Lambda_X \right|^+$$

$$+ \sum_{\substack{V_{UXY\tilde{Y}Z} \\ \in \mathcal{V}^r_{Y,n}}} 2^{-n[D(V_{Z|XYU}||W|V) + I(\tilde{Y} \wedge Z|XYU)]} \left| N_Y - \sum_{\substack{V_{UXY\tilde{Y}\hat{Y}Z}: \\ V_{UXY\hat{Y}Z} = V_{UXY\tilde{Y}Z}}} 2^{nI(\hat{Y} \wedge Z|UXY\tilde{Y})} \Lambda_Y \right|^+$$

$$+ \sum_{\substack{V_{UXY\tilde{X}\tilde{Y}Z} \\ \in \mathcal{V}^r_{XY,n}}} 2^{-n[D(V_{Z|XYU}||W|V) + I(\tilde{X}\tilde{Y} \wedge Z|XYU)]}$$

$$\cdot \left| N_{XY} - \sum_{\substack{V_{UXY\tilde{X}\hat{X}\tilde{Y}\hat{Y}Z}: \\ V_{UXY\hat{X}\hat{Y}Z} = V_{UXY\tilde{X}\tilde{Y}Z}}} 2^{nI(\hat{X}\hat{Y} \wedge Z|UXY\tilde{X}\tilde{Y})} \Lambda_{XY} \right|^+ . \tag{3.59}$$

This completes our task of relating the average probability of error of any code C in terms of the first and the second order packing functions. We next proceed toward obtaining lower bounds on the error exponents. The expressions for the error exponents that we derive are conceptually very similar to those derived for the point-to-point channels. However, since we have to deal with a bigger class of error events, the expressions for the error exponents become longer. To state our results concisely, in the next subsection, we define certain functions of information quantities and transmission rates. We will express our results in terms of these functions. The reader can skip this subsection, and move to the next subsection without losing the flow of the exposition. The reader can come back to it when we refer to it in the subsequent discussions.

3.2.3 Definition of Information Functions

In the following, we consider five definitions which are mainly used for conciseness.

Definition 3.2.2. *For any fix rate pair $R_X, R_Y \geq 0$, and any distribution $V_{UXY\tilde{X}\tilde{Y}} \in \mathcal{P}(\mathcal{U} \times (\mathcal{X} \times \mathcal{Y})^2)$, we define*

$$F_U(V_{UXY}) \triangleq I(X \wedge Y|U), \tag{3.60a}$$

$$F_X(V_{UXY\tilde{X}}) \triangleq I(X \wedge Y|U) + I_V(\tilde{X} \wedge XY|U) - R_X, \tag{3.60b}$$

$$F_Y(V_{UXY\tilde{Y}}) \triangleq I(X \wedge Y|U) + I(\tilde{Y} \wedge XY|U) - R_Y, \tag{3.60c}$$

$$F_{XY}(V_{UXY\tilde{X}\tilde{Y}}) \triangleq I(X \wedge Y|U) + I(\tilde{X} \wedge \tilde{Y}|U) + I(\tilde{X}\tilde{Y} \wedge XY|U) - R_X - R_Y. \tag{3.60d}$$

Moreover, for any $V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}} \in \mathcal{P}(\mathcal{U} \times (\mathcal{X} \times \mathcal{Y})^3)$, we define

$$E_S^X(V_{UXY\tilde{X}\hat{X}}) \triangleq I(\hat{X} \wedge XY\tilde{X}|U) + I(\tilde{X} \wedge XY|U) + I(X \wedge Y|U) - 2R_X, \tag{3.61a}$$

$$E_S^Y(V_{UXY\tilde{Y}\hat{Y}}) \triangleq I(\hat{Y} \wedge XY\tilde{Y}|U) + I(\tilde{Y} \wedge XY|U) + I(X \wedge Y|U) - 2R_Y, \tag{3.61b}$$

$$\begin{aligned} E_S^{XY}(V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}) \triangleq\ & I(\hat{X}\hat{Y} \wedge XY\tilde{X}\tilde{Y}|U) + I(\tilde{X}\tilde{Y} \wedge XY|U) + I(X \wedge Y|U) \\ & + I(\tilde{X} \wedge \tilde{Y}|U) + I(\hat{X} \wedge \hat{Y}|U) - 2R_X - 2R_Y. \end{aligned} \tag{3.61c}$$

Definition 3.2.3. *For any given $R_X, R_Y \geq 0$, $P_{XYU} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$, we define* *the sets of distributions* $\mathcal{V}_X^r$, $\mathcal{V}_Y^r$ *and* $\mathcal{V}_{XY}^r$ *as follows:*

$$\mathcal{V}_X^r \triangleq \{V_{UXY\tilde{X}Z} : \alpha(V_{UXYZ}) \geq \alpha(V_{U\tilde{X}YZ}), V_{UX} = V_{U\tilde{X}} = P_{UX}, V_{UY} = P_{UY}\}, \tag{3.62a}$$

$$\mathcal{V}_Y^r \triangleq \{V_{UXY\tilde{Y}Z} : \alpha(V_{UXYZ}) \geq \alpha(V_{UX\tilde{Y}Z}), V_{UX} = P_{UX}, V_{UY} = V_{U\tilde{Y}} = P_{UY}\}, \tag{3.62b}$$

$$\mathcal{V}_{XY}^r \triangleq \left\{ V_{UXY\tilde{X}\tilde{Y}Z} \middle| \begin{array}{c} \alpha(V_{UXYZ}) \geq \alpha(V_{U\tilde{X}\tilde{Y}Z}) \\ V_{UX} = V_{U\tilde{X}} = P_{UX}, V_{UY} = V_{U\tilde{Y}} = P_{UY} \end{array} \right\}. \tag{3.62c}$$

Moreover, $\mathcal{V}_X^{r,L}$, $\mathcal{V}_Y^{r,L}$ *and* $\mathcal{V}_{XY}^{r,L}$ *are sets of distributions and defined as*

$$\mathcal{V}_X^{r,L} \triangleq \left\{ V_{UXY\tilde{X}Z} \in \mathcal{V}_X^r : I(\tilde{X} \wedge XYZ|U) \geq R_X \right\}, \tag{3.63a}$$

$$\mathcal{V}_Y^{r,L} \triangleq \left\{ V_{UXY\tilde{Y}Z} \in \mathcal{V}_Y^r : I(\tilde{Y} \wedge XYZ|U) \geq R_Y \right\}, \tag{3.63b}$$

$$\mathcal{V}_{XY}^{r,L} \triangleq \left\{ V_{UXY\tilde{X}\tilde{Y}Z} \in \mathcal{V}_{XY}^r : I(\tilde{X}\tilde{Y} \wedge XYZ|U) + I(\tilde{X} \wedge \tilde{Y}) \geq R_X + R_Y \right\}. \tag{3.63c}$$

Definition 3.2.4. *For any given* $R_X, R_Y \geq 0$, $P_{XYU} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$, *we define the sets of distributions* $\mathcal{V}_X^T$, $\mathcal{V}_Y^T$, *and* $\mathcal{V}_{XY}^T$ *as follows*

$$\mathcal{V}_X^T \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{X}} : & V_{XU} = V_{\tilde{X}U} = P_{XU}, V_{YU} = P_{YU} \\ & F_U(V_{UXY}), F_U(V_{U\tilde{X}Y}) \leq R_X + R_Y \\ & F_X(V_{UXY\tilde{X}}) \leq R_X + R_Y \\ & \alpha(V_{UXYZ}) \geq \alpha(V_{U\tilde{X}YZ}) \end{array} \right\} \tag{3.64a}$$

$$\mathcal{V}_Y^T \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{Y}} : & V_{XU} = P_{XU}, V_{YU} = V_{\tilde{Y}U} = P_{YU} \\ & F_U(V_{UXY}), F_U(V_{UX\tilde{Y}}) \leq R_X + R_Y \\ & F_Y(V_{UXY\tilde{Y}}) \leq R_X + R_Y \\ & \alpha(V_{UXYZ}) \geq \alpha(V_{UX\tilde{Y}Z}) \end{array} \right\} \tag{3.64b}$$

$$\mathcal{V}_{XY}^{T} \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{X}\tilde{Y}}: & V_{XU} = V_{\tilde{X}U} = P_{XU}, V_{YU} = V_{\tilde{Y}U} = P_{YU} \\ & F_U(V_{UXY}), F_U(V_{U\tilde{X}Y}), F_U(V_{UX\tilde{Y}}), F_U(V_{U\tilde{X}\tilde{Y}}) \leq R_X + R_Y \\ & F_X(V_{UXY\tilde{X}}), F_X(V_{UX\tilde{Y}\tilde{X}}) \leq R_X + R_Y \\ & F_Y(V_{UXY\tilde{Y}}), F_Y(V_{U\tilde{X}Y\tilde{Y}}) \leq R_X + R_Y \\ & F_{XY}(V_{UXY\tilde{X}\tilde{Y}}), F_{XY}(V_{U\tilde{X}YX\tilde{Y}}) \leq R_X + R_Y \\ & \alpha(V_{UXYZ}) \geq \alpha(V_{U\tilde{X}\tilde{Y}Z}) \end{array} \right\} \tag{3.64c}$$

Moreover, $\mathcal{V}_X^{T,L}$, $\mathcal{V}_Y^{T,L}$, *and* $\mathcal{V}_{XY}^{T,L}$ *are sets of distributions and defined as*

$$\mathcal{V}_X^{T,L} \triangleq \left\{ V_{UXY\tilde{X}Z} \in \mathcal{V}_X^T : I(\tilde{X} \wedge XYZ|U) \geq R_X \right\}, \tag{3.65a}$$

$$\mathcal{V}_Y^{T,L} \triangleq \left\{ V_{UXY\tilde{Y}Z} \in \mathcal{V}_Y^T : I(\tilde{Y} \wedge XYZ|U) \geq R_Y \right\}, \tag{3.65b}$$

$$\mathcal{V}_{XY}^{T,L} \triangleq \left\{ V_{UXY\tilde{X}\tilde{Y}Z} \in \mathcal{V}_{XY}^T : I(\tilde{X}\tilde{Y} \wedge XYZ|U) + I(\tilde{X} \wedge \tilde{Y}) \geq R_X + R_Y \right\}. \tag{3.65c}$$

Definition 3.2.5. *For any given* $R_X, R_Y \geq 0$, $P_{XYU} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$, *we define the sets of distributions* $\mathcal{V}_X^{ex}$, $\mathcal{V}_Y^{ex}$, *and* $\mathcal{V}_{XY}^{ex}$ *as follows*

$$\mathcal{V}_X^{ex} \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{X}}: & V_{XU} = V_{\tilde{X}U} = P_{XU}, V_{YU} = P_{YU} \\ & F_U(V_{UXY}), F_U(V_{U\tilde{X}Y}) \leq \min\{R_X, R_Y\} \\ & F_X(V_{UXY\tilde{X}}) \leq \min\{R_X, R_Y\} \\ & \alpha(V_{UXYZ}) \geq \alpha(V_{U\tilde{X}YZ}) \end{array} \right\} \tag{3.66a}$$

$$\mathcal{V}_Y^{ex} \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{Y}}: & V_{XU} = P_{XU}, V_{YU} = V_{\tilde{Y}U} = P_{YU} \\ & F_U(V_{UXY}), F_U(V_{UX\tilde{Y}}) \leq \min\{R_X, R_Y\} \\ & F_Y(V_{UXY\tilde{Y}}) \leq \min\{R_X, R_Y\} \\ & \alpha(V_{UXYZ}) \geq \alpha(V_{UX\tilde{Y}Z}) \end{array} \right\} \tag{3.66b}$$

$$\mathcal{V}_{XY}^{ex} \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{X}\tilde{Y}} : & V_{XU} = V_{\tilde{X}U} = P_{XU}, V_{YU} = V_{\tilde{Y}U} = P_{YU} \\ & F_U(V_{UXY}), F_U(V_{U\tilde{X}Y}) \leq \min\{R_X, R_Y\} \\ & F_U(V_{UX\tilde{Y}}), F_U(V_{U\tilde{X}\tilde{Y}}) \leq \min\{R_X, R_Y\} \\ & F_X(V_{UXY\tilde{X}}), F_X(V_{UX\tilde{Y}\tilde{X}}) \leq \min\{R_X, R_Y\} \\ & F_Y(V_{UXY\tilde{Y}}), F_Y(V_{U\tilde{X}Y\tilde{Y}}) \leq \min\{R_X, R_Y\} \\ & F_{XY}(V_{UXY\tilde{X}\tilde{Y}}), F_{XY}(V_{U\tilde{X}\tilde{Y}X\tilde{Y}}) \leq \min\{R_X, R_Y\} \\ & \alpha(V_{UXYZ}) \geq \alpha(V_{U\tilde{X}\tilde{Y}Z}) \end{array} \right\} \quad (3.66c)$$

Definition 3.2.6. *For any given $R_X, R_Y \geq 0$, $P_{XYU} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$, and $V_{UXY\tilde{X}\tilde{Y}} \in \mathcal{P}(\mathcal{U} \times (\mathcal{X} \times \mathcal{Y})^2)$, we define the following quantities*

$$\begin{aligned} E_X(R_X, R_Y, W, P_{XYU}, V_{UXY\tilde{X}}) &\triangleq D(V_{Z|XYU}||W|V_{XYU}) + I_V(X \wedge Y|U) \\ &\quad + |I(\tilde{X} \wedge XYZ|U) - R_X|^+, \quad (3.67a) \\ E_Y(R_X, R_Y, W, P_{XYU}, V_{UXY\tilde{Y}}) &\triangleq D(V_{Z|XYU}||W|V_{XYU}) + I_V(X \wedge Y|U) \\ &\quad + |I(\tilde{Y} \wedge XYZ|U) - R_Y|^+, \quad (3.67b) \\ E_{XY}(R_X, R_Y, W, P_{XYU}, V_{UXY\tilde{X}\tilde{Y}}) &\triangleq D(V_{Z|XYU}||W|V_{XYU}) + I_V(X \wedge Y|U) \\ &\quad + |I(\tilde{X}\tilde{Y} \wedge XYZ|U) + I_V(\tilde{X} \wedge \tilde{Y}|U) - R_X - R_Y|^+. \quad (3.67c) \end{aligned}$$

Moreover, we define

$$\begin{aligned} E_X^L(R_X, R_Y, W, P_{XYU}, V_{UXY\tilde{X}}) &\triangleq D(V_{Z|XYU}||W|V_{XYU}) + I_V(X \wedge Y|U) \\ &\quad + I(\tilde{X} \wedge XYZ|U) - R_X, \quad (3.68a) \\ E_Y^L(R_X, R_Y, W, P_{XYU}, V_{UXY\tilde{Y}}) &\triangleq D(V_{Z|XYU}||W|V_{XYU}) + I_V(X \wedge Y|U) \\ &\quad + I(\tilde{Y} \wedge XYZ|U) - R_Y, \quad (3.68b) \\ E_{XY}^L(R_X, R_Y, W, P_{XYU}, V_{UXY\tilde{X}\tilde{Y}}) &\triangleq D(V_{Z|XYU}||W|V_{XYU}) + I_V(X \wedge Y|U) \\ &\quad + I(\tilde{X}\tilde{Y} \wedge XYZ|U) + I_V(\tilde{X} \wedge \tilde{Y}|U) - R_X - R_Y, \quad (3.68c) \end{aligned}$$

and,

$$E_\beta^\alpha(R_X, R_Y, W, P_{XYU}, \mathcal{V}_\beta^\alpha) \triangleq \min_{V_{UXY\tilde{\beta}Z} \in \mathcal{V}_\beta^\alpha} E_\beta(R_X, R_Y, W, P_{XYU}, V_{UXY\tilde{\beta}}), \tag{3.69a}$$

$$E_\beta^{\alpha,L}(R_X, R_Y, W, P_{XYU}, \mathcal{V}_\beta^\alpha) \triangleq \min_{V_{UXY\tilde{\beta}Z} \in \mathcal{V}_\beta^{\alpha,L}} E_\beta^L(R_X, R_Y, W, P_{XYU}, V_{UXY\tilde{\beta}}), \tag{3.69b}$$

for $\alpha \in \{r, T, ex\}$*, and* $\beta \in \{X, Y, XY\}$*.*

3.2.4 Packing Lemmas

As we did in the point-to-point case, here we perform random coding and derive bounds on the packing functions. The results will be stated as three lemmas, one for the average and one for the typical performance of the ensemble, and finally one for the expurgated ensemble. These results will be used in conjunction with the relation between the packing functions and the probability of error established in Section 3.2.2 to obtain the bounds on the error exponents.

Lemma 3.2.1. *Fix a finite set* $\mathcal{U}$*,* $P_{XYU} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ *such that* $X - U - Y$*,* $R_X \geq 0$*,* $R_Y \geq 0$ *,* $\delta > 0$*,* $2^{n(R_X - \delta)} \leq M_X \leq 2^{nR_X}$*,* $2^{n(R_Y - \delta)} \leq M_Y \leq 2^{nR_Y}$*, and* $\mathbf{u} \in T_{P_U}$*. Let* $X^{M_X} \triangleq \{X_1, X_2, ..., X_{M_X}\}$ *and* $Y^{M_Y} \triangleq \{Y_1, Y_2, ..., Y_{M_Y}\}$ *are independent, and* X_i*s and* Y_j*s are uniformly distributed over* $T_{P_{X|U}}(\mathbf{u})$ *and* $T_{P_{Y|U}}(\mathbf{u})$ *respectively. For every joint type* $V_{UXY\tilde{X}\tilde{Y}} \in \mathcal{P}_n(\mathcal{U} \times (\mathcal{X} \times \mathcal{Y})^2)$*, the expectation of the packing functions over the random code* $X^{M_X} \times Y^{M_Y}$ *are bounded by*

$$2^{-n[F_U(V_{UXY})+\delta]} \leq \mathbb{E}\Big[N_U(X^{M_X} \times Y^{M_Y}, V_{UXY})\Big] \leq 2^{-n[F_U(V_{UXY})-2\delta]}, \tag{3.70a}$$

$$2^{-n[F_X(V_{UXY\tilde{X}})+3\delta]} \leq \mathbb{E}\Big[N_X(X^{M_X} \times Y^{M_Y}, V_{UXY\tilde{X}})\Big] \leq 2^{-n[F_X(V_{UXY\tilde{X}})-4\delta]}, \tag{3.70b}$$

$$2^{-n[F_Y(V_{UXY\tilde{Y}})+3\delta]} \leq \mathbb{E}\Big[N_Y(X^{M_X} \times Y^{M_Y}, V_{UXY\tilde{Y}})\Big] \leq 2^{-n[F_Y(V_{UXY\tilde{Y}})-4\delta]}, \tag{3.70c}$$

$$2^{-n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})+4\delta]} \leq \mathbb{E}\Big[N_{XY}(X^{M_X} \times Y^{M_Y}, V_{UXY\tilde{X}\tilde{Y}})\Big] \leq 2^{-n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})-4\delta]}, \tag{3.70d}$$

whenever $n \geq n_0(|\mathcal{U}|, |\mathcal{X}|, |\mathcal{Y}|, \delta)$*. Moreover, for any* $V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}} \in \mathcal{P}_n(\mathcal{U} \times (\mathcal{X} \times \mathcal{Y})^3)$

$$\mathbb{E}\Big[\Lambda_X(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{X}\hat{X}})\Big] \le 2^{-n\left(E_S^X(V_{UXY\tilde{X}\hat{X}})-4\delta\right)}, \tag{3.71a}$$

$$\mathbb{E}\Big[\Lambda_Y(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{Y}\hat{Y}})\Big] \le 2^{-n\left(E_S^Y(V_{UXY\tilde{Y}\hat{Y}})-4\delta\right)}, \tag{3.71b}$$

$$\mathbb{E}\Big[\Lambda_{XY}(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}})\Big] \le 2^{-n\left(E_S^{XY}(V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}})-6\delta\right)}, \tag{3.71c}$$

whenever $n \ge n_0(|\mathcal{U}|, |\mathcal{X}|, |\mathcal{Y}|, \delta)$.

Proof. The proof is provided in Section 3.4.2. □

Lemma 3.2.2. *Fix a finite set* $\mathcal{U}$, $P_{XYU} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}\times\mathcal{U})$ *such that* $X - U - Y$, $R_X \ge 0$, $R_Y \ge 0$, $\delta > 0$, $2^{n(R_X-\delta)} \le M_X \le 2^{nR_X}$, $2^{n(R_Y-\delta)} \le M_Y \le 2^{nR_Y}$, *and* $\mathbf{u} \in T_{P_U}$. *Almost every multi-user code* $C = C_X \times C_Y$, $C_X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{M_X}\} \subset T_{P_{X|U}}(\mathbf{u})$ *and* $C_Y = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{M_Y}\} \subset T_{P_{Y|U}}(\mathbf{u})$, *in the constant composition code ensemble,* $\mathcal{C}$, *satisfies the following inequalities:*

$$2^{-n[F_U(V_{UXY})+3\delta]} \le N_U(C, V_{UXY}) \le 2^{-n[F_U(V_{UXY})-3\delta]}, \tag{3.72a}$$

$$2^{-n[F_X(V_{UXY\tilde{X}})+5\delta]} \le N_X(C, V_{UXY\tilde{X}}) \le 2^{-n[F_X(V_{UXY\tilde{X}})-5\delta]}, \tag{3.72b}$$

$$2^{-n[F_Y(V_{UXY\tilde{Y}})+5\delta]} \le N_Y(C, V_{UXY\tilde{Y}}) \le 2^{-n[F_Y(V_{UXY\tilde{X}\tilde{Y}})-5\delta]}, \tag{3.72c}$$

$$2^{-n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})+5\delta]} \le N_{XY}(C, V_{UXY\tilde{X}\tilde{Y}}) \le 2^{-n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})-5\delta]}, \tag{3.72d}$$

for all $V_{UXY\tilde{X}\tilde{Y}} \in \mathcal{P}_n(\mathcal{U}\times(\mathcal{X}\times\mathcal{Y})^2)$, *and*

$$\Lambda_X(C, V_{UXY\tilde{X}\hat{X}}) \le 2^{-n\left(E_S^X(V_{UXY\tilde{X}\hat{X}})-5\delta\right)}, \tag{3.73a}$$

$$\Lambda_Y(C, V_{UXY\tilde{Y}\hat{Y}}) \le 2^{-n\left(E_S^Y(V_{UXY\tilde{Y}\hat{Y}})-5\delta\right)}, \tag{3.73b}$$

$$\Lambda_{XY}(C, V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}) \le 2^{-n\left(E_S^{XY}(V_{UXY\tilde{X}\hat{X}})-7\delta\right)}. \tag{3.73c}$$

for all $V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}} \in \mathcal{P}_n\left(\mathcal{U}\times(\mathcal{X}\times\mathcal{Y})^3\right)$, *whenever* $n \ge n_0(|\mathcal{U}|, |\mathcal{X}|, |\mathcal{Y}|, \delta)$.

Proof. The proof is provided in 3.4.2. □

Lemma 3.2.3. *For every finite set* $\mathcal{U}$, $P_{XYU} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}\times\mathcal{U})$ *such that* $X - U - Y$, $R_X \ge 0$, $R_Y \ge 0$, $\delta > 0$, *and* $\mathbf{u} \in T_{P_U}$, *there exist a multi-user code* $C^* = C_X^* \times C_Y^*$,

$C_X^* = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{M_X^*}\} \subset T_{P_{X|U}}(\mathbf{u})$ and $C_Y^* = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{M_Y^*}\} \subset T_{P_{Y|U}}(\mathbf{u})$ with $M_X^* \geq \frac{2^{n(R_X-\delta)}}{2}$, $M_Y^* \geq \frac{2^{n(R_Y-\delta)}}{2}$, such that for every joint type $V_{UXY\tilde{X}\tilde{Y}} \in \mathcal{P}_n(\mathcal{U} \times (\mathcal{X} \times \mathcal{Y})^2)$,

$$N_U(C^*, V_{UXY}) \leq 2^{-n[F_U(V_{UXY})-6\delta]} \tag{3.74a}$$

$$N_X(C^*, V_{UXY\tilde{X}}) \leq 2^{-n[F_X(V_{UXY\tilde{X}})-6\delta]} \tag{3.74b}$$

$$N_Y(C^*, V_{UXY\tilde{Y}}) \leq 2^{-n[F_Y(V_{UXY\tilde{Y}})-6\delta]} \tag{3.74c}$$

$$N_{XY}(C^*, V_{UXY\tilde{X}\tilde{Y}}) \leq 2^{-n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})-6\delta]} \tag{3.74d}$$

and for any $1 \leq i \leq M_X^*$, and any $1 \leq j \leq M_Y^*$,

$$1_{T_{V_{UXY}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j) \leq 2^{-n[F_U(V_{UXY})-\min\{R_X,R_Y\}-6\delta]} \tag{3.75a}$$

$$\sum_{k\neq i} 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k) \leq 2^{-n[F_X(V_{UXY\tilde{X}})-\min\{R_X,R_Y\}-6\delta]} \tag{3.75b}$$

$$\sum_{l\neq j} 1_{T_{V_{UXY\tilde{Y}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{y}_l) \leq 2^{-n[F_Y(V_{UXY\tilde{Y}})-\min\{R_X,R_Y\}-6\delta]} \tag{3.75c}$$

$$\sum_{k\neq i}\sum_{l\neq j} 1_{T_{V_{UXY\tilde{X}\tilde{Y}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k, \mathbf{y}_l) \leq 2^{-n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})-\min\{R_X,R_Y\}-6\delta]}, \tag{3.75d}$$

whenever

$$n \geq n_0(|\mathcal{U}|, |\mathcal{X}|, |\mathcal{Y}|, \delta).$$

Proof. The proof is provided in 3.4.2. □

As it is shown in the Section 3.4.2, the above property is derived by the method of expurgation. Unlike the point-to-point case, expurgation in the MAC is not a trivial procedure. To see that, observe that expurgating bad pairs of codewords results in a code with correlated messages, which is hard to analyze. Instead, what we do is a sort of "partial" expurgation. Roughly speaking, we start with a code whose existence is proved in Lemma 3.2.1 and eliminate some of the bad codewords from the code with the larger rate (as opposed to codeword pairs). By doing that, all messages in the new code are independent, and such a code is easier to analyze.

3.2.5 Error exponent bounds

We can now proceed in a fashion that is similar to the point-to-point case and derive a series of exponential bounds based on Lemmas 3.2.1, 3.2.2, and 3.2.3. In the following, we present three lower bounds, the random coding, the typical random coding, and the expurgated bounds. As in the case of point-to-point channels, here too, all the lower bounds are expressed in terms of the optimization of a single objective function under different constraint sets.

Theorem 3.2.1. *Fix a finite set* $\mathcal{U}$, $P_{XYU} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ *such that* $X - U - Y$, $R_X \geq 0$, $R_Y \geq 0$, $\delta > 0$, $2^{n(R_X-\delta)} \leq M_X \leq 2^{nR_X}$, $2^{n(R_Y-\delta)} \leq M_Y \leq 2^{nR_Y}$, *and* $\mathbf{u} \in T_{P_U}$. *Consider the ensemble,* $\mathcal{C}$, *of multi-user codes consisting of all pair of codebooks* (C_X, C_Y), *where* $C_X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{M_X}\} \subset T_{P_{X|U}}(\mathbf{u})$ *and* $C_Y = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{M_Y}\} \subset T_{P_{Y|U}}(\mathbf{u})$. *The expectation of the average probability of error over* $\mathcal{C}$ *is bounded by*

$$2^{-n[E_{rL}(R_X,R_Y,W,P_{XYU})+8\delta]} \leq \bar{P}_e \leq 2^{-n[E_r(R_X,R_Y,W,P_{XYU})-6\delta]} \tag{3.76}$$

whenever $n \geq n_1(|\mathcal{Z}|, |\mathcal{X}|, |\mathcal{Y}|, |\mathcal{U}|, \delta)$, *where*

$$E_r(R_X, R_Y, W, P_{XYU}) \triangleq \min_{\beta=X,Y,XY} E_\beta^r(R_X, R_Y, W, P_{UXY}, \mathcal{V}_\beta^r), \tag{3.77}$$

$$E_{rL}(R_X, R_Y, W, P_{XYU}) \triangleq \min_{\beta=X,Y,XY} E_\beta^{r,L}(R_X, R_Y, W, P_{UXY}, \mathcal{V}_\beta^{r,L}). \tag{3.78}$$

Proof. The proof is provided in 3.4.2. □

Corollary 3.2.1. *In the low rate regime,*

$$E_{rL}(R_X, R_Y, W, P_{XYU}) = E_r(R_X, R_Y, W, P_{XYU}). \tag{3.79}$$

We call this rate region as the critical region for W.

Proof. The proof is similar to the proof of Corollary 3.1.1 and is omitted. □

Theorem 3.2.2. *Fix a finite set* $\mathcal{U}$, $P_{XYU} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ *such that* $X - U - Y$, $R_X \geq 0$, $R_Y \geq 0$, $\delta > 0$, $2^{n(R_X-\delta)} \leq M_X \leq 2^{nR_X}$, $2^{n(R_Y-\delta)} \leq M_Y \leq 2^{nR_Y}$, *and $\mathbf{u} \in T_{P_U}$. The average probability of error for almost all multi-user codes $C = C_X \times C_Y$, $C_X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{M_X}\} \subset T_{P_{X|U}}(\mathbf{u})$ and $C_Y = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{M_Y}\} \subset T_{P_{Y|U}}(\mathbf{u})$, in ensemble $\mathcal{C}$, satisfies the following inequalities*

$$2^{-n[E_{TL}(R_X,R_Y,W,P_{XYU})+7\delta]} \leq e(C,W) \leq 2^{-n[E_T(R_X,R_Y,W,P_{XYU})-6\delta]} \tag{3.80}$$

whenever $n \geq n_1(|\mathcal{Z}|, |\mathcal{X}|, |\mathcal{Y}|, |\mathcal{U}|, \delta)$, where

$$E_T(R_X, R_Y, W, P_{XYU}) \triangleq \min_{\beta=X,Y,XY} E_\beta^T(R_X, R_Y, W, P_{UXY}, \mathcal{V}_\beta^T) \tag{3.81}$$

$$E_{TL}(R_X, R_Y, W, P_{XYU}) \triangleq \min_{\beta=X,Y,XY} E_\beta^{T,L}(R_X, R_Y, W, P_{UXY}, \mathcal{V}_\beta^{T,L}). \tag{3.82}$$

Proof. The proof is provided in 3.4.2. □

Corollary 3.2.2. *For every finite set $\mathcal{U}$, $\mathcal{P}_{XYU} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ such that $X - U - Y$, $R_X \geq 0$, $R_Y \geq 0$,*

$$E_T(R_X, R_Y, P_{XYU}, W) \leq E_{av}^T(R_X, R_Y) \leq E_{TL}(R_X, R_Y, P_{XYU}, W). \tag{3.83}$$

Proof. The proof is very similar to the proof of Corollary 3.1.2. □

Corollary 3.2.3. *In the low rate regime,*

$$E_{TL}(R_X, R_Y, P_{XYU}, W) = E_T(R_X, R_Y, P_{XYU}, W). \tag{3.84}$$

Proof. The proof is similar to the proof of Corollary 3.1.1 and is omitted. □

Theorem 3.2.3. *For every finite set $\mathcal{U}$, $\mathcal{P}_{XYU} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ such that $X - U - Y$, $R_X \geq 0$, $R_Y \geq 0$, $\delta > 0$, and $\mathbf{u} \in T_{P_U}$, there exists a multi-user code*

$$C = \{(\mathbf{x}_i, \mathbf{y}_j, D_{ij}) : i = 1, ..., M_X^*, j = 1, ..., M_Y^*\} \tag{3.85}$$

with $\mathbf{x}_i \in T_{P_{X|U}}(\mathbf{u})$, $\mathbf{y}_j \in T_{P_{Y|U}}(\mathbf{u})$ for all i and j, $M_X^* \geq \frac{2^{n(R_X-\delta)}}{2}$, and $M_Y^* \geq \frac{2^{n(R_Y-\delta)}}{2}$, *such that for every MAC* $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$

$$e(C, W) \leq 2^{-n[E_{ex}(R_X, R_Y, W, P_{XYU}) - 5\delta]} \tag{3.86}$$

whenever $n \geq n_1(|\mathcal{Z}|, |\mathcal{X}|, |\mathcal{Y}|, |\mathcal{U}|, \delta)$, *where*

$$E_{ex}(R_X, R_Y, W, P_{XYU}) \triangleq \min_{\beta=X,Y,XY} E_\beta^{ex}(R_X, R_Y, W, P_{UXY}, \mathcal{V}_\beta^{ex}). \tag{3.87}$$

Proof. The proof is provided in 3.4.2. □

This exponential error bound can be universally obtained for all MAC's with given input and output alphabets, since the choice of the codewords does not depend on the channel.

In the following, we show that the bounds in Theorems 3.2.1, 3.2.2, 3.2.3 are at least as good as the best known random coding bound, found in [38]. For this purpose, let us use the minimum equivocation decoding rule.

Definition 3.2.7. *Given* $\mathbf{u}$, *for a multiuser code*

$$C = \{(\mathbf{x}_i, \mathbf{y}_j, D_{ij}) : i = 1, ..., M_X, j = 1, ..., M_Y\}$$

we say that the D_{ij} *are minimum equivocation decoding sets for* $\mathbf{u}$, *if* $\mathbf{z} \in D_{ij}$ *implies*

$$H(\mathbf{x}_i\mathbf{y}_j|\mathbf{z}\mathbf{u}) = \min_{k,l} H(\mathbf{x}_k\mathbf{y}_l|\mathbf{z}\mathbf{u}).$$

It can be easily observed that these sets are equivalent to α*-decoding sets, where* $\alpha(\mathbf{u}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ *is defined as*

$$\alpha(V_{UXYZ}) \triangleq H_V(XY|ZU). \tag{3.88}$$

Here, V_{UXYZ} *is the joint empirical distribution of* $(\mathbf{u}, \mathbf{x}, \mathbf{y}, \mathbf{z})$.

Theorem 3.2.4. *For every finite set* $\mathcal{U}$, $\mathcal{P}_{XYU} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$, $R_X \geq 0$, $R_Y \geq 0$, *and* $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$*, and an appropriate* α*-decoder (minimum equivocation),*

$$E_\beta^r(R_X, R_Y, W, P_{XYU}) \geq E_{r\beta}^{Liu}(R_X, R_Y, W, P_{XYU}) \qquad \beta = X, Y, XY, \tag{3.89a}$$

$$E_\beta^T(R_X, R_Y, W, P_{XYU}) \geq E_{r\beta}^{Liu}(R_X, R_Y, W, P_{XYU}) \qquad \beta = X, Y, XY, \tag{3.89b}$$

$$E_\beta^{ex}(R_X, R_Y, W, P_{XYU}) \geq E_{r\beta}^{Liu}(R_X, R_Y, W, P_{XYU}) \qquad \beta = X, Y, XY. \tag{3.89c}$$

Hence

$$E_r(R_X, R_Y, W, P_{XYU}) \geq E_r^{Liu}(R_X, R_Y, W, P_{XYU}), \tag{3.90a}$$

$$E_T(R_X, R_Y, W, P_{XYU}) \geq E_r^{Liu}(R_X, R_Y, W, P_{XYU}), \tag{3.90b}$$

$$E_{ex}(R_X, R_Y, W, P_{XYU}) \geq E_r^{Liu}(R_X, R_Y, W, P_{XYU}), \tag{3.90c}$$

for all $P_{XYU} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ *satisfying* $X - U - Y$*. Here,* E_r^{Liu} *is the random coding exponent of [38] which is defined in* (2.23)*.*

Proof. The proof is provided in 3.4.2. □

We expect our typical random coding and expurgated bound to be strictly better than the one in [38] at low rates. This is so, because all inequalities in (3.64a)-(3.64c) and (3.66a)-(3.66c) will be active at zero rates, and thus (due to continuity) at sufficiently low rates. Although we have not been able to prove this fact rigorously, in the next section, we show that this is true by numerically evaluating the expurgated bound for different rate pairs.

3.3 Numerical result

In this section, we calculate the exponent derived in Theorem 3.2.3 for a multiple-access channel very similar to the one used in [38]. This example shows that strict inequality can hold in (3.89c). Consider a discrete memoryless MAC with $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$ and the transition probability given in Table 3.1. First, we choose some time-sharing alphabet $\mathcal{U}$ of size $|\mathcal{U}| = 4$. Then some channel input distribution

x	y	z	$W(z\|xy)$
0	0	0	0.99
0	0	1	0.01
0	1	0	0.01
0	1	1	0.99
1	0	0	0.01
1	0	1	0.99
1	1	0	0.50
1	1	1	0.50

Table 3.1: Channel Statistics

$P_U P_{X|U} P_{Y|U}$ is chosen randomly. Table 3.2 gives numerical values of the random coding exponent of [38], and the expurgated exponent we have obtained for selected rate pairs. As we see in the table, in the low rate regime, we have strictly better results in comparison with the results of [38]. For larger rate pairs, the inequalities containing $\min\{R_X, R_Y\}$ will not be active anymore, thus, we will end up with result similar to [38].

3.4 Proof of Theorems

3.4.1 Point to Point Proofs

This section contains the proof of all lemmas and theorems related to point to point result.

Proof. **(Lemma 3.1.1)** We use the method of random selection. Define M such that

$$2^{n(R-\delta)} \leq M \leq 2^{nR}. \tag{3.91}$$

In the following, we obtain the expectation of the packing functions over the constant composition code ensemble. The expectation of $\pi(X^M, V_{X\tilde{X}})$ can be obtained as

R_X	R_Y	$E_{ex}(R_X, R_Y, W, P_{UXY})$	$E_r^{Liu}(R_X, R_Y, W, P_{UXY})$
0.01	0.01	0.2672	0.2330
0.01	0.02	0.2671	0.2330
0.01	0.03	0.2671	0.2330
0.02	0.01	0.2458	0.2230
0.02	0.02	0.2379	0.2230
0.02	0.05	0.2379	0.2230
0.03	0.01	0.2279	0.2130
0.03	0.03	0.2183	0.2130
0.04	0.01	0.2123	0.2030
0.04	0.04	0.2040	0.2030
0.05	0.05	0.1930	0.1930
0.06	0.01	0.1856	0.1830
0.06	0.06	0.1830	0.1830
0.07	0.01	0.1740	0.1730
0.07	0.07	0.1730	0.1730

Table 3.2: E_{ex} vs. E_r^{Liu}

follows:

$$\begin{aligned}\mathbb{E}\left(\pi(X^M, V_{X\tilde{X}})\right) &= \frac{1}{M}\sum_{i=1}^{M}\sum_{j\neq i}\mathbb{E}\left(1_{T_{V_{X\mathcal{X}}}}(X_i, X_j)\right) = \frac{1}{M}\sum_{i=1}^{M}\sum_{j\neq i}\mathbb{P}\left(X_j \in T_{V_{\tilde{X}|X}}(X_i)\right)\\ &= (M-1)\mathbb{P}\left(X_2 \in T_{V_{\tilde{X}|X}}(X_1)\right) \leq 2^{n(R-I_V(X\wedge\tilde{X})+\delta)}. \end{aligned} \tag{3.92}$$

Similarly, it can be shown that for sufficiently large n,

$$\mathbb{E}\left(\pi(X^M, V_{X\tilde{X}})\right) \geq 2^{n(R-I_V(X\wedge\tilde{X})-\delta)}. \tag{3.93}$$

The expectation of λ over the ensemble can be written as

$$\mathbb{E}\left(\lambda(X^M, V_{X\tilde{X}\hat{X}})\right) = \frac{1}{M}\sum_{i=1}^{M}\sum_{j\neq i}\sum_{k\neq i,j}\mathbb{P}\left((X_i, X_j, X_k) \in T_{V_{X\tilde{X}\hat{X}}}\right). \tag{3.94}$$

Since

$$\frac{2^{n[H(\tilde{X}\hat{X}|X)-\delta]}}{2^{nH(\tilde{X})}2^{nH(\hat{X})}} \leq \mathbb{P}\left((X_i, X_j, X_k) \in T_{V_{X\tilde{X}\hat{X}}}\right) \leq \frac{2^{nH(\tilde{X}\hat{X}|X)}}{2^{n[H(\tilde{X})-\delta]}2^{n[H(\hat{X})-\delta]}}, \tag{3.95}$$

it can be concluded that

$$2^{n[E_S(R,V_{X\tilde{X}\hat{X}})-2\delta]} \leq \mathbb{E}\left(\lambda(X^M, V_{X\tilde{X}\hat{X}})\right) \leq 2^{n[E_S(R,V_{X\tilde{X}\hat{X}})+2\delta]}, \tag{3.96}$$

where E_S is defined as follows,

$$E_S(R, V_{X\tilde{X}\hat{X}}) \triangleq 2R - I(X \wedge \tilde{X}) - I(\hat{X} \wedge \tilde{X}X). \tag{3.97}$$

By using (3.92) and markov inequality, it can be concluded that

$$\mathbb{P}\left(\pi(X^M, V_{X\tilde{X}}) \geq 2^{n(R-I_V(X\wedge\tilde{X})+2\delta)} \text{ for some } V_{X\tilde{X}}\right) \leq \sum_{V_{X\tilde{X}}} \frac{\mathbb{E}\left(\pi(X^M, V_{X\tilde{X}})\right)}{2^{n(R-I_V(X\wedge\tilde{X})+2\delta)}} \leq 2^{-n\frac{\delta}{2}}, \tag{3.98}$$

therefore, there exists at least one code, C^r, with M codewords satisfying

$$\pi(C^r, V_{X\tilde{X}}) \leq 2^{n(R-I_V(X\wedge\tilde{X})+2\delta)}. \tag{3.99}$$

□

Proof. **(Lemma 3.1.2)** To prove that a specific property holds for almost all codes, with certain number of codewords, in the constant composition code ensemble, we use a second-order argument method. We already have obtained upper and lower bounds on the expectation of the desired function over the entire ensemble. In the following, we derive an upper bound on the variance of the packing function. Finally, by using the Chebychev's inequality, we prove that the desired property holds for almost all codes in the ensemble.

