Structural Results for Coding Over Communication Networks

by

Farhad Shirani Chaharsooghi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2017

Doctoral Committee:

Professor S. Sandeep Pradhan, Chair
Associate Professor Achilleas Anastasopoulos
Professor Prakash Narayan
Professor David Neuhoff
Professor Martin Strauss

To my family.

ACKNOWLEDGEMENTS

It is my pleasure to express my gratitude to the people who have made this thesis possible. I would like to sincerely thank my research advisor Professor Sandeep Pradhan for his ceaseless support and encouragement, for providing me with the opportunity to conduct independent research, for allowing me to explore freely, and for his continued guidance along the way. Sandeep's insistence on pursuing high quality research, as well as his strict dedication to ethical conduct in research are attributes I wish to emulate in my future career as well.

I would like to thank my dissertation committee members, Professor Prakash Narayan, Professor David Neuhoff, Professor Achilleas Anastasapolous, and Professor Martin Strauss for accepting to be on my dissertation committee. I am grateful to Professor Narayan for making the time to provide me with his invaluable advice, despite his busy schedule. I am thankful to Professor Neuhoff for sharing his deep insight into the field, as well as introducing me to the basics of information theory in my first course in our department. I am deeply grateful to Professor Anastasapolous for teaching me many of the concepts I used in this research through his courses on digital communication and coding theory, as well as for his knowledgeable advice during my proposal and defense meetings. I am thankful to Professor Strauss for providing a contrasting perspective on my research as the cognate committee member.

The staff at my department have been extremely helpful throughout the past five years and their supportive efforts are hereby acknowledged. I would like to thank Becky Turanski, Anne Rhodes, Michele Feldkamp, Karen Liska, Beth Lawson, Ann Pace, José-Antonio

Rubio, and Kevin Calhoun for their efficient and friendly conduct which helped me navigate the convoluted administrative tasks during my PhD.

I would also like to thank my friends in Ann Arbor for making such indelible, joyful memories in the past five years of my life. I am thankful to my dear friend and roommate Hamid Aghasi for twenty two years of friendship and for his constant selfless support. I wish to thank my officemate Mohsen Heidari for his unabating help both professionally and as a friend. I am thankful to my cousin Saloomeh Mokhtari for her continued support during my stay in the United States. I wish to express my deep gratitude to Parisa Ghaderi, Mehrzad Samadi, Avish Kosari, Armin Jam, Payam Mirshams, Mohammadreza Imani, Aria Ghasemian, Ali Nazari, Arun Padakandla, Hamidreza Tavafoghi, Parinaz Naghizadeh, Azadeh Ansari, and Mahmood Barangi for the help they provided during the years, and for making my stay in Ann Arbor a pleasurable and exciting journey.

Lastly, I am honored to express my deepest appreciation to my family, my father Ebrahim, my mother Mahnaz, and my sister Maryam, who have been the source of constant endorsement and motivation which helped me weather the difficulties of graduate life, and provided me with the tranquility without which this work would not be possible.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES ix

LIST OF TABLES xi

LIST OF APPENDICES xii

LIST OF ABBREVIATIONS xiii

ABSTRACT xiv

CHAPTER

I. Introduction 1

- 1.1 Codebooks 2
- 1.2 Codes over Asymptotically large blocklengths 5
- 1.3 Codes over Constant blocklengths 7

II. Linear Structure for Multiple-Descriptions Coding 10

- 2.1 Introduction 10
- 2.2 Definitions and Notation 15
- 2.3 Previous Work 17
  - 2.3.1 Problem Statement 17
  - 2.3.2 Prior Results 19
- 2.4 Improvements Using Unstructured Codes 24
  - 2.4.1 Improvements to the RD Region Using Unstructured Codes 24
  - 2.4.2 Improvements Due to additional codebooks 29
  - 2.4.3 Improvements Due to Binning 32
- 2.5 Linear Coding Examples 32

2.5.1 Gains Due to Linear Quantizers 33
2.5.2 Gains Due to Linear Binning 35
2.5.3 Correlated Quantizations of a Source 38
2.6 Achievable RD Region using Structured Codes 44
2.6.1 Stage 1: Achievability of the SSC Region Using Nested Coset Codes 44
2.6.2 Stage 2: Reconstruction of a summation of two codebooks 45
2.6.3 Stage 3: Reconstruction of a summation of arbitrary number of codebooks 49
2.6.4 Stage 4: Reconstruction of an Arbitrary Number of Summations of Arbitrary Lengths 52

III. Lattice Construction for Multi-terminal Source Coding 55

3.1 Introduction 55
3.2 General Information Measures 56
3.3 Lattice Codes for PtP Source Coding 58
3.4 Multiple-Descriptions Coding for Continuous Sources 63
3.5 Improvements Using Nested Lattice Quantizers 65
3.5.1 Example 1 65
3.5.2 Example 2 66
3.6 An Improved Achievable Region Using Nested Lattice Codes . . . 67

IV. Quasi-Linear Structures for Source and Channel Coding 70

4.1 Introduction 70
4.2 Notation 72
4.3 New Codebook Structures 73
4.3.1 Quasi Linear Codes 73
4.3.2 Nested Quasi Linear Codes 74
4.4 Fundamental Properties of QLC's 76
4.4.1 The Addition of QLC's 77
4.4.2 Mutual Covering Bounds for NQLC's 78
4.5 Gains in the MD Problem 82
4.6 The Interference Channel 86
4.7 The Interference Alignment Tradeoff 87
4.7.1 Example 1 87
4.7.2 Example 2 89
4.8 New Achievable Rate Region for the IC 91

V. The Necessity of Finite Block-length Coding 94

5.1 Introduction 94
5.2 Notation 95
5.3 The *Effective Length* of an Encoder 96

5.4 Correlation Preservation in Arbitrary Encoders 107
5.5 Correlation in Single Letter Coding Schemes 114

VI. Application of Finite Block-length Codes to Distributed Source Coding 125

6.1 Preliminaries 128
6.2 Binary One Help One Example 128
6.2.1 Finite Length Quantizer Scheme 129
6.2.2 Comparison with the Common Component Scheme . . . 134
6.3 Large blocklength Quantization of Binary Variables 135
6.4 Simulations for Hamming codes 136
6.5 The New Coding Strategy 137

APPENDICES 146
A.1 Proofs for Section 2.4 147
A.1.1 Proof of Lemma 17 147
A.1.2 Proof of Lemma 19 148
A.1.3 Proof of Lemma 26 148
A.1.4 Proof of Lemma 24 for l>3 150
A.1.5 Proof of Lemma 28 150
A.2 Proofs for Section 2.5 152
A.2.1 Proof of Lemma 31 152
A.2.2 Proof of Lemma 32 152
A.2.3 Proof of Lemma 35 163
A.2.4 Proof of Lemma 39 165
A.2.5 Proof of Lemma 42 169
A.2.6 Proof of Lemma 43 171
A.3 Proofs for Section 2.6 185
A.3.1 Proof of lemma 46 185
B.1 Proofs for Chapter III 186
B.1.1 Proof of Lemma 13 186
B.1.2 Proof of Theorem III.14 187
B.1.3 Proof of Theorem III.15 188
B.1.4 Proof of Lemma 18 189
C.1 Proofs for chapter IV 190
C.1.1 Proof of Lemma 26 190
C.1.2 Proof of Lemma 29 193
D.1 Proofs for Chapter V 195
D.1.1 Proof of Proposition 15 195
D.1.2 Proof of Proposition 16 196
D.1.3 Proof of Lemma 25 199
E.1 Proofs for Chapter VI 202
E.1.1 Proof of Lemma 5 202
E.1.2 Proof of Lemma 14 203
E.1.3 Proof of Lemma 15 203

BIBLIOGRAPHY 205

LIST OF FIGURES

Figure

2.1 The Two-Descriptions Setup 12

2.2 The structure of CMSB codebooks in the three-descriptions problem . . . 23

2.3 The SSC codebooks present in the three-descriptions problem 25

2.4 Three Descriptions Setup Showing $C_{\{1,2\},\{3\}}$ is not redundant. 30

2.5 Example Showing Improvements Due to Binning 33

2.6 Three-Descriptions Example with a Vector Binary Source 34

2.7 An Example Showing the Gains Due to Linear Binning 36

2.8 Scalar Source Example with Correlated Quantization 39

2.9 Codebook Construction for Lemma 39 41

3.1 Point-to-point source coding for continuous sources 60

3.2 Example for Vector Gaussian Source 66

4.1 A setup where interference alignment is beneficial to user 3 but harmful for user 2. 72

4.2 A three-descriptions example where NQLC's give gains 82

4.3 A Three User IC Where Alignment Is Strictly Beneficial 88

4.4 A Three User IC Where Alignment Results in a Tradeoff 89

4.5 The LHS random variables are the ones sent by each encoder, the RHS random variables are the ones decoded at each decoder. 92

5.1 Correlated Boolean decision functions. 108

5.2 Point-to-point source coding example 115

6.1 General Lossy Distributed Source Coding 125

6.2 The binary one help one example 129

6.3 A block-diagram of the scheme 130

6.4 Comparison of the new scheme using Hamming codes and [50], $\delta_1 = 0.1, p = 0.3, \epsilon = 10^{-10}$ 137

A.1 Maximum value of $I(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1, X_3; X) + I(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_2, X_3; X)$ 184

B.1 Three-Descriptions Codebook Structure 188

LIST OF TABLES

Table

2.1 Joint distribution on X, $V_{\{1\}}$ and $V_{\{2\}}$. 42

4.1 The joint distribution P_{X,V_1,V_2}. 83

A.1 Joint probability distribution of $P_{\hat{X}_{12},\hat{X}_1,\hat{X}_2}$. 173

A.2 Joint probability distribution of $P_{X,\hat{X}_1,\hat{X}_2}$ 174

A.3 The optimal joint distribution 176

A.4 The optimal joint distribution 178

LIST OF APPENDICES

Appendix

A. Proofs for Chapter II 147

B. Proofs for Chapter III 186

C. Proofs for Chapter IV 190

D. Proofs for Chapter V 195

E. Proofs for Chapter VI 202

LIST OF ABBREVIATIONS

PtP Point-to-Point

MD Multiple Descriptions

RD Rate Distortion

QLC Quasi Linear Code

CMSB Combinatorial Message Sharing with Binning

ZB Zhang Berger

SSC Sperner Set Coding

BC Broadcast Channel

EGC El Gamal - Cover

SCEC Source-Channel Erasure Code

MDS Maximum Distance Separable

BSS Binary Symmetric Source

IC Interference Channel

NLC Nested Linear Code

NQLC Nested Quasi Linear Code

IBL Infinite Block Length

SLC Single-Letter Coding

DSC Distributed Source Coding

DMS Discrete Memoryless Source

BBE Binary Block Encoder

BOHO Binary-One-Help-One

BT Berger-Tung

CC Common Component

i.i.d independent, identically distributed

ABSTRACT

Structural Results for Coding Over Communication Networks

by

Farhad Shirani Chaharsooghi

Chair: S. Sandeep Pradhan

We study the structure of optimality achieving codes in network communications. The thesis consists of two parts: in the first part, we investigate the role of algebraic structure in the performance of communication strategies. In chapter two, we provide a linear coding scheme for the multiple-descriptions source coding problem which improves upon the performance of the best known unstructured coding scheme. In chapter three, we propose a new method for lattice-based codebook generation. The new method leads to a simplification in the analysis of the performance of lattice codes in continuous-alphabet communication. In chapter four, we show that although linear codes are necessary to achieve optimality in certain problems, loosening the closure restriction in the codebook leads to gains in other network communication settings. We introduce a new class of structured codes called quasi-linear codes (QLC). These codes cover the whole spectrum between unstructured codes and linear codes. We develop coding strategies in the interference channel and the multiple-descriptions problems using QLCs which outperform the previous schemes.

In the second part, which includes the last two chapters, we consider a different structural restriction on codes used in network communication. Namely, we limit the 'effective length' of these codes. First, we consider an arbitrary pair of Boolean functions which operate on two sequences of correlated random variables. We derive a new upper-bound on the correlation between the outputs of these functions. The upper-bound is presented as a function of the 'dependency spectrum' of the corresponding Boolean functions. Next, we investigate binary block-codes (BBC). A BBC is defined as a vector of Boolean functions. We consider BBCs which are generated randomly, and using single-letter distributions. We characterize the vector of dependency spectrums of these BBCs. This gives an upper-bound on the correlation between the outputs of two distributed BBCs. Finally, the upper-bound is used to show that the large blocklength single-letter coding schemes in the literature are sub-optimal in various multiterminal communication settings.

CHAPTER I

Introduction

Ever since the inception of Shannon theory, the problem of approaching the performance limits for multiterminal communications has been of great interest. However, contrary to the Point-to-Point (PtP) communication settings, characterizing the optimal performance in multiterminal communications has remained an open problem. Although the exact asymptotic limits to the performance are not known in general, considerable progress was made. Optimal coding strategies were proposed for special classes of multiterminal problems, and general upper and lower bounds were derived. Initial attempts at solving the problem mainly included the application of Shannon's unstructured random coding techniques over large blocklengths, coupled with superposition coding and binning [54], [25], [16], [45]. Later, it was observed that random coding over ensembles of codes with specific structure results in improved performance in certain communication problems. More specifically, it was shown that in the distributed source coding problem, the interference channel, and the broadcast channel, randomly generated linear codes outperform previous known coding strategies [18], [29],[28], [20], [23]. Similar results were derived for codes constructed over weaker structures such as rings and groups [35], [22]. These findings suggest that the key to constructing optimal random coding strategies might lie in characterizing the structure of optimal encoders. In this work we investigate this idea further. In our efforts to find optimal encoding functions, we uncover several new coding structures. We divide these structures into two main categories: 1) Coding structures which are generated randomly, and whose effective length is asymptotically large and 2) Codes with constant finite effective length. In the next sections, we first discuss codebooks and their role in the communication strategies considered in this work, next we explain the two categories of code ensembles mentioned above.

1.1 Codebooks

The goal of the communications engineer is to design and implement algorithms for efficient and reliable communication. The coding system designed for performing the task of communication consists of encoders and decoders. Ever since Shannon's work on point-to-point communications [38], codebooks have played a key role in the operation of the encoding and decoding. A codebook is an ordered collection of sequences of a specific length. The elements of this collection are called codewords. Shannon proposed the use of such structures as a prime component in the construction of encoders and decoders. In the ensuing decades, most of the coding strategies which were developed for multiterminal communications have utilized codebooks in constructing the encoders and decoders. Furthermore, all of the communication strategies investigated in this work are largely based on the concept of a codebook. In this thesis we are mostly concerned with new designs for, strategies for, and uses of codebooks.

The role of codebooks varies with the specific communication task. As an example, in point-to-point channel coding, the codebook consists of the set of possible outputs of the encoder. The decoder is tasked with identifying which channel codeword was the input to the channel. That is, by observing the output of the channel, which is affected by the input codeword as well as the channel noise, the decoder must decide which codeword was transmitted. The codebook is designed in a way as to facilitate the task of reliable communication. In the case of point-to-point channel coding, this requires the codewords to be as distinct as possible, to allow decoding with low probability of error. For efficient channel coding, the codebook should be large, to allow for communication at higher rates. While in information theory, the efficiency of the communication method is measured based on the rate of communication, in coding theory there is an additional issue. That is that codebooks should ideally have structure that enables the encoding and decoding to have as low implementation complexity as possible in terms of arithmetic operations per channel symbol and required auxiliary storage.

Another example is the problem of point-to-point lossy source coding. In this problem, the codebook is the set of possible outputs of the decoder. Given an input sequence, the encoder transmits an index to the decoder. The index points to the codeword which is the most suitable reconstruction of the input sequence. The suitability of the reconstruction depends on the distortion criterion in the source coding problem. In this case, the codebook should be designed such that there will be codewords that are satisfactory reconstructions for the source sequence as often as possible. In this problem efficient communication required smaller codebooks and small average distortion at the decoder. So, the codebook should be designed to optimize the rate-distortion tradeoff. Similar to the channel coding case, coding theory is further concerned with the computational and storage complexity of the resulting communication algorithms.

The use of codebooks was extended to the multiterminal communication scenarios such as distributed source coding, multiple-descriptions source coding, and communicating over the interference channel and the broadcast channel. As an example, in the two user distributed source coding problem, the goal is to separately encode and jointly decode, two correlated sources. This task is to be performed with low transmission rates and low average distortion at the decoder. The Berger-Tung strategy [45] is the best known strategy for this setup in terms of optimizing the rate-distortion tradeoff. In this strategy, there are two codebooks, one associated with each encoder. The encoding operation involves two steps. First, encoders quantize the source sequence using their corresponding codebooks. This step is similar to the encoding in point-to-point source coding explained above. In the second step, the quantized outputs are binned and the bin numbers are sent to the decoder. Binning is the operation of grouping codewords together and indexing these groups. The binning step results in a reduction of rate due to the correlation between the outputs of the two quantizers. The decoder recovers the quantizations at each encoder using the bin numbers it receives. Here, the objective is to design codebooks and binning functions which optimize the rate-distortion tradeoff.

In Shannon's original point-to-point communication theory, the codebooks are randomly generated according to some distribution. The choice of the distribution is essential in the efficiency of the resulting coding strategy. Moreover, the performance of point-to-point systems improves as the length of the codewords increases. Shannon's method does not enforce any additional structural restriction on the randomly generated codebooks. As a result these codebooks are called randomly generated unstructured codebooks. Most of the multiterminal communication strategies in the literature build upon Shannon's work, and utilize randomly generated unstructured codebooks with asymptotically large blocklengths.

Initial interest in constructing random coding strategies which produce codebooks with specific algebraic structure was due to the computational efficiency of such algorithms. Unfortunately, the complexity of coding systems based on randomly generated unstructured codebooks grows exponentially with the codeword length. For this reason, ever since Shannon's original work, the main focus of point-to-point communication has been on codebooks with "structure". That enables encoding and decoding to be done by algorithms requiring far fewer operations and storage than required for unstructured codebooks.

Recently, it was found that in various multiterminal communication problems, structure not only decreases the encoding/decoding complexity, but also improves performance in terms of achievable rates and distortions. In this thesis we introduce several different codebook structures, and analyze their performance. We show that these new codebook structures give improved performance in many network communication settings. We also discover the striking fact that in some multiterminal communication problems, and some common codebook design strategies, performance is not made best with asymptotically large codeword lengths. That is, there is a best codeword length.

1.2 Codes over Asymptotically large blocklengths

In the first part of the thesis - which includes chapters two, three, and four - we investigate the role of algebraic structure in the performance of communication strategies. In this part, we follow the common approach in multiterminal communication, which involves using single-letter distributions and random coding schemes to generate codebooks. The blocklength of these randomly generated codebooks are then taken to approach infinity, and the resulting performance is characterized using single-letter expressions.

The necessity of algebraic structure in codebooks used for multiterminal communication was first shown in [18]. Korner and Marton [18] observed that in the special case of lossless distributed transmission of the sum of two correlated binary sources, randomly generated linear codes achieve a larger rate-distortion (RD) region than unstructured codes. The phenomenon turned-out to be prevalent in multiterminal problems, and such gains were also observed in other multiterminal problems such as multiple-access channel with states available at the transmitters [30], computation over multiple-access channels [26], the interference channel [29], and the broadcast channel [28]. In the large body of work dedicated to this topic various types of structured codes have been considered. The most well-studied of these codes are linear codes. These codes are constructed over finite fields and are closed with respect to the linear operation associated with the field. The gains due to linear coding are twofolds. First, it turns out that due to their structure, linear codes can compress and transmit sums of random variables more efficiently than unstructured codes, particularly when the compression is done in a distributed fashion. Second, the rate of the addition of a linear code with itself remains the same. This is particularly useful in problems involving interference alignment, when the size of the interfering set of sequences needs to be small. Based on these observations it is expected that utilizing linear codes is also advantageous in the multiple-descriptions (MD) problems when more than two descriptions are transmitted. In the second chapter, we investigate this idea for the MD problem with discrete memoryless sources. We prove that the application of linear codes gives gains in the MD problem. Then, we generalize the idea and provide a new scheme for the MD problem with an arbitrary number of descriptions. This gives a new RD region for the MD problem which improves upon the previous known RD regions. This is proved analytically for several different classes of three and four descriptions problems. In Chapter 3, we further extend the results to the problem of compression of continuous sources. Previously, inner-bounds to the achievable regions in various multiterminal continuous source coding problems was provided using lattices [31], [43], [42]. The first contribution in this chapter is that, we provide a new method for lattice construction which is considerably simpler than the previous methods. Using the new method, we derive a new RD region for the MD problem with general continuous sources (i.e. sources which are not necessarily Gaussian.). The second contribution in this chapter is that we show that using identical lattices results in larger achievable RD regions in the Gaussian MD problem.

In Chapter 4, we show that loosening the structure of linear codes, leads to further gains in some multiterminal settings. Korner and Marton suggested the use of identical linear codes to effect binning of two correlated binary sources when the objective is to reconstruct the modulo-two sum of the sources at the decoder. They showed that such an approach leads to optimality. However, if the objective is to have the complete reconstruction of both the sources at the decoder (Slepian-Wolf setting), then it was shown that for certain sources, using identical binning can be strictly suboptimal [21]. In general, to achieve the Slepian-Wolf performance limit, one needs to use either binning of the two sources using two independent linear codes or use independent unstructured binning of the two sources. Moreover, there is no known method based on unstructured codes which achieves optimality for the reconstruction of the modulo-two sum. In summary, the former requires only identical binning, whereas the latter requires only independent binning.

This leads to the following question: (i) is there a spectrum of strategies involving partially independent binning of the two sources that lie between these two extremes, and (ii) is there a class of problems for which such strategies give gains in asymptotic performance? In other words, is there a *trade-off* between structured coding and unstructured coding. Based on this intuition, in Chapter 4, we propose a new class of codes called Quasi Linear Codes (QLC). These codes are not fully closed under any algebraic structure but maintain a degree of *"closedness"*. More precisely, the addition of a QLC with itself does not result in the same code as in the case of a linear code, rather, the resulting set of codewords has rate less than twice the rate of the original code. In that sense, these codes breach the gap between codes with algebraic structure and unstructured codes. Using QLCs we provide new schemes for the IC and MD problems. We show analytically that these new schemes result in achievable regions that are strictly larger than the original regions.

1.3 Codes over Constant blocklengths

In his paper, "A Mathematical Theory of Communications" [38] - often regarded as the Magna Carta of digital communications - Shannon pointed out that in order to exploit the redundancy of the source in data compression, it is necessary to compress large blocks of the source simultaneously. More precisely, optimality is only approached as the effective length of the coding blocks approaches infinity. The same observation was made in the case of PtP channel coding. As a result, a common feature of the coding schemes used in PtP communication is that they have large effective lengths. Loosely speaking, this means that each output element in these schemes is a function of the entire input sequence, where the length of the input sequence is asymptotically large. In the source coding problem, by compressing large blocks at the same time, one can exploit the redundancy in the source. In the channel coding problem, transmitting the input message over large blocks allows the decoder to exploit the typicality of the noise vector which results from the laws of large numbers. This remains unchanged by the multiterminal nature of the problem in coding over networks. However, in multiterminal communication it is often desirable to maintain correlation amongst the compressed sequences at different nodes. This requirement can be due to explicit constraints in the problem statement such as joint distortion measures in multiterminal source coding, or it can be due to implicit factors such as the need for interference alignment in multiterminal channel coding, or it can be due to the nature of the shared communication channel. In the latter case, correlation between the outputs is necessary as a means for further cooperation among the transmitters. It turns out that pairs of encoders with large effective lengths are inefficient in coordinating their outputs. This is due to the fact that such encoding functions are ineffective in preserving correlation. The loss of correlation undermines the encoders' ability to conspire to take advantage of the multiterminal nature of the problem. In PtP communication problems, where there is only one transmitter, the necessity for cooperation does not manifest itself. For this reason, although encoders with asymptotically large effective lengths are optimal in PtP communications, they are sub-optimal in the network communication case.

In Chapter 5, we show that as the effective length of the code increases, the outputs of the quantizers at each terminal become less correlated. The proof involves three steps. First, we consider an arbitrary pair of Boolean functions which operate on two sequences of correlated random variables. We derive a new upper-bound on the correlation between the outputs of these functions. The upper-bound is presented as a function of the 'dependency spectrum' of the corresponding Boolean functions. Next, we investigate binary block-codes (BBC) as defined in [53]. A BBC is defined as a vector of Boolean functions. We consider BBCs which are generated randomly, and using single-letter distributions. We characterize the vector of dependency spectrums of these BBCs. This gives an upper-bound on the correlation between the outputs of two distributed BBCs. Using the upper bound, it is shown that random coding over large blocklengths is detrimental to the ability of the encoders to coordinate their outputs. Hence, it is sometimes advantageous, that in the interest of cooperation, smaller blocklength random codes be used instead [1]. In summary, the first contribution in the second part of the thesis is that we show that the encoders' ability to preserve correlation has an inverse relation with their 'effective length'. The second contribution is that we show that the single-letter coding schemes used in the literature produce encoding functions which have large effective lengths. This leads us to conjecture that such schemes are sub-optimal in network communication problems. We investigate this idea further in Chapter 6 where we show the sub-optimality of the Berger-Tung[20] scheme in the distributed source coding problem, and come up with a new achievable rate-distortion region for the problem using finite blocklength schemes. We show analytically that this achievable RD region improves upon the BT region.

[1] It is well known that the performance of block-codes is super-additive, meaning that the best performance of block-codes of a certain length is an increasing function of the blocklength. This is true since a concatenation of smaller block codes gives the same performance as the original code. However, here we are discussing the performance of randomly generated block-codes.

CHAPTER II

Linear Structure for Multiple-Descriptions Coding

2.1 Introduction

In this chapter, we consider the problem of multiple descriptions (MD) source coding. The Multiple-Descriptions (MD) source coding problem arises naturally in a number of applications such as transmission of video, audio and speech over packet networks and fading channels [15][52]. The multiple-descriptions (MD) source coding setup describes a communications setting consisting of one encoder and several decoders. The encoder receives a discrete memoryless source and wishes to compress it into several *descriptions*. Each decoder receives a specific subset of these descriptions through noiseless links, and produces a reconstruction of the source vector with respect to its own distortion criterion. The parameters of interest are the rates required for transmitting each description, and the resulting distortions at the decoders. The objective is to design communications schemes which result in the optimal asymptotic trade-off between these two groups of parameters. The problem has been studied extensively [27][1][13][54][46][2], however, the optimal asymptotically achievable rate-distortion (RD) is not known even for the most elementary case when only two descriptions are considered. The two-descriptions setup is depicted in Figure 2.1. Evidently, for the individual decoders (which receive only one description) to perform optimally the encoder must transmit the two-descriptions according to the optimal Point-to-Point (PtP) source coding schemes. This may require the two-descriptions to be similar to each other. On the other hand, if the descriptions are similar, one of them would be redundant at the central decoder (which receive two descriptions). In fact, this decoder requires the two-descriptions to be different from one another in order to yield a better reconstruction. The main challenge in the MD problem is to strike a balance between these two situations. The best known achievable region for the this communications setting is due to Zhang and Berger [54]. In the Zhang-Berger (ZB) strategy, the encoder in the first step sends a common and coarsely quantized version of the source on both descriptions, then in the next step, the encoder sends individual refinements for each decoder on the corresponding descriptions.

The ZB scheme, utilizes three codebooks at the encoder. One codebook to produce the common quantization which is sent on all of the descriptions, and two codebooks for refining the original quantization. This refinement is sent on each individual description. The first codebook is used similar to the codebook in a point-to-point source coding problem. The source is quantized using this codebook, and this quantized version is sent on both descriptions. As a result all three decoders receive this quantized version of the source. The other two codebooks are superimposed on the first codebook. These codebooks are used for the purpose of refining the reconstruction at each decoder. The encoder uses these codebooks to quantize the source sequence conditioned on the common quantization from the first step.

The ZB coding strategy was generalized in [46] for the case where there are more than two descriptions. In this strategy, first, a common coarsely quantized version of the source is sent to all the decoders, then in the next step, several refinement layers are transmitted. As a result this strategy uses $l + 1$ codebooks. For the symmetric $l-$descriptions[1] problem, a coding scheme based on random binning was considered in [44] which outperforms the VKG scheme. This involves generation of independent codebooks followed by random binning. Although the MD problem has a centralized encoder, the strategy involving ran-

[1]Here l is an arbitrary natural number whose value is greater than one.

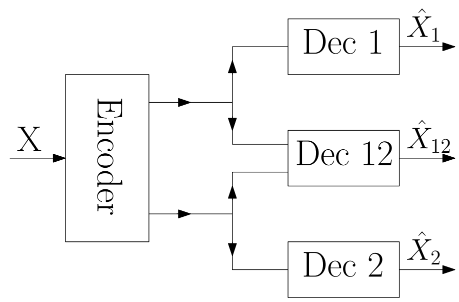

Figure 2.1: The Two-Descriptions Setup

dom binning was proved to be useful. This was further improved upon by a new coding scheme in [33] based on certain parity-check codes. However all the three schemes do not fully exploit the common-information among every subset of individual descriptions. For example in the three-descriptions problem, there can be common-information between the first and second descriptions which is not common with the third description. A new coding scheme called Combinatorial Message Sharing with Binning (CMSB) was considered in [2, 49] which provided a unified achievable RD region for the general l-descriptions problem. This scheme provided a grand unification of the schemes based on conditional codebooks and the schemes based on random binning, which in turn results in the largest achievable RD region for the problem and enlarges the achievable RD region for the previous coding schemes. The name is due to the combinatorial number of common-component codebooks present. It can be noted that CMSB scheme is based on a construction of random codes where the codewords are mutually independent, and where the codebooks do not have any algebraic structure. In the first part of this chapter, we show that this strategy can be improved upon using more general unstructured quantizers and a more general unstructured binning method. In the second part, we use several examples to prove that if linear codes are used instead, the resulting rate-distortion region can be improved even further. Furthermore, we show that structured binning also yields improvements. These improvements are in addition to the ones derived in the first part. This suggests that structured coding is essential when coding over more than two descriptions. Using the ideas developed through these examples we provide a new unified coding strategy by considering several structured coding layers. Finally, we characterize its performance in the form of an inner bound to the optimal rate-distortion region using computable single-letter information quantities. The new RD region strictly contains all of the previous known achievable regions.

We provide a new coding strategy for the general l-descriptions problem which strictly subsumes CMSB strategy which is the best known in the literature till now. The coding strategy is based on the common-information perspective. Taking a cue from the two-descriptions ZB strategy, we propose that for the general l-descriptions problem the encoder constructs a common constituent codebook for each subset of the $2^l - 1$ decoders. So, for each subset of the decoders there is one common component in the overall coding scheme. This implies that the number of constituent codebooks grows double-exponentially in l. However, we prove that only an asymptotically exponential number of the codebooks are necessary in terms of contributing to the rate-distortion region, and the rest are redundant. This significantly simplifies the coding strategy. As an example, for the $l = 3$ case, there are $2^{2^3-1} = 128$ possible common code components, but only 17 of the corresponding codebooks are non-redundant. It turns out that one can identify all of the non-redundant codebooks by associating them with the Sperner families of sets [3]. As a result, we call the new scheme the Sperner Set Coding (SSC) scheme. The CMSB scheme utilizes 14 codebooks for the 3-descriptions problem. We prove analytically that the addition of the 3 new codebooks in the SSC scheme results in an improved achievable RD region. In other words, we show analytically that the CMSB scheme is not complete. Additionally, we propose a generalized binning approach which improves upon the CMSB scheme and further enhances the SSC scheme. We characterize the asymptotic performance of this coding scheme using computable single-letter information quantities. This forms the first part of the chapter. Similar to the coding scheme of CMSB, the SSC scheme uses random unstructured codes.

It has been observed in several other multiterminal communications settings such as the Broadcast Channel (BC) [28], Interference Channel (IC) [29], variations of the MAC channel [26][30] and the Distributed Source Coding (DSC) problem [20], that the application of algebraic structured codes results in improvements over random unstructured codes in the asymptotic performance limits. Based on the inherent dualities between the multiterminal communication problems and the corresponding coding schemes, these observations suggest that one may get such gains in performance even in the MD problem.

In the second part we show that SSC coding scheme which is based on unstructured codes as mentioned above is not complete. We provide several specific examples of 3- and 4-description problems and example-specific coding schemes based on random linear codes that perform strictly better than the above SSC coding scheme. Subsequently, we supplement the above SSC scheme with new coding layers which have algebraic structure. We restrict our attention to the algebraic structure associated with finite fields. We present a unified coding scheme which works for arbitrary sources and distortion measures. We characterize the asymptotic performance of this coding scheme using computable single-letter information quantities. We interpret the SSC coding as capturing the common-information components among $2^l - 1$ decoders using univariate functions, and the algebraic coding supplement as capturing common information among $2^l - 1$ decoders using bivariate and multivariate functions.

The rest of the chapter is organized as follows. Section 2.2 explains the notation used in this chapter. Section 2.3 provides an overview of the ideas developed in previous works and provides the groundwork for the next sections. In Section 2.4, we present a new unstructured coding strategy which improves upon the CMSB scheme. We show that there are two different types of gains compared to the previous scheme: the first is due to the addition of several common-component codebook layers, the second is due to a more generalized binning method. In Section 2.5, we identify examples where improvements due to structured coding materialize in the MD setup. In this section, we investigate three different examples. In two of the examples the achievable RD region is improved via using linear quantizers, and in the other example the gains are due to linear binning. In Section 2.6 we generalize the ideas in the previous section and provide an achievable RD region for the general l-descriptions problem. Since the characterization of RD region is involved and complicated we provide the final RD region through several steps, adding new coding layers in each step.

2.2 Definitions and Notation

In this section we introduce the notation used in the chapter. While most of the new notation is clarified when it is first used in the next sections, we provide a summary of the notation here as well, as a reference-point for the reader.

We restrict ourselves to finite alphabet random variables. We denote random variables by capital letters such as X, U and their corresponding alphabets (finite) by sans-serif typeface X, U, respectively. Numbers are denoted by small letters such as l, k. Sets of numbers are also denoted by the sans-serif typeface such as M, N. Specifically, we denote the set of natural numbers by $\mathbb{N}$, and the field of size q by $\mathbb{F}_q$. The set of numbers $\{1, 2, \ldots, m\}$ is also denoted by $[1, m]$. α_{M} is used to express the vector $(\alpha_1, \alpha_2, ..., \alpha_m)$ where $\mathsf{M} = \{1, 2, \ldots, m\}$. A collection whose elements are sets is called a family of sets and is denoted by the calligraphic typeface $\mathcal{M}$. For a given family of sets $\mathcal{M}$ we define a set $\widetilde{\mathcal{M}} = \bigcup_{\mathsf{M} \in \mathcal{M}} \mathsf{M}$ as the set of numbers which are the elements of the sets in $\mathcal{M}$. The family of sets containing all subsets of M is denoted by 2^{M}. A collection whose elements are families of sets is denoted by the bold typeface $\mathbf{M}$. The collection of families of sets $\{\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_m\}$ is also represented by $\mathcal{A}_{\mathsf{M}}$. Random variables are indexed by families of sets as in $U_{\mathcal{M}}$. For the purposes of brevity we will write $U_{\mathsf{M}_1, \mathsf{M}_2, \ldots, \mathsf{M}_n}$ instead of $U_{\mathcal{M}}$ where $\mathcal{M} = \{\mathsf{M}_1, \mathsf{M}_2, \ldots, \mathsf{M}_n\}$ wherever the notation doesn't cause ambiguity. $U^n_{\mathcal{M}}$ denotes a vector of length n of random variables, each distributed according to the distribution $P_{U_{\mathcal{M}}}$. For $\epsilon > 0$ and $n \in \mathbb{N}$, we denote the set of n-length vectors which are ϵ-typical with respect to $P_{U_{\mathcal{M}}}$ by $A^n_\epsilon(U_{\mathcal{M}})$. We use the definition of frequency typicality as given in [9].

We denote a set of random variables as follows $U_{\mathbf{M}} = \{U_{\mathcal{M}} | \mathcal{M} \in \mathbf{M}\}$. For two collections of families $\mathbf{M}_1$ and $\mathbf{M}_2$, we write $[U, V]_{(\mathbf{M}_1, \mathbf{M}_2)}$ to denote the unordered collection of random variables $\{U_{\mathbf{M}_1}, V_{\mathbf{M}_2}\}$. Let $\mathbf{N}_i \subset \mathbf{M}_i, i = 1, 2$, and define $\overline{\mathbf{N}} = (\mathbf{N}_1, \mathbf{N}_2)$. We express this as $\overline{\mathbf{N}} \subset (\mathbf{M}_1, \mathbf{M}_2)$. Unions, intersections and complements are defined for $(\mathbf{M}_1, \mathbf{M}_2)$ in the same manner. A family of sets is called a Sperner family of sets if none of its elements is a subset of another element. In other words a family of sets $\mathcal{S}$ is a Sperner family if $\nexists \mathsf{N}, \mathsf{N}' \in \mathcal{S}, \mathsf{N} \subsetneq \mathsf{N}'$. For any given set M, the three families $\phi, \{\phi\}$ and $\{\mathsf{M}\}$ are all Sperner families. For a set M, we define the collection of families of sets $\mathbf{S}_{\mathsf{M}}$ as the set of all Sperner families whose elements are subsets of M except for the three trivial Sperner families mentioned above. So we have $\mathbf{S}_{\mathsf{M}} = \{\mathcal{S} | \nexists \mathsf{N}, \mathsf{N}' \in \mathcal{S}, \mathsf{N} \subsetneq \mathsf{N}'\} \backslash \{\phi, \{\phi\}, \{\mathsf{M}\}\}$.