To find the variance of the packing function, let us define $U_{ij} \triangleq 1_{T_{V_{X\tilde{X}}}}(X_i, X_j)$, and $Y_{ij} \triangleq U_{ij} + U_{ji}$. We can rewrite $\pi(X^M, V_{X\tilde{X}})$ as

$$\pi(X^M, V_{X\tilde{X}}) = \frac{1}{M}\sum_{i=1}^{M}\sum_{j\neq i} U_{ij} = \frac{1}{M}\sum_{i=1}^{M}\sum_{j<i}(U_{ij} + U_{ji}) = \frac{1}{M}\sum_{i=1}^{M}\sum_{j<i} Y_{ij}. \tag{3.100}$$

It is easy to check that Y_{ij}'s are identically distributed pairwise independent random variables. Therefore, the variance of $\pi(X^M, V_{X\tilde{X}})$ can be written as

$$Var\left(\pi(X^M, V_{X\tilde{X}})\right) = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j<i} Var(Y_{ij}) = \frac{1}{M^2}\binom{M}{2} Var(Y_{21}). \tag{3.101}$$

To find the variance of Y_{21}, let us consider the following two cases for $V_{X\tilde{X}}$:

- $V_{X\tilde{X}}$ is a symmetric distribution. In this case $U_{12} = U_{21}$, therefore,

$$Y_{21} = \begin{cases} 2 & \text{with probability } p \leq 2^{-n[I(X\wedge\tilde{X})-\delta]} \\ 0 & \text{with probability } 1-p \end{cases},$$

  and the variance is upper bounded by

$$Var(Y_{21}) \leq E(Y_{21}^2) = 4 \times 2^{-n[I(X\wedge\tilde{X})-\delta]}. \tag{3.102}$$

- $V_{X\tilde{X}}$ is not a symmetric distribution. In this case, if $U_{ij} = 1 \Rightarrow U_{ji} = 0$. Therefore,

$$\begin{aligned}\mathbb{P}\left(Y_{12} = 1\right) = \mathbb{P}\left(U_{12} = 1 \text{ or } U_{21} = 1\right) &= \mathbb{P}\left(U_{12} = 1\right) + \mathbb{P}\left(U_{21} = 1\right) \\ &\leq 2 \times 2^{-n[I(X\wedge\tilde{X})-\delta]}, \end{aligned} \tag{3.103}$$

  therefore,

$$Var(Y_{21}) \leq E(Y_{21}^2) = 2 \times 2^{-n[I(X\wedge\tilde{X})-\delta]}. \tag{3.104}$$

By using (3.102), and (3.104), we have

$$Var\left(\pi(X^M, V_{X\tilde{X}})\right) \leq \frac{1}{M^2}\binom{M}{2} 4 \times 2^{-n[I(X\wedge\tilde{X})-\delta]} \leq 2 \times 2^{-n[I(X\wedge\tilde{X})-\delta]}, \tag{3.105}$$

for any $V_{X\tilde{X}} \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$. Now, by using Chebychev's inequality,

$$\begin{aligned}
&\mathbb{P}\left(\left|\pi(X^M, V_{X\tilde{X}}) - \mathbb{E}\left(\pi(X^M, V_{X\tilde{X}})\right)\right| \geq 2^{n\delta} \text{ for some } V_{X\tilde{X}}\right) \\
&\quad \leq \sum_{V_{X\tilde{X}}} \mathbb{P}\left(\left|\pi(X^M, V_{X\tilde{X}}) - \mathbb{E}\left(\pi(X^M, V_{X\tilde{X}})\right)\right| \geq 2^{n\delta}\right) \\
&\quad \leq \sum_{V_{X\tilde{X}}} \frac{Var\left(\pi(X^M, V_{X\tilde{X}})\right)}{2^{2n\delta}} \leq \sum_{V_{X\tilde{X}}} \frac{2 \times 2^{-n[I(X\wedge\tilde{X})-\delta]}}{2^{2n\delta}} \\
&\quad = \sum_{V_{X\tilde{X}}} 2 \times 2^{-n(I(X\wedge\tilde{X})+\delta)} \leq 2^{-n\frac{\delta}{2}}, \qquad \text{for sufficiently large } n. \qquad (3.106)
\end{aligned}$$

Moreover, by using (3.96) and Markov's inequality, it can be concluded that

$$\begin{aligned}
\mathbb{P}\left(\lambda(X^M, V_{X\tilde{X}\hat{X}}) \geq 2^{n[E_S(R,V_{X\tilde{X}\hat{X}})+4\delta]} \text{ for some } V_{X\tilde{X}\hat{X}}\right) &\leq \sum_{V_{X\tilde{X}\hat{X}}} \frac{\mathbb{E}\lambda(X^M, V_{X\tilde{X}\hat{X}})}{2^{n[E_S(R,V_{X\tilde{X}\hat{X}})+4\delta]}} \\
&\leq 2^{-n\delta}. \qquad (3.107)
\end{aligned}$$

Now, by combining (3.106) and (3.107) and using the bound on $\mathbb{E}\left(\pi(X^M, V_{X\tilde{X}})\right)$, we conclude that for any $V_{X\tilde{X}} \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$, any $V_{X\tilde{X}\hat{X}} \in \mathcal{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{X})$, for sufficiently large n

$$\begin{aligned}
2^{n(R-I(X\wedge\tilde{X})-\delta)} \leq \pi(X^M, V_{X\tilde{X}}) \leq 2^{n(R-I(X\wedge\tilde{X})+\delta)}, \\
\lambda(X^M, V_{X\tilde{X}\hat{X}}) \leq 2^{n[E_S(R,V_{X\tilde{X}\hat{X}})+4\delta]}, \qquad (3.108)
\end{aligned}$$

with probability $> 1 - 2 \times 2^{-n\frac{\delta}{2}}$. We put all the codebooks satisfying (3.108) in a set called $\mathcal{C}^T$.

□

Proof. **(Lemma 3.1.3)** Consider the code $C^r \triangleq \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M\}$ whose existence is asserted in random coding packing lemma. Let us define

$$\Pi(C^r) \triangleq \sum_{V_{X\tilde{X}}} 2^{-n(R-I_V(X\wedge\tilde{X})+3\delta)}\pi(C^r, V_{X\tilde{X}}). \qquad (3.109)$$

Note that using Lemma 3.1.1 and using the fact that

$$\Pi(C^r) = \frac{1}{M}\sum_{i=1}^{M}\left\{\sum_{V_{X\tilde{X}}} |T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C^r| 2^{-n(R-I_V(X\wedge\tilde{X})+3\delta)}\right\}, \tag{3.110}$$

it can be concluded that

$$\Pi(C^r) \leq \sum_{V_{X\tilde{X}}} 2^{-n(R-I_V(X\wedge\tilde{X})+3\delta)} 2^{n(R-I_V(X\wedge\tilde{X})+2\delta)} < \frac{1}{2}. \tag{3.111}$$

As a result, it can be concluded that there exists $M^* \geq \frac{M}{2}$ codewords in C^r satisfying

$$\sum_{V_{X\tilde{X}}} |T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C^r| 2^{-n(R-I_V(X\wedge\tilde{X})+3\delta)} < 1. \tag{3.112}$$

Let us call this subset of the code as C^{ex}. Without loss of generality, we assume C^{ex} contains the first M^* sequences of C^r, i.e., $C^{ex} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{M^*}\}$. Since

$$|T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C^{ex}| \leq |T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C^r| \quad \forall \mathbf{x}_i \in C^{ex}, \tag{3.113}$$

it can be concluded that for all $\mathbf{x}_i \in C^{ex}$,

$$\sum_{V_{X\tilde{X}}} |T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C^{ex}| 2^{-n(R-I_V(X\wedge\tilde{X})+3\delta)} < 1. \tag{3.114}$$

Since all the terms in the summation are non-negative terms, we conclude that

$$|T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C^{ex}| < 2^{n(R-I_V(X\wedge\tilde{X})+3\delta)}, \tag{3.115}$$

for all $V_{X\tilde{X}} \in \mathcal{P}(\mathcal{X}\times\mathcal{X})$, and all $\mathbf{x}_i \in C^{ex}$. Also, by (3.115), it can be concluded that

$$\pi(C^{ex}, V_{X\tilde{X}}) = \frac{1}{M^*}\sum_{i=1}^{M^*} |T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C^{ex}| \leq 2^{n(R-I_V(X\wedge\tilde{X})+3\delta)}, \tag{3.116}$$

for all $V_{X\tilde{X}} \in \mathcal{P}(\mathcal{X}\times\mathcal{X})$. □

Proof. **(Fact 3.1.1)** We will use the result of Lemma 3.1.1 and the relation between the probability of error and the packing functions. Let $X^M \triangleq \Big(X_1, X_2, ..., X_M\Big)$ be independent sequences of independent random variable, where X_is are uniformly distributed on T_P.

(Upper Bound): Taking expectation on both sides of (3.15), using Lemma 3.1.1 and using the continuity of information measures, it can be concluded that

$$\begin{aligned}\mathbb{E}\left(e(X^M, W)\right) &\leq \sum_{V_{X\tilde{X}Y} \in \mathcal{P}_n^r} 2^{-n[D(V_{Y|X}||W|P)+|I(\tilde{X}\wedge XY)-R|^+-\delta]} \\ &\leq 2^{-n[E_r(R,P,W)-2\delta]} \end{aligned} \tag{3.117}$$

whenever $n \geq n_1(|\mathcal{X}|, |\mathcal{Y}|, \delta)$, where

$$E_r(R, P, W) \triangleq \min_{V_{X\tilde{X}Y} \in \mathcal{P}^r} D(V_{Y|X}||W|P) + |I_V(XY \wedge \tilde{X}) - R|^+, \tag{3.118}$$

and $\mathcal{P}^r$ is defined in (3.29).

(Lower Bound): Taking expectation on both sides of (3.19), and using Lemma 3.1.1 we have

$$\bar{P}_e = \mathbb{E}e(X^M, W) \geq \sum_{V_{X\tilde{X}Y} \in \mathcal{P}_n^r} 2^{-n[D(V_{Y|X}||W|P)+I_V(\tilde{X}\wedge Y|X)+\delta]} \Bigg| 2^{n(R-I(X\wedge\tilde{X})-\delta)} - \sum_{\substack{V_{X\tilde{X}\hat{X}Y}: \\ V_{X\hat{X}Y}=V_{X\tilde{X}Y}}} 2^{-n[I_V(\hat{X}\wedge Y|X\tilde{X})]} 2^{n(2R-I(X\wedge\tilde{X})-I(\hat{X}\wedge X\tilde{X})+4\delta)} \Bigg|^+ \tag{3.119}$$

$$= \sum_{V_{X\tilde{X}Y} \in \mathcal{P}_n^r} 2^{-n[D(V_{Y|X}||W|P)+I_V(\tilde{X}\wedge XY)-R+2\delta]} \left| 1 - \sum_{\substack{V_{X\tilde{X}\hat{X}Y}: \\ V_{X\hat{X}Y}=V_{X\tilde{X}Y}}} 2^{-n[I_V(\hat{X}\wedge X\tilde{X}Y)-R+3\delta]} \right|^+ \tag{3.120}$$

Toward further simplification of this expression, we have the following lemma.

Lemma 3.4.1.

$$\min_{\substack{V_{X\tilde{X}\hat{X}Y}: \\ V_{X\hat{X}Y}=V_{X\tilde{X}Y}}} I(\hat{X} \wedge X\tilde{X}Y) = I(\tilde{X} \wedge XY). \tag{3.121}$$

Proof. Note that, for any $V_{X\tilde{X}\hat{X}Y}$,

$$I(\hat{X} \wedge X\tilde{X}Y) = I(\hat{X} \wedge XY) + I(\hat{X} \wedge \tilde{X}|XY) \geq I(\hat{X} \wedge XY), \tag{3.122}$$

therefore,

$$\min_{\substack{V_{X\tilde{X}\hat{X}Y}: \\ V_{X\hat{X}Y}=V_{X\tilde{X}Y}}} I(\hat{X} \wedge X\tilde{X}Y) \geq I(\hat{X} \wedge XY) = I(\tilde{X} \wedge XY). \tag{3.123}$$

Now, consider $V^*_{X\tilde{X}\hat{X}Y}$ defined as

$$V^*_{X\tilde{X}\hat{X}Y}(x, \tilde{x}, \hat{x}, y) = V_{\tilde{X}|XY}(\tilde{x}|x, y)V_{\tilde{X}|XY}(\hat{x}|x, y)V_{XY}(x, y). \tag{3.124}$$

Note that $V^*_{X\hat{X}Y} = V^*_{X\tilde{X}Y}$, and $\tilde{X} - (X, Y) - \hat{X}$. Therefore,

$$I_{V^*}(\hat{X} \wedge X\tilde{X}Y) = I_V(\hat{X} \wedge XY) = I_V(\tilde{X} \wedge XY). \tag{3.125}$$

By combining (3.123) and (3.125), the proof is complete. □

Therefore, using the above lemma, (3.120) can be rewritten as

$$\bar{P}_e \geq \sum_{\substack{V_{X\tilde{X}Y} \in \mathcal{P}_n^r \\ I(\tilde{X} \wedge XY) > R+3\delta}} 2^{-n[D(V_{Y|X}||W|P)+I_V(\tilde{X}\wedge XY)-R+3\delta]}. \tag{3.126}$$

By using the continuity of information measures, it can be concluded that

$$\mathbb{E}\left(e(X^M, W)\right) \geq 2^{-n[E_L(R,P,W)+4\delta]}, \qquad \text{for sufficient large } n \tag{3.127}$$

where

$$E_L(R,P,W) \triangleq \min_{\substack{V_{X\tilde{X}Y}\in\mathcal{P}^r \\ I(\tilde{X}\wedge XY)\geq R}} D(V_{Y|X}||W|P) + I_V(XY\wedge\tilde{X}) - R. \tag{3.128}$$

Now, by using Markov inequality and (3.117), we conclude that

$$\mathbb{P}\left(e(X^M,W) \geq 2^{-n[E_r(R,P,W)-3\delta]}\right) \leq \frac{\mathbb{E}\left(e(X^M,W)\right)}{2^{-n[E_r(R,P,W)-3\delta]}} \leq 2^{-n\delta}. \tag{3.129}$$

Therefore, with probability greater than $1-2^{-n\delta}$, any selected code with M codewords form the constant composition code ensemble satisfies the desired property. Let us call one of these codebooks as C^r. □

Proof. **(Corollary 3.1.1)** Consider the input distribution $P^* \in \mathcal{P}(\mathcal{X})$ maximizing the random coding bound, i.e.,

$$P^* \triangleq \arg\max_{P\in\mathcal{P}(\mathcal{X})} E_r(R,P,W). \tag{3.130}$$

Let us define

$$V^*_{X\tilde{X}Y} \triangleq \arg\min_{V_{X\tilde{X}Y}} E_r(R,P^*,W). \tag{3.131}$$

For any $R \leq R_{crit}$, the random coding bound is a straight line with slope -1, and the term in $|\cdot|^+$ is active. Therefore,

$$E_r(R,P^*,W) = D(V^*_{Y|X}||W|P^*) + I_{V^*}(\tilde{X}\wedge XY) - R. \tag{3.132}$$

Here, $I_{V^*}(\tilde{X}\wedge XY) \geq R$. It is clear that $V^*_{X\tilde{X}Y}$ is the minimizing distribution in $E_{rL}(R,P^*,W)$, and as a result

$$E_{rL}(R,P^*,W) = E_r(R,P^*,W). \tag{3.133}$$

□

Proof. **(Theorem 3.1.1)** In the proof of Fact 3.1.1, we used the lower and upper bounds on the expected value of he first-order packing functions and an upper bound on the expected value of the second-order packing functions. In the following, we use similar techniques on the packing function of almost every codebook in the ensemble by using the bounds obtained in Lemma 3.1.2. Consider the code C whose existence is asserted in the typical random coding packing lemma. For all $V_{X\tilde{X}} \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$, we have

$$\frac{1}{M}\sum_{i=1}^{M} |T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C| \leq 2^{n(R-I_V(X\wedge\tilde{X})+2\delta)}. \tag{3.134}$$

By multiplying both sides of inequality (3.134) by M, and using the proper upper bound on the number of sequences in C, we conclude that

$$|T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C| \leq 2^{n(2R-I_V(X\wedge\tilde{X})+2\delta)} \qquad \forall i = 1, ..., M, \tag{3.135}$$

for all $V_{X\tilde{X}} \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$. We will obtain a higher error exponent for almost all codes by removing certain types from the constraint set $\mathcal{P}_n^r$. Consider any $V_{X\tilde{X}} \in \mathcal{P}(\mathcal{X}\times\mathcal{X})$ satisfying $I_V(X \wedge \tilde{X}) > 2(R+\delta)$. By (3.135),

$$|T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C| = 0 \text{ for all } i \;\Rightarrow \pi(C, V_{X\tilde{X}}) = 0. \tag{3.136}$$

Upper bound: Hence, by using (3.15) on C, and by using the result of Lemma 3.1.2, we have

$$e(C, W) \leq \sum_{V_{X\tilde{X}Y} \in \mathcal{P}_n^T(\delta)} 2^{-n[D(V_{Y|X}||W|P)+|I_V(XY\wedge\tilde{X})-R|^+-2\delta]}$$

where

$$\mathcal{P}_n^T(\delta) \triangleq \; \{V_{X\tilde{X}Y} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X}\times\mathcal{Y}) : \; V_X = V_{\tilde{X}} = P, \; I_V(X\wedge\tilde{X}) \leq 2R+2\delta$$
$$\alpha(P, V_{Y|\tilde{X}}) \leq \alpha(P, V_{Y|X})\}. \tag{3.137}$$

Using the continuity of information measures, the upper bound as given by the theorem follows.

Lower bound: Using (3.19) on C and using Lemma 3.1.2, we have

$$
\begin{aligned}
& e(C,W) \\
& \geq \sum_{V_{X\tilde{X}Y}\in\mathcal{P}_n^r} 2^{-n[D(V_{Y|X}||W|P)+I_V(\tilde{X}\wedge Y|X)+\delta]} \Bigg| \pi(C,V_{X\tilde{X}}) \\
& \qquad\qquad - \sum_{\substack{V_{X\tilde{X}\hat{X}Y}: \\ V_{X\hat{X}Y}=V_{X\tilde{X}Y}}} 2^{-n[I_V(\hat{X}\wedge Y|X\tilde{X})]} \lambda(C,V_{X\tilde{X}\hat{X}}) \Bigg|^+ \\
& \geq \sum_{V_{X\tilde{X}Y}\in\mathcal{P}_n^T(\delta)} 2^{-n[D(V_{Y|X}||W|P)+I_V(\tilde{X}\wedge Y|X)+\delta]} \Bigg| 2^{n(R-I(X\wedge\tilde{X})-\delta)} \\
& \qquad\qquad - \sum_{\substack{V_{X\tilde{X}\hat{X}Y}: \\ V_{X\hat{X}Y}=V_{X\tilde{X}Y}}} 2^{-n[I_V(\hat{X}\wedge Y|X\tilde{X})]} 2^{n(2R-I(X\wedge\tilde{X})-I(\hat{X}\wedge X\tilde{X})+2\delta)} \Bigg|^+ \\
& = \sum_{V_{X\tilde{X}Y}\in\mathcal{P}_n^T(\delta)} 2^{-n[D(V_{Y|X}||W|P)+I_V(\tilde{X}\wedge XY)-R+2\delta]} \left| 1 - \sum_{\substack{V_{X\tilde{X}\hat{X}Y}: \\ V_{X\hat{X}Y}=V_{X\tilde{X}Y}}} 2^{-n[I_V(\hat{X}\wedge X\tilde{X}Y)-R-3\delta]} \right|^+ \qquad (3.138) \\
& \geq \sum_{\substack{V_{X\tilde{X}Y}\in\mathcal{P}_n^T(\delta) \\ I(\tilde{X}\wedge XY)>R+5\delta}} 2^{-n[D(V_{Y|X}||W|P)+I_V(\tilde{X}\wedge XY)-R+3\delta]},
\end{aligned}
$$

Here, the last inequality follows from Lemma 3.4.1.

By using the continuity argument, and for sufficient large n,

$$e(C,W) \geq 2^{-n[E_{LT}(R,P,W)+4\delta]}, \qquad (3.139)$$

where

$$E_{LT}(R,P,W) \triangleq \min_{\substack{V_{X\tilde{X}Y}\in\mathcal{P}^T \\ I(\tilde{X}\wedge XY)\geq R}} D(V_{Y|X}||W|P) + I_V(XY\wedge\tilde{X}) - R. \qquad (3.140)$$

$\square$

Proof. (**Corollary 3.1.2**) Fix $R \geq 0$, $\delta > 0$. By the result of Theorem 3.1.1 and for sufficiently large n, there exists a collection of codes, $\mathcal{C}^*$, with length n and rate R, such that

- $\mathbb{P}(\mathcal{C}^*) \geq 1 - \delta$,
- $2^{-n[E_{TL}(R,P,W)+4\delta]} \leq e(C, W) \leq 2^{-n[E_T(R,P,W)-3\delta]}$ for all $C \in \mathcal{C}^*$.

Note that

$$\max_{\tilde{\mathcal{C}}:\mathbb{P}(\tilde{\mathcal{C}})>1-\delta} \min_{C\in\tilde{\mathcal{C}}} -\frac{1}{n} \log e(C, W) \geq \min_{C\in\mathcal{C}^*} -\frac{1}{n} \log e(C, W) \geq E_T(R, P, W) - 3\delta. \quad (3.141)$$

Now, consider any high probability collection of codes with length n and rate R. Let us call this collection as $\hat{\mathcal{C}}$. Note that

$$\left.\begin{array}{r} \mathbb{P}(\mathcal{C}^*) \geq 1 - \delta \\ \mathbb{P}(\hat{\mathcal{C}}) \geq 1 - \delta \end{array}\right\} \Rightarrow \mathbb{P}(\mathcal{C}^* \cap \hat{\mathcal{C}}) \geq 1 - 2\delta \Rightarrow \mathcal{C}^* \cap \hat{\mathcal{C}} \neq \phi. \quad (3.142)$$

Consider a code $C(\hat{\mathcal{C}}) \in \mathcal{C}^* \cap \hat{\mathcal{C}}$. It can be concluded that

$$\max_{\tilde{\mathcal{C}}:\mathbb{P}(\tilde{\mathcal{C}})>1-\delta} \min_{C\in\tilde{\mathcal{C}}} -\frac{1}{n} \log e(C, W) \leq \max_{\tilde{\mathcal{C}}:\mathbb{P}(\tilde{\mathcal{C}})>1-\delta} -\frac{1}{n} \log e(C(\tilde{\mathcal{C}}), W) \leq E_{LT}(R, P, W) + 4\delta. \quad (3.143)$$

The last inequality follows from the fact that $C(\hat{\mathcal{C}}) \in \mathcal{C}^*$. By combining (3.141) and (3.143), and by letting δ goes to zero and n goes to infinity, it can be concluded that

$$E_T(R, P, W) \leq E^T_{av}(R) \leq E_{TL}(R, P, W). \quad (3.144)$$

□

Proof. **(Fact 3.1.2)** First, we prove the following lemma.

Lemma 3.4.2. *Let C^{ex} be the collection of the codewords whose existence is asserted in Lemma 3.1.3. For any distribution $V_{X\tilde{X}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X})$, satisfying $I_V(X \wedge \tilde{X}) > R+\delta$,* *the following holds:*

$$\pi(C^{ex}, V_{X\tilde{X}}) = 0. \tag{3.145}$$

Proof. By (3.25),

$$|T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C^{ex}| \leq 2^{n(R - I_V(X\wedge\tilde{X}) + 2\delta)}, \tag{3.146}$$

for every $\mathbf{x}_i \in C^{ex}$. Since $I_V(X \wedge \tilde{X}) > R + 2\delta$, it can be concluded that

$$|T_{V_{\tilde{X}|X}}(\mathbf{x}_i) \cap C^{ex}| = 0 \text{ for every } \mathbf{x}_i \in C^{ex} \Rightarrow \pi(C^{ex}, V_{X\tilde{X}}) = 0 \tag{3.147}$$

□

The rest of the proof is identical to the proof of random coding bound. □

3.4.2 MAC Proofs

Proof. **(Lemma 3.2.1)** In this proof, we use a similar random coding argument that Pokorny and Wallmeier used in [41]. The main difference is that our lemma uses a different code ensemble which results in a tighter bound. Instead of choosing our sequences from T_{P_X} and T_{P_Y}, we choose our random sequences uniformly from $T_{P_{X|U}}(\mathbf{u})$, and $T_{P_{Y|U}}(\mathbf{u})$ for a given $\mathbf{u} \in T_{P_U}$. In [38], we see a similar random code ensemble, however, their packing lemma incorporates the channel output $\mathbf{z}$ into the packing inequalities. One can easily show that, by using this packing lemma and considering the minimum equivocation decoding rule, we would end up with the random coding bound derived in [38].

Fix any $\mathcal{U}$, $P_{XYU} \in \mathcal{P}_n(\mathcal{U} \times \mathcal{X} \times \mathcal{Y})$ such that $X - U - Y$, $R_X \geq 0$, $R_Y \geq 0$, $\delta > 0$, and $\mathbf{u} \in T_{P_U}$. Define M_X, M_Y such that

$$2^{n(R_X-\delta)} \leq M_X \leq 2^{nR_X}, \qquad 2^{n(R_Y-\delta)} \leq M_Y \leq 2^{nR_Y}.$$

First, we find upper bounds on the expectations of packing functions for a fixed α and $V_{UXY\tilde{X}\tilde{Y}}$, with respect to the random variables X_i and Y_j. Since X_is and Y_js are i.i.d random sequences, we have

$$\begin{aligned}
\mathbb{E}\Big[N_U(X^{M_X} \times Y^{M_Y}, V_{UXY})\Big] &\triangleq \mathbb{E}\Big[\frac{1}{M_X M_Y}\sum_{i,j} 1_{T_{V_{UXY}}}(\mathbf{u}, X_i, Y_j)\Big] \\
&= \mathbb{E}\big[1_{T_{V_{UXY}}}(\mathbf{u}, X_1, Y_1)\big] \\
&= \sum_{\mathbf{x},\mathbf{y}} 1_{T_{V_{XY|U}}}(\mathbf{x}, \mathbf{y}|\mathbf{u})\mathbb{P}(X_1 = \mathbf{x}|\mathbf{u})\mathbb{P}(Y_1 = \mathbf{y}|\mathbf{u}) \\
&\leq \sum_{(\mathbf{x},\mathbf{y})\in T_{V_{XY|U}}(\mathbf{u})} 2^{-n[H_V(X|U)-\delta]}2^{-n[H_V(Y|U)-\delta]} \\
&\leq 2^{nH_V(XY|U)}2^{-n[H_V(X|U)-\delta]}2^{-n[H_V(Y|U)-\delta]} \\
&= 2^{-n[I_V(X\wedge Y|U)-2\delta]} = 2^{-n[F_U(V_{UXY})-2\delta]}. \qquad (3.148)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}\Big[N_U(X^{M_X} \times Y^{M_Y}, V_{UXY})\Big] &= \sum_{\mathbf{x},\mathbf{y}} 1_{T_{V_{XY|U}}}(\mathbf{x}, \mathbf{y}|\mathbf{u})\mathbb{P}(X_1 = \mathbf{x}|\mathbf{u})\mathbb{P}(Y_1 = \mathbf{y}|\mathbf{u}) \\
&\geq \sum_{(\mathbf{x},\mathbf{y})\in T_{V_{XY|U}}(\mathbf{u})} 2^{-nH_V(X|U)}2^{-nH_V(Y|U)} \\
&\geq 2^{n[H_V(XY|U)-\delta]}2^{-nH_V(X|U)}2^{-nH_V(Y|U)} \\
&= 2^{-n[I_V(X\wedge Y|U)+\delta]} = 2^{-n[F_U(V_{UXY})+\delta]}. \qquad (3.149)
\end{aligned}$$

Therefore, by (3.148) and (3.149),

$$2^{-n[F_U(V_{UXY})+\delta]} \leq \mathbb{E}\Big[N_U(X^{M_X} \times Y^{M_Y}, V_{UXY})\Big] \leq 2^{-n[F_U(V_{UXY})-2\delta]}. \qquad (3.150)$$

By using a similar argument,

$$\mathbb{E}\Big[N_X(X^{M_X} \times Y^{M_Y}, V_{UXY\tilde{X}})\Big] \leq 2^{-n[F_X(V_{UXY\tilde{X}})-4\delta]}. \qquad (3.151)$$

On the other hand,

$$
\begin{aligned}
\mathbb{E}\Big[N_X(X^{M_X}\times Y^{M_Y},V)\Big] &\geq (M_X-1)\,\mathbb{E}\Big[1_{T_{V_{UXY}}}(\mathbf{u},X_1,Y_1)1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u},X_1,Y_1,X_2)\Big]\\
&= (M_X-1)\sum_{\mathbf{x},\mathbf{y}}\mathbb{P}(X_1=\mathbf{x}|\mathbf{u})\mathbb{P}(Y_1=\mathbf{y}|\mathbf{u})1_{T_{V_{UXY}}}(\mathbf{u},\mathbf{x},\mathbf{y})\\
&\qquad\qquad \cdot\sum_{\tilde{\mathbf{x}}}\mathbb{P}(X_2=\tilde{\mathbf{x}}|\mathbf{u})1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u},\mathbf{x},\mathbf{y},\tilde{\mathbf{x}})\\
&\geq (M_X-1)\sum_{\mathbf{x},\mathbf{y}\in T_{V_{XY|U}}(\mathbf{u})}2^{-nH_V(X|U)}2^{-nH_V(Y|U)}\\
&\qquad \sum_{\tilde{\mathbf{x}}\in T_{V_{\tilde{X}|UXY}}(\mathbf{u},\mathbf{x},\mathbf{y})}2^{-nH_V(\tilde{X}|U)}\\
&\geq (M_X-1)\,2^{n[H(XY|U)-\delta]}2^{-nH_V(X|U)}2^{-nH_V(Y|U)}\\
&\qquad\qquad \cdot 2^{n[H_V(\tilde{X}|UXY)-\delta]}2^{-nH_V(\tilde{X}|U)}\\
&\geq 2^{-n[I_V(X\wedge Y|U)+I_V(\tilde{X}\wedge Y|U)+I_V(\tilde{X}\wedge X|UY)-R_X+3\delta]}\\
&= 2^{-n[F_X(V_{UXY\tilde{X}})+3\delta]}.
\end{aligned}
\tag{3.152}
$$

Therefore, by (3.151) and (3.152),

$$
2^{-n[F_X(V_{UXY\tilde{X}})+3\delta]}\leq\mathbb{E}\Big[N_X(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}})\Big]\leq 2^{-n[F_X(V_{UXY\tilde{X}})-4\delta]}. \tag{3.153}
$$

By using a similar argument for $N_Y(.)$ and $N_{XY}(.)$, we can show that

$$
2^{-n[F_Y(V_{UXY\tilde{Y}})+3\delta]}\leq\mathbb{E}\Big[N_Y(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{Y}})\Big]\leq 2^{-n[F_Y(V_{UXY\tilde{Y}})-4\delta]}, \tag{3.154}
$$

$$
2^{-n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})+4\delta]}\leq\mathbb{E}\Big[N_{XY}(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}\tilde{Y}})\Big]\leq 2^{-n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})-4\delta]}. \tag{3.155}
$$

We can obtain an upper bound for $\mathbb{E}\left[\Lambda_{XY}(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}})\right]$ as follows

$$
\begin{aligned}
&\mathbb{E}\left[\Lambda_{XY}(X^{M_X},Y^{M_Y},V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}})\right]\\
&\quad=\mathbb{E}\left[\frac{1}{M_XM_Y}\sum_{i,j}\sum_{\substack{k\neq i\\ l\neq j}}\sum_{\substack{k'\neq i,k\\ l'\neq j,l}}1_{T_{V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}}}(\mathbf{u},X_i,Y_j,X_k,Y_l,X_{k'},Y_{l'})\right]\\
&\quad\leq M_X^2M_Y^2\mathbb{E}\Big[1_{T_{V_{UXY}}}(\mathbf{u},X_1,Y_1)1_{T_{V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}}}(\mathbf{u},X_1,Y_1,X_2,Y_2,X_3,Y_3)\Big]\\
&\quad=M_X^2M_Y^2\sum_{\mathbf{x},\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{y}},\hat{\mathbf{x}},\hat{\mathbf{y}}}\mathbb{P}(X_1=\mathbf{x},Y_1=\mathbf{y},X_2=\tilde{\mathbf{x}},Y_2=\tilde{\mathbf{y}},X_3=\hat{\mathbf{x}},Y_3=\hat{\mathbf{y}}|\mathbf{u})\\
&\qquad\qquad\cdot 1_{T_{V_{UXY}}}(\mathbf{u},\mathbf{x},\mathbf{y}).1_{T_{V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}}}(\mathbf{u},\mathbf{x},\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{y}},\hat{\mathbf{x}},\hat{\mathbf{x}})\\
&\quad=M_X^2M_Y^2\sum_{\mathbf{x},\mathbf{y}}\mathbb{P}(X_1=\mathbf{x}|\mathbf{u})Pr(Y_1=\mathbf{y}|\mathbf{u})\cdot 1_{T_{V_{UXY}}}(\mathbf{u},\mathbf{x},\mathbf{y})\\
&\qquad\cdot\sum_{\tilde{\mathbf{x}}}\mathbb{P}(X_2=\tilde{\mathbf{x}}|\mathbf{u})1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u},\mathbf{x},\mathbf{y},\tilde{\mathbf{x}})\cdot\sum_{\tilde{\mathbf{y}}}\mathbb{P}(Y_2=\tilde{\mathbf{y}}|\mathbf{u})1_{T_{V_{UXY\tilde{X}\tilde{Y}}}}(\mathbf{u},\mathbf{x},\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{y}})\\
&\qquad\cdot\sum_{\hat{\mathbf{x}}}\mathbb{P}(X_3=\hat{\mathbf{x}}|\mathbf{u})1_{T_{V_{UXY\tilde{X}\tilde{Y}\hat{X}}}}(\mathbf{u},\mathbf{x},\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{y}},\hat{\mathbf{x}})\\
&\qquad\qquad\cdot\sum_{\hat{\mathbf{y}}}\mathbb{P}(Y_3=\hat{\mathbf{y}}|\mathbf{u})1_{T_{V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}}}(\mathbf{u},\mathbf{x},\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{y}},\hat{\mathbf{x}},\hat{\mathbf{y}})\\
&\quad\leq M_X^2M_Y^2\sum_{\mathbf{x},\mathbf{y}\in T_{V_{XY|U}}(\mathbf{u})}2^{-n[H_V(X|U)-\delta]}2^{-n[H_V(Y|U)-\delta]}\sum_{\tilde{\mathbf{x}}\in T_{V_{\tilde{X}|UXY}}(\mathbf{u},\mathbf{x},\mathbf{y})}2^{-n[H_V(\tilde{X}|U)-\delta]}\\
&\qquad\cdot\sum_{\tilde{\mathbf{y}}\in T_{V_{\tilde{Y}|UXY\tilde{X}}}(\mathbf{u},\mathbf{x},\mathbf{y},\tilde{\mathbf{x}})}2^{-n[H_V(\tilde{Y}|U)-\delta]}\sum_{\hat{\mathbf{x}}\in T_{V_{\hat{X}|UXY\tilde{X}\tilde{Y}}}(\mathbf{u},\mathbf{x},\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{y}})}2^{-n[H_V(\hat{X}|U)-\delta]}\\
&\qquad\cdot\sum_{\hat{\mathbf{y}}\in T_{V_{\hat{Y}|UXY\tilde{X}\tilde{Y}\hat{X}}}}(\mathbf{u},\mathbf{x},\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{y}},\hat{\mathbf{x}})2^{-n[H_V(\hat{Y}|U)-\delta]}\\
&\quad\leq M_X^2M_Y^2\cdot 2^{nH(XY|U)}2^{-n[H_V(X|U)-\delta]}2^{-n[H_V(Y|U)-\delta]}2^{nH_V(\tilde{X}|UXY)}2^{-n[H_V(\tilde{X}|U)-\delta]}\\
&\qquad\cdot 2^{nH_V(\tilde{Y}|UXY\tilde{X})}2^{-n[H_V(\tilde{Y}|U)-\delta]}2^{nH_V(\hat{X}|UXY\tilde{X}\tilde{Y})}2^{-n[H_V(\hat{X}|U)-\delta]}2^{nH_V(\hat{Y}|UXY\tilde{X}\tilde{Y}\hat{X})}\\
&\qquad\cdot 2^{-n[H_V(\hat{Y}|U)-\delta]}\\
&\quad\leq 2^{-n[I(\tilde{X}\tilde{Y}\wedge XY|U)+I(\hat{X}\hat{Y}\wedge XY\tilde{X}\tilde{Y}|U)+I(X\wedge Y|U)+I(\tilde{X}\wedge\tilde{Y}|U)+I(\hat{X}\wedge\hat{Y}|U)-2R_X-2R_Y-6\delta]}\\
&\quad=2^{-n[E_S^{XY}(V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}})-6\delta]}.
\end{aligned}\tag{3.156}
$$

By using a similar argument, we can obtain the following bounds

$$\mathbb{E}\left[\Lambda_X(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{X}\hat{X}})\right] \leq 2^{-n[E_S^X(V_{UXY\tilde{X}\hat{X}})-4\delta]} \tag{3.157}$$

$$\mathbb{E}\left[\Lambda_Y(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{Y}\hat{Y}})\right] \leq 2^{-n[E_S^Y(V_{UXY\tilde{Y}\hat{Y}})-4\delta]} \tag{3.158}$$

Here, E_S^X, E_S^Y and E_S^{XY} are defined in (3.61a)-(3.61c).