For the general l-descriptions problem, we define the set $\mathsf{L} \triangleq [1, l]$, and this set represents the set of all descriptions. Each decoder receives a subset of these descriptions. Let $l_i \in \mathsf{L}, i \in [1, n]$ for some n. We denote the decoder which receives descriptions $l_1, l_2, \ldots, l_n$ by the set $\mathsf{N} = \{l_1, l_2, .., l_n\}$. Define the family of sets $\mathscr{L} \triangleq 2^{\mathsf{L}} - \{\phi\}$. This family of sets corresponds to the set of all possible decoders. We further explain the notation through an example. Consider the three-descriptions problem. In this case we have $l = 3$, the set of descriptions are $\mathsf{L} = \{1, 2, 3\}$. There are seven possible decoders. The set of all decoders is $\mathscr{L} = \Big\{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\Big\}$. Consider the two families of sets $\mathcal{M}_1 = \Big\{\{1, 2\}, \{1, 3\}\Big\}$ and $\mathcal{M}_2 = \Big\{\{1\}, \{3\}, \{1, 2\}\Big\}$. In this case, $\widetilde{\mathcal{M}_i} = \{1, 2, 3\}, i \in \{1, 2\}$. Define the set $\mathbf{M} = \{\mathcal{M}_1, \mathcal{M}_2\}$. The set of random variables $\{U_{\mathcal{M}_1}, U_{\mathcal{M}_2}\}$ is denoted by $U_{\mathbf{M}} = U_{\mathcal{M}_1, \mathcal{M}_2}$. Here $\mathcal{M}_1$ is a Sperner family, but $\mathcal{M}_2$ is not a Sperner family since $\{1\}, \{1, 2\} \in \mathcal{M}_2$ and $\{1\} \subsetneq \{1, 2\}$, furthermore $\mathcal{M}_1 \in \mathbf{S}_{\mathsf{L}}$ but $\mathcal{M}_2 \notin \mathbf{S}_{\mathsf{L}}$.

The second part of the chapter involves the application of linear codes and their cosets to the MD problem. The following give the definitions of a codebook:

Definition 1. *An (n, R) codebook constructed on a finite alphabet X is an ordered collection of n-length vectors, where the components of these n-length vectors take value from the* *alphabet X; the size of this collection is 2^{nR}.*[2] *Each vector in the codebook is called a codeword.*

A linear code is a codebook which is linearly closed. The following gives a formal definition for such codebooks,

Definition 2. *Let q be a prime number. A (k, n) linear code*[3] *, C is characterized by its generator matrix $G_{k\times n}$ defined on $\mathbb{F}_q$. C is defined as follows: $C \triangleq \{\mathbf{u}G|\mathbf{u} \in \mathbb{F}_q^k\}$. A coset code C' is a shifted version of a linear code and is characterized by a generator matrix $G_{k\times n}$ and a dither $\mathbf{b}^n$ defined on $\mathbb{F}_q$. C' is defined as follows: $C' \triangleq \{\mathbf{u}G + \mathbf{b}|\mathbf{u} \in \mathbb{F}_q^k\}$.*

We will make frequent use of nested linear codes. A pair of nested linear codes is defined as follows,

Definition 3. *For natural numbers $k_i < k_o < n$, let $G_{k_i\times n}$, and $\Delta G_{(k_o-k_i)\times n}$ be matrices on $\mathbb{F}_q$. Define C_i, C_o as the linear codes generated by G, $[G|\Delta G]$, respectively. (C_i, C_o) is called a pair of nested linear codes with the inner code C_i and the outer code C_o. Nested coset codes are defined as shifted versions of nested linear codes.*

2.3 Previous Work

2.3.1 Problem Statement

The general l-descriptions problem is described in this section. The setup is characterized by a discrete memoryless source with probability distribution $P_X(x), x \in \mathsf{X}$, where X is a finite set, and the distortion functions $d_{\mathsf{N}} : \mathsf{X} \times \hat{\mathsf{X}}_{\mathsf{N}} \to \mathbb{R}^+, \mathsf{N} \in \mathcal{L}$, where $\hat{\mathsf{X}}_{\mathsf{N}}$ is the reconstruction alphabet. We assume that the distortion functions are bounded, and that the

[2]When the blocklength n and the alphabet X is clear from the context, we refer to the codebook by its size 2^{nR}.

[3]It is worth noting that the term linear codes has been used to refer to linear codebooks. While it is more precise to use the latter term, instead due to the wide usage of the former, we call such codebooks linear codes.

distortion for the n-length sequence $(x^n, \hat{x}^n)$ is given by the average distortion of the components $(x_i, \hat{x}_i)$. The discrete, memoryless source X is fed into an encoder. The encoder upon receiving a block of length n of source symbols produces l different indices called *descriptions* of the source. These descriptions are sent to the decoders. Each decoder receives a specific subset of the descriptions. Decoder $\mathsf{N}, \mathsf{N} \in \mathcal{L}$ receives description i for all $i \in \mathsf{N}$. Based on the descriptions it has received, the decoder produces a reconstruction of the source vector.

Definition 4. *An $(n, \Theta_1, \Theta_2, \ldots, \Theta_l)$ multiple-descriptions code consist of an encoder and $|\mathcal{L}|$ decoders:*

$$e_i : \mathsf{X}^n \to [1, \Theta_i], i \in \mathsf{L},$$
$$f_{\mathsf{N}} : \prod_{i \in \mathsf{N}} [1, \Theta_i] \to \hat{\mathsf{X}}^n_{\mathsf{N}}, \mathsf{N} \in \mathcal{L}.$$

For a given $(n, \Theta_1, \Theta_2, \ldots, \Theta_l)$ multiple-descriptions code, the achievable RD vector is defined as $(R_i, D_{\mathsf{N}})_{i \in \mathsf{L}, \mathsf{N} \in \mathcal{L}}$, where

1. $\frac{\log \Theta_i}{n} = R_i, \forall i \in \mathsf{L}$,

2. $E_{X^n}\Big[d_{\mathsf{N}}\big(f_{\mathsf{N}}((e_i(X^n))_{i \in \mathsf{N}}), X^n\big)\Big] = D_{\mathsf{N}}, \forall \mathsf{N} \in \mathcal{L}$.

The achievable rate-distortion (RD) region is defined as follows,

Definition 5. *The RD vector $(R_i, D_{\mathsf{N}})_{i \in \mathsf{L}, \mathsf{N} \in \mathcal{L}}$ is said to be achievable if for all $\epsilon > 0$ and sufficiently large n, there exists an $(n, \Theta_1, \Theta_2, \ldots, \Theta_l)$ multiple-descriptions code such that the following constraints are satisfied:*

1. $\frac{\log \Theta_i}{n} \leq R_i + \epsilon, \forall i \in \mathsf{L}$,

2. $E_{X^n}\Big[d_{\mathsf{N}}\big(f_{\mathsf{N}}((e_i(X^n))_{i \in \mathsf{N}}), X^n\big)\Big] \leq D_{\mathsf{N}} + \epsilon, \forall \mathsf{N} \in \mathcal{L}$.

The achievable RD region for the $l-$descriptions problem is the set of all achievable RD vectors.

Remark 6. *Although the reconstruction alphabet can be different from the source alphabet, throughout this work we assume that the two alphabets are the same for the ease of notation. The results hold for the general case.*

2.3.2 Prior Results

In this section we present a brief description of some of the previous known schemes, and state the corresponding inner bounds developed for the achievable RD region. One of the early strategies for coding over two descriptions was the El Gamal - Cover (EGC) strategy [13]. Similar to all the other strategies explained in this section, the EGC scheme relies on random, unstructured codebook generation. The following theorem describes the corresponding inner bound to the achievable RD region which results from the EGC scheme. Note that this is an alternative way to characterize the inner bound described in [13].

Definition 7. *For a joint distribution P on random variables $(U_{\{1\}}, U_{\{2\}}, U_{\{1,2\}}, X, Q)$ and a set of reconstruction functions $g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}_{\mathsf{N}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$, the set $\mathcal{RD}_{EGC}(P, g_{\mathcal{L}})$ is defined as the set of RD vectors satisfying the following bounds:*

$$R_1 \geq I(U_{\{1\}}; X|Q), \quad R_2 \geq I(U_{\{2\}}; X|Q), \tag{2.1}$$

$$R_1 + R_2 \geq I(U_{\{1\}}, U_{\{2\}}; X|Q) + I(U_{\{1\}}; U_{\{2\}}|Q) + I(U_{\{1,2\}}; X|U_{\{1\}}, U_{\{2\}}, Q), \tag{2.2}$$

$$D_{\mathsf{N}} \geq E(d_{\mathsf{N}}(g_{\mathsf{N}}(U_{\mathsf{N}}, Q), X)), \mathsf{N} \in \mathcal{L}. \tag{2.3}$$

Theorem II.8 (EGC)**.** *The RD vector $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$ is achievable for the two descriptions problem, if there exists a distribution P and reconstruction functions $g_{\mathcal{L}}$ such that $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}) \in \mathcal{RD}_{EGC}(P, g_{\mathcal{L}})$.*

Since the results in this work build upon schemes such as the EGC scheme, we provide a summary of the coding strategy, and the role of the codebooks present in the scheme[4]:

[4]In this outline we have neglected time-sharing for brevity, as a result the time-sharing random variable

Codebook Generation: Fix blocklength n, and positive reals r_1, r_2, r_{12}. Define $\rho_{12,i} \triangleq R_i - r_i, i \in [1,2]$. For $i \in \{1,2\}$, generate a codebook $C_{\{i\}}$ based on the marginal $P_{U_{\{i\}}}$ with size 2^{nr_i}. That is, randomly generate 2^{nr_i} n-length sequences of vectors on the alphabet $\mathsf{U}^n_{\{i\}}$, where each element of each vector is generated based on the distribution $P_{U_{\{i\}}}$. The set of these vectors is denoted by $C_{\{i\}}$. For each pair of codewords $c^n_1, c^n_2 \in C_{\{1\}} \times C_{\{2\}}$, generate a codebook $C_{\{1,2\}}$ with size $2^{nr_{\{1,2\}}}$ super-imposed on $C_{\{1\}}$, and $C_{\{2\}}$ based on the conditional distribution $P_{U_{\{1,2\}}|U_{\{1\}},U_{\{2\}}}$. Index the codewords in the codebooks $C_{\{1\}}$ and $C_{\{2\}}$ by the numbers $[1, 2^{nr_1}]$ and $[1, 2^{nr_2}]$, respectively. Also, index the codewords in $C_{\{1,2\}}$ by the pairs $[1, 2^{n\rho_{12,1}}] \times [1, 2^{n\rho_{12,2}}]$.

Encoding: Upon receiving the source vector X^n, the encoder finds a jointly-typical set of codewords c^n_1, c^n_2, and $c^n_{1,2}$ in the set $C_{\{1\}} \times C_{\{2\}} \times C_{\{1,2\}}$ with respect to the distribution $P_{U_{\{1\}},U_{\{2\}},U_{\{1,2\}}}$. Description one carries the index of c^n_1 and the first element of the index of $c^n_{1,2}$. Description two carries the index of c^n_2 and the second element of the index of $c^n_{1,2}$.

Decoding: Decoder one reconstructs c^n_1, and produces the reconstruction $\hat{X}^n = g_{\{1\}}(c^n_1)$. Decoder two reconstructs c^n_2, and produces the reconstruction $\hat{X}^n = g_{\{2\}}(c^n_2)$. Finally, the joint decoder reconstructs $(c^n_1, c^n_2, c^n_{1,2})$, and produces the reconstruction $\hat{X}^n = g_{\{1,2\}}(c^n_1, c^n_2, c^n_{1,2})$.

As explained above, in the EGC scheme, two codebooks $C_{\{1\}}$ and $C_{\{2\}}$ are generated independently based on the marginals $P_{U_{\{1\}}}$ and $P_{U_{\{2\}}}$. The two codebooks should be large enough so that the encoder can find a pair of jointly typical codevectors in the two codebooks. If the codebooks were generated jointly based on the joint distribution $P_{U_{\{1\}},U_{\{2\}}}$, $R_1 + R_2 \geq I(U_{\{1\}}, U_{\{2\}}; X|Q) + I(U_{\{1,2\}}; X|U_{\{1\}}, U_{\{2\}}, Q)$ would ensure the existence of such jointly typical codevectors, however in the EGC scheme, since the codebooks are generated independently, a *rate-penalty* is inflicted on the encoder. The term $I(U_{\{1\}}; U_{\{2\}}|Q)$ in (2.2) is a manifestation of this *rate-penalty*. Towards reducing the *rate-penalty* a new coding strategy was introduced. The resulting achievable RD region is called the Zhang-Berger (ZB) region. The region is given in the following theorem:

Q does not appear in the formulas.

Remark 9. *In Theorem II.8, the random variable Q is the time sharing random variable. The codewords corresponding to codebooks related to random variables $U_{\{i\},i\in\{1,2\}}$ are reconstructed at any decoder receiving description i, and the codeword corresponding to the random variable $U_{\{1,2\}}$ is reconstructed at the central decoder as a refinement.*

Definition 10. *For a joint distribution P on random variables $(U_{\{1\},\{2\}}, U_{\{1\}}, U_{\{2\}}, U_{\{1,2\}}, X)$ and set of reconstruction functions $g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}_{\mathsf{N}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$, the set $\mathcal{RD}_{ZB}(P, g_{\mathcal{L}})$ is defined as the set of RD vectors satisfying the following bounds:*

$$
\begin{aligned}
&R_1 \geq I(U_{\{1\},\{2\}}, U_{\{1\}}; X), \quad R_2 \geq I(U_{\{1\},\{2\}}, U_{\{2\}}; X),\\
&R_1 + R_2 \geq I(U_{\{1\},\{2\}}; X) + I(U_{\{1\},\{2\}}, U_{\{1,2\}}, U_{\{1\}}, U_{\{2\}}; X) + I(U_{\{1\}}; U_{\{2\}}|U_{\{1\},\{2\}}),\\
&D_{\mathsf{N}} \geq E(d_{\mathsf{N}}(g_{\mathsf{N}}(U_{\mathsf{N}}), X)), \mathsf{N} \in \mathcal{L}.
\end{aligned}
$$

Theorem II.11 (ZB)**.** *The RD vector $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$ is achievable for the two descriptions problem, if there exists a distribution P and reconstruction functions $g_{\mathcal{L}}$ such that $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}) \in \mathcal{RD}_{ZB}(P, g_{\mathcal{L}})$.*

The closure of the union of all the achievable vectors is called the ZB rate-distortion region and is denoted by $\mathcal{RD}_{ZB}$:

$$\mathcal{RD}_{ZB} = cl\left(\bigcup_{P, g_{\mathcal{L}}} \mathcal{RD}_{ZB}(P, g_{\mathcal{L}})\right).$$

The scheme differs from the EGC strategy in the introduction of the random variable $U_{\{1\},\{2\}}$. There is an additional codebook constructed based on this random variable. In the ZB scheme the codeword corresponding to the codebook relating to the new random variable $U_{\{1\},\{2\}}$ is decoded at all decoders. In the EGC scheme, in order to send $U_{\{1\}} = (\widetilde{U}_{\{1\}}, U_{\{1\},\{2\}})$ and $U_{\{2\}} = (\widetilde{U}_{\{2\}}, U_{\{1\},\{2\}})$, one has to pay the following *rate-penalty*:

$$I(U_{\{1\}}; U_{\{2\}}) = H(U_{\{1\},\{2\}}) + I(\widetilde{U}_{\{1\}}; \widetilde{U}_{\{2\}}|U_{\{1\},\{2\}}).$$

But in the ZB scheme the *rate-penalty* is reduced to:

$$I(U_{\{1\},\{2\}};X) + I(U_{\{1\}};U_{\{2\}}|U_{\{1\},\{2\}}) = I(U_{\{1\},\{2\}};X) + I(\widetilde{U}_{\{1\}};\widetilde{U}_{\{2\}}|U_{\{1\},\{2\}}).$$

As mentioned above, in the ZB scheme the codeword corresponding to the codebook relating to the random variable $U_{\{1\},\{2\}}$ is decoded at all decoders, and is the 'common information' among the random vectors reconstructed in these decoders. The following definition provides a characterization of the common-component between two random variables,

Definition 12. *Let $X_{\{1\}}$ and $X_{\{2\}}$ be two random variables. W is called a common-component between $X_{\{1\}}$ and $X_{\{2\}}$, if there exist functions $h_i : \mathsf{X}_{\{i\}} \to \mathsf{W}, i = 1,2$ such that $W = h_1(X_{\{1\}}) = h_2(X_{\{2\}})$ with probability one, and the entropy of W is positive.*

It was shown in [54] that in a certain two-descriptions setup, the addition of $U_{\{1\},\{2\}}$ enlarges the RD region. We call such a random variable non-redundant. The following definition gives a formal description of a non-redundant random variable:

Definition 13. *In a given achievable RD region for the $l-$descriptions setup, characterized by a collection of auxiliary random variables, an auxiliary random variable U is called non-redundant if the RD region strictly reduces when U is set as constant.*

Example 14. *We provide an overview of the example in [54] where the ZB rate-distortion region is strictly better than EGC rate-distortion region, since it is used extensively in the following sections. Consider the two-descriptions setting. Here X is a binary symmetric source (BSS), and the side decoders intend to reconstruct X with Hamming distortion. The central decoder needs a lossless reconstruction of the source. In [54], it is shown that the rate distortion vector $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}) = (0.629, 0.629, 0.11, 0.11, 0)$ is achievable using the ZB scheme but not the EGC scheme.*

Typically, in a given RD region achievable by a specific coding scheme, each random variable in the single-letter characterization, is associated with an underlying codebook *used in that scheme. We call a codebook non-redundant if it is associated with a non-redundant random variable. In the ZB coding scheme, the codebook corresponding to $U_{\{1\},\{2\}}$ is non-redundant.*

The idea of constructing a codebook associated with the common-component between the two random variables is the foundation of most of the schemes proposed for the general $l-$descriptions problem. One can even interpret the main difference between these schemes to be the way the common-component between different random variables are exploited.

As explained in the introduction, the best known achievable RD region for the $l-$descriptions problem is the CMS with binning (CMSB) strategy. In this strategy a combinatorial number of *common-component* random variables are considered. We explain the codebook structure for the three-descriptions case. The codebook structure is shown in Figure 2.2. There are two layers of codebooks, a layer of Maximum-Distance Separable (MDS) codes and a layer of Source Channel Erasure Codes (SCEC's). The codebook $C_{\mathcal{M}}$ is decoded at decoder N if $\exists \mathsf{N}' \in \mathcal{M}, \mathsf{N}' \subset \mathsf{N}$[5]. The codebooks are binned[6] independently, and the bin numbers for the MDS code $C_{\mathcal{M}}$ are carried by description i if $i \in \bigcup_{N \in \mathcal{M}} N$. Whereas the bin number for each SCEC is carried by only one description i where $i \in \bigcap_{N \in \mathcal{M}} N$. Let $\mathcal{RD}_{CMSB}$ denote the resulting RD region achievable using CMSB strategy (see [2, 48]).

MDS Codes		SCEC	
	$C_{\{2\},\{3\}}$		
	$C_{\{1\},\{3\}}$		
	$C_{\{1\},\{2\}}$		
	$C_{\{2,3\}}$	$C_{\{1,3\},\{2,3\}}$	$C_{\{3\}}$
$C_{\{1,2\},\{1,3\},\{2,3\}}$	$C_{\{1,3\}}$	$C_{\{1,2\},\{2,3\}}$	$C_{\{2\}}$
$C_{\{1\},\{2\},\{3\}}$	$C_{\{1,2\}}$	$C_{\{1,2\},\{1,3\}}$	$C_{\{1\}}$

Figure 2.2: The structure of CMSB codebooks in the three-descriptions problem

[5]More precisely, the codeword in $C_{\mathcal{M}}$ which is used in the corresponding block of transmission is decoded if this condition is satisfied.

[6]We explain the binning operation in detail in the proof of Theorem II.16.

2.4 Improvements Using Unstructured Codes

Our objective is to provide a new achievable RD region for the l-descriptions problem, which improves upon the RD region given by the CMSB strategy. This is based on a new coding scheme involving both unstructured and structured codes. The achievable RD region and the corresponding coding scheme is presented pedagogically in two steps. In the first step, presented in this section, we provide an RD region achievable using unstructured codes. This region is strictly better than the CMSB region. In other words this is an improvement upon the CMSB region using only unstructured codes. In the second step, presented in the next two sections, this is enhanced with a structured coding layer which improves the performance even further. In other words we show that the codebooks associated with the structured coding layer are non-redundant.

2.4.1 Improvements to the RD Region Using Unstructured Codes

We describe the key ideas for the case $l = 3$. There are 7 distinct decoders, one associated with every non-empty subset of $\mathsf{L} = \{1, 2, 3\}$. That is, we identify the set of decoders with $\mathcal{L} = 2^{\mathsf{L}} \backslash \phi$. The new achievable RD region that we provide improves upon the CMSB rate-distortion region on two factors. The first comes by adding extra codebooks, and the second comes by a more general binning method. Using the common-component perspective, we associate with every non-empty subset $\mathcal{M}$ of these 7 decoders an auxiliary random variable and a corresponding codebook. That is, we identify the collection of auxiliary variables (and their codebooks) with $2^{\mathcal{L}} \backslash \phi$. Each codebook is binned multiple times, once for each description. Each description carries the bin number of the codewords in each codebook, which correspond to its own binning function. If a description is received by at least one decoder in $\mathcal{M}$, then a bin index of the codebook associated with $\mathcal{M}$ is sent on that description.

Although it appears that the strategy involves the generation of a doubly-exponential number of codebooks (in l), we show that most of these codebooks are redundant, leaving only an asymptotically exponential number of non-redundant codebooks. While the remaining codebooks are generally non-redundant, only a small number of them are such in most of the examples we consider here.

$$\begin{array}{lllll}
 & C_{\{2\},\{3\}} & & & \\
 & C_{\{1\},\{3\}} & & & \\
 & C_{\{1\},\{2\}} & & & \\
 & C_{\{2,3\}} & C_{\{1,3\},\{2,3\}} & C_{\{3\}} & C_{\{3\},\{1,2\}} \\
C_{\{1,2\},\{1,3\},\{2,3\}} & C_{\{1,3\}} & C_{\{1,2\},\{2,3\}} & C_{\{2\}} & C_{\{2\},\{1,3\}} \\
C_{\{1\},\{2\},\{3\}} & C_{\{1,2\}} & C_{\{1,2\},\{1,3\}} & C_{\{1\}} & C_{\{1\},\{2,3\}}
\end{array}$$

Figure 2.3: The SSC codebooks present in the three-descriptions problem

It turns out that a codebook is non-redundant if and only if it is associated with a a family of sets in $\mathbf{S}_{\mathsf{L}}$. So, instead of 63 codebooks, we have just 17. Since the indices of the codebooks are associated with the Sperner families of sets, we call the scheme the Sperner Set Coding (SSC) scheme. A schematic of the codebook collection is shown in Figure 2.3. We start from the left and from the top. The first two codebooks can be identified as $(3, 2)$ and $(3, 1)$ MDS codes. The next six codebooks can be identified as three $(2, 1)$ MDS codes, and three $(2, 2)$ MDS codes associated with decoders which get two descriptions. The next three can be identified as $(3, 2)$ source-channel erasure codes (SCEC). The next three can be identified as $(3, 1)$ SCEC's (similar to the codebooks used in the EGC rate region). All these 14 codebooks are considered in deriving the CMSB rate region. The final set of codebooks are new. They can be identified as three $(2, 1)$ MDS codes associated with decoders that receive disjoint subsets of descriptions. Next we provide the main result of this sections. The following theorem characterizes the achievable RD region for the SSC scheme:

Definition 15. *For a joint distribution P on random variables $U_{\mathcal{M}}$, $\mathcal{M} \in \mathbf{S}_{\mathsf{L}}$ and X and a set of reconstruction functions $g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}_{\mathsf{N}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$, the set $\mathcal{RD}_{SSC}(P, g_{\mathcal{L}})$ is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers*

$(\rho_{\mathcal{M},i}, r_{\mathcal{M}})_{i\in\widetilde{\mathcal{M}},\mathcal{M}\in\mathbf{S}_{\mathsf{L}}}$:

$$H(U_{\mathbf{M}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (H(U_{\mathcal{M}}) - r_{\mathcal{M}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \tag{2.4}$$

$$H(U_{\mathbf{M}_{\mathsf{N}}}|U_{\mathbf{L}\cup\widetilde{\mathbf{M}}_{\mathsf{N}}}) \leq \sum_{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash(\mathbf{L}\cup\widetilde{\mathbf{M}}_{\mathsf{N}})} (H(U_{\mathcal{M}}) + \sum_{i\in\widetilde{\mathcal{M}}} \rho_{\mathcal{M},i} - r_{\mathcal{M}}), \forall \mathbf{L} \subset \mathbf{M}_{\mathsf{N}}, \forall \mathsf{N} \in \mathcal{L}, \tag{2.5}$$

$$r_{\mathcal{M}} \leq H(U_{\mathcal{M}}), \forall \mathcal{M} \in \mathbf{S}_{\mathsf{L}},$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \quad D_{\mathsf{N}} = E\{d_{\mathsf{N}}(g_{\mathsf{N}}(U_{\mathsf{N}}), X)\}, \tag{2.6}$$

where $\mathbf{M}_{\mathsf{N}}$ *is the set of all codebooks decoded at decoder* N*, that is* $\mathbf{M}_{\mathsf{N}} \triangleq \{\mathcal{M} \in \mathbf{S}_{\mathsf{L}} | \exists \mathsf{N}' \subset \mathsf{N}, \mathsf{N}' \in \mathcal{M}\}$*, and* $\widetilde{\mathbf{M}}_{\mathsf{N}}$ *denotes the set of all codebooks decoded at decoders* $\mathsf{N}_p \subsetneq \mathsf{N}$ *which receive subsets of descriptions received by* N*, that is* $\widetilde{\mathbf{M}}_{\mathsf{N}} \triangleq \bigcup_{\mathsf{N}_p\subsetneq\mathsf{N}} \mathbf{M}_{\mathsf{N}_p}$.

Theorem II.16. *The RD vector* $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},\mathsf{N}\in\mathcal{L}}$ *is achievable for the* $l-$*descriptions problem, if there exists a distribution* P *and reconstruction functions* $g_{\mathcal{L}}$ *such that* $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},N\in\mathcal{L}} \in \mathcal{RD}_{SSC}(P, g_{\mathcal{L}})$.

The closure of the union of all such achievable vectors is called the SSC achievable rate-distortion region and is denoted by $\mathcal{RD}_{SSC}$,

$$\mathcal{RD}_{SSC} = cl\left(\bigcup_{P,g_{\mathcal{L}}} \mathcal{RD}_{SSC}(P, g_{\mathcal{L}})\right).$$

In order to clarify the notation we explain the random variables decoded at each decoder[7] in the three-descriptions problem. When $l = 3$, we know $\mathbf{S}_{\mathsf{L}}$ has 17 elements. In the formulas, $\mathbf{M}_{\mathsf{N}}$ corresponds to the set of random variables decoded at decoder N, whereas $\widetilde{\mathbf{M}}_{\mathsf{N}}$ corresponds to the set of random variables which are decodable if we have access to strict subsets of the descriptions received by N. Here are the random variables decoded at

[7]More precisely, the random variables relating to the codebooks whose corresponding codewords are decoded at each decoder.

decoders $\{1\}$ and $\{2,3\}$:

$$\text{decoder } \{1\}\text{: } U_{\{1\},\{2\},\{3\}}, U_{\{1\},\{2\}}, U_{\{1\},\{3\}}, U_{\{1\},\{2,3\}}, U_{\{1\}}$$

$$\text{decoder } \{2,3\}\text{: } U_{\{1\},\{2\},\{3\}}, U_{\{1,2\},\{1,3\},\{2,3\}}, U_{\{1\},\{2\}}, U_{\{1\},\{3\}}, U_{\{2\},\{3\}},$$

$$U_{\{1\},\{2,3\}}, U_{\{2\},\{1,3\}} U_{\{3\},\{1,2\}}, U_{\{1,2\},\{2,3\}}, U_{\{1,3\},\{2,3\}}, U_{\{2\}}, U_{\{3\}}, U_{\{2,3\}}$$

So as an example $\mathbf{M}_{\{1\}} = \Big\{\big\{\{1\},\{2\},\{3\}\big\},\big\{\{1\},\{2\}\big\},\big\{\{1\},\{3\}\big\},\big\{\{1\},\{2,3\}\big\},\big\{\{1\}\big\}\Big\}$ which are all the codebooks decoded at decoder $\{1\}$.Also $\widetilde{\mathbf{M}}_{\{2,3\}} = \Big\{\big\{\{1\},\{2\},\{3\}\big\}, \big\{\{1\},\{2\}\big\}, \big\{\{1\},\{3\}\big\}, \big\{\{2\},\{3\}\big\}, \big\{\{2\},\{1,3\}\big\}, \big\{\{3\},\{1,2\}\big\}, \big\{\{2\}\big\}, \big\{\{3\}\big\}\Big\}$, and these are all the codebooks which are decoded at decoders $\{2\}$ and $\{3\}$.

Lemma 17. *The SSC rate-distortion region is convex.*

Proof. See Section A.1.1 in the appendix. □

Remark 18. *For every decoder $\mathsf{N} \in \mathcal{L}$, we have defined the reconstruction as a function of the random variable U_{N}. However, decoder N decodes all random variables $U_{\mathcal{M}}$ where $\mathcal{M} \in \mathbf{M}_{\mathsf{N}}$. The following lemma shows that the RD region does not improve if the reconstruction function is defined as a function of $U_{\mathbf{M}_{\mathsf{N}}}$ instead.*

Lemma 19. *The RD region in Theorem II.16 does not change if the reconstruction function at decoder N is defined as a function of $U_{\mathbf{M}_{\mathsf{N}}}$.*

Proof. See Section A.1.2 in the appendix. □

Remark 20. *In the scheme proposed in Theorem II.16 there are $|\mathbf{S}_{\mathsf{L}}|$ codebooks. We know that the size of $\mathbf{S}_{\mathsf{L}}$ is the number of Sperner families on L minus three. The number of Sperner families is called the Dedekind numbers [17]. There has been a large body of work in determining the values of Dedekind numbers for different l. It is known that these numbers grow exponentially in l. As an example the number of codebooks necessary for $l = 2, 3$ and 4 are 3, 17 and 165. However in all of the examples in this work it turns* *out that many of the codebooks become redundant and only a small subset are used in the scheme.*

Proof. Before proceeding to a more detailed description of the coding strategy we provide a brief outline. For each family of sets $\mathcal{M} \in \mathbf{S}_{\mathsf{L}}$ the encoder generates a codebook $C_{\mathcal{M}}$ based on the marginal $P_{U_{\mathcal{M}}}$ independently of the other codebooks. Intuitively, this codebook is the common-component among all the decoders N such that $\mathsf{N} \in \mathcal{M}$, and it is decoded in all decoders $N' \supset N$. Codebook $C_{\mathcal{M}}$ is binned independently and uniformly for each description i if $i \in \widetilde{\mathcal{M}}$. The description will carry the corresponding bin number for the codewords in each of the corresponding codebooks. Each decoder reconstructs its corresponding codewords by finding a unique set of jointly typical codevectors in the bins it has received. The existence of the jointly typical set of codewords is ensured at the encoder by the way of satisfaction of (2.4), whereas at the decoder unique reconstruction is warranted by (2.5).

Codebook Generation: Fix blocklength n and positive reals $(\rho_{\mathcal{M},i}, r_{\mathcal{M}})_{i\in\widetilde{\mathcal{M}},\mathcal{M}\in\mathbf{S}_{\mathsf{L}}}$. For every $\mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, generate a codebook $C_{\mathcal{M}}$ based on the marginal $P_{U_{\mathcal{M}}}$ with size $2^{nr_{\mathcal{M}}}$. For the ith description, if $i \in \widetilde{M}$, bin the codebook $C_{\mathcal{M}}$ randomly and uniformly into $2^{n\rho_{\mathcal{M},i}}$ bins (i.e. randomly and uniformly assign an index $[1, 2^{n\rho_{\mathcal{M},i}}]$ to each codeword in $C_{\mathcal{M}}$, and the index is called the *bin-index*.).

Encoding: Upon receiving the source vector X^n, the encoder finds a jointly-typical set of codewords $u^n_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$. Each description carries the *bin-indices* of all the codewords corresponding to its own binning function.

Decoding: Having received the bin-indices from descriptions $i \in \mathsf{N}$, decoder N tries to reconstruct the codeword corresponding to $C_{\mathcal{M}}$ if $\mathcal{M} \in \mathbf{M}_{\mathsf{N}}$. In other words the decoder finds a unique vector $(u^n_{\mathcal{M}})_{\mathsf{N}\in\mathcal{M}}$ of jointly typical sequences in the corresponding bins. If the vector does not exist or is not unique, the decoder declares error.

Covering Bounds: Since codebooks are generated randomly and independently, to find a set of vectors $U^n_{\mathcal{M}}$ that is jointly typical with the source vector X^n, the mutual covering bounds (2.4) are necessary based on the mutual covering lemma [14].

Packing Bounds: For decoder N, description i is received if $i \in \mathsf{N}$. Since binning is done independently and uniformly, to find a unique set of jointly typical sequences $(u^n_{\mathcal{M}})_{\mathsf{N}\in\mathcal{M}}$, the mutual packing bounds (2.5) are required by the mutual packing lemma [14].

□

Remark 21. *There are two main differences between the new scheme and the previous CMSB scheme. First there are additional codebooks present. As an example in Figure 2.3, the three codebooks in the right column are not present in the CMSB scheme. Second, description i bins all of the codebooks $\mathcal{M}$ such that $i \in \widetilde{M}$. We will show in the next sections that these additional codebooks contribute to an enlargement of the achievable RD region. In other words we prove that all of the additional codebooks are non-redundant. Also we show that the new binning strategy improves the achievable RD region.*

2.4.2 Improvements Due to additional codebooks

Consider the general l-descriptions problem. In this section we prove that a codebook $C_{\mathcal{M}}$ is non-redundant if $\mathcal{M} \in \mathbf{S}_{\mathsf{L}}$.

Remark 22. *It is straightforward to see that addition of a codebook $C_{\mathcal{M}}$ where $\mathcal{M} \notin \mathbf{S}_{\mathsf{L}}$ is not going to result in a larger achievable RD region. To see this consider the three descriptions problem and assume we add the codebook $C_{\{1\},\{1,2\}}$. By our definition this new codebook is decoded if we either receive description 1 or both descriptions 1 and 2. In this case the codebook is decoded in exactly those decoders where $C_{\{1\}}$ is decoded. This means that merging these two codebooks does not change the packing bounds whereas it may relax the covering bounds. So such a codebook would be redundant. This is the reason why we consider only those codebooks which are associated with Sperner families.*

Remark 23. *There are three Sperner families for which we do not construct codebooks: $\{\phi, \{\phi\}, \{\mathsf{L}\}\}$. It is clear that U_ϕ and $U_{\{\phi\}}$ are not necessary since they are not decoded at* *any decoder. Furthermore one can use the proof provided in [51] to show that U_{L} is also redundant.*

The next lemma proves that the random variables considered in Theorem II.16 are non-redundant.

Lemma 24. *The random variable $U_{\mathcal{M}}$ is non-redundant for every $\mathcal{M} \in \mathbf{S}_{\mathsf{L}}$.*

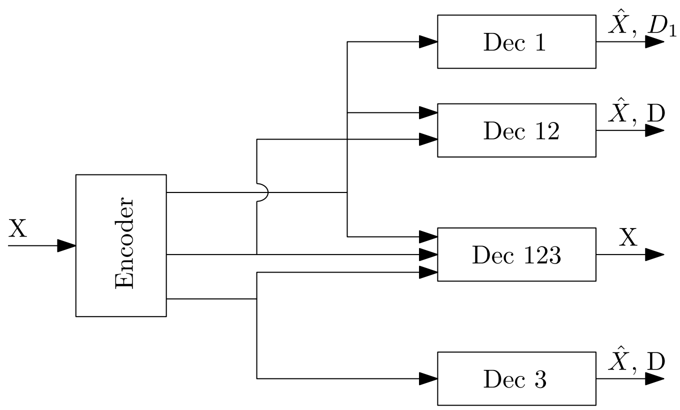

Figure 2.4: Three Descriptions Setup Showing $C_{\{1,2\},\{3\}}$ is not redundant.

Proof. We provide the proof for the $l = 3$ case and give an outline of how the proof is generalized for $l > 3$. The codebooks $C_{\{1\}}$, $C_{\{2\}}$, $C_{\{3\}}$, $C_{\{1,2\}}$, $C_{\{1,3\}}$, $C_{\{2,3\}}$, $C_{\{1\},\{2\}}$, $C_{\{1\},\{3\}}$, $C_{\{2\},\{3\}}$, $C_{\{1,2\},\{1,3\}}$, $C_{\{1,2\},\{2,3\}}$, $C_{\{1,3\},\{2,3\}}$, $C_{\{1\},\{2\},\{3\}}$, $C_{\{1,2\},\{1,3\},\{2,3\}}$ are all present in the CMSB scheme and it was shown that they are non-redundant. The new codebooks are $C_{\{1,2\},\{3\}}$, $C_{\{1,3\},\{2\}}$ and $C_{\{23\},\{1\}}$. We prove that $C_{\{1,2\},\{3\}}$ is non-redundant using the following example, the two other codebooks are non-redundant by symmetry.