By using Markov inequality, it can be concluded that

$$\mathbb{P}\left(N_U(X^{M_X}\times Y^{M_Y}, V_{UXY}) \geq 2^{-n[F_U(V_{UXY})-3\delta]} \text{ for some } V_{UXY}\right)$$
$$\leq \sum_{\substack{V_{UXY}:\\ V_{UX}=P_{UX}\\ V_{UY}=P_{UY}}} \frac{\mathbb{E}\left(N_U(X^{M_X}\times Y^{M_Y}, V_{UXY})\right)}{2^{-n[F_U(V_{UXY})-3\delta]}} \leq \sum_{\substack{V_{UXY}:\\ V_{UX}=P_{UX}\\ V_{UY}=P_{UY}}} 2^{-n\delta} \leq 2^{-n\frac{\delta}{2}} \tag{3.159}$$

Similarly, it can be shown that

$$\mathbb{P}\left(N_X(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{X}}) \geq 2^{-n[F_X(V_{UXY\tilde{X}})-5\delta]} \text{ for some } V_{UXY\tilde{X}}\right) \leq 2^{-n\frac{\delta}{2}}, \tag{3.160}$$

$$\mathbb{P}\left(N_Y(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{Y}}) \geq 2^{-n[F_Y(V_{UXY\tilde{Y}})-5\delta]} \text{ for some } V_{UXY\tilde{Y}}\right) \leq 2^{-n\frac{\delta}{2}}, \tag{3.161}$$

$$\mathbb{P}\left(N_{XY}(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{X}\tilde{Y}}) \geq 2^{-n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})-5\delta]} \text{ for some } V_{UXY\tilde{X}\tilde{Y}}\right) \leq 2^{-n\frac{\delta}{2}}. \tag{3.162}$$

Now, by combining (3.159)-(3.162), and using the union bound, it can be concluded that

$$\begin{aligned}\mathbb{P}\Big(&N_U(X^{M_X}\times Y^{M_Y}, V_{UXY}) \geq 2^{-n[F_U(V_{UXY})-3\delta]} \text{ for some } V_{UXY} \text{ or}\\ &N_X(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{X}}) \geq 2^{-n[F_X(V_{UXY\tilde{X}})-5\delta]} \text{ for some } V_{UXY\tilde{X}} \text{ or}\\ &N_Y(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{Y}}) \geq 2^{-n[F_Y(V_{UXY\tilde{Y}})-5\delta]} \text{ for some } V_{UXY\tilde{Y}} \text{ or}\\ &N_{XY}(X^{M_X}\times Y^{M_Y}, V_{UXY\tilde{X}\tilde{Y}}) \geq 2^{-n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})-5\delta]} \text{ for some } V_{UXY\tilde{X}\tilde{Y}}\Big)\\ &\leq 4\times 2^{-n\frac{\delta}{2}},\end{aligned} \tag{3.163}$$

therefore, there exists at least a multi-user code with the desired properties mentioned in (3.70)-(3.71). □

Proof. **(Lemma 3.2.2)** To prove that a specific property holds for almost all codes, with certain number of codewords, in the constant composition code ensemble, we use a second order argument method. We already have obtained upper and lower bounds on the expectation of the desired function over the entire ensemble. In the following, we derive an upper bound on the variance of the packing function. Finally, by using the Chebychev's inequality, we prove that the desired property holds for almost all codes in the ensemble. To find the variance of $N_U(X^{M_X} \times Y^{M_Y}, V_{UXY})$, let us define $W_{ij} \triangleq 1_{T_{V_{UXY}}}(\mathbf{u}, X_i, Y_j)$. Therefore, the variance of $N_U(X^{M_X} \times Y^{M_Y}, V_{UXY})$ can be written as

$$\begin{aligned} Var\left(N_U(X^{M_X} \times Y^{M_Y}, V_{UXY})\right) &= Var\left(\frac{1}{M_X M_Y}\sum_{i,j} 1_{T_{V_{UXY}}}(\mathbf{u}, X_i, Y_j)\right) \\ &= \frac{1}{M_X^2 M_Y^2} Var\left(\sum_{i,j} W_{ij}\right). \end{aligned} \tag{3.164}$$

Since W_{ij}'s are pairwise independent random variables, (3.164) can be written as

$$\begin{aligned} Var\left(N_U(X^{M_X} \times Y^{M_Y}, V_{UXY})\right) &= \frac{1}{M_X^2 M_Y^2}\sum_{i,j} Var\left(W_{ij}\right) \\ &\leq \frac{1}{M_X^2 M_Y^2}\sum_{i,j} \mathbb{E}\left(W_{ij}\right) \\ &\leq \frac{1}{M_X M_Y} \cdot 2^{-n[F_U(V_{UXY})-2\delta]} \leq 2^{-n[F_U(V_{UXY})+R_X+R_Y-2\delta]}. \end{aligned} \tag{3.165}$$

By defining $Q_{ik}^j \triangleq 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, X_i, Y_j, X_k)$, the variance of $N_X(X^{M_X} \times Y^{M_Y}, V_{UXY\tilde{X}})$

can be upper-bounded as follows

$$\begin{aligned}Var\left(N_X(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}})\right) &= Var\left(\frac{1}{M_XM_Y}\sum_{i,j}\sum_{k\neq i}1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u},X_i,Y_j,X_k)\right)\\ &= \frac{1}{M_X^2M_Y^2}Var\left(\sum_{i,j}\sum_{k\neq i}1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u},X_i,Y_j,X_k)\right)\\ &= \frac{1}{M_X^2M_Y^2}Var\left(\sum_{j}\sum_{i}\sum_{k\neq i}Q_{ik}^j\right)\\ &= \frac{1}{M_X^2M_Y^2}Var\left(\sum_{j}\sum_{i}\sum_{k<i}Q_{ik}^j+Q_{ki}^j\right)\\ &= \frac{1}{M_X^2M_Y^2}Var\left(\sum_{j}\sum_{i}\sum_{k<i}J_{i,k}^j\right), \end{aligned}\tag{3.166}$$

where $J_{i,k}^j \triangleq Q_{ik}^j+Q_{ki}^j$, $k<i$. One can show that $J_{i,k}^j$'s are identically pairwise independent random variables. Therefore, the $Var\left(N_X(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}})\right)$ can be written as

$$Var\left(N_X(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}})\right)=\frac{1}{M_X^2M_Y^2}\sum_{j}\sum_{i}\sum_{k<i}Var\left(J_{i,k}^j\right)\leq\frac{1}{2M_Y}Var\left(J_{2,1}^1\right).\tag{3.167}$$

To find the variance of $J_{2,1}^1$, let us consider the following two cases for $V_{UXY\tilde{X}}$:

- $V_{UXY\tilde{X}}$ is a symmetric distribution, i.e., $V_{UXY\tilde{X}}=V_{U\tilde{X}YX}$. In this case $Q_{12}^1=Q_{21}^1$, therefore,

$$J_{2,1}^1=\begin{cases}2 & \text{with probability } p\approx 2^{-n[I_V(X\wedge Y|U)+I_V(\tilde{X}\wedge XY|U)]}\\ 0 & \text{with probability } 1-p\end{cases},$$

  and the variance is upper bounded by

$$Var(J_{2,1}^1)\leq E({J_{2,1}^1}^2)=4\times 2^{-n[I_V(X\wedge Y|U)+I_V(\tilde{X}\wedge XY|Y)]},\tag{3.168}$$

- $V_{UXY\tilde{X}}$ is not a symmetric distribution. In this case, if $Q_{ik}^j=1\Rightarrow Q_{ki}^j=0$.

Therefore,

$$\mathbb{P}\left(J_{2,1}^{1}=1\right)=\mathbb{P}\left(Q_{12}^{1}=1 \text{ or } Q_{21}^{1}=1\right)=\mathbb{P}\left(Q_{12}^{1}=1\right)+\mathbb{P}\left(Q_{21}^{1}=1\right)$$
$$\leq 2\times 2^{-n[I_V(X\wedge Y|U)+I_V(\tilde{X}\wedge XY|U)]}, \tag{3.169}$$

therefore,

$$Var(J_{2,1}^{1})\leq E({J_{2,1}^{1}}^{2})=2\times 2^{-n[I_V(X\wedge Y|U)+I_V(\tilde{X}\wedge XY|U)]}. \tag{3.170}$$

By combining the results in (3.167)-(3.169), it can be concluded that

$$Var\left(N_X(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}}\right)\leq 2^{-n[I_V(X\wedge Y|U)+I_V(\tilde{X}\wedge XY|U)+R_Y-3\delta]}. \tag{3.171}$$

Similarly, it can be shown that

$$Var\left(N_Y(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{Y}})\right)\leq 2^{-n[I_V(X\wedge Y|U)+I_V(\tilde{Y}\wedge YX|U)+R_X-3\delta]}. \tag{3.172}$$

By defining

$$R_{ik}^{jl}\triangleq 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u},X_i,Y_j,X_k,Y_l), \tag{3.173}$$

the variance of $N_{XY}(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}\tilde{Y}})$ can be upper-bounded as follows

$$Var\left(N_{XY}(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}\tilde{Y}})\right)$$
$$=Var\left(\frac{1}{M_XM_Y}\sum_{i,j}\sum_{\substack{k\neq i\\ l\neq j}}1_{T_{V_{UXY\tilde{X}\tilde{Y}}}}(\mathbf{u},X_i,Y_j,X_k,Y_l)\right)$$

$$= \frac{1}{M_X^2 M_Y^2} Var\left(\sum_{i,j}\sum_{\substack{k\neq i\\ l\neq j}} 1_{T_{V_{UXY\tilde{X}\tilde{Y}}}}(\mathbf{u}, X_i, Y_j, X_k, Y_l)\right)$$

$$= \frac{1}{M_X^2 M_Y^2} Var\left(\sum_{i}\sum_{j}\sum_{k\neq i}\sum_{l\neq j} R_{ik}^{jl}\right)$$

$$= \frac{1}{M_X^2 M_Y^2} Var\left(\sum_{i}\sum_{j}\sum_{k<i}\sum_{l\neq j} \left\{R_{ik}^{jl} + R_{ki}^{jl} + R_{ik}^{lj} + R_{ki}^{lj}\right\}\right)$$

$$= \frac{1}{M_X^2 M_Y^2} Var\left(\sum_{i}\sum_{j}\sum_{k<i}\sum_{j<l} S_{i,k}^{j,l}\right), \tag{3.174}$$

where $S_{i,k}^{j,l} \triangleq R_{ik}^{jl} + R_{ki}^{jl} + R_{ik}^{lj} + R_{ki}^{lj}$, $k < i$, $l < j$. It is easy to check that $S_{i,k}^{j,l}$'s are identically pairwise independent random variables. Therefore, the $Var\left(N_{XY}(.)\right)$ can be written as

$$\begin{aligned} Var\left(N_{XY}(X^{M_X} \times Y^{M_Y}, V_{UXY\tilde{X}\tilde{Y}})\right) &= \frac{1}{M_X^2 M_Y^2} \sum_{i}\sum_{j}\sum_{k<i}\sum_{l<j} Var\left(S_{i,k}^{j,l}\right) \\ &\leq \frac{1}{4} Var\left(S_{1,2}^{1,2}\right). \end{aligned} \tag{3.175}$$

By using a similar argument to (3.168)-(3.169), the variance of To find the variance of $S_{1,2}^{1,2}$,can be upper bounded by

$$Var\left(S_{1,2}^{1,2}\right) \leq 16 \times 2^{-n[I_V(X\wedge Y|U)+I_V(\tilde{X}\wedge\tilde{Y}|U)+I_V(\tilde{X}\tilde{Y}\wedge XY|U)-4\delta]}, \tag{3.176}$$

and therefore,

$$Var\left(N_{XY}(X^{M_X} \times Y^{M_Y}, V_{UXY\tilde{X}\tilde{Y}})\right) \leq 4 \times 2^{-n[I_V(X\wedge Y|U)+I_V(\tilde{X}\wedge\tilde{Y}|U)+I_V(\tilde{X}\tilde{Y}\wedge XY|U)-4\delta]}. \tag{3.177}$$

Now, by using the Chebychev's inequality, we can obtain the following

$$\begin{aligned}
&\mathbb{P}\left(\left|N_U(X^{M_X}\times Y^{M_Y},V_{UXY})-\mathbb{E}\left(N_U(X^{M_X}\times Y^{M_Y},V_{UXY})\right)\right|\geq 2^{2n\delta}\text{ for some }V_{UXY}\right)\\
&\quad\leq\sum_{V_{UXY}}\mathbb{P}\left(\left|N_U(X^{M_X}\times Y^{M_Y},V_{UXY})-\mathbb{E}\left(N_U(X^{M_X}\times Y^{M_Y},V_{UXY})\right)\right|\geq 2^{2n\delta}\right)\\
&\quad\leq\sum_{V_{UXY}}\frac{Var\left(N_U(X^{M_X}\times Y^{M_Y},V_{UXY})\right)}{2^{4n\delta}}\\
&\quad\leq\sum_{V}2^{-n[F_U(V)+R_X+R_Y+2\delta]}\leq 2^{-n\delta}.
\end{aligned}\tag{3.178}$$

Similarly, it can be shown that

$$\mathbb{P}\begin{pmatrix}\left|N_X(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}})-\mathbb{E}N_X(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}})\right|\geq 2^{2n\delta}\\ \text{for some }V_{UXY\tilde{X}}\end{pmatrix}\leq 2^{-n\delta}\tag{3.179}$$

$$\mathbb{P}\begin{pmatrix}\left|N_Y(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{Y}})-\mathbb{E}N_Y(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{Y}})\right|\geq 2^{2n\delta}\\ \text{for some }V_{UXY\tilde{Y}}\end{pmatrix}\leq 2^{-n\delta}\tag{3.180}$$

$$\mathbb{P}\begin{pmatrix}\left|N_{XY}(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}\tilde{Y}})-\mathbb{E}N_{XY}(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}\tilde{Y}})\right|\geq 2^{2n\delta}\\ \text{for some }V_{UXY\tilde{X}\tilde{Y}}\end{pmatrix}\leq 2^{-n\delta}\tag{3.181}$$

Now, by using the result of Lemma 3.2.1 and Markov's inequality, it can be concluded that

$$\begin{aligned}
&\mathbb{P}\left(\Lambda_X(X^{M_X}\times^{M_Y},V_{UXY\tilde{X}\hat{X}})\geq 2^{-n\left(E_S^X(V_{UXY\tilde{X}\hat{X}})-5\delta\right)}\text{ for some }V_{UXY\tilde{X}\hat{X}}\right)\\
&\quad\leq\sum_{V_{UXY\tilde{X}\hat{X}}}\mathbb{P}\left(\Lambda_X(X^{M_X}\times^{M_Y},V_{UXY\tilde{X}\hat{X}})\geq 2^{-n\left(E_S^X(V_{UXY\tilde{Y}\hat{Y}})-5\delta\right)}\right)\\
&\quad\leq\sum_{V_{UXY\tilde{X}\hat{X}}}\frac{\mathbb{E}\left(\Lambda_X(X^{M_X}\times^{M_Y},V_{UXY\tilde{X}\hat{X}})\right)}{2^{-n\left(E_S^X(V_{UXY\tilde{X}\hat{X}})-5\delta\right)}}\leq\sum_{V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}}2^{-n\delta}\leq 2^{-n\frac{\delta}{2}}.
\end{aligned}\tag{3.182}$$

Similarly,

$$\mathbb{P}\left(\Lambda_Y(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{Y}\hat{Y}})\geq 2^{-n\left(E_S^Y(V_{UXY\tilde{Y}\hat{Y}})-5\delta\right)}\text{ for some }V_{UXY\tilde{Y}\hat{Y}}\right)\leq 2^{-n\frac{\delta}{2}}, \tag{3.183}$$

and

$$\mathbb{P}\left(\begin{array}{c}\Lambda_{XY}(X^{M_X}\times Y^{M_Y},V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}})\geq 2^{-n\left(E_S^{XY}(V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}})-7\delta\right)}\\ \text{for some }V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}\end{array}\right)\leq 2^{-n\frac{\delta}{2}}. \tag{3.184}$$

Therefore, with probability $>1-7\times 2^{-n\frac{\delta}{2}}$, a code $C=C_X\times C_Y$ from random code ensemble satisfies the conditions given in the lemma. □

Proof. **(Lemma 3.2.3)** Let $C_X^r=\{\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_{M_X}\}$ and $C_Y^r=\{\mathbf{y}_1,\mathbf{y}_2,...,\mathbf{y}_{M_Y}\}$ be the collections of codewords whose existence is asserted in Lemma 3.2.1. Let us define

$$\begin{aligned}\Pi(C_X^r\times C_Y^r)\triangleq\sum_{V_{UXY\tilde{X}\tilde{Y}}}&\Big\{N_U(C_X^r\times C_Y^r,V_{UXY})2^{n[F_U(V_{UXY})-6\delta]}\\&+N_X(C_X^r\times C_Y^r,V_{UXY\tilde{X}})2^{n[F_X(V_{UXY\tilde{X}})-6\delta]}\\&+N_Y(C_X^r\times C_Y^r,V_{UXY\tilde{Y}})2^{n[F_Y(V_{UXY\tilde{Y}})-6\delta]}\\&+N_{XY}(C_X^r\times C_Y^r,V_{UXY\tilde{X}\tilde{Y}})2^{n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})-6\delta]}\Big\} &(3.185)\\ \leq\sum_{V_{UXY\tilde{X}\tilde{Y}}}&4\times 2^{-n\delta}<\frac{1}{2} &(3.186)\end{aligned}$$

For $C^r=C_X^r\times C_Y^r$, and the sequence $\mathbf{u}$ defined in random coding packing lemma, we define

$$\begin{aligned}L_U(C^r,V_{UXY},i,j)&\triangleq 1_{T_{V_{UXY}}}(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j), &(3.187)\\ L_X(C^r,V_{UXY\tilde{X}},i,j)&\triangleq\sum_{k\neq i}1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{x}_k), &(3.188)\\ L_Y(C^r,V_{UXY\tilde{Y}},i,j)&\triangleq\sum_{l\neq j}1_{T_{V_{UXY\tilde{Y}}}}(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{y}_l), &(3.189)\\ L_{XY}(C^r,V_{UXY\tilde{X}\tilde{Y}},i,j)&\triangleq\sum_{k\neq i}\sum_{l\neq j}1_{T_{V_{UXY\tilde{X}\tilde{Y}}}}(\mathbf{u},\mathbf{x}_i,\mathbf{y}_j,\mathbf{x}_k,\mathbf{y}_l). &(3.190)\end{aligned}$$

By definition of N_α, (3.185) can be written as

$$\Pi(C^r) = \frac{1}{M_X} \sum_{i=1}^{M_X} G(i), \qquad \text{for } \alpha = U, X, Y, XY, \tag{3.191}$$

where $G(i)$ is defined as follows:

$$\begin{aligned} G(i) \triangleq \frac{1}{M_Y} \sum_{j=1}^{M_Y} \sum_{V_{UXY\tilde{X}\tilde{Y}}} \Big\{ & L_U(C^r, V_{UXY}, i, j) 2^{n[F_U(V_{UXY})-6\delta]} \\ & + L_X(C^r, V_{UXY\tilde{X}}, i, j) 2^{n[F_X(V_{UXY\tilde{X}})-6\delta]} \\ & + L_X(C^r, V_{UXY\tilde{Y}}, i, j) 2^{n[F_Y(V_{UXY\tilde{Y}})-6\delta]} \\ & + L_{XY}(C^r, V_{UXY\tilde{X}\tilde{Y}}, i, j) 2^{n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})-6\delta]} \Big\}. \end{aligned} \tag{3.192}$$

By using (3.186), we see that the average of $G(i)$ over C_X^r is upper bounded by $\frac{1}{2}$, therefore, there must exist $\hat{M}_X \geq \frac{M_X}{2}$ codewords, $\mathbf{x}_i \in C_X^r$, for which

$$G(i) < 1. \tag{3.193}$$

Let us call this set of codewords as C_X^{ex}. Without loss of generality, we assume C_X^{ex} contains the first $\hat{M}_X$ sequences of C_X^r, i.e., $C_X^{ex} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{\hat{M}_X}\}$. Consider the multiuser code $C_1^{ex} \triangleq C_X^{ex} \times C_Y$. By definition of L_α, $\alpha = U, X, Y, XY$,

$$L_\alpha(C_1^{ex}, V, i, j) \leq L_\alpha(C^r, V, i, j) \quad \forall\, (\mathbf{x}_i.\mathbf{y}_j) \in C_1^{ex}. \tag{3.194}$$

By combining (3.193) and (3.194), we conclude that for all $i \in \{1, 2, ..., \hat{M}_X\}$

$$\begin{aligned} \frac{1}{M_Y} \sum_{j=1}^{M_Y} \sum_{V_{UXY\tilde{X}\tilde{Y}}} \Big\{ & L_U(C_1^{ex}, V_{UXY}, i, j) 2^{n[F_U(V_{UXY})-6\delta]} \\ & + L_X(C_1^{ex}, V_{UXY\tilde{X}}, i, j) 2^{n[F_X(V_{UXY\tilde{X}})-6\delta]} \\ & + L_X(C_1^{ex}, V_{UXY\tilde{Y}}, i, j) 2^{n[F_Y(V_{UXY\tilde{Y}})-6\delta]} \\ & + L_{XY}(C_1^{ex}, V_{UXY\tilde{X}\tilde{Y}}, i, j) 2^{n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}})-6\delta]} \Big\} < 1, \end{aligned} \tag{3.195}$$

which results in

$$\sum_{j=1}^{M_Y} \sum_{V_{UXY\tilde{X}\tilde{Y}}} \Big\{ L_U(C_1^{ex}, V_{UXY}, i, j) 2^{n[F_U(V_{UXY}) - R_Y - 6\delta]}$$
$$+ L_X(C_1^{ex}, V_{UXY\tilde{X}}, i, j) 2^{n[F_X(V_{UXY\tilde{X}}) - R_Y - 6\delta]}$$
$$+ L_X(C_1^{ex}, V_{UXY\tilde{Y}}, i, j) 2^{n[F_Y(V_{UXY\tilde{Y}}) - R_Y - 6\delta]}$$
$$+ L_{XY}(C_1^{ex}, V_{UXY\tilde{X}\tilde{Y}}, i, j) 2^{n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}}) - R_Y - 6\delta]} \Big\} < 1. \tag{3.196}$$

Since all terms in the summation are non-negative, we conclude that

$$L_\alpha(C_1^{ex}, V, i, j) 2^{-n[F_\alpha(V) - R_Y - 6\delta]} < 1 \tag{3.197}$$

for all $i \in \{1, 2, ..., \hat{M}_X\}$, $j \in \{1, 2, ..., M_Y\}$, all $V \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{Y})$, and all $\alpha = U, X, Y, XY$. Therefore,

$$L_\alpha(C_1^{ex}, V, i, j) < 2^{-n[F_\alpha(V) - R_Y - 6\delta]}. \tag{3.198}$$

On the other hand, (3.186) can also be written as

$$\Pi(C^r) = \frac{1}{M_Y} \sum_{j=1}^{M_Y} H(j), \qquad \text{for } \alpha = U, X, Y, XY, \tag{3.199}$$

where $H(j)$ is defined as

$$H(j) \triangleq \frac{1}{M_X} \sum_{i=1}^{M_X} \sum_{V_{UXY\tilde{X}\tilde{Y}}} \Big\{ L_U(C^r, V_{UXY}, i, j) 2^{n[F_U(V_{UXY}) - 6\delta]}$$
$$+ L_X(C^r, V_{UXY\tilde{X}}, i, j) 2^{n[F_X(V_{UXY\tilde{X}}) - 6\delta]}$$
$$+ L_X(C^r, V_{UXY\tilde{Y}}, i, j) 2^{n[F_Y(V_{UXY\tilde{Y}}) - 6\delta]}$$
$$+ L_{XY}(C^r, V_{UXY\tilde{X}\tilde{Y}}, i, j) 2^{n[F_{XY}(V_{UXY\tilde{X}\tilde{Y}}) - 6\delta]} \Big\}. \tag{3.200}$$

By a similar argument as we did before, we can show that there exist $\hat{M}_Y \geq \frac{M_Y}{2}$

codewords, $\mathbf{y}_j \in C_Y^r$, for which

$$H(j) < 1. \tag{3.201}$$

Let us call this set of codewords as C_Y^{ex}. Without loss of generality, we assume C_Y^{ex} contains the first $\hat{M}_Y$ sequences of C_Y^r, i.e., $C_Y^{ex} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{\hat{M}_Y}\}$. Consider the multiuser code $C_2^{ex} \triangleq C_X \times C_Y^{ex}$. By definition of L_α, $\alpha = U, X, Y, XY$, we have

$$L_\alpha(C_2^{ex}, V, i, j) \leq L_\alpha(C^r, V, i, j) \quad \forall\, (\mathbf{x}_i . \mathbf{y}_j) \in C_2^{ex}. \tag{3.202}$$

By a similar argument as we did before, we can show that

$$L_\alpha(C_2^{ex}, V, i, j) < 2^{-n[F_\alpha(V) - R_X - 6\delta]}. \tag{3.203}$$

for all $i \in \{1, 2, ..., M_X\}$, $j \in \{1, 2, ..., \hat{M}_Y\}$, all $V \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{Y})$, and all $\alpha = U, X, Y, XY$.

By combining (3.198) and (3.203), we conclude that, there exists a multiuser code $C^{ex} = C_X^* \times C_Y^*$ with $M_X^* \times M_Y^*$ messages

$$M_X^* \geq \frac{2^{n(R_X - \delta)}}{2}, \qquad M_Y^* \geq \frac{2^{n(R_Y - \delta)}}{2}, \qquad M_X^* \times M_Y^* \geq \frac{2^{n(R_X + R_Y - 2\delta)}}{2} \tag{3.204}$$

such that for any pair of messages $(\mathbf{x}_i, \mathbf{y}_j) \in C^{ex}$, all $V \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{Y})$, and all $\alpha = U, X, Y, XY$,

$$L_\alpha(C^{ex}, V, i, j) < 2^{-n[F_\alpha(V) - \min\{R_X, R_Y\} - 6\delta]}. \tag{3.205}$$

It is easy to check that

$$\Pi(C^{ex}) \leq 2 \times \Pi(C^r) < 1, \tag{3.206}$$

therefore, C^{ex}, satisfies all the constraints in (3.74a)-(3.74d).

Here, by method of expurgation, we end up with a code with a similar average bound as we had for the original code. However, all pairs of codewords in the new code also satisfy (3.75a)-(3.75d). Therefore, we did not lose anything in terms of average performance, however, as we will see in Theorem 3.2.1, we would end up with a tighter bound since we have more constraints on any particular pair of codewords in our codebook pair. $\square$

Proof. **(Theorem 3.2.1)** Let us do random coding. Fix any $\mathcal{U}$, $P_{XYU} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ such that $X - U - Y$, $R_X \geq 0$, $R_Y \geq 0$, $\delta > 0$, and $\mathbf{u} \in T_{P_U}$. Define M_X, M_Y such that

$$2^{n(R_X-\delta)} \leq M_X \leq 2^{nR_X} \qquad 2^{n(R_Y-\delta)} \leq M_Y \leq 2^{nR_Y}$$

Let $X^{M_X} \triangleq \left(X_1, X_2, ..., X_{M_X}\right)$ and $Y^{M_Y} \triangleq \left(Y_1, Y_2, ..., Y_{M_Y}\right)$ be independent random variables, where X_is are uniformly distributed on $T_{P_{X|U}}(\mathbf{u})$, and Y_js are uniformly distributed on $T_{P_{Y|U}}(\mathbf{u})$.

Upper bound: By taking expectation over (3.57), applying Lemma 3.2.1, and using the continuity of information measures, we get the desired upper bound.

Lower bound: By taking expectation over (3.59), applying Lemma 3.2.1, we get

$$\begin{aligned}
\mathbb{E}e(C,W) \geq & \sum_{\substack{V_{UXY\tilde{X}Z} \\ \in \mathcal{V}^r_{X,n}}} 2^{-n(E^L_X+4\delta)} \left[1 - \sum_{\substack{V_{UXY\tilde{X}\hat{X}Z}: \\ V_{UXY\hat{X}Z}=V_{UXY\tilde{X}Z}}} 2^{-n(I_V(\hat{X}\wedge XY\tilde{X}Z|U)-R_x-7\delta)}\right] \\
& + \sum_{\substack{V_{UXY\tilde{Y}Z} \\ \in \mathcal{V}^r_{Y,n}}} 2^{-n(E^L_Y+4\delta)} \left[1 - \sum_{\substack{V_{UXY\tilde{Y}\hat{Y}Z}: \\ V_{UXY\hat{Y}Z}=V_{UXY\tilde{Y}Z}}} 2^{-n(I_V(\hat{Y}\wedge XY\tilde{Y}Z|U)-R_Y-7\delta)}\right] \\
+ & \sum_{\substack{V_{UXY\tilde{X}\tilde{Y}Z} \\ \in \mathcal{V}^r_{XY,n}}} 2^{-n(E^L_{XY}+4\delta)} \left[1 - \sum_{\substack{V_{UXY\tilde{X}\hat{X}\tilde{Y}\hat{Y}Z}: \\ V_{UXY\hat{X}\hat{Y}Z}=V_{UXY\tilde{X}\tilde{Y}Z}}} 2^{-n(I_V(\hat{X}\hat{Y}\wedge XY\tilde{X}\tilde{Y}Z|U)-R_X-R_Y-7\delta)}\right]
\end{aligned} \tag{3.207}$$

Toward further simplification of this expression, we use the following lemma.

Lemma 3.4.3.

$$\min_{\substack{V_{UXY\tilde{X}\hat{X}Z}: \\ V_{UXY\hat{X}Z}=V_{UXY\tilde{X}Z}}} I_V(\hat{X} \wedge XY\tilde{X}Z|U) = I_V(\tilde{X} \wedge XYZ|U) \tag{3.208}$$

Proof. Note that, for any $V_{UXY\tilde{X}\hat{X}Z}$,

$$I_V(\hat{X} \wedge XY\tilde{X}Z|U) = I_V(\hat{X} \wedge XYZ|U) + I_V(\tilde{X} \wedge \hat{X}|UXYZ), \tag{3.209}$$

therefore,

$$\min_{\substack{V_{UXY\tilde{X}\hat{X}Z}: \\ V_{UXY\hat{X}Z}=V_{UXY\tilde{X}Z}}} I_V(\hat{X} \wedge XY\tilde{X}Z|U) \geq I_V(\hat{X} \wedge XYZ|U) = I_V(\tilde{X} \wedge XYZ|U). \tag{3.210}$$

Now, consider $V^*_{UX\tilde{X}\hat{X}YZ}$ defined as

$$V^*_{UXY\tilde{X}\hat{X}Z}(u,x,y,\tilde{x},\hat{x},z) = V_{\tilde{X}|UXYZ}(\tilde{x}|u,x,y,z)V_{\tilde{X}|UXYZ}(\hat{x}|u,x,y,z)V_{UXYZ}(u,x,y,z) \tag{3.211}$$

Note that $V^*_{UX\hat{X}YZ} = V^*_{UX\tilde{X}YZ}$, and $\tilde{X} - (U,X,Y,Z) - \hat{X}$. Therefore,

$$I_{V^*}(\hat{X} \wedge XY\tilde{X}Z|U) = I_V(\hat{X} \wedge XYZ|U) = I_V(\tilde{X} \wedge XYZ|U). \tag{3.212}$$

By combining (3.210) and (3.212), the proof is complete. □

Using the above lemma, the average probability of error can be bounded from below as

$$\bar{P}_e \geq \sum_{\substack{V_{UXY\tilde{X}Z} \in \mathcal{V}^r_{X,n} \\ I(\tilde{X} \wedge XYZ|U) > R_X + 12\delta}} 2^{-nE^L_X} + \sum_{\substack{V_{UXY\tilde{Y}Z} \in \mathcal{V}^r_{Y,n} \\ I(\tilde{Y} \wedge XYZ|U) > R_Y + 12\delta}} 2^{-nE^L_Y}$$
$$+ \sum_{\substack{V_{UXY\tilde{X}\tilde{Y}Z} \in \mathcal{V}^r_{XY,n} \\ I_V(\tilde{X}\tilde{Y} \wedge XY|U) + I_V(\tilde{X} \wedge \tilde{Y}|U) > \\ R_X + R_Y + 14\delta}} 2^{-nE^L_{XY}} \tag{3.213}$$

Using the continuity argument, the lower bound on the average error probability follows. □

Proof. **(Theorem 3.2.2)** As was done in Theorem 3.1.1 for the point-to-point case, here, we will obtain higher error exponents for almost all codes by removing certain types from the constraint sets $\mathcal{V}^r_X$, $\mathcal{V}^r_Y$ and $\mathcal{V}^r_{XY}$. Let us define the sets of n-types $\mathcal{V}^t_X$, $\mathcal{V}^t_X$ and $\mathcal{V}^t_{XY}$ as follows:

$$\mathcal{V}^t_{X,n} \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{X}}: & V_{XU} = V_{\tilde{X}U} = P_{XU}, V_{YU} = P_{YU} \\ & F_U(V_{UXY}), F_U(V_{U\tilde{X}Y}) \leq R_X + R_Y \\ & F_X(V_{UXY\tilde{X}}) \leq R_X + R_Y \end{array} \right\} \tag{3.214}$$

$$\mathcal{V}^t_{Y,n} \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{Y}}: & V_{XU} = P_{XU}, V_{YU} = V_{\tilde{Y}U} = P_{YU} \\ & F_U(V_{UXY}), F_U(V_{UX\tilde{Y}}) \leq R_X + R_Y \\ & F_Y(V_{UXY\tilde{Y}}) \leq R_X + R_Y \end{array} \right\} \tag{3.215}$$

$$\mathcal{V}^t_{XY,n} \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{X}\tilde{Y}}: & V_{UXY\tilde{X}}, V_{UX\tilde{Y}\tilde{X}} \in \mathcal{V}^t_X, \quad V_{UXY\tilde{Y}}, V_{U\tilde{X}Y\tilde{Y}} \in \mathcal{V}^t_Y \\ & F_{XY}(V_{UXY\tilde{X}\tilde{Y}}), F_{XY}(V_{U\tilde{X}\tilde{Y}X\tilde{Y}}) \leq R_X + R_Y \end{array} \right\} \tag{3.216}$$

Lemma 3.4.4. *Let $C = C_X \times C_Y$ be one of the multiuser codes whose existence is asserted in the Typical random coding packing lemma. The following hold:*

$$\textit{If } V_{UXY\tilde{X}} \in (\mathcal{V}^t_{X,n})^c \Rightarrow N_X(C, V_{UXY\tilde{X}}) = 0, \tag{3.217}$$

$$\textit{If } V_{UXY\tilde{Y}} \in (\mathcal{V}^t_{Y,n})^c \Rightarrow N_Y(C, V_{UXY\tilde{Y}}) = 0, \tag{3.218}$$

$$\textit{If } V_{UXY\tilde{X}\tilde{Y}} \in (\mathcal{V}^t_{XY,n})^c \Rightarrow N_{XY}(C, V_{UXY\tilde{X}\tilde{Y}}) = 0. \tag{3.219}$$

Proof. Consider $V_{UXY\tilde{X}} \in (\mathcal{V}^t_{X,n})^c$. If $V_{XU} \neq P_{XU}$ or $V_{\tilde{X}U} \neq P_{XU}$ or $V_{YU} \neq P_{YU}$, it is clear that

$$N_X(C, V_{UXY\tilde{X}}) = 0. \tag{3.220}$$

Now, let us assume $F_U(V_{UXY}) > R_X + R_Y + 3\delta$. In this case, by using (3.72a), we conclude that

$$\begin{aligned} N_U(C, V_{UXY}) < 2^{-n(R_X+R_Y)} &\Rightarrow \sum_{i=1}^{M_X}\sum_{j=1}^{M_Y} 1_{T_{V_{UXY}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j) < 1 \\ &\Rightarrow \sum_{i=1}^{M_X}\sum_{j=1}^{M_Y} 1_{T_{V_{UXY}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j) = 0, \end{aligned} \tag{3.221}$$

and as a result, $N_U(C, V_{UXY}) = 0$. Now, note that

$$\begin{aligned} N_X(C, V_{UXY\tilde{X}}) &= \frac{1}{M_X M_Y}\sum_{i=1}^{M_X}\sum_{j=1}^{M_Y}\sum_{k\neq i} 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k) \\ &\leq \frac{1}{M_X M_Y}\sum_{i=1}^{M_X}\sum_{j=1}^{M_Y}\sum_{k\neq i} 1_{T_{V_{UXY}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j) \\ &= 2^{nR_X} N_U(C, V_{UXY}) = 0, \end{aligned} \tag{3.222}$$

therefore, $N_X(C, V_{UXY\tilde{X}}) = 0$. Similarly, if $F_U(V_{U\tilde{X}Y}) > R_X + R_Y + 3\delta$,

$$\begin{aligned} N_U(C, V_{U\tilde{X}Y}) < 2^{-n(R_X+R_Y)} &\Rightarrow \sum_{i=1}^{M_X}\sum_{j=1}^{M_Y} 1_{T_{V_{U\tilde{X}Y}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j) < 1 \\ &\Rightarrow \sum_{i=1}^{M_X}\sum_{j=1}^{M_Y} 1_{T_{V_{U\tilde{X}Y}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j) = 0, \end{aligned} \tag{3.223}$$

and as a result, $N_U(C_X, C_Y, V_{U\tilde{X}Y}) = 0$. Also, note that

$$N_X(C, V_{UXY\tilde{X}}) = \frac{1}{M_X M_Y} \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} \sum_{k \neq i} 1_{T_{V_{UXY\tilde{X}}}}(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_k)$$
$$\leq \frac{1}{M_X M_Y} \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} \sum_{k \neq i} 1_{T_{V_{U\tilde{X}Y}}}(\mathbf{u}, \mathbf{x}_k, \mathbf{y}_j) = 0, \tag{3.224}$$

therefore, $N_X(C, V_{UXY\tilde{X}}) = 0$. If $F_X(V_{UXY\tilde{X}}) > R_X + R_Y + 5\delta$, by the property of the code derived in Lemma 3.2.2, we observe that $N_X(C_X, C_Y, V_{UXY\tilde{X}}) = 0$. Similarly, by doing a similar argument, it can be concluded that

$$\text{If } V_{UXY\tilde{Y}} \in (\mathcal{V}^t_{Y,n})^c \Rightarrow N_Y(C, V_{UXY\tilde{Y}}) = 0, \tag{3.225}$$

and

$$\text{If } V_{UXY\tilde{X}\tilde{Y}} \in (\mathcal{V}^t_{XY,n})^c \Rightarrow N_{XY}(C, V_{UXY\tilde{X}\tilde{Y}}) = 0. \tag{3.226}$$

□

Upper bound: We will follow the techniques used in Theorem 3.2.1 to provide lower and upper bounds on the average probability of error of almost all codes in the random coding ensemble. For this, we will use the results of Lemma 3.2.3. Consider any typical two-user code $C = C_X \times C_Y$ whose existence was established in Lemma 3.2.2. Applying (3.57) on C, and using the continuity argument, we conclude that

$$\begin{aligned} e(C, W) \leq & \sum_{V_{UXY\tilde{X}Z} \in \mathcal{V}^r_{X,n} \cap \mathcal{V}^t_{X,n}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU}) + I_V(X \wedge Y|U) + |I_V(\tilde{X} \wedge XYZ|U) - R_X|^+ - 5\delta]} \\ & + \sum_{V_{UXY\tilde{Y}Z} \in \mathcal{V}^r_{Y,n} \cap \mathcal{V}^t_{Y,n}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU}) + I_V(X \wedge Y|U) + |I_V(\tilde{Y} \wedge XYZ|U) - R_Y|^+ - 5\delta]} \\ & + \sum_{\substack{V_{UXY\tilde{X}\tilde{Y}Z} \\ \in \mathcal{V}^r_{XY,n} \cap \mathcal{V}^t_{XY,n}}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU}) + I_V(X \wedge Y|U) + |I_V(\tilde{X} \wedge \tilde{Y}|U) + I_V(\tilde{X}\tilde{Y} \wedge XYZ|U) - R_X - R_Y|^+ - 5\delta]} \\ & \leq 2^{-n[E_T(R_X, R_Y, W, P_{UXY}) - 6\delta]} \end{aligned} \tag{3.227}$$

whenever $n \geq n_1(|\mathcal{Z}|, |\mathcal{X}|, |\mathcal{Y}|, |\mathcal{U}|, \delta)$, where $E_T(R_X, R_Y, W, P_{XYU})$ is defined in the statement of the theorem.