We build on Example 14 to construct a three-descriptions example as shown in Figure 2.4. As explained in the previous section, it is known that $U_{\{1\},\{2\}}$ is non-redundant. Let $R_i = 0.629, i \in \{1, 2\}$, and $D_{\{1,2\}} = 0$. Let

$$D^* = \min_D \{D | (0.629, 0.629, D, D, 0) \in \mathcal{RD}_{ZB}\}. \tag{2.7}$$

Let P be the set of probability distributions $P_{U_{\{1\},\{2\}},U_{\{1\}},U_{\{2\}},X}$, such that $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$ $= (0.629, 0.629, D^*, D^*, 0)$ belongs to $\mathcal{RD}_{ZB}(P_{U_{\{1\},\{2\}},U_{\{1\}},U_{\{2\}},X}, g_{\mathcal{L}})$ for some $g_{\mathcal{L}}$ as given in

Theorem II.11. Define the joint distribution $P^*_{U_{\{1\},\{2\}},U_{\{1\}},U_{\{2\}},X}$ as follows:

$$P^*_{U_{\{1\},\{2\}},U_{\{1\}},U_{\{2\}},X} \triangleq \underset{P_{U_{\{1\},\{2\}},U_{\{1\}},U_{\{2\}},X}\in\mathsf{P}}{\operatorname{arginf}} I(U_{\{1\},\{2\}};X).$$

Let $P^*_{U_{\{1\},\{2\}},X}$ be the marginal distribution of $U_{\{1\},\{2\}}$ and X. Define a random variable W that is correlated with X such that $P_{W,X} = P^*_{U_{\{1\},\{2\}},X}$. Let N_δ be a binary random variable independent of X and W with $P(N_\delta = 1) = \delta, \delta \in (0, 0.5)$. Define $\hat{W} = U_{\{1\},\{2\}} \wedge N_\delta$ where $\wedge$ denotes the logical AND function. Let $P_{\hat{W},X}, P_{X|\hat{W}}$ be the induced joint and conditional distributions, respectively.

Example 25. *We proceed by explaining the new example. The source X is a BSS, decoders $\{1,2\}$ and $\{3\}$ want to reconstruct the source with respect to Hamming distortion and the central decoder wants to reconstruct the source losslessly. Decoder $\{1\}$ wants to reconstruct the source with respect to the distortion function given by:*

$$d_{\{1\}}(x,\hat{x}) = -\log(P_{X|\hat{W}}(x|\hat{x}))$$

Lemma 26. *The following RD vector does not belong to $\mathcal{RD}_{CMSB}$, where $U_{\{1,2\},\{3\}}$ is constant. The vector belongs to $\mathcal{RD}_{SSC}$ given in Theorem II.16 which is achievable using the SSC scheme:*

$$(R_1, R_2, R_3, D_{\{1\}}, D_{\{1,2\}}, D_{\{3\}}, D_{\{1,2,3\}}) = (I(X;\hat{W}), 0.629 - I(X;\hat{W}), 0.629, D', D^*, D^*, 0),$$

where $D' = E(d_{\{1\}}(X,\hat{W}))$.

Proof. We provide the intuition behind the proof first. In the coding scheme in Theorem II.16, the only codebooks capable of carrying the common-component between decoders $\{1,2\}$ and $\{3\}$ are $C_{\{1\},\{3\}}$, $C_{\{2\},\{3\}}$, $C_{\{1\},\{2\},\{3\}}$ and $C_{\{1,2\},\{3\}}$. We have set the distortion constraint at decoder $\{1\}$ such that this common message can't be carried exclusively on either of the descriptions 1 and 2, but rather both descriptions are necessary for the reconstruction of the common codebook. So the codebook $C_{\{1,2\},\{3\}}$ can't be empty. The proof is provided in Section A.1.3 in the appendix. □

So far we have shown that the additional codebooks are non-redundant when $l = 3$. The argument can be extended to the case when $l > 3$, an outline of the general argument is provided in appendix A.1.4. □

2.4.3 Improvements Due to Binning

The second factor contributing to the gains in the SSC rate-distortion region is the binning method. In the SSC scheme all descriptions $i \in \widetilde{\mathcal{M}}$ carry independent bin indices of codebook $C_{\mathcal{M}}$. This is different from the CMSB strategy where each codebook is binned by a specific subset of the descriptions based on whether the codebook is a SCEC or an MDS codebook. We prove through a three-descriptions example that the RD region enlarges due to binning in the SSC scheme, even with the three additional codebooks. We show in the following example that the bin indices of $C_{\{1,2\},\{1,3\}}$ should be carried by all descriptions.

Example 27. *The example is generated by modifying Example 25 and is illustrated in Figure 2.5. The source X is BSS. $d_{\{1\}}(X, \hat{W})$ is defined as in Example 25. Decoders $\{1,2\}$ and $\{1,3\}$ want to reconstruct the source with Hamming distortion and decoder $\{1,2,3\}$ wants to reconstruct the source losslessly.*

Lemma 28. *In order to achieve $(R_1, R_2, R_3, D_{\{1\}}, D_{\{1,2\}}, D_{\{1,3\}}, D_{\{1,2,3\}}) = (I(\hat{W}; X), R-I(\hat{W}; X), R - I(\hat{W}; X), D', D, D, 0)$ we must have $\rho_{\{1,2\},\{1,3\},2} + \rho_{\{1,2\},\{1,3\},3} > 0$.*

Proof. See Section A.1.5 in the appendix. □

2.5 Linear Coding Examples

Before providing a unified RD region which uses both unstructured and structured codes (step 2), in this section, for pedagogical reasons, we look at three examples of

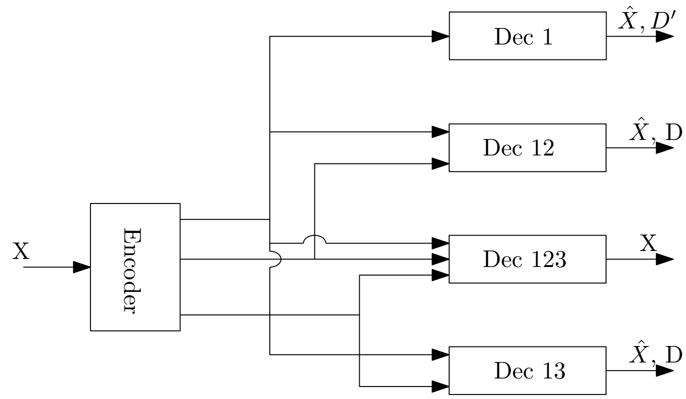

Figure 2.5: Example Showing Improvements Due to Binning

$l-$descriptions problems and provide example-specific coding schemes based on linear codes that perform strictly better than the SSC scheme which is based on unstructured codes. This shows that the SSC region is not complete and a structured coding layer is necessary. These coding schemes are unified and presented in the next section.

2.5.1 Gains Due to Linear Quantizers

We create a three-descriptions setting where reconstructions of bivariate functions are necessary.

Example 29. *Consider the three-descriptions example in Figure 2.6. Here X and Z are independent BSS. Decoder* $\{1\}, \{2\}$ *and* $\{3\}$ *wish to reconstruct X, Z and X + Z, respectively, with Hamming distortion. Decoders* $\{1, 2\}, \{1, 3\}$*, and {2,3} wish to reconstruct the pair* (X, Z) *with distortion function*

$$d_{XZ}((\hat{X}, \hat{Z}), (X, Z)) = d_H(\hat{X}, X) + d_H(\hat{Z}, Z).$$

We are interested in achieving the following RD vector:

$$R_i = 1 - h_b(\delta), i \in \{1, 2, 3\}, D_{\{1\}} = D_{\{2\}} = \delta, , D_{\{3\}} = \delta * \delta, D_{\{1,2\}} = D_{\{1,3\}} = D_{\{2,3\}} = 2\delta. \tag{2.8}$$

First we argue that in this example, description 3 should carry a bivariate function of descriptions 1 and 2. Decoders {1} and {2} operate at the optimal PtP rate-distortion function. So the corresponding descriptions have to allocate all of their rates to satisfy their individual decoder's distortion criteria. Since the distortion constraint at decoder {1} only relates to X, this description only carries a quantization of X, and by the same argument description 2 carries a quantization of Z. Then description 3 has to carry the sum of these two quantizations so that the joint decoders' distortion constraints are all satisfied. Since structured codes are efficient for transmitting bivariate summations of random variables, we expect that using structured codes would give gains in this example as opposed to unstructured codes. First, we prove that the RD vector is achievable using linear codes.

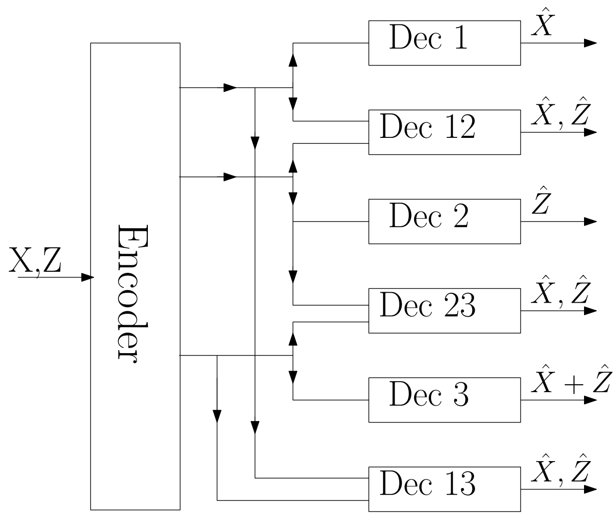

Figure 2.6: Three-Descriptions Example with a Vector Binary Source

Lemma 30. *The RD vector in (2.8) is achievable.*

Proof. **Encoding:** Construct a sequence of random linear codes C^n of rate $1 - h_b(\delta) + \epsilon_n$, where ϵ_n is going to 0. It is well known that such a sequence of linear codes can be used to quantize a BSS to Hamming distortion δ. Define the following:

$$\hat{X}^n = argmin_{c^n \in C^n} d_H(x^n, c^n)$$

$$\hat{Z}^n = argmin_{c^n \in C^n} d_H(z^n, c^n)$$

Since $\hat{X}^n$ and $\hat{Z}^n$ are codewords and the codebook is linear, $\hat{X}^n + \hat{Z}^n$ is also a codeword. Description 1 carries the index of $\hat{X}^n$, description 2 carries the index of $\hat{Z}^n$ and description 3 carries the index of $\hat{X}^n + \hat{Z}^n$.

Decoding: Decoders $\{1\}$ and $\{2\}$, receive $\hat{X}^n$ and $\hat{Z}^n$, respectively, so they satisfy their distortion constraints. Decoder $\{3\}$ reconstructs $\hat{X}^n + \hat{Z}^n$. Lemma 31 shows that the distortion criteria at this decoder is satisfied.

Lemma 31. *In the above setting, we have* $\frac{1}{n}E(d_H(\hat{X}^n + \hat{Z}^n, X^n + Z^n)) \rightarrow \delta * \delta$.

Proof. See Section A.2.1 in the appendix.

□

Decoder $\{1, 2\}$ receives $\hat{X}^n$ and $\hat{Z}^n$, so it satisfies its distortion requirements. Also decoders $\{1, 3\}$ and $\{2, 3\}$ can recover $\hat{X}^n$ and $\hat{Z}^n$ by adding $\hat{X}^n + \hat{Z}^n$ to $\hat{X}^n$ and $\hat{Z}^n$, respectively. This shows that the RD vector in (2.8) is achievable using linear codes.

□

Next we show that the SSC scheme cannot achieve this RD vector.

Lemma 32. *The RD vector in (2.8) does not belong to $\mathcal{RD}_{SSC}$, i.e., it is not achievable using the SSC scheme.*

Proof. See Section A.2.2 in the appendix. □

2.5.2 Gains Due to Linear Binning

In the SSC scheme, there are two stages in the codebook generation phase. In the first stage unstructured codebooks are generated randomly and independently, and in the second stage these codebooks are binned randomly in an unstructured fashion for each description. In the previous example it was shown that in the first stage, it is beneficial to generate codebooks with a linear structure. However in that example there was no need for binning. In the next example, we show that the binning operation needs to be carried out in a structured manner as well. This is analogous to the gains observed in the distributed source coding problem [18] where the bin structure needs to be linear. Consider the four-descriptions example in Figure 2.7.

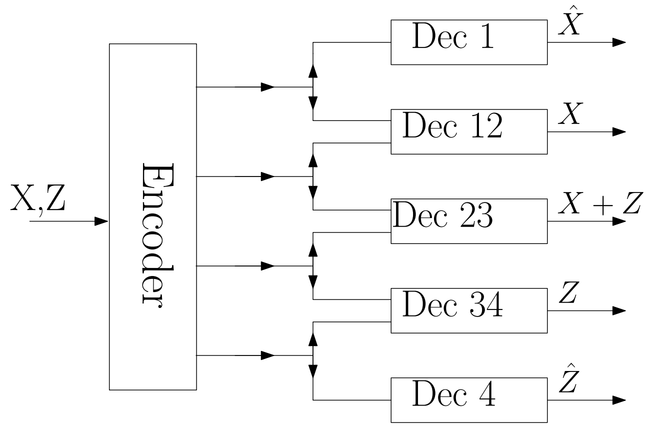

Figure 2.7: An Example Showing the Gains Due to Linear Binning

Example 33. *X and Z are BSS's. X and Z are not independent, and they are related to each other through a binary symmetric channel with bias $p \in (0, \frac{1}{2})$. In other words $X = Z + N_p$ where $N_p \sim Be(p)$ is independent of X and Z. Decoders $\{1\}$ and $\{4\}$ wish to decode X and Z, respectively, with Hamming distortion. Decoders $\{1, 2\}$, $\{3, 4\}$ and $\{2, 3\}$ require a lossless reconstruction of X, Z and X + Z, respectively. We are interested in achieving the following RD vector:*

$$R_1 = R_4 = 1 - h_b(p), R_2 = R_3 = h_b(p), D_{\{1\}} = D_{\{4\}} = p \tag{2.9}$$

We show that the RD vector in (2.9) is achievable using structured codebooks and linear binning in the next lemma.

Lemma 34. *The RD vector in (2.9) is achievable.*

Proof. **Codebook Generation:** Take an arbitrary sequence of positive numbers ϵ_n, where $\epsilon_n \to 0$ as $n \to \infty$. For any large $n \in \mathbb{N}$, fix $r_{i,n} = 1 - h_b(q) - \epsilon_n$ and $r_{o,n} = 1 - h_b(q) + \epsilon_n$. Construct a family of nested coset codes (C_i^n, C_o^n) where $C_i^n \subset C_o^n$ such that the rate of the outer code is $r_{o,n}$ and the rate of the inner code is $r_{i,n}$. Choose C_i^n such that it is a good channel code for a BSC(p), and choose C_o^n such that it is a good source code for quantizing a BSS to Hamming distortion p. The existence of such nested coset codes is well-known from random coding arguments [12]. Next we bin the space $\mathbb{F}_2^n$ into shifted versions (cosets) of C_i^n. Let $\mathcal{P}_i$ be the Voronoi region of the codeword 0^n in C_i^n. Any vector $x^n \in \mathbb{F}_2^n$ can be written in the form $x^n = v^n + c_i^n, v^n \in \mathcal{P}_i, c_i^n \in C$. Define the ith bin as $v^n + C_i^n$. This operation bins the space into $|\mathcal{P}_i| = 2^{n(h_b(p)+\epsilon_n)}$ bins. The bin number associated with an arbitrary vector x^n determines exactly the quantization noise resulting from quantizing the vector using C_i^n with the minimum Hamming distortion criterion. We denote the bin number of x^n as $B_i(x^n)$. A similar binning operation can be performed using C_o^n. Denote the bin number of x^n obtained using shifted versions of C_o^n by $B_o(x^n)$.

Encoding: The encoder quantizes x^n and z^n using C_o^n to $Q_o(x^n)$, and $Q_o(z^n)$, respectively. It also finds the bin number of the two source sequences $B_i(x^n)$ and $B_i(z^n)$. $Q_o(x^n)$ is transmitted on the first description, $B_i(x^n)$ is transmitted on the second description, $B_i(z^n)$ is transmitted on the third description, and $Q_o(z^n)$ is transmitted on the fourth description.

Decoding: Since the outer codes are good source codes, the distortion constraints at decoders $\{1\}$ and $\{4\}$ are satisfied.

We argue that the Voronoi region of 0^n in C_o^n is a subset of the one for C_i^n. This is true since $C_i^n \subset C_o^n$. Hence, having $B_i(x^n)$, decoders $\{1,2\}$ and $\{3,4\}$ can calculate $B_o(x^n)$. As mentioned above the bin number determines the quantization noise, so the decoders can reconstruct the source losslessly using the bin number and the quantization vector. Decoder $\{2,3\}$ receives $B_i(x^n)$ and $B_i(z^n)$. We have $x^n = Q_i(x^n) + B_i(x^n)$ and $z^n = Q_i(z^n) + B_i(z^n)$, so $B_i(x^n) + B_i(z^n) = x^n + z^n + Q_i(x^n) + Q_i(z^n)$. Since C_i^n is linear, $Q_i(x^n) + Q_i(z^n)$ is a codeword, and $x^n + z^n$ can be thought of as the noise vector for a $BSC(p)$. We constructed C_i^n such that it is a good channel code for BSC(p), so the decoder can recover $Q(x^n) + Q(z^n)$ from $x^n + z^n + Q_i(x^n) + Q_i(z^n)$. Then by subtracting the two vectors it can get $x^n + z^n$. □

Although we have used linear codes for quantization as well as binning, the linearity of the binning codebook C_i^n is critical in this example. In fact, it can be similarly shown that one can achieve the RD vector in (2.9) with C_o^n chosen to be a union of random cosets of C_i^n. This is in contrast with the previous example where the quantizing codebook was required to be linear.

Lemma 35. *The RD vector in (2.9) is not achievable using the SSC scheme.*

Proof. See Section A.2.3 in the appendix. □

2.5.3 Correlated Quantizations of a Source

It can be noted that in the case of SSC scheme, the unstructured quantizers are generated randomly and independently. As observed in these two examples, in order to efficiently reconstruct the bivariate summation, it is beneficial to use the same linear code for quantizing the source. However, in the two examples the source was a vector with two components which were separately quantized using identical linear codes, and the analysis of the coding scheme required only standard PtP covering and packing bounds for linear codes. In the more general case, evaluation of the performance of identical, and more generally, correlated linear codes for MD quantization, requires new covering and packing bounds. This is illustrated through the following scalar source example which is depicted in Figure 2.8. The setup is constructed based on the no-excess rate example described in [54] for the two-descriptions problem. In the two-descriptions example, the source X is BSS, and the distortion functions at all decoders is Hamming distortion. For the special case, called no-excess rate regime, when $R_1 = R_2 = \frac{1-h(D_0)}{2}$, it was shown that the EGC region is tight. Here D_0 is the distortion $D_{\{1,2\}}$ at decoder $\{1, 2\}$, and the minimum side distortion

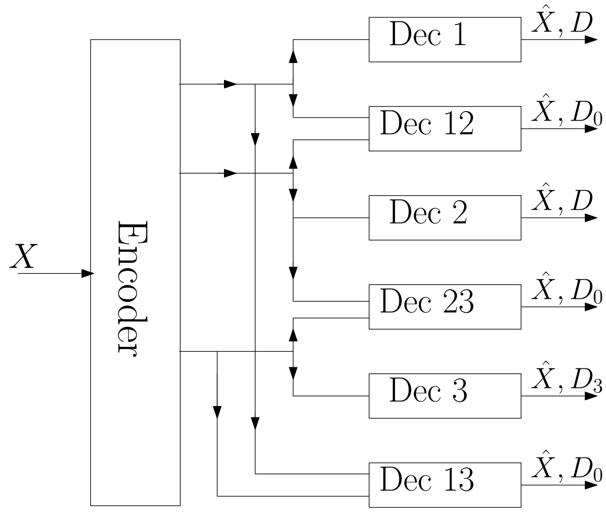

Figure 2.8: Scalar Source Example with Correlated Quantization

$D_{\{1\}} = D_{\{2\}}$ achievable was shown to be $\frac{1}{2}(1 - (1 - 2D_0)(2 - \sqrt{2}))$. The three-descriptions example is given as follows.

Example 36. *The source X is BSS, the distortion functions at decoders $\{1\}$, $\{2\}$, $\{1,2\}$, $\{1,3\}$ and $\{2,3\}$ are Hamming distortions, and the distortion function at decoder $\{3\}$ is the following general distortion function,*

$$d_{\{3\}}(x,\hat{x}) = \begin{cases} 0 & \textit{if } x = \hat{x} \\ \alpha & \textit{if } x = 0, \hat{x} = 1 \\ \beta & \textit{if } x = 1, \hat{x} = 0 \end{cases}$$

where α and β are positive real numbers. We are interested in achieving the RD vectors with the following projections:

$$R_1 = R_2 = \frac{1 - h_b(D_0)}{2}, D_{\{1\}} = D_{\{2\}} = \frac{1}{2}(1 - (1 - 2D_0)(2 - \sqrt{2})),$$
$$D_{\{1,2\}} = D_{\{1,3\}} = D_{\{2,3\}} = D_0, \tag{2.10}$$

Our objective is to evaluate the optimal $(R_3, D_{\{3\}})$ trade-off. The following lemma provides the RD vectors achievable using linear codes.

Lemma 37. *The RD vector in (4.14) is achievable using linear codes, as long as the following constraints are satisfied:*

$$R_3 \geq \frac{1}{2} + h_b\left(\sqrt{2}-1\right) - h_b\left(\frac{\sqrt{2}}{2}\right) - \frac{h_b\left(D_0\right)}{2} \tag{2.11}$$

$$D_{\{3\}} \geq \alpha\left(\sqrt{2}-1\right)D_0 + \beta\left(\left(\frac{3-2\sqrt{2}}{2}\right)(1-D_0) + \frac{D_0}{2}\right) \tag{2.12}$$

$$h_b\left(D_0\right) + 2h_b\left(\frac{\sqrt{2}}{2}\right) + h_b\left(2\left(\sqrt{2}-1\right)D_0\right) + h_b\left(2\left(\sqrt{2}-1\right)(1-D_0)\right) \geq 1. \tag{2.13}$$

Proof. Consider the following definition.

Definition 38. *Let $\mathbb{F}_q$ be a field. Consider 3 random variables X, U and V, where X is defined on an arbitrary finite set X, and U and V are defined on $\mathbb{F}_q$. Fix a PMF $P_{X,U,V}$ on $\mathsf{X} \times \mathbb{F}_q \times \mathbb{F}_q$. A sequence of code pairs (C_1, C_2), where $C_j \subset \mathbb{F}_q^n$ for $j = 1, 2$, is called P_{XUV}-covering if $\forall \epsilon > 0$,*

$$P(\{x^n | \exists (u^n, v^n) \in A_\epsilon^n(U, V|x^n) \cap C_1 \times C_2\}) \to 1 \text{ as } n \to \infty.$$

First, we derive new covering and packing bounds for joint quantization of a general source X (i.e. not necessarily binary), using two pairs of nested coset codes. Let (C_i, C_o) and (C_i, C'_o) be two pairs of nested coset codes with generator matrices G_1 and G_2 shown in Figure 2.9 which share the inner code C_i. If $r_i = 0$, the two codebooks are generated independently. On the other hand, if $r_o = r'_o = r_i$, the two codebooks are the same, so this construction generalizes the previous constructions.

Lemma 39 (Covering Lemma)**.** *For any P_{XUV} on $\mathsf{X}\times\mathbb{F}_q\times\mathbb{F}_q$ and rates r_o, r'_o and r_i satisfying (2.14)-(2.17), there exists a sequence of two pairs of nested coset codes (C_o, C_i) and (C'_o, C_i)*

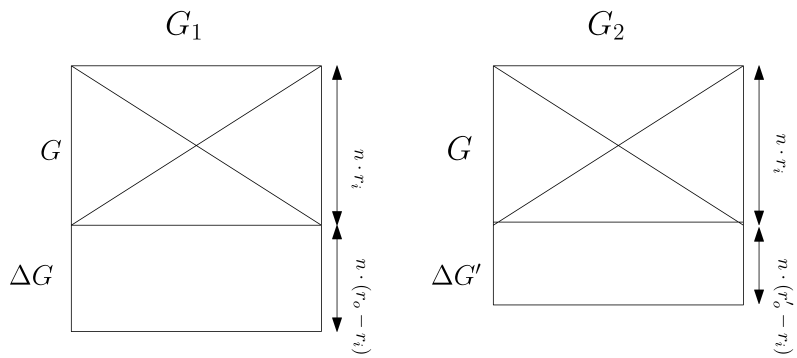

Figure 2.9: Codebook Construction for Lemma 39

which are P_{XUV}-covering.

$$r_o \geq \log q - H(U|X) \tag{2.14}$$

$$r_o' \geq \log q - H(V|X) \tag{2.15}$$

$$r_o + r_o' \geq 2\log q - H(U, V|X) \tag{2.16}$$

$$r_o + r_o' - r_i \geq \log q - H(\alpha U \oplus_q \beta V|X), \forall \alpha, \beta \in \mathbb{F}_q \backslash \{0\}, \tag{2.17}$$

Proof. See Section A.2.4 in the appendix. □

Remark 40. *The only difference between the new mutual covering bounds and the ones for independent codebook generation is the presence of the constraint (2.17). If $r_i = 0$, (2.17) is redundant, so we recover the mutual covering bounds for independent codebook generation as expected. If $r_i \neq 0$, (2.17) is non-redundant. There is an intuitive explanation for this additional bound. Define $C_3 = \alpha C_1 \oplus_q \beta C_2$. C_3 is a coset code with generator matrix $G_3 = [G^t\ G'^t\ \Delta G^t]^t$, and the size of this codebook is $2^{n(r_o + r_o' - r_i)}$. Suppose there are codevetors $\mathbf{c}_u \in C_1$ and $\mathbf{c}_v \in C_2$ jointly typical with $\mathbf{x}$ with respect to P_{UVX}, then $\alpha\mathbf{c}_1 \oplus_q \beta\mathbf{c}_2 \in C_3$ is jointly typical with $\mathbf{x}$ with respect to $P_{\alpha U \oplus_q \beta V, X}$. This implies that C_3 should have size at least $2^{n(\log q - H(\alpha U \oplus_q \beta V|X))}$ by the converse source coding theorem.*

Definition 41. *Let $\mathbb{F}_q$, U, V and X be as in Definition 14. A sequence of code pairs (C_1, C_2)*

and bin functions $B_i : C_i \to [1, 2^{n\rho_i}], i \in \{1, 2\}$ *is called* P_{XUV}*-packing if for all* $\epsilon > 0$,

$$P\left(\left\{x^n \,\middle|\, \begin{array}{c} \exists(c_u^n, c_v^n) \neq (c'^n_u, c'^n_v), \\ (c_u^n, c_v^n) \in A_\epsilon^n(U, V|X) \bigcap C_1 \times C_2, \\ (c'^n_u, c'^n_v) \in A_\epsilon^n(U, V) \bigcap C_1 \times C_2, \\ B_1(c_u^n) = B_1(c'^n_u), \;\; B_2(c_v^n) = B_2(c'^n_v) \end{array}\right\}\right) \to 0 \text{ as } n \to \infty.$$

Lemma 42 (Packing Lemma)**.** *For any* P_{XUV} *on* $\mathsf{X} \times \mathbb{F}_q \times \mathbb{F}_q$*, there exists a sequence of two pairs of nested coset codes* (C_o, C_i) *and* (C'_o, C_i) *and bin function* $B_i, i \in \{1, 2\}$ *which are* P_{XUV}*-packing, if* r_o, r'_o, ρ_1 *and* ρ_2 *satisfy*

$$r_o - \rho_1 \leq \log q - H(U|V), \tag{2.18}$$

$$r'_o - \rho_2 \leq \log q - H(V|U), \tag{2.19}$$

$$(r_o - \rho_1) + (r'_o - \rho_2) \leq 2\log q - H(U, V). \tag{2.20}$$

Proof. See Section A.2.5 in the appendix. □

We proceed with explaining the achievability scheme. Define the joint distribution as in the following tableon random variables $V_{\{1\}}, V_{\{2\}}$ and X.

X \ $V_{\{1\}}, V_{\{2\}}$	00	01	10	11
0	$\frac{1}{2}(1 - D_0)$	$\frac{\sqrt{2}-1}{2}D_0$	$\frac{\sqrt{2}-1}{2}D_0$	$\frac{3-2\sqrt{2}}{2}D_0$
1	$\frac{1}{2}D_0$	$\frac{\sqrt{2}-1}{2}(1 - D_0)$	$\frac{\sqrt{2}-1}{2}(1 - D_0)$	$\frac{3-2\sqrt{2}}{2}(1-D_0)$

Table 2.1: Joint distribution on X, $V_{\{1\}}$ and $V_{\{2\}}$.

Codebook Generation: Set $r = r_o = r'_o = r_i = 1 - \frac{H(V_{\{1\}},V_{\{2\}}|X)}{2} + \epsilon$, and $\rho_1 = \rho_2 = H(V_{\{1\}}) - \frac{H(V_{\{1\}},V_{\{2\}}|X)}{2} + \epsilon$ and $\rho_3 = H(V_{\{1\}} \oplus V_{\{2\}}) - \frac{H(V_{\{1\}},V_{\{2\}}|X)}{2} + \epsilon$. Construct a family of coset codes C with rate r. Also, construct three binning functions $B_i : C^n \to [1, 2^{n\rho_i}], i \in \{1, 2, 3\}$.

Encoding: Upon receiving source sequence x^n, the encoder finds $\mathbf{c}_1^n$ and $\mathbf{c}_2^n$ in the codebook, such that they are jointly typical with x^n with respect to $P_{V_{\{1\}},V_{\{2\}},X}$. Such a pair of codewords exists as long as the covering bounds in Lemma 39 are satisfied. In the case at hand it can be readily checked that r_o, r'_o and r_i satisfy the bounds. Description 1 carries the bin index of $\mathbf{c}_1^n$ using B_1, description 2 carries the bin index of $\mathbf{c}_2^n$ using B_2 and description 3 carries the bin index of $\mathbf{c}_1^n + \mathbf{c}_2^n$ using B_3.

Decoding: Decoder {1} receives the bin index carried by description 1, and reconstructs $\mathbf{c}_1^n$ as long as there is a unique codeword in the bin which is typical with respect to $P_{V_{\{1\}}}$. The following packing bound ensures correct decoding with arbitrarily small error:

$$H(V_{\{1\}}) \leq 1 - \rho_1 + r_i.$$

By the same arguments decoder {2} reconstructs $\mathbf{c}_2^n$ correctly. Decoder {3} reconstructs $\mathbf{c}_1^n + \mathbf{c}_2^n$ with arbitrarily small error since the following packing bound is satisfied:

$$H(V_{\{1\}} + V_{\{2\}}) \leq 1 - \rho_3 + r_i.$$

We conclude that all the decoders which receive two descriptions would have access to $\mathbf{c}_1^n$ and $\mathbf{c}_2^n$. Decoders {1}, {2} and {3} announce their decoded codewords as their reconstruction of the source. The reconstruction function at the decoders which receive two descriptions is given as follows:

$$\hat{x}_i = \begin{cases} 0 & c_{1i} = c_{2i} = 0 \\ 1 & \text{Otherwise} \end{cases}$$

This implies that the RD vector stated in the lemma is achieved from strong typicality.

□

The following lemma shows that some of the RD vectors in Lemma 21 are not achievable using the SSC scheme.

Lemma 43. *The RD vector in (4.14) is not achievable using the SSC scheme for the following values of α and β and when the equality holds in (4.16):*

$$\alpha = \log_2 \frac{1 - 2(\sqrt{2} - 1)D_0}{2(2 - \sqrt{2})D_0}, \quad \beta = -\log_2 \frac{1 - 2(\sqrt{2} - 1)(1 - D_0)}{2(2 - \sqrt{2})(1 - D_0)}$$

For example, $D_0 = 0.035, \alpha = 4.566$ and $\beta = 2.495$ satisfy the above constraints, where we have rounded the parameters up to the third decimal place.

Proof. See Section A.2.6 in the appendix. □

2.6 Achievable RD Region using Structured Codes

In this section, we provide a new achievable RD region for the general $l-$descriptions problem by enhancing the SSC coding scheme with a structured coding layer. We present this region in four stages. In the first stage, we prove that the SSC region can also be achieved using structured codes. In particular, we use independent nested coset codes for each auxiliary random variable, and exploit the pairwise independence of the codewords to show the achievability of the SSC region. In the subsequent stages, we add coding layers that facilitates the reconstruction of multi-variate functions of the auxiliary random variables. The improvements due to these additional layers comes from exploiting the algebraic structure of the codebooks. In the second stage, we only allow the reconstruction of a bivariate summation of codewords. In the third stage we extend this to a multi-variate summation of the codewords. In the fourth stage, we consider the general case involving the reconstruction of an arbitrary number of multi-variate summations at the decoders.

2.6.1 Stage 1: Achievability of the SSC Region Using Nested Coset Codes

Definition 44. *For a joint distribution P on random variables $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$ and X, and a set of reconstruction functions $g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}_{\mathsf{N}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$, the set $\mathcal{RD}_1(P, g_{\mathcal{L}})$ is defined as* *the set of RD vectors satisfying the following bounds for some non-negative real numbers* $(\rho_{\mathcal{M},i}, r_{o,\mathcal{M}})_{i\in\widetilde{\mathcal{M}},\mathcal{M}\in\mathbf{S}_{\mathsf{L}}}$ *:*

$$H(U_{\mathbf{M}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}} \tag{2.21}$$

$$H(U_{\mathbf{M}_{\mathsf{N}}}|U_{\mathbf{L}\cup\widetilde{\mathbf{M}}_{\mathsf{N}}}) \leq \sum_{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash\mathbf{L}\cup\widetilde{\mathbf{M}}_{\mathsf{N}}} (\log q + \sum_{j\in[1:L]} \rho_{\mathcal{M},j} - r_{o,\mathcal{M}}), \forall \mathbf{L} \subset \mathbf{M}_{\mathsf{N}}, \forall \mathsf{N} \in \mathcal{L} \tag{2.22}$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \qquad D_{\mathsf{N}} = E\{d_{\mathsf{N}}(h_{\mathsf{N}}(U_{\mathsf{N}}, X))\}.$$

where $r_{o,\mathcal{M}} \leq \log q, \forall \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$.

Theorem II.45. *The RD vector* $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},\mathsf{N}\in\mathcal{L}}$ *is achievable for the* $l-$*descriptions problem using nested coset codes, if there exists a distribution P and reconstruction functions* $g_{\mathcal{L}}$ *such that* $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},N\in\mathcal{L}} \in \mathcal{RD}_1(P, g_{\mathcal{L}})$.

Proof. The encoding and decoding steps are exactly the same as the ones in the proof of Theorem II.16. The only difference is in the codebook generation phase. In this phase, for every $\mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, we generate a coset code $C_{\mathcal{M}}$ with rate $r_{\mathcal{M}}$, generator matrix $G_{\mathcal{M}}$, and dither $b_{\mathcal{M}}$. $G_{\mathcal{M}}$ and $b_{\mathcal{M}}$ are generated randomly and uniformly for every $\mathcal{M}$. The bounds in (2.21) are the mutual covering bounds for independently generated coset codes. These bounds ensure encoding can be carried out without error. The bounds in (2.22) are the mutual packing bounds in each decoder. They ensure errorless decoding.

□

Lemma 46. *The RD region in Theorem II.45 is equal to the SSC RD region.*

Proof. See Section A.3.1 in the appendix. □

2.6.2 Stage 2: Reconstruction of a summation of two codebooks

In the first stage we constructed one codebook for each subset of the decoders. However, only the codebooks corresponding to the Sperner families of sets are shown to be non-redundant. We interpret this using the notion of common-information as defined by Gacs, Körner, Witsenhausen [11] [53]. Let $K(A_1; A_2)$ denote the common information between any two random variables A_1 and A_2. The common information among m random variables $A_1, A_2, \ldots, A_m$ is a vector of length $(2^m - m - 1)$ of information that is common among every subset of m random variables of size at least two. When $m = 3$, the common information is given by

$$[K(A_1; A_2; A_3), K(A_1; A_2), K(A_1; A_3), K(A_2; A_3)].$$

This was referred to as univariate common information in [29], as each of these components are characterized using univariate functions. We interpret the scheme in the first stage (SSC scheme) as capturing the common-information components among the random variables associated with $2^l - 1$ decoders using univariate functions.

For $m = 3$, this notion of common information was generalized using bivariate functions to the following seven-dimensional vector in [29]:

$$[K(A_1; A_2; A_3), K(A_1; A_2), K(A_1; A_3), K(A_2; A_3), K(A_1; A_2, A_3), K(A_2; A_1, A_3), K(A_3; A_1, A_2)].$$

There are seven degrees of freedom in having information common among 3 random variables. The latter three are called bivariate common information components as they are characterized using bivariate functions of random variables. In this sense, the addition of the structured coding layers in the next stages can be thought of as capturing the common-information among $2^l - 1$ decoders using bivariate and, more generally, multivariate functions.

We extend the notion of bivariate common information to $m > 3$ random variables as follows. To characterize a bivariate common information component, we consider three subsets of $\mathsf{N}_1, \mathsf{N}_2$ and N_3 of $\{1, 2, 3, \ldots, m\}$. Define $K(\tilde{A}_{\mathsf{N}_1}; \tilde{A}_{\mathsf{N}_2}, \tilde{A}_{\mathsf{N}_3})$ as a bivariate common information component among $A_1, A_2, \ldots, A_m$, where $\tilde{A}_{\mathsf{N}_i}$ is the information that is common among A_{N_i}. For example for $m = 4$, let $\mathsf{N}_1 = \{4\}$, $\mathsf{N}_2 = \{1,2\}$ and $\mathsf{N}_3 = \{3\}$. This characterizes the information in A_4 that can be computed by a conference via a bivariate function of (i) the information common between A_1 and A_2, and (ii) the information in A_3. This concept can be extended to define multivariate common information among m random variables.