Lower bound: In the following, we obtain a lower bound on the average error probability of code $C = C_X \times C_Y$. Applying (3.59) on C, then using (a) Lemma 3.2.2 and (b) the fact that for $V \notin V^t_{X,n}$, we have $A^X_{i,j} \geq 0$, and similar such facts about A^Y and A^{XY}, we get

$$\begin{aligned} e(C,W) \geq & \sum_{\substack{V_{UXY\tilde{X}Z} \\ \in \mathcal{V}^r_{X,n} \cap V^t_{X,n}}} 2^{-n(E^L_X + 4\delta)} \left| 1 - \sum_{\substack{V_{UXY\tilde{X}\hat{X}Z}: \\ V_{UXY\hat{X}Z} = V_{UXY\tilde{X}Z}}} 2^{-n(I_V(\hat{X} \wedge XY\tilde{X}Z|U) - R_x - 7\delta)} \right|^+ \\ & + \sum_{\substack{V_{UXY\tilde{Y}Z} \\ \in \mathcal{V}^r_{Y,n} \cap V^t_{Y,n}}} 2^{-n(E^L_Y + 4\delta)} \left| 1 - \sum_{\substack{V_{UXY\tilde{Y}\hat{Y}Z}: \\ V_{UXY\hat{Y}Z} = V_{UXY\tilde{Y}Z}}} 2^{-n(I_V(\hat{Y} \wedge XY\tilde{Y}Z|U) - R_Y - 7\delta)} \right|^+ \\ & + \sum_{\substack{V_{UXY\tilde{X}\tilde{Y}Z} \\ \in \mathcal{V}^r_{XY,n} \cap V^t_{XY,n}}} 2^{-n(E^L_{XY} + 4\delta)} \left| 1 - \sum_{\substack{V_{UXY\tilde{X}\hat{X}\tilde{Y}\hat{Y}Z}: \\ V_{UXY\hat{X}\hat{Y}Z} = V_{UXY\tilde{X}\tilde{Y}Z}}} 2^{-n(I_V(\hat{X}\hat{Y} \wedge XY\tilde{X}\tilde{Y}Z|U) - R_X - R_Y - 7\delta)} \right|^+ \end{aligned} \tag{3.228}$$

This expression can be simplified as follows.

$$\begin{aligned} e(C,W) \geq & \sum_{\substack{V_{UXY\tilde{X}Z} \in \mathcal{V}^r_{X,n} \cap V^t_{X,n} \\ I(\tilde{X} \wedge XYZ|U) > R_X + 12\delta}} 2^{-nE^L_X} + \sum_{\substack{V_{UXY\tilde{Y}Z} \in \mathcal{V}^r_{Y,n} \cap V^t_{Y,n} \\ I(\tilde{Y} \wedge XYZ|U) > R_Y + 12\delta}} 2^{-nE^L_Y} \\ & + \sum_{\substack{V_{UXY\tilde{X}\tilde{Y}Z} \in \mathcal{V}^r_{XY,n} \cap V^t_{XY,n} \\ I_V(\tilde{X}\tilde{Y} \wedge XY|U) + I_V(\tilde{X} \wedge \tilde{Y}|U) > R_X + R_Y + 14\delta}} 2^{-nE^L_{XY}} \end{aligned} \tag{3.229}$$

Using the continuity argument, the lower bound on the average error probability follows. □

Proof. **(Theorem 3.2.3)** Fix $\mathcal{U}$, $\mathcal{P}_{XYU} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ with $X - U - Y$, $R_X \geq 0$, $R_Y \geq 0$, $\delta > 0$, and $\mathbf{u} \in T_{P_U}$. Let $C^* = C^*_X \times C^*_Y$ be the multiuser code whose existence is asserted in Lemma 3.2.3. Taking into account the given $\mathbf{u}$, the α-decoding yields the decoding sets

$$D_{ij} = \{\mathbf{z} : \alpha(\mathbf{u}, \mathbf{x}_i, \mathbf{y}_j, \mathbf{z}) \leq \alpha(\mathbf{u}, \mathbf{x}_k, \mathbf{y}_l, \mathbf{z}) \text{ for all } (k, l) \neq (i, j)\}.$$

Let us define the collection of n-types $\mathcal{V}^x_{X,n}$, $\mathcal{V}^x_{Y,n}$ and $\mathcal{V}^x_{XY,n}$ as follows:

$$\mathcal{V}^x_{X,n} \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{X}} : & V_{XU} = V_{\tilde{X}U} = P_{XU}, V_{YU} = P_{YU} \\ & F_U(V_{UXY}), F_U(V_{U\tilde{X}Y}) \leq \min\{R_X, R_Y\} \\ & F_X(V_{UXY\tilde{X}}) \leq \min\{R_X, R_Y\} \end{array} \right\} \tag{3.230}$$

$$\mathcal{V}^x_{Y,n} \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{Y}} : & V_{XU} = P_{XU}, V_{YU} = V_{\tilde{Y}U} = P_{YU} \\ & F_U(V_{UXY}), F_U(V_{UX\tilde{Y}}) \leq \min\{R_X, R_Y\} \\ & F_Y(V_{UXY\tilde{Y}}) \leq \min\{R_X, R_Y\} \end{array} \right\} \tag{3.231}$$

$$\mathcal{V}^x_{XY,n} \triangleq \left\{ \begin{array}{ll} V_{UXY\tilde{X}\tilde{Y}} : & V_{UXY\tilde{X}}, V_{UX\tilde{Y}\tilde{X}} \in \mathcal{V}^x_X, \quad V_{UXY\tilde{Y}}, V_{U\tilde{X}Y\tilde{Y}} \in \mathcal{V}^x_Y \\ & F_{XY}(V_{UXY\tilde{X}\tilde{Y}}), F_{XY}(V_{U\tilde{X}\tilde{Y}X\tilde{Y}}) \leq \min\{R_X, R_Y\} \end{array} \right\} \tag{3.232}$$

Lemma 3.4.5. *For the multiuser code $C^* = C^*_X \times C^*_Y$, the following holds:*

$$\textit{If } V_{UXY\tilde{X}} \in (\mathcal{V}^x_{X,n})^c \Rightarrow N_X(C^*, V_{UXY\tilde{X}}) = 0, \tag{3.233}$$

$$\textit{If } V_{UXY\tilde{Y}} \in (\mathcal{V}^x_{Y,n})^c \Rightarrow N_Y(C^*, V_{UXY\tilde{Y}}) = 0, \tag{3.234}$$

$$\textit{If } V_{UXY\tilde{X}\tilde{Y}} \in (\mathcal{V}^x_{XY,n})^c \Rightarrow N_{XY}(C^*, V_{UXY\tilde{X}\tilde{Y}}) = 0. \tag{3.235}$$

Proof. The proof is very similar to the proof of lemma 3.4.4. □

The average error probability of C^* can be obtained as follows in a similar way that used in the proof of Theorem 3.2.1 and Theorem 3.2.2.

$$
\begin{aligned}
e(C^*, W) \leq & \sum_{V_{UXY\tilde{X}Z} \in \mathcal{V}^r_{X,n} \cap \mathcal{V}^x_{X,n}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU})+I_V(X\wedge Y|U)-3\delta]} \\
& + \sum_{V_{UXY\tilde{Y}Z} \in \mathcal{V}^r_{Y,n} \cap \mathcal{V}^x_{Y,n}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU})+I_V(X\wedge Y|U)-3\delta]} \\
& + \sum_{\substack{V_{UXY\tilde{X}\tilde{Y}Z} \\ \in \mathcal{V}^r_{XY,n} \cap \mathcal{V}^x_{XY,n}}} 2^{-n[D(V_{Z|XYU}||W|V_{XYU})+I_V(X\wedge Y|U)-3\delta]}. \qquad (3.236)
\end{aligned}
$$

Now using the continuity argument the statement of the theorem follows. □

Proof. **(Theorem 3.2.4)** For any $V_{UXY\tilde{X}Z} \in \mathcal{V}^r_X$,

$$H_V(XY|ZU) \geq H_V(\tilde{X}Y|ZU), \qquad (3.237)$$

therefore, by subtracting $H_V(Y|ZU)$ form both sides of (3.237), we can conclude that

$$H_V(X|U) - I_V(X \wedge YZ|U) \geq H_V(\tilde{X}|U) - I_V(\tilde{X} \wedge YZ|U), \qquad (3.238)$$

Since $V_{XU} = V_{\tilde{X}U} = P_{XU}$, the last inequality is equivalent to

$$I_V(X \wedge YZ|U) \leq I_V(\tilde{X} \wedge YZ|U). \qquad (3.239)$$

Since $I_V(\tilde{X} \wedge XYZ|U) \geq I_V(\tilde{X} \wedge YZ|U)$, it can be seen that for any $V_{UXY\tilde{X}Z} \in \mathcal{V}^r_X$

$$I_V(\tilde{X} \wedge XYZ|U) \geq I_V(X \wedge YZ|U). \qquad (3.240)$$

Moreover, since

$$\mathcal{V}^r_X \subseteq \{V_{UXY\tilde{X}Z} : V_{UXYZ} \in \mathcal{V}(P_{UXY})\} \qquad (3.241)$$

it can be easily concluded that

$$E_X^r(R_X, R_Y, W, P_{XYU}) \geq E_{rX}^{Liu}(R_X, R_Y, W, P_{XYU}).$$

Similarly, for any $V_{UXY\tilde{Y}Z} \in \mathcal{V}_Y^r$,

$$H_V(XY|ZU) \geq H_V(X\tilde{Y}|ZU). \tag{3.242}$$

By using the fact that, $V_{YU} = V_{\tilde{Y}U} = P_{YU}$, it can be concluded that

$$I_V(\tilde{Y} \wedge XYZ|U) \geq I_V(Y \wedge XZ|U). \tag{3.243}$$

Since

$$\mathcal{V}_Y^r \subseteq \left\{V_{UXY\tilde{Y}Z} : V_{UXYZ} \in \mathcal{V}(P_{UXY})\right\}, \tag{3.244}$$

we conclude that

$$E_Y^r(R_X, R_Y, W, P_{XYU}) \geq E_{rY}^{Liu}(R_X, R_Y, W, P_{XYU}). \tag{3.245}$$

Similarly, we can conclude that, for any $V_{UXY\tilde{X}\tilde{Y}Z} \in \mathcal{V}_{XY}^r$,

$$I_V(\tilde{X}\tilde{Y} \wedge XYZ|U) + I(\tilde{X} \wedge \tilde{Y}|U) \geq I_V(XY \wedge Z|U) + I(X \wedge Y|U). \tag{3.246}$$

Since

$$\mathcal{V}_{XY}^r \subseteq \left\{V_{UXY\tilde{X}\tilde{Y}Z} : V_{UXYZ} \in \mathcal{V}(P_{UXY})\right\}, \tag{3.247}$$

it can be concluded that

$$E_{XY}^r(R_X, R_Y, W, P_{XYU}) \geq E_{rXY}^{Liu}(R_X, R_Y, W, P_{XYU}). \tag{3.248}$$

By combining (3.4.2), (3.245) and (3.248), we conclude that (3.90a) holds. Similarly, we can prove that (3.90b) and (3.90c) hold. $\square$

CHAPTER 4

Typicality Graphs and Their Properties

The concept of typicality and typical sequences is central to information theory. It has been used to develop computable performance limits for several communication problems. In studying the performance of channel block codes for discrete memoryless channels, it is observed that the composition of the codewords play a crucial role. In particular, to obtain upper and lower bounds on the reliability of the channel, the method of types not only simplifies the derivation of the bounds but also provides us more with intuition about the system. In the study of an arbitrary channel, it has been shown that it is sufficient to study constant composition codes, the codes for which all codewords have a similar composition. The idea behind the method of types is to partition the codewords of an arbitrary code into classes according to their composition. The error event is then partitioned into its intersections with these type classes, and the error probability can be obtained by adding up the probabilities of these intersections. In [16], it is shown that the number of type classes grows polynomially as a function of the blocklength, implying that the error probability has the same exponential asymptotics as the largest one among the probabilities of these intersections. In other words, one of the types plays a crucial role in determining the performance of the code. Note that to obtain an upper bound on the reliability function of the channel, we need to study the performance of the best code. It is observed that for the best code, the composition which dominates the error event must be a dominant type of the code. Otherwise, one can eliminate all codewords with this particular composition, and the resulting code, which has the same transmission rate, outperforms the best code, which causes a contradiction. Therefore, to obtain an upper bound on the reliability function of a DMC, we need to study the compositions that can be a dominant type of the best code. In particular, we need to answer the following question: "At a fixed transmission rate, R, which composition is the dominant type of the best code?," or as a more general question, one might ask "At a fixed transmission rate, R, which compositions can potentially be the dominant type of an arbitrary code?". For single user codes, the answer to this question is straight forward, and it is clear that as long as the number of sequences of type P is larger than the number of codewords in the code, P can be a dominant type of the code. Therefore, P could be a dominant type of a code of rate R, if and only if $H(P) \geq R$.

Now, consider any (n, M_X, M_Y) code C. Suppose all the messages of any source are equiprobable and the sources are sending data independently. Assuming these conditions, all $M_X M_Y$ pairs are occuring with the same probability. Thus, at the input of the channel, all possible $M_X M_Y$ (an exponentially increasing function of n) pairs of input sequences can be observed. However, we also know that the number of possible joint types on $\mathcal{X} \times \mathcal{Y}$ is a polynomial function of n of degree at most equal to $|\mathcal{X}||\mathcal{Y}|$ [16]. Thus, for at least one joint type, the number of pairs of sequences in the multi user code, sharing that particular type, should be an exponential function of n with the rate almost equal to the rate of the multi user code, C. We call the subcode consisting of these pairs of sequences as a dominant subcode of C. As a result, we obtain:

Fact 4.0.1. *Fix any $\delta > 0$, $n \geq n_1(|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}|, \delta)$. For any multi user code, C, with parameters $(n, 2^{nR_X}, 2^{nR_Y})$, there exists a joint composition $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ such that*

$$R(C, P_{XY}) \geq R_X + R_Y - 3\delta, \tag{4.1}$$

where $R(C, P_{XY})$ is defined in (5.1). Any joint composition satisfying (4.1) is called a dominant joint type of C.

Hence, for any multi-user code, there must exist at least one joint type which dominates the codebook. The dominant type of the best code plays a crucial role in determining its performance. Therefore, to obtain an upper bound on the reliability function of a DM-MAC, we need to characterize the possible dominant joint compositions of multi-user codes at certain transmission rate pair. In particular, we can ask the following question: "For a multiuser code, with rate pair (R_X, R_Y), which joint types can be its dominant type?" As shown in chapter 5, the answer to this question helps us to characterize tighter upper bounds on the error exponent of multiple access channels.

Consider a pair of correlated discrete memoryless information sources X and Y characterized by a generic joint distribution p_{XY} defined on the product of two finite sets $\mathcal{X} \times \mathcal{Y}$. A length n X-sequence x^n is typical if the empirical histogram of x^n is close to p_X. A pair of length n sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ is said to be jointly typical if the empirical joint histogram of (x^n, y^n) is close to the joint distribution p_{XY}. The set of all jointly typical sequence pairs is called the typical set of p_{XY}.

Given a sequence length n, the typical set can be represented in terms of the following undirected, bipartite graph. The left vertices of the graph are all the typical X-sequences, and the right vertices are all the typical Y-sequences. In accordance with the properties of typical sets, there are (approximately) $2^{nH(X)}$ left vertices and $2^{nH(Y)}$ right vertices. A left vertex is connected to a right vertex through an edge if the corresponding X and Y-sequences are *jointly* typical. From the properties of joint typicality, we know that the number of edges in this graph is roughly $2^{nH(X,Y)}$. Additionally, every left vertex (a typical X-sequence) has degree roughly equal to $2^{nH(Y|X)}$, i.e., it is jointly typical with $2^{nH(Y|X)}$ Y-sequences. Similarly, each right vertex has degree roughly equal to $2^{nH(X|Y)}$.

In this chapter, we formally characterize the typicality graph and look at some subgraph containment problems. In particular, we answer three questions concerning the typicality graph:

- When can we find subgraphs such that the left and right vertices of the subgraph have specified degrees, say R'_X and R'_Y, respectively ?

- What is the maximum size of subgraphs that are complete, i.e., every left vertex is connected to every right vertex? One of the main contributions of this chapter is providing a sharp answer to this question.

- If we create a subgraph by randomly picking a specified number of left and right vertices, what is the probability that this subgraph has far fewer edges than expected?

These questions arise in a variety of multiuser communication problems. Transmitting correlated information over a multiple-access channel (MAC) [42] and communicating over a MAC with feedback [48], are two problems in which the first question plays an important role. The techniques used to answer the second question will be applied in the following chapter to develop tighter upper bounds on the error exponents of discrete memoryless multiple-access channels. The third question arises within the context of transmitting correlated information over a broadcast channel [11]. Moreover, the evaluation of performance limits of a multiuser communication problem can be thought of as characterizing certain properties of typicality graphs of random variables associated with the problem.

4.1 Preliminaries

In this section, we provide a concise review of some of the results available in the literature on typical sequences, δ-typical sets and their properties [16].

Definition 4.1.1. *A sequence $x^n \in \mathcal{X}^n$ is X-typical with constant δ if*

1. $\left|\frac{1}{n}N(a|x^n) - P_X(a)\right| \leq \delta, \quad \forall a \in \mathcal{X}$

2. *No $a \in \mathcal{X}$ with $P_X(a) = 0$ occurs in x^n.*

The set of such sequences is denoted by $T^n_\delta(P_X)$ or $T^n_\delta(X)$, when the distribution being used is unambiguous.

Definition 4.1.2. *Given a conditional distribution $P_{Y|X}$, a sequence $y^n \in \mathcal{Y}^n$ is conditionally $P_{Y|X}$-typical with $x^n \in \mathcal{X}^n$ with constant δ if*

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

2. $N(a,b|x^n,y^n) = 0$ *whenever* $P_{Y|X}(b|a) = 0$.

The set of such sequences is denoted $T^n_\delta(P_{Y|X}|x^n)$ or $T^n_\delta(Y|x^n)$, when the distribution being used is unambiguous.

We will repeatedly use the following results, which we state below as facts:

Fact 4.1.1. *[16, Lemma 2.10]*
(a) If $x^n \in T^n_\delta(X)$ and $y^n \in T^n_{\delta'}(Y|x^n)$, then $(x^n,y^n) \in T^n_{\delta+\delta'}(X,Y)$ and $y^n \in T^n_{(\delta+\delta')|\mathcal{X}|}(Y)$. [1]
(b) If $x^n \in T^n_\delta(X)$ and $(x^n,y^n) \in T^n_\epsilon(X,Y)$, then $y^n \in T^n_{\delta+\epsilon}(Y|x^n)$.

Fact 4.1.2. *[16, Lemma 2.13]* [2]*: There exists a sequence $\epsilon_n \to 0$ depending only on $|\mathcal{X}|$ and $|\mathcal{Y}|$ such that for every joint distribution $P_X \cdot P_{Y|X}$ on $\mathcal{X} \times \mathcal{Y}$,*

$$\begin{aligned} \left|\frac{1}{n}\log|T^n(X)| - H(X)\right| &\le \epsilon_n \\ \left|\frac{1}{n}\log|T^n(Y|x^n)| - H(Y|X)\right| &\le \epsilon_n, \quad \forall x^n \in T^n(X). \end{aligned} \tag{4.2}$$

The next fact deals with the continuity of entropy with respect to probability distributions.

Fact 4.1.3. *[16, Lemma 2.7] If P and Q are two distributions on X such that*

$$\sum_{x\in\mathcal{X}} |P(x) - Q(x)| \le \epsilon \le \frac{1}{2} \tag{4.3}$$

then

$$|H(P) - H(Q)| \le -\epsilon \log \frac{\epsilon}{|\mathcal{X}|} \tag{4.4}$$

[1] The typical sets are with respect to distributions $P_X, P_{Y|X}$ and P_{XY}, respectively.

[2] The constants of the typical sets for each n, when suppressed, are understood to be some δ_n with $\delta_n \to 0$ and $\sqrt{n} \cdot \delta_n \to \infty$ (delta convention).

4.2 Typicality graphs

Definition 4.2.1. *For any any joint distribution $P_X \cdot P_{Y|X}$ on $\mathcal{X}, \times \mathcal{Y}$, and any $\epsilon_{1n}, \epsilon_{2n}, \lambda_n \to 0$, the sequence of typicality graphs $G_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ is defined as follows: for every n, G_n is a bipartite graph, with its left vertices consisting of all $x^n \in T^n_{\epsilon_{1n}}(X)$ and the right vertices consisting of all $y^n \in T^n_{\epsilon_{2n}}(Y)$. A vertex on the left (say $\tilde{x}^n$) is connected to a vertex on the right (say $\tilde{y}^n$) iff $(\tilde{x}^n, \tilde{y}^n) \in T^n_{\lambda_n}(X, Y)$.*

Remark 4.2.1. *Henceforth, we will assume that the sequences $\epsilon_{1n}, \epsilon_{2n}, \lambda_n$ satisfy the 'delta convention' [16, Convention 2.11], i.e.,*

$$\epsilon_{1n} \to 0, \quad \sqrt{n} \cdot \epsilon_{1n} \to \infty \text{ as } n \to \infty$$

with similar conditions for ϵ_{2n} and λ_n as well. The delta convention ensures that the typical sets have 'large probability'.

We will use the notation $V_X(.), V_Y(.)$ to denote the vertex sets of any bipartite graph. Some properties of the typicality graph are:

1. From Fact 4.1.2, we know that for any sequence of typicality graphs $\{G_n\}$, the cardinality of the vertex sets satisfies

$$\left| \frac{1}{n} \log |V_X(G_n)| - H(X) \right| \leq \epsilon_n, \quad \left| \frac{1}{n} \log |V_Y(G_n)| - H(Y) \right| \leq \epsilon_n \tag{4.5}$$

   for some sequence $\epsilon_n \to 0$.

2. The degree of each each vertex $i \in V_X(G_n)$ and $j \in V_Y(G_n)$ satisfies

$$\begin{aligned} \text{degree}(x^n) &\leq 2^{n(H(Y|X)+\epsilon_n)}, \quad \forall x^n \in V_X(G_n) \\ \text{degree}(y^n) &\leq 2^{n(H(X|Y)+\epsilon_n)}, \quad \forall y^n \in V_Y(G_n) \end{aligned} \tag{4.6}$$

   for some $\epsilon_n \to 0$.

The second property gives upper bounds on the degree of each vertex in the typicality graph. Since, we have not imposed any relationships between the typicality constants

$\epsilon_{1n}, \epsilon_{2n}$ and λ_n, in general, it cannot be assumed that the degree of *every* X-vertex (resp. Y-vertex) is close to $2^{nH(Y|X)}$ (resp. $2^{nH(X|Y)}$). However, such an assertion holds for *almost* every vertex in G_n . Specifically, we can show that the above degree conditions hold for a subgraph with exponentially the same size as G_n.

Theorem 4.2.1. *Every sequence of typicality graphs $G_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ has a sequence of subgraphs $A_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ satisfying the following properties for some $\delta_n \to 0$.*

1. *The vertex set sizes $|V_X(A_n)|$ and $|V_Y(A_n)|$, denoted θ_X^n and θ_Y^n, respectively, satisfy*
$$\left|\frac{1}{n}\log\theta_X^n - H(X)\right| \le \delta_n, \quad \left|\frac{1}{n}\log\theta_Y^n - H(Y)\right| \le \delta_n \quad \forall n$$
2. *The degree of each X-vertex x^n, denoted ${\theta'}^n(x^n)$ satisfies*
$$\left|\frac{1}{n}\log{\theta'}^n(x^n) - H(Y|X)\right| \le \delta_n \quad \forall x^n \in V_X(A_n).$$
3. *The degree of each Y-vertex y^n, denoted ${\theta'}^n(y^n)$, satisfies*
$$\left|\frac{1}{n}\log{\theta'}^n(y^n) - H(X|Y)\right| \le \delta_n \quad \forall y^n \in V_Y(A_n).$$

Proof. The proof is provided in section 4.4 □

4.3 Sub-graphs contained in typicality graphs

In this section, we study the subgraphs contained in a sequence of typicality graphs.

4.3.1 Subgraphs of general degree

Definition 4.3.1. *A sequence of typicality graphs $G_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ is said to contain a sequence of subgraphs Γ_n of rates (R_X, R_Y, R'_X, R'_Y) if for each n, there exists a sequence $\delta_n \to 0$ such that*

1. *The vertex sets of the subgraphs have sizes (denoted Δ_X^n and Δ_Y^n) that satisfy*

$$\left|\frac{1}{n}\log\Delta_X^n - R_X\right| \leq \delta_n, \quad \left|\frac{1}{n}\log\Delta_Y^n - R_Y\right| \leq \delta_n, \; \forall n. \tag{4.7}$$

2. *The degree of each vertex x^n in $V_X(\Gamma_n)$, denoted $\Delta'^n(x^n)$ satisfies*

$$\left|\frac{1}{n}\log\Delta'^n(x^n) - R'_Y\right| \leq \delta_n, \quad \forall x^n \in V_X(\Gamma_n), \; \forall n. \tag{4.8}$$

3. *The degree of each vertex y^n in the $V_Y(\Gamma_n)$, denoted $\Delta'^n(y^n)$ satisfies*

$$\left|\frac{1}{n}\log\Delta'^n(y^n) - R'_X\right| \leq \delta_n, \quad \forall y^n \in V_Y(\Gamma_n), \; \forall n. \tag{4.9}$$

The following theorem gives a characterization of the rate-tuple of a sequence of subgraphs in the sequence of typicality graphs of P_{XY}.

Theorem 4.3.1. *Let $G_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ be a sequence of typicality graphs of P_{XY}. Define*

$$\mathcal{R} \triangleq \{(R_X, R_Y, R'_X, R'_Y) : G_n \text{ contains subgraphs of rates } (R_X, R_Y, R'_X, R'_Y)\} \tag{4.10}$$

Then,

$$\mathcal{R} \supseteq \left\{ (R_X, R_Y, R'_X, R'_Y) : \begin{array}{c} R_X + R'_Y = R_Y + R'_X, \\ R_X \leq H(X|U),\; R_Y \leq H(Y|U),\; R'_X \leq H(X|YU), \\ R'_Y \leq H(Y|XU) \text{ for some } P_{U|XY} \end{array} \right\}. \tag{4.11}$$

Proof. The proof is provided in Section 4.4. □

4.3.2 Nearly complete subgraphs

A complete bipartite graph is one in which each vertex of the first set is connected with every vertex on the other set. We next consider a specific class of subgraphs, namely nearly complete subgraphs. For this class of subgraphs, we have a converse result that fully characterizes the set of nearly complete subgraphs present in any typicality graph.

Definition 4.3.2. *A sequence of typicality graphs $G_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ is said to contain a sequence of nearly complete subgraphs $\Gamma_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ of rates (R_X, R_Y) if for each n, there exists a sequence $\delta_n \to 0$ such that*

1. *The sizes of the vertex sets of the subgraphs, denoted Δ_X^n and Δ_Y^n, satisfy*

$$\left|\frac{1}{n}\log \Delta_X^n - R_X\right| \leq \delta_n, \quad \left|\frac{1}{n}\log \Delta_Y^n - R_Y\right| \leq \delta_n, \, \forall n. \tag{4.12}$$

2. *The degree of each vertex x^n in the X-set, denoted $\Delta'^n(x^n)$ satisfies*

$$\frac{1}{n}\log \Delta'^n(x^n) \geq R_Y - \delta_n, \quad \forall x^n \in V_X(\Gamma_n), \, \forall n. \tag{4.13}$$

3. *The degree of each vertex y^n in the Y-set, denoted $\Delta'^n(x^n)$ satisfies*

$$\frac{1}{n}\log \Delta'^n(y^n) \geq R_X - \delta_n, \quad \forall y^n \in V_Y(\Gamma_n), \, \forall n. \tag{4.14}$$

Theorem 4.3.2. *Let $G_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ be a sequence of typicality graphs for P_{XY}. Define*

$$\mathcal{R} \triangleq \left\{ (R_X, R_Y) : \begin{array}{c} G_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n) \textit{ contains nearly complete subgraphs} \\ \textit{of rates } (R_X, R_Y) \end{array} \right\}, \tag{4.15}$$

Then,

$$\mathcal{R} \supseteq \{(R_X, R_Y) : R_X \leq H(X|U), \; R_Y \leq H(Y|U) \textit{ for some } P_{U|XY} \textit{ s.t. } X - U - Y\},^3 \tag{4.16}$$

and for all sequences of nearly complete subgraphs of G_n such that the sequence δ_n (in Definition 4.3.2) converges to 0 faster than $1/\log n$ (more precisely, $\delta_n = o(\frac{1}{\log n})$

[3] X, U, Y form a Markov chain, in that order.

or $\lim_{n\to\infty} \delta_n \log n = 0$*), the rates of the subgraph* (R_X, R_Y) *satisfy*

$$R_X \leq H(X|U),\ R_Y \leq H(Y|U) \text{ for some } P_{U|XY} \text{ s.t. } X - U - Y$$

Proof. The proof is provided in Section 4.4. □

4.3.3 Nearly Empty Subgraphs

So far, we have discussed properties of subgraphs of the typicality graph G_n such as the containment of nearly complete subgraphs and subgraphs of general degree. Now, we turn our attention to the presence of nearly empty subgraphs in the typicality graph. Our approach towards this problem differs slightly from the approach we took in Sections 4.3.1 and 4.3.2. While in previous sections we characterized the subgraphs based on the degrees of their vertices, in this section we characterize nearly empty subgraphs by the total number of edges present in such graphs. In this section, we take a different approach than the one used in previous sections and analyze the probability that a randomly chosen subgraph of the typicality graph has far fewer edges than expected. In particular, we focus on the case of a random subgraph with no edges.

Consider a pair (X, Y) of discrete memoryless stationary correlated sources with finite alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Suppose we sample 2^{nR_1} sequences from the typical set of X, $T^n_{\epsilon_{1n}}(X)$, independently with replacement and similarly sample 2^{nR_2} sequences from the typical set of Y, $T^n_{\epsilon_{2n}}(Y)$. The underlying typicality graph $G_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ induces a bipartite graph on these $2^{nR_1} + 2^{nR_2}$ sequences. We provide a characterization of the probability that this graph is sparser than expected. This characterization is obtained using a version of Suen's inequalities [34] and the Lovasz local lemma [7] listed below.

Lemma 4.3.1. *[34] Let* $I_i \in Be(p_i), i \in \mathcal{I}$ *be a family of Bernoulli random variables. Their dependency graph* L *is formed in the following manner. Denote the random variable* I_i *by a vertex* i *and join vertices* i *and* j *by an edge if the corresponding ran-*

dom variables are dependent. Let $X = \sum_i \mathbb{E}(I_i)$ *and* $\Gamma = \mathbb{E}(X) = \sum_i p_i$*. Moreover, write* $i \sim j$ *if* (i, j) *is an edge in the dependency graph* L *and let* $\Theta = \frac{1}{2}\sum_i \sum_{j\sim i} \mathbb{E}(I_i I_j)$ *and* $\theta = \max_i \sum_{j\sim i} p_j$*. Then, Suen's inequalities state that for any* $0 \le a \le 1$*,*

$$P(X \le a\Gamma) \le \exp\left\{-\min\left((1-a)^2 \frac{\Gamma^2}{8\Theta + 2\Gamma}, (1-a)\frac{\Gamma}{6\theta}\right)\right\} \tag{4.17}$$

Putting $a = 0$*, this can be further tightened to*

$$P(X = 0) \le \exp\left\{-\min\left(\frac{\Gamma^2}{8\Theta}, \frac{\Gamma}{2}, \frac{\Gamma}{6\theta}\right)\right\} \tag{4.18}$$

Lemma 4.3.2. *[7] Let* L *be the dependency graph for events* $\varepsilon_1, \ldots, \varepsilon_n$ *in a probability space and let* $E(L)$ *be the edge set of* L*. Suppose there exists* $x_i \in [0, 1], 1 \le i \le n$ *such that*

$$P(\varepsilon_i) \le x_i \prod_{(i,j)\in E(L)} (1 - x_j). \tag{4.19}$$

Then, we have

$$P(\cap_{i=1}^n \overline{\varepsilon_i}) \ge \prod_{i=1}^n (1 - x_i). \tag{4.20}$$

Another version of the local lemma is as given below. Let $\phi(x), 0 \le x \le e^{-1}$ *be the smallest root of the equation* $\phi(x) = e^{x\phi(x)}$*. With definitions of* Γ *and* θ *as in Lemma 4.3.1 and defining* $\tau \triangleq \max_i P(\varepsilon_i)$*, we have*

$$P\left(\cap_{i=1}^n \overline{\varepsilon_i}\right) \ge \exp\left\{-\Gamma\phi(\theta + \tau)\right\} \tag{4.21}$$

With these preliminaries, we are ready to state the main result of this section.

Theorem 4.3.3. *Suppose* X *and* Y *are correlated finite alphabet memoryless random variables with joint distribution* $p(x, y)$*. Let* $\epsilon_{1n}, \epsilon_{2n}, \lambda_n$ *satisfy the 'delta convention' and* R_1, R_2 *be any positive real numbers such that* $R_1 + R_2 > I(X; Y)$*. Let* C_X *be a collection of* 2^{nR_1} *sequences picked independently and with replacement from* $T^n_{\epsilon_{1n}}(X)$

and let C_Y be defined similarly. Let U be the cardinality of the set

$$\mathcal{U} \triangleq \{(x^n, y^n) \in C_X \times C_Y \colon (x^n, y^n) \in T^n_{\lambda_n}(X, Y)\} \tag{4.22}$$

Assume, without loss of generality that $R_1 \geq R_2$. Then, for any $\gamma \geq 0$, we have

$$\lim_{n\to\infty} \frac{1}{n} \log\log \left[\mathbb{P}\left(\frac{\mathbb{E}(U) - U}{\mathbb{E}(U)} \geq e^{-n\gamma}\right)\right]^{-1} \geq \begin{cases} R_1 + R_2 - I(X;Y) - \gamma & \text{if } R_1 < I(X;Y) \\ R_2 - \gamma & \text{if } R_1 \geq I(X;Y) \end{cases} \tag{4.23}$$

Setting $\gamma = 0$ in the above equation gives us

$$\lim_{n\to\infty} \frac{1}{n} \log\log \frac{1}{\mathbb{P}(U = 0)} \geq \min\left(R_2, R_1 + R_2 - I(X;Y)\right) \tag{4.24}$$

This inequality holds with equality when $R_2 \leq R_1 \leq I(X;Y)$.

Proof. The proof is provided in Section 4.4. □

4.4 Proof of Theorems

Proof. (**Theorem 4.2.1**) The vertex sets $V_X(G_n)$ and $V_Y(G_n)$ are the ϵ_{1n}-typical and ϵ_{2n}-typical sets of P_X and P_Y, respectively. To define the subgraphs A_n, we would like to choose the sequences with type P_X and P_Y, respectively as the vertex sets of the subgraph, with an edge connecting two sequences if they have joint type P_{XY}. However, the values taken by the joint pmfs P_{XY}, P_X, P_Y may be any real number between 0 and 1, whereas the joint type of two n-sequences is always a rational number(with denominator n). Therefore, we choose the subgraph A_n as follows:

- For each n, approximate the values of P_{XY} to rational numbers with denominator n to obtain pmf $\tilde{P}_{XY}$, respectively. Clearly, $\tilde{P}_{XY}$ is a valid joint type of length n and the maximum approximation error is bounded by $\frac{1}{n}$. In fact, for all sufficiently large n:

$$|P_{XY}(x,y) - \tilde{P}_{XY}(x,y)| < \frac{1}{n} << \frac{1}{\sqrt{n}} < \lambda_n \quad \forall (x,y), \tag{4.25}$$

where the last inequality follows from the delta convention. Using Fact 4.1.1, we also have

$$|P_X(x) - \tilde{P}_X(x)| < |\mathcal{Y}| \cdot \frac{1}{n} << \frac{1}{\sqrt{n}} < \epsilon_{1n} \tag{4.26}$$

$$|P_Y(y) - \tilde{P}_Y(y)| < |\mathcal{X}| \cdot \frac{1}{n} << \frac{1}{\sqrt{n}} < \epsilon_{2n} \tag{4.27}$$

- The left vertex set of A_n is $T_0^n(\tilde{P}_X)$, i.e., the set of x^n sequences with type $\tilde{P}_X$. The right vertex set of A_n is $T_0^n(\tilde{P}_Y)$- the set of y^n sequences with type $\tilde{P}_Y$. A vertex in $V_X(A_n)$, say a^n is connected to a vertex in $V_Y(A_n)$, say b^n iff $(a^n, b^n) \in T_0^n(\tilde{P}_{X,Y})$, i.e., (a^n, b^n) have joint type $\tilde{P}_{XY}$.

From (4.25),(4.26) and (4.27), we have

$$T_0^n(\tilde{P}_X) \subset T_{\epsilon_{1n}}^n(P_X), \quad T_0^n(\tilde{P}_Y) \subset T_{\epsilon_{2n}}^n(P_Y), \quad T_0^n(\tilde{P}_{X,Y}) \subset T_{\lambda_n}^n(P_{X,Y}). \tag{4.28}$$

Hence A_n is a subgraph of G_n, as required. From [16, Lemma 2.3], we have

$$\left|\frac{1}{n}\log|T_0^n(\tilde{P}_X)| - H(\tilde{P}_X)\right| \le \delta_{1n}, \quad \left|\frac{1}{n}\log|T_0^n(\tilde{P}_Y)| - H(\tilde{P}_Y)\right| \le \delta_{2n} \quad \forall n, \tag{4.29}$$

where $\delta_{1n} = (n+1)^{-|\mathcal{X}|}$ and $\delta_{2n} = (n+1)^{-|\mathcal{Y}|}$. Fact 4.1.3 establishes the continuity of entropy with respect to the probability distribution. Using Fact 4.1.3 along with (4.25),(4.26) and (4.27), we obtain

$$\left|\frac{1}{n}\log|T_0^n(\tilde{P}_X)| - H(P_X)\right| \le \delta_{1n}, \quad \left|\frac{1}{n}\log|T_0^n(\tilde{P}_Y)| - H(P_Y)\right| \le \delta_{2n} \quad \forall n, \tag{4.30}$$

where we have reused δ_{1n}, δ_{2n} with some abuse of notation. This proves the first property.

We now note that $x^n \in V_X(A_n) = T_0^n(\tilde{P}_X)$ and $y^n \in T_0^n(\tilde{P}_{Y|X}|x^n)$ implies a) $(x^n, y^n) \in T_0^n(\tilde{P}_{X,Y})$ and b) $y^n \in T_0^n(\tilde{P}_Y) = V_Y(A_n)$ (Fact 4.1.1). This implies

$$\text{degree}(x^n) \geq |T_0^n(\tilde{P}_{Y|X}|x^n)|, \forall x^n \in V_X(A_n). \tag{4.31}$$

From [16, Lemma 2.5], we know that

$$|T_0^n(\tilde{P}_{Y|X})| \geq 2^{n(H(\tilde{P}_{Y|X})-\delta_{3n})} \tag{4.32}$$

where $\delta_{3n} = |\mathcal{X}||\mathcal{Y}|\frac{\log(n+1)}{n}$. In the above, $H(\tilde{P}_{Y|X})$ stands for $H(Y|X)$ computed under the joint distribution $\tilde{P}_{XY}$. Combining this with (4.31), we get a lower bound on the degree of each $x^n \in V_X(A_n)$:

$$\text{degree}(x^n) \geq 2^{n(H(\tilde{P}_{Y|X})-\delta_{3n})} \tag{4.33}$$

From (4.25) and (4.26), one can deduce that $\forall x, y$

$$|P_{Y|X}(y|x) - \tilde{P}_{Y|X}(y|x)| < \gamma_n$$

for some $\gamma_n \to 0$. Combining this with Fact 4.1.3, (4.33) can be written as

$$\text{degree}(x^n) \geq 2^{n(H(P_{Y|X})-\delta_{3n})}, \tag{4.34}$$

where we reuse the symbol δ_{3n}.

Furthermore, (4.6) gives an upper bound on the degree of each vertex in G_n. Hence we have

$$\left|\frac{1}{n}\log \theta'^n(x^n) - H(Y|X)\right| \leq \max(\delta_{3n}, \epsilon_n) \quad \forall x^n \in V_X(A_n) \tag{4.35}$$

Similarly, we can bound the degree of each vertex in $V_Y(A_n)$ as

$$\left|\frac{1}{n}\log \theta'^n(y^n) - H(X|Y)\right| \leq \max(\delta_{4n}, \epsilon_n) \quad \forall y^n \in V_Y(A_n) \tag{4.36}$$

Finally, we can set $\delta_n = \max(\delta_{1n}, \delta_{2n}, \delta_{3n}, \delta_{4n}, \epsilon_n)$ to complete the proof of the proposition.

$\square$

Proof. **(Theorem 4.3.1)** The proof is along the same lines as the proof of Theorem 4.2.1), but we will repeat it for completeness. For every n, we shall demonstrate the existence of a subgraph Γ_n with the required rates contained within $A_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ (A_n is the subgraph specified by Theorem 4.2.1).

Definition of Γ_n. Consider any conditional distribution $P_{U|XY}$. This fixes the joint distribution $P_{XYU} = P_{XY}P_{U|XY}$. We construct Γ_n as follows.

- For each n, approximate the values of P_{UXY} to rational numbers with denominator n to obtain pmf $\tilde{P}_{UXY}$, respectively. Clearly $\tilde{P}_{UXY}$ is a valid joint type of length n and the maximum approximation error is bounded by $\frac{1}{n}$. By marginalizing the joint pmf, for all x, y we also have

$$|P_{XY}(x,y) - \tilde{P}_{XY}(x,y)| < |\mathcal{U}| \cdot \frac{1}{n} << \frac{1}{\sqrt{n}} < \lambda_n, \tag{4.37a}$$

$$|P_X(x) - \tilde{P}_X(x)| < |\mathcal{Y}| \cdot |\mathcal{U}| \cdot \frac{1}{n} << \frac{1}{\sqrt{n}} < \epsilon_{1n} \tag{4.37b}$$

$$|P_Y(y) - \tilde{P}_Y(y)| < |\mathcal{X}| \cdot |\mathcal{U}| \cdot \frac{1}{n} << \frac{1}{\sqrt{n}} < \epsilon_{2n}, \tag{4.37c}$$

  where the last inequality in each equation follows from the delta convention. Furthermore, for all u

$$|P_U(u) - \tilde{P}_U(u)| < |\mathcal{Y}| \cdot |\mathcal{X}| \cdot \frac{1}{n}. \tag{4.38}$$

- Pick any length n sequence u^n with type $\tilde{P}_U$, i.e., $u^n \in T_0^n(\tilde{P}_U)$. Consider a bipartite graph Γ_n with X-vertices consisting of all $x^n \in T_0^n(\tilde{P}_{X|U}|u^n)$, Y-vertices consisting of all $y^n \in T_0^n(\tilde{P}_{Y|U}|u^n)$. In other words, having fixed u^n, the X-vertex sets and Y-vertex sets consist of all length n sequences having conditional type $\tilde{P}_{X|U}$ and $\tilde{P}_{Y|U}$, respectively. Vertices $x^n \in V_X(\Gamma_n)$ and $y^n \in$

$V_Y(\Gamma_n)$ are connected in Γ_n iff $(x^n, y^n) \in T_0^n(\tilde{P}_{XY|U}|u^n)$, i.e., if they have the conditional joint type $P_{XY|U}$ given u^n.

First, let us verify that Γ_n is a subgraph of G_n. From Fact 4.1.1, if $u^n \in T_0^n(\tilde{P}_U)$ and $x^n \in T_0^n(\tilde{P}_{X|U}|u^n)$, then $(x^n, u^n) \in T_0^n(\tilde{P}_{X,U})$. Consequently, $x^n \in T_0^n(\tilde{P}_X)$. Similarly, all $y^n \in T_0^n(\tilde{P}_{Y|U}|u^n)$ belong to $T_0^n(\tilde{P}_Y)$. On the same lines, if $u^n \in T_0^n(\tilde{P}_U)$ and $(x^n, y^n) \in T_0^n(\tilde{P}_{XY|U}|u^n)$, then $(x^n, y^n, u^n) \in T_0^n(\tilde{P}_{X,Y,U})$. This implies $(x^n, y^n) \in T_0^n(\tilde{P}_{X,Y})$. Furthermore, from (4.37a),(4.37b) and (4.37c), we know

$$\begin{aligned} T_0^n(\tilde{P}_X) \subset T_{\epsilon_{1n}}^n(P_X) = V_X(G_n), \quad T_0^n(\tilde{P}_Y) \subset T_{\epsilon_{2n}}^n(P_Y) = V_Y(G_n) \quad \text{and} \\ T_0^n(\tilde{P}_{X,Y}) \subset T_{\lambda_n}^n(P_{X,Y}). \end{aligned} \tag{4.39}$$

Hence, for all sufficiently large n, Γ_n is a subgraph of the typicality graph G_n.