We return to our discussion on the achievable RD region for the MD problem, where $m = 2^l - 1$. In the second stage, we aim to capture the bivariate common information among random variables associated with $2^l - 1$ decoders. In particular, we reconstruct a summation of two codebooks. From the above arguments, instead of one codebook for each subset of decoders as in the first stage, in this stage we need to construct one codebook for every triple of subsets of the decoders. For a given triple of sets of decoders, the third set of decoders reconstruct a bivariate summation of a random variable corresponding to the first subset and a random variable corresponding to the second subset of decoders. This is explained in more detail next. We add two new codebooks to the SSC scheme. The underlying random variables for these two codebooks are denoted by $V_{\mathcal{A}_1}$ and $V_{\mathcal{A}_2}$ where $\mathcal{A}_i \in \mathbf{S}_{\mathsf{L}}, i \in \{1,2\}$, $\mathcal{A}_1 \neq \mathcal{A}_2$. We construct two pairs of nested coset codes for these two random variables. The two nested coset codes have the same inner code. The codebook corresponding to $V_{\mathcal{A}_i}$ is decoded at decoder N if $\mathcal{A}_i \in \mathbf{M}_{\mathsf{N}}$, furthermore, the sum of the two codebooks is decoded at decoder N if $\mathcal{A}_3 \in \mathbf{M}_{\mathsf{N}} \backslash \{\mathcal{A}_1, \mathcal{A}_2\}$, where $\mathcal{A}_3$ is an element of $\mathbf{S}_{\mathsf{L}}$. For example, let us choose $\mathcal{A}_i = \{\{i\}\}, i \in \{1,2,3\}$. In this case the first codebook is decoded whenever description 1 is received, the second codebook is decoded if description 2 is received, and the sum is decoded whenever description 3 is received. This corresponds to the coding schemes we presented for example 29, where $V_{\mathcal{A}_1} = X + N_\delta$ and $V_{\mathcal{A}_2} = Z + N'_\delta$. The following theorem describes the achievable RD region using this scheme.

Definition 47. *For any three distinct families $\mathcal{A}_i \in \mathbf{S}_{\mathsf{L}}, i = 1,2,3$, and for a joint distribution P on random variables $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, $V_{\mathcal{A}_j}, j \in \{1,2\}$, and X, where the underlying alphabet for all auxiliary random variables is the field $\mathbb{F}_q$, and a set of reconstruction functions*

$g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}_{\mathsf{N}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$*, the set* $\mathcal{RD}_2(P, g_{\mathcal{L}})$ *is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers* $(\rho_{\mathcal{M},i}, r_{o,\mathcal{M}})_{i\in\widetilde{\mathcal{M}},\mathcal{M}\in\mathbf{S}_{\mathsf{L}}}$ *and* $\rho_{o,\mathcal{A}_j,i}, r'_{o,\mathcal{A}_j}, i \in \widetilde{\mathcal{A}}_i, j \in \{1,2,3\}$ *and* r_i*:*

$$H(U_{\mathbf{M}}V_{\mathbf{E}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{E}\in\mathbf{E}} (\log q - r'_{o,\mathcal{E}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \mathbf{E} \subset \mathbf{A} \tag{2.23}$$

$$H(U_{\mathbf{M}}, W_{\mathcal{A}_3,\alpha,\beta}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \log q - r'_{o,\mathcal{A}_3}, \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \forall \alpha,\beta \in \mathbb{F}_q\backslash\{0\} \tag{2.24}$$

$$H([U,V,W]_{\overline{\mathbf{M}}_{\mathsf{N}}}|[U,V,W]_{\widehat{\mathbf{M}}_{\mathsf{N}}\cup\overline{\mathbf{L}}}) \leq$$

$$\sum_{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash\widetilde{\mathbf{M}}_{\mathsf{N}}\cup\mathbf{L}} (\log q + \sum_{j\in\widetilde{\mathcal{M}}} \rho_{\mathcal{M},j} - r_{o,\mathcal{M}}) + \sum_{\substack{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash\widetilde{\mathbf{M}}_{\mathsf{N}}\cup\overline{\mathbf{L}} \\ \bigcap\{\mathcal{A}_i|i\in[1,3]\}}} (\log q + \sum_{j\in\widetilde{\mathcal{M}}} \rho_{o,\mathcal{M},j} - r'_{o,\mathcal{M}}), \quad \forall \overline{\mathbf{L}} \subset \overline{\mathbf{M}}_{\mathsf{N}} \tag{2.25}$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \qquad D_{\mathsf{N}} = E\{d_{\mathsf{N}}(h_{\mathsf{N}}(U_{\mathsf{N}}, X))\}. \tag{2.26}$$

where (a) $\mathbf{A} \triangleq \{\mathcal{A}_1, \mathcal{A}_2\}$*, (b)* $\overline{\mathbf{M}}_{\mathsf{N}} \triangleq (\mathbf{M}_{\mathsf{N}}, \{\mathcal{A}_j, j \in \{1,2\}|\mathcal{A}_j \in \mathbf{M}_{\mathsf{N}}\}, \{\{\mathcal{A}_3, 1, 1\}|\mathcal{A}_3 \in \mathbf{M}_{\mathsf{N}}\})$*, (c)* $\widehat{\mathbf{M}}_{\mathsf{N}} \triangleq \bigcup_{\mathsf{N}'\subsetneq\mathsf{N}} \overline{\mathbf{M}}_{\mathsf{N}'}$*, (d)* $r'_{o,\mathcal{A}_3} \triangleq r'_{o,\mathcal{A}_1} + r'_{o,\mathcal{A}_2} - r_i$*, (e)* $r_{o,\mathcal{M}} \leq \log q$*, and (f)* $W_{\mathcal{A}_3,\alpha,\beta} \triangleq \alpha V_{\mathcal{A}_1} + \beta V_{\mathcal{A}_2}$[8].

Theorem II.48. *The RD vector* $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},\mathsf{N}\in\mathcal{L}}$ *is achievable for the* $l-$*descriptions problem, if there exists a distribution P and reconstruction functions* $g_{\mathcal{L}}$ *such that* $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},N\in\mathcal{L}} \in \mathcal{RD}_2(P, g_{\mathcal{L}})$*.*

Before providing the proof we explain the bounds in the new RD region. (4.17) and (4.19) are the mutual covering and packing bounds which are also present in the Theorem II.16, respectively. (4.18) is a generalization of the additional covering bound derived in the Lemma in 39. Note that the common component among decoders $\mathsf{N} \in \mathcal{A}_1$ is the pair $(U_{\mathcal{A}_1}, V_{\mathcal{A}_1})$, and similarly for $\mathcal{A}_2$. The common component among decoders $\mathsf{N} \in \mathcal{A}_3$ is the pair $(U_{\mathcal{A}_1}, W_{\mathcal{A}_3,1,1})$, and observe that $W_{\mathcal{A}_3,1,1} = V_{\mathcal{A}_1} + V_{\mathcal{A}_2}$.

Proof. Given a joint distribution $P_{\mathbf{U},\mathbf{V},X}$, and codebook and binning rates satisfying the

[8] We have used the script $\mathcal{A}$ to denote subscripts of random variables throughout the chapter. However, the collection $\{\mathcal{A}_3, \alpha, \beta\}$ is used as the subscript for W since the random variable is defined using α and β.

bounds in the theorem we prove achievability of the RD vector in (4.20).

Codebook Generation: Fix blocklength n. For every $\mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, independently generate a linear code $C_{\mathcal{M}}$ with size $2^{nr_{o,\mathcal{M}}}$. Also generate two nested coset codes $C_{\mathcal{A}_j} = (C_i, C_{o,\mathcal{A}_j})$, $j \in \{1, 2\}$ where the inner code has rate r_i and the outer codes have rates $r'_{o,\mathcal{A}_j}$. Define the set of codewords $C_{o,\mathcal{A}_3} \triangleq C_{o,\mathcal{A}_1} + C_{o,\mathcal{A}_2}$. The size of $C_{o,\mathcal{A}_3}$ is $2^{nr'_{o,\mathcal{A}_3}}$, where $r'_{o,\mathcal{A}_3} = r'_{o,\mathcal{A}_1} + r'_{o,\mathcal{A}_2} - r_i$. For the ith description bin the codebook $C_{\mathcal{M}}$ randomly and uniformly with rate $2^{n\rho_{\mathcal{M},i}}$.

Encoding: Upon receiving the source vector X^n, the encoder finds a jointly-typical set of codewords $c_{\mathcal{M}}$. Each description carries the *bin-indices* of all of the corresponding codewords. The encoder declares an error if there is no jointly typical set of codewords available.

Decoding: Having received the bin-indices from descriptions $i \in \mathsf{N}$, decoder N tries to find a set of jointly typical codewords $c_{\mathcal{M}}, \mathcal{M} \in \overline{\mathbf{M}}_{\mathsf{N}}$. If the set of codewords is not unique, the decoder declares error.

In order for the encoder to find a set of jointly typical codewords, the mutual covering bounds (4.17) and (4.18) should hold. This is a generalization of the result in lemma 39 and we omit the proof for brevity. The bounds in (4.19) are the mutual packing bounds at each decoder.

□

Remark 49. *Here we have considered the general case where $\mathcal{A}_i$ are chosen arbitrarily from $\mathbf{S}_{\mathsf{L}}$. It turns out that only certain choices of $\mathcal{A}_i$ would give non-redundant codebooks and thus provide improvements over the SSC scheme. One can show that the codebooks are redundant if $\exists N \in \mathcal{A}_1 \cup \mathcal{A}_2, N' \in \mathcal{A}_3$ such that $N \subset N'$. For example take $\mathcal{A}_1 = \{\{1\}, \{3\}\}$, $\mathcal{A}_2 = \{\{2\}\}$ and $\mathcal{A}_3 = \{\{2, 3\}\}$.*

2.6.3 Stage 3: Reconstruction of a summation of arbitrary number of codebooks

In this section we reconstruct a multi-variate summation of an arbitrary number m of random variables at one decoder where $m \in \mathsf{L}$ and the summation is with respect to a finite field $\mathbb{F}_q$. Following the steps in the previous section, we add m new codebooks to the original SSC scheme. Let $\mathsf{M} \triangleq [1, m]$. The underlying random variables for these codebooks are denoted by $V_{\mathcal{A}_k}, k \in \mathsf{M}$. The random variable $V_{\mathcal{A}_k}$ is decoded at decoder N if $\mathcal{A}_k \in \mathbf{M}_{\mathsf{N}}$. We take the families $\mathcal{A}_k, k \in \mathsf{M}$ to be distinct. The random variable $\sum_{i\in\mathsf{M}} V_{\mathcal{A}_k}$ is decoded at decoder N if $\mathcal{A}_{m+1} \in \mathbf{M}_{\mathsf{N}}$, where $\mathcal{A}_{m+1}$ is an element of $\mathbf{S}_{\mathsf{L}}$. The following theorem describes the achievable RD region:

Definition 50. *For any $m \in \mathsf{L}$, and $m+1$ distinct families $\mathcal{A}_i \in \mathbf{S}_{\mathsf{L}}, i \in [1, m+1]$, and for a joint distribution P on random variables $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, $V_{\mathcal{A}_k}, k \in \mathsf{M}$ and X, where the underlying alphabet for the auxiliary random variables is the field $\mathbb{F}_q$, and a set of reconstruction functions $g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}_{\mathsf{N}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$, the set $\mathcal{RD}_3(P, g_{\mathcal{L}})$ is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers $(\rho_{\mathcal{M},i}, r_{o,\mathcal{M}})_{i\in\widetilde{\mathcal{M}},\mathcal{M}\in\mathbf{S}_{\mathsf{L}}}$ and $\rho_{o,\mathcal{A}_k,i}, \rho_{o,\mathcal{A}_{m+1},i}, r'_{o,\mathcal{A}_k}, i \in \widetilde{A}_k, k \in [1, m+1]$ and $r_{i,\alpha_J}, J \subset \mathsf{M}$:*

$$H(U_{\mathbf{M}}V_{\mathbf{E}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{E}\in\mathbf{E}} (\log q - r'_{o,\mathcal{E}}) +, \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \mathbf{E} \subset \mathbf{A}, \tag{2.27}$$

$$H(U_{\mathbf{M}}W_{\mathbf{F}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{F}\in\mathbf{F}} (\log q - r'_{o,\mathcal{F}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \mathbf{F} \subset \mathbf{B}, \tag{2.28}$$

$$H([U, V, W]_{\overline{\mathbf{M}}_{\mathsf{N}}}|[U, V, W]_{\widehat{\mathbf{M}}_{\mathsf{N}}\cup\overline{\mathbf{L}}}) \leq$$

$$\sum_{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash\widehat{\mathbf{M}}_{\mathsf{N}}\cup\mathbf{L}} (\log q + \sum_{j\in\widetilde{\mathcal{M}}} \rho_{\mathcal{M},j} - r_{o,\mathcal{M}}) + \sum_{\substack{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash\widehat{\mathbf{M}}_{\mathsf{N}}\cup\overline{\mathbf{L}} \\ \bigcap\{\mathcal{A}_i|i\in[1,m+1]\}}} (\log q + \sum_{j\in\widetilde{\mathcal{M}}} \rho_{o,\mathcal{M},j} - r'_{o,\mathcal{M}}), \quad \forall \overline{\mathbf{L}} \subset \overline{\mathbf{M}}_{\mathsf{N}} \tag{2.29}$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \qquad D_{\mathsf{N}} = E\{d_{\mathsf{N}}(h_{\mathsf{N}}(U_{\mathsf{N}}, X))\}. \tag{2.30}$$

where (a) $\mathbf{A} = \{\mathcal{A}_k, k \in \mathsf{M}\}$, (b) $\mathbf{B} = \{(\mathcal{A}_{m+1}, \alpha_{\mathsf{M}})|\alpha_{\mathsf{M}} \in \mathbb{F}_q^m\}$, (c) $r'_{o,\mathcal{A}_{m+1},\alpha_{\mathsf{M}}} = \sum_{k\in\mathsf{J}} r'_{o,\mathcal{A}_k} - r_{i,\alpha_{\mathsf{J}}}, \mathsf{J} = \{k|\alpha_k \neq 0\}$, (d) $\sum_{\mathsf{J}':\mathsf{J}\subset\mathsf{J}'} r_{i,\alpha_{\mathsf{J}'}} \leq r_{i,\alpha_{\mathsf{J}}}, \forall \mathsf{J} \subset \mathsf{M}$,
(e) $\overline{\mathbf{M}}_{\mathsf{N}} = (\mathbf{M}_{\mathsf{N}}, \{\mathcal{A}_k|\mathcal{A}_k \in \mathbf{M}_{\mathsf{N}}\}, \{(\mathcal{A}_{m+1}, \alpha_{\mathsf{M}})|\mathcal{A}_{m+1} \in \mathbf{M}_{\mathsf{N}}, \alpha_i = 1, i \in \mathsf{M}\})$, (f) $\widehat{\mathbf{M}}_{\mathsf{N}} = \bigcup_{\mathsf{N}'\subsetneq\mathsf{N}} \overline{\mathbf{M}}_{\mathsf{N}'}$, (g) $r_{o,\mathcal{M}} \leq \log q$, and (h) $W_{\mathcal{A}_{m+1},\alpha_{\mathsf{M}}} = \sum_{i\in\mathsf{M}} \alpha_i V_{\mathcal{A}_k}$.

Theorem II.51. *The RD vector $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},\mathsf{N}\in\mathcal{L}}$ is achievable for the l−descriptions problem, if there exists a distribution P and reconstruction functions $g_{\mathcal{L}}$ such that $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},N\in\mathcal{L}} \in$*

$\mathcal{RD}_3(P, g_{\mathcal{L}})$.

Toward proving the theorem we need the following definition.

Definition 52. *A set of m coset codes $C^n_{\mathcal{A}_k}, k \in \mathsf{M}$ is called an ensemble of nested coset codes with parameter $(r_J)_{J\subset \mathsf{M}}$ if the size of the intersection $C_{\mathcal{A}_\mathsf{J}} \triangleq \bigcap_{k\in \mathsf{J}} C_{\mathcal{A}_k}$ is equal to 2^{nr_J} for all $\mathsf{J} \subset \mathsf{M}$.*

It is straightforward to show that one can always generate an ensemble of nested coset codes $C_{\mathcal{A}_k}, k \in \mathsf{M}$ with parameter $(r_{i,\alpha_J})_{J\subset\mathsf{M}}$ as long as $\sum_{J':J\subset J'} r_{i,\alpha_{J'}} \leq r_{i,\alpha_J}, \forall J \subset \mathsf{M}$. It is enough to choose the rows of the generator matrices of $C_{\mathcal{A}_k}, k \in J$ such that they have nr_{i,α_J} common rows, similar to the case of Figure 2.9.

Proof. We provide an outline of the proof. The codebook generation for codebooks $C_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_\mathsf{L}$ is similar to the previous scheme. For random variables $V_{\mathcal{A}_k}, k \in \mathsf{M}$ we construct an ensemble of nested coset codes $C_{\mathcal{A}_k}, k \in \mathsf{M}$ with parameter $(r_{i,\alpha_J})_{J\subset\mathsf{M}}$. The encoder chooses a set of codewords from all the codebooks that is jointly typical with the source sequence. The following is a generalized covering lemma which shows that if (2.27) and (2.28) is satisfied such a set of codewords exists.

Definition 53. *Let $\mathbb{F}_q$ be a field and define $\mathcal{M} \triangleq \{\{1\}, \{2\}, \ldots, \{m\}\}$. Consider $m + 1$ random variables X, $V_{\{i\}}, i \in M$, where X is defined on an arbitrary finite set X and $V_{\{i\}}$ are defined on $\mathbb{F}_q$. Fix a PMF $P_{X,V_\mathcal{M}}$ on $\mathsf{X}\times\mathbb{F}_q^m$. A sequence of m-tuples of codebooks $(C^n_{\{i\}})_{\{i\}\in\mathcal{M}}$ is called $P_{XV_\mathcal{M}}$-covering if:*

$$\forall\epsilon > 0, P(\{x^n|\exists v^n_\mathcal{M} \in A^n_\epsilon(V_\mathcal{M}|x^n) \cap \Pi_{\{i\}\in\mathcal{M}} C_{\{i\}}\}) \to 1 \textit{ as } n \to \infty.$$

Lemma 54 (Covering Lemma)**.** *For any $P_{X,V_\mathcal{M}}$ on $\mathsf{X}\times\mathbb{F}_q^\mathcal{M}$ and rates $r_{o,\{j\}}, \{j\} \in \mathcal{M}$ satisfying (2.31)-(2.33), there exists a sequence of ensemble of nested coset codes $C^n_\mathcal{M}$ with parameter $(r_{i,\mathsf{J}})_{\mathsf{J}\subset\mathsf{M}}$, which are $P_{X,V_\mathcal{M}}$-covering.*

$$H(V_{\mathcal{J}}|X) \geq \sum_{\{j\}\in\mathcal{J}} (\log q - r_{o,\{j\}}), \forall \mathcal{J} \subset \mathcal{M} \tag{2.31}$$

$$H(W_{\mathcal{K}}|X) \geq \sum_{\alpha_{\mathsf{M}}\in\mathcal{K}} (\log q - r_{o,\alpha_{\mathsf{M}}}), \forall \mathcal{J} \subset \mathcal{M}, \mathcal{K} \subset \mathcal{N} \tag{2.32}$$

$$\sum_{\mathsf{J}':\mathsf{J}\subset\mathsf{J}'} r_{i,\mathsf{J}'} \leq r_{i,\mathsf{J}}, \forall \mathsf{J} \subset \mathsf{M}, \tag{2.33}$$

where, (a) $\mathcal{N} \triangleq \{\alpha_{\mathsf{M}} \in \mathbb{F}_q^m\}$*, (b)* $W_{\alpha_{\mathsf{M}}} \triangleq \sum_{j\in\mathsf{M}} \alpha_j V_{\{j\}}$ *and (c)* $r_{o,\alpha_{\mathsf{M}}} \triangleq \sum_{j\in\mathsf{J}} r_{o,\{j\}} - r_{i,\mathsf{J}}, \mathsf{J} = \{k|\alpha_k \neq 0\}$.

Proof. The proof of the lemma follows the same steps as in lemma 39. We provide the intuition behind the proof. Given that there is a set of codewords in the codebooks $C_{\{j\}}, j \in \mathsf{M}$ which are jointly typical with the source sequence, for any linear combination $C \triangleq \sum_{j\in\mathsf{M}} \alpha_j C_{\{j\}}$ there is a codeword which is jointly typical with the random variables $X, V_{\mathcal{M}}, W_{\mathcal{N}}, \forall \mathcal{M}, \mathcal{N}$. From a PtP perspective, the rate of codebook C must satisfy (2.31) and (2.32). This rate can be calculated by counting the number of rows in the generator matrix of C which is $nr_{o,\alpha_{\mathsf{M}}}$. □

The packing bounds at each encoder can be written in the same way as in the previous section and are given in (2.29). **B** is defined such that $W_{\mathbf{B}}$ is the set of all possible linear combinations of $V_{\mathcal{A}_k}$'s.

□

2.6.4 Stage 4: Reconstruction of an Arbitrary Number of Summations of Arbitrary Lengths

In this section for completeness, we provide a coding scheme where we reconstruct multi-variate summations of random variables at an arbitrary number of decoders, and these summations each have arbitrary lengths. Of course, due to the large number of random variables the coding scheme becomes extremely complicated. Let the number of the summations be s, and for each summation, let the length of the summation be denoted by $m_i \in L, i \in [1, s]$. Define the sets $\mathsf{S} \triangleq [1, s]$ and $\mathsf{M}_i \triangleq [1, m_i], i \in \mathsf{S}$. Following the steps in the previous sections, we add m_i new codebooks for each summation. The underlying random variables for these codebooks are denoted by $V_{\mathcal{A}_{k,i}}, k \in \mathsf{M}_i, i \in \mathsf{S}$. The random variable $V_{\mathcal{A}_{k,i}}$ is decoded at decoder N if $\mathcal{A}_{k,i} \in \mathbf{M}_{\mathsf{N}}$. Fix the prime number q_i. The random variable $\sum_{j\in\mathsf{M}_i} V_{\mathcal{A}_{k,i}}$ is decoded at decoder N if $\mathcal{A}_{m_i+1,i} \in \mathbf{M}_{\mathsf{N}}$, where the summation is carried out in the finite field $\mathbb{F}_{q_i}$. The following theorem describes the achievable RD region.

Definition 55. *For a joint distribution P on random variables $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, $V_{\mathcal{A}_{k,i}}, k \in \mathsf{M}_i, i \in \mathsf{S}$ and X, where the underlying alphabet the auxiliary random variables is the field $\mathbb{F}_q$, and a set of reconstruction functions $g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}_{\mathsf{N}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$, the set $\mathcal{RD}_{linear}(P, g_{\mathcal{L}})$ is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers $(\rho_{\mathcal{M},i}, r_{o,\mathcal{M}})_{i\in\widetilde{\mathcal{M}},\mathcal{M}\in\mathbf{S}_{\mathsf{L}}}$ and $\rho_{o,\mathcal{A}_{k,i},j_k}, \rho_{o,\mathcal{A}_{m_i+1,i},j_k}, r_{o,\mathcal{A}_{k,i}}, j_k \in \widetilde{A}_{k,i}, k \in [1, m_i+1], i \in \mathsf{S}$ and $r_{i,\alpha_{J,i}}, J \subset \mathsf{M}_i, i \in \mathsf{S}$:*

$$H(U_{\mathbf{M}}, V_{\mathbf{E}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{E}\in\mathbf{E}} (\log q - r_{o,\mathcal{E}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \mathbf{E} \subset \mathbf{A}, \tag{2.34}$$

$$H(U_{\mathbf{M}}, W_{\mathbf{F}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{F}\in\mathbf{F}} (\log q - r_{o,\mathcal{F}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \mathbf{F} \subset \mathbf{B}, \tag{2.35}$$

$$\begin{aligned} &H([U, V, W]_{\overline{\mathbf{M}}_{\mathsf{N}}}|[U, V, W]_{\widehat{\mathbf{M}}_{\mathsf{N}}\cup\overline{\mathbf{L}}}) \leq \\ &\sum_{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash\widehat{\mathbf{M}}_{\mathsf{N}}\cup\mathbf{L}} (\log q + \sum_{j\in\widetilde{\mathcal{M}}} \rho_{\mathcal{M},j} - r_{o,\mathcal{M}}) + \sum_{\substack{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash\widehat{\mathbf{M}}_{\mathsf{N}}\cup\overline{\mathbf{L}} \\ \bigcap\{\mathcal{A}_{\{i\},s}|i\in[1,m_s+1],s\in\mathsf{S}\}}} (\log q + \sum_{j\in\widetilde{\mathcal{M}}} \rho_{o,\mathcal{M},j} - r'_{o,\mathcal{M}}), \quad \forall \overline{\mathbf{L}} \subset \overline{\mathbf{M}}_{\mathsf{N}} \end{aligned} \tag{2.36}$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \quad D_{\mathsf{N}} = E\{d_{\mathsf{N}}(h_{\mathsf{N}}(U_{\mathsf{N}}, X))\}. \tag{2.37}$$

where (a) $\mathbf{A} = \bigcup_{i\in\mathsf{S}}\{\mathcal{A}_{k,i}|k \in \mathsf{M}_i\}$, (b) $\mathbf{B} = \bigcup_{i\in\mathsf{S}}\{(\mathcal{A}_{m_i+1,i}, \alpha_{\mathsf{M}_i,i})|\alpha_{j,i} \in \mathbb{F}_{q_i}\}$, (c) $r_{o,\mathcal{A}_{m_i+1},\alpha_{\mathsf{M}_i,i}} = \sum_{k\in\mathsf{J}} r_{o,\mathcal{A}_{k,i}} - r_{i,\alpha_{\mathsf{J}_i,i}}, \mathsf{J}_i = \{k|\alpha_{k,i} \neq 0\}, i \in \mathsf{S}$, (d) $\sum_{\mathsf{J}':\mathsf{J}\subset\mathsf{J}'} r_{i,\alpha_{\mathsf{J}',i}} \leq r_{i,\alpha_{\mathsf{J},i}}, \forall \mathsf{J} \subset \mathsf{M}_i, i \in \mathsf{S}$, (e) $\overline{\mathbf{M}}_{\mathsf{N}} = (\mathbf{M}_{\mathsf{N}}, \{\mathcal{A}_{k,i}|k \in \mathsf{M}_i, i \in \mathsf{S}, \mathcal{A}_{k,i} \in \mathbf{M}_{\mathsf{N}}\}, \{(\mathcal{A}_{m_i+1,i}, \alpha_{\mathsf{M}_i,i})|\mathcal{A}_{m_i+1} \in \mathbf{M}_{\mathsf{N}}, \alpha_{k,i} = 1\})$, (f) $\widehat{\mathbf{M}}_{\mathsf{N}} = \bigcup_{\mathsf{N}'\subsetneq\mathsf{N}} \overline{\mathbf{M}}_{\mathsf{N}'}$, (g) $r_{o,\mathcal{M}} \leq \log q$ and (h) $W_{\mathcal{A}_{m+1,i},\alpha_{\mathsf{M}_i,i}} = \sum_{j=1}^{m_i} \alpha_{j,i} V_{\mathcal{A}_{k,i}}$.

Theorem II.56. *The RD vector $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},\mathsf{N}\in\mathcal{L}}$ is achievable for the $l-$descriptions problem,* *if there exists a distribution P and reconstruction functions $g_{\mathcal{L}}$ such that* $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L}, N\in\mathcal{L}} \in \mathcal{RD}_{linear}(P, g_{\mathcal{L}})$.

Proof. This is a straightforward generalization of the previous step, since the proof is similar, it is omitted. □

Remark 57. *Similar to Theorem 24 one can identify the non-redundant codebooks in the above scheme. One can show that a large number of possible codebooks become redundant in this case as well.*

CHAPTER III

Lattice Construction for Multi-terminal Source Coding

3.1 Introduction

Lattice quantizers have been of great interest in compression of continuous real-valued sources [31, 43]. A lattice quantizer is a quantizer, whose set of outputs are closed under additions and subtractions. Similar to linear codes, in the point-to-point (PtP) communication settings, the interest towards such codes is mainly due to reduced complexity of encoding and decoding. In multiterminal communications, the significance of lattice codes is augmented because they give performance gains over unstructured codes in terms of achievable rate-distortion (RD) regions. The improvement is due to the structure of lattice quantizers. Encoders utilizing such quantizers can transmit summations of different quantizations more efficiently than those using unstructured quantizers. These gains are observed in a variety of multiterminal settings [31], [43], [36], and they are analogous to those seen when using linear codes for quantizing discrete sources [30], [29], [20] as discussed in the previous chapter.

Direct performance analysis of lattice coding techniques in general multiterminal source coding setups turns out to be difficult. The analysis is usually carried out for quantization of Gaussian sources with Gaussian test channels. Hence, characterizations of inner bounds to the optimal RD region for general sources and test channels are not available. In this section we provide a new method for lattice generation and lattice quantization. The advantage of this new method is that it allows for straightforward transfer of the schemes in the discrete communication regime to the continuous problems. The method involves two steps. First the source is discretized using a clipping and quantization operation. Next, the discrete problems solved using linear codes, similar to the previous chapter. The main difficulty in analyzing such a scheme is to prove the convergence of the achievable rate-distortion region for the discrete problem, to the continuous one as a finer quantization is used in the first step. In Section 3.2, we provide a summary of results on general information measures for continuous sources. Next, in Section 3.3, we show the convergence for the PtP source coding problem. Lastly, as an example of how to extend the results to multiterminal source coding, we investigate the multiple descriptions problem. As discussed in the previous chapter, in the discrete alphabet case, if a pair of nested coset quantizers with the same inner code are used in the encoder, there would be a strict improvement in the achievable RD region for the L-descriptions problem when $L \geq 3$. In other words, structured codes provide asymptotic RD performance gains over unstructured codes. Here, we consider continuous sources and provide a new achievable RD region for the L-descriptions problem using random lattice codes. These are new multiterminal lattice codes. This means that these lattice quantizers induce correlated quantization noises; they can not be decomposed as a collection of PtP lattice quantizers. An example of such codes is pair of nested lattice quantizers (i.e. quantizers whose output sequences have a nested lattice structure) with a shared inner lattice code. We show that using a pair of nested lattice quantizers with the same inner code gives strict improvements over the SSC region in the continuous source case.

3.2 General Information Measures

First, let us define the extensions of information theoretic measures such as mutual information, entropy, and divergence to general continuous sources. The definitions and results provided here can be found in [32]. The following gives the definition of the probability density function (PDF) of a random variable:

Definition 1. *Let S be a continuous random variable defined by the probability space $(P_S, \mathbb{F}_S, \mathcal{S})$. Also, assume that the probability measure P_S is absolutely continuous with respect to the Labesgue measure L, denoted by $P_S \ll L$, the PDF is defined as the Radon-Nikodym derivative of P_S with respect to L:*

$$f_S \triangleq \frac{\delta P_S}{\delta L}.$$

Remark 2. *The PDF may not generally exist. Particularly, if the probability measure P_S is not absolutely continuous with respect to the Labesgue measure L, then the above definition does not provide the formula for the PDF of S.*

For the random variable S, if the PDF exists, the entropy is defined by the following natural extension of the discrete entropy:

$$h(S) \triangleq -\int f_S(s) \log f_S(s) ds,$$

if the integral exists. The following gives the extension of the Kullback-Leibler divergence:

Definition 3. *Consider the pair of continuous random variables (S, T). Let P_S (Q_T) be the probability measure corresponding to S (T), respectively. Assume that the joint PDF f_{ST} exists. Also, assume that $P_S \ll Q_T$, i.e. $f_S(x) = 0$ whenever $f_T(x) = 0$. The relative entropy between P_S and Q_T is defined as follows:*

$$D(P \| Q) = \int f_S(x) \log \frac{f_S(x)}{f_T(x)} dx,$$

if the integral exists.

The mutual information between S and T is defined as:

$$I(S;T) \triangleq D(P_{ST} \| P_S Q_T).$$

So far, we have defined the information measures for the random variables with the assumption that the PDF exists. The following definition gives an alternative characterization of the relative entropy which does not require the existence of the PDF.

Definition 4. *Let (S,T) be a pair of random variables, both defined on the measurable space $(\mathcal{X}, \mathbb{F}_X)$, where $\mathcal{X}$ is a set, and $\mathbb{F}$ is a σ-algebra defined on $\mathcal{X}$. Let the corresponding probability measures be P_S and Q_T. Also, assume $P_S \ll Q_T$. The relative entropy is defined as follows:*

$$D(P_S \| Q_T) \triangleq \sup_q \sum_{i=1}^{K_q} P_S(E_i) \log \frac{P_S(E_i)}{Q_T(E_i)},$$

where the supremum is over all finite measurable partitions $q = \{E_1, E_2, \cdots, E_{K_q}\}$ of $\mathcal{X}$.

In [32] it is shown that Definition 4 is an extension of Definition 3. In other words, the formula in the above definition is equal to the one in Definition 3, when $P_S \ll Q_T \ll L$.

3.3 Lattice Codes for PtP Source Coding

In this section we prove the existence of optimality achieving lattice codes for the problem of PtP source coding. The methods developed in this section are used in the rest of the chapter to construct coding schemes for multiterminal source coding.

Before proceeding with our results, we give a brief summary of nested lattice constructions which are used in this section. We start by explaining coset codes, and nested coset codes constructed over discrete alphabets. Let q be a prime number. Let $\mathbb{Z}_q$ denote the unique finite field of size q. From the previous chapter, a coset code is a shifted version of a linear code and is characterized by a generator matrix $G_{k\times n}$ and a dither B^n:

$$C = \{u^k G_{k\times n} + B^n | u^k \in \mathbb{Z}_q^k\}.$$

A pair of coset codes (C_i, C_o), are called nested if C_i lies inside C_o. C_o and C_i are called the outer and inner codes, respectively. A nested coset code is characterized by two generator matrices $G_{k\times n}$ and $\Delta G_{l\times n}$ and a dither B^n. Here $(G_{k\times n}, B^n)$ is a characterization for C_i and $([G, \Delta G]^t, B^n)$ characterizes C_o. Each of these shifted version of C_i is called a bin of C_o and is denoted by $\mathcal{B}_m$:

$$\mathcal{B}_m = \{a\mathbf{G} + m\mathbf{\Delta G} + B^n | a \in \mathbb{Z}_q^k\}.$$

A lattice code is a subset of $\mathbb{R}^n$ which is closed under integer addition. In the previous section, we presented a method to generate lattice quantizers using linear codes. For a coset code C, grid step γ, and discrete alphabet size q, the corresponding lattice would be characterized follows:

$$\bar{\Lambda}(C, \gamma, q) = \bigcup_{v\in\gamma p\mathbb{Z}^n} \{v + \Lambda(C, \gamma, q)\},$$

where

$$\Lambda(C, \gamma, q) = \{\gamma(c_i - \frac{q-1}{2})_{i=1}^n | (c_i)_{i=1}^n \in C\}.$$

Nested lattice codes are the continuous duals of coset codes, and are also defined in a similar fashion by constructing a pair (Λ_i, Λ_o) from an underlying pair of nested coset codes (C_i, C_o). Similar to nested coset codes, for $m \in \mathbb{Z}_q^l$, bin m can be defined as:

$$\bar{\mathcal{B}}_m = \{\gamma(c_i - \frac{q-1}{2})_{i=1}^n | (c_i)_{i=1}^n \in \mathcal{B}_m\} \tag{3.1}$$

where $\mathcal{B}_m$ is a bin in the underlying nested coset code.

We proceed with explaining the PtP source coding problem. Consider the PtP source coding problem depicted in Figure 5.2. The continuous memoryless source X and its recon-

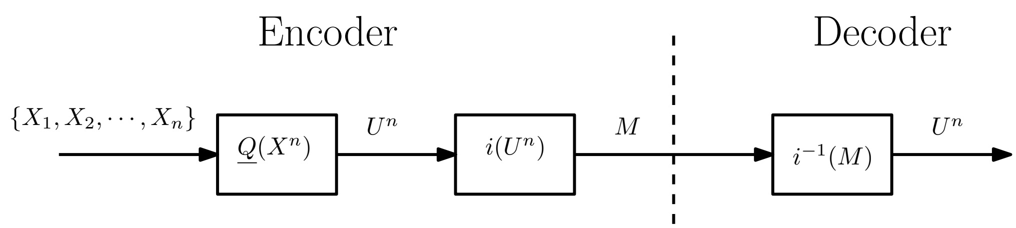

Figure 3.1: Point-to-point source coding for continuous sources

struction U are defined on the probability space $(\mathbb{R}, \mathbb{B}_X, P_X)$, where $\mathbb{B}$ is the Borel σ-algebra defined on the set of real numbers $\mathbb{R}$. The source X is being fed to an encoder. The encoder utilizes the mapping $\underline{Q} : \mathcal{X}^n \to \mathcal{U}^n$ to compress the source sequence. The image of $\underline{Q}$ is indexed by the bijection $i : Im(\underline{Q}) \to [1, |Im(\underline{Q})|]$. The index $M \triangleq i(Q(X^n))$ is sent to the decoder. The decoder reconstructs the compressed sequence $U^n \triangleq i^{-1}(M) = Q(X^n)$. The efficiency of the reconstruction is evaluated based on the separable distortion criteria $d_n : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty)$. The separability property means that $d_n(x^n, u^n) = \sum_{i\in[1,n]} d(x_i, u_i)$, where $d(x, u) \triangleq d_1(x, u)$. The rate of transmission is defined as $R \triangleq \frac{1}{n} \log |Im(\underline{Q})|$, and the average distortion is defined as $\frac{1}{n}\mathbb{E}(d_n(X^n, U^n))$. The goal is to choose $\underline{Q}$ such that the rate-distortion tradeoff is optimized. Note that the choice of the bijection 'i' is irrelevant to the performance of the system. The following Theorem, gives an achievable RD region for this setup.

Theorem III.5. *Let X be a continuous memoryless source and let $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$, be a separable distortion criteria. Let U be a random variable defined on the probability space $(\mathbb{R}, \mathbb{B})$, such that its probability measure P_U is absolutely continuous with respect to P_X. Furthermore, assume the distortion criteria $d : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ satisfies the following properties:*

1) $d(x, u)$ is a jointly continuous function for all x, and u.

2) For any given source distribution P_X, the expected distortion $\mathbb{E}(d(X, u))$ is finite whenever u is finite.

The rate-distortion pair $(R, D) = (r, \mathbb{E}_{X,U}(d(X, U)))$ is achievable for all $r \geq I(U; X)$ using *coset-lattice codebooks*[1]*, where achievability is defined in the usual way.*

Remark 6. *The conditions in Theorem III.5 can be explained as follows. The first condition is a smoothness condition on the distortion function. The second condition is usually assumed in source coding. As an example the rth power difference distortion function $d(x, u) = |x - u|^r$ satisfies this assumption.*

Proof. The proof involves two steps: 1) Discretization, 2) Quantization.

Step 1: In this step we turn the problem into a discrete source coding problem. The discretization process involves two steps. First, the source is clipped so that its value is bounded. Next, it is finely quantized using a uniform scalar grid. Fix $n \in \mathbb{N}$, and reals $l_1, l_2 > 0$. Define the set $\mathbb{Z}_n \triangleq \frac{1}{n}\mathbb{Z}$. The gridding and clipping functions are defined below:

Definition 7. *The gridding function $G_n : \mathbb{R} \to \mathbb{Z}_n$ is given as follows:*

$$G_n(x) = \underset{a \in \mathbb{Z}_n}{argmin} |x - a|.$$

The clipping function $C_{l,u} : \mathbb{R} \to \mathbb{R}$ is defined as follows:

$$C_{l_1,l_2}(x) = max\{min\{l_2, x\}, -l_1\}.$$

Define the random variable X_d as $G_n(C_{l_1,l_2}(X))$. The index of the discretized version of the reconstruction U is defined similarly and denoted by U_d. The random variables X_d and U_d take values from the set $\mathbb{Z}_{l_1,l_2,n} \triangleq [-l_1, l_2] \bigcap \mathbb{Z}_n$. Define $L \triangleq \lceil (l_1 + l_2)2^n \rceil - 1$ The quantization cells $A_{l_1,l_2,n}(i), i \in [0, L]$ of this discretization process are:

$$A_{l_1,l_2,n}(i) = \begin{cases} \left(-\infty, \left\lceil \frac{l_1}{2^n} \right\rceil + \frac{1}{2^{n+1}}\right], & i = 0 \\ \left(\left\lceil \frac{l_1}{2^n} \right\rceil + \frac{2i-1}{2^{n+1}}, \left\lceil \frac{l_1}{2^n} \right\rceil + \frac{2i+1}{2^{n+1}}\right], & i \in [1, L-1] \\ \left(\left\lceil \frac{l_2}{2^n} \right\rceil - \frac{1}{2^{n+1}}, \infty\right), & i = L. \end{cases}$$

[1]This means that the set of reconstruction vectors used at the decoder have a nested-lattice structure.

Also for a quantization cell $A_{l_1,l_2,n} = (a_1, a_2]$, define the quantization reconstruction $a_{l_1,l_2,n}$ as $\frac{a_1+a_2}{2}$. For the first cell, the quantization reconstruction is defined as $\left\lceil \frac{l_1}{2^n} \right\rceil$, for the last cell it is defined as $\left\lfloor \frac{l_2}{2^n} \right\rfloor$.

Step 2: Define the discrete distortion function $d_Q(i, j) \triangleq d(a_{l,u,n}(i), a_{l,u,n}(j)), i, j \in [1, L]$. Also, define the discrete sources $X_Q \triangleq a_{l_1,l_2,n}(X_d)$, and $U_Q \triangleq a_{l_1,l_2,n}(U_d)$. The joint distribution between the discrete random variables (X_Q, U_Q) is defined based on the Markov chain $X_Q \leftrightarrow X \leftrightarrow U \leftrightarrow U_Q$. The Markov chain completely characterizes the joint distribution since P_{X,X_Q}, P_{U,U_Q}, and $P_{X,U}$ are defined in the above steps. Consider the discrete PtP source coding problem with source X_Q, and distortion criterion d_Q. Using nested coset codes as in the previous chapter, we can achieve the rate distortion vector (r, d), where $r > I(X_Q; U_Q)$, and $d > \mathbb{E}(d_Q(X_Q, U_Q))$. In the quantization step we transmit U_Q to the decoder with rate r.

The coding scheme can be summarized in the following way. The encoder receives a sequence X^m, it then produces X_Q^m as the discretized version of the source sequence. It uses the test channel defined by the joint distribution P_{U_Q,X_Q} to produce the quantized version U_Q^m using nested coset codes with rate r similar to the previous chapter. Note that by this construction the Markov chain $X^m \leftrightarrow X_Q^m \leftrightarrow U_Q^m$ holds. The decoder receives the vector U_Q^m as the reconstruction of X^m. In order to complete the proof we need to show that as $n, l_1, l_2 \to \infty$, we have $I(X_Q; U_Q) \leq I(X, U)$, and $\mathbb{E}(d(X^m, U_Q^m)) \leq \mathbb{E}(d(X^m, U^m))$, as $m \to \infty$. First we prove the convergence of the expected distortion.

Lemma 8. *For every $\epsilon > 0$, there exist l_1, l_2, n large enough, such that*

$$\frac{1}{m}\sum_{i=1}^{m} \mathbb{E}(d(X_i, U_{Q,i})) \leq \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}(d(X_i, U_i)) + \epsilon.$$

Proof. We provide an outline of the proof. Using the assumption that the distortion function is jointly continuous, we conclude the following:

$$\forall i \in [1, m], \forall \delta > 0, \exists l_1, l_2, n, |\mathbb{E}(d(X_i, U_{Q,i})) - \mathbb{E}(d(X_i, U_i)| < \delta.$$

In summary, the above inequality is proved using 1) The distortion function is bounded for bounded inputs, 2) $P(X_i > l_2) \to 0$ as $l_2 \to \infty$, 3) $P(X_i < -l_1) \to 0$ as $l_1 \to \infty$, and finally 4) $\forall x, y \in A_{l_1,l_2,n}(j), |x-y| \to 0$ as $n \to \infty$. Summing the above over i completes the proof. □

The next lemma proves the mutual informations converge as well.

Lemma 9. *For every $\epsilon > 0$, there exist l_1, l_2, n large enough, such that*

$$I(X_Q; U_Q) \leq I(X; U) + \epsilon.$$

Proof. The proof follows from 4, since $I(X; U)$ is the supremum over all quantizers and X_Q, U_Q are calculated by using a specific quantizer the statement is correct. □

□

Remark 10. *Note that using this method, the set of possible reconstructions of the source sequence at the decoder are closed under additions and subtractions. So, we have constructed a lattice quantization method. The method circumvents the usual lattice construction complexities by using the linear codes available for discrete alphabets.*

3.4 Multiple-Descriptions Coding for Continuous Sources

In this section we extend the method presented in the previous section to provide an achievable RD region for the MD problem with general continuous sources. First, we prove that the extension of the SSC rate-distortion region is achievable. Next, we use lattice codes to improve upon the extension of the SSC scheme similar to the previous chapter. The following theorem proves the achievability of the SSC scheme:

Definition 11. *Consider the MD problem with the continuous source X, and separable distortion criteria $d_{\mathsf{N}}, \mathsf{N} \in \mathcal{L}$ satisfying the conditions in Theorem III.5. For a joint distribution P on random variables $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$ and X and a set of reconstruction functions*

$g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}_{\mathsf{N}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$*, the set* $\mathcal{RD}_{SSCCTS}(P, g_{\mathcal{L}})$ *is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers* $(\rho_{\mathcal{M},i}, r_{\mathcal{M}})_{i\in\widetilde{\mathcal{M}},\mathcal{M}\in\mathbf{S}_{\mathsf{L}}}$ *:*

$$H(U_{\mathbf{M}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (H(U_{\mathcal{M}}) - r_{\mathcal{M}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \tag{3.2}$$

$$H(U_{\mathbf{M}_{\mathsf{N}}}|U_{\mathbf{L}\cup\widetilde{\mathbf{M}}_{\mathsf{N}}}) \leq \sum_{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash(\mathbf{L}\cup\widetilde{\mathbf{M}}_{\mathsf{N}})} (H(U_{\mathcal{M}}) + \sum_{i\in\widetilde{\mathcal{M}}} \rho_{\mathcal{M},i} - r_{\mathcal{M}}), \forall \mathbf{L} \subset \mathbf{M}_{\mathsf{N}}, \forall \mathsf{N} \in \mathcal{L}, \tag{3.3}$$

$$r_{\mathcal{M}} \leq H(U_{\mathcal{M}}), \forall \mathcal{M} \in \mathbf{S}_{\mathsf{L}},$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \quad D_{\mathsf{N}} = E\{d_{\mathsf{N}}(g_{\mathsf{N}}(U_{\mathsf{N}}), X)\}, \tag{3.4}$$

where $\mathbf{M}_{\mathsf{N}}$ *is the set of all codebooks decoded at decoder* N*, that is* $\mathbf{M}_{\mathsf{N}} \triangleq \{\mathcal{M} \in \mathbf{S}_{\mathsf{L}} | \exists \mathsf{N}' \subset \mathsf{N}, \mathsf{N}' \in \mathcal{M}\}$*, and* $\widetilde{\mathbf{M}}_{\mathsf{N}}$ *denotes the set of all codebooks decoded at decoders* $\mathsf{N}_p \subsetneq \mathsf{N}$ *which receive subsets of descriptions received by* N*, that is* $\widetilde{\mathbf{M}}_{\mathsf{N}} \triangleq \bigcup_{\mathsf{N}_p \subsetneq \mathsf{N}} \mathbf{M}_{\mathsf{N}_p}$*.*

Theorem III.12. *The RD vector* $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},\mathsf{N}\in\mathcal{L}}$ *is achievable for the* $l-$*descriptions problem, if there exists a distribution* P *and reconstruction functions* $g_{\mathcal{L}}$ *such that* $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},N\in\mathcal{L}} \in \mathcal{RD}_{SSCCTS}(P, g_{\mathcal{L}})$*.*

The proof follows by a similar coding scheme as in the previous section. The source is first discretized, then the coding scheme in chapter 2 is used to transmit the discrete source. We showed in the previous section that mutual informations converge. However, the bounds in the above region are not in terms of mutual informations. The next lemma shows that after the Fourier-Motzkin elimination these bounds can be written in terms of mutual informations, and hence convergence follows.

Lemma 13. *The bounds in the improved CMS region can be written in terms of mutual informations.*

Proof. In the appendix. □

The above scheme uses independently generated linear codes, which leads to independently generated lattices. We will show in the next section that this scheme is sub-optimal compared to a scheme which utilizes nested lattices.

3.5 Improvements Using Nested Lattice Quantizers

In this section, we provide two jointly Gaussian examples in which using nested lattice codes gives gains in terms of achievable rate-distortion.

3.5.1 Example 1

The set-up is shown in figure 3.2. Here $(X, Z) \sim N(0, 1)$. The distortion function for the individual decoders is mean squared error. Decoder 1 and 2 want to reconstruct their respective source with MSE less than or equal P, and decoder 3 wants to reconstruct $Y = X + Z$ with distortion $2P$. The distortion constraint in the joint decoders is as follows:

$$E(([X,Z] - [\hat{X}_{12}, \hat{Z}_{12}])^t([X,Z] - [\hat{X}_{12}, \hat{Z}_{12}])) \leq \begin{bmatrix} P & 1 \\ 1 & P \end{bmatrix}$$

$$E(([X,Y] - [\hat{X}_{13}, \hat{Y}_{13}])^t([X,Y] - [\hat{X}_{13}, \hat{Y}_{13}])) \leq \begin{bmatrix} P & P \\ P & 2P \end{bmatrix}$$

$$E(([X,Y] - [\hat{X}_{23}, \hat{Y}_{23}])^t([X,Y] - [\hat{X}_{23}, \hat{Y}_{23}])) \leq \begin{bmatrix} P & P \\ P & 2P \end{bmatrix}$$

Where $\hat{A}_{ij}, A \in \{X, Z, Y\}, i, j \in \{1, 2, 3\}$ is the reconstruction of A at decoder ij.

Theorem III.14. *The following rate triple is achievable for the distortions mentioned above using the nested lattice structure.*

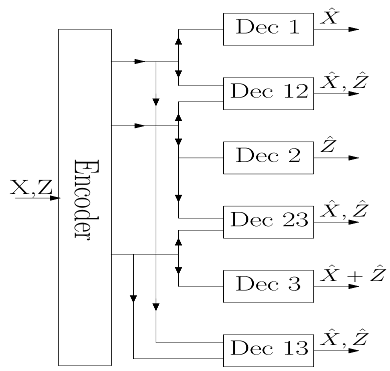

Figure 3.2: Example for Vector Gaussian Source

$$(R_1, R_2, R_3) = (\frac{1}{2}log(\frac{1}{p}), \frac{1}{2}log(\frac{1}{p}), \frac{1}{2}log(\frac{2}{p}))$$

Proof. In the appendix. □

Theorem III.15. *The above rates are not achievable using the enhanced CMS scheme.*

Proof. In the appendix. □

Note that in the above example, as a result of the independence between the two source components X and Z, the extra covering bound in (2.17) is redundant. But this is not always the case, to illustrate this point we investigate another example.

3.5.2 Example 2

Assume the source X is scalar Gaussian with distribution $N(0, 1)$. Consider the random variables U and V which are jointly Gaussian with X and have the following covariance matrix ($\frac{1}{2} < P < \frac{2}{3}$):

$$Cov([X, U, V]) = \begin{pmatrix} 1 & 1-P & 1-P \\ 1-P & 1-P & 0 \\ 1-P & 0 & 1-P \end{pmatrix}$$

We intend to transmit U on the first description, V on the second description and $U+V$ on the third description. In this case the new covering bound is not redundant. To see this note that from (B.5) the covering bound is non-redundant if $I(U{+}V;V|X) - I(U;V|X) < 0$. Simplifying the inequality, the bound is non-redundant if $Var(V|XU) < Var(V|X, U+V)$. From the formula for conditional variance we have:

$$\begin{aligned}
&Var(V|X,U) = \frac{p(1-p)}{2}, Var(V|X, U+V) = \frac{1-p}{2},\\
&Var(V|XU) < Var(V|X, U+V) \iff p < \frac{2}{3}
\end{aligned}$$

Which shows the bound is non-redundant in this setting. We calculate the achievable rates using nested lattices. FME gives:

$$\begin{aligned}
R_1 =& R_2 = \max\{\frac{I(UV;X)}{2}, I(U;X) - I(\alpha U+\beta V;V|X) + I(U;V|X)\},\\
&R_3 = R_1 - H(U) + H(U+V)
\end{aligned}$$

We have:

$$\begin{aligned}
&I(UV;X) = \frac{1}{2}\log(\frac{1}{2p-1}), \quad I(U;X) = \frac{1}{2}\log\frac{1}{p},\\
&-I(\alpha U+\beta V;V|X) + I(U;V|X) = \frac{1}{2}\log(\frac{\alpha^2 p}{\alpha^2+\beta^2-(\alpha+\beta)^2(1-p)})
\end{aligned}$$

Hence $R_1 = R_2 = \max(\frac{1}{2}\log(\frac{p^2}{p+(1-p)^2}), \frac{1}{4}\log(\frac{1}{2p-1})), R_3 = R_1 + \frac{1}{2}$. And the distortions are $D_1 = D_2 = p, D_3 = 2p, D_{12} = D_{13} = D_{23} = 2p-1$.

3.6 An Improved Achievable Region Using Nested Lattice Codes

In this section we generalize the gains observed in the previous section, and provide a new achievable RD region for the MD problem with continuous sources. The following theorem provides the new RD region.

Definition 16. *Consider a joint distribution P on continuous random variables $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, $V_{\mathcal{A}_{k,i}}, k \in \mathsf{M}_i, i \in \mathsf{S}$ and X satisfying the conditions in Theorem 4, and a set of reconstruction functions $g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}_{\mathsf{N}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$, the set $\mathcal{RD}_{linearCTS}(P, g_{\mathcal{L}})$ is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers $(\rho_{\mathcal{M},i}, r_{o,\mathcal{M}})_{i\in\widetilde{\mathcal{M}},\mathcal{M}\in\mathbf{S}_{\mathsf{L}}}$ and $\rho_{o,\mathcal{A}_{k,i},j_k}, \rho_{o,\mathcal{A}_{m_i+1,i},j_k}, r_{o,\mathcal{A}_{k,i}}, j_k \in \widetilde{A}_{k,i}, k \in [1, m_i + 1], i \in \mathsf{S}$ and $r_{i,\alpha_{J,i}}, J \subset \mathsf{M}_i, i \in \mathsf{S}$, when $q \to \infty$:*

$$H(U_{\mathbf{M}}, V_{\mathbf{E}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{E}\in\mathbf{E}} (\log q - r_{o,\mathcal{E}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \mathbf{E} \subset \mathbf{A}, \tag{3.5}$$

$$H(U_{\mathbf{M}}, W_{\mathbf{F}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{F}\in\mathbf{F}} (\log q - r_{o,\mathcal{F}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \mathbf{F} \subset \mathbf{B}, \tag{3.6}$$

$$H([U, V, W]_{\overline{\mathbf{M}}_{\mathsf{N}}}|[U, V, W]_{\widehat{\mathbf{M}}_{\mathsf{N}}\cup\overline{\mathbf{L}}}) \leq$$

$$\sum_{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash\widetilde{\mathbf{M}}_{\mathsf{N}}\cup\mathbf{L}} (\log q + \sum_{j\in\widetilde{\mathcal{M}}} \rho_{\mathcal{M},j} - r_{o,\mathcal{M}}) + \sum_{\substack{\mathcal{M}\in\mathbf{M}_{\mathsf{N}}\backslash\widetilde{\mathbf{M}}_{\mathsf{N}}\cup\overline{\mathbf{L}} \\ \bigcap\{\mathcal{A}_{\{i\},s}|i\in[1,m_s+1],s\in\mathsf{S}\}}} (\log q + \sum_{j\in\widetilde{\mathcal{M}}} \rho_{o,\mathcal{M},j} - r'_{o,\mathcal{M}}), \quad \forall \overline{\mathbf{L}} \subset \overline{\mathbf{M}}_{\mathsf{N}} \tag{3.7}$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \quad D_{\mathsf{N}} = E\{d_{\mathsf{N}}(h_{\mathsf{N}}(U_{\mathsf{N}}, X))\}. \tag{3.8}$$

where (a) $\mathbf{A} = \bigcup_{i\in\mathsf{S}}\{\mathcal{A}_{k,i}|k \in \mathsf{M}_i\}$, (b) $\mathbf{B} = \bigcup_{i\in\mathsf{S}}\{(\mathcal{A}_{m_i+1,i}, \alpha_{\mathsf{M}_i,i})|\alpha_{j,i} \in \mathbb{F}_{q_i}\}$, (c) $r_{o,\mathcal{A}_{m_i+1},\alpha_{\mathsf{M}_i,i}} = \sum_{k\in\mathsf{J}} r_{o,\mathcal{A}_{k,i}} - r_{i,\alpha_{\mathsf{J}_i,i}}, \mathsf{J}_i = \{k|\alpha_{k,i} \neq 0\}, i \in \mathsf{S}$, (d) $\sum_{\mathsf{J}':\mathsf{J}\subset\mathsf{J}'} r_{i,\alpha_{\mathsf{J}',i}} \leq r_{i,\alpha_{\mathsf{J},i}}, \forall \mathsf{J} \subset \mathsf{M}_i, i \in \mathsf{S}$, (e) $\overline{\mathbf{M}}_{\mathsf{N}} = (\mathbf{M}_{\mathsf{N}}, \{\mathcal{A}_{k,i}|k \in \mathsf{M}_i, i \in \mathsf{S}, \mathcal{A}_{k,i} \in \mathbf{M}_{\mathsf{N}}\}, \{(\mathcal{A}_{m_i+1,i}, \alpha_{\mathsf{M}_i,i})|\mathcal{A}_{m_i+1} \in \mathbf{M}_{\mathsf{N}}, \alpha_{k,i} = 1\})$, (f) $\widehat{\mathbf{M}}_{\mathsf{N}} = \bigcup_{\mathsf{N}'\subsetneq\mathsf{N}} \overline{\mathbf{M}}_{\mathsf{N}'}$, (g) $r_{o,\mathcal{M}} \leq \log q$ and (h) $W_{\mathcal{A}_{m+1,i},\alpha_{\mathsf{M}_i,i}} = \sum_{j=1}^{m_i} \alpha_{j,i} V_{\mathcal{A}_{k,i}}$.

Theorem III.17. *The RD vector $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},\mathsf{N}\in\mathcal{L}}$ is achievable for the $l-$descriptions problem, if there exists a distribution P and reconstruction functions $g_{\mathcal{L}}$ such that $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},N\in\mathcal{L}} \in \mathcal{RD}_{linearCTS}(P, g_{\mathcal{L}})$.*

Similar to the previous chapter, since codes with the same inner codes are used, the mutual covering bounds are not enough to ensure the existence of jointly typical sequences at the encoder and the uniqueness of the decoded sequences at the decoders. We need to use the covering bounds in Lemma 39, as $q \to \infty$. It follows by the same arguments as in the previous section that the region is achievable using lattice quantizers. To see this, note that there is only one additional covering bound in the new system of inequalities, but the difference between this bound and the previous bounds can be written in terms of mutual informations.

Lemma 18. *The new bounds for the achievable RD region using nested coset codes can be written in terms of mutual informations.*

Proof. In the appendix. □

So, the bounds from the discretized version converge to the continuous bounds. Since the RD vectors in Example 3.5.1 are inside the RD region described above, this provides a larger RD region than that of the SSC scheme. Furthermore, the RD regions presented in this chapter are achievable for sources with arbitrary non-Gaussian probability measures.

CHAPTER IV

Quasi-Linear Structures for Source and Channel Coding

4.1 Introduction

So far, we have used codebooks which are closed with respect to linear combinations to improve the performance in different multiterminal problems. In this chapter, we introduce a new class of structured code ensembles called QLC's whose 'closedness' with respect to an operation can be controlled. A QLC is a subset of a linear code. It is difficult to analyze the performance of arbitrary subsets of linear codes. Instead, we provide a method for constructing specific subsets of these codes by putting single-letter distributions on the indices of the codewords. We analyze the performance of the resulting ensemble. We are able to characterize the asymptotic performance using single-letter information quantities. By choosing the single-letter distribution on the indices one can operate anywhere in the spectrum between the two extremes: linear codes and unstructured codes. First, we show that QLC's achieve the Shannon rate-distortion function for discrete memoryless sources with bounded additive distortion functions. Next, we show through an MD source coding example that application of QLC's gives a better inner bound to the optimal achievable RD region compared to the best known MD coding strategies. We provide a new inner bound to the optimal achievable RD region for the general MD problem. The method builds upon the Sperner Set Coding (SSC) scheme introduced in chapter 2. Next, we apply QLC's to the problem of transmission of massages in the interference channel.

The interference channel problem describes a setup where multiple pairs of transmitters and receivers share a communication medium. Each receiver is only interested in decoding the message from its corresponding transmitter. However, since the channel is shared, signals from other senders interfere with the desired signal at each decoder. The presence of interfering signals adds new dimensions to this problem in terms of strategies that can be used as compared to point-to-point (PtP) communication. For example, the encoders can *cooperate* with each other by choosing their channel inputs in a way that would facilitate their joint communication. It turns out that, often, this cooperation requires an encoder to employ a strategy which may be sub-optimal from its own PtP communications perspective. In this chapter, we investigate this tradeoff and develop a new class of codes which allow for more efficient cooperation between the transmitters.

Characterizing the capacity region for the general IC has been a challenge for decades. Even in the simplest case of the two user IC, the capacity region is only known in special cases [37][8]. The best known achievable region for the IC was due to Han and Kobayashi [16]. However, recently it was shown that the Han-Kobayashi (HK) rate region is suboptimal [5][29]. Particularly, when there are more than two transmitter-receiver pairs, the natural generalization of the HK strategy can be improved upon by inducing structure in the codebooks used in the scheme [29]. Structured codes such as linear codes and group codes enable the encoders to align their signals more efficiently. This in turn reduces interference at the decoders. Such codebook structures have also proven to give gains in other multiterminal communication problems [18]-[41].

The idea of interference alignment was proposed for managing interference when there are three or more users. Initially, the technique was proposed by Maddah-Ali et. al. [23] for the MIMO X channel, and for the multi-user IC by Jafar and Cadambe [6]. The interference alignment strategy was developed for cases of additive interference and uniform channel inputs over finite fields. However, it turns out that alignment is not always beneficial to the users in terms of achievable rates. Consider the example in Figure 4.1. Intuitively, it would

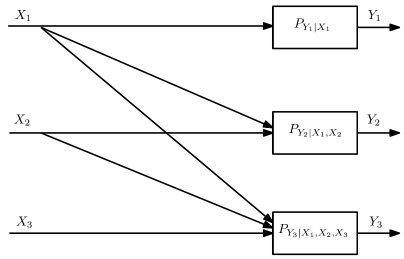

Figure 4.1: A setup where interference alignment is beneficial to user 3 but harmful for user 2.

be beneficial to align the input from users 1 and 2 to reduce interference at decoder 3. However, if users 1 and 2 align their signals, it becomes harder for decoder 2 to distinguish between the two inputs. One might suggest that the problem could be alleviated if users 1 and 2 designed their codebooks in a way that they would "look" aligned at decoder 3 based on $P_{Y_3|X_1,X_2,X_3}$, but at the same time they would seem different at decoder 2 based on $P_{Y_2|X_1,X_2}$. In this chapter, we show that linear codes lack the necessary flexibility for such a strategy. Based on this intuition, we propose a new class of structured codes. Using these codes we derive an achievable rate region which improves upon the best known achievable region for the three user IC given in [29].

4.2 Notation

Random variables are denoted by capital letters such as X, U, their realizations by small letters x, u, and their corresponding alphabets (finite) by sans-serif typeface X, U, respectively. Numbers are represented by small letters such as l, k. We denote the the field of size q by $\mathbb{F}_q$. We represent the field addition by $\oplus$ and the addition on real numbers by $+$. The set of numbers $\{1, 2, \ldots, m\}$ is represented by $[1, m]$. Vectors are represented by the bold type-face such as $\mathbf{u}, \mathbf{b}$. For a random variable X, $A_\epsilon^n(X)$ denotes the set of ϵ-typical sequence of length n with respect to P_X, where we use the definition of frequency typicality. Let q be a prime number. For $l \in \mathbb{N}$, consider $U_i, i \in [1, m]$ i.i.d random variables with distribution P_U defined on a field $\mathbb{F}_q$. $U^{\otimes l}$ denotes a random variable which has the same distribution as

$\sum_{i\in[1,l]} U_i$ where the summation is over $\mathbb{F}_q$.

4.3 New Codebook Structures

In this section, we define our new coding structures and provide the foundations for their analysis. These coding structures are used in the next section to derive larger achievable regions for a variety of multiterminal communication problems.

4.3.1 Quasi Linear Codes

First, we define a new ensemble of codes called QLC's. The ensemble is defined over a finite field $\mathbb{F}_q$ where q is a prime number. The codebooks are constructed by first generating the coset of a linear code called a coset code.

Definition 1. *A (k, n) coset code C is characterized by a generator matrix $G_{k\times n}$ and a dither $\mathbf{b}^n$ defined on the field $\mathbb{F}_q$. C is defined as follows:*

$$C \triangleq \{\mathbf{u}G + \mathbf{b}|\mathbf{u} \in \mathbb{F}_q^k\}.$$

The rate of the codebook is defined as $R = \frac{k}{n}\log q$.

A QLC is a subset of a linear code, the following provides the definition of a QLC:

Definition 2. *A (k, n) QLC is characterized by a generator matrix $G_{k\times n}$, a dither $\mathbf{b}^n$ and a set U defined on $\mathbb{F}_q$. The codebook is defined as follows:*

$$C \triangleq \{\mathbf{u}G + \mathbf{b}|\mathbf{u} \in \mathsf{U}\}.$$

If G is injective on U, then the rate of the code is given by $R = \frac{1}{n}\log|C| = \frac{1}{n}\log|\mathsf{U}|$.

It is difficult to derive single-letter characterizations for the performance of coding schemes using QLCs with general sets U. In this work, we focus on the case when U is a cartesian product of typical sets. More precisely, let $m \in \mathbb{N}$, $\epsilon \in \mathbb{R}^+$, and $U_1, U_2, ..., U_m$ be random variables defined on $\mathbb{F}_q$. Consider natural numbers $k_i, i \in [1, m]$ such that $\sum_{i \in [1,m]} k_i = k$. Construct generator matrices G_i with dimension $k_i \times n$. We are interested in analyzing the performance of codebooks of the following form:

$$C \triangleq \{ \sum_{i \in [1,m]} \mathbf{u}_i G_i + \mathbf{b} | \mathbf{u}_i \in A_\epsilon^{k_i}(U_i)\},$$

where $A_\epsilon^{k_i}(U_i)$ is the set of frequency ϵ-typical sequences of length k_i with respect to distribution P_{U_i}. In this case, the rate of the code is $R = \sum_{i \in [1,m]} \frac{1}{n} \log |A_\epsilon^{k_i}(U_i)|$ which approaches $\sum_{i \in [1,m]} \frac{k_i}{n} H(U_i)$ as $n \to \infty, \epsilon \to 0$.

Remark 3. *In the notation of Definition 2,* $G = [G_1^t | G_2^t | ... | G_m^t]^t$ *and* $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m)$.

Remark 4. *While we concentrate on the case when* $\mathsf{U} = \bigotimes_{i \in [1,m]} A_\epsilon^{k_i}(U_i)$*, it is possible to carry out performance analysis of such an ensemble of codebooks when* U *is taken to be more general. For example, a more general result can be obtained by taking* U *to be a joint typical set of vectors of correlated random variables* $U_1, U_2, ..., U_m$.

Remark 5. *A* (k, n) *linear code is only defined for* $k \leq n$*. When constructing a QLC, we take* $R \leq \log q$*. This ensures that for a randomly and uniformly generated matrix* G*, the resulting mapping is injective on* U *with high probability. However, there is no additional restrictions on the* k_i*'s. As an example, let* $m = 1$*, one can take* $k_1 > n$ *and* U_1 *such that* $\frac{k_1}{n} H(U_1) < \log q$*. In this case* $\{\mathbf{u}_1 G_1 + B | \mathbf{u}_1 \in A_\epsilon^{k_1}(U_1)\}$ *is a codebook whose rate is close to* $\frac{k_1}{n} H(U_1)$ *for large* n*. Note that* G_1 *is not injective on the vector space* $\mathbb{F}_q^{k_1}$.

4.3.2 Nested Quasi Linear Codes

In this section, we define Nested Quasi Linear Codes (NQLC). The following gives the definition for a pair of Nested Linear Codes (NLC):

Definition 6. *For natural numbers $k_i < k_o, k'_o < n$, let $G_{k_i \times n}, \Delta G_{(k_o - k_i)\times n}$ and $\Delta G'_{(k'_o - k_i)\times n}$ be matrices on $\mathbb{F}_q$. Define C_i, C_o and C'_o as the linear codes generated by G, $[G|\Delta G]$ and $[G|\Delta G']$, respectively. C_o and C'_o are called a pair of NLC's with inner code C_i. We denote the outer rates as $r_o = \frac{k_o}{n}$ and $r'_o = \frac{k'_o}{n}$, and the inner rate $r_i = \frac{k_i}{n}$.*

A pair of NQLC's are defined as follows:

Definition 7. *For natural numbers $k_1, k_2, \cdots, k_m$, let $G_{k_i \times n}, i \in [1, m]$ be matrices on $\mathbb{F}_q$, and let $\mathbf{b_j}, j \in \{1, 2\}$ be two dithers on the field. Also, let $(U_1, U_2, \cdots, U_m)$ and $(U'_1, U'_2, \cdots, U'_m)$ be a pair of random vectors on $\mathbb{F}_q$. The pair of QLC's characterized by the matrices $G_{k_i \times n}, i \in [1, m]$, and each of the two vectors of random variables and dithers are called a pair of NQLC's.*

The definition of the NQLC's is a generalization of NLC's. To see this, consider an arbitrary pair of NLC's with the parameters as in Definition 6. These two codes are a pair of NQLC's with parameters $m = 3$, U_1, U_2 and U'_1, U'_3 uniform, U_3 and U'_2 constants and $k_1 = k'_1 = k_i$ and $k_2 = k_o - k_i, k'_3 = k'_o - k_i$. It was shown in [41] that in the general MD problem, it is beneficial to use m-tuples of NLC's called an ensemble of NLC's. The following gives the definition for an ensemble of NLC's:

Definition 8. *A set of l linear codes $C^n_k, k \in [1, l]$ is called an ensemble of nested linear codes with parameter $(r_{\mathsf{J}})_{\mathsf{J} \subset [1,l]}$ if the size of the intersection $\bigcap_{k \in \mathsf{J}} C_k$ is equal to $2^{nr_{\mathsf{J}}}$ for all $\mathsf{J} \subset \mathsf{M}$.*

From the above discussions we can define an ensemble of NQLC's as follows:

Definition 9. *Let $l \in \mathbb{N}$. For natural numbers $k_1, k_2, \cdots, k_m$, let $G_{k_i \times n}, i \in [1, m]$ be matrices on $\mathbb{F}_q$, and $\mathbf{b}_j, j \in [1, l]$ dithers on the field. Also, let $(U_{i,1}, U_{i,2}, \cdots, U_{i,m}), i \in [1, l]$ be vectors of random variables on $\mathbb{F}_q$. The ensemble of QLC's characterized by the matrices $G_{k_i \times n}, i \in [1, m]$ and each of the vectors of random variables and the dithers is called an ensemble of NQLC's.*

Once more it is straightforward to check that this is a generalization of the definition for ensembles of NLC's. Consequently, any achievability results derived using NLC's can be obtained via NQLC's as well. As an example, the next lemma proves that combined with binning, application of these codes can achieve Shannon's RD function for PtP communication.

Lemma 10. *NQLC's achieve Shannon's RD function for PtP source coding for arbitrary source distributions and bounded additive distortion functions.*

Proof. We provide an outline of the proof for an arbitrary source X defined on $\mathbb{F}_q$. Let $p(y|x)$ be an optimizing test channel for Shannon's RD function. Take $m = 1$, and $R_o = \frac{k_1}{n_1} = \log q - H(Y|X)$. Construct a QLC with these parameters. Bin the code randomly and uniformly with rate $\log q - H(Y)$. For each source sequence x^n, the encoder finds a codeword typical with x^n. The encoder transmits the bin index. The decoder finds the unique codeword in the bin which is typical with respect to $p(y)$. It is straightforward to check that with the above rates transmission can be carried out with probability of error going to 0. $\square$

4.4 Fundamental Properties of QLC's

As mentioned in the introduction, the application of NLC's gives gains in different multiterminal source coding problems. These gains are a result of the fact that linear codes are closed under addition. More precisely, the sum of a pair of NLC's has a smaller size than that of two randomly generated unstructured codes of the same rates. As a result, for the two codebooks C_1 and C_2 it takes less rate to transmit $C_1 + C_2$ if the two codes are nested linear codes. However, it has been shown that this closure property has its downsides as well. In fact, a tradeoff has been observed between using NLC's and unstructured codes in different communication setups. The drawback of using NLC's manifests itself in the derivation of the mutual covering bounds for these coding structures. It turns out that unstructured codes satisfy their covering constraint more easily (i.e. their covering bounds are satisfied for lower rates). The idea behind defining QLC's is to breach this gap between NLC's and unstructured codes.

This section is divided into two parts. First, we analyze the addition of QLC's. We show that for the two codebooks C_1 and C_2, the sum $C_1 + C_2$ has a higher rate than the sum of two linear codes of the same rate and a smaller rate than that of two unstructured codes. Then, in the second part, we derive the covering bounds associated with QLC's. In this part, we show that the covering bounds for QLC's are less strict than those for NLC's and more strict than the ones for unstructured codes. Using these two results, we can analyze the tradeoff mentioned above for the application of QLC's.

4.4.1 The Addition of QLC's

QLC's are not linearly closed but at the same time maintain a degree of "closedness" in their structure. Notice that if we repeatedly add a QLC with itself, the resulting set of codevectors will be a subset of the linear code generated by $[G_1|G_2|\cdots|G_m]$, where the G_i's are the generator matrices for the QLC. Whereas, if a random unstructured code is added with itself repeatedly, the resulting space would converge to the whole vector space. In the following lemma we investigate the addition of l copies of a QLC with each other:

Lemma 11. *For $R \in (0, \log q)$, let C_Q be a QLC with parameters $m, n, k_1, k_2, ..., k_m \in \mathbb{N}$, $U_i, i \in [1, m]$, matrices $G_i, i \in [1, m]$, and dither $\mathbf{b}$, such that the code has rate R, where the G_i's and $\mathbf{b}$ are generated randomly and uniformly on $\mathbb{F}_q$. The probability of the following events goes to one as $n \to \infty$:*

1. $\frac{1}{n} \log |\sum_{i\in[1,l]} C_Q| \doteq \sum_{i\in[1,m]} \frac{k_i}{n} H(U_i^{\otimes l})$,

2. $R \leq \frac{1}{n} \log |\sum_{i\in[1,l]} C_Q| \leq \min(\log q, l \times R)$ *where equality is achieved on the left hand side by taking U_i's to be uniform.*

Proof. The proof follows standard typicality arguments and is omitted. $\square$

Remark 12. *As mentioned in the lemma, equality on the left hand side of condition 2 can be achieved by taking the U_i's to be uniform. In this case the QLC becomes a coset code. On the right hand side, one can approach equality by taking $k_i > n$ and U_i to be very low entropy random variables. Observe that if the random variable U_i has low entropy, then $H(U_i^{\otimes l}) \approx lH(U_i)$.*

Remark 13. *For an arbitrary n-length codebook C with rate R, it is straightforward to show that $R \leq \frac{1}{n}\log|\sum_{i\in[1,l]} \mathsf{C}| \leq \min(\log q, l \times R)$. For linear codes equality always holds on the left-hand side. For random codes, equality always holds on the right-hand side. Whereas, QLC's achieve all of the possible values allowed for $\frac{1}{n}\log|\sum_{i\in[1,l]} \mathsf{C}|$.*

4.4.2 Mutual Covering Bounds for NQLC's

We proceed with deriving the mutual covering bounds for the NQLC's. The covering bounds are useful for determining inner bounds to achievable RD regions in different source coding settings. In this work, we concentrate on the MD problem. The following gives a formal definition for a $P_{XV_1V_2}$-covering pair of codes.