Properties of Γ_n. From [16, Lemma 2.3], we have

$$\left|\frac{1}{n}\log|T_0^n(\tilde{P}_{X|U}|u^n)| - H(\tilde{P}_{X|U})\right| \leq \delta_{1n}, \quad \left|\frac{1}{n}\log|T_0^n(\tilde{P}_{Y|U}|u^n)| - H(\tilde{P}_{Y|U})\right| \leq \delta_{2n} \ \ \forall n, \tag{4.40}$$

where $\delta_{1n} = (n+1)^{-|\mathcal{X}||\mathcal{U}|}$ and $\delta_{2n} = (n+1)^{-|\mathcal{Y}||\mathcal{U}|}$. Using (4.37b), (4.37c) with (4.38), we know that $\tilde{P}_{X|U}, \tilde{P}_{Y|U}$ are close to $P_{X|U}, P_{Y|U}$, respectively. Using Fact 4.1.3, we know that the entropies $H(\tilde{P}_{X|U}), H(\tilde{P}_{Y|U})$ must close to $H(P_{X|U}), H(P_{Y|U})$, respectively. Thus, we can write (4.40) as (reusing δ_{1n}, δ_{2n})

$$\left|\frac{1}{n}\log|T_0^n(\tilde{P}_{X|U}|u^n)| - H(P_{X|U})\right| \leq \delta_{1n}, \quad \left|\frac{1}{n}\log|T_0^n(\tilde{P}_{Y|U}|u^n)| - H(P_{Y|U})\right| \leq \delta_{2n} \ \ \forall n, \tag{4.41}$$

Thus, the vertex sets of Γ_n have rates $R_X = H(X|U)$ and $R_Y = H(Y|U)$, as required.

Using Fact 4.1.1, for any $x^n \in V_X(\Gamma_n)$, every $y^n \in T_0^n(\tilde{P}_{Y|XU}|x^n, u^n)$ will satisfy a) $(x^n, y^n) \in T_0^n(\tilde{P}_{X,Y|U}|u^n)$ and b) $y^n \in T_0^n(\tilde{P}_{Y|U}|u^n)$. Hence

$$\text{degree}(x^n) \geq |T_0^n(\tilde{P}_{Y|XU}|x^n, u^n)| \geq 2^{n(H(\tilde{P}_{Y|XU}) - \delta_{3n})}, \tag{4.42}$$

where $\delta_{3n} = |\mathcal{X}||\mathcal{Y}||\mathcal{U}|\frac{\log(n+1)}{n}$. We can also upper bound the degree of x^n by noting that $x^n \in T_0^n(\tilde{P}_{X|U}|u^n)$ and $(x^n, y^n) \in T_0^n(\tilde{P}_{X,Y|U}|u^n)$ implies $y^n \in T_0^n(\tilde{P}_{Y|XU}|x^n, u^n)$. From [16, Lemma 2.5],

$$|T_0^n(\tilde{P}_{Y|XU}|x^n, u^n)| \leq 2^{nH(\tilde{P}_{Y|XU})}.$$

Combining this with (4.42), we have

$$\left|\frac{1}{n}\log \Delta'^n(x^n) - H(\tilde{P}_{Y|XU})\right| \leq \delta_{3n}, \quad \forall x^n \in V_X(\Gamma_n),\ \forall n. \tag{4.43}$$

In a similar fashion, we can show that

$$\left|\frac{1}{n}\log \Delta'^n(y^n) - H(\tilde{P}_{X|YU})\right| \leq \delta_{4n}, \quad \forall y^n \in V_Y(\Gamma_n),\ \forall n. \tag{4.44}$$

Since the distributions $\tilde{P}_{Y|XU}$ and $\tilde{P}_{X|YU}$ are close to $P_{Y|XU}$ and $P_{X|YU}$, respectively, Fact 3 enables us to replace $H(\tilde{P}_{Y|XU}), H(\tilde{P}_{X|YU})$ with $H(P_{Y|XU}), H(P_{X|YU})$, respectively in the two preceding equations.

Taking $\delta_n = \max(\delta_{1n}, \delta_{2n}, \delta_{3n}, \delta_{4n})$, we have shown the existence of a sequence of subgraphs Γ_n with rates $(H(X|U), H(Y|U), H(Y|XU), H(X|YU))$. Since we can simply exclude edges from Γ_n to obtain subgraphs with smaller rates, it is clear that all rate tuples characterized by

$$\left\{ (R_X, R_Y, R'_X, R'_Y) : \begin{array}{c} R_X + R'_Y = R_Y + R'_X \\ R_X \leq H(X|U),\ R_Y \leq H(Y|U), \\ R'_X \leq H(X|YU),\ R'_Y \leq H(Y|XU) \end{array} \right\} \tag{4.45}$$

are achievable for every conditional distribution $P_{U|XY}$. Note that the first equality results from the fixed number of edges, regardless of whether they are counted from the left or right side. □

Proof. **(Theorem 4.3.2)** The first part of the theorem follows directly from Theorem 4.3.1 by choosing $P_{U|XY}$ such that $X - U - Y$ form a Markov chain. We now prove the converse under the stated assumption that the sequence δ_n satisfies

$\lim_{n\to\infty} \delta_n \log n = 0$.

Suppose that a sequence of typicality graphs $G_n(\epsilon_{1n}, \epsilon_{2n}, \lambda_n)$ contains nearly complete subgraphs Γ_n of rates R_X, R_Y. The total number of edges in Γ_n can be lower bounded as

$$
\begin{aligned}
|\text{Edges}(\Gamma_n)| &\geq \Delta_X^n \cdot \text{ minimum degree of a vertex in } V_X(\Gamma_n) \\
&\geq \Delta_X^n \cdot 2^{n(R_Y - \delta_n)} \\
&\geq \Delta_X^n \cdot 2^{n(R_Y - \delta_n)} \Delta_Y^n \cdot 2^{-n(R_Y + \delta_n)} \\
&= \Delta_X^n \cdot \Delta_Y^n \cdot 2^{-2n\delta_n}.
\end{aligned} \tag{4.46}
$$

Each of these edges represent a pair (x^n, y^n) that is jointly λ_n-typical with respect to the distribution P_{XY}. In other words, each of these pairs (x^n, y^n) belongs to a joint type [16] that is 'close' to P_{XY}. Since the number of joint types of a pair of sequences of length n is at most $(n+1)^{|\mathcal{X}||\mathcal{Y}|}$, the number of edges belonging to the dominant joint type, say $\bar{P}_{XY}$ satisfies

$$
|\text{Edges}(\Gamma_n) \text{ having joint type } \bar{P}_{XY}| \geq \frac{\Delta_X^n \cdot \Delta_Y^n 2^{-2n\delta_n}}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}}. \tag{4.47}
$$

Define a subgraph $\mathcal{A}_n$ of Γ_n consisting only of the edges having joint type $\bar{P}_{XY}$. A word about the notation used in the sequel: We will use i, j to index the vertices in $V_X(\Gamma_n), V_Y(\Gamma_n)$, respectively. Thus $i \in \{1, \ldots, \Delta_X^n\}$ and $j \in \{1, \ldots, \Delta_Y^n\}$. The actual sequences corresponding to these vertices will be denoted $x^n(i), y^n(j)$ etc. Using this notation,

$$
\mathcal{A}_n \triangleq \{(i,j) : i \in V_X(\Gamma_n), j \in V_Y(\Gamma_n) \text{ s.t. } (x^n(i), y^n(j)) \text{ has joint type } \bar{P}_{XY}\} \tag{4.48}
$$

From (4.47),

$$
|\mathcal{A}_n| \geq \frac{\Delta_X^n \cdot \Delta_Y^n 2^{-2n\delta_n}}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} \tag{4.49}
$$

We will prove the converse result using a series of lemmas concerning $\mathcal{A}_n$. Some of the lemmas are similar to those required to prove Theorem 5.3.2. We only sketch the proofs of such lemmas, referring the reader to Chapter 5 for details.

Define random variables X'^n, Y'^n with pmf

$$\mathbb{P}\left((X'^n, Y'^n) = (x^n(i), y^n(j)\right) = \frac{1}{|\mathcal{A}_n|}, \quad \text{if } (i,j) \in \mathcal{A}_n. \tag{4.50}$$

Lemma 4.4.1. *$I(X'^n \wedge Y'^n) \leq 2n\delta_n + |\mathcal{X}||\mathcal{Y}| \log(n+1)$.*

Proof. Follow steps similar to the proof of Lemma 5.6.2, using (4.49) to lower bound the size of $\mathcal{A}_n$. □

Let us apply Lemma 5.6.3 to random variables X'^n and Y'^n. Lemma 4.4.1 indicates $\sigma = 2n\delta_n + |\mathcal{X}||\mathcal{Y}| \log(n+1)$, and δ shall be specified later. Hence, for some

$$k \leq \frac{2\sigma}{\delta} = \frac{2(n\delta_n + |\mathcal{X}||\mathcal{Y}| \log(n+1))}{\delta}, \tag{4.51}$$

there exist $\bar{x}_{t_1}, \bar{y}_{t_1}, \bar{x}_{t_2}, \bar{y}_{t_2}, ..., \bar{x}_{t_k}, \bar{y}_{t_k}$ such that

$$I(X'_t \wedge Y'_t | X'_{t_1} = \bar{x}_{t_1}, Y'_{t_1} = \bar{y}_{t_1}, ..., X'_{t_k} = \bar{x}_{t_k}, Y'_{t_k} = \bar{y}_{t_k}) \leq \delta \qquad \text{for } t = 1, 2, ..., n. \tag{4.52}$$

We now define a subgraph of $\mathcal{A}_n$ consisting of all edges (X'^n, Y'^n) that have

$$X'_{t_1} = \bar{x}_{t_1}, Y'_{t_1} = \bar{y}_{t_1}, ..., X'_{t_k} = \bar{x}_{t_k}, Y'_{t_k} = \bar{y}_{t_k}$$

The subgraph denoted as $\bar{\mathcal{A}}_n$ is given by: [4]

$$\bar{\mathcal{A}}_n \triangleq \{(i,j) \in \mathcal{A}_n : X'_{t_1}(i) = \bar{x}_{t_1}, Y'_{t_1}(j) = \bar{y}_{t_1}, ..., X'_{t_k}(i) = \bar{x}_{t_k}, Y'_{t_k}(j) = \bar{y}_{t_k}\}. \tag{4.53}$$

On the same lines as Lemma 5.6.4, we have

$$|\bar{\mathcal{A}}_n| \geq \left(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma - \delta)}\right)^k |\mathcal{A}_n|. \tag{4.54}$$

[4] The heirarchy of subgraphs is $G_n \supset \Gamma_n \supset \mathcal{A}_n \supset \bar{\mathcal{A}}_n$

Let us define random variables $\bar{X}^n$, $\bar{Y}^n$ on $\mathcal{X}^n$ resp. $\mathcal{Y}^n$ by

$$\mathbb{P}((\bar{X}^n, \bar{Y}^n) = (x^n(i), y^n(j)) = \frac{1}{|\bar{\mathcal{A}}_n|} \text{if } (i,j) \in \bar{\mathcal{A}}_n. \tag{4.55}$$

If we denote $\bar{X}^n = (\bar{X}_1, ..., \bar{X}_n)$, $Y^n = (\bar{Y}_1, ..., \bar{Y}_n)$, the Fano-distribution of the graph $\bar{\mathcal{A}}_n$ induces a distribution $P_{\bar{X}_t, \bar{Y}_t}$ on the random variables $\bar{X}_t\bar{Y}_t, t = 1, \ldots, n$. One can show that

$$\begin{aligned} &\mathbb{P}(\bar{X}_t = x, \bar{Y}_t = y) \\ &\quad = \mathbb{P}(X'_t = x, \bar{Y}'_t = y | X'_{t_1}(i) = \bar{x}_{t_1}, Y'_{t_1}(j) = \bar{y}_{t_1}, ..., X'_{t_k}(i) = \bar{x}_{t_k}, Y'_{t_k}(j) = \bar{y}_{t_k}), \ \forall t. \end{aligned} \tag{4.56}$$

Using (4.56) in Lemma 5.6.4, we get the bound $I(\bar{X}_t \wedge \bar{Y}_t) < \delta$. Applying Pinsker's inequality for I-divergences [24], we have

$$\sum_{x,y} |\mathbb{P}(\bar{X}_t = x, \bar{Y}_t = y) - \mathbb{P}(\bar{X}_t = x)\mathbb{P}(\bar{Y}_t = y)| \leq 2\delta^{1/2}, \quad 1 \leq t \leq n. \tag{4.57}$$

Also define

$$\bar{\mathcal{C}}(i) = \{(i,j) : (i,j) \in \bar{\mathcal{A}}_n, 1 \leq j \leq \Delta_Y^n\}, \tag{4.58a}$$

$$\bar{\mathcal{B}}(j) = \{(i,j) : (i,j) \in \bar{\mathcal{A}}_n, 1 \leq i \leq \Delta_X^n\}. \tag{4.58b}$$

We are now ready to present the final lemma required to complete the proof of the converse.

Lemma 4.4.2.

$$R_X \leq \frac{1}{n}\sum_{t=1}^{n} H(\bar{X}_t|\bar{Y}_t) + \delta_{1n} \tag{4.59a}$$

$$R_Y \leq \frac{1}{n}\sum_{t=1}^{n} H(\bar{Y}_t|\bar{X}_t) + \delta_{2n} \tag{4.59b}$$

$$R_X + R_Y \leq \frac{1}{n}\sum_{t=1}^{n} H(\bar{X}_t\bar{Y}_t) + +\delta_{3n} \tag{4.59c}$$

for some $\delta_{1n}, \delta_{2n}, \delta_{3n} \to 0$ *and the distributions of the RV's are determined by the Fano-distribution on the codewords* $\{(x^n(i), y^n(j)) : (i,j) \in \bar{\mathcal{A}}_n\}$.

Proof. We use a strong converse result for non-stationary discrete memoryless channels, found in [1]. Consider a DMC with input A_t and output B_t ($t = 1, \dots, n$), with average error probability $\lambda\,(0 \le \lambda < 1)$. The result states that the size of the message set M is upper-bounded as

$$\log M < \sum_{t=1}^{n} I(A_t \wedge B_t) + \frac{3}{1-\lambda}|\mathcal{A}|n^{1/2}, \tag{4.60}$$

where the distributions of the RV's are determined by the Fano-distribution on the codewords.

We apply the above result to three noiseless DMCs ($B_t = A_t, \lambda = 0$) as follows. Fix $\bar{Y}^n = y^n(j)$ for some $j \in \bar{\mathcal{A}}_n$ and let the input be $\bar{X}_t$, $t = 1, \cdots, n$. Then, from (4.60) we have

$$\log|\bar{\mathcal{B}}(j)| \le \sum_{t=1}^{n} H(\bar{X}_t|\bar{Y}_t = y_t(j)) + 3|\mathcal{X}|n^{1/2}. \tag{4.61}$$

Similarly,

$$\log|\bar{\mathcal{C}}(i)| \le \sum_{t=1}^{n} H(\bar{Y}_t|\bar{X}_t = x_t(i)) + 3|\mathcal{Y}|n^{1/2}, \tag{4.62}$$

$$\log|\bar{\mathcal{A}}_n| \le \sum_{t=1}^{n} H(\bar{X}_t\bar{Y}_t) + 3|\mathcal{X}||\mathcal{Y}|n^{1/2}. \tag{4.63}$$

Noting that $\mathbb{P}(\bar{Y}_t = y) = |\bar{\mathcal{A}}|^{-1}\sum_{(i,j)\in\bar{\mathcal{A}}_n} 1_{\{y_t(j)=y\}}$, we can sum both sides of (4.61) over all $(i,j) \in \bar{\mathcal{A}}_n$ to obtain

$$|\bar{\mathcal{A}}_n|^{-1} \sum_{(i,j)\in\bar{\mathcal{A}}_n} \log|\bar{\mathcal{B}}(j)| \le \sum_{t=1}^{n} H(\bar{X}_t|\bar{Y}_t) + 3|\mathcal{X}|n^{1/2}. \tag{4.64}$$

Define

$$B^* \triangleq \frac{2^{-2n\delta_n}}{n} \frac{\Delta_X^n}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} \left(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma-\delta)}\right)^k. \tag{4.65}$$

Then,

$$
\begin{aligned}
|\bar{\mathcal{A}}_n|^{-1} \sum_{(i,j)\in\bar{\mathcal{A}}_n} \log|\bar{\mathcal{B}}(j)| &= |\bar{\mathcal{A}}_n|^{-1} \sum_{j} |\bar{\mathcal{B}}(j)| \log|\bar{\mathcal{B}}(j)| \\
&\geq |\bar{\mathcal{A}}_n|^{-1} \sum_{j:|\bar{\mathcal{B}}(j)|\geq B^*} |\bar{\mathcal{B}}(j)| \log|\bar{\mathcal{B}}(j)| \\
&\geq |\bar{\mathcal{A}}_n|^{-1} \log(B^*) \sum_{j:|\bar{\mathcal{B}}(j)|\geq B^*} |\bar{\mathcal{B}}(j)| \\
&\geq |\bar{\mathcal{A}}_n|^{-1} \log(B^*)(|\bar{\mathcal{A}}_n| - \Delta_Y^n B^*). \qquad (4.66)
\end{aligned}
$$

Combining (4.54), (4.49) and the definition of B^*, we also have

$$
\Delta_Y^n B^* \leq \frac{1}{n}|\bar{\mathcal{A}}_n|. \qquad (4.67)
$$

Therefore, (4.66) can be written as

$$
\begin{aligned}
|\bar{\mathcal{A}}_n|^{-1} \sum_{(i,j)\in\bar{\mathcal{A}}_n} \log|\bar{\mathcal{B}}(j)| &\geq |\bar{\mathcal{A}}_n|^{-1} \log(B^*) \left(|\bar{\mathcal{A}}_n| - \frac{1}{n}|\bar{\mathcal{A}}_n|\right) \qquad (4.68) \\
&= (1 - \frac{1}{n}) \log\left(\frac{2^{-2n\delta_n}}{n} \frac{\Delta_X^n}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} (\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma - \delta)})^k\right).
\end{aligned}
$$

Using (4.64) in the above we have

$$
\begin{aligned}
\log \Delta_X^n \leq \frac{n}{n-1}\left(\sum_{t=1}^{n} H(\bar{X}_t|\bar{Y}_t) + 3|\mathcal{X}|n^{1/2}\right) + 2n\delta_n + \log n + |\mathcal{X}||\mathcal{Y}|\log(n+1) \\
+ k\log(\frac{|\mathcal{X}||\mathcal{Y}|2\sigma}{\delta}) \quad (4.69)
\end{aligned}
$$

Analogously,

$$
\begin{aligned}
\log \Delta_Y^n \leq \frac{n}{n-1}\left(\sum_{t=1}^{n} H(\bar{Y}_t|\bar{X}_t) + 3|\mathcal{Y}|n^{1/2}\right) + 2n\delta_n + \log n + |\mathcal{X}||\mathcal{Y}|\log(n+1) \\
+ k\log(\frac{|\mathcal{X}||\mathcal{Y}|2\sigma}{\delta}) \quad (4.70)
\end{aligned}
$$

Next, we find an upper bound for $\log \Delta_X^n \Delta_Y^n$. From (4.54), we get

$$\begin{aligned}
\log |\bar{\mathcal{A}}_n| &\geq \log |\mathcal{A}_n| + k \log(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma - \delta)}) \\
&\geq \log |\mathcal{A}_n| + k \log(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|2\sigma}) \\
&= \log |\mathcal{A}_n| - k \log(\frac{2\sigma}{\delta}) - k \log(|\mathcal{X}||\mathcal{Y}|) \\
&\overset{(a)}{\geq} \log(\Delta_X^n \Delta_Y^n) - |\mathcal{X}||\mathcal{Y}| \log(n+1) - 2n\delta_n - k \log(\frac{|\mathcal{X}||\mathcal{Y}|2\sigma}{\delta}), \qquad (4.71)
\end{aligned}$$

where (a) is obtained by using (4.49). Using (4.63), the above inequality becomes

$$\log(\Delta_X^n \Delta_Y^n) \leq \sum_{t=1}^{n} H(\bar{X}_t \bar{Y}_t) + 3|\mathcal{X}||\mathcal{Y}|n^{1/2} + |\mathcal{X}||\mathcal{Y}| \log(n+1) + 2n\delta_n + k \log(\frac{2\sigma}{\delta})$$
$$+ k \log(|\mathcal{X}||\mathcal{Y}|). \quad (4.72)$$

Using the lower bounds on the sizes of Δ_X, Δ_Y from 4.3.2, we can rewrite (4.69),(4.70) and (4.72) as

$$\begin{aligned}
R_X - \delta_n &\leq \frac{1}{n-1} \sum_{t=1}^{n} H(\bar{X}_t|\bar{Y}_t) + 3|\mathcal{X}|\frac{n^{1/2}}{n-1} + 2\delta_n + \frac{\log n + |\mathcal{X}||\mathcal{Y}| \log(n+1)}{n} \\
&\quad + \frac{k}{n} \log(\frac{2|\mathcal{X}||\mathcal{Y}|\sigma}{\delta}), \qquad (4.73\text{a}) \\
R_Y - \delta_n &\leq \frac{1}{n-1} \sum_{t=1}^{n} H(\bar{Y}_t|\bar{X}_t) + 3|\mathcal{Y}|\frac{n^{1/2}}{n-1} + 2\delta_n + \frac{\log n + |\mathcal{X}||\mathcal{Y}| \log(n+1)}{n} \\
&\quad + \frac{k}{n} \log(\frac{2|\mathcal{X}||\mathcal{Y}|\sigma}{\delta}), \qquad (4.73\text{b}) \\
R_X + R_Y - 2\delta_n &\leq \frac{1}{n} \sum_{t=1}^{n} H(\bar{X}_t \bar{Y}_t) + 3|\mathcal{X}||\mathcal{Y}|\frac{n^{1/2}}{n-1} + |\mathcal{X}||\mathcal{Y}|\frac{\log(n+1)}{n} + 2\delta_n \\
&\quad + \frac{k}{n} \log(\frac{2|\mathcal{X}||\mathcal{Y}|\sigma}{\delta}). \qquad (4.73\text{c})
\end{aligned}$$

For our proof, we would like all the terms on the right hand side of the above equations (except the entropies) to converge to 0 as $n \to \infty$. This will happen if

$$\frac{k}{n} \log(\frac{2\sigma}{\delta}) \to 0.$$

Recall from Lemma 4.4.1 that $\sigma = 2n\delta_n + |\mathcal{X}||\mathcal{Y}| \log(n+1)$ and $k < \frac{2\sigma}{\delta}$. Hence we need to choose δ such that

$$\frac{2\sigma}{n\delta} \log(\frac{2\sigma}{\delta}) \sim \frac{\delta_n + \frac{\log n}{n}}{\delta} (\log(n\delta_n + \log n) - \log \delta) \to 0. \tag{4.74}$$

From our assumption in the beginning, we have $\delta_n \log n \to 0$. By setting

$$\delta = (\delta_n \log n)^{1/2} \tag{4.75}$$

(4.74) becomes asymptotically equal to

$$\frac{\delta_n^{1/2}}{(\log n)^{1/2}} \left[\log(n\delta_n + \log n) - \log(\delta_n^{1/2}) - \log\log n\right]. \tag{4.76}$$

We separately consider each of the terms in the equation above

1. If $\log(n\delta_n + \log n) \sim \log(n\delta_n)$ for large n, then

$$\begin{aligned} \frac{\delta_n^{1/2}}{(\log n)^{1/2}} \log(n\delta_n + \log n) \sim \frac{\delta_n^{1/2}}{(\log n)^{1/2}} \log(n\delta_n) &= \frac{\delta_n^{1/2}}{(\log n)^{1/2}} [\log n + \log \delta_n] \\ &= (\delta_n \log n)^{1/2} + \frac{\delta_n^{1/2} \log \delta_n}{(\log n)^{1/2}} \to 0, \text{ since } \delta_n \to 0. \end{aligned} \tag{4.77}$$

   If $\log(n\delta_n + \log n) \sim \log(\log n)$ for large n, then

$$\frac{\delta_n^{1/2}}{(\log n)^{1/2}} \log(n\delta_n + \log n) \sim \frac{\delta_n^{1/2}}{(\log n)^{1/2}} \log(\log n) \to 0. \tag{4.78}$$

2. $\frac{\delta_n^{1/2}}{(\log n)^{1/2}} \log(\delta_n^{1/2}) \to 0$ because $x \log x \to 0$ when $x \to 0$.

3. $\frac{\delta_n^{1/2}}{(\log n)^{1/2}} \log\log n = (\delta_n \log n)^{1/2} \frac{\log\log n}{\log n} \to 0$.

Hence, the term in (4.76) converges to 0 as $n \to \infty$, completing the proof of the lemma. $\square$

We can rewrite Lemma 4.4.2 using new variables $\bar{X}, \bar{Y}, Q$, where $Q = t \in \{1, ..., n\}$ with probability $\frac{1}{n}$ and $P_{\bar{X}\bar{Y}|Q=t} = P_{\bar{X}_t\bar{Y}_t}$. Therefore, we now have (for all sufficiently large n),

$$
\begin{aligned}
R_X &\leq H(\bar{X}|\bar{Y}, Q) + \delta_{1n} && \text{(4.79a)}\\
R_Y &\leq H(\bar{Y}|\bar{X}, Q) + \delta_{2n} && \text{(4.79b)}\\
R_X + R_Y &\leq H(\bar{X}, \bar{Y}|Q) + \delta_{3n}, && \text{(4.79c)}
\end{aligned}
$$

for some $\delta_{1n}, \delta_{2n}, \delta_{3n} \to 0$.
Finally, using (4.57), we also have

$$
\begin{aligned}
&|\mathbb{P}(\bar{X} = x, \bar{Y} = y|Q = t) - \mathbb{P}(\bar{X} = x|Q = t)\mathbb{P}(\bar{Y} = y|Q = t)| \\
&= |\mathbb{P}(\bar{X}_t = x, \bar{Y}_t = y) - \mathbb{P}(\bar{X}_t = x)\mathbb{P}(\bar{Y}_t = y)| \\
&\leq 2\delta^{1/2} = 2(\delta_n \log n)^{1/4} \to 0 \text{ as } n \to \infty.
\end{aligned}
\tag{4.80}
$$

In other words, for all t, $\bar{X}_t, \bar{Y}_t$ are almost independent for large n. Consequently, using the continuity of mutual information with respect to the joint distribution, Lemma 4.4.2 holds with for any joint distribution $P_Q P_{\bar{X}|Q} P_{\bar{Y}|Q}$ such that the marginal on $(\bar{X}, \bar{Y})$ is $P_{\bar{X}\bar{Y}}$. Recall that $P_{\bar{X}\bar{Y}}$ is the dominant joint type that is λ_n-close to P_{XY}. Using suitable continuity arguments, we can now argue that Lemma 4.4.2 holds for any joint distribution $P_Q P_{X|Q} P_{Y|Q}$ such that the marginal on (X, Y) is P_{XY}, completing the proof of the converse. □

Proof. **(Theorem 4.3.3)** Let $X^n(i)$ and $Y^n(j)$ denote the ith and jth codewords in the random codebooks C_X and C_Y respectively. For $1 \leq i \leq 2^{nR_1}$ and $1 \leq j \leq 2^{nR_2}$, define the indicator random variables

$$
U_{ij} \triangleq \begin{cases} 1 & \text{if } (X^n(i), Y^n(j)) \in T^n_{\lambda_n}(X, Y) \\ 0 & \text{else} \end{cases}
\tag{4.81}
$$

The cardinality of the set $\mathcal{U}$ is then

$$U = \sum_{i=1}^{2^{nR_1}} \sum_{j=1}^{2^{nR_2}} U_{ij} \tag{4.82}$$

We derive upper bounds on the probability of the lower tail of U using Suen's inequality. To do this, we first set up the dependency graph of the indicator random variables U_{ij}. The vertex set of the graph is indexed by the ordered pair $(i,j), 1 \leq i \leq 2^{nR_1}, 1 \leq j \leq 2^{nR_2}$. From the nature of the random experiment, it is clear that the indicator random variables U_{ij} and $U_{i'j'}$ are independent if and only if $i \neq i'$ and $j \neq j'$. Thus, each vertex (i,j) is connected to exactly $2^{nR_1} + 2^{nR_2} - 2$ vertices of the form $(i,j'), j' \neq j$ or $(i',j), i' \neq i$. If vertices (i,j) and (k,l) are connected, we denote it by $(i,j) \sim (k,l)$.

In order to estimate Γ, Θ and θ as defined in Lemma 4.3.1, define the following quantities. Let $\alpha_{ij} \triangleq \mathbb{P}(U_{ij} = 1)$ and $\beta_{\{ij\}\{kl\}} \triangleq \mathbb{E}(U_{ij}U_{kl})$ where $(i,j) \sim (k,l)$. Using Fact 4.1.1 and Fact 4.1.2, uniform bounds can be derived for these quantities as

$$\alpha \triangleq 2^{-n(I(X;Y)+\epsilon_{3n})} \leq \alpha_{ij} \leq 2^{-n(I(X;Y)-\epsilon_{3n})} \triangleq \alpha^{'} \tag{4.83}$$

where ϵ_{3n} is a continuous positive function of $\epsilon_{1n}, \epsilon_{2n}$ and λ_n that goes to 0 as $n \to \infty$. Similarly, a uniform bound on $\beta_{\{ij\}\{kl\}}$ can be derived as

$$2^{-2n(I(X;Y)+2\epsilon_{4n})} \leq \beta_{\{ij\}\{kl\}} \leq 2^{-2n(I(X;Y)-2\epsilon_{4n})} \triangleq \beta \tag{4.84}$$

where ϵ_{4n} is a continuous positive function of $\epsilon_{1n}, \epsilon_{2n}$ and λ_n that goes to 0 as $n \to \infty$.

The quantities involved in Suen's inequality can now be estimated.

$$\Gamma \triangleq \mathbb{E}(U) \geq 2^{n(R_1+R_2)}\alpha \tag{4.85}$$

$$\Theta \triangleq \frac{1}{2} \sum_{(i,j)} \sum_{(k,l)\sim(i,j)} \mathbb{E}(U_{ij}U_{kl}) \leq \frac{1}{2} 2^{n(R_1+R_2)}(2^{nR_1} + 2^{nR_2} - 2)\beta \tag{4.86}$$

$$\theta \triangleq \max_{(i,j)} \sum_{(k,l)\sim(i,j)} \mathbb{E}(U_{kl}) \leq (2^{nR_1} + 2^{nR_2} - 2)\alpha^{'} \tag{4.87}$$

Substituting these bounds into equations (4.18) and (4.17) proves the claims made in equations (4.23) and (4.24) of Theorem 4.3.3.

A lower bound on the probability of the empty induced random subgraph can be derived by employing the Lovasz local lemma on the $2^{n(R_1+R_2)}$ events $\{U_{ij} = 1\}, 1 \leq i \leq 2^{nR_1}, 1 \leq j \leq 2^{nR_2}$. Symmetry considerations imply that all x_i can be set identically to x in Lemma 4.3.2. Then, the Lovasz lemma states that if there exists $x \in [0,1]$ such that $\alpha \leq \mathbb{P}(U_{ij} = 1) \leq x(1-x)^{(2^{nR_1}+2^{nR_2}-2)}$, then $\mathbb{P}(U = 0) \geq (1-x)^{2^{n(R_1+R_2)}}$. It is easy to verify that for such an x to exist, we need $R_2 \leq R_1 < I(X;Y)$ and if so, $x = 2^{-nR_1}$ satisfies the condition. Therefore, we have

$$\mathbb{P}(U=0) \geq \exp\left(-\left(2^{nR_2}+1\right)\right) \qquad R_2 \leq R_1 < I(X;Y) \tag{4.88}$$

We can derive a similar bound using the second version of the local lemma given in Lemma 4.3.2. While Γ and θ are same as estimated earlier, $\tau = \max_{(i,j)} \mathbb{P}(U_{ij} = 1)$ is upper bounded by $\alpha^{'}$ as defined in equation (4.83). Hence,

$$\mathbb{P}(U=0) \geq \exp\left(-\Gamma\phi(\theta+\tau)\right). \tag{4.89}$$

Under the same assumption $R_2 \leq R_1 < I(X;Y)$, $\theta + \tau \leq (2^{nR_1} + 2^{nR_2} - 2)\alpha^{'} \to 0$ as $n \to \infty$ and hence $\phi(\theta+\tau) \to 1$. Combining equations (4.88) and (4.89), taking logarithms and letting $n \to \infty$, we get

$$\lim_{n\to\infty} \frac{1}{n} \log\log \frac{1}{P(U=0)} \leq \min\left(R_2, R_1 + R_2 - I(X;Y)\right). \tag{4.90}$$

Comparing this to equation (4.24) shows that this expression is asymptotically tight in the regime $R_2 \leq R_1 < I(X;Y)$. $\square$

CHAPTER 5

Upper Bounds on the Error Exponent of Multiple-Access Channels

In this chapter, we develop two new upper bounds on the reliability function of DM-MACs. Towards this goal, we first revisit the point-to-point case and examine the techniques used for obtaining the upper bounds on the optimum error exponent. The techniques employed to obtain the sphere packing bound can be broadly classified into three categories. The first is known as the Gallager technique [28]. Although this yields expressions for the error exponents that are computationally easier to evaluate than others, the expressions themselves are much more difficult to interpret. The Method of Types technique, introduced by Csiszar [14], comprises the second category. This technique uses more intuitive expressions for the error exponents in terms of the optimization of an objective function involving information quantities over probability distributions. It results in a sphere packing bound for the average probability of error and is more amenable for multi-user channels. The third category consists of the Strong Converse technique, introduced by Csiszar-Korner [16]. This technique results in an expression identical to the result of the Method of Types technique. The only difference between the two is that the third technique results in a sphere packing bound for the maximal probability of error, and not the average. However, in point to point scenario, by purging the worst half of the codewords in any codebook, it can be easily shown that the average and maximal performance are the same at any transmission rate. As a result, the sphere packing bound derived by using the strong converse technique is as strong as the one obtained by using the method of types technique.

In developing our first sphere packing bound for multiple-access channels, we use a technique very similar to the method of types technique. We start by partitioning the error event into its intersection with disjoint type classes. Following, the error probability of the code can be obtained by adding up the probabilities of these intersections. By deriving lower bounds on the probability of these sets, a lower bound on the average probability of error can be obtained. The result of this step is a sphere packing bound which is identical to the well-known sphere packing bound derived by Haroutunian [33]. Our approach provides more intuition than Haroutunian's result. Based on this intuition, and using some properties of typicality graphs, we can obtain a sphere packing bound outperforming Haroutunian's result especially at high rates.

In developing the second sphere packing bound for DM-MACs, we introduce a new technique for deriving the sphere packing exponent for point-to-point channels by using a strong converse theorem for codes with a specified dominant composition. The new converse theorem not only determines a lower bound on the error probability of an individual codeword, but also provides a lower bound on the number of codewords with that error probability. Using this converse theorem, we directly derive the well known sphere packing bound for the average probability of error for DMCs without the elimination of codewords as the final step. Toward extending this technique to MACs, we start by deriving a strong converse theorem for codes with a particular input joint empirical distribution. By using this theorem and the technique developed for point-to-point channels [16], we develop a tighter sphere packing bound for the average error exponent of DM-MACs.

Next, we derive a new upper bound on the maximal error exponent for multiple-access channels by studying the Bhattacharyya distance distribution of multi-user codes. This bound, called the minimum distance bound, is deriven by establishing a link between the minimum Bhattacharyya distance and maximal probability of decoding error; the upper bound on the Bhattacharyya distance can then be used to infer the lower bound on the probability of decoding error. At zero rate pair, this upper bound has a similar structure to the partial expurgated bound derived in Chapter 3. However, the two bounds are not necessarily equal. By using a conjecture about the structure of typicality graphs, we derive a tighter conjectured minimum distance bound for the maximal error exponent. Later on in this chapter, we study the relationship between average and maximal error probabilities for a two user (DM) MAC and develop a method to obtain new bounds on the average/maximal error exponent by using known bounds on the maximal/average error exponent. It is observed that at zero rate, the bounds on average error exponent are valid bounds on the maximal error exponent and vice versa. As a result, the comparison between the conjectured minimum distance bound and the expurgated bound is indeed a valid comparison at zero rate. By comparing these bounds at zero rate, it is shown that the expurgated and the conjectured minimum distance bound are tight bounds at rate zero.

The chapter is organized as follows: Some preliminaries are introduced in Section 5.1. The two sphere packing bounds on the average probability of error for DM-MACs are studied in Section 5.2 and 5.3. Another central result of this chapter is a minimum distance bound for the maximal error exponent for MAC, obtained in Section 5.4. In Section 5.4.1, by using a conjecture about the structure of the typicality graph, a tighter minimum distance bound is derived and shown to be tight at zero rate. In Section 5.5, by using a known upper bound on the maximum error exponent function, we derive an upper bound on the average error exponent function and vice versa. The proof of some of these results are given in Section 5.6 .

5.1 Preliminaries

Definition 5.1.1. *For a multiuser code $C = C_X \times C_Y$ with codewords of length n, and for any joint composition $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$, we define the P_{XY}-rate of C as*

$$R(C, P_{XY}) \triangleq \frac{1}{n} \log |C \cap T_{P_{XY}}|. \tag{5.1}$$

Definition 5.1.2. *For a specified channel $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, the Bhattacharyya distance between the channel input letter pairs (x, y), and $(\tilde{x}, \tilde{y})$ is defined by*

$$d_B\big((x,y),(\tilde{x},\tilde{y})\big) \triangleq -\log\big(\sum_{z\in\mathcal{Z}} \sqrt{W(z|x,y)W(z|\tilde{x},\tilde{y})}\big). \tag{5.2}$$

A channel for which $d_B\big((x,y),(\tilde{x},\tilde{y})\big) \neq \infty$ for all (x,y) and $(\tilde{x},\tilde{y})$, is called as an indivisible channel. *An indivisible channel for which the matrix $A_{(i,j),(k,l)} = 2^{sd_B\big((i,j),(k,l)\big)}$ is nonnegative-definite for all $s > 0$ is called a* nonnegative-definite *channel.*

For a block channel W^n, the normalized Bhattacharyya distance between two channel input block pairs $(\mathbf{x},\mathbf{y})$, and $(\tilde{\mathbf{x}},\tilde{\mathbf{y}})$ is given by:

$$d_B\big((\mathbf{x},\mathbf{y}),(\tilde{\mathbf{x}},\tilde{\mathbf{y}})\big) = -\frac{1}{n}\log\big(\sum_{\mathbf{z}\in\mathcal{Z}^n} \sqrt{W^n(\mathbf{z}|\mathbf{x},\mathbf{y})W^n(\mathbf{z}|\tilde{\mathbf{x}},\tilde{\mathbf{y}})}\big). \tag{5.3}$$

If W^n is a memoryless channel, it can be easily shown that the Bhattacharyya distance between two pairs of codewords $(\mathbf{x},\mathbf{y})$ and $(\tilde{\mathbf{x}},\tilde{\mathbf{y}})$, with joint empirical distribution $P_{XY\tilde{X}\tilde{Y}}$, is

$$d_B\big((\mathbf{x},\mathbf{y}),(\tilde{\mathbf{x}},\tilde{\mathbf{y}})\big) = \sum_{\substack{x,\tilde{x}\in\mathcal{X}\\ y,\tilde{y}\in\mathcal{Y}}} P_{XY\tilde{X}\tilde{Y}}(x,y,\tilde{x},\tilde{y}) d_B\big((x,y),(\tilde{x},\tilde{y})\big). \tag{5.4}$$

As it can be seen from (5.4)*, for a fixed channel, the Bhattacharyya distance between two pairs of codewords depends only on their joint composition. The minimum Bhattacharyya distance for a code C is defined as:*

$$d_B(C) \triangleq \min_{\substack{(\mathbf{x},\mathbf{y}),(\tilde{\mathbf{x}},\tilde{\mathbf{y}})\in C\\ (\mathbf{x},\mathbf{y})\neq(\tilde{\mathbf{x}},\tilde{\mathbf{y}})}} d_B\big((\mathbf{x},\mathbf{y}),(\tilde{\mathbf{x}},\tilde{\mathbf{y}})\big). \tag{5.5}$$

For a fixed rate pair (R_X, R_Y) and blocklength n, the best possible minimum distance is defined as

$$d_B^*\big(R_X, R_Y, n\big) \triangleq \max_C d_B(C), \tag{5.6}$$

where the maximum is over all multi user codes with parameters $(n, 2^{nR_X}, 2^{nR_Y})$.

Finally, we define

$$d_B^*(R_X, R_Y) \triangleq \limsup_{n\to\infty} d_B^*\big(R_X, R_Y, n\big). \tag{5.7}$$

Note that since any multi-user code with repeated codewords has at least two identical codeword pairs, it can be concluded that the minimum distance for such a code is equal to zero. Therefore, in order to find an upper bound for the best possible minimum distance, $d_B^*(R_X, R_Y)$, we only need to concentrate on codes without repetition.

For a fixed joint composition $P_{XY} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y})$, using the structure of Bhattacharyya distance function, we can define spheres in $T_{P_{XY}}$. For any $(\mathbf{x}, \mathbf{y}) \in T_{P_{XY}}$, the sphere about $(\mathbf{x}, \mathbf{y})$, of radius r, is given by

$$S \triangleq \left\{(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) : d_B\big((\mathbf{x}, \mathbf{y}), (\tilde{\mathbf{x}}, \tilde{\mathbf{y}})\big) \le r\right\}.$$

Every point, $(\mathbf{x}, \mathbf{y}) \in T_{P_{XY}}$, is surrounded by a set consisting of all pairs with which it shares some given joint type $V_{XY\tilde{X}\tilde{Y}}$. Basically, any pair of sequences, $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \in T_{P_{XY}}$, sharing a common joint type with some given pair of sequences, $(\mathbf{x}, \mathbf{y}) \in T_{P_{XY}}$, belongs to the surface of a sphere with center $(\mathbf{x}, \mathbf{y})$ and radius $r = d_B\big((\mathbf{x}, \mathbf{y}), (\tilde{\mathbf{x}}, \tilde{\mathbf{y}})\big)$. The set of these pairs is called a spherical collection about $(\mathbf{x}, \mathbf{y})$ defined by $P_{\mathbf{x},\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{y}}}$.

5.2 Sphere Packing Bound on the Average Error Exponent (Method of Types Technique)

The focal point of this section is an upper (*sphere packing*) bound for the average error exponent for discrete memoryless multiple access channels. To obtain this bound, we use the method of types. Using the method of types, Csiszar [14] derived a sphere packing bound for the average error exponent of discrete memoryless channels. The idea behind the method of types is to partition the n-length sequences into classes according to their empirical distribution. In [14], the average error probability of the code is partitioned into its intersection with the type classes, and the probability of error is obtained by adding up the probabilities of the intersections. Since the number of type classes grows polynomially with n, the average probability of error has the same exponent as the largest among the probabilities of the above intersections. The second key idea of the method of types is that sequences of the same type have the same probability under a memoryless channel. Therefore, to bound the probabilities of intersections, it is sufficient to bound their cardinalities. Toward extending this technique to MACs, we follow a two-step approach. First, we derive a lower bound on the average error probability for a multi-user code with a specified dominant joint composition. Since the dominant joint type for an arbitrary two-user code is unknown, to obtain a lower bound on the error probability of the best code, we need to minimize the aforementioned lower bound over all possible joint input distributions. The result of this step is a sphere packing bound for the average error exponent identical to the Haroutunian's result [33]. As the second step, we use the properties of typicality graphs to restrict the set of possible dominant joint compositions. Since the minimization is taken over a smaller set, the new sphere packing bound is tighter than Haroutunian's result.

Theorem 5.2.1. *For any $R_X, R_Y \geq 0$, $\delta > 0$ and any DM-MAC, $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, every (n, M_X, M_Y) code, C, with a dominant type $P^n_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ and rate pair satisfying*

$$\frac{1}{n} \log M_X \geq R_X + \delta \tag{5.8a}$$

$$\frac{1}{n} \log M_Y \geq R_Y + \delta, \tag{5.8b}$$

has average probability of error

$$e(C, W) \geq \frac{1}{2} e^{-n[E^T_{sp}\left(R_X, R_Y, W, P^n_{XY}\right) + \delta]}, \tag{5.9}$$

where

$$E^T_{sp}\left(R_X, R_Y, W, P^n_{XY}\right) \triangleq \min_{V_{Z|XY}} D\left(V_{Z|XY} || W | P^n_{XY}\right). \tag{5.10}$$

Here, the minimization is over all possible conditional distributions $V_{Z|XY} : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$, which satisfy at least one of the following conditions

$$I_V(X \wedge Z|Y) \leq R_X \tag{5.11a}$$

$$I_V(Y \wedge Z|X) \leq R_Y \tag{5.11b}$$

$$I_V(XY \wedge Z) \leq R_X + R_Y, \tag{5.11c}$$

and all mutual informations are calculated based on $P^n_{XY}(x,y)V_{Z|XY}(z|x,y)$.

Proof. The proof is provided in Section 5.6.1. □

In Theorem 5.2.1, we have obtained a sphere packing bound on the average error exponent for a multiuser code with a certain dominant type. For an arbitrary code, the dominant joint type in unknown. However, using the properties of the typicality graph obtained in Chapter 4, the necessary and sufficient condition for a joint type to be a dominant type of a code with certain parameters is known. By combining the result of theorem 5.2.1 and the result of Chapter 4, we can obtain the following sphere packing bound for any multiuser code:

Theorem 5.2.2. *For any $R_X, R_Y \geq 0$, $\delta > 0$ and any DM-MAC, $W : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$, every (n, M_X, M_Y) code, C, with*

$$\frac{1}{n}\log M_X \geq R_X + \delta \tag{5.12a}$$

$$\frac{1}{n}\log M_Y \geq R_Y + \delta, \tag{5.12b}$$

has average probability of error

$$e(C,W) \geq \frac{1}{2}e^{-n[E^T_{sp}(R_X,R_Y,W)+\delta]}, \tag{5.13}$$

where

$$E^T_{sp}(R_X,R_Y,W) \triangleq \max_{P_{XY}\in\mathcal{B}(R_X,R_Y)} E^T_{sp}(R_X,R_Y,W,P_{XY}). \tag{5.14}$$

where $\mathcal{B}(R_X, R_Y)$ is defined as follows:

$$\mathcal{B}(R_X, R_Y) \triangleq \left\{ \begin{array}{c} P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \ R_X \leq H\left(X|U\right), \ R_Y \leq H\left(Y|U\right) \\ X - U - Y \quad \text{for some } U \in \{1, 2, 3, 4\} \end{array} \right\} \tag{5.15}$$

5.3 Sphere Packing Bound on the Average Error Exponent (Strong Converse Technique)

5.3.1 Point to Point Case

The main result of this section is an upper (*sphere packing*) bound for the average error exponent for discrete memoryless channels. By using the strong converse theorem for DMCs and applying the method of types idea, the authors in [16] derived a sphere packing bound on the maximal error exponent. For point to point transmission systems, it is unimportant whether we work with average or maximal errors. As a result, for all transmission rates, the sphere packing bound of [16] is an upper bound on the average error exponent of DMCs. In this section, we use an approach very similar to [16]. First, we obtain a strong converse theorem for codes with a specified good dominant composition, meaning most of the codewords with this dominant composition have small error probability. This strong converse theorem is a generalized version of the well-known converse theorem for discrete memoryless channels in the sense that it not only determines a lower bound for the error probability of the individual codewords, but also provides a lower bound on the number of codewords with that error probability. Since we are using a stronger converse theorem, we can obtain a sphere packing bound on the average probability of error without expurgating any codeword.

Definition 5.3.1. *For any discrete memoryless channel, W, for any joint distribution $P \in \mathcal{P}(\mathcal{X})$, any $0 \leq \lambda < 1$, and any (n, M) code, C, define*

$$\mathcal{E}_W(C, P, \lambda) \triangleq \left\{ \mathbf{x}_i \in C : W^n(D_i|\mathbf{x}_i) \geq \frac{1-\lambda}{2}, \mathbf{x}_i \in T_P \right\}. \tag{5.16}$$

Theorem 5.3.1. *Consider any* (n, M) *code* C. *For every* $P^* \in \mathcal{P}_n(\mathcal{X})$ *and every* $0 \le \lambda < 1$, *the condition* $|\mathcal{E}_W(C, P^*, \lambda)| \ge \frac{1}{(n+1)^{|\mathcal{X}|}}\left(1 - \frac{2\lambda}{1+\lambda}\right) M$ *implies*

$$\frac{1}{n}\log M \le I(P^*, W) + \epsilon_n(\lambda, |\mathcal{X}|). \tag{5.17}$$

Here, $\epsilon_n \to 0$ *as* $n \to \infty$.

Proof. The proof is provided in Section 5.6. □

Fact 5.3.1. ***(Sphere Packing Bound)*** *For any* $R \ge 0$, $\delta > 0$ *and any discrete memoryless channel,* $W : \mathcal{X} \to \mathcal{Z}$, *every* (n, M) *code,* C, *with*

$$\frac{1}{n}\log M \ge R + \delta \tag{5.18}$$

has average probability of error

$$e(C, W) \ge \frac{1}{2}e^{-n[E_{sp}(R,W)(1+\delta)+\delta]}, \tag{5.19}$$

where

$$E_{sp}(R, W) \triangleq \max_{P\in\mathcal{P}(\mathcal{X})} \min_{V:I(P,V)\le R} D(V||W|P). \tag{5.20}$$

Proof. The proof is provided in Section 5.6. □

5.3.2 MAC Case

The main result of this section is a lower (*sphere packing*) bound for the average error probability of a DM-MAC. To state the new bound we need an intermediate result that has the form of a strong converse for the MAC. We state this result here and relegate the proof to Section 5.6.

Definition 5.3.2. *For any DM-MAC,* W, *for any joint distribution* $P \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, *any* $0 \le \lambda < 1$, *and any* (n, M_X, M_Y) *code,* C, *define*

$$\mathcal{E}_W(C, P, \lambda) \triangleq \left\{(\mathbf{x}_i, \mathbf{y}_j) \in C : W(D_{ij}|\mathbf{x}_i, \mathbf{y}_j) \ge \frac{1-\lambda}{2}, (\mathbf{x}_i, \mathbf{y}_j) \in T_P\right\}. \tag{5.21}$$

Theorem 5.3.2. *Fix $0 \leq \lambda < 1$. Consider any (n, M_X, M_Y) code C. For every $P^n_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$, such that $|\mathcal{E}_W(C, P^n_{XY}, \lambda)| \geq \frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} \left(1 - \frac{2\lambda}{1+\lambda}\right) M_X M_Y$, the following holds*

$$\left(\frac{1}{n}\log M_X, \frac{1}{n}\log M_Y\right) \in C^n_W\left(P^n_{XY}\right) \tag{5.22}$$

where $C^n_W(P)$ is defined as the closure of the set of all (R_1, R_2) pairs satisfying

$$R_X \leq I(X \wedge Z|YU) + \epsilon_n, \tag{5.23a}$$

$$R_Y \leq I(Y \wedge Z|XU) + \epsilon_n, \tag{5.23b}$$

$$R_X + R_Y \leq I(XY \wedge Z|U) + \epsilon_n, \tag{5.23c}$$

for some choice of random variables U defined on $\{1, 2, 3, 4\}$, and joint distribution $p(u)p(x|u)p(y|u)W(z|x, y)$, with marginal distribution $p(x, y) = P^n(x, y)$. Here, $\epsilon_n \to 0$ as $n \to \infty$.

Proof. The proof is provided in Section 5.6. □

We further define $C_W(P)$ $\big($the *limiting* version of the sets $C^n_W(P)\big)$ as the closure of the set of all (R_X, R_Y) pairs satisfying

$$R_X \leq I(X \wedge Z|YU), \tag{5.24a}$$

$$R_Y \leq I(Y \wedge Z|XU), \tag{5.24b}$$

$$R_X + R_Y \leq I(XY \wedge Z|U), \tag{5.24c}$$

for some choice of random variables U defined on $\{1, 2, 3, 4\}$, and joint distribution $p(u)p(x|u)p(y|u)W(z|x, y)$, with marginal distribution $p(x, y) = P(x, y)$.

Theorem 5.3.3. ***(Sphere Packing Bound)*** *For any $R_X, R_Y \geq 0$, $\delta > 0$ and any*

DM-MAC, $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, every (n, M_X, M_Y) code, C, with

$$\frac{1}{n} \log M_X \geq R_X + \delta \tag{5.25a}$$

$$\frac{1}{n} \log M_Y \geq R_Y + \delta, \tag{5.25b}$$

has average probability of error

$$e(C, W) \geq \frac{1}{2} e^{-n[E_{sp}(R_X, R_Y, W)(1+\delta)+\delta]}, \tag{5.26}$$

where

$$E_{sp}(R_X, R_Y, W) \triangleq \max_{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})} \min_{V : (R_X, R_Y) \notin C_V(P_{XY})} D(V||W|P_{XY}). \tag{5.27}$$

Proof. The proof is provided in Section 5.6. □

5.4 A Minimum Distance on the Maximal Error Exponent

In this section, we present an upper (minimum distance) bound for the maximal error exponent for a DM-MAC. The idea behind the derivation of this bound is the connection between the minimum distance of the code and the maximal probability of decoding error. Intuitively, the closer the codewords are, the more confusion exists in decoding. An arbitrary channel, $W(\cdot|\cdot,\cdot)$, is used to define the Bhattacharyya distance. To derive an upper bound on the error exponent at rate (R_X, R_Y), we need to show that for any code with parameter (R_X, R_Y), there exist at least two pairs of codewords which are very close to each other in terms of Bhattacharyya distance. In other words, we need to find an upper bound on the minimum distance of codes with parameter $\left(n, 2^{nR_X}, 2^{nR_Y}\right)$. Consider any arbitrary multi-user code, C, with parameters $\left(n, 2^{nR_X}, 2^{nR_Y}\right)$ with a dominant joint type P_{XY}. We concentrate on the dominant subset corresponding P_{XY}, i.e. all codeword pairs sharing P_{XY} as their joint type. We study the minimum distance of this subset and in particular we prove that there exist at least two pairs of codewords at a certain Bhattacharrya distance. As a result, we find an upper bound for the minimum distance of this subset of the code. Clearly, this bound is still a valid upper bound for the minimum distance of the original multi user code. To obtain this upper bound, we show that there exist a spherical collection about a pair of sequences, not necessarily codeword pairs, with exponentially many codeword pairs on it. Intuitively, since exponentially many codeword pairs are located on this spherical collection, all of these pairs cannot be far from each other. We study the distance structure of this collection, and find the average distance of this subset. It can be concluded that there must exist at least two pairs of codewords with distance at most as large as the average distance previously found. Next, by relating the maximal error probability of code to its minimum distance, we derive a lower bound on the maximal error probability of any multiuser code satisfying some rate constraints.

In Theorem 5.6.1, we derive an upper bound on the minimum distance of all multi user codes with certain rate pair. In Theorem 5.6.2, we show the connection between the maximal probability of error to the upper bound we derive in Theorem 5.6.1. Finally, by combining these results, in the following theorem, we end up with the main result of this section.

Theorem 5.4.1. *For any indivisible nonnegative-definite channel, W, the maximal error reliability function, $E_m^*(R_X, R_Y)$, satisfies*

$$E_m^*(R_X, R_Y) \leq E_U(R_X, R_Y, W). \tag{5.28}$$

where $E_U(R_X, R_Y, W)$ is defined as

$$E_U(R_X, R_Y, W) \triangleq \max_{P_{UXY}} \min_{\beta = X,Y,XY} E_U^{\beta}(R_X, R_Y, W, P_{XYU}). \tag{5.29}$$

The maximum is taken over all $P_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y})$ such that $X - U - Y$, and $R_X \leq H(X|U)$ and $R_Y \leq H(Y|U)$. The functions $E_U^{\beta}(R_X, R_Y, W, P_{XYU})$ are defined *as follows:*

$$
\begin{aligned}
E_U^X(R_X, R_Y, W, P_{XYU}) &\triangleq \min_{V_{X\tilde{X}\hat{X}YZ} \in \mathcal{V}_X^U} D(V_{Z|\tilde{X}Y}||W|P_{XY}) + I(\hat{X} \wedge Z|\tilde{X}Y),\\
E_U^Y(R_X, R_Y, W, P_{XYU}) &\triangleq \min_{V_{XY\tilde{Y}\hat{Y}Z} \in \mathcal{V}_Y^U} D(V_{Z|X\tilde{Y}}||W|P_{XY}) + I(\hat{Y} \wedge Z|X\tilde{Y}),\\
E_U^{XY}(R_X, R_Y, W, P_{XYU}) &\triangleq \min_{V_{XY\tilde{X}\tilde{Y}\hat{X}\hat{Y}Z} \in \mathcal{V}_{XY}^U} D(V_{Z|\tilde{X}\tilde{Y}}||W|P_{XY}) + I(\hat{X}\hat{Y} \wedge Z|\tilde{X}\tilde{Y}).
\end{aligned}
\tag{5.30}
$$

where

$$
\begin{aligned}
\mathcal{V}_X^U \triangleq \Big\{ & V_{X\tilde{X}\hat{X}YZ} : V_{\tilde{X}Y} = V_{\hat{X}Y} = V_{XY} = P_{XY},\ \hat{X} - XY - \tilde{X}\\
& V_{\tilde{X}|XY} = V_{\hat{X}|XY},\ I(X \wedge \tilde{X}|Y) = I(X \wedge \hat{X}|Y) \leq R_X,\\
& \alpha(V_{\hat{X}YZ}) < \alpha(V_{\tilde{X}YZ}) \Big\},
\end{aligned}
\tag{5.31}
$$

$$
\begin{aligned}
\mathcal{V}_Y^U \triangleq \Big\{ & V_{XY\tilde{Y}\hat{Y}Z} : V_{X\tilde{Y}} = V_{X\hat{Y}} = V_{XY} = P_{XY},\ \hat{Y} - XY - \tilde{Y}\\
& V_{\tilde{Y}|XY} = V_{\hat{Y}|XY},\ I(Y \wedge \tilde{Y}|X) = I(Y \wedge \hat{Y}|X) \leq R_Y,\\
& \alpha(V_{X\hat{Y}Z}) < \alpha(V_{X\tilde{Y}Z}) \Big\},
\end{aligned}
\tag{5.32}
$$

$$
\begin{aligned}
\mathcal{V}_{XY}^U \triangleq \Big\{ & V_{XY\tilde{X}\tilde{Y}\hat{X}\hat{Y}Z} : V_{\tilde{X}\tilde{Y}} = V_{\hat{X}\hat{Y}} = V_{XY} = P_{XY},\ \hat{X}\hat{Y} - XY - \tilde{X}\tilde{Y}\\
& V_{\tilde{X}\tilde{Y}|XY} = V_{\hat{X}\hat{Y}|XY},\ I(XY \wedge \tilde{X}\tilde{Y}) = I(XY \wedge \hat{X}\hat{Y}) \leq R_X + R_Y,\\
& \alpha(V_{\hat{X}\hat{Y}Z}) < \alpha(V_{\tilde{X}\tilde{Y}Z}) \Big\}.
\end{aligned}
\tag{5.33}
$$

Proof. The proof is provided in Section 5.6. □

5.4.1 A Conjectured Tighter Upper Bound

Conjecture 5.4.1. *For all sequences of nearly complete subgraphs of a particular type graph $T_{P_{XY}}$, the rates of the subgraph (R_X, R_Y) satisfy*

$$
R_X \leq H(X|U),\ R_Y \leq H(Y|U) \tag{5.34}
$$

for some $P_{U|XY}$ such that $X - U - Y$. Moreover, there exists $\mathbf{u} \in T_{P_U}$ such that the intersection of the fully connected subgraph with $T_{P_{XY|U}}(\mathbf{u})$ has the rate (R_X, R_Y).

Based on the result of the previous lemma, and by following a similar argument as proof of Theorem 5.4.1 , we can conclude the following result:

Theorem 5.4.2. *For any indivisible nonnegative-definite channel, W, the maximal error reliability function, $E_m^*(R_X, R_Y)$, satisfies*

$$E_m^*(R_X, R_Y) \leq E_C(R_X, R_Y, W). \tag{5.35}$$

where $E_C(R_X, R_Y, W)$ is defined as

$$\max_{P_{UXY}} \min_{\beta = X,Y,XY} E_C^{\beta}(R_X, R_Y, W, P_{XYU}) \tag{5.36}$$

The maximum is taken over all $P_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y})$ such that $X - U - Y$, and $R_X \leq H(X|U)$ and $R_Y \leq H(Y|U)$. The functions $E_C^{\beta}(R_X, R_Y, W, P_{XYU})$ are defined as follows:

$$\begin{aligned}
E_C^X(R_X, R_Y, W, P_{XYU}) &\triangleq \min_{V_{X\tilde{X}\hat{X}YZ} \in \mathcal{V}_X^C} D(V_{Z|U\tilde{X}Y}||W|V_{U\tilde{X}Y}) + I(\hat{X} \wedge Z|U\tilde{X}Y), \\
E_C^Y(R_X, R_Y, W, P_{XYU}) &\triangleq \min_{V_{XY\tilde{Y}\hat{Y}Z} \in \mathcal{V}_Y^C} D(V_{Z|UX\tilde{Y}}||W|V_{UX\tilde{Y}}) + I(\hat{Y} \wedge Z|UX\tilde{Y}), \\
E_C^{XY}(R_X, R_Y, W, P_{XYU}) &\triangleq \min_{V_{XY\tilde{X}\tilde{Y}\hat{X}\hat{Y}Z} \in \mathcal{V}_{XY}^C} D(V_{Z|U\tilde{X}\tilde{Y}}||W|V_{U\tilde{X}\tilde{Y}}) + I(\hat{X}\hat{Y} \wedge Z|U\tilde{X}\tilde{Y}).
\end{aligned} \tag{5.37}$$

where

$$
\begin{aligned}
\mathcal{V}_X^C \triangleq \Big\{ & V_{UX\tilde{X}\hat{X}YZ} : V_{U\tilde{X}Y} = V_{U\hat{X}Y} = V_{UXY} = P_{UXY}, \\
& \hat{X} - UXY - \tilde{X} \ , \ V_{\tilde{X}|XYU} = V_{\hat{X}|XYU}, \\
& I(X \wedge \tilde{X}|YU) = I(X \wedge \hat{X}|YU) \leq R_X \ \ , \ \ \alpha(V_{U\hat{X}YZ}) < \alpha(V_{U\tilde{X}YZ}) \Big\} \qquad (5.38) \\
\mathcal{V}_Y^C \triangleq \Big\{ & V_{UXY\tilde{Y}\hat{Y}Z} : V_{UX\tilde{Y}} = V_{UX\hat{Y}} = V_{UXY} = P_{UXY}, \\
& \hat{Y} - UXY - \tilde{Y}, V_{\tilde{Y}|XYU} = V_{\hat{Y}|XYU}, \\
& I(Y \wedge \tilde{Y}|UX) = I(Y \wedge \hat{Y}|UX) \leq R_Y, \ \ \alpha(V_{UX\hat{Y}Z}) < \alpha(V_{UX\tilde{Y}Z}) \Big\} \qquad (5.39) \\
\mathcal{V}_{XY}^C \triangleq \Big\{ & V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}Z} : V_{U\tilde{X}\tilde{Y}} = V_{U\hat{X}\hat{Y}} = V_{UXY} = P_{UXY}, \\
& \hat{X}\hat{Y} - UXY - \tilde{X}\tilde{Y} V_{\tilde{X}\tilde{Y}|UXY} = V_{\hat{X}\hat{Y}|UXY}, \\
& I(XY \wedge \tilde{X}\tilde{Y}|U) = I(XY \wedge \hat{X}\hat{Y}|U) \leq R_X + R_Y, \ \ \alpha(V_{U\hat{X}\hat{Y}Z}) < \alpha(V_{U\tilde{X}\tilde{Y}Z}) \Big\}
\end{aligned} \tag{5.40}
$$

Let us focus on the case where both codebooks have rate zero, $R_X = R_Y = 0$. Any $V_{UX\tilde{X}\hat{X}Y} \in \mathcal{V}_X^C$ satisfies the following:

$$X - UY - \tilde{X}, \quad X - UY - \hat{X}, \tag{5.41}$$

therefore, any $V_{UX\tilde{X}\hat{X}YZ} \in \mathcal{V}_X^C$ can be written as

$$P_{X|U} P_{X|U} P_{X|U} P_{Y|U} P_U V_{Z|UXY\tilde{X}\hat{X}}. \tag{5.42}$$

Similarly, any $V_{UXY\tilde{Y}\hat{Y}} \in \mathcal{V}_Y^C$ can be written as

$$P_{X|U} P_{Y|U} P_{Y|U} P_{Y|U} P_U V_{Z|UXY\tilde{Y}\hat{Y}}, \tag{5.43}$$

and any $V_{UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}} \in \mathcal{V}_{XY}^C$ can be written as

$$P_{X|U} P_{Y|U} P_{X|U} P_{Y|U} P_{X|U} P_{Y|U} P_U V_{Z|UXY\tilde{X}\tilde{Y}\hat{X}\hat{Y}}. \tag{5.44}$$

Hence, E_C^X, E_C^Y, and E_C^{XY} would be equal to

$$E_C^X(0,0,P_{XYU}) = \min_{\substack{V_{Z|UXY\tilde{X}}P_{X|U}P_{X|U}P_{Y|U}P_U: \\ \alpha(V_{U\tilde{X}YZ})<\alpha(V_{UXYZ})}} D(V_{Z|UXY}||W|V_{UXY}) + I(\tilde{X} \wedge Z|UXY), \tag{5.45}$$

$$E_C^Y(0,0,P_{XYU}) = \min_{\substack{V_{Z|UXY\tilde{Y}}P_{X|U}P_{Y|U}P_{Y|U}P_U: \\ \alpha(V_{UX\tilde{Y}Z})<\alpha(V_{UXYZ})}} D(V_{Z|UXY}||W|V_{UXY}) + I(\tilde{Y} \wedge Z|UXY), \tag{5.46}$$

$$E_C^{XY}(0,0,P_{XYU}) = \min_{\substack{V_{Z|UXY\tilde{X}\tilde{Y}}P_{X|U}P_{Y|U}P_{X|U}P_{Y|U}P_U: \\ \alpha(V_{U\tilde{X}\tilde{Y}Z})<\alpha(V_{UXYZ})}} D(V_{Z|UXY}||W|V_{UXY}) + I(\tilde{X}\tilde{Y} \wedge Z|UXY). \tag{5.47}$$

Theorem 5.4.3. *At rate $R_X = R_Y = 0$,*

$$E_C(0,0,P_{XYU}) = E_\beta(0,0,W,P_{XYU}), \quad for\ \beta \in \{ex, T\} \tag{5.48}$$

where $E_\beta(R_X, R_Y, W, P_{XYU})$ for $\beta \in \{ex, T\}$ are defined in Chapter 3.

5.5 The Maximal Error Exponent vs. The Average Error Exponent

In point to point communication systems, one can show that a lower/upper bound for the maximal error probability of the best code is also a lower/upper bound on the average probability of error for such a code. This is not the case in multiuser communications. For example, it has been shown that for multiuser channels, in general, the maximal error capacity region is smaller than the average error capacity region [18]. The minimum distance bound, we obtained in the previous section, is a valid bounds for the maximal error exponent, but not the average. On the other hand, all the known lower bounds in [38][41][40][39], are only valid for the average error exponent, not the maximal. As a result, despite of the point to point case, comparing these upper and lower bounds does not give us any information about how good these bounds are. In the following, we illustrate an approach that derives a lower/upper bound on the average/maximal error exponent by using a known lower/upper bound for the maximal/average error exponent.

Theorem 5.5.1. *Fix any DM-MAC* $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, $R_X \geq 0$, $R_Y \geq 0$. *The following inequalities hold*

$$E_{av}^{*}\left(R_X, R_Y\right) - R \leq E_{m}^{*}\left(R_X, R_Y\right) \leq E_{av}^{*}\left(R_X, R_Y\right) \leq E_{m}^{*}\left(R_X, R_Y\right) + R, \tag{5.49}$$

where $R = \min\{R_X, R_Y\}$.

Proof. The proof is provided in Section 5.6. □

Corollary 5.5.1. *If* $\min\{R_X, R_Y\} = 0$, *i.e.,* $R_X = 0$ *or* $R_Y = 0$,

$$E_{m}^{*}(R_X, R_Y) = E_{av}^{*}(R_X, R_Y) \tag{5.50}$$

Corollary 5.5.2. *Fix any DM-MAC* $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, $R_X \geq 0$, $R_Y \geq 0$. *Assume that the maximal reliability function is bounded as follows:*

$$E_{m}^{L}\left(R_X, R_Y\right) \leq E_{m}^{*}\left(R_X, R_Y\right) \leq E_{m}^{U}\left(R_X, R_Y\right), \tag{5.51}$$

therefore, the average reliability function can be bounded by

$$E_{m}^{L}\left(R_X, R_Y\right) \leq E_{av}^{*}\left(R_X, R_Y\right) \leq E_{m}^{U}\left(R_X, R_Y\right) + R, \tag{5.52}$$

where $R = \min\{R_X, R_Y\}$. *Similarly, if the average reliability function is bounded as follows:*

$$E_{av}^{L}\left(R_X, R_Y\right) \leq E_{av}^{*}\left(R_X, R_Y\right) \leq E_{av}^{U}\left(R_X, R_Y\right), \tag{5.53}$$

it can be concluded that the maximal reliability function satisfies the following constraint

$$E_{av}^{L}\left(R_X, R_Y\right) - R \leq E_{m}^{*}\left(R_X, R_Y\right) \leq E_{av}^{U}\left(R_X, R_Y\right). \tag{5.54}$$

5.6 Proof of Theorems

5.6.1 Proof of Theorem 5.2.1

For a given MAC, $W : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$, and a multi user code $C = C_X \times C_Y$, where $C_X = \{ \mathbf{x}_i \in \mathcal{X}^n : \quad i = 1, ..., M_X\}$ and $C_Y = \{ \mathbf{y}_j \in \mathcal{Y}^n : \quad j = 1, ..., M_Y\}$, with decoding sets $D_{ij} \subset \mathcal{Z}^n$, the average error probability can be written as

$$e(C, W) = \frac{1}{M_X M_Y} \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} W^n \left(D_{ij}^c | \mathbf{x}_i, \mathbf{x}_j \right) \tag{5.55}$$

$$= \frac{1}{M_X M_Y} \sum_{P_{XY}} \frac{M_{XY}}{M_{XY}} \sum_{(i,j) \in C_{XY}} W^n \left(D_{ij}^c | \mathbf{x}_i, \mathbf{x}_j \right) \tag{5.56}$$

where C_{XY} is the set of all codewords pairs sharing fix joint composition P_{XY}, i.e., $C_{XY} = (C_X \times C_Y) \cap T_{P_{XY}}$. The cardinality of this set is shown by M_{XY}, and R_{XY} denotes the rate of this set, i.e., $R_{XY} = \frac{1}{n} \log M_{XY}$. For a fixed pair (i, j), $T_V(\mathbf{x}_i, \mathbf{x}_j)$s are disjoint subsets of $\mathcal{Z}^n$ for different conditional types $V : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$. Therefore, the average error probability of the code can be written as

$$e(C, W) = \frac{1}{M_X M_Y} \sum_{P_{XY}} \frac{M_{XY}}{M_{XY}} \sum_{(i,j) \in C_{XY}} \sum_{V} W^n \left(D_{ij}^c \cap T_V(\mathbf{x}_i, \mathbf{y}_j) | \mathbf{x}_i, \mathbf{y}_j \right)$$

$$= \frac{1}{M_X M_Y} \sum_{P_{XY}} M_{XY} \sum_{V} \frac{1}{M_{XY}} \sum_{(i,j) \in C_{XY}} W^n \left(T_V(\mathbf{x}_i, \mathbf{y}_j) | i, j \right) \frac{|D_{ij}^c \cap T_V(\mathbf{x}_i, \mathbf{y}_j)|}{|T_V(\mathbf{x}_i, \mathbf{y}_j)|}$$

$$= \frac{1}{M_X M_Y} \sum_{P_{XY}} M_{XY} \sum_{V} 2^{-nD(V||W|P_{XY})} \left[1 - \frac{1}{M_{XY}} \sum_{(i,j) \in C_{XY}} \frac{|D_{ij} \cap T_V(\mathbf{x}_i, \mathbf{y}_j)|}{|T_V(\mathbf{x}_i, \mathbf{y}_j)|} \right]$$

$$\geq \frac{1}{M_X M_Y} \sum_{P_{XY}} M_{XY} \sum_{V} 2^{-nD(V||W|P_{XY})} \left[1 - \frac{1}{M_{XY}} \sum_{(i,j) \in C_{XY}} \frac{|D_{ij} \cap T_V(\mathbf{x}_i, \mathbf{y}_j)|}{2^{nH(Z|X,Y)}} \right]$$

$$
\begin{aligned}
&= \frac{1}{M_X M_Y} \sum_{P_{XY}} M_{XY} \sum_{V} 2^{-nD(V||W|P_{XY})} \left[1 - \frac{1}{M_{XY}} \frac{|\bigcup_{(i,j)\in C_{XY}} D_{ij} \cap T_V\left(\mathbf{x}_i, \mathbf{y}_j\right)|}{2^{nH(Z|X,Y)}}\right] \\
&\geq \frac{1}{M_X M_Y} \sum_{P_{XY}} M_{XY} \sum_{V} 2^{-nD(V||W|P_{XY})} \left[1 - \frac{1}{M_{XY}} \frac{|T_Z|}{2^{nH(Z|X,Y)}}\right] \\
&\geq \frac{1}{M_X M_Y} \sum_{P_{XY}} M_{XY} \sum_{V} 2^{-nD(V||W|P_{XY})} \left[1 - \frac{1}{M_{XY}} \frac{2^{nH(Z)}}{2^{nH(Z|X,Y)}}\right] \\
&= \frac{1}{M_X M_Y} \sum_{P_{XY}} M_{XY} \sum_{V} 2^{-nD(V||W|P_{XY})} \left[1 - 2^{-n[R_{XY} - I_V(XY \wedge Z)]}\right]. \qquad (5.57)
\end{aligned}
$$

By defining

$$
V_{bad}^{XY} = \{V : R_{XY} \geq I_V\left(XY \wedge Z\right)\}, \qquad (5.58)
$$

the average error probability of the code can be further lower bounded by

$$
\begin{aligned}
e\left(C, W\right) &\geq \frac{1}{M_X M_Y} \sum_{P_{XY}} M_{XY} \sum_{V \in V_{bad}^{XY}} \frac{1}{2} 2^{-nD(V||W|P_{XY})} \\
&\geq \frac{1}{M_X M_Y} \sum_{P_{XY}} \frac{M_{XY}}{2} 2^{-n[\min_{V \in V_{bad}^{XY}} D(V||W|P_{XY})]} \\
&= \frac{1}{M_X M_Y} \sum_{P_{XY}} \frac{1}{2} 2^{-n[\min_{V \in V_{bad}^{XY}} D(V||W|P_{XY}) - R_{XY}]} \\
&\geq \frac{1}{2 M_X M_Y} 2^{-n[\min_{P_{XY}} \min_{V \in V_{bad}^{XY}} D(V||W|P_{XY}) - R_{XY}]}. \qquad (5.59)
\end{aligned}
$$

Thus,

$$
e\left(C, W\right) \geq \frac{1}{2} 2^{-n[\min_{P_{XY}} \min_{V \in V_{bad}^{XY}} D(V||W|P_{XY}) + R_X + R_Y - R_{XY}]} \qquad (5.60)
$$

On the other hand, by using the fact that $D_{ij}^c \subseteq \bigcup_{j'} \bigcup_{i' \neq i} D_{i'j'}$, the average probability of error can be lower bounded by

$$
\begin{aligned}
e\left(C,W\right) &= \frac{1}{M_X M_Y} \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} W^n \left(D_{ij}^c | \mathbf{x}_i, \mathbf{y}_j\right) \\
&\geq \frac{1}{M_Y} \sum_{j=1}^{M_Y} \frac{1}{M_X} \sum_{i=1}^{M_X} W^n \left(\bigcup_{j'} \bigcup_{i' \neq i} D_{i'j'} | \mathbf{x}_i, \mathbf{y}_j \right) \\
&= \frac{1}{M_X M_Y} \sum_{P_{XY}} \sum_{i} \sum_{j:(i,j) \in C_{XY}} \sum_{V} W^n \left(D_i^c \cap T_V\left(\mathbf{x}_i, \mathbf{y}_j\right) | \mathbf{x}_i, \mathbf{y}_j\right) \\
&= \frac{1}{M_X M_Y} \sum_{P_{XY}} \sum_{V} 2^{-nD(V||W|P_{XY})} \sum_{i} \sum_{j:(i,j) \in C_{XY}} \frac{|D_i^c \cap T_V\left(\mathbf{x}_i, \mathbf{y}_j\right)|}{|T_V\left(\mathbf{x}_i, \mathbf{y}_j\right)|} \\
&= \sum_{P_{XY}} \sum_{V} 2^{-nD(V||W|P_{XY})} \frac{1}{M_X M_Y} \sum_{i} \sum_{j:(i,j) \in C_{XY}} \left[1 - \frac{|D_i \cap T_V\left(\mathbf{x}_i, \mathbf{y}_j\right)|}{|T_V\left(\mathbf{x}_i, \mathbf{y}_j\right)|}\right] \\
&= \sum_{P_{XY}} \sum_{V} 2^{-nD(V||W|P_{XY})} \frac{M_{XY}}{M_X M_Y} \left[1 - \frac{1}{M_{XY}} \sum_{i} \sum_{j:(i,j) \in C_{XY}} \frac{|D_i \cap T_V\left(\mathbf{x}_i, \mathbf{y}_j\right)|}{|T_V\left(\mathbf{x}_i, \mathbf{y}_j\right)|}\right] \\
&\geq \sum_{P_{XY}} \sum_{V} 2^{-nD(V||W|P_{XY})} \frac{M_{XY}}{M_X M_Y} \left[1 - \frac{1}{M_{XY}} \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} \frac{|D_i \cap T_V\left(\mathbf{x}_i, \mathbf{y}_j\right)|}{|T_V\left(\mathbf{x}_i, \mathbf{y}_j\right)|}\right] \\
&\geq \sum_{P_{XY}} \sum_{V} 2^{-nD(V||W|P_{XY})} \frac{M_{XY}}{M_X M_Y} \left[1 - \frac{1}{M_{XY}} \sum_{j=1}^{M_Y} \sum_{i=1}^{M_X} \frac{|D_i \cap T_V\left(\mathbf{x}_i, \mathbf{y}_j\right)|}{2^{nH(Z|X,Y)}}\right] \\
&\geq \sum_{P_{XY}} \sum_{V} 2^{-nD(V||W|P_{XY})} \frac{M_{XY}}{M_X M_Y} \left[1 - \frac{1}{M_{XY}} \sum_{j=1}^{M_Y} \frac{2^{nH(Z,X|Y)}}{2^{nH(Z|X,Y)}}\right] \\
&\geq \sum_{P_{XY}} \sum_{V} 2^{-nD(V||W|P_{XY})} \frac{M_{XY}}{M_X M_Y} \left[1 - 2^{-n[R_{XY} - R_Y - I_V(Z \wedge X|Y)]}\right].
\end{aligned} \tag{5.61}
$$

By defining

$$
V_{bad}^X \triangleq \{V : R_{XY} - R_Y \geq I_V\left(Z \wedge X|Y\right)\}, \tag{5.62}
$$

and using (5.61), it can be concluded that

$$
\begin{aligned}
e\left(C,W\right) &\geq \frac{1}{M_X M_Y}\sum_{P_{XY}} M_{XY} \sum_{V\in V_{bad}^{X}} \frac{1}{2} 2^{-nD(V||W|P_{XY})} \\
&\geq \frac{1}{M_X M_Y}\sum_{P_{XY}} \frac{M_{XY}}{2} 2^{-n[\min_{V\in V_{bad}^{X}} D(V||W|P_{XY})]} \\
&= \frac{1}{M_X M_Y}\sum_{P_{XY}} \frac{1}{2} 2^{-n[\min_{V\in V_{bad}^{X}} D(V||W|P_{XY})-R_{XY}]} \\
&\geq \frac{1}{2M_X M_Y} 2^{-n[\min_{P_{XY}} \min_{V\in V_{bad}^{X}} D(V||W|P_{XY})-R_{XY}]}.
\end{aligned} \tag{5.63}
$$

Therefore, the average error probability can be lower bounded by

$$
e\left(C,W\right) \geq \frac{1}{2} 2^{-n[\min_{P_{XY}} \min_{V\in V_{bad}^{X}} D(V||W|P_{XY})+R_X+R_Y-R_{XY}]}. \tag{5.64}
$$

Similarly, it can be shown that

$$
e\left(C,W\right) \geq \frac{1}{2} 2^{-n[\min_{P_{XY}} \min_{V\in V_{bad}^{Y}} D(V||W|P_{XY})+R_X+R_Y-R_{XY}]}, \tag{5.65}
$$

where

$$
V_{bad}^{Y} = \{V : R_{XY} - R_X \geq I_V\left(Z \wedge Y|X\right)\}. \tag{5.66}
$$

By combining (5.60), (5.60), (5.60), we conclude that

$$
e\left(C,W\right) \geq \frac{1}{2} 2^{-n[\min_{P_{XY}} \min_{V\in V_{bad}^{X}\cup V_{bad}^{Y}\cup V_{bad}^{XY}} D(V||W|P_{XY})+R_X+R_Y-R_{XY}]}. \tag{5.67}
$$

Equivalently, for the exponent of $e\left(C,W\right)$, which is denoted by $E\left(C,W\right)$, can be upper bounded by

$$
E\left(C,W\right) \leq \min_{P_{XY}} \min_{V\in V_{bad}^{X}\cup V_{bad}^{Y}\cup V_{bad}^{XY}} D\left(V\|W|P_{XY}\right) + R_X + R_Y - R_{XY}. \tag{5.68}
$$

By defining $V_{bad} \triangleq V_{bad}^X \cup V_{bad}^Y \cup V_{bad}^{XY}$, the previous inequality can be simplified to

$$\begin{aligned} E(C,W) &\leq \max_{C} \quad \min_{P_{XY}} \min_{V \in V_{bad}} D(V\|W|P_{XY}) + R_X + R_Y - R_{XY} && (5.69) \\ &= \max_{\underline{R}\in\mathcal{R}} \quad \min_{P_{XY}} \min_{V \in V_{bad}} D(V\|W|P_{XY}) + R_X + R_Y - R_{XY} && (5.70) \end{aligned}$$

Where $\underline{R}$ is a vector with elements $R(C,P_{XY})$ and $\mathcal{R}$ is the set of all possible vectors $\underline{R}$. The last equality follows from the fact that $E(C,W)$ is only a function of R_{XY}s. By using the fact that P_{XY}^n is a dominant type of the code, we conclude that

$$\begin{aligned} E(C,W) &\leq \max_{\underline{R}\in\mathcal{R}} \min_{V \in V_{bad}} D(V\|W|P_{XY}^n) + R_X + R_Y - R(C,P_{XY}^n) && (5.71) \\ &= \max_{\underline{R}\in\mathcal{R}} \min_{V \in V_{bad}} D(V\|W|P_{XY}^n). && (5.72) \end{aligned}$$

since this expression does not depend on $\underline{R}$, we conclude that

$$E(C,W) \leq \min_{V \in V_{bad}} D(V\|W|P_{XY}^n), \tag{5.73}$$

where

$$V_{bad} = \{V : I_V(XY \wedge Z) \leq R_X + R_Y \text{ or } I_V(Y \wedge Z|X) \leq R_Y \text{ or } I_V(X \wedge Z|Y) \leq R_X\} \tag{5.74}$$

5.6.2 Proof of Theorem 5.3.1

Our approach makes use of Agustin's [1] strong converse theorem for one-way channels which is stated in the following:

Lemma 5.6.1. *[1]: For a (n,M,λ) code $\{(\mathbf{x}_i, D_i) : 1 \leq i \leq M\}$ and a non-stationary DMC $(W_t)_{t=1}^{\infty}$*

$$\log M < \sum_{t=1}^{n} I(X_t \wedge Z_t) + \frac{3}{1-\lambda}|\mathcal{X}|n^{1/2}, \tag{5.75}$$

where the distribution of the RV's are determined by the Fano-distribution on the codewords.