Definition 14. *Let $\mathbb{F}_q$ be a field. Consider 3 random variables X, V_1 and V_2, where X is defined on an arbitrary finite set X and V_1 and V_2 are defined on $\mathbb{F}_q$. Fix a PMF P_{X,V_1,V_2} on $\mathsf{X} \times \mathbb{F}_q \times \mathbb{F}_q$. A sequence of code pairs (C_1^n, C_2^n) is called $P_{XV_1V_2}$-covering if:*

$$\forall \epsilon > 0, P(\{x^n | \exists (v_1^n, v_2^n) \in A_\epsilon^n(V_1, V_2|x^n) \cap C_1 \times C_2\}) \to 1,$$

as $n \to \infty$.

As mentioned in chapter 2, deriving the achievable RD region for the MD setup using the SSC scheme involves obtaining the mutual covering bounds for independently generated codebooks. The following lemma characterizes these bounds for a pair of unstructured codes.

Lemma 15. *[14] For any distribution $P_{XV_1V_2}$ on $\mathsf{X} \times \mathbb{F}_q \times \mathbb{F}_q$ and rates r_1, r_2 satisfying (4.1)-(4.3), there exists a sequence of pairs of unstructured codes C_1^n and C_2^n which are $P_{XV_1V_2}$-covering.*

$$r_1 \geq H(V_1) - H(V_1|X) \tag{4.1}$$

$$r_2 \geq H(V_2) - H(V_2|X) \tag{4.2}$$

$$r_1 + r_2 \geq H(V_1, V_2) - H(V_1, V_2|X) \tag{4.3}$$

When using ensembles of NLC's, new covering bounds are necessary since the codebooks are not independently generated (e.g. they share a common inner code.). The next lemma presents the bounds for a pair of NLC's.

Lemma 16. *[41] For any $P_{XV_1V_2}$ on $\mathsf{X} \times \mathbb{F}_q \times \mathbb{F}_q$ and rates $r_o = r_1$, $r'_o = r_2$ and r_i satisfying 4.4-4.7, there exists a sequence of pairs of NLC's C_1^n and C_2^n which are $P_{XV_1V_2}$-covering.*

$$r_1 \geq \log q - H(V_1|X) \tag{4.4}$$

$$r_2 \geq \log q - H(V_2|X) \tag{4.5}$$

$$r_1 + r_2 \geq 2\log q - H(V_1, V_2|X) \tag{4.6}$$

$$r_1 + r_2 - r_i \geq \max_{\alpha,\beta \in \mathbb{F}_q \setminus \{0\}} (\log q - H(\alpha V_1 + \beta V_2|X)), \tag{4.7}$$

In the process of deriving the inner bound to the achievable RD region, the entropy terms in Lemma 15 and $\log q$ terms in Lemma 16 vanish in the Fourier-Motzkin elimination and only the conditional entropy terms would remain on the RHS [41]. So, the only consequential difference between the two bounds lies in the introduction of inequality (4.7). First, we argue that this inequality can not be eliminated by a more precise error analysis. We use a converse coding argument to prove this point. Assume the existence of a pair of NLC's C_1 and C_2 which are $P_{XV_1V_2}$-covering. Then, for any typical sequence $\mathbf{x}^n$, one can find sequences $\mathbf{c}_i^n \in C_i, i \in \{1, 2\}$ which are typical with $\mathbf{x}^n$ with respect to $P_{XV_1V_2}$. From the Markov Lemma [45], $\mathbf{x}^n$ is typical with $\alpha\mathbf{c}_1^n + \beta\mathbf{c}_2^n$ with respect to $P_{X(\alpha V_1+\beta V_2)}$ since $\alpha V_1 + \beta V_2 \leftrightarrow V_1, V_2 \leftrightarrow X$. So, by the converse source coding theorem, $\frac{1}{n}\log|\alpha C_1 + \beta C_2| \geq \log q - H(\alpha V_1 + \beta V_2)$ which gives (4.7). The following lemma characterizes the covering bounds for a pair of NQLC's.

Lemma 17. *For any $P_{XV_1V_2}$ on $\mathsf{X} \times \mathbb{F}_q \times \mathbb{F}_q$, parameters $m, n, k_1, k_2, \cdots, k_m$ and random vectors $(U_{1,i})_{i\in[1,m]}, (U_{2,i})_{i\in[1,m]}$ satisfying (4.8)-(4.11), there exists a sequence of pairs of NQLC's C_1^n and C_2^n which are $P_{XV_1V_2}$-covering.*

$$\sum_{i\in[1,m]} \frac{k_i}{n} H(U_{1,i}) \geq \log q - H(V_1|X) \tag{4.8}$$

$$\sum_{i\in[1,m]} \frac{k_i}{n} H(U_{2,i}) \geq \log q - H(V_2|X) \tag{4.9}$$

$$\sum_{i\in[1,m]} \frac{k_i}{n} (H(U_{1,i}) + H(U_{2,i})) \geq 2\log q - H(V_1, V_2|X) \tag{4.10}$$

$$\sum_{i\in[1,m]} \frac{k_i}{n} H(\alpha U_{1,i} + \beta U_{2,i}) \geq \log q - H(\alpha V_1 + \beta V_2|X), \forall \alpha, \beta \in \mathbb{F}_q\backslash\{0\}. \tag{4.11}$$

Proof. Let X be a discrete memoryless source, for typical sequence x with respect to P_X, define the following:

$$\begin{aligned}\theta(x) &= \sum_{u\in C_1, v\in C_2} \mathbb{1}\{(\mathbf{v}_1, \mathbf{v}_2) \in A_\epsilon^n(V_1, V_2|\mathbf{x})\} \\ &= \sum_{\substack{\mathbf{u}_{1,i}, \mathbf{u}_{2,i} \in \mathbb{Z}_q^{k_i}, \\ i\in[1,m]}} \sum_{(\mathbf{v}_1, \mathbf{v}_2)\in A_\epsilon^n(V_1, V_2|\mathbf{x})} \mathbb{1}\{\sum_{i\in[1,m]} \mathbf{u}_{1,i} G_i + \mathbf{b}_1 = \mathbf{v}_1, \sum_{i\in[1,m]} \mathbf{u}_{2,i} G_i + \mathbf{b}_2 = \mathbf{v}_2\}\end{aligned}$$

Here, $\mathbf{G}_i, \mathbf{b}_1$, and $\mathbf{b}_2$ are chosen randomly and uniformly. For $\mathbf{u}_{1,i} \in \mathbb{Z}_q^{k_i}$, define $g(\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \cdots, \mathbf{u}_{1,m}) \triangleq \sum_{i\in[1,m]} \mathbf{u}_{1,i} G_i + \mathbf{b}_1$. Similarly define $g(\mathbf{u}_{2,1}, \mathbf{u}_{2,2}, \cdots, \mathbf{u}_{2,m}) \triangleq \sum_{i\in[1,m]} \mathbf{u}_{2,i} G_i + \mathbf{b}_2$ for $\mathbf{u}_{2,i} \in \mathbb{Z}_q^{k_i}$.

Lemma 18. *The following hold:*

1. *$g(\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \cdots, \mathbf{u}_{1,m})$ and $g'(\mathbf{u}_{2,1}, \mathbf{u}_{2,2}, \cdots, \mathbf{u}_{2,m})$ are uniform over $\mathbb{F}_q^n$.*

2. *$g(\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \cdots, \mathbf{u}_{1,m})$ is independent of $g(\tilde{\mathbf{u}}_{1,1}, \tilde{\mathbf{u}}_{1,2}, \cdots, \tilde{\mathbf{u}}_{1,m})$ when $(\mathbf{u}_{1,i})_{i\in[1,m]} \neq (\tilde{\mathbf{u}}_{1,i})_{i\in[1,m]}$.*

3. *$g'(\mathbf{u}_{2,1}, \mathbf{u}_{2,2}, \cdots, \mathbf{u}_{2,m})$ is independent of $g'(\tilde{\mathbf{u}}_{2,1}, \tilde{\mathbf{u}}_{2,2}, \cdots, \tilde{\mathbf{u}}_{2,m})$ when $(\mathbf{u}_{2,i})_{i\in[1,m]} \neq (\tilde{\mathbf{u}}_{2,i})_{i\in[1,m]}$.*

4. *If $\mathbf{b}_1$ and $\mathbf{b}_2$ independent uniform over $\mathbb{F}_q^n$, then $g(\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \cdots, \mathbf{u}_{1,m})$ and $g'(\mathbf{u}_{2,1}, \mathbf{u}_{2,2}, \cdots, \mathbf{u}_{2,m})$ are independent.*

Proof. Similar to the proof of the covering lemma in [42]. □

We want to use the Chebyshev's inequality to obtain:

$$P\{\theta(\mathbf{x}) = 0\} \leq \frac{4var\{\theta(\mathbf{x})\}}{\mathbb{E}\{\theta(\mathbf{x})\}^2} \to 0$$

We calculate the expected value of $\theta(\mathbf{x})$:

$$\begin{aligned}\mathbb{E}\{\theta(\mathbf{x})\} &= \sum_{\mathbf{x}\in A_\epsilon^n(X)} \sum_{\substack{\mathbf{u_{1,i}}\neq\mathbf{u}_{2,i},\\ i\in[1,m]}} \sum_{(\mathbf{v}_1,\mathbf{v}_2)\in A_\epsilon^n(V_1,V_2|\mathbf{x})} P(\mathbf{x})P\{g(\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \cdots, \mathbf{u}_{1,m}) = \mathbf{v}_1, g'(\mathbf{u}_{2,1}, \mathbf{u}_{2,2}, \cdots, \mathbf{u}_{2,m}) = \mathbf{v}_2\} \\ &= \sum_{\mathbf{x}\in A(X)} \sum_{\mathbf{u}_{1,i}\neq\mathbf{u}_{2,i}} |A_\epsilon^n(V_1, V_2|\mathbf{x})| P(\mathbf{x}) \frac{1}{q^{2n}} \\ &= 2^{n(-\sum_{i\in[1,m]} \frac{k_i}{n}\left(H(U_{1,i})+H(U_{2,i})\right)+H(V_1,V_2|X)+O(\epsilon))} \end{aligned} \tag{4.12}$$

The following lemma bounds $\frac{var\{\theta(\mathbf{x})\}}{\mathbb{E}\{\theta(\mathbf{x})\}^2}$.

Lemma 19.

$$\begin{aligned}&\frac{var\{\theta(\mathbf{x})\}}{\mathbb{E}\{\theta(\mathbf{x})\}^2} \\ &\leq 2^{-n(-\sum_{i\in[1,m]} \frac{k_i}{n}\left(H(U_{1,i})+H(U_{2,i})\right)+H(V_1,V_2|X))} + 2^{-n(-\sum_{i\in[1,m]} \frac{k_i}{n} H(U_{1,i})+H(V_1|X))} + 2^{-n(-\sum_{i\in[1,m]} \frac{k_i}{n} H(U_{2,i})+H(V_2|X))} \\ &+ \sum_{\alpha\in\mathbb{F}_q\setminus\{0\}} 2^{n(-\sum_{i\in[1,m]} \frac{k_i}{n} H(U_{1,i}+\alpha U_{2,i})+H(V_1,V_2|X)-H(V_1,V_2|X,V_1+\alpha V_2))}\end{aligned}$$

The proof of the lemma follows from arguments similar to the ones used in [42]. Setting the above to go to 0, we get the covering bounds mentioned in Lemma 17

□

Remark 20. *Inequalities (4.8)-(4.10) are exactly the same bounds on the codebook rates as in (4.4)-(4.6) (note that* $\sum_{i\in[1,m]} \frac{k_i}{n} H(U_{j,i})$ *is the rate of* $C_j, j \in \{1,2\}$*.). (4.11) can also be written as* $\frac{1}{n}\log|\alpha C_1 + \beta C_2| \geq \log q - H(\alpha V_1 + \beta V_2)$*. By the same argument as in the previous lemma, the bounds can not be tightened by a finer error analysis. The main difference between inequality* (4.11) *and (4.7) is that in the new bound, the LHS changes as a function of* α *and* β*. This provides new degrees of freedom which in turn result in improvements in the MD problem as shown in the next section.*

4.5 Gains in the MD Problem

In this section, we first present an example in which a scheme based on NQLC's gives improvements in terms of achievable RD's compared to the SSC scheme. The example is constructed by slightly altering example 6 in [41]. The setup is depicted in Figure 4.2. Here, X is a binary symmetric source. The distortion constraints at all decoders are binary Hamming distortions except for decoder {3}. Assume that the distortion constraint at decoder {3} is such that it needs to reconstruct the ternary addition $\hat{X}_1 \oplus_3 2\hat{X}_2$, where $\hat{X}_i, i \in \{1,2\}$ are the reconstructions at decoders {1} and {2}. We are interested in achieving

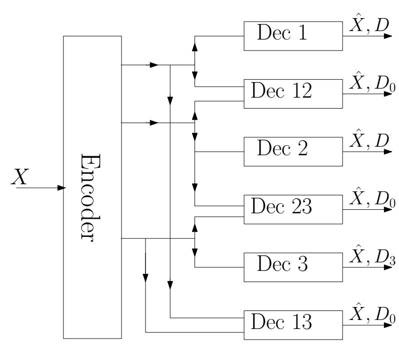

Figure 4.2: A three-descriptions example where NQLC's give gains

the RD vectors with the following projections:

$$R_1 = R_2 = \frac{1 - h_b(D_0)}{2}, D_{\{1\}} = D_{\{2\}} = \frac{1}{2}(1 - (1 - 2D_0)(2 - \sqrt{2})), \tag{4.13}$$

$$D_{\{1,2\}} = D_{\{1,3\}} = D_{\{2,3\}} = D_0, \tag{4.14}$$

Our objective is to minimize R_3 subject to these constraints. The following lemma gives the RD vector achievable using NQLC's which is not present in the RD region in [41].

Lemma 21. *There exists $\epsilon > 0$, such that the RD vector in (4.14) is achievable using NQLC's, if the following hold:*

$$R_3 \geq H(V_1 \oplus_3 2V_2) - H(V_1 \oplus_3 V_2|X) - \epsilon \tag{4.15}$$

$$h_b(D_0) + 2h_b\left(\frac{\sqrt{2}}{2}\right) + h_b\left(2\left(\sqrt{2} - 1\right)D_0\right) + h_b\left(2\left(\sqrt{2} - 1\right)(1 - D_0)\right) = 1, \tag{4.16}$$

where the joint distribution between X, V_1 and V_2 is given in table I.

X \ V_1, V_2	00	01	10	11
0	$\frac{1}{2}(1 - D_0)$	$\frac{\sqrt{2}-1}{2} D_0$	$\frac{\sqrt{2}-1}{2} D_0$	$\frac{3-2\sqrt{2}}{2} D_0$
1	$\frac{1}{2} D_0$	$\frac{\sqrt{2}-1}{2}(1 - D_0)$	$\frac{\sqrt{2}-1}{2}(1 - D_0)$	$\frac{3-2\sqrt{2}}{2}(1-D_0)$

Table 4.1: The joint distribution P_{X,V_1,V_2}.

Furthermore, the RD vector is not achievable using the linear coding scheme stated in [41].

Proof. We provide a scheme which achieves the RD vector for $\epsilon = 10^{-4}$ using NQLC's. Let n be large and λ a small positive number. construct a pair of $P_{XV_1V_2}$-covering NQLC's with parameters $m = 2$, $\frac{k_1}{n} = 0.8$, and $\frac{k_2}{n} = 0.2665$ where U_1 and U_1' are ternary and uniform. U_2 and U_2' have the following distributions,

	0	1	2
U_1	0.33	0.48	0.19

	0	1	2
U_2	0.33	0.19	0.48

Given the above parameters, it is straightforward to check that the constraints in Lemma 17 are satisfied. Description 1 carries the bin index of C with bin size $\log 3 - \frac{H(V_1,V_2)}{2} - \lambda$, also, description 2 carries the index for C' with the same bin size. Description 3 carries the index for $C \oplus_3 2C'$ with bin size $\log 3 - H(V_1 \oplus 2V_2) - \lambda$. Then,

$$R_1 = R_2 = \frac{k_1}{n} + \frac{k_2}{n} H(U_1) - (\log 3 - \frac{H(V_1, V_2)}{2} - \lambda)$$
$$R_3 = \frac{k_1}{n} + \frac{k_2}{n} H(U_1 \oplus_3 2U_2) - (\log 3 - H(V_1 \oplus 2V_2) - \lambda).$$

Direct calculation shows that the above rates are equal to the ones stated in the lemma. We provide an outline of the proof that the scheme in [41] can not achieve these rates. By the same arguments as in the proof of Example 6 in [41], it can be shown that the only non-redundant codebooks in the scheme are $C_{\{1\}}$, $C_{\{2\}}$, and $C_{o,\{3\}}$ (this follows from optimality at decoders $\{1, 2\}$ and $\{3\}$, and the uniqueness of the optimizing distribution at decoder $\{1, 2\}$ shown in [41]). Then, in order to satisfy the constraints at decoder $\{3\}$, we need to set $V_{o,\{3\}} = V_1 + 2V_2$. Checking the bounds in [41] it can be seen that the above rates are not achievable. . $\square$

Next, we provide a new achievable RD region for the general MD problem using NQLC's. For brevity, we have only considered the case where a summation of two codebooks decoded at decoders $\{1\}$ and $\{2\}$ is to be decoded at decoder $\{3\}$. So, to achieve the RD region we use all of the codebooks present in the SSC scheme with the addition of a pair of NQLC's. One of the NQLC's is decoded at $\{1\}$, the other at $\{2\}$, and a linear combination of the two is decoded at $\{3\}$ as was the case in the previous example. This RD region could be improved upon by considering the reconstruction of an arbitrary number of summations of arbitrary lengths at the decoders as done for the NLC's in [41]. The notation used in the next definition is the same as in [41].

Definition 22. *Fix the prime number q. For a joint distribution P on random variables $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, $V_{\{j\}}, j \in \{1,2\}$, and X, where the underlying alphabet for all auxiliary random variables is the field $\mathbb{F}_q$, and a set of reconstruction functions $g_{\mathcal{L}} = \{g_{\mathsf{N}} : \mathsf{U}_{\mathsf{N}} \to \mathsf{X}, \mathsf{N} \in \mathcal{L}\}$, the set $\mathcal{RD}(P, g_{\mathcal{L}})$ is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers $(\rho_{\mathcal{M},i}, r_{o,\mathcal{M}})_{i\in\widetilde{\mathcal{M}},\mathcal{M}\in\mathbf{S}_{\mathsf{L}}}$, $\rho_{o,\{j\},i}, i \in \{1,2,3\}$, and parameters $(m, n, k_1, k_2, \cdots, k_m)$ and vectors of random variables $(A_{i,j})_{j\in[1,m]}, i \in \{1,2\}$:*

$$H(U_{\mathbf{M}}V_{\mathbf{E}}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{E}\in\mathbf{E}} (\log q - r_{o,\mathcal{E}}), \tag{4.17}$$

$$H(U_{\mathbf{M}}, W_{3,\alpha,\beta}|X) \geq \sum_{\mathcal{M}\in\mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \log q - r_{o,3,\alpha,\beta} \tag{4.18}$$

$$H([U,V,W]_{\overline{\mathbf{M}}_{\mathsf{N}}}|[U,V,W]_{\overline{\mathbf{M}}_{\mathsf{N}}\cup\overline{\mathbf{L}}}) \leq \sum_{\mathcal{M}\in\overline{\mathbf{M}}_{\mathsf{N}}\backslash\overline{\mathbf{L}}} (\log q + \sum_{j\in\widetilde{\mathcal{M}}} \rho_{\mathcal{M},j} - r_{o,\mathcal{M}}) + \sum_{\substack{\mathcal{M}\in\{\mathcal{A}_1,\mathcal{A}_2,\mathcal{A}_3\}\bigcap\overline{\mathbf{M}}_{\mathsf{N}},\\ j\in\widetilde{\mathcal{M}}}} \rho_{o,\mathcal{M},j}, \tag{4.19}$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \qquad D_{\mathsf{N}} = E\{d_{\mathsf{N}}(h_{\mathsf{N}}(U_{\mathsf{N}}, X))\}. \tag{4.20}$$

Where (a) $\overline{\mathbf{M}}_{\mathsf{N}} \triangleq (\mathbf{M}_{\mathsf{N}}, \{\{j\}|\{j\} \in \mathbf{M}_{\mathsf{N}}\}, \{\{\{3\},1,1\}|\{3\} \in \mathbf{M}_{\mathsf{N}}\})$, (b) $\widetilde{\mathbf{M}}_{\mathsf{N}} \triangleq \bigcup_{\mathsf{N}'\subsetneq\mathsf{N}} \overline{\mathbf{M}}_{\mathsf{N}'}$, (c) $r_{o,\{3\},\alpha,\beta} \triangleq \sum_{j\in[1,m]} \frac{k_i}{n} H(\alpha V_{1,j} + \beta V_{2,j})$, (d) $r_{o,\mathcal{M}} \leq \log q$, and (e) $W_{\{3\},\alpha,\beta} \triangleq \alpha V_{\{1\}} + \beta V_{\{2\}}$, and the bounds should hold for all $\mathbf{M} \subset \mathbf{S}_{\mathsf{L}}, \mathbf{E} \subset \{\{1\},\{2\}\}$ and $\overline{\mathbf{L}} \subset \overline{\mathbf{M}}_{\mathsf{N}}$.

The main difference between this scheme and the one in [41] is that the rate $r_{o,\{3\},\alpha,\beta}$ is now defined according to the size of the linear combination of NQLC's rather than NLC's.

Theorem IV.23. *RD vectors in $cl\,(\mathcal{RD}(P, g_{\mathcal{L}}))$ are achievable. Where $cl(\mathsf{A})$ is the closure of set A.*

Proof. Given a joint distribution $P_{\mathbf{U},\mathbf{V},X}$, and codebook and binning rates satisfying the bounds in the theorem we prove achievability of the RD vector in (4.20).

Codebook Generation: Fix blocklength n. For every $\mathcal{M} \in \mathbf{S}_{\mathsf{L}}$, independently generate a linear code $C_{\mathcal{M}}$ with size $2^{nr_{o,\mathcal{M}}}$. Also generate a pair of NQLC's $C_{\{j\}}, j \in \{1,2\}$ with parameters as in Definition 7 and random variables $(V_{1,j})$ and $(V_{2,j})$, $j \in [1,m]$, respectively

. Define the set of codewords $C_{o,\{3\},\alpha,\beta} \triangleq \alpha C_{o,\{1\}} + \beta C_{o,\{2\}}$. The size of $C_{o,\{3\},\alpha,\beta}$ is $2^{nr_{o,\{3\},\alpha,\beta}}$ where $r_{o,\{3\},\alpha,\beta} = \sum_{j\in[1,m]} \frac{k_i}{n} H(\alpha V_{1,j} + \beta V_{2,j})$. For the ith description bin the codebook $C_{\mathcal{M}}$ randomly and uniformly with rate $2^{n\rho_{\mathcal{M},i}}$.

Encoding: Upon receiving the source vector X^n, the encoder finds a jointly-typical set of codewords $c_{\mathcal{M}}$. Each description carries the *bin-indices* of all of the corresponding codewords. The encoder declares an error if there is no jointly typical set of codewords available.

Decoding: Having received the bin-indices from descriptions $i \in \mathsf{N}$, decoder N tries to find a set of jointly typical codewords $c_{\mathcal{M}}, \mathcal{M} \in \overline{\mathbf{M}}_{\mathsf{N}}$. If the set of codewords is not unique, the decoder declares error.

In order for the encoder to find a set of jointly typical codewords, the mutual covering bounds (4.17) and (4.18) should hold. This is a generalization of the result in lemma 17 and we omit the proof for brevity. The bounds in (4.19) are the mutual packing bounds at each decoder.

□

4.6 The Interference Channel

We proceed with formally defining the three user IC problem. A three user IC consists of three input alphabets $\mathsf{X}_i, i \in \{1,2,3\}$, three output alphabets $\mathsf{Y}_i, i \in \{1,2,3\}$, and a transition probability matrix $P_{\mathbf{Y}|\mathbf{X}}$. A code for this setup is defined as follows.

Definition 24. *A three user IC code* $(n, \mathsf{M}_1, \mathsf{M}_2, \mathsf{M}_3, \mathbf{e}, \mathbf{d})$ *consists of (1) Three sets of message indices* M_i *(2) Three encoder mappings* $e_i : \mathsf{M}_i \to \mathsf{X}_i^n, i \in [1,3]$*, without loss of generality, these maps are assumed to be injective (3) and three decoding functions* $d_i : \mathsf{Y}_i^n \to \mathsf{M}_i, i \in [1,3]$*. We define the codebook corresponding to the encoding map* e_i *as* $\mathbb{C}_i = \{e_i(m_i)|m_i \in \mathsf{M}_i\}, i \in [1,3]$*. The rate of user i is defined as* $r_i = \frac{1}{n}\log|\mathbb{C}_i|$.

Definition 25. *A rate-triple* (R_1, R_2, R_3) *is said to be achievable if for every* $\epsilon > 0$*, there* *exists a code* $(n, \mathsf{M}_1, \mathsf{M}_2, \mathsf{M}_3, \mathbf{e}, \mathbf{d})$ *such that* (1) $r_i \geq R_i - \epsilon, i \in [1,3]$*, and (2)* $P(d(\mathbf{Y}^n) = \mathbf{M}|\mathbf{e}(\mathbf{M}) = \mathbf{X}^n) \geq 1 - \epsilon$.

4.7 The Interference Alignment Tradeoff

In this section, we investigate the interference alignment tradeoff mentioned in the introduction in more detail. We show that in certain three user interference setups, on the one hand, alignment is beneficial to one of the users, while on the other hand, the rates achieved by the aligning users is reduced due to the alignment. We investigate the phenomenon in two examples. The first example involves a three user interference setup. In this example, the first two encoders use linear codes to manage the interference for the third user. This gives a strictly improved achievable rate region. It is well-known that interference alignment can be induced efficiently by the application of structured codes. Additional to this, we show the stronger statement that the only ensemble of codes which achieve the desired rate-triples in this example, are the ones with specific linearity properties. Next, we build upon the first example to create a setup where alignment is beneficial to one of the users and harmful for the other one. This second example provides the motivation for our new codebook constructions in the next section.

4.7.1 Example 1

Consider the example shown in Figure 4.3. All of the inputs are q-ary and the additions are defined on the field $\mathbb{F}_q$. The three outputs of the channel are $Y_i = X_i \oplus N_1 \oplus N_3, i \in \{1,2\}$, $Y_3 = X_1 \oplus X_2 \oplus X_3 \oplus N_3$. We are interested in achieving the following rates for the first and second users:

$$R_1 = R_2 = \log q - H(N_1 \oplus N_3).$$

Given these rates, we want to maximize R_3. The following lemma shows that linear codes achieve the optimum R_3 for this setup. Furthermore, we show that if an ensemble of codes

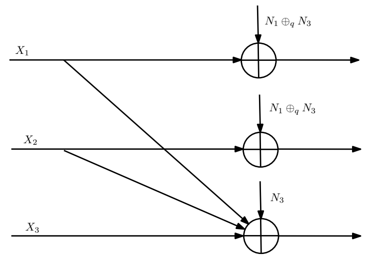

Figure 4.3: A Three User IC Where Alignment Is Strictly Beneficial

achieves the optimum R_3, then the codes corresponding to the first two users are "almost" the same coset code.

Lemma 26. *For a given family of codes* $(n, \mathsf{M}_1, \mathsf{M}_2, \mathsf{M}_3, \mathbf{e}, \mathbf{d}), n \in \mathbb{N}$ *satisfying the rate and error constraints at decoders 1 and 2, user* 3 *can achieve the rate* $R_3 = H(N_1 \oplus N_3) - H(N_3)$ *iff there exists a dither* $\mathbf{b}$ *such that for every random variable N defined on* $\mathbb{F}_q$ *with positive entropy, the following holds:*

$$P\Big(e_1(M_1) \oplus e_2(M_2) \in (\mathbb{C}_1 \bigcup \mathbb{C}_2) \oplus A_\epsilon^n(N) \oplus \mathbf{b}\Big) \to 1, \text{ as } n \to \infty. \tag{4.21}$$

Equivalently, the optimal rate is achieved iff there exists another family of codes $(n, \mathsf{M}'_1, \mathsf{M}'_2, \mathsf{M}'_3, \mathbf{e}', \mathbf{d}'), n \in \mathbb{N}$ *for which 1)* $P(\mathbf{e}'_i(M'_i) \in \mathbb{C}_i \oplus A_\epsilon^n(N)) \to 1$ *as* $n \to \infty$*, 2)* $\mathbb{C}'_1 = \mathbb{C}'_2$ *is a coset code, and 3) they also achieve the rate triple* $(R_1, R_2, R_3) = (\log q - H(N_1 \oplus N_3), \log q - H(N_1 \oplus N_3), H(N_1 \oplus N_3) - H(N_3))$.

Proof. We provide an outline of the proof in Appendix C.1.1. □

The lemma proves that even if we expand our search to arbitrary n-length codebook constructions (as opposed to the usual random codebook generation based on single-letter distributions), coset codes are the only efficient ensemble of codes for the classes of interference channels under consideration up to small perturbations. This is a stronger assertion than the well-known result that linear codes are useful for aligning the interfering signals. The lemma can be used to provide a converse result proving that schemes involving random unstructured codes (e.g. the generalized version of the single-letter HK scheme), can't achieve the desired rate-triple without directly analyzing the bounds corresponding to their achievable rate region as done in [29].

4.7.2 Example 2

Next, we consider an example where interference alignment results in a tradeoff between two of the users. Consider the setup in Figure 4.4. Similar to the previous example, all input alphabets, output alphabets, and additions are defined on the field $\mathbb{F}_q$. The outputs of the channel are $Y_1 = X_1 \oplus_q N_1 \oplus_q N_2 \oplus_q N_3$, $Y_2 = X_1 \oplus_q X_2 \oplus_q N_2 \oplus N_3$, and $Y_3 = 2X_1 \oplus_q X_2 \oplus_q X_3 \oplus_q N_3$. Following our arguments in the previous example, for user 3 to be able to transmit its messages at rate $R_3 = H(X_3 \oplus N_3) - H(N_3)$, the inputs for users 1 and 2 must align. However, if these two users align their inputs, user 2 would not be able to decode its message which is being corrupted by its aligned interfering signal coming from user 1. Hence, we have a tradeoff. We proceed with evaluating the rate-triples achievable in this example. The following lemma proves that we must have $R_1 + R_2 \leq \log q - H(N_2 \oplus N_3)$, otherwise the rate-triple (R_1, R_2, R_3) is not achievable.

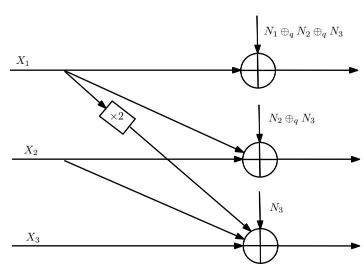

Figure 4.4: A Three User IC Where Alignment Results in a Tradeoff

Lemma 27. *Given that* (R_1, R_2, R_3) *is achievable, we must have* $R_1 + R_2 \leq log q - H(N_2 \oplus N_3)$.

Proof. Since the rate triple is achievable, there exists a family of codes $(n, \mathsf{M}_1, \mathsf{M}_2, \mathsf{M}_3, \mathbf{e}, \mathbf{d})$ for which the codewords sent by the second user, X_2^n, is decoded at decoder 2 with error probability approaching 0. Assuming errorless decoding at decoder 2, the decoder has access to $X_2^n, X_1^n \oplus X_2^n \oplus N_2^n$. The decoder can subtract X_2^n from $X_1^n \oplus X_2^n \oplus N_2^n$ to get $X_1^n \oplus N_2^n$.

Now, since by assumption decoder 1 can decode X_1^n from $X_1^n \oplus N_1^n \oplus N_2^n$ with error going to 0, decoder 2 can use $X_1^n \oplus N_2^n$ to recover X_1^n with small error. So decoder 2 has access to M_1 and M_2. By the converse of the point-to-point channel coding theorem, we must have $\frac{1}{n} \log |\mathsf{M}_1 \times \mathsf{M}_2| \leq \log q - H(N_2)$, which completes the proof. $\square$

We want to achieve the rate $R_1 = \log q - H(N_1 \oplus N_2 \oplus N_3)$. In other words, encoder 1 is to operate at PtP optimality. The goal is to optimize the linear combination $R_2 + R_3$. We argue that the linear coding scheme presented in [29] can't achieve the triple (R_1, R_2, R_3) for $R_2 + R_3 > H(N_1 \oplus N_2 \oplus N_3) - H(N_3)$.

Lemma 28. *Given $R_1 = \log q - H(N_1)$, the scheme in [29] can't achieve $R_2 + R_3 > H(N_1 \oplus N_2 \oplus N_3) - H(N_3)$.*

Proof. We provide the intuition behind the proof here. Let us use two NCL's C_1 and C_2 as defined in Definition 6 to transmit the messages at encoders 1 and 2. Let the rate of $C_j, j \in \{1, 2\}$ be r_j and let the inner code have rate r_i. If we assume that the coding scheme exists which achieves the rate-triple, then by the proof of Lemma 27, one should be able to recover X_1, X_2 from $X_1 \oplus X_2 \oplus N_2 \oplus N_3$ with small error probability. Also, at decoder 3, the decoder can reconstruct X_3 with low error probability and by subtraction it can have $2X_1 \oplus X_2 \oplus N_3$. Note that in the linear coding scheme, both $2X_1 \oplus X_2$ and $X_1 \oplus X_2$ come from randomly and uniformly generated linear codes of rate $r_1 + r_2 - r_i$. So, given that X_1 and X_2 can be recovered from $X_1 \oplus X_2 \oplus N_2 \oplus N_3$, decoder 3 must be able to recover X_1 and X_2 from $2X_1 \oplus X_2 \oplus N_3$. Then similar to the proof of Lemma 26, by the point-to-point channel coding converse, we must have $R_1 + R_2 + R_3 < \log q - H(N_3)$. $\square$

The arguments in the proof of the previous lemma suggest that NLC's lack the necessary flexibility when it comes to determining the size of different linear combinations of such codes. We explain this in more detail. Consider two NLC's, C and C', with rates r_o and r'_o, respectively, and with inner code rate r_i. The rate of any linear combination of the two, $\alpha C \oplus \beta C', \alpha, \beta \in \mathbb{F}_q \backslash \{0\}$, is equal to $r_o + r'_o - r_i$. Whereas in settings such as the one at hand, it is desirable to have different rates for different values of α and β. In this setup, decoder 2 requires $C_1 \oplus C_2$ to be large (since by Lemma 27 in order to increase R_2 it needs to increase the rate of this linear combination) and decoder 3 wants the size of the interfering codebook $2C_1 \oplus C_2$ to be small, so that it can decode the interference. In the next section, we provide a new class of codes. The new construction allows for different rates for different linear combinations of such codes. This in turn results in higher achievable sum-rates.

Lemma 29. *There exists achievable rate-triples* $(R_1, R_2, R_3) = (\log q - H(N_1 \oplus N_2 \oplus N_3), r_2, r_3)$ *such that* $r_2 + r_3 > H(N_1 \oplus N_2 \oplus N_3) - H(N_3)$.

Proof. Refer to Appendix C.1.2.

□

So far we have proved that NQLC's outperform NLC's in this specific example. It is straightforward to show that NQLC's are a generalization of NLC's. To see this, consider an arbitrary pair of NLC's with the parameters as in definition 6. These two codes are a pair of NQLC's with parameters $m = 3$, U_1, U_2 and U'_1, U'_3 uniform, U_3 and U'_2 constants and $k_1 = k'_1 = k_i$ and $k_2 = k_o - k_i, k'_3 = k'_o - k_i$. So, any rate region achievable by NLC's is also achievable using NQLC's.

4.8 New Achievable Rate Region for the IC

In this section, we provide a general achievable rate region for the three user IC. The scheme is similar to the one presented in [29] (Theorem 2). The main difference is that here instead of NLC's we use NQLC's. The random variables involved in the coding scheme are depicted in Figure 4.5. Note that in contrast with the scheme in [29], decoder 2 reconstructs a linear combination of U_1 and U_2. By setting $\alpha_2 = 0, \beta_2 = 1$, we recover the random variables in [29]. The next theorem provides the achievable rate region.

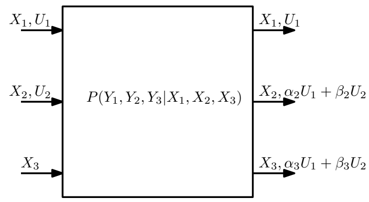

Figure 4.5: The LHS random variables are the ones sent by each encoder, the RHS random variables are the ones decoded at each decoder.