Consider any $P^* \in \mathcal{P}_n(\mathcal{X})$, such that $|\mathcal{E}_W(C, P^*, \lambda)| \geq \frac{1}{(n+1)^{|\mathcal{X}|}} \left(1 - \frac{2\lambda}{1+\lambda}\right) M$. The code $\mathcal{E}_W(C, P^*, \lambda)$ is an $\left(n, |\mathcal{E}_W(C, P^*, \lambda)|, \frac{1+\lambda}{2}\right)$ code. Let us define $\lambda' \triangleq \frac{1+\lambda}{2}$. Therefore, by the result of Lemma 5.6.1, we conclude that

$$\log\left(|\mathcal{E}_W(C, P^*, \lambda)|\right) < \sum_{t=1}^{n} I(X_t \wedge Z_t) + \frac{3}{1-\lambda'}|\mathcal{X}|\sqrt{n}, \tag{5.76}$$

where the distribution of RV's are determined by the Fano-distribution on the codewords. By using the lower bound on the size of $\mathcal{E}_W(C, P^*, \lambda)$, it can be concluded that

$$\frac{1}{n}\log M \leq \frac{1}{n}\sum_{t=1}^{n} I(X_t \wedge Z_t) + \frac{3|\mathcal{X}|}{(1-\lambda')\sqrt{n}} + |\mathcal{X}|\frac{\log(n+1)}{n} + \frac{1}{n}\log\left(\frac{1+\lambda}{1-\lambda}\right). \tag{5.77}$$

The last three terms on the right hand side of (5.77) are approaching zero for sufficiently large n. Let us focus on the first term. In the following, we prove that the first term is bounded from above by:

$$\frac{1}{n}\sum_{t=1}^{n} I(X_t \wedge Z_t) \leq I(P^*, W). \tag{5.78}$$

First, note that

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} I(X_t \wedge Z_t)\\
&\quad= \frac{1}{n}\sum_{t=1}^{n}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} \mathbb{P}(X_t = x)\mathbb{P}(Z_t = z|X_t = x)\log\left(\frac{\mathbb{P}(Z_t = z|X_t = x)}{\mathbb{P}(Z_t = z)}\right)\\
&\quad= \frac{1}{n}\sum_{t=1}^{n}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} \mathbb{P}(X_t = x)W(z|x)\log\left(\frac{W(z|x)}{\mathbb{P}(Z_t = z)}\right)\\
&\quad= \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} W(z|x)\log(W(z|x))\frac{1}{n}\sum_{t=1}^{n}\mathbb{P}(X_t = x)\\
&\qquad\qquad - \frac{1}{n}\sum_{t=1}^{n}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} \mathbb{P}(X_t = x)W(z|x)\log(\mathbb{P}(Z_t = z))
\end{aligned}$$

$$= \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} P^*(x)W(z|x)\log(W(z|x))$$
$$- \frac{1}{n}\sum_{t=1}^{n}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} \mathbb{P}(X_t = x)W(z|x)\log(\mathbb{P}(Z_t = z)). \tag{5.79}$$

The last equality holds because $\mathcal{E}_W(C, P^*, \lambda)$ is a constant composition code with composition P^*. The second term in (5.79) can be written as

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} \mathbb{P}(X_t = x)W(z|x)\log(\mathbb{P}(Z_t = z))$$
$$= \sum_{z\in\mathcal{Z}}\frac{1}{n}\sum_{t=1}^{n}\left(\log\left(\sum_{x'\in\mathcal{X}}\mathbb{P}(X_t = x')W(z|x')\right)\sum_{x\in\mathcal{X}}\mathbb{P}(X_t = x)W(z|x)\right) \tag{5.80}$$

In the right hand side of (5.80), the summands are of the form of $u\log(u)$, which is a convex function of u. Thus,

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} \mathbb{P}(X_t = x)W(z|x)\log(\mathbb{P}(Z_t = z))$$
$$\geq \sum_{z\in\mathcal{Z}}\left(\log\left(\frac{1}{n}\sum_{t=1}^{n}\sum_{x'\in\mathcal{X}}\mathbb{P}(X_t = x')W(z|x')\right)\frac{1}{n}\sum_{t=1}^{n}\sum_{x\in\mathcal{X}}\mathbb{P}(X_t = x)W(z|x)\right)$$
$$= \sum_{z\in\mathcal{Z}}\left(\log\left(\sum_{x'\in\mathcal{X}}\frac{1}{n}\sum_{t=1}^{n}\mathbb{P}(X_t = x')W(z|x')\right)\sum_{x\in\mathcal{X}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{P}(X_t = x)W(z|x)\right)$$
$$= \sum_{z\in\mathcal{Z}}\left(\log\left(\sum_{x'\in\mathcal{X}}P^*(x')W(z|x')\right)\sum_{x\in\mathcal{X}}P^*(x)W(z|x)\right). \tag{5.81}$$

Finally, by combining (5.79) and (5.81), it can be concluded that

$$\frac{1}{n}\sum_{t=1}^{n} I(X_t \wedge Z_t) \leq \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} P^*(x)W(z|x)\log(W(z|x))$$
$$- \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} P^*(x)W(z|x)\log\left(\sum_{x'\in\mathcal{X}}P^*(x')W(z|x')\right)$$
$$= I(P^*, W), \tag{5.82}$$

which completes the proof.

5.6.3 Proof of Fact 5.3.1

Since code C is an (n, M) code, it can be concluded that it must have at least a dominant type, $P^* \in \mathcal{P}_n(\mathcal{X})$. Consider an arbitrary discrete memoryless channel $V : \mathcal{X} \to \mathcal{Z}$, such that $R > I(P^*, V)$. Choose $\bar{\lambda} < 1$ satisfying

$$\frac{1+\bar{\lambda}}{2} > 1 - \frac{\delta}{2}. \tag{5.83}$$

Since P^* is a dominant type of code C,

$$|C \cap T_{P^*}| \geq \frac{1}{(n+1)^{|\mathcal{X}|}} M. \tag{5.84}$$

On the other hand, since $R > I(P^*, V)$, it can be concluded from Theorem 5.3.1 that

$$\left|\mathcal{E}_V(C, P^*, \bar{\lambda})\right| < \frac{1}{(n+1)^{|\mathcal{X}|}} \left(1 - \frac{2\bar{\lambda}}{1+\bar{\lambda}}\right) M. \tag{5.85}$$

By combining (5.84) and (5.85), it can be concluded that

$$\left|\mathcal{D}_V(C, P^*, \bar{\lambda})\right| \geq \frac{1}{(n+1)^{|\mathcal{X}|}} \left(\frac{2\bar{\lambda}}{1+\bar{\lambda}}\right) M, \tag{5.86}$$

where $\mathcal{D}_V(C, P^n_{XY}, \bar{\lambda})$ is defined as

$$\mathcal{D}_V(C, P^*, \bar{\lambda}) \triangleq (C \cap T_{P^*}) \,/\, \mathcal{E}_V(C, P^*, \bar{\lambda}) = \big\{\mathbf{x}_i \in C \cap T_{P^*} : V(D_i^c|\mathbf{x}_i) > \frac{1+\bar{\lambda}}{2}\big\}. \tag{5.87}$$

By combining (5.83), (5.87) and using the same method as Csiszar in [16, pp. 167], we have

$$
\begin{aligned}
W^n(D_i^c|\mathbf{x}_i) &\geq \exp\Big\{-\frac{D(V||W|P^*)+h(1-\frac{\delta}{2})}{1-\frac{\delta}{2}}\Big\} \\
&\geq \frac{1}{2}\exp\{-nD(V||W|P^*)(1+\delta)\} \qquad \text{for all } \mathbf{x}_i \in \mathcal{D}_V(C,P^*,\bar{\lambda}), \quad (5.88)
\end{aligned}
$$

for small enough δ satisfying $h(1-\frac{\delta}{2}) < 1-\frac{\delta}{2}$. The average error probability of the code C over the channel W can be written as

$$
\begin{aligned}
e(C,W) &= \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{M_Y} W^n(D_i^c|\mathbf{x}_i) \\
&\geq \frac{1}{M}\sum_{\substack{i:\\ \mathbf{x}_i\in\mathcal{D}_V(C,P^*,\bar{\lambda})}} \frac{1}{2}\exp\{-nD(V||W|P^*)(1+\delta)\} \\
&\geq \frac{1}{(n+1)^{|\mathcal{X}|}}\left(\frac{\bar{\lambda}}{1+\bar{\lambda}}\right)\exp\{-nD(V||W|P^*)(1+\delta)\}. \qquad (5.89)
\end{aligned}
$$

Since the inequality (5.89) holds for all $V:\mathcal{X}\to\mathcal{Z}$ satisfying $I(P^*,V)<R$, it can be concluded that

$$
\begin{aligned}
e(C,W) &\geq \max_{V:I(P^*,V)<R} \exp\{-n[D(V||W|P^*)(1+\delta)+\delta]\} \\
&= \exp\{-n[\min_{V:I(P^*,V)<R} D(V||W|P^*)(1+\delta)+\delta]\}, \qquad (5.90)
\end{aligned}
$$

for sufficiently large n. As we mentioned earlier, P^* is any dominant type of the code. We can further lower bound the average error probability as follows

$$
\begin{aligned}
e(C,W) &\geq \min_{P^*\in\mathcal{P}_n(\mathcal{X})} \exp\{-n[\min_{V:I(P^*,V)<R} D(V||W|P^*)(1+\delta)+\delta]\} \\
&\geq \min_{P\in\mathcal{P}(\mathcal{X})} \exp\{-n[\min_{V:I(P,V)<R} D(V||W|P)(1+\delta)+\delta]\}, \qquad (5.91)
\end{aligned}
$$

where the last inequality follows by a continuity argument.

5.6.4 Proof of Theorem 5.3.2

The basic idea of the proof is wringing technique which was used for the first time, by Ahlswede [2]. Consider any $P^n_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$, such that $|\mathcal{E}_W(C, P^n_{XY}, \lambda)| \geq \frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}}(1 - \frac{2\lambda}{1+\lambda})M_X M_Y$. Let $\mathcal{A}$ contains all codewords pairs with joint composition P^n_{XY}, and with small probability error:

$$\bar{\mathcal{A}} \triangleq \left\{(i,j) : W(D_{ij}|\mathbf{x}_i, \mathbf{y}_j) \geq \frac{1-\lambda}{2}, (\mathbf{x}_i, \mathbf{y}_j) \in T_{P^n_{XY}}\right\}. \tag{5.92}$$

Since $|\bar{\mathcal{A}}| = |\mathcal{E}_W(C, P^n_{XY}, \lambda)|$, we conclude that

$$|\bar{\mathcal{A}}| \geq \frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} \left(1 - \frac{2\lambda}{1+\lambda}\right) M_X M_Y. \tag{5.93}$$

Define

$$\bar{\mathcal{C}}(i) = \left\{(i,j) : (i,j) \in \bar{\mathcal{A}}, 1 \leq j \leq M_Y\right\} \tag{5.94a}$$

$$\bar{\mathcal{B}}(j) = \left\{(i,j) : (i,j) \in \bar{\mathcal{A}}, 1 \leq i \leq M_X\right\}. \tag{5.94b}$$

Consider the subcode $\left\{(\mathbf{x}_i, \mathbf{y}_j, D_{ij}) : (i,j) \in \bar{\mathcal{A}}\right\}$ and define random variables $\bar{X}^n$, $\bar{Y}^n$

$$\mathbb{P}\left((\bar{X}^n, \bar{Y}^n) = (\mathbf{x}_i, \mathbf{y}_j)\right) = \frac{1}{|\bar{\mathcal{A}}|} \qquad \text{if } (i,j) \in \bar{\mathcal{A}}. \tag{5.95}$$

Lemma 5.6.2. *For random variables $\bar{X}^n$, $\bar{Y}^n$ defined in* (5.95)*, the mutual information satisfies the following inequality:*

$$I(\bar{X}^n \wedge \bar{Y}^n) \leq -\log\left(1 - \frac{2\lambda}{1+\lambda}\right) + |\mathcal{X}||\mathcal{Y}|\log(n+1). \tag{5.96}$$

Proof. This is a generalization of the proof by Dueck in [19]. Note that

$$H(\bar{Y}^n|\bar{X}^n) = H(\bar{X}^n, \bar{Y}^n) - H(\bar{X}^n) = \log|\bar{\mathcal{A}}| - H(\bar{X}^n) \geq \log|\bar{\mathcal{A}}| - \log(M_X). \tag{5.97}$$

By (5.93), we conclude that

$$H(\bar{Y}^n|\bar{X}^n) \geq \quad \log M_Y + \log\left(1 - \frac{2\lambda}{1+\lambda}\right) - |\mathcal{X}||\mathcal{Y}|\log(n+1). \tag{5.98}$$

Finally,

$$\begin{aligned} I(\bar{X}^n \wedge \bar{Y}^n) = H(\bar{Y}^n) - H(\bar{Y}^n|\bar{X}^n) &\leq \log M_Y - H(\bar{Y}^n|\bar{X}^n) \\ &\leq -\log\left(1 - \frac{2\lambda}{1+\lambda}\right) + |\mathcal{X}||\mathcal{Y}|\log(n+1), \end{aligned} \tag{5.99}$$

which concludes the proof. □

The next lemma is Ahlswede's version of the 'wringing' technique. Roughly speaking, if it is known that the mutual information between two random sequences is small, then the lemma gives an upper bound on the per-letter mutual information terms (conditioned on some values).

Lemma 5.6.3. *[5] Let X^n, Y^n be RV's with values in $\mathcal{X}^n$, $\mathcal{Y}^n$ resp. and assume that*

$$I(X^n \wedge Y^n) \leq \sigma \tag{5.100}$$

Then, for any $0 < \delta < \sigma$ there exist $t_1, t_2, ..., t_k \in \{1, ..., n\}$ where $0 \leq k < \frac{2\sigma}{\delta}$ such that for some $\bar{x}_{t_1}, \bar{y}_{t_1}, \bar{x}_{t_2}, \bar{y}_{t_2}$,
$..., \bar{x}_{t_k}, \bar{y}_{t_k}$

$$I(X_t \wedge Y_t | X_{t_1} = \bar{x}_{t_1}, Y_{t_1} = \bar{y}_{t_1}, ..., X_{t_k} = \bar{x}_{t_k}, Y_{t_k} = \bar{y}_{t_k}) \leq \delta \qquad \text{for } t = 1, 2, ..., n, \tag{5.101}$$

and

$$\mathbb{P}(X_{t_1} = \bar{x}_{t_1}, Y_{t_1} = \bar{y}_{t_1}, ..., X_{t_k} = \bar{x}_{t_k}, Y_{t_k} = \bar{y}_{t_k}) \geq \left(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma - \delta)}\right)^k. \tag{5.102}$$

Proof. The proof is provided in [5]. □

Consider the subcode $\{(\mathbf{x}_i, \mathbf{y}_j, D_{ij}) : (i,j) \in \mathcal{A}\}$, where

$$\mathcal{A} \triangleq \left\{(i,j) \in \bar{\mathcal{A}} : (\mathbf{x}_i)_{t_l} = \bar{x}_{t_l}, (\mathbf{y}_j)_{t_l} = \bar{y}_{t_l} \;\; 1 \leq l \leq k\right\} \tag{5.103}$$

and define

$$\mathcal{C}(i) = \{(i,j) : (i,j) \in \mathcal{A}, 1 \leq j \leq M_Y\} \tag{5.104a}$$

$$\mathcal{B}(j) = \{(i,j) : (i,j) \in \mathcal{A}, 1 \leq i \leq M_X\}\,. \tag{5.104b}$$

Lemma 5.6.4. *The subcode* $\{(\mathbf{x}_i, \mathbf{y}_j, D_{ij}) : (i,j) \in \mathcal{A}\}$ *is a subcode with maximal error probability of at most* $\frac{1+\lambda}{2}$*, and*

$$|\mathcal{A}| \geq \left(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma - \delta)}\right)^k |\bar{\mathcal{A}}|. \tag{5.105}$$

Moreover,

$$\sum_{x,y} |\mathbb{P}(X_t = x, Y_t = y) - \mathbb{P}(X_t = x)\mathbb{P}(Y_t = y)| \leq 2\delta^{1/2}, \tag{5.106}$$

where $X^n = (X_1, ..., X_n)$*,* $Y^n = (Y_1, ..., Y_n)$ *are distributed according to the Fano-distribution of the subcode* $\{(\mathbf{x}_i, \mathbf{y}_j, D_{ij}) : (i,j) \in \mathcal{A}\}$*.*

Proof. Since $\mathcal{A} \subset \bar{\mathcal{A}}$, the maximal probability of error for the corresponding code is at most $\frac{1+\lambda}{2}$. The second part of Lemma 5.6.3, immediately yields (5.105). On the other hand,

$$\begin{aligned}
\mathbb{P}_{\bar{\mathcal{A}}}(\bar{X}_t = x, \bar{Y}_t = y | \bar{x}_{t_1}, \bar{y}_{t_1}, \bar{x}_{t_2}, \bar{y}_{t_2}, ..., \bar{x}_{t_k}, \bar{y}_{t_k}) \\
&= \frac{\mathbb{P}_{\bar{\mathcal{A}}}(\bar{X}_t = x, \bar{Y}_t = y, \bar{x}_{t_1}, \bar{y}_{t_1}, \bar{x}_{t_2}, \bar{y}_{t_2}, ..., \bar{x}_{t_k}, \bar{y}_{t_k})}{\mathbb{P}_{\bar{\mathcal{A}}}(\bar{x}_{t_1}, \bar{y}_{t_1}, \bar{x}_{t_2}, \bar{y}_{t_2}, ..., \bar{x}_{t_k}, \bar{y}_{t_k})} \\
&= \frac{N_{\bar{\mathcal{A}}}(\bar{X}_t = x, \bar{Y}_t = y, \bar{x}_{t_1}, \bar{y}_{t_1}, \bar{x}_{t_2}, \bar{y}_{t_2}, ..., \bar{x}_{t_k}, \bar{y}_{t_k})}{N_{\mathcal{A}}(\bar{x}_{t_1}, \bar{y}_{t_1}, \bar{x}_{t_2}, \bar{y}_{t_2}, ..., \bar{x}_{t_k}, \bar{y}_{t_k})} \\
&= \frac{N_{\mathcal{A}}(X_t = x, Y_t = y)}{|\mathcal{A}|} \\
&= \mathbb{P}_{\mathcal{A}}(X_t = x, Y_t = y). \qquad (5.107)
\end{aligned}$$

Therefore, by the first part of Lemma 5.6.3, we conclude that

$$I_{\mathcal{A}}(X_t \wedge Y_t) \leq \delta, \quad \text{for } 1 \leq t \leq n. \qquad (5.108)$$

Since $I_{\mathcal{A}}(X_t \wedge Y_t)$ is an *I-divergence*, Pinsker's inequality implies [24]

$$\sum_{x,y} |\mathbb{P}(X_t = x, Y_t = y) - \mathbb{P}(X_t = x)\mathbb{P}(Y_t = y)| \leq 2\delta^{1/2}, \qquad \text{for } t = 1, 2, ..., n. \qquad (5.109)$$

□

Now, let us define random variables X^n, Y^n on $\mathcal{X}^n$, $\mathcal{Y}^n$ respectively by

$$Pr((X^n, Y^n) = (\mathbf{x}_i, \mathbf{y}_j)) = \frac{1}{|\mathcal{A}|} \qquad \text{if } (i,j) \in \mathcal{A}. \qquad (5.110)$$

Lemma 5.6.5. *For any $0 \leq \lambda < 1$, any (n, M_X, M_Y) code $C \triangleq \{(\mathbf{x}_i, \mathbf{y}_j, D_{ij}) : 1 \leq i \leq M_X, 1 \leq j \leq M_Y\}$ any MAC, W, and any $P^n_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ satisfying*

$|\mathcal{E}_W(C, P_{XY}^n, \lambda)| \geq \frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} \left(1 - \frac{2\lambda}{1+\lambda}\right) M_X M_Y$,*the following holds*

$$\log M_X \leq \sum_{t=1}^{n} I(X_t \wedge Z_t|Y_t) + c_1(\lambda)n^{1/2} + c_1 k \log(\frac{2\sigma}{\delta}),$$
$$\log M_Y \leq \sum_{t=1}^{n} I(Y_t \wedge Z_t|X_t) + c_2(\lambda)n^{1/2} + c_2 k \log(\frac{2\sigma}{\delta}),$$
$$\log (M_X M_Y) \leq \sum_{t=1}^{n} I(X_t Y_t \wedge Z_t) + c_3(\lambda)n^{1/2} + c_3 k \log(\frac{2\sigma}{\delta}),$$

where the distributions of the RV's are determined by the Fano-distribution on the codewords $\{(\mathbf{x}_i, \mathbf{y}_j) : (i,j) \in \mathcal{A}\}$*. Here,* $c_i(\lambda)$ *and* c_i *are suitable functions of* λ*.*

Proof. For any fixed j, consider $(n, |\mathcal{B}(j)|)$ code $\{(\mathbf{x}_i, D_{ij}) : (i,j) \in \mathcal{B}(j)\}$. For channel W, any pair of codewords in this code has probability of error at most equal to $\frac{1+\lambda}{2}$. Let us define $\lambda' \triangleq \frac{1+\lambda}{2}$. It follows from lemma 5.6.1 that

$$\log |\mathcal{B}(j)| \leq \sum_{t=1}^{n} I(X_t \wedge Z_t | Y_t = (\mathbf{y}_j)_t) + \frac{3}{1-\lambda'}|\mathcal{X}|n^{1/2}. \tag{5.111}$$

Similarly, it can be shown that

$$\log |\mathcal{C}(i)| \leq \sum_{t=1}^{n} I(Y_t \wedge Z_t | X_t = (\mathbf{x}_i)_t) + \frac{3}{1-\lambda'}|\mathcal{Y}|n^{1/2}, \tag{5.112}$$
$$\log |\mathcal{A}| \leq \sum_{t=1}^{n} I(X_t Y_t \wedge Z_t) + \frac{3}{1-\lambda'}|\mathcal{X}||\mathcal{Y}|n^{1/2}. \tag{5.113}$$

Since $\mathbb{P}(Y_t = y) = \frac{\sum_{(i,j)\in\mathcal{A}} 1_{\{\mathbf{y}_{jt}=y\}}}{|\mathcal{A}|}$,

$$\begin{aligned}
\frac{1}{|\mathcal{A}|}\sum_{(i,j)\in\mathcal{A}} \log |\mathcal{B}(j)| &\leq \sum_{(i,j)\in\mathcal{A}} \sum_{t=1}^{n} I(X_t \wedge Z_t | Y_t = (\mathbf{y}_j)_t) \frac{\sum_y 1_{\{(\mathbf{y}_j)_t = y\}}}{|\mathcal{A}|} + \frac{3}{1-\lambda'}|\mathcal{X}|n^{1/2} \\
&= \sum_{t=1}^{n} \sum_{y} I(X_t \wedge Z_t | Y_t = y)\mathbb{P}(Y_t = y) + \frac{3}{1-\lambda'}|\mathcal{X}|n^{1/2} \\
&= \sum_{t=1}^{n} I(X_t \wedge Z_t | Y_t) + \frac{3}{1-\lambda'}|\mathcal{X}|n^{1/2}.
\end{aligned} \tag{5.114}$$

The left hand side of (5.114) can be lower-bounded as follows

$$
\begin{aligned}
\frac{1}{|\mathcal{A}|}\sum_{(i,j)\in\mathcal{A}}\log|\mathcal{B}(j)| &= \frac{1}{|\mathcal{A}|}\sum_{j}|\mathcal{B}(j)|\log|\mathcal{B}(j)| \\
&\geq \frac{1}{|\mathcal{A}|}\sum_{j:|\mathcal{B}(j)|\geq B^*}|\mathcal{B}(j)|\log|\mathcal{B}(j)| \\
&\geq \frac{1}{|\mathcal{A}|}\log(B^*)\sum_{j:|\mathcal{B}(j)|\geq B^*}|\mathcal{B}(j)| \\
&\geq \frac{|\mathcal{A}|-M_Y B^*}{|\mathcal{A}|}\log(B^*), \qquad (5.115)
\end{aligned}
$$

where λ^* and B^* are defined as follows

$$\lambda^* \triangleq \frac{2\lambda}{1+\lambda}, \qquad (5.116)$$

$$B^* \triangleq \frac{1-\lambda^*}{n}\frac{M_X}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}}\left(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma-\delta)}\right)^k. \qquad (5.117)$$

By using (5.93) and the result of Lemma 5.6.4, it can be concluded that

$$M_Y B^* \leq \frac{1}{n}|\mathcal{A}|. \qquad (5.118)$$

Therefore,

$$
\begin{aligned}
\frac{1}{|\mathcal{A}|}\sum_{(i,j)\in\mathcal{A}}\log|\mathcal{B}(j)| &\geq \frac{|\mathcal{A}|-\frac{1}{n}|\mathcal{A}|}{|\mathcal{A}|}\log(B^*) \\
&= (1-\frac{1}{n})\log\left(\frac{1-\lambda^*}{n}\frac{M_X}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}}\left(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma-\delta)}\right)^k\right).
\end{aligned}
\qquad (5.119)
$$

By (5.114), (5.119)

$$\begin{aligned}\log M_X \le (1+\frac{2}{n})&\left(\sum_{t=1}^{n} I(X_t \wedge Z_t|Y_t) + \frac{3}{1-\lambda'}|\mathcal{X}|n^{1/2}\right)\\ &\quad - \log(1-\lambda^*) + \log n + |\mathcal{X}||\mathcal{Y}|\log(n+1)\\ &\quad + k\log\left(\frac{|\mathcal{X}||\mathcal{Y}|2\sigma}{\delta}\right)\\ &\le \sum_{t=1}^{n} I(X_t \wedge Z_t|Y_t) + c_1(\lambda')n^{1/2} + c_1 k\log(\frac{2\sigma}{\delta}) + 2|\mathcal{Z}|. \end{aligned} \tag{5.120}$$

Analogously,

$$\log M_Y \le \sum_{t=1}^{n} I(Y_t \wedge Z_t|X_t) + c_2(\lambda')n^{1/2} + c_2 k\log(\frac{2\sigma}{\delta}) + 2|\mathcal{Z}|. \tag{5.121}$$

To find an upper bound for $\log(M_X M_Y)$, we first try to find a lower bound on the $\log|\mathcal{A}|$. By Lemma 5.6.4

$$\begin{aligned}\log|\mathcal{A}| &\ge \log|\bar{\mathcal{A}}| + k\log\left(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|(2\sigma-\delta)}\right)\\ &\ge \log|\bar{\mathcal{A}}| + k\log\left(\frac{\delta}{|\mathcal{X}||\mathcal{Y}|2\sigma}\right)\\ &= \log|\bar{\mathcal{A}}| - k\log\left(\frac{2\sigma}{\delta}\right) - k\log\left(|\mathcal{X}||\mathcal{Y}|\right)\\ &\ge \log(M_X M_Y) - |\mathcal{X}||\mathcal{Y}|\log(n+1) + \log\left(1-\frac{2\lambda}{1+\lambda}\right) - k\log\left(\frac{2\sigma}{\delta}\right)\\ &\qquad - k\log\left(|\mathcal{X}||\mathcal{Y}|\right). \end{aligned} \tag{5.122}$$

Therefore,

$$\begin{aligned}\log(M_X M_Y) \le \log|\mathcal{A}| + c_3 k\log(\frac{2\sigma}{\delta}) + |\mathcal{X}||\mathcal{Y}|\log(n+1) + k\log\left(|\mathcal{X}||\mathcal{Y}|\right)\\ - \log\left(1-\frac{2\lambda}{1+\lambda}\right). \end{aligned} \tag{5.123}$$

Using (5.113),

$$\log\left(M_X M_Y\right) \leq \sum_{t=1}^{n} I(X_t Y_t \wedge Z_t) + c_3(\lambda') n^{1/2} + c_3 k \log(\frac{2\sigma}{\delta}) + |\mathcal{X}||\mathcal{Y}| \log(n+1). \quad (5.124)$$

$\square$

Note that, in general X_t and Y_t are not independent. In the following, we prove that they are nearly independent. In the following, we combine (5.96) and the result of Lemma 5.6.4. For an (n, M_X, M_Y) code $\{(\mathbf{x}_i, \mathbf{y}_j, D_{ij}) : 1 \leq i \leq M_X, 1 \leq j \leq M_Y\}$ which has the particular property mentioned in Theorem 5.3.2, define $\bar{\mathcal{A}}$, $\mathcal{A}$ as defined before. Apply Lemma 5.6.4 with parameter $\delta = n^{-1/2}$. By using $\sigma = -\log\left(1 - \frac{2\lambda}{1+\lambda}\right) + |\mathcal{X}||\mathcal{Y}| \log(n+1)$, we conclude that

$$k \leq \frac{2\sigma}{\delta} = 2\sqrt{n}\left(-\log\left(1 - \frac{2\lambda}{1+\lambda}\right) + |\mathcal{X}||\mathcal{Y}| \log(n+1)\right) \sim O(\sqrt{n}\log n), \quad (5.125)$$

and

$$\left|\mathbb{P}(X_t = x, Y_t = y) - \mathbb{P}(X_t = x)\mathbb{P}(Y_t = y)\right| \leq 2n^{-1/4}, \quad (5.126)$$

for any $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $t = 1, ..., n$. By dividing both sides of equations (5.120), (5.121) and (5.124) by n, and defining appropriate functions, the following can be concluded

$$\frac{1}{n}\log M_X \leq \frac{1}{n}\sum_{t=1}^{n} I(X_t \wedge Z_t | Y_t) + C(\lambda)\frac{o(n)}{n}, \quad (5.127a)$$

$$\frac{1}{n}\log M_Y \leq \frac{1}{n}\sum_{t=1}^{n} I(Y_t \wedge Z_t | X_t) + C(\lambda)\frac{o(n)}{n}, \quad (5.127b)$$

$$\frac{1}{n}\log\left(M_X M_Y\right) \leq \frac{1}{n}\sum_{t=1}^{n} I(X_t Y_t \wedge Z_t) + C(\lambda)\frac{o(n)}{n}. \quad (5.127c)$$

The expressions in (5.127a)-(5.127c) are the averages of the mutual informations calculated at the empirical distributions in the column t of the mentioned subcode. We can rewrite these equations with the new random variable U, where U is distributed uniformly on $\{1, 2, ..., n\}$. Using the same method as Cover [13, pp. 402], we obtain the result. The only thing remained to be found is the distribution under which we calculate the mutual informations. However, by (5.126)

$$\begin{aligned}
&|\mathbb{P}(X_U = x, Y_U = y|U = u) - \mathbb{P}(X_U = x|U = u)\mathbb{P}(Y_U = y|U = u)| \\
&\qquad = |\mathbb{P}(X_u = x, Y_u = y) - \mathbb{P}(X_u = x)\mathbb{P}(Y_u = y)| \leq 2n^{-1/4}. \quad (5.128)
\end{aligned}$$

Using the continuity of conditional mutual information with respect to distributions, and by using the idea of [6, pp. 722], we conclude that, if two distributions are close, the conditional mutual informations, calculated based on them, cannot be too far. More precisely, we can say that there exists a sequence $\{\delta_n\}_{n=1}^{\infty}$, $\delta_n \to 0$ as $n \to \infty$, such that,

$$\frac{1}{n}\log M_X \leq \frac{1}{n}\sum_{t=1}^{n} I(X_t \wedge Z_t|Y_t, U) + C(\lambda)\frac{o(n)}{n} + \delta_n, \quad (5.129a)$$

$$\frac{1}{n}\log M_Y \leq \frac{1}{n}\sum_{t=1}^{n} I(Y_t \wedge Z_t|X_t, U) + C(\lambda)\frac{o(n)}{n} + \delta_n, \quad (5.129b)$$

$$\frac{1}{n}\log (M_X M_Y) \leq \frac{1}{n}\sum_{t=1}^{n} I(X_t Y_t \wedge Z_t|U) + C(\lambda)\frac{o(n)}{n} + \delta_n. \quad (5.129c)$$

By defining new random variables $X \triangleq X_U$, $Y \triangleq Y_U$ and $Z \triangleq Z_U$, whose distributions depend on U in the same way as the distributions of X_t, Y_t and Z_t depend on t, (5.129a)-(5.129c) can be written as

$$\frac{1}{n}\log M_X \leq I(X \wedge Z|Y, U) + C(\lambda)\frac{o(n)}{n} + \delta_n, \quad (5.130a)$$

$$\frac{1}{n}\log M_Y \leq I(Y \wedge Z|X, U) + C(\lambda)\frac{o(n)}{n} + \delta_n, \quad (5.130b)$$

$$\frac{1}{n}\log (M_X M_Y) \leq I(XY \wedge Z|U) + C(\lambda)\frac{o(n)}{n} + \delta_n. \quad (5.130c)$$

Here, the mutual informations are calculated based on $p(u)p(x|u)p(y|u)W(z|x,y)$. On the other hand, the joint probability distribution of X and Y is

$$\begin{aligned}\mathbb{P}(X=x,Y=y) &= \sum_{(i,j)\in\mathcal{A}} \mathbb{P}\left(X(W_1)=x, Y(W_2)=y, W_1=i, W_2=j\right)\\ &= \sum_{(i,j)\in\mathcal{A}} \mathbb{P}\left(X(i)=x, Y(j)=y\right)\mathbb{P}(i,j)\\ &= \frac{1}{|\mathcal{A}|}\sum_{(i,j)\in\mathcal{A}} \mathbb{P}\left(X(i)=x, Y(j)=y\right)\\ &= \frac{1}{|\mathcal{A}|}\sum_{(i,j)\in\mathcal{A}} \frac{1}{n}\sum_{u=1}^{n} 1\{X_u(i)=x, Y_u(j)=y\}. \end{aligned} \tag{5.131}$$

However, all codeword pairs have the same joint type P^n_{XY}, hence,

$$\sum_{u=1}^{n} 1\{X_u(i)=x, Y_u(j)=y\} = nP^n_{XY}(x,y). \tag{5.132}$$

By combining (5.131) and (5.132), it can be concluded that

$$\mathbb{P}(X=x,Y=y) = P^n_{XY}(x,y). \tag{5.133}$$

Finally, we can conclude that

$$P(u,x,y,z) = p(u)p(x|u)p(y|u)W(z|x,y), \tag{5.134}$$

in which the marginal distribution of X and Y is $P^n_{XY}(x,y)$.

The cardinality bound on the time-sharing random variable, U, is the consequence of Carathéodory's theorem on the convex set [21], [32], [13].

5.6.5 Proof of Theorem 5.3.3

To show the result, we must first prove the following theorem:

Lemma 5.6.6. *For any fixed $P \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, rate pair (R_X, R_Y),*

$$\lim_{n\to\infty} \min_{V \in D^n(P)} D(V||W|P) = \min_{V \in D(P)} D(V||W|P), \tag{5.135}$$

where

$$\begin{aligned} D(P) &\triangleq \{V : (R_X, R_Y) \notin C_V(P)\} \\ D^n(P) &\triangleq \{V : (R_X, R_Y) \notin C_V^n(P)\}. \end{aligned} \tag{5.136}$$

Proof. Define $\alpha_n \triangleq \min_{V \in D^n(P)} D(V||W|P)$, and $\alpha^* \triangleq \min_{V \in D(P)} D(V||W|P)$. Moreover, suppose α^* is achieved by V^*. Since $\{\alpha_n\}_{n=1}^{\infty}$ is a decreasing sequence and it is bounded from below ($\alpha_n \geq \alpha^*$), therefore it has a limit. Suppose the limit is not equal to α^*. Therefore, there exist a $\delta > 0$, such that for infinitely many n,

$$|\alpha_n - \alpha^*| \geq \delta. \tag{5.137}$$

Hence, for infinitely many n,

$$D(V||W|P) - \alpha^* \geq \delta \qquad\qquad \forall\, V \in D^n(P) \tag{5.138}$$

which implies that V^* cannot belong to $D^n(P)$ for infinitely many n, i.e., for infinitely many n,

$$(R_X, R_Y) \in C_{V^*}^n(P). \tag{5.139}$$

Since $V^* \in D(P)$,

$$(R_X, R_Y) \notin C_{V^*}(P). \tag{5.140}$$

Therefore $C_{V^*}^n(P)$ cannot converge to $C_{V^*}(P)$, which is a contradiction. $\square$

Since C is an $(n, M_X =, M_Y)$ multi-user code, it can be concluded that it must have at least a dominant type, $P^n_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$. Consider an arbitrary DM-MAC $V : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, such that $(R_X, R_Y) \notin C^n_V(P^n_{XY})$. By Theorem 5.3.2, for channel V, C cannot be an (n, M_X, M_Y, λ)-code for any $0 \le \lambda < 1$. Choose $\bar{\lambda} < 1$ satisfying

$$\frac{1+\bar{\lambda}}{2} > 1 - \frac{\delta}{2}. \tag{5.141}$$

Since P^n_{XY} is a dominant type of code C,

$$\left| C \cap T_{P^n_{XY}} \right| \ge \frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} M_X M_Y. \tag{5.142}$$

On the other hand, since $(R_X, R_Y) \notin C^n_V(P^n_{XY})$, it can be concluded that

$$\left| \mathcal{E}_V(C, P^n_{XY}, \bar{\lambda}) \right| < \frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} \left(1 - \frac{2\bar{\lambda}}{1+\bar{\lambda}} \right) M_X M_Y. \tag{5.143}$$

By combining (5.142) and (5.143), it can be concluded that

$$\left| \mathcal{D}_V(C, P^n_{XY}, \bar{\lambda}) \right| \ge \frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} \left(\frac{2\bar{\lambda}}{1+\bar{\lambda}} \right) M_X M_Y, \tag{5.144}$$

where $\mathcal{D}_V(C, P^n_{XY}, \bar{\lambda})$ is defined as

$$\begin{aligned} \mathcal{D}_V(C, P^n_{XY}, \bar{\lambda}) &\triangleq \left(C \cap T_{P^n_{XY}} \right) / \, \mathcal{E}_V(C, P^n_{XY}, \bar{\lambda}) \\ &= \left\{ (\mathbf{x}_i, \mathbf{y}_j) \in C \cap T_{P^n_{XY}} : V(D^c_{ij} | \mathbf{x}_i, \mathbf{y}_j) > \frac{1+\bar{\lambda}}{2} \right\}. \end{aligned} \tag{5.145}$$

By combining (5.141), (5.145) and using the same method as Csiszar in [16, pp. 167], we have

$$\begin{aligned} W^n(D^c_{ij} | \mathbf{x}_i, \mathbf{y}_j) &\ge \exp\left\{ - \frac{D(V||W|P^n_{XY}) + h(1-\frac{\delta}{2})}{1 - \frac{\delta}{2}} \right\} \\ &\ge \frac{1}{2} \exp\left\{ -nD(V||W|P^n_{XY})(1+\delta) \right\} \quad \text{for all } (\mathbf{x}_i, \mathbf{y}_j) \in \mathcal{D}_V(C, P^n_{XY}, \bar{\lambda}), \end{aligned} \tag{5.146}$$

for small enough δ satisfying $h(1-\frac{\delta}{2}) < 1 - \frac{\delta}{2}$. The average error probability of the code C over the channel W can be written as

$$\begin{aligned} e(C,W) &= \frac{1}{M_X M_Y} \sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} W^n(D_{ij}^c|\mathbf{x}_i, \mathbf{y}_j) \\ &\geq \frac{1}{M_X M_Y} \sum_{\substack{i,j: \\ (\mathbf{x}_i,\mathbf{y}_j)\in\mathcal{D}_V(C,P_{XY}^n,\bar{\lambda})}} \frac{1}{2} \exp\{-nD(V||W|P_{XY}^n)(1+\delta)\} \\ &\geq \frac{1}{(n+1)^{|\mathcal{X}||\mathcal{Y}|}} \left(\frac{\bar{\lambda}}{1+\bar{\lambda}}\right) \exp\{-nD(V||W|P_{XY}^n)(1+\delta)\}. \end{aligned} \tag{5.147}$$

Since the inequality (5.147) holds for all $V : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ satisfying $(R_X, R_Y) \notin C_V^n(P_{XY}^n)$, it can be concluded that

$$\begin{aligned} e(C,W) &\geq \max_{V:(R_X,R_Y)\notin C_V^n(P_{XY}^n)} \exp\{-n[D(V||W|P_{XY}^n)(1+\delta)+\delta]\} \\ &= \exp\{-n[\min_{V:(R_X,R_Y)\notin C_V^n(P_{XY}^n)} D(V||W|P_{XY}^n)(1+\delta)+\delta]\}, \\ &\geq \min_{P_{XY}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y}):} \exp\{-n[\min_{V:(R_X,R_Y)\notin C_V^n(P_{XY})} D(V||W|P_{XY})(1+\delta)+\delta]\}, \end{aligned} \tag{5.148}$$

for sufficiently large n. Using Lemma 5.6.6, we conclude that for sufficiently large n,

$$e(C,W) \geq \min_{P_{XY}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})} \exp\{-n[\min_{V:(R_X,R_Y)\notin C_V(P_{XY})} D(V||W|P_{XY})(1+\delta)+\delta]\}, \tag{5.149}$$

which completes the proof.