Definition 30. *For a given three user IC problem with q-ary inputs and outputs, define the set $\mathcal{R}_{3\text{-}IC}$ as the set of rate triples (R_1, R_2, R_3) such that there exist 1) a joint probability distribution $P_{U_1,X_1}P_{U_2,X_3}P_{X_3}$, 2) A vector of positive reals $(K_1, K_2, L_1, L_2, T_1, T_2)$, and 3) a vector of parameters $(m, n, k_1, k_2, \cdots, k_m)$ and pair of vectors of random variables $(V_{ij})_{j\in[1,m]}, i \in \{1, 2\}$, such that the following inequalities are satisfied:*

$$R_1 = L_1 + T_1, R_2 = L_2 + \frac{I(U_2; \alpha_2 U_1 \oplus \alpha_2 U_2)}{H(U_2)} T_2 \tag{4.22}$$

$$r_{1,0} - T_1 \geq \log q - H(U_1), r_{0,1} - T_2 \geq \log q - H(U_2), \tag{4.23}$$

$$K_1 + r_{1,0} - T_1 \geq \log q + H(X_1) - H(X_1, U_1) \tag{4.24}$$

$$K_2 + r_{0,1} - T_2 \geq \log q + H(X_2) - H(X_2, U_2) \tag{4.25}$$

$$r_{0,1} \leq \log q - H(U_1|X_1, Y_1) \tag{4.26}$$

$$r_{0,1} + L_1 + K_1 \leq \log q + H(X_1) - H(U_1, X_1|Y_1) \tag{4.27}$$

$$L_1 + K_1 \leq I(X_1; U_1 Y_1) \tag{4.28}$$

$$r_{\alpha_2,\beta_2} \leq \log q - H(\alpha_2 U_1 \oplus \beta_2 | X_2, Y_2) \tag{4.29}$$

$$r_{\alpha_2,\beta_2} + L_2 + K_2 \leq \log q + H(X_2) - H(\alpha_2 U_1 \oplus \beta_2 U_2, X_2|Y_2) \tag{4.30}$$

$$L_2 + K_2 \leq I(X_2; U_2 Y_2) \tag{4.31}$$

$$r_{\alpha_3,\beta_3} + L_3 + K_3 \leq \log q + H(X_3) - H(\alpha_3 U_1 \oplus \beta_3 U_2, X_3|Y_3) \tag{4.32}$$

$$R_3 \leq I(X_3; Y_3, \alpha_3 U_1 \oplus \beta_3 U_2), \tag{4.33}$$

where $r_{\alpha,\beta} \triangleq \sum_{i\in[1,m]} \frac{k_i}{n} H(\alpha V_1 \oplus \beta V_2), \forall \alpha, \beta \in \mathbb{F}_q$.

Theorem IV.31. *A rate triple (R_1, R_2, R_3) is achievable if it belongs to cl($\mathcal{R}_{3\text{-}IC}$).*

Proof. We provide an outline of the proof. The coding scheme is similar to the one in [29]. Except that 1) decoder 2 also decodes a linear combination $\alpha_2 U_1 + \beta_2 U_2$, 2) The underlying codes for U_1 and U_2 are QNLC's instead of nested coset codes, and 2) There is an outer code on U_2 which allows decoder 2 to decode U_2 from $\alpha_2 U_1 + \beta_2 U_2$. As a result the rate region is similar to the one in [29] except for a few changes. Bounds (4.23)-(4.25) ensure the existence of jointly typical codewords at each encoder. These bounds are the same with the ones in [29]. Bounds (4.26)-(4.28) ensure errorless decoding at decoder 1, they also remain the same. Inequalities (4.29)-(4.31) correspond to the error events at decoder 2, these bounds are altered to ensure reconstruction of $\alpha_2 U_1 + \beta_2 U_2$, also the rate R_2 is changed and the linear coding rate T_2 is multiplied by $\frac{I(U_2;\alpha_2 U_1 \oplus \alpha_2 U_2)}{H(U_2)}$, which is due to the outer code. Lastly, (4.32)-(4.33) are for the error events at decoder 3, which is also similar to the ones in [29]. □

Remark 32. *For ease of notation, we have dropped the time-sharing random variable Q. The scheme can be enhanced by adding the variable in the standard way.*

Remark 33. *By taking $\alpha_2 = 0$ and $\beta_2 = 1$ and choosing the NQLC parameters so that the codes become a pair of NLC's we recover the bound in [29] as expected.*

Remark 34. *Following the generalizations in [29], this coding scheme can be enhanced by adding additional layers containing the public message codebooks corresponding to the HK strategy.*

CHAPTER V

The Necessity of Finite Block-length Coding

5.1 Introduction

A critical feature of the coding schemes used in PtP communication is that they have large effective lengths. Loosely speaking, this means that each output element in these schemes is a function of the entire input sequence, where the length of the input sequence is asymptotically large. By compressing large blocks at the same time, one can exploit the redundancy in the source. This remains unchanged by the multiterminal nature of the problem in source coding over networks. However, in multiterminal communication it is often desirable to maintain correlation amongst the compressed sequences at each node. Using this fact, we prove that in various network communication problems, the optimal encoding functions which have constant, finite effective lengths (i.e. each output element can be approximated with high precision using a constant, finite number of input elements. This number does not increase with blocklength.). We prove this claim in several steps. First, in Section 5.3 we introduce the notion of the effective length of an encoder. The effective length of the encoder is to be interpreted as the average number of input elements necessary to estimate an output element of that encoder with high accuracy. In Section 5.4, we consider two arbitrary binary block codes (BBC) as defined in [11]. The two encoding functions are applied to two correlated Discrete, Memoryless Sources (DMS). We define the correlation between the outputs of these encoding functions as the average probability that any two output-bits are equal, where the average is over the elements of the output vector. We derive a general upper-bound on the correlation preserving properties of arbitrary pairs of binary encoding functions. Using this bound we conclude that single-letter codes are incapable of sustaining the correlation between their respective outputs. More precisely, we show that as the blocklength increases, the outputs of the quantizers at each terminal become less correlated.

In Section 5.5, we first characterize a set of coding schemes called the Single-Letter Coding (SLC) schemes. We show that the SLC coding schemes are a subset of the IBL coding schemes. The set of SLC schemes is quite general and includes the well-known coding algorithms in multiterminal communications such as the Burger-Tung scheme for Distributed Source Coding (DSC) [45], the Zhang-Berger Strategy for the Multiple-Descriptions (MD) problem [54], the Han-Kobayashi strategy for the Interference Channel (IC) [16], and Marton's coding scheme for the Broadcast channel (BC) [25]. Applying the results from Section 5.4, we conclude that single-letter coding is detrimental to the ability of the encoders to coordinate their outputs. Hence, it is sometimes advantageous, that in the interest of cooperation, finite blocklength codes be used instead. This leads us to hypothesize that the SLC schemes are sub-optimal in multiterminal coding. This hypothesis is proved in the case of the distributed source coding problem in the next chapter.

5.2 Notation

In this section, we introduce the notation used in this chapter. We represent random variables by capital letters such as X, U. Sets are denoted by calligraphic letters such as $\mathcal{X}, \mathcal{U}$. Particularly, the set of natural numbers and real numbers are shown by $\mathbb{N}$, and $\mathbb{R}$, respectively. For a random variable X, the corresponding probability space is $(\mathcal{X}, \mathbf{F}_X, P_X)$, where $\mathbf{F}$ is the underlying σ-field. The set of all subsets of $\mathcal{X}$ is written as $2^{\mathcal{X}}$. There are three different notations used for different classes of vectors. For random variables, the n-length vector $(X_1, X_2, \cdots, X_n), X_i \in \mathcal{X}$ is denoted by $X^n \in \mathcal{X}^n$. For the vector of functions $(e_1(X), e_2(X), \cdots, e_n(X))$ we use the notation $\underline{e}(X)$. The binary string $(i_1, i_2, \cdots, i_n), i_j \in \{0,1\}$ is written as $\mathbf{i}$. As an example, the set of functions $\{\underline{e}_{\mathbf{i}}(X^n)|\mathbf{i} \in \{0,1\}^n\}$ is the set of n-length vectors of functions $(e_{1,\mathbf{i}}, e_{2,\mathbf{i}}, \cdots, e_{n,\mathbf{i}})$ operating on the vector $(X_1, X_2, \cdots, X_n)$ each indexed by an n-length binary string $(i_1, i_2, \cdots, i_n)$. The vector of binary strings $(\mathbf{i}_1, \mathbf{i}_2, \cdots, \mathbf{i}_n)$ denotes the standard basis for the n-dimensional space (e.g. $\mathbf{i}_1 = (0, 0, \cdots, 0, 1)$). The vector of random variables $(X_{j_1}, X_{j_2}, \cdots, X_{j_k})$, $j_i \in [1, n]$, $j_i \neq j_k$, is denoted by $X_{\mathbf{i}}$, where $i_{j_l} = 1, \forall l \in [1, k]$. For example, take $n = 3$, the vector (X_1, X_3) is denoted by X_{101}, and the vector (X_1, X_2) by X_{110}. Particularly, $X_{\mathbf{i}_j} = X_j$, $j \in [1, n]$. Also, for $\mathbf{t} = \underline{1}$, the all-ones vector, $X_{\mathbf{t}} = X^n$. For two binary strings $\mathbf{i}, \mathbf{j}$, we write $\mathbf{i} < \mathbf{j}$ if and only if $i_k < j_k, \forall k \in [1, n]$. For a binary string $\mathbf{i}$ we define $N_{\mathbf{i}} \triangleq w_H(\mathbf{i})$, where w_H denotes the Hamming weight. Lastly, the vector $\sim \mathbf{i}$ is the element-wise complement of $\mathbf{i}$.

5.3 The *Effective Length* of an Encoder

In this chapter, we strive to show that in various network communication problems, the optimal encoding functions operate with finite'*effective length*'. More specifically, we show that in these encoders, each output element is (almost) determined by a finite number of the input elements. To this end, we first define a set of parameters called the dependency spectrum which measure the effective length of an encoding function. The effective length is to be interpreted as the average number of the input elements needed to estimate an output element with high precision. As an initial attempt at characterizing the effective length, let us look at the following example.

Example 1. *Consider the BBC $\underline{e}$ which takes n-length blocks of the input X and produces output elements $e_i(X^n), i \in [1, n]$ using the following mapping:*

$$e_i(X^n) = \begin{cases} X_i + X_{i+1}, & i \neq n, \\ X_n + X_1. & i = n. \end{cases} \tag{5.1}$$

Clearly, each output element is completely determined by the value of X_i and $X_{i\oplus_n 1}$. Based on our interpretation, the effective length of $\underline{e}(X^n)$ is equal to two.

The example suggests a simple way to define the effective length of an encoder:

Definition 2. *For a Boolean function $e : \{0,1\}^n \to \{0,1\}$ defined by $e(X^n) = \sum_{i\in\mathsf{J}} X_i, \mathsf{J} \subset [1,n]$, where the addition operator is the binary addition, the effective length is defined as the cardinality of the set J.*

So, the effective length is a scalar, and the value is equal to the minimum number of input elements whose function gives the output element. However, this elementary definition proves to be frivolous when considering more complicated encoding functions. For a generic encoding function, most of the elements in X^n are correlated with each output $e_i(X^n), i \in [1,n]$. Hence, if we were to define the effective length as described above, the value would be essentially trivial and of little use. For an arbitrary encoding function, it would be more meaningful to ask questions such as how strongly does the first element X_1 affect the output of $e_i(X^n)$? Is this effect amplified when we take X_2 into account as well? One can ask the same question about the effect of an arbitrary subset of random variables $\mathcal{A} = \{X_{k_1}, X_{k_2}, \cdots, X_{k_l}\} \subset \{X_1, X_2, \cdots, X_n\}$. Is there a subset of random variables that (almost) determines the value of the output, similar to the simple case discussed above? In this section, we formulate these questions in mathematical terms, and find a characterization of the dependency spectrum of an encoding function which is a generalization of the above definition. The dependency spectrum is a vector which captures the correlation between different subsets of the input elements with each element of the output. In the next step, we calculate this vector of correlations in terms of $\underline{e}$. With this goal in mind, in this section, we take the intermediate step of finding a decomposition of $\underline{e}$ into a set of functions $\underline{e}_{\mathbf{i}}, \mathbf{i} \in \{0,1\}^n$. In this decomposition, each $\underline{e}_{\mathbf{i}}$ is only a function of a specific subset of the input random variables.

We proceed by formally defining the problem described above. We assume that two correlated DMS's are being fed to two arbitrary encoders, and analyze the correlation between the outputs of these encoders as a function of the dependency spectrum. The following gives the formal definition for DMS's.

Definition 3. *(X, Y) is called a pair of DMS's if we have $P_{X^n,Y^n}(x^n, y^n) = \prod_{i\in[1,n]} P_{X_i,Y_i}(x_i, y_i), \forall n \in \mathbb{N}, x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n$.*

Akin to the results presented in [11, 53], we restrict our attention to the binary block encoders (BBE), which are defined below. The interested reader may refer to [53] for a discussion on extending the analysis to multi-valued (i.e. non-binary) block encoders.

Definition 4. *A Binary-Block-Encoder is characterized by the triple $(\underline{e}, \mathcal{X}, n)$, where e is a mapping $\underline{e} : \mathcal{X}^n \to \{0, 1\}^n$, $\mathcal{X}$ is a set, and n is an integer.*

We refer to a BBE by its corresponding mapping $\underline{e}$. The mapping $\underline{e}$ can be viewed as a vector of functions $(e_i)_{i\in[1,n]}$, where $e_i : \mathcal{X}^n \to \{0, 1\}$. Furthermore, we assume binary input alphabets, i.e. $\mathcal{X} = \{0, 1\}$. Following the method presented in [53], we convert the problem of analyzing a BBE into one where the encoder is a binary real-valued function. Converting the discrete-valued encoding function into a real-valued one is crucial since it allows us to use the rich set of tools available in functional analysis. We present a summary of the functional analysis apparatus used in this work.

Fix a discrete memoryless source X, and a BBE defined by $\underline{e} : \{0, 1\}^n \to \{0, 1\}^n$. Let $P(e_i(X^n) = 1) = q_i$. The real-valued function corresponding to e_i is represented by $\tilde{e}_i$, and is defined as follows:

$$\tilde{e}_i(X^n) = \begin{cases} 1 - q_i, & e_i(X^n) = 1, \\ -q_i. & \text{otherwise.} \end{cases} \tag{5.2}$$

Remark 5. *Note that $\tilde{e}_i, i \in [1, n]$ has zero mean and variance $q_i(1 - q_i)$.*

The random variable $\tilde{e}_i(X^n)$ has finite variance on the probability space $(\mathcal{X}^n, 2^{\mathcal{X}^n}, P_{X^n})$. The set of all such functions is denoted by $\mathcal{H}_{X,n}$. More precisely, we define $\mathcal{H}_{X,n} \triangleq$

$L_2(\mathcal{X}^n, 2^{\mathcal{X}^n}, P_{X^n})$ as the separable Hilbert space of all measurable functions $\tilde{h} : \mathcal{X}^n \to \mathbb{R}$. Since X is a DMS, the isomorphy relation

$$\mathcal{H}_{X,n} = \mathcal{H}_{X,1} \otimes \mathcal{H}_{X,1} \cdots \otimes \mathcal{H}_{X,1} \tag{5.3}$$

holds [34], where $\otimes$ indicates the tensor product.

Example 6. *Let n=1. The Hilbert space $\mathcal{H}_{X,1}$ is the space of all measurable functions $\tilde{h} : \mathcal{X} \to \mathbb{R}$. The space is spanned by the two linearly independent functions $\tilde{h}_1(X) = \mathbb{1}(X)$ and $\tilde{h}_2(X) = \mathbb{1}(\bar{X})$, where $\bar{X} = X \oplus 1$. We conclude that the space is two-dimensional.*

Remark 7. *The tensor operation in $\mathcal{H}_{X,n}$ is real multiplication (i.e. $f_1, f_2 \in \mathcal{H}_{X,1} : f_1(X_1) \otimes f_2(X_2) \triangleq f_1(X_1)f_2(X_2)$). So, if $\{f_i(X)|i \in [1, d]\}$ is a basis for $\mathcal{H}_{X,1}$, a basis for $\mathcal{H}_{X,n}$ would be the set of all the real multiplications of these basis elements: $\{\Pi_{j\in[1,n]} f_{i_j}(X_j), i_j \in [1, d]\}$.*

Example 6 gives a decomposition of the space $\mathcal{H}_{X,1}$. Next, we introduce another decomposition of $\mathcal{H}_{X,1}$ which turns out to be very useful. Let $\mathcal{I}_{X,1}$ be the subset of all measurable functions of X which have 0 mean, and let $\gamma_{X,1}$ be the set of constant real functions of X. $\mathcal{I}_{X,1}$ and $\gamma_{X,1}$ are linear subspaces of $\mathcal{H}_{X,1}$. $\mathcal{I}_{X,1}$ is the nullity of the functional which takes an arbitrary function $\tilde{f} \in \mathcal{H}_{X,1}$ to its expected value $\mathbb{E}_X(\tilde{f})$. The nullity of any non-zero functional is a hyper-space in $\mathcal{H}_{X,1}$. So, $\mathcal{I}_{X,1}$ is a one-dimensional subspace of $\mathcal{H}_{X,1}$. From Remark 5, $\tilde{e}_1 \in \mathcal{I}_{X,1}$. We conclude that any element of $\mathcal{I}_{X,1}$ can be written as $c\tilde{e}_1(X^n), c \in \mathbb{R}$. $\gamma_{X,1}$ is also one dimensional. It is spanned by the function $\tilde{g}(X) = 1$. Consider an arbitrary element $\tilde{f} \in \mathcal{H}_{X,1}$. One can write $\tilde{f} = \tilde{f}_1 + \tilde{f}_2$ where $\tilde{f}_1 = \tilde{f} - \mathbb{E}_X(\tilde{f}) \in \mathcal{I}_{X,1}$, and $\tilde{f}_2 = \mathbb{E}_X(\tilde{f}) \in \gamma_{X,1}$. Hence, $\mathcal{H}_{X,1} = \mathcal{I}_{X,1} \oplus \gamma_{X,1}$ gives a decomposition of $\mathcal{H}_{X,1}$. Replacing $\mathcal{H}_{X,1}$ with $\mathcal{I}_{X,1} \oplus \gamma_{X,1}$ in (5.3), we have:

$$\mathcal{H}_{X,n} = \otimes_{i=1}^n \mathcal{H}_{X,1} = \otimes_{i=1}^n (\mathcal{I}_{X,1} \oplus \gamma_{X,1}) \stackrel{(a)}{=} \oplus_{\mathbf{i}\in\{0,1\}^n} (\mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}), \tag{5.4}$$

where

$$\mathcal{G}_j = \begin{cases} \gamma_{X,1} & j = 0, \\ \mathcal{I}_{X,1} & j = 1, \end{cases}$$

and, in (a), we have used the distributive property of tensor products over direct sums.

Remark 8. *Equation* (5.4)*, can be interpreted as follows: for any $\tilde{e} \in \mathcal{H}_{X,n}, n \in \mathbb{N}$, we can find a decomposition $\tilde{e} = \sum_{\mathbf{i}} \tilde{e}_{\mathbf{i}}$, where $\tilde{e}_{\mathbf{i}} \in \mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}$. $\tilde{e}_{\mathbf{i}}$ can be viewed as the component of $\tilde{e}$ which is only a function of $\{X_{i_j} | i_j = 1\}$. In this sense, the collection $\{\tilde{e}_{\mathbf{i}} | \sum_{j\in[1,n]} i_j = k\}$, is the set of k-letter components of $\tilde{e}$.*

In order clarify the notation, we provide the following two examples.

Example 9. *Let (X_1, X_2) be two independent symmetric binary random variables. Assume $e(X_1, X_2) = X_1 \oplus X_2$ is the binary addition function. In this example $P(e = 1) = \frac{1}{2}$. The corresponding real function is given as follows:*

$$\tilde{e}(X_1, X_2) = \begin{cases} -\frac{1}{2} & X_1 + X_2 \in \{0, 2\}, \\ \frac{1}{2} & X_1 + X_2 = 1, \end{cases}$$

Using Lagrange interpolation we can write $\tilde{e}$ as follows:

$$\begin{aligned} \tilde{e} &= -\frac{1}{2}(X_1 + X_2 - 2)(X_1 + X_2) - \frac{1}{4}(X_1 + X_2 - 1)(X_1 + X_2 - 2) - \frac{1}{4}(X_1 + X_2)(X_1 + X_2 - 1) \\ &= -X_1^2 - X_2^2 - 2X_1X_2 + 2X_1 + 2X_2 - \frac{1}{2}. \end{aligned}$$

The decomposition of $\tilde{e}$ in the form given in (5.4) *is*

$$\tilde{e}_{1,1} = X_1 + X_2 - 2X_1X_2 - \frac{1}{2} = -\frac{1}{2}(1-2X_1)(1-2X_2),$$
$$\tilde{e}_{1,0} = -X_1^2 + X_1 = X_1(1-X_1) \overset{(a)}{=} 0, \tilde{e}_{0,1} = -X_2^2 - X - 2 = X_2(1-X_2) \overset{(a)}{=} 0,$$
$$\tilde{e}_{0,0} = 0.$$

Where (a) holds since the input is chosen from $\{0,1\}$*. In this simple example, the decomposition can be derived directly, by finding functions in the corresponding vector spaces. This is not possible for general n-letter functions. In this section we provide a formula to derive each of these components. Note that $\tilde{e}$ has a single non-zero component in its decomposition. This component is the two-letter function $\tilde{e}_{1,1} = \mathcal{I}_{X,1} \otimes \mathcal{I}_{X,1}$. This is to be expected since the binary addition of two symmetric variables is independent of each variable. So there are no single-letter components. In fact one can verify this directly as follows:*

$$\mathbb{E}_{X_2|X_1}(\tilde{e}|X_1) = X_1 - X_1 = 0, \mathbb{E}_{X_1|X_2}(\tilde{e}|X_2) = X_2 - X_2 = 0.$$

Also, as expected, $\tilde{e}$ can be written as a product of a function of X_1 and another function of X_2, namely $-\frac{1}{2}(1-2X_1)$*, and* $(1-2X_2)$*, respectively. From these three properties, we conclude that $\tilde{e} \in \mathcal{I}_{X,1} \otimes \mathcal{I}_{X,1}$.*

Remark 10. *In the previous example, we found that the binary summation of two independent binary symmetric variables is a two-letter function (i.e. it only has a two-letter component). However, this is not true when the source is not symmetric. When* $P(X=1) \neq P(X=0)$*, the output of the summation is not independent of each of the inputs. One can show that the single-letter components of the summation are non-zero in this case.*

Example 11. *Let $e(X_1, X_2) = X_1 \wedge X_2$ be the binary 'and' function. The corresponding real* *function is:*

$$\tilde{e}(X_1, X_2) = \begin{cases} -\frac{1}{4} & (X_1, X_2) \neq (1,1), \\ \frac{3}{4} & (X_1, X_2) = (1,1). \end{cases}$$

Lagrange interpolation gives $\tilde{e} = X_1X_2 - \frac{1}{4}$. *The decomposition is given by:*

$$\tilde{e}_{1,1} = (X_1 - \frac{1}{2})(X_2 - \frac{1}{2}), \qquad \tilde{e}_{1,0} = \frac{1}{2}(X_1 - \frac{1}{2}), \qquad \tilde{e}_{0,1} = \frac{1}{2}(X_2 - \frac{1}{2}), \qquad \tilde{e}_{0,0} = 0.$$

The variances of these functions are given below:

$$Var(\tilde{e}) = \frac{3}{16}, \qquad Var(\tilde{e}_{0,1}) = Var(\tilde{e}_{1,0}) = Var(\tilde{e}_{1,1}) = \frac{1}{16}.$$

As we shall see in the next sections, these variances play a major role in determining the correlation preserving properties of $\tilde{e}$. *The vector whose elements include these variances is called the dependency spectrum of* e. *In the perspective of the effective length, the function* $\tilde{e}$ *has* $\frac{2}{3}$ *of its variance distributed between* $\tilde{e}_{0,1}$, *and* $\tilde{e}_{1,0}$ *which are single-letter functions, and* $\frac{1}{3}$ *of the variance is on* $\tilde{e}_{1,1}$ *which is a two-letter function.*

Similar to the previous examples, for arbitrary $\tilde{e} \in \mathcal{H}_{X,n}, n \in \mathbb{N}$, we can find a decomposition $\tilde{e} = \sum_{\mathbf{i}} \tilde{e}_{\mathbf{i}}$, where $\tilde{e}_{\mathbf{i}} \in \mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}$. We can characterize $\tilde{e}_{\mathbf{i}}$ in terms of products of the basis elements of $\mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}$ using the following result in linear algebra:

Lemma 12 ([34])**.** *Let* $\mathcal{H}_i, i \in [1,n]$ *be vector spaces over a field* F. *Also, let* $\mathcal{B}_i = \{v_{i,j} | j \in [1, d_i]\}$ *be the basis for* $\mathcal{H}_i$ *where* d_i *is the dimension of* $\mathcal{H}_i$. *Then, any element* $v \in \otimes_{i \in [1,n]} \mathcal{H}_i$ *can be written as* $v = \sum_{j_1 \in [1,d_1]} \sum_{j_2 \in [1,d_2]} \cdots \sum_{j_n \in [1,d_n]} c_{j_1, j_2, \cdots, j_n} v_{j_1} \otimes v_{j_2} \cdots \otimes v_{j_n}$.

Since $\mathcal{G}_{i_j}$'s, $j \in [1,n]$ take values from the set $\{\mathcal{I}_{X,1}, \gamma_{X,1}\}$, they are all one-dimensional.

Let $\tilde{h}$ be defined as follows:

$$\tilde{h}(X) = \begin{cases} 1-q, & \text{if } X = 1, \\ -q. & \text{if } X = 0. \end{cases} \tag{5.5}$$

Then, the single element set $\{\tilde{h}(X)\}$ is a basis for $\mathcal{I}_{X,1}$. Also, the function $\tilde{h}(X) = 1$ spans $\gamma_{X,1}$. So, using Lemma 12, $\tilde{e}_{\mathbf{i}}(X^n) = c_{\mathbf{i}} \prod_{t:i_t=1} \tilde{h}(X_{i_t}), c_i \in \mathbb{R}$. We are interested in the variance of $\tilde{e}_{\mathbf{i}}$'s. In the next proposition, we show that the $\tilde{e}_{\mathbf{i}}$'s are uncorrelated and we derive an expression for the variance of $\tilde{e}_{\mathbf{i}}$.

Proposition 13. *Define* $\mathbf{P}_{\mathbf{i}}$ *as the variance of* $\tilde{e}_{\mathbf{i}}$. *The following hold:*
1) $\mathbb{E}(\tilde{e}_{\mathbf{i}}\tilde{e}_{\mathbf{j}}) = 0, \mathbf{i} \neq \mathbf{j}$, *in other words* $\tilde{e}_{\mathbf{i}}$*'s are uncorrelated.*
2) $\mathbf{P}_{\mathbf{i}} = \mathbb{E}(\tilde{e}_{\mathbf{i}}^2) = c_{\mathbf{i}}^2$.

Proof. 1) follows by direct calculation. 2) holds from the independence of X_i's. □

In the next lemma we find the characterization of $\tilde{e}_{\mathbf{i}}, \mathbf{i} \in \{0,1\}^n$ for general $\tilde{e}$.

Lemma 14. $\tilde{e}_{\mathbf{i}} = \mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}}) - \sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_{\mathbf{j}}$ *gives the unique orthogonal decomposition of* $\tilde{e}$ *into the Hilbert spaces* $\mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \cdots \otimes \mathcal{G}_{i_n}, i^n \in \{0,1\}^n$.

Proof. The uniqueness of such a decomposition follows from the isomorphy relation stated in equation (5.4). We prove that the $\tilde{e}_{\mathbf{i}}$ given in the lemma are indeed the decomposition into the components of the direct sum. Equivalently, we show that 1) $\tilde{e} = \sum_{\mathbf{i}} \tilde{e}_{\mathbf{i}}$, and 2) $\tilde{e}_{\mathbf{i}} \in \mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}, \forall \mathbf{i} \in \{0,1\}^n$.

First we check the equality $\tilde{e} = \sum_{\mathbf{i}} \tilde{e}_{\mathbf{i}}$. Let $\mathbf{t}$ denote the n-length vector whose elements are all ones. We have:

$$\tilde{e}_{\mathbf{t}} = \mathbb{E}_{X^n|X_{\mathbf{t}}}(\tilde{e}|X_{\mathbf{t}}) - \sum_{\mathbf{i}<\mathbf{t}} \tilde{e}_{\mathbf{i}} \overset{(a)}{\Rightarrow} \tilde{e}_{\mathbf{t}} + \sum_{\mathbf{i}<\mathbf{t}} \tilde{e}_{\mathbf{i}} = \tilde{e} \overset{(b)}{\Rightarrow} \tilde{e} = \sum_{\mathbf{i}\in\{0,1\}^n} \tilde{e}_{\mathbf{i}},$$

where in (a) we have used 1) $X_{\mathbf{t}} = X^n$ and 2) for any function $\tilde{f}$ of X^n, $\mathbb{E}_{X^n|X^n}(\tilde{f}|X^n) = \tilde{f}$, and

(b) holds since $\mathbf{i} < \mathbf{t} \Leftrightarrow \mathbf{i} \neq \mathbf{t}$. . It remains to show that $\tilde{e}_{\mathbf{i}} \in \mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}, \forall \mathbf{i} \in \{0,1\}^n$. The next proposition provides a means to verify this property.

Proposition 15. *Fix $\mathbf{i} \in \{0,1\}^n$, define $\mathcal{A}_0 \triangleq \{s|i_s = 0\}$, and $\mathcal{A}_1 \triangleq \{s|i_s = 1\}$. $\tilde{f}$ is an element of $\mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}$ if and only if* (1) *it is constant in all X_s, $s \in \mathcal{A}_0$, and* (2) *it has* 0 *mean on all X_s, when $s \in \mathcal{A}_1$.*

Proof. Please refer to the appendix. □

Returning to the original problem, it is enough to show that $\tilde{e}_{\mathbf{i}}$'s satisfy the conditions in Proposition 15. We prove the stronger result presented in the next proposition.

Proposition 16. *The following hold:*

1) $\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{i}})=0$.

2) $\forall \mathbf{i} \leq \mathbf{k}$, *we have* $\mathbb{E}_{X^n|X_{\mathbf{j}}}(\tilde{e}_{\mathbf{i}}|X_{\mathbf{k}}) = \tilde{e}_{\mathbf{i}}$.

3) $\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{i}}\tilde{e}_{\mathbf{k}}) = 0$*, for* $\mathbf{i} \neq \mathbf{k}$.

4) $\forall \mathbf{k} \leq \mathbf{i} : \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{\mathbf{i}}|X_{\mathbf{k}}) = 0$.

Proof. Please refer to the appendix. □

Remark 17. *The second condition above is equivalent to condition (2) in Proposition 15. The fourth condition is equivalent to (1) in Proposition 15.*

Using propositions 15 and 16, we conclude that $\tilde{e}_{\mathbf{i}} \in \mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}, \forall \mathbf{i} \in \{0,1\}^n$. This completes the proof of Lemma 14. □

The following example clarifies the notation used in Lemma 14.

Example 18. *Consider the case where $n = 2$. We have the following decomposition of $\mathcal{H}_{X,2}$:*

$$\mathcal{H}_{X,2} = (\mathcal{I}_{X,1} \otimes \mathcal{I}_{X,1}) \oplus (\mathcal{I}_{X,1} \otimes \gamma_{X,1}) \oplus (\gamma_{X,1} \otimes \mathcal{I}_{X,1}) \oplus (\gamma_{X,1} \otimes \gamma_{X,1}). \tag{5.6}$$

Let $\tilde{e}(X_1, X_2)$ be an arbitrary function in $\mathcal{H}_{X,2}$. The unique decomposition of $\tilde{e}$ in the form given in (5.6) *is as follows:*

$$
\begin{aligned}
\tilde{e} &= \tilde{e}_{1,1} + \tilde{e}_{1,0} + \tilde{e}_{0,1} + \tilde{e}_{0,0}, \\
\tilde{e}_{1,1} &= \tilde{e} - \mathbb{E}_{X_2|X_1}(\tilde{e}|X_1) - \mathbb{E}_{X_1|X_2}(\tilde{e}|X_2) + \mathbb{E}_{X_1,X_2}(\tilde{e}) \in \mathcal{I}_{X,1} \otimes \mathcal{I}_{X,1}, \\
\tilde{e}_{1,0} &= \mathbb{E}_{X_2|X_1}(\tilde{e}|X_1) - \mathbb{E}_{X_1,X_2}(\tilde{e}) \in \mathcal{I}_{X,1} \times \gamma_{X,1}, \\
\tilde{e}_{0,1} &= \mathbb{E}_{X_1|X_2}(\tilde{e}|X_2) - \mathbb{E}_{X_1,X_2}(\tilde{e}) \in \gamma_{X,1} \otimes \mathcal{I}_{X,1}, \\
\tilde{e}_{0,0} &= \mathbb{E}_{X_1,X_2}(\tilde{e}) \in \gamma_{X,1} \otimes \gamma_{X,1}.
\end{aligned}
$$

It is straightforward to show that each of the $\tilde{e}_{i,j}$'s, $i, j \in \{0, 1\}$, belong to their corresponding subspaces. For instance, $\tilde{e}_{0,1}$ is constant in X_1, and is a 0 *mean function of X_2 (i.e.* $\mathbb{E}_{X_2}\left(\tilde{e}_{0,1}(x_1, X_2)\right) = 0, x_1 \in \{0, 1\}$*), so $\tilde{e}_{0,1} \in \gamma_{X,1} \otimes \mathcal{I}_{X,1}$.*

Lastly, we derive an expression for $\mathbf{P_i}$ using Lemma 14:

Lemma 19. *For arbitrary $e : \{0, 1\}^n \to \{0, 1\}$, let $\tilde{e}$ be the corresponding real function, and let $\tilde{e} = \sum_{\mathbf{i}} \tilde{e}_{\mathbf{i}}$ be the decomposition in the form of Equation* (5.4)*. The variance of each component in the decomposition is given by the following recursive formula* $\mathbf{P_i} = \mathbb{E}_{X_{\mathbf{i}}}(\mathbb{E}^2_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})) - \sum_{\mathbf{j}<\mathbf{i}} \mathbf{P_j}, \forall \mathbf{i} \in \mathbb{F}_2^n$*, where* $\mathbf{P}_{\underline{0}} \triangleq 0$.

Proof.

$$
\begin{aligned}
\mathbf{P_i} &= Var_{X_\mathbf{i}}(\tilde{e}_\mathbf{i}(X^n)) = \mathbb{E}_{X_\mathbf{i}}(\tilde{e}^2_\mathbf{i}(X^n)) - \mathbb{E}^2_{X_\mathbf{i}}(\tilde{e}_\mathbf{i}(X^n)) \\
&\overset{(a)}{=} \mathbb{E}_{X_\mathbf{i}}\left(\left(\mathbb{E}_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i}) - \sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_\mathbf{j}\right)^2\right) - 0 \\
&= \mathbb{E}_{X_\mathbf{i}}\left(\mathbb{E}^2_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})\right) - 2\sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X_\mathbf{i}}\left(\mathbb{E}_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})\tilde{e}_\mathbf{j}\right) + \mathbb{E}_{X_\mathbf{i}}((\sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_\mathbf{j})^2) \\
&\overset{(b)}{=} \mathbb{E}_{X_\mathbf{i}}\left(\mathbb{E}^2_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})\right) - 2\sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X_\mathbf{i}}\left(\mathbb{E}_{X^n|X_\mathbf{i}}(\sum_{\mathbf{l}} \tilde{e}_\mathbf{l}|X_\mathbf{i})\tilde{e}_\mathbf{j}\right) + \mathbb{E}_{X_\mathbf{i}}((\sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_\mathbf{j})^2) \\
&\overset{(c)}{=} \mathbb{E}_{X_\mathbf{i}}\left(\mathbb{E}^2_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})\right) - 2\sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X_\mathbf{i}}\left(\sum_{\mathbf{l}} \mathbb{E}_{X^n|X_\mathbf{i}}(\tilde{e}_\mathbf{l}|X_\mathbf{i})\tilde{e}_\mathbf{j}\right) + \mathbb{E}_{X_\mathbf{i}}((\sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_\mathbf{j})^2) \\
&\overset{(d)}{=} \mathbb{E}_{X_\mathbf{i}}\left(\mathbb{E}^2_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})\right) - 2\sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X_\mathbf{i}}\left(\sum_{\mathbf{l}} \mathbb{1}(\mathbf{l} \leq \mathbf{i})\mathbb{E}_{X^n|X_\mathbf{i}}(\tilde{e}_\mathbf{l}|X_\mathbf{i})\tilde{e}_\mathbf{j}\right) + \mathbb{E}_{X_\mathbf{i}}((\sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_\mathbf{j})^2) \\
&\overset{(e)}{=} \mathbb{E}_{X_\mathbf{i}}\left(\mathbb{E}^2_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})\right) - 2\sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X_\mathbf{i}}\left(\sum_{\mathbf{l}<\mathbf{i}} \tilde{e}_\mathbf{l}\tilde{e}_\mathbf{j}\right) + \mathbb{E}_{X_\mathbf{i}}((\sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_\mathbf{j})^2) \\
&\overset{(f)}{=} \mathbb{E}_{X_\mathbf{i}}\left(\mathbb{E}^2_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})\right) - 2\sum_{\mathbf{j}<\mathbf{i}}\sum_{\mathbf{l}<\mathbf{i}} \mathbb{1}(\mathbf{j} = \mathbf{l})\mathbb{E}_{X_\mathbf{i}}\left(\tilde{e}_\mathbf{l}\tilde{e}_\mathbf{j}\right) + \mathbb{E}_{X_\mathbf{i}}((\sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_\mathbf{j})^2) \\
&= \mathbb{E}_{X_\mathbf{i}}\left(\mathbb{E}^2_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})\right) - 2\sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X_\mathbf{j}}(\tilde{e}^2_\mathbf{j}) + \mathbb{E}_{X_\mathbf{i}}((\sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_\mathbf{j})^2) \\
&= \mathbb{E}_{X_\mathbf{i}}(\mathbb{E}^2_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})) - 2\sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X_\mathbf{j}}(\tilde{e}^2_\mathbf{j}) + \sum_{\mathbf{j}<\mathbf{i}}\sum_{\mathbf{k}<\mathbf{i}} \mathbb{E}_{X_\mathbf{i}}(\tilde{e}_\mathbf{j}\tilde{e}_\mathbf{k}) \\
&\overset{(g)}{=} \mathbb{E}_{X_\mathbf{i}}(\mathbb{E}^2_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})) - 2\sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X_\mathbf{j}}(\tilde{e}^2_\mathbf{j}) + \sum_{\mathbf{j}<\mathbf{i}}\sum_{\mathbf{k}<\mathbf{i}} \mathbb{1}(\mathbf{j} = \mathbf{k})\mathbb{E}_{X_\mathbf{i}}(\tilde{e}^2_\mathbf{j}) \\
&= \mathbb{E}_{X_\mathbf{i}}(\mathbb{E}^2_{X^n|X_\mathbf{i}}(\tilde{e}|X_\mathbf{i})) - \sum_{\mathbf{j}<\mathbf{i}} \mathbf{P_j},
\end{aligned}
$$

where (a) follows from 1) in Proposition 16, b) follows from the decomposition in Equation (5.4), (c) uses linearity of expectation, (d) uses 4) in Proposition 16, (e) holds from 2) in 16, and in (f) and (g) we have used 1) in Proposition 16. □

Corollary 20. *For an arbitrary* $e : \{0,1\}^n \to \{0,1\}$ *with corresponding real function* $\tilde{e}$*, and decomposition* $\tilde{e} = \sum_\mathbf{i} \tilde{e}_\mathbf{i}$*. Let the variance of* $\tilde{e}$ *be denoted by* $\mathbf{P}$*. Then,* $\mathbf{P} = \sum_\mathbf{i} \mathbf{P_i}$*. The corollary is a special case of Lemma 19, where we have taken* $\mathbf{i}$ *to be the all ones vector.*

The following provides a definition of the dependency spectrum of a BCC:

Definition 21 (Dependency Spectrum)**.** *For an encoding function e, the vector of variances* $(P_{\mathbf{i}})_{\mathbf{i}\in\{0,1\}^n}$ *is called the dependency spectrum of e.*

So far we have found the decomposition of $\underline{\tilde{e}}$ into components $\underline{\tilde{e}}_{\mathbf{i}}$. In the next section, we find that the answers to the questions posed in the beginning of the section, are related to the variance of each component in this decomposition. We can reiterate the claims made at the beginning of the section based on our new understanding of the dependency spectrum: In various network communication problems, for an optimal encoding function $\underline{e}$, the variance of $\underline{\tilde{e}}_{\mathbf{i}}$ is small when $\underline{\tilde{e}}_{\mathbf{i}}$ is a function of a large subset of the input. We prove this claim in the next three sections.