5.6.6 Proof of Theorem 5.4.1

Theorem 5.6.1. *For any nonnegative-definite channel, W, the minimum distance of any multiuser code, $C = C_X \times C_Y$, with rate pair (R_X, R_Y) satisfies*

$$d_B(C) \leq E_M(R_X, R_Y, W), \tag{5.150}$$

where $E_M(R_X, R_Y, W)$ is defined as

$$E_M(R_X, R_Y, W) \triangleq \max_{P_{UXY}} \min_{\beta=X,Y,XY} E_{U,1}^{\beta}(R_X, R_Y, W, P_{XYU}). \tag{5.151}$$

The maximum is taken over all $P_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y})$ such that $X - U - Y$, and $R_X \leq H(X|U)$ and $R_Y \leq H(Y|U)$. The functions $E_M^{\beta}(R_X, R_Y, W, P_{XYU})$ are defined as follows:

$$\begin{aligned} E_M^X(R_X, R_Y, W, P_{XYU}) &\triangleq \min_{V_{X\tilde{X}\hat{X}Y} \in \mathcal{V}_X^M} \mathbb{E} d_W\big((\hat{X}, Y), (\tilde{X}, Y)\big), \\ E_M^Y(R_X, R_Y, W, P_{XYU}) &\triangleq \min_{V_{XY\tilde{Y}\hat{Y}} \in \mathcal{V}_Y^M} \mathbb{E} d_W\big((X, \hat{Y}), (X, \tilde{Y})\big), \\ E_M^{XY}(R_X, R_Y, W, P_{XYU}) &\triangleq \min_{V_{XY\tilde{X}\tilde{Y}\hat{X}\hat{Y}} \in \mathcal{V}_{XY}^M} \mathbb{E} d_W\big((\hat{X}, \hat{Y}), (\tilde{X}, \tilde{Y})\big). \end{aligned} \tag{5.152}$$

where

$$\begin{aligned} \mathcal{V}_X^M \triangleq \Big\{ & V_{X\tilde{X}\hat{X}Y} : V_{\tilde{X}Y} = V_{\hat{X}Y} = V_{XY} = P_{XY},\ \hat{X} - XY - \tilde{X} \\ & V_{\tilde{X}|XY} = V_{\hat{X}|XY},\ I(X \wedge \tilde{X}|Y) = I(X \wedge \hat{X}|Y) \leq R_X \Big\}, \end{aligned} \tag{5.153}$$

$$\begin{aligned} \mathcal{V}_Y^M \triangleq \Big\{ & V_{XY\tilde{Y}\hat{Y}} : V_{X\tilde{Y}} = V_{X\hat{Y}} = V_{XY} = P_{XY},\ \hat{Y} - XY - \tilde{Y} \\ & V_{\tilde{Y}|XY} = V_{\hat{Y}|XY},\ I(Y \wedge \tilde{Y}|X) = I(Y \wedge \hat{Y}|X) \leq R_Y \Big\}, \end{aligned} \tag{5.154}$$

$$\begin{aligned} \mathcal{V}_{XY}^M \triangleq \Big\{ & V_{XY\tilde{X}\tilde{Y}\hat{X}\hat{Y}} : V_{\tilde{X}\tilde{Y}} = V_{\hat{X}\hat{Y}} = V_{XY} = P_{XY},\ \hat{X}\hat{Y} - XY - \tilde{X}\tilde{Y} \\ & V_{\tilde{X}\tilde{Y}|XY} = V_{\hat{X}\hat{Y}|XY},\ I(XY \wedge \tilde{X}\tilde{Y}) = I(XY \wedge \hat{X}\hat{Y}) \leq R_X + R_Y \Big\}. \end{aligned} \tag{5.155}$$

Proof. Consider any joint composition $V_{XY\tilde{X}\tilde{Y}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{Y})$ with marginal distributions $V_{XY} = V_{\tilde{X}\tilde{Y}} = P_{XY}$. In the following lemma, we find the average number of pairs of codewords in a spherical collection defined by joint type $V_{XY\tilde{X}\tilde{Y}}$ about an arbitrary pair of sequences $(\mathbf{x}, \mathbf{y}) \in T_{P_{XY}}$. For such $(\mathbf{x}, \mathbf{y})$, which is not necessarily a pair of codewords, let us define the following sets:

- $A_X(\mathbf{x}, \mathbf{y}) \triangleq \big\{(\mathbf{x}, \tilde{\mathbf{y}}) \in C : (\mathbf{x}, \mathbf{y}, \mathbf{x}, \tilde{\mathbf{y}}) \in T_{V_{XY\tilde{X}\tilde{Y}}}\big\}$
- $A_Y(\mathbf{x}, \mathbf{y}) \triangleq \big\{(\tilde{\mathbf{x}}, \mathbf{y}) \in C : (\mathbf{x}, \mathbf{y}, \tilde{\mathbf{x}}, \mathbf{y}) \in T_{V_{XY\tilde{X}\tilde{Y}}}\big\}$

- $A_{XY}(\mathbf{x},\mathbf{y}) \triangleq \left\{(\tilde{\mathbf{x}},\tilde{\mathbf{y}}) \in C : (\mathbf{x},\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{y}}) \in T_{V_{XY\tilde{X}\tilde{Y}}}\right\}$

Note that, if $\mathbf{x} \notin C_X$ or $X \neq \tilde{X}$, the first set would be empty. Similarly, if $\mathbf{y} \notin C_Y$ or $Y \neq \tilde{Y}$, the second one would be an empty set.

Lemma 5.6.7. *Consider the multi-user code, C with a dominant joint type P_{XY}. Additionally, consider any distribution $V_{XY\tilde{X}\tilde{Y}} \in \mathcal{P}\big((\mathcal{X} \times \mathcal{Y})^2\big)$, satisfying $V_{XY} = V_{\tilde{X}\tilde{Y}} = P_{XY}$. Then, there exists a pair of sequences $(\mathbf{x},\mathbf{y}) \in T_{P_{XY}}$ such that*

$$|A_{XY}(\mathbf{x},\mathbf{y})| \geq \exp\{n[R_X + R_Y - I(\tilde{X}\tilde{Y} \wedge XY)]\}. \tag{5.156}$$

Also, for any distribution $V_{XY\tilde{X}} \in \mathcal{P}\big(\mathcal{X} \times \mathcal{Y} \times \mathcal{X}\big)$ satisfying $V_{XY} = V_{\tilde{X}Y} = P_{XY}$, and any $\mathbf{y} \in C_Y \cap T_{P_Y}$, there exists a $\mathbf{x} \in T_{P_X}$, such that $(\mathbf{x},\mathbf{y}) \in T_{P_{XY}}$, and

$$|A_Y(\mathbf{x},\mathbf{y})| \geq \exp\{n[R_X - I(\tilde{X} \wedge X|Y)]\}. \tag{5.157}$$

Similarly, for any distribution $V_{XY\tilde{Y}} \in \mathcal{P}\big(\mathcal{X} \times \mathcal{Y} \times \mathcal{Y}\big)$ satisfying $V_{XY} = V_{X\tilde{Y}} = P_{XY}$, and any $\mathbf{x} \in C_X \cap T_{P_X}$, there exists a sequence $\mathbf{y} \in T_{P_Y}$ such that $(\mathbf{x},\mathbf{y}) \in T_{P_{XY}}$, and

$$|A_X(\mathbf{x},\mathbf{y})| \geq \exp\{n[R_Y - I(\tilde{Y} \wedge Y|X)]\}. \tag{5.158}$$

Proof. For a fixed $V_{XY\tilde{X}\tilde{Y}}$, let us study the spherical collection consisting of all pairs of codewords sharing composition $V_{XY\tilde{X}\tilde{Y}}$ with some arbitrary pair of sequences in $T_{P_{XY}}$. Consider such spherical collection for every pair of sequences. Since each of the codeword pairs shares joint composition $V_{XY\tilde{X}\tilde{Y}}$ with $\exp\{H(\tilde{X}\tilde{Y}|XY)\}$ pair of sequences, it must belong to $\exp\{H(\tilde{X}\tilde{Y}|XY)\}$ different spherical collections. Therefore,

$$\sum_{(\mathbf{x},\mathbf{y}) \in T_{P_{XY}}} |A_{XY}(\mathbf{x},\mathbf{y})| \approx \exp\{n[R_X + R_Y + H(\tilde{X}\tilde{Y}|XY)]\}$$

Hence, by dividing both sides of the previous equality by $|T_{P_{XY}}|$, we conclude that

$$\frac{1}{|T_{P_{XY}}|} \sum_{(\mathbf{x},\mathbf{y}) \in T_{P_{XY}}} |A_{XY}(\mathbf{x},\mathbf{y})| \approx 2^{n[R_X + R_Y - I(\tilde{X}\tilde{Y} \wedge XY)]}.$$

Thus, there must exist a pair of sequence, $(\mathbf{x}, \mathbf{y}) \in T_{P_{XY}}$, with

$$|A_{XY}(\mathbf{x}, \mathbf{y})| \gtrapprox \exp\{n[R_X + R_Y - I(\tilde{X}\tilde{Y} \wedge XY)]\}. \tag{5.159}$$

By a similar argument, we can conclude (5.157) and (5.158). $\square$

Lemma 5.6.8. *Fix $\epsilon > 0$. Let W be a nonnegative-definite channel. Let $C = C_X \times C_Y$ be any multi-user code with dominant composition nP_{XY} and rate pair (R_X, R_Y). Consider any distribution $V_{XY\tilde{X}\tilde{Y}} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{Y})$ satisfying the following constraints:*

- $V_{XY} = V_{\tilde{X}\tilde{Y}} = P_{XY}$
- $I_V(XY \wedge \tilde{X}\tilde{Y}) \leq R_X + R_Y - \epsilon,$

Then, C has two pairs of codewords, $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ and $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$, such that

$$d_B\big((\tilde{\mathbf{x}}, \tilde{\mathbf{y}}), (\hat{\mathbf{x}}, \hat{\mathbf{y}})\big) \leq (1+\epsilon)\mathbb{E} d_B\big((\tilde{X}, \tilde{Y}), (\hat{X}, \hat{Y})\big), \tag{5.160}$$

where the expectation is calculated based on $V_{XY\tilde{X}\tilde{Y}\hat{X}\hat{Y}} \in \mathcal{P}((\mathcal{X} \times \mathcal{Y})^3)$ satisfying

- $V_{XY} = V_{\tilde{X}\tilde{Y}} = V_{\hat{X}\hat{Y}} = P_{XY}$
- $\tilde{X}\tilde{Y} - XY - \hat{X}\hat{Y}$
- $V_{\tilde{X}\tilde{Y}|XY} = V_{\hat{X}\hat{Y}|XY}$
- $I_V(XY \wedge \tilde{X}\tilde{Y}) \leq R_X + R_Y - \epsilon.$

Moreover, for any $V_{XY\tilde{X}} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{X})$ satisfying the following constraints:

- $V_{XY} = V_{\tilde{X}Y} = P_{XY}$
- $I_V(X \wedge \tilde{X}|Y) \leq R_X - \epsilon,$

C has two pairs of codewords, $(\tilde{\mathbf{x}}, \mathbf{y})$ and $(\hat{\mathbf{x}}, \mathbf{y})$, such that

$$d_B\big((\tilde{\mathbf{x}}, \mathbf{y}), (\hat{\mathbf{x}}, \mathbf{y})\big) \leq (1+\epsilon)\mathbb{E} d_B\big((\tilde{X}, Y), (\hat{X}, Y)\big), \tag{5.161}$$

where the expectation is calculated based on $V_{XY\tilde{X}\hat{X}} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{X})$ *satisfying*

- $V_{XY} = V_{\tilde{X}Y} = V_{\hat{X}Y} = P_{XY}$
- $\tilde{X} - XY - \hat{X}$
- $V_{\tilde{X}|XY} = V_{\hat{X}|XY}$
- $I_V(X \wedge \tilde{X}|Y) \leq R_X - \epsilon.$

Similarly, for any $V_{XY\tilde{Y}} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Y})$ *satisfying the following constraints:*

- $V_{XY} = V_{\tilde{Y}} = P_{XY}$
- $I_V(Y \wedge \tilde{Y}|X) \leq R_Y - \epsilon.$

C *has two pairs of codewords,* $(\mathbf{x}, \tilde{\mathbf{y}})$ *and* $(\mathbf{x}, \hat{\mathbf{y}})$*, such that*

$$d_B\big((\mathbf{x}, \tilde{\mathbf{y}}), (\mathbf{x}, \hat{\mathbf{y}})\big) \leq (1+\epsilon)\mathbb{E}d_B\big((X, \tilde{Y}), (X, \hat{Y}))\big), \tag{5.162}$$

where the expectation is calculated based on $V_{XY\tilde{Y}\hat{Y}} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Y} \times \mathcal{Y})$ *satisfying*

- $V_{XY} = V_{X\tilde{Y}} = V_{X\hat{Y}} = P_{XY}$
- $\tilde{Y} - XY - \hat{Y}$
- $V_{\tilde{Y}|XY} = V_{\hat{Y}|XY}$
- $I_V(Y \wedge \tilde{Y}|X) \leq R_Y - \epsilon.$

Proof. Consider the joint type $V_{XY\tilde{X}\tilde{Y}}$ for which we have the following properties

- $V_{XY} = V_{\tilde{X}\tilde{Y}} = P_{XY}.$
- $I(\tilde{X}\tilde{Y} \wedge XY) \leq R_X + R_Y - \delta.$

For the moment, let us assume that $X \neq \tilde{X}$ and $Y \neq \tilde{Y}$. Let us choose $(\mathbf{x}, \mathbf{y}) \in T_{P_{XY}}$ whose existence is asserted in the previous lemma. Let us call the spherical collection about $(\mathbf{x}, \mathbf{y}) \in T_{P_{XY}}$, which is defined by $V_{XY\tilde{X}\tilde{Y}}$, as $\mathbf{S}_{XY}$. Also, call the cardinality of this set by T_{XY}, i.e. $|\mathbf{S}_{XY}| = T_{XY}$. From this point, we are going to study the distance structure of the pairs of codewords that lie in $\mathbf{S}_{XY}$. Since we have so many codewords in this spherical collection, they cannot be far from one another. First, we calculate the average distance between any two pairs in this spherical collection. The average distance is given by

$$d_{av}^{XY} = \frac{1}{T_{XY}(T_{XY}-1)} d_{tot}$$

where d_{tot} is obtained by adding up all unordered distances between any two not necessarily distinct pairs of codewords in $\mathbf{S}_{XY}$. In the other words, d_{tot} is defined as

$$d_{tot}^{XY} = \sum_{(\hat{\mathbf{x}},\hat{\mathbf{y}})\in\mathbf{S}_{XY}} \sum_{(\tilde{\mathbf{x}},\tilde{\mathbf{y}})\in\mathbf{S}_{XY}} d_B\big((\hat{\mathbf{x}},\hat{\mathbf{y}}),(\tilde{\mathbf{x}},\tilde{\mathbf{y}})\big)$$

where $(\tilde{\mathbf{x}},\tilde{\mathbf{y}})$ and $(\hat{\mathbf{x}},\hat{\mathbf{y}})$ are not necessarily distinct pairs. Therefore,

$$d_{av}^{XY} = \frac{1}{n}\frac{1}{T_{XY}(T_{XY}-1)} \sum_{\substack{(\hat{\mathbf{x}},\hat{\mathbf{y}}),\\(\tilde{\mathbf{x}},\tilde{\mathbf{y}})\in\mathbf{S}_{XY}}} \sum_{\substack{i,k\in\mathcal{X}\\j,l\in\mathcal{Y}}} n_{\hat{\mathbf{x}}\hat{\mathbf{y}}\tilde{\mathbf{x}}\tilde{\mathbf{y}}}(i,j,k,l)\, d_B\big((i,j),(k,l)\big)$$

where $n_{\hat{\mathbf{x}}\hat{\mathbf{y}}\tilde{\mathbf{x}}\tilde{\mathbf{y}}}(i,j,k,l) \triangleq nP_{\hat{\mathbf{x}}\hat{\mathbf{y}}\tilde{\mathbf{x}}\tilde{\mathbf{y}}}(i,j,k,l)$, and $P_{\hat{\mathbf{x}}\hat{\mathbf{y}}\tilde{\mathbf{x}}\tilde{\mathbf{y}}}$ is the joint composition of $(\hat{\mathbf{x}},\hat{\mathbf{y}})$ and $(\tilde{\mathbf{x}},\tilde{\mathbf{y}})$. Furthermore, define the variable $n_{\hat{\mathbf{x}}\hat{\mathbf{y}}\tilde{\mathbf{x}}\tilde{\mathbf{y}}}(i,j,k,l|p)$ as follows:

$$n_{\hat{\mathbf{x}}\hat{\mathbf{y}}\tilde{\mathbf{x}}\tilde{\mathbf{y}}}(i,j,k,l|p) = \begin{cases} 1 & \text{if } (\hat{\mathbf{x}})_p = i, (\hat{\mathbf{y}})_p = j, (\tilde{\mathbf{x}})_p = k, (\tilde{\mathbf{y}})_p = l \\ 0 & \text{otherwise} \end{cases}$$

Hence, the average distance can be written as

$$d_{av}^{XY} = \frac{1}{n}\frac{1}{T_{XY}(T_{XY}-1)} \sum_{p} \sum_{\substack{(\hat{\mathbf{x}},\hat{\mathbf{y}}),\\(\tilde{\mathbf{x}},\tilde{\mathbf{y}})\in\mathbf{S}_{XY}}} \sum_{\substack{i,k\in\mathcal{X}\\j,l\in\mathcal{Y}}} n_{\hat{\mathbf{x}}\hat{\mathbf{y}}\tilde{\mathbf{x}}\tilde{\mathbf{y}}}(i,j,k,l|p)\, d_B\big((i,j),(k,l)\big)$$

Let $T_{(i,j)|p}$ be the number of $(\mathbf{x},\mathbf{y}) \in \mathbf{S}_{XY}$ with $(\mathbf{x})_p = i$, and $(\mathbf{y})_p = j$. It can be shown that the previous can be written as

$$d_{av}^{XY} = \frac{1}{n}\frac{T_{XY}}{T_{XY}-1}\sum_{p}\sum_{\substack{i,k\in\mathcal{X}\\ j,l\in\mathcal{Y}}}\frac{T_{(i,j)|p}T_{(k,l)|p}}{T_{XY}^2}d_B\big((i,j),(k,l)\big).$$

Moreover, Let us define $\lambda_{(i,j)|p}$ as the fraction of the pairs in $\mathbf{S}_{XY}$ with an (i,j) in their p-th component, i.e.,

$$\lambda_{(i,j)|p} \triangleq \frac{T_{(i,j)|p}}{T_{XY}}. \tag{5.163}$$

Therefore, d_{av}^{XY} can be written as

$$d_{av}^{XY} = \frac{1}{n}\frac{T_{XY}}{T_{XY}-1}\sum_{p}\sum_{\substack{i,k\in\mathcal{X}\\ j,l\in\mathcal{Y}}}\lambda_{(i,j)|p}\lambda_{(k,l)|p}d_B\big((i,j),(k,l)\big). \tag{5.164}$$

In general, λ is an unknown function. However, it must satisfy the following equality

$$\sum_{i\in\mathcal{X},j\in\mathcal{Y}}\lambda_{(i,j)|p} = 1 \qquad \text{for all } p. \tag{5.165}$$

For the center of the sphere, $(\mathbf{x},\mathbf{y})$, we define $\gamma_{(i,j)|p}$ as

$$\gamma_{(i,j)|p} = \begin{cases} 1 & \text{if } (\mathbf{x})_p = i,\ (\mathbf{y})_p = j \\ 0 & \text{otherwise} \end{cases}.$$

On the other hand, a valid λ must satisfy the following constraint:

$$\sum_{p}\lambda_{(i,j)|p}\gamma_{(k,l)|p} = n_{XY\tilde{X}\tilde{Y}}(k,l,i,j) \tag{5.166}$$

for all $i,k\in\mathcal{X}$ and all $j,l\in\mathcal{Y}$. Therefore, we can upper bound d_{av} with

$$d_{av}^{XY} \le \frac{1}{n}\frac{T_{XY}}{T_{XY}-1}\max_{\lambda}\sum_{p}\sum_{\substack{i,k\in\mathcal{X}\\ j,l\in\mathcal{Y}}}\lambda_{(i,j)|p}\lambda_{(k,l)|p}d_B\big((i,j),(k,l)\big). \tag{5.167}$$

where the maximization is taken over all λ satisfying (5.165) and (5.166). In the following lemma, we will find the maximum.

Lemma 5.6.9. *Suppose that W is a nonnegative-definite channel. The average distance between the T_{XY} pairs of codewords in the spherical collection, defined by joint composition $V_{XY\tilde{X}\tilde{Y}}$, satisfies*

$$d_{av}^{XY} \leq \frac{T_{XY}}{T_{XY}-1} \sum_{\substack{i,k\in\mathcal{X}\\ j,l\in\mathcal{Y}}} \sum_{\substack{r\in\mathcal{X},\\ s\in\mathcal{Y}}} f_{XY}\big((i,j),(k,l),(r,s)\big). \tag{5.168}$$

where $f_{XY}\big((i,j),(k,l),(r,s)\big)$ is defined as

$$f_{XY}\big((i,j),(k,l),(r,s)\big) \triangleq \frac{n_{XY\tilde{X}\tilde{Y}}(r,s,i,j)n_{XY\tilde{X}\tilde{Y}}(r,s,k,l)}{n.n_{XY}(r,s)} d_B\big((i,j),(k,l)\big). \tag{5.169}$$

Proof. Let

$$\lambda^*_{(i,j)|p} = \sum_{k\in\mathcal{X},l\in\mathcal{Y}} \frac{n_{XY\tilde{X}\tilde{Y}}(k,l,i,j)}{n_{XY}(k,l)} \gamma_{(k,l)|p} \tag{5.170}$$

We are going to prove that λ^* achieves the maximum. It is easy to clarify that λ^* satisfies (5.165) and (5.166). Moreover, for all λ satisfying (5.165) and (5.166),

$$\begin{aligned} \sum_p \lambda^*_{(i,j)|p}\lambda_{(k,l)|p} &= \sum_{r\in\mathcal{X},s\in\mathcal{Y}} \frac{n_{XY\tilde{X}\tilde{Y}}(r,s,i,j)}{n_{XY}(r,s)} \sum_p \gamma_{(r,s)|p}\lambda_{(k,l)|p} \\ &= \sum_{r\in\mathcal{X},s\in\mathcal{Y}} \frac{n_{XY\tilde{X}\tilde{Y}}(r,s,i,j)n_{XY\tilde{X}\tilde{Y}}(r,s,k,l)}{n.n_{XY}(r,s)} \end{aligned} \tag{5.171}$$

By assuming that the channel is nonnegative definite, and by using a similar argument as [10, Lemma 6], we can show that λ^* achieves the maximum. Substituting this value for λ completes the proof. □

Now, let us fix a joint type $V_{XY\tilde{X}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}\times\mathcal{X})$ for which we have the following properties

- $V_{XY} = V_{\tilde{X}Y} = P_{XY}$
- $I(\tilde{X} \wedge X|Y) \leq R_X - \delta$

Let us choose any $\mathbf{y} \in C_Y \cap T_{P_Y}$. By Lemma 5.6.7, there exists a sequence $\mathbf{x} \in T_{P_X}$ such that $(\mathbf{x}, \mathbf{y}) \in T_{P_{XY}}$, and the spherical collection about $(\mathbf{x}, \mathbf{y})$, defined by $V_{XY\tilde{X}}$ has many pairs of codewords. Let us call such a sphere as $\mathbf{S}_Y$. Assume that $|\mathbf{S}_Y| = T_Y$. We denote the average distance between any two pairs of codeword belonging to this spherical collection by d_{av}^Y. Using a similar argument to the one in Lemma 5.6.9, it can be shown that d_{av}^Y is bounded from above by

$$d_{av}^Y \leq \frac{T_Y}{T_Y - 1} \sum_{i,k \in \mathcal{X}} \sum_{r \in \mathcal{X}, s \in \mathcal{Y}} f_Y\big(i, k, (r, s)\big). \tag{5.172}$$

where $f_Y\big(i, k, (r, s)\big)$ is defined as

$$f_Y\big(i, k, (r, s)\big) \triangleq \frac{n_{XY\tilde{X}}(r, s, i) n_{XY\tilde{X}}(r, s, k)}{n.n_{XY}(r, s)} d_B\big((i, j), (k, j)\big). \tag{5.173}$$

Similarly, let's fix a joint type $V_{XY\tilde{Y}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{Y})$ for which we have the following properties

- $V_{XY} = V_{X\tilde{Y}} = P_{XY}$
- $I(\tilde{Y} \wedge Y|X) \leq R_Y - \delta$

Choose any $\mathbf{x} \in C_X \cap T_{P_X}$. By Lemma 5.6.7, there exist a sequence $\mathbf{y} \in T_{P_Y}$ such that $(\mathbf{x}, \mathbf{y}) \in T_{P_{XY}}$ and the spherical collection about $(\mathbf{x}, \mathbf{y})$ defined by $V_{XY\tilde{Y}}$ has many pairs of codewords. Let us call such a sphere as $\mathbf{S}_X$. Assume that $|\mathbf{S}_X| = T_X$. We denote the average distance between any two pairs of codewords belonging to this spherical collection by d_{av}^X. By doing a similar argument as we did before, we can find an upper bound on the d_{av}^X. It can be easily shown that

$$d_{av}^X \leq \frac{T_X}{T_X - 1} \sum_{j,l \in \mathcal{Y}} \sum_{r \in \mathcal{X}, s \in \mathcal{Y}} f_X\big(j, l, (r, s)\big). \tag{5.174}$$

where $f_X\big(j, l, (r, s)\big)$ is defined as

$$f_X\big(j, l, (r, s)\big) \triangleq \frac{n_{XY\tilde{Y}}(r, s, j) n_{XY\tilde{Y}}(r, s, l)}{n.n_{XY}(r, s)} d_B\big((i, j), (i, l)\big), \tag{5.175}$$

which completes the proof of Lemma 5.6.8. □

□

As a result, it can be concluded that for any $V_{XY\tilde{X}\tilde{Y}}$ satisfying the aforementioned constraints, there exists a pair of sequences $(\mathbf{x}, \mathbf{y}) \in T_{P_{XY}}$, such that the spherical collection about $(\mathbf{x}, \mathbf{y})$ and defined by $V_{XY\tilde{X}\tilde{Y}}$ has exponential many codeword pairs around. Therefore, for sufficiently large n,

$$\frac{T_{XY}}{T_{XY}-1} \leq 1+\epsilon \tag{5.176}$$

Therefore by substituting this upper bound, and by simplifying the result of Lemma 5.6.8, we observe that

$$d_{av}^{XY} \leq (1+\epsilon)\mathbb{E}d_B\big((\tilde{X}, \tilde{Y}), (\hat{X}, \hat{Y})\big) \tag{5.177}$$

The expectation is calculated based on $V_{\tilde{X}\tilde{Y}|XY}V_{\hat{X}\hat{Y}|XY}V_{XY}$. Since the average distance between the pairs in $\mathbf{S}^{XY}$ is greater than some number, there must exist at least two pairs of codewords in $\mathbf{S}^{XY}$ satisfying the same constraints. By a similar argument, we can show the correctness of the second and third part of the theorem.

Lemma 5.6.10. *For $\beta = X, Y, XY$, the following quantities are equivalent*

$$E_M^{\beta}(R_X, R_Y, W, P_{XYU}) = E_U^{\beta}(R_X, R_Y, W, P_{XYU}), \tag{5.178}$$

where E_U^{β}s and E_M^{β}s are defined in (5.30) *and* (5.152).

Proof. The proof is a generalized version of the result of [15]. □

Theorem 5.6.2. *For any indivisible channel*

$$E_m^*(R_X, R_Y) \leq d_B^*(R_X, R_Y), \tag{5.179}$$

where $E_m^*(R_X, R_Y)$ is the maximal error reliability function at rate pair (R_X, R_Y).

Proof. The proof is very similar to [10]. □

Therefore, by combining the result of Theorem 5.6.1 and Theorem 5.6.2, the result of Theorem 5.4.1 is concluded.

5.6.7 Proof of Theorem 5.5.1

Without loss of generality, let us assume $R_X \leq R_Y$. The average error probability of any code is always less than or equal to its maximal probability of error. As a result,

$$E_m^* \left(R_X, R_Y\right) \leq E_{av}^* \left(R_X, R_Y\right). \tag{5.180}$$

On the hand, for any $\delta > 0$ and sufficiently large n, there exists an (n, R_X, R_Y) code, $C = C_X \times C_Y$, satisfying the following inequality

$$e(C, W) \leq 2^{-n(E_{av}^*(R_X, R_Y) - \delta)}, \tag{5.181}$$

which can be written as

$$\frac{1}{M_Y} \sum_{j=1}^{M_Y} \left\{ \frac{1}{M_X} \sum_{i=1}^{M_X} e_{ij}(C, W) \right\} \leq 2^{-n(E_{av}^*(R_X, R_Y) - \delta)}. \tag{5.182}$$

It can be concluded that, for $M_Y^* \geq \frac{M_Y}{2}$ codewords in C_Y, the following holds

$$\frac{1}{M_X} \sum_{i=1}^{M_X} e_{ij}(C, W) \leq 2 \times 2^{-n(E_{av}^*(R_X, R_Y) - \delta)}, \qquad \text{for all } j = 1, 2, ..., M_Y^*. \tag{5.183}$$

Here, without loss of generality, we assumed that these codewords are the first M_Y^* codewords in C_Y. By using (5.183), it can be concluded that

$$e_{ij}(C, W) \leq 2 \times 2^{-n(E_{av}^*(R_X, R_Y) - R_X - \delta)}, \qquad \text{for all } j = 1, 2, ..., M_Y^*, i = 1, 2, ..., M_X, \tag{5.184}$$

and therefore

$$e_m(C^*, W) \leq 2 \times 2^{-n(E_{av}^*(R_X,R_Y)-R_X-\delta)}, \tag{5.185}$$

where

$$C^* \triangleq \{(\mathbf{x}_i, \mathbf{y}_j) : i = 1, 2, ..., M_X, j = 1, 2, ..., M_Y^*\}. \tag{5.186}$$

Note that,

$$e_m(C^*, W) \geq 2^{-n(E_m^*(R_X,R_Y-\delta)+\delta)} \geq 2^{-n(E_m^*(R_X,R_Y)+2\delta)}. \tag{5.187}$$

By combining (5.185) and (5.187), we conclude that

$$E_m^*(R_X, R_Y) \geq E_{av}^*(R_X, R_Y) - R_X. \tag{5.188}$$

Similarly, it can be shown that

$$E_{av}^*(R_X, R_Y) \leq E_m^*(R_X, R_Y) + R_X. \tag{5.189}$$

CHAPTER 6

Conclusions

This work addresses the problem of communication over a multiple-access channel (MAC) without feedback in the discrete memoryless setting. We consider the error exponents for this channel model and obtain upper and lower bounds on the channel reliability function.

In Chapter 3, we study a unified framework to obtain all known lower bounds (random coding, typical random coding and expurgated bound) on the reliability function of a point-to-point discrete memoryless channel. We show that the typical random coding bound is the typical performance of the constant composition code ensemble. By using a similar idea with a two-user discrete memoryless multiple-access channel, we derive three lower bounds on the reliability function. The first one (random coding) is identical to the best known lower bound on the reliability function of DM-MAC. We also showed that the random coding bound is the average performance of the constant composition code ensemble. The second bound (typical random coding) is the typical performance of the constant composition code ensemble. To derive the third bound (expurgated), we eliminate some of the codewords from the codebook with a larger rate. This is the first bound of its type that explicitly uses the method of expurgation in a multi-user transmission system. We show that the exponent of the typical random coding and expurgated bounds are greater than or equal to the exponent of the known random coding bounds for all rate pairs. By numerical evaluation of the random coding and the expurgated bounds for a simple symmetric MAC, we show that, at low rates, the expurgated bound is strictly larger. We also show that all these bounds can be universally obtained for all discrete memoryless MACs with given input and output alphabets.

To obtain upper bounds on the reliability function for DM-MACs, in Chapter 4, we formally characterize the typicality graph and look at some subgraph containment problems. In particular, we answer three questions concerning the typicality graph:

- When can we find subgraphs such that the left and right vertices of the subgraph have specified degrees, say R'_X and R'_Y, respectively ?
- What is the maximum size of subgraphs that are complete, i.e., every left vertex is connected to every right vertex? One main contribution of this chapter is to provide a complete answer to this question.
- If we create a subgraph by randomly picking a specified number of left and right vertices, what is the probability that this subgraph has far fewer edges than expected?

Finally, in Chapter 5, two new upper bounds on the error exponent of a two-user discrete memoryless (DM) multiple-access channel (MAC) are derived. The first bound (sphere packing) is an upper bound on the average error exponent and is the first bound of this type that explicitly imposes independence of the users' input distributions (conditioned on the time-sharing auxiliary variable) and, thus, results in a tighter sphere-packing exponent when compared to the tightest known exponent derived by Haroutunian. The second bound (minimum distance) is an upper bound on the maximal error exponent, not the average. To obtain this bound, we first derive an upper bound on the minimum Bhattacharyya distance between codeword pairs. For a certain large class of two-user (DM) MAC, an upper bound on the maximal error exponent is derived as a consequence of the upper bound on Bhattacharyya distance. Using a conjecture about the structure of the multi-user code, a tighter minimum distance bound for the maximal error exponent is derived and shown to be tight at zero rates. Finally, the relationship between average and maximal error probabilities for a two user (DM) MAC is studied. As a result, a method to derive new bounds on the average/maximal error exponent by using known bounds on the maximal/average error exponent is obtained.

BIBLIOGRAPHY

[1] U. Agustin. Gedachtnisfreie kannale for diskrete zeit. *Z. Wahrscheinlichkelts theory verw*, pages 10–61, 1966.

[2] R. Ahlswede. On two-way communication channels and a problem by Zarankiewics. In *Probl. of Control and Inform. Theory.*

[3] R. Ahlswede. Multi-way communication channels. In *Proc. International Symposium on Information Theory*, 1971.

[4] R. Ahlswede. The capacity region of a channel with two senders and two receivers. *Academic Press*, pages 805–814, 1972.

[5] R. Ahlswede. An elementary proof of the strong converse theorem for the multiple access channel. *Journal of Combinatorics, Information and System sciences*, pages 216–230, 1982.

[6] R. Ahlswede. The rate-distortion for multiple description without excess rates. *IEEE Trans. Information Theory*, 31(6):721–726, Nov. 1985.

[7] N. Alon and J. H. Spencer. *The Probabilistic Method*. John Wiley and Sons, Inc., 1992.

[8] A. Barg and D. Forney. Random codes: Minimum distances and error exponents. *IEEE Trans. Information Theory*, 48(9):2568–2573, Sept. 2002.

[9] E. R. Berlekamp. *Algebraic Coding Theory*. 1968.

[10] R. E. Blahut. Composition bounds for channel block codes. *IEEE Trans. Information Theory*, 23(6):656–674, Nov. 1977.

[11] S. Choi and S. S. Pradhan. A graph-based framework for transmission of correlated sources over broadcast channels. *IEEE Transactions on Information Theory*, 54(7):2841–2856, 2008.

[12] G. Como. Group codes outperform binary coset codes on non-binary symmetric memoryless channels. *IEEE Trans. Information Theory (Submitted)*.

[13] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.

[14] I. Csiszar. The method of types. *IEEE Trans. Information Theory*, 44(6):2505–2523, Oct. 1998.

[15] I. Csiszar and J. Korner. Graph decomposition: A new key to coding theorems. *IEEE Trans. Information Theory*, 1:5–12, Jan. 1981.

[16] I. Csiszar and J. Korner. *Information theory: Coding theorems for Discrete memoryless Systems.* 1981.

[17] R. L. Dobrusin. Asymptotic bounds of the probability of error for the transmission of messages over a discrete memoryless channel with a symmetric transition probability matrix. *Teor. Veroyatnost. i Primenen*, pages 283–311, 1962.

[18] G. Dueck. Maximal error capacity regions are smaller than average error capacity regions for multi-user channels. *Probl. of Control and Inform. Theory*, pages 11–19, 1978.

[19] G. Dueck. The strong converse of the coding theorem for the multi-user channels. *Journal of Combinatorics, Information and System sciences*, pages 187–196, 1981.

[20] A. G. Dyachkov. Random constant composition codes for multiple-access channels. *Probl. of Control and Inform. Theory*, pages 357–369, 1984.

[21] H. G. Eggleston. *Convexity*. Cambridge University Press, Cambridge, UK, 1969.

[22] P. Elias. Coding for noisy channels. *IRE Convention Record*, pages 37–46, 1955.

[23] R. M. Fano. *Transmission of Information: A Statistical Theory of Communication*. MIT Press, 1961.

[24] A. Fedotov, P. Harremoës, and F. Topsøe. Refinements of Pinsker's inequality. *IEEE Trans. Information Theory*, 49(6):1491–1498, June 2003.

[25] A. Feinstein. Error bounds in noisy channels without memory,. *IEEE Trans. Information Theory*, 1(2):13–14, Sept. 1955.

[26] R. G. Gallager. *Transmission of Information: A Statistical Theory of Communica- tion*. MIT Press, 1961.

[27] R. G. Gallager. A simple derivation of the coding theorem and some applications. *IEEE Trans. Information Theory*, 11(1):3–18, Jan. 1965.

[28] R. G. Gallager. *Information theory and Reliable Communications*. John Wiley & Sons, New York, 1968.

[29] R. G. Gallager. The random coding bound is tight for the average code. *IEEE Trans. Information Theory*, 23(2):244–246, Mar. 1973.

[30] R. G. Gallager. A perspective on multi-access channels. *IEEE Trans. Information Theory*, 31(2):124–142, Mar. 1985.

[31] E. N. Gilbert. A comparison of signaling alphabets. *bell Syst. tech. J.*, 31:504–522, May 1952.

[32] B. Grünbaum. *Convex Polytopes*. McGraw-Hill, New York, 1976.

[33] E. A. Haroutunian. Lower bound for the error probability of multiple-access channels. *Problemy Peredachi Informatsii*, 11:23–36, June 1975.

[34] S. Janson. New versions of Suen's correlation inequality. *Random Structures Algorithms*, 13:467–483, 1998.

[35] R. J. McEliece L. R. Welch and H. Rumsey. A low rate improvement on the elias bound. *IEEE Trans. Information Theory*, 20:676–678, Sept. 1974.

[36] V. I. Levenshtein. On the minimal redundance of binary errdrcorrecting codes. *Problemy Peredachi Informatsii*, 10:26–42, June 1974.

[37] H. Liao. A coding theorem for multiple-access communications. In *Proc. International Symposium on Information Theory*.

[38] Y. Liu and B. L. Hughes. A new universal random coding bound for the multiple-access channels. *IEEE Trans. Information Theory*, 42(2):376–386, Mar. 1996.

[39] A. Nazari, A. Anastasopoulos, and S. S. Pradhan. Error exponent for multiple-access channels: Lower bounds. 2010. http://arxiv.org/abs/1010.1303.

[40] A. Nazari, S. S. Pradhan, and A. Anastasopoulous. New bounds on the maximal error exponent for multiple-access channels. In *Proc. IEEE Int. Symp. Inf. Theory*, July 2009.

[41] J. Pokorney and H. S. Wallmeier. Random coding bounds and codes produced by permutations for the multiple-access channels. *IEEE Trans. Information Theory*, 31(6):741–750, Nov. 1985.

[42] S. S. Pradhan, S. Choi, and K. Ramchandran. A graph-based framework for transmission of correlated sources over multiple-access channels. *IEEE Trans Inf Theory*, 53(12):4583–4604, 2007.

[43] P. Gacs R. Ahlswede and J. Korner. Bounds on conditional probabilities with application in multi-user communication. *Springer: Probability Theory and Related Fields*, pages 157–177, 1976.

[44] C. E. Shannon. Two-way communication channels. In *Proc. Berkeley Symposium on Mathematics, Statistics, and Probablity*, volume 10, pages 611–644, 1961.

[45] C. E. Shannon, R. G. Gallager, and E. Berlekamp. Lower bounds on errro probability for coding on discrete memoryless channels (part i). *Inform. and Control*, pages 65–103, 1967.

[46] C. E. Shannon, R. G. Gallager, and E. Berlekamp. Lower bounds on errro probability for coding on discrete memoryless channels (part ii). *Inform. and Control*, pages 522–552, 1967.

[47] D. Slepian and J. K. Wolf. A coding theorem for multiple access channels with correlated sources. *bell Syst. tech. J.*, 52:1037–1076, 1973.

[48] R. Venkataramanan and S. S. Pradhan. A new achievable rate region for the discrete memoryless multiple-access channel with feedback. In *Proc. IEEE Int. Symp. on Inf. Theory*, June 2009.