5.4 Correlation Preservation in Arbitrary Encoders

In this work, we prove that in various multiterminal communication problems, encoders whose effective blocklength is asymptotically large are sub-optimal. This is in contrast with PtP communications where optimality is achieved only as the length of the encoder approaches infinity. The secret to this apparent discrepancy is that in multiterminal communications, it is often required that the encoders, having received correlated inputs, produce outputs which are correlated with each other. This requirement can be due to explicit constraints in the problem statement such as joint distortion measures, or it can be due to implicit factors such as the need for interference alignment, or the nature of the shared communication channel. In the latter case, correlation between the outputs is necessary as a means for further cooperation between the transmitters. It turns out that pairs of encoders with large effective lengths are inefficient in coordinating their outputs. This is due to the fact that such encoding functions are ineffective in preserving correlation. The loss of correlation undermines the encoders' ability to conspire to take advantage of the multiterminal nature of the problem. In PtP communication problems, where there is only one

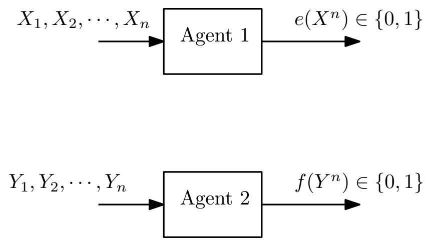

Figure 5.1: Correlated Boolean decision functions.

transmitter, the necessity for cooperation does not manifest itself. For this reason, although encoders with asymptotically large blocklengths are optimal in PtP communications, they are sub-optimal in the network communication case. When transmitting data over networks, there is a trade-off between the sender's need to transmit in a PtP optimal manner, and the networks' requirements for coordination among the transmitters. This results in the so-called 'sweet-spot' for the length of an encoder. In this section, we show that to achieve a fixed correlation between the outputs of the encoders, most of the power in the dependency spectrum is distributed on the decomposition elements with lower effective lengths. Alternatively, we derive a bound on the correlation between the outputs of two arbitrary encoding functions given their dependency spectrums.

Our goal is to bound the correlation preserving properties of general n-length encoding functions. As a first step, we derive bounds on the correlation between the outputs of two arbitrary Boolean functions (i.e. functions whose output is a binary scalar). The result that follows can be viewed as the solution to a more fundamental problem than what we discussed so far. This general problem is the main motivation for the work in [53]. We explain a summary of the setup here. Consider the two distributed agents shown in figure 5.1. The generality of the setup is in that these agents can be two encoders in the distributed source coding problem, or two transmitters in the interference channel problem, or Alice and Bob in a secret key-generation problem, or they can be two agents in a distributed control problem. They each recieve a binary string. For simplicity assume that the strings are produced based on a memoryless distribution. The strings are correlated with each other. The first agent is to make a binary decision based on its input. The second aims to guess the other's decision. In [53], the author considers the problem when 1) The only constraint on the agents is that the entropy of the binary output is fixed (e.g. they can not output constants to improve estimation accuracy.) 2) The agents are completely cooperative in the sense that barring their personal constraints, they choose the Boolean decision function which maximizes estimation accuracy, and 3) prior to the start of the process each agent is made aware of the other's decision function. It was shown that the best strategy is for both users to output a single element of the string without further processing (e.g. each user outputs the first element of its corresponding string). This means that further processing of the binary strings by the two agents can not induce additional correlation. The result was used extensively in a variety of areas such as information theory, security, and control [4, 10, 24]. However, the first assumption proves to be too restrictive in many cases. Here, we relax this assumption by assuming that the agents may have additional constraints. Particularly, we assume that the users have constraints on the effective length of their decision functions. This is a valid assumption, for instance, in the case of communication systems, the users have restrictions on their effective lengths due to the rate-distortion requirements in the problem.

We proceed with presenting the main result of this section. Let (X, Y) be a pair of DMS's. Consider two arbitrary Boolean functions $e : \mathcal{X}^n \to \{0, 1\}$ and $f : \mathcal{Y}^n \to \{0, 1\}$. Let $q \triangleq P(e = 1)$, $r \triangleq P(f = 1)$. The following theorem provides an upper-bound on the probability of equality between the functions $e(X^n)$ and $f(Y^n)$. . The proof uses some of the ideas used in [53].

Theorem V.22. *Let $\epsilon \triangleq P(X \neq Y)$, the following bound holds:*

$$2\sqrt{\sum_{\mathbf{i}} \mathbf{P_i}}\sqrt{\sum_{\mathbf{i}} \mathbf{Q_i}} - 2\sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}} \leq P(e(X^n) \neq f(Y^n))$$
$$\leq 1 - 2\sqrt{\sum_{\mathbf{i}} \mathbf{P_i}}\sqrt{\sum_{\mathbf{i}} \mathbf{Q_i}} + 2\sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}},$$

where $C_{\mathbf{i}} \triangleq (1-2\epsilon)^{N_{\mathbf{i}}}$, $\mathbf{P}_{\mathbf{i}}$ *is the variance of* $\tilde{e}_{\mathbf{i}}$, *and* $\underline{\tilde{e}}$ *is the real function corresponding to* $\underline{e}$, *and* $\mathbf{Q}_{\mathbf{i}}$ *is the variance of* $\tilde{f}_{\mathbf{i}}$, *and finally,* $N_{\mathbf{i}} \triangleq w_H(\mathbf{i})$.

Remark 23. *The value* $C_{\mathbf{i}} = (1-2\epsilon)^{N_{\mathbf{i}}}$ *is decreasing with* $N_{\mathbf{i}}$. *So, in order to increase* $P(e(X^n) \neq f(Y^n))$, *most of the variance* $\mathbf{P}_{\mathbf{i}}$ *should be distributed on* $\tilde{e}_{\mathbf{i}}$ *which have lower* $N_{\mathbf{i}}$ *(i.e. operate on smaller blocks). Particularly, the lower bound is minimized by setting*

$$\mathbf{P}_{\mathbf{i}} = \begin{cases} 1 & \mathbf{i} = \mathbf{i}_1, \\ 0 & otherwise. \end{cases}$$

This recovers the result in [53].

Remark 24. *For fixed* $\mathbf{P}_{\mathbf{i}}$, *the lower-bound is minimized by taking* $\tilde{e}$, *and* $\tilde{f}$ *to be the same functions.*

Proof. The proof involves three main steps. In the first two steps we prove the lower bound. First, we bound the Pearson correlation between the real-valued functions $\tilde{e}$, and $\tilde{f}$. In the second step, we relate the correlation to the probability that the two functions are equal and derive the necessary bounds. Finally, in the third step we use the lower bound proved in the first two steps to derive the upper bound.

Step 1: From Remark 5, the expectation of both functions is 0. So, the Pearson correlation is given by $\frac{\mathbb{E}_{X^n,Y^n}(\tilde{e}\tilde{f})}{(rq(1-q)(1-r))^{\frac{1}{2}}}$. Our goal is to bound this value. We have:

$$\mathbb{E}_{X^n,Y^n}(\tilde{e}\tilde{f}) \overset{(a)}{=} \mathbb{E}_{X^n,Y^n}\left((\sum_{\mathbf{i}\in\{0,1\}^n} \tilde{e}_{\mathbf{i}})(\sum_{\mathbf{k}\in\{0,1\}^n} \tilde{f}_{\mathbf{k}}) \right) \overset{(b)}{=} \sum_{\mathbf{i}\in\{0,1\}^n} \sum_{\mathbf{k}\in\{0,1\}^n} \mathbb{E}_{X^n,Y^n}(\tilde{e}_{\mathbf{i}}\tilde{f}_{\mathbf{k}}). \tag{5.7}$$

In (a) we have used Remark 8, and in (b) we use linearity of expectation. Using the fact that $\tilde{e}_{\mathbf{i}} \in \mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}$ and Lemma 12, we have:

$$\tilde{e}_{\mathbf{i}} = c_{\mathbf{i}} \prod_{t:i_t=1} \tilde{e}_t(X_{i_t}), \tilde{f}_{\mathbf{k}} = d_{\mathbf{k}} \prod_{t:k_t=1} \tilde{f}_t(X_{k_t}). \tag{5.8}$$

We replace $\tilde{e}_{\mathbf{i}}$ and $\tilde{f}_{\mathbf{k}}$ in (5.7):

$$\mathbb{E}_{X^n,Y^n}(\tilde{e}_{\mathbf{i}}\tilde{f}_{\mathbf{k}}) \overset{(5.8)}{=} \mathbb{E}_{X^n,Y^n}\left(\left(c_{\mathbf{i}}\prod_{t:i_t=1}\tilde{e}(X_{i_t})\right)\left(d_{\mathbf{k}}\prod_{s:k_s=1}\tilde{f}(Y_{k_s})\right)\right) \overset{(a)}{=} c_{\mathbf{i}}d_{\mathbf{k}}\mathbb{E}_{X^n,Y^n}\left(\prod_{t:i_t=1}\tilde{e}(X_{i_t})\prod_{s:k_s=1}\tilde{f}(Y_{k_s})\right)$$
$$\overset{(b)}{=} c_{\mathbf{i}}d_{\mathbf{k}}\mathbb{E}_{X^n,Y^n}\left(\prod_{t:i_t=1,k_t=1}\tilde{e}(X_{i_t})\tilde{f}(Y_{k_t})\right)\mathbb{E}_{X^n}\left(\prod_{t:i_t=1,k_t=0}\tilde{e}(X_{i_t})\right)\mathbb{E}_{Y^n}\left(\prod_{t:i_t=0,k_t=1}\tilde{f}(Y_{k_t})\right)$$
$$\overset{(c)}{=} \mathbb{1}(\mathbf{i}=\mathbf{k})c_{\mathbf{i}}d_{\mathbf{k}}\prod_{t:i_t=1}\mathbb{E}_{X^n,Y^n}\left(\tilde{e}(X_{i_t})\tilde{f}(Y_{i_t})\right) \overset{(d)}{\leq} \mathbb{1}(\mathbf{i}=\mathbf{k})c_{\mathbf{i}}d_{\mathbf{k}}(1-2\epsilon)^{N_{\mathbf{i}}}\prod_{t:i_t=1}\mathbb{E}^{\frac{1}{2}}_{X^n}\left(\tilde{e}^2(X_{i_t})\right)\mathbb{E}^{\frac{1}{2}}_{Y^n}\left(\tilde{f}^2_{t,m}(Y)\right)$$
$$\overset{(e)}{=} \mathbb{1}(\mathbf{i}=\mathbf{k})(1-2\epsilon)^{N_{\mathbf{i}}}c_{\mathbf{i}}d_{\mathbf{k}} \overset{(f)}{=} \mathbb{1}(\mathbf{i}=\mathbf{k})(1-2\epsilon)^{N_{\mathbf{i}}}\mathbf{P}^{\frac{1}{2}}_{\mathbf{i}}\mathbf{Q}^{\frac{1}{2}}_{\mathbf{i}} = \mathbb{1}(\mathbf{i}=\mathbf{k})C_{\mathbf{i}}\mathbf{P}^{\frac{1}{2}}_{\mathbf{i}}\mathbf{Q}^{\frac{1}{2}}_{\mathbf{i}}. \tag{5.9}$$

(a) follows from linearity of expectation. In (b) we have used the fact that in a pair of DMS's, X_i and Y_j are independent for $i \neq j$. (c) holds since from Proposition 16, $\mathbb{E}(\tilde{e}_i) = \mathbb{E}(\tilde{f}_i) = 0, \forall i \in [1, n]$. We prove (d) in Lemma 25 below. In (e) and (f) we have used proposition 13.

Lemma 25. *Let $g(X)$ and $h(Y)$ be two arbitrary zero-mean, real valued functions, then:*

$$\mathbb{E}_X(g(X)h(Y)) \leq (1-2\epsilon)\mathbb{E}^{\frac{1}{2}}_X(g^2(X))\mathbb{E}^{\frac{1}{2}}_Y(h^2(Y)).$$

Proof. Please refer to the appendix. □

Using equations (5.7) and (5.9) we get:

$$\mathbb{E}_X(\tilde{e}\tilde{f}) \leq \sum_{\mathbf{i}} C_{\mathbf{i}}\mathbf{P}^{\frac{1}{2}}_{\mathbf{i}}\mathbf{Q}^{\frac{1}{2}}_{\mathbf{i}}.$$

Step 2: We use the results from step one to derive a bound on $P(e \neq f)$. Define $a \triangleq P(e(X^n) = 1, f(Y^n) = 1)$, $b \triangleq P(e(X^n) = 0, f(Y^n) = 1)$, $c \triangleq P(e(X^n) = 1, f(Y^n) = 0)$, and

$d \triangleq P(e(X^n) = 0, f(Y^n) = 0)$, then

$$\mathbb{E}_{X^n,Y^N}(\tilde{e}(X^n)\tilde{f}(Y^n)) = a(1-q)(1-r) - bq(1-r) - c(1-q)r + dqr, \tag{5.10}$$

We write this equation in terms of $\sigma \triangleq P(f \neq g)$, q, and r using the following relations:

$$1)a + c = q, \qquad 2)b + d = 1 - q, \qquad 3)a + b = r, \qquad 4)c + d = 1 - r, \qquad 5)b + c = \sigma.$$

Solving the above we get:

$$a = \frac{q + r - \sigma}{2}, \qquad b = \frac{r + \sigma - q}{2}, \qquad c = \frac{q - r + \sigma}{2}, \qquad d = 1 - \frac{q + r + \sigma}{2}. \tag{5.11}$$

We replace a, b, c, and d in (5.10) by their values in (5.11):

$$\frac{\sigma}{2} \geq (\frac{q+r}{2})(1-q)(1-r) + (\frac{q-r}{2})q(1-r) + (\frac{r-q}{2})(1-q)r + qr(1 - \frac{q+r}{2}) - \sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}}$$
$$\Rightarrow \sigma \geq q + r - 2rq - 2 \sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}}$$
$$\Rightarrow \sigma \geq (\sqrt{q(1-r)} - \sqrt{r(1-q)})^2 + 2\sqrt{q(1-q)r(1-r)} - 2 \sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}}$$
$$\Rightarrow \sigma \geq 2\sqrt{q(1-q)r(1-r)} - 2 \sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}}$$

On the other hand $\mathbb{E}_X(\tilde{e}^2) = q(1-q) = \sum_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}$, where the last equality follows from the fact that $\tilde{e}_{\mathbf{i}}$'s are uncorrelated. This proves the lower bound. Next we use the lower bound to derive the upper bound.

Step 3: The upper-bound can be derived by considering the function $h(Y^n)$ to be the complement of $f(Y^n)$ (i.e. $h(Y^n) \triangleq 1 \oplus_2 f(Y^n)$.) In this case $P(h(Y^n) = 1) = P(f(Y^n) = 0) = 1 - r$.

The corresponding real function for $h(Y^n)$ is:

$$\tilde{h}(Y^n) = \begin{cases} r & h(Y^n) = 1, \\ -(1-r) & h(Y^n) = 0, \end{cases} = \begin{cases} r & f(Y^n) = 0, \\ -(1-r) & f(Y^n) = 1, \end{cases} \Rightarrow \tilde{h}(Y^n) = -\tilde{f}(Y^n).$$

So, $\tilde{h}(Y^n) = -\sum_{\mathbf{i}} \tilde{f}_{\mathbf{i}}$. Using the same method as in the previous step, we have:

$$\begin{aligned} \mathbb{E}_{X^n,Y^n}(\tilde{e}\tilde{h}) = -\mathbb{E}_{X^n,Y^n}(\tilde{e}\tilde{f}) \leq \sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}} \\ \Rightarrow P(e(X^n) \neq h(Y^n)) \geq 2\sqrt{\sum_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}}\sqrt{\sum_{\mathbf{i}} \mathbf{Q}_{\mathbf{i}}} - 2\sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}} \end{aligned}$$

On the other hand $P(e(X^n) \neq h(Y^n)) = P(e(X^n) \neq 1 \oplus f(Y^n)) = P(e(X^n) = f(Y^n)) = 1 - P(e(X^n) \neq f(Y^n)$. So,

$$\begin{aligned} 1 - P(e(X^n) \neq f(Y^n)) \geq 2\sqrt{\sum_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}}\sqrt{\sum_{\mathbf{i}} \mathbf{Q}_{\mathbf{i}}} - 2\sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}} \\ \Rightarrow P(e(X^n) \neq f(Y^n)) \leq 1 - 2\sqrt{\sum_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}}\sqrt{\sum_{\mathbf{i}} \mathbf{Q}_{\mathbf{i}}} + 2\sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}}. \end{aligned}$$

This completes the proof. □

Corollary 26. *We can simplify the bound in Theorem V.22 as follows:*

$$2\sum_{\mathbf{i}} (1 - C_{\mathbf{i}}) \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}} \leq P(e(X^n) \neq f(Y^n)) \leq 1 - 2\sum_{\mathbf{i}} (1 - C_{\mathbf{i}}) \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}}$$

Proof.

$$\sigma \geq 2\sqrt{\sum_{\mathbf{i}} \mathbf{P_i}}\sqrt{\sum_{\mathbf{i}} \mathbf{Q_i}} - 2\sum_{\mathbf{i}} C_{\mathbf{i}}\mathbf{P}_{\mathbf{i}}^{\frac{1}{2}}\mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}}$$
$$\overset{(a)}{\Rightarrow} \sigma \geq 2\sum_{\mathbf{i}} \mathbf{P}_i^{\frac{1}{2}}\mathbf{Q}_i^{\frac{1}{2}} - 2\sum_{\mathbf{i}} C_{\mathbf{i}}\mathbf{P}_{\mathbf{i}}^{\frac{1}{2}}\mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}}$$
$$\Rightarrow \sigma \geq 2\sum_{\mathbf{i}}(1 - C_{\mathbf{i}})\mathbf{P}_{\mathbf{i}}^{\frac{1}{2}}\mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}}.$$

In (a) we have used the Cauchy-Schwarz inequality. □

Next, we relate the previous theorem with BBE's.

Definition 27. *Consider two BBE's characterized by* $(\underline{e}, \mathcal{X}, n)$ *and* $(\underline{f}, \mathcal{Y}, n)$*. Define the Expected Single-Letter Correlation (ESLC) between* e *and* f *with respect to the sources* X *and* Y *as:*

$$\mathcal{E}(\underline{e}(X^n), \underline{f}(Y^n)) \triangleq \frac{1}{n}\sum_{j=1}^{n} P(e_j(X^n) = f_j(Y^n)).$$

Here each (e_j, f_j) is a pair of Boolean functions. So, using Theorem V.22, we can derive a bound on the ESLC between e and f as well. We use this bound in the next section to prove the sub-optimality of the single-letter coding schemes.

5.5 Correlation in Single Letter Coding Schemes

In this section, we first characterize a group of coding strategies we call Single-letter Random Coding Schemes (SLCS). These coding strategies include a broad range of strategies in the literature including Shannon's PtP source coding and channel coding schemes, the Berger-Tung coding strategy in distributed source coding [45], the Han-Kobayashi scheme in the interference channel [16], Marton's coding strategy for the broadcast channel [25], and the Zhang-Berger multiple-descriptions coding scheme [54]. Using the results from the previous sections, we bound the correlation preserving properties of the SLCS's. We show that encoding functions generated by SLCS's have most of their variance $\mathbf{P_i}$ dis-

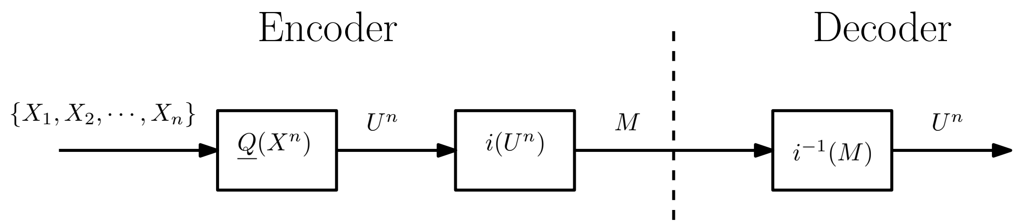

Figure 5.2: Point-to-point source coding example

tributed among $\tilde{e}_{\mathbf{i}}$'s which operate on large blocks. This along with Theorem V.22 proves that such schemes are inefficient in preserving correlation. These results are used in the next section to show the sub-optimality of the SLCS's in various problems. First, we provide our definition of a coding scheme. We explain the term through the following example before providing the definition.

Example 28. *Consider the PtP source coding problem depicted in Figure 5.2. This along with the PtP channel coding problem are the two fundamental problems in information theory. While this is a basic setup, it possesses the complexities involved in the more advanced setups considered in the next section. The problem was first solved by Shannon [38]. A discrete memoryless source X is being fed to an encoder. The encoder utilizes the mapping* $\underline{Q} : \mathcal{X}^n \to \mathcal{U}^n$ *to compress the source sequence. The image of* $\underline{Q}$ *is indexed by the bijection* $i : Im(\underline{Q}) \to [1, |Im(\underline{Q})|]$. *The index* $M \triangleq i(Q(X^n))$ *is sent to the decoder. The decoder reconstructs the compressed sequence* $U^n \triangleq i^{-1}(M) = Q(X^n)$. *The efficiency of the reconstruction is evaluated based on the separable distortion criteria* $d_n : \mathcal{X}^n \times \mathcal{U}^n \to [0, \infty)$. *The separability property means that* $d_n(x^n, u^n) = \sum_{i \in [1,n]} d_1(x_i, u_i)$. *We assume that the alphabets* $\mathcal{X}$ *and* $\mathcal{U}$ *are both binary. The rate of transmission is defined as* $R \triangleq \frac{1}{n} \log |Im(\underline{Q})|$, *and the average distortion is defined as* $\frac{1}{n}\mathbb{E}(d_n(X^n, U^n))$. *The goal is to choose* $\underline{Q}$ *such that the rate-distortion tradeoff is optimized. Note that the choice of the bijection 'i' is irrelevant to the performance of the system. The following Lemma gives the achievable RD region for this setup.*

Lemma 29. *[38] For the source X and distortion criteria* $d_1 : \{0, 1\} \times \{0, 1\} \to [0, \infty)$, *fix a conditional distribution $p_{U|X}(u|x), x, u \in \{0,1\}$. The rate-distortion pair $(R, D) = (r, \mathbb{E}_{X,U}(d_1(X,U)))$ is achievable for all $r < I(U;X)$.*

Proof. In order to characterize the properties of the SLCS's, we give an outline of the scheme used to achieve the above RD region. Fix $n \in \mathbb{N}$, and $\epsilon > 0$. Define $P_U(u) = \mathbb{E}_X\{P_{U|X}(u|X))\}$. Proving achievability is equivalent to showing the existence of a suitable encoding function $\underline{Q}(X^n)$. In [38], a randomly generated encoding function is constructed with the aid of a set of vectors called the codebook, and an assignment rule called typicality encoding. We construct the codebook C as follows. Let $A_\epsilon^n(U) \triangleq \{u^m \big| \frac{1}{n}|w_H(u^m) - P_U(1)| < \epsilon\}$ be the set of n-length binary vectors which are ϵ-typical with respect to P_U. Choose $\lceil 2^{nR} \rceil$ vectors from $A_\epsilon^n(U)$ randomly and uniformly. Let $C \subset A_\epsilon^n(U)$ be the set of these vectors. The encoder constructs the encoding function $\underline{Q}(X^n)$ as follows. For an arbitrary sequence $x^n \in \{0,1\}^n$, define $A_\epsilon^n(U|x^n)$ as the set of vectors in C which are jointly ϵ-typical with x^n based on $P_{U|X}$. The vector $\underline{Q}(x^n)$ is chosen randomly and uniformly from $A_\epsilon^n(U|x^n) \cap C$. The probabilistic choice of the codewords as well as the quantization, puts a distribution on the random function $\underline{Q}$. It can be shown that as n becomes larger codes produced based on this distribution $P(\underline{Q})$ achieve the rate-distortion vector (R,D) with probability approaching one. □

Remark 30. *It is well-known that in the above scheme, the codebook generation process could be altered in the following way. Instead of choosing the codewords randomly and uniformly from the set of typical sequences $A_\epsilon^n(U)$, the encoder can produce each codeword independent of the others and with the distribution $P_{U^n}(u^n) = \Pi_{i\in[1,n]} P_U(u_i)$. However, the discussion that follows remains unchanged regardless of which of these codebook generation methods are used.*

In the previous example, the distribution $P(\underline{Q})$ completely characterizes the coding scheme in the sense that any two coding schemes with the same $P(\underline{Q})$, would have the same performance in terms of achievable rate-distortion. Based on this notion, we provide our definition of a coding scheme:

Definition 31. *A Coding Scheme $\mathcal{S}$ is characterized by a probability distribution $P_{\mathcal{S}}(\underline{e})$ on the set of functions $\underline{e} : \{0,1\}^n \to \{0,1\}^n$.*

Whenever the choice of coding scheme is clear we drop the subscript $\mathcal{S}$ to denote the distribution by $P_{\underline{E}}(\underline{e})$. We can make the following observations about the encoding scheme in Example 28.

1) Codewords are chosen pairwise independently. So, for two input sequences x^n, and y^n given that $A_\epsilon^n(U|x^n) \cap A_\epsilon^n(V|y^n) = \phi$, the two vectors are not compressed to the same codeword, hence they are compressed to independently generated codewords (i.e. $\underline{Q}(x^n)$ is chosen independently of $\underline{Q}(y^n)$).

2) As n becomes large, the ith output element $Q_i(X^n)$ is correlated with the input sequence X^n only through the ith input element X_i:

$$\forall \delta > 0, \exists n \in \mathbb{N} : m > n \Rightarrow \forall x^m \in \{0,1\}^m, v \in \{0,1\},$$

$$|P_{\mathcal{S}}(Q_i(X^m) = v|X^m = x^m) - P_{\mathcal{S}}(Q_i(X^m) = v|X_i = x_i)| < \epsilon.$$

3) The encoder is insensitive to permutations. Due to typicality encoding the probability that a vector x^n is mapped to y^n depends only on their joint type and is equal to the probability that $\pi(x^n)$ is mapped to $\pi(y^n)$.

Remark 32. *The second property demands more explanation. Note that for a fixed quantization function $\underline{q} : \{0,1\}^m \to \{0,1\}^m$, $\underline{q}(X^m)$ is a function of X^m. However, without the knowledge that which encoding function is used, $Q_i(X^m)$ is related to X^m only through X_i. In other words, averaged over all encoding functions, the effects of the rest of the elements diminishes. We provide a proof of this statement below:*

Proof. First, we are required to provide some definitions relating to the joint type of pairs of sequences. For binary strings u^m, x^m, define $N(a,b|u^m,x^m) \triangleq |\{j|u_j = a, x_j = b\}|$, that is the number of indices j for which the value of the pair (u_j, x_j) is (a, b). For $s, t \in \{0, 1\}$, define $l_{s,t} \triangleq N(s, t|u^m, x^m)$, the vector $(l_{0,0}, l_{0,1}, l_{1,0}, l_{1,1})$ is called the joint type of (u^m, x^m). For fixed x^m The set of sequences $T(l_{0,0}, l_{0,1}, l_{1,0}, l_{1,1}) = \{u^m | N(s, t|u^m, x^m) = l_{s,t}, s, t \in \{0, 1\}\}$, is the set of vectors which have joint type $(l_{0,0}, l_{0,1}, l_{1,0}, l_{1,1})$ with the sequence x^m. Fix $m, \epsilon > 0$, and define $\mathcal{L}_{\epsilon,n} \triangleq \{(l_{0,0}, l_{0,1}, l_{1,0}, l_{1,1}) || \frac{l_{s,t}}{m} - P_{U,X}(s, t)| < \epsilon\}$. Then for the conditional typical set $A_\epsilon^n(U|x^m)$ defined above we can write

$$A_\epsilon^n(U|x^m) = \cup_{(l_{0,0},l_{0,1},l_{1,0},l_{1,1})\in\mathcal{L}_{\epsilon,n}} T(l_{0,0}, l_{0,1}, l_{1,0}, l_{1,1}).$$

The type of x^m, denoted by (l_0, l_1) is defined in a similar manner. Since $Q_i(X^m)$ are chosen uniformly from the set $A_\epsilon^n(U|x^m)$, we have:

$$\begin{aligned}
&P_{\mathcal{S}}(Q_i(X^m) = v | X^m = x^m) = \frac{|\{u^m | u_1 = v, u^m \in A_\epsilon^n(U|x^m)\}|}{|\{u^m | u^m \in A_\epsilon^n(U|x^m)\}|} \\
&= \frac{\sum_{(l_{0,0},l_{0,1},l_{1,0},l_{1,1})\in\mathcal{L}_{\epsilon,n}} |\{u^m | u_1 = v, u^m \in T(l_{0,0}, l_{0,1}, l_{1,0}, l_{1,1})\}|}{\sum_{(l_{0,0},l_{0,1},l_{1,0},l_{1,1})\in\mathcal{L}_{\epsilon,n}} |\{u^m | u^m \in T(l_{0,0}, l_{0,1}, l_{1,0}, l_{1,1})\}|} \\
&= \frac{\sum_{(l_{0,0},l_{0,1},l_{1,0},l_{1,1})\in\mathcal{L}_{\epsilon,n}} \binom{l_{x_1}-1}{l_{u_1,x_1}-1}\binom{l_{\bar{x}_1}}{l_{u_1,\bar{x}_1}}}{\sum_{(l_{0,0},l_{0,1},l_{1,0},l_{1,1})\in\mathcal{L}_{\epsilon,n}} \binom{l_{x_1}}{l_{u_1,x_1}}\binom{l_{\bar{x}_1}}{l_{u_1,\bar{x}_1}}} \\
&= \frac{\sum_{(l_{0,0},l_{0,1},l_{1,0},l_{1,1})\in\mathcal{L}_{\epsilon,n}} \frac{(l_{x_1}-1)!}{(l_{u_1,x_1}-1)!(l_{x_1}-l_{u_1,x_1})!} \frac{l_{\bar{x}_1}!}{l_{u_1,\bar{x}_1}!(l_{\bar{x}_1}-l_{u_1,\bar{x}_1})!}}{\sum_{(l_{0,0},l_{0,1},l_{1,0},l_{1,1})\in\mathcal{L}_{\epsilon,n}} \frac{l_{x_1}!}{l_{u_1,x_1}!(l_{x_1}-l_{u_1,x_1})!} \frac{l_{\bar{x}_1}!}{l_{u_1,\bar{x}_1}!(l_{\bar{x}_1}-l_{u_1,\bar{x}_1})!}} \\
&\overset{(a)}{=} \frac{\sum_{(l_{0,0},l_{0,1},l_{1,0},l_{1,1})\in\mathcal{L}_{\epsilon,n}} l_{u_1,x_1} \frac{1}{l_{u_1,x_1}!(l_{x_1}-l_{u_1,x_1})!} \frac{1}{l_{u_1,\bar{x}_1}!(l_{\bar{x}_1}-l_{u_1,\bar{x}_1})!}}{l_{x_1} \sum_{(l_{0,0},l_{0,1},l_{1,0},l_{1,1})\in\mathcal{L}_{\epsilon,n}} \frac{1}{l_{u_1,x_1}!(l_{x_1}-l_{u_1,x_1})!} \frac{1}{l_{u_1,\bar{x}_1}!(l_{\bar{x}_1}-l_{u_1,\bar{x}_1})!}} \\
&\overset{(b)}{\Rightarrow} \frac{P_{U,X}(u_1, x_1) - \epsilon}{P_X(x_1) + \epsilon} \leq P_{\mathcal{S}}(Q_i(X^m) = v | X^m = x^m) \leq \frac{P_{U,X}(u_1, x_1) + \epsilon}{P_X(x_1) - \epsilon} \\
&\Rightarrow \exists m, \epsilon > 0 : |P_{\mathcal{S}}(Q_i(X^m) = v | X^m = x^m) - P_{U|X}(u_1|x_1)| \leq \delta.
\end{aligned}$$

In (a), we use the fact that for fixed x^m, $(l_{x_1}, l_{\bar{x}_1})$ is fixed to simplify the numerators. In (b) we have used that for jointly typical ϵ-sequences (u^m, x^m), $l_{u_1,x_1} \in [n(P_{U,X}(u_1, x_1) - \epsilon), n(P_{U,X}(u_1, x_1) + \epsilon)]$, and $l_{x_1} \in [n(P_X(x_1) - \epsilon), n(P_X(x_1) + \epsilon)]$. □

We generalize these conditions to define what we call SLCS's:

Definition 33. *The coding scheme characterized by $P_{\underline{E}}$ is called an SLCS if its corresponding probability distribution satisfies the following constraints:*

1) $\forall x^n, \exists B_n(x^n)$ such that $\forall y^n \notin B_n(x^n), P(\underline{E}(x^n) = \underline{c}, \underline{E}(y^n) = \underline{c}') = P(\underline{E}(x^n) = \underline{c})P(\underline{E}(y^n) = \underline{c}')$, where $P^n_X(B_n(x^n)) \leq 2^{-n\delta_X}, \delta_X > 0$.

2) $\forall \delta > 0, \exists n \in \mathbb{N} : m > n \Rightarrow \forall x^m \in \{0,1\}^m, v \in \{0,1\}, |P_{\mathcal{S}}(Q_i(X^m) = v|X^m = x^m) - P_{\mathcal{S}}(Q_i(X^m) = v|X_i = x_i)| < \epsilon$.

3) $\forall \pi \in S_n : P(\underline{E}) = P(\underline{E}_\pi)$, where $\underline{E}_\pi(X^n) = \pi^{-1}(\underline{E}(\pi(X^n)))$.

Our goal is to analyze the correlation preserving properties of SLCS's. For a randomly generated encoding function $\underline{E} = (E_1, E_2, \cdots, E_n)$, denote the decomposition of the real function corresponding to the kth element into the form in Equation 5.4 as $\tilde{E}_k = \sum_{\mathbf{i}} \tilde{E}_{k,\mathbf{i}}, k \in [1, n]$. Let $\mathbf{P}_{j,\mathbf{i}}$ be the variance of $\tilde{E}_{k,\mathbf{i}}$. The next theorem states the main result of this section.

Theorem V.34. *For any $k \in [1, n], m \in \mathbb{N}, \gamma > 0$, $P_{\mathcal{S}}(\sum_{\mathbf{i}:N_{\mathbf{i}} \leq m, \mathbf{i} \neq \mathbf{i}_k} \mathbf{P}_{k,\mathbf{i}} \geq \gamma) \to 0$, as $n \to \infty$. Where, $\mathbf{i}_k$ is the kth standard basis element.*

Remark 35. *Theorem V.34 shows that SLCS's distribute most of the variance of $\tilde{E}_k$ on $\tilde{E}_{k,\mathbf{i}}$'s which operate on large blocks. Hence, the encoders generated using such schemes have high expected variance for decomposition elements with large effective lengths. This along with Theorem V.22 gives an upper bound on the correlation preserving properties of SLCS's.*

Proof. The following proposition shows that the probability $P_{\mathcal{S}}(\sum_{\mathbf{i}:N_{\mathbf{i}} \leq m, \mathbf{i} \neq \mathbf{i}_k} \mathbf{P}_{k,\mathbf{i}} \geq \gamma)$ is independent of the index k. This is due to property 1) in the Definition of SLCS's.

Proposition 36. *$P(\sum_{\mathbf{i}:N_{\mathbf{i}} \leq m, \mathbf{i} \neq 00\cdots01} \mathbf{P}_{k,\mathbf{i}} \geq \gamma)$ is constant in k.*

Proof. Fix $k, k' \in \mathbb{N}$. Define the permutation $\pi_{k \to k'} \in S_n$ as the permutation which switches the kth and k'th elements and fixes all other elements. Also, let $\mathcal{E}$ be the set of all mappings

$e : \{0,1\}^n \to \{0,1\}^n$.

$$P_{\mathcal{S}}(\sum_{\mathbf{i}:N_{\mathbf{i}}\le m,\mathbf{i}\neq\mathbf{i}_k}\mathbf{P}_{k,\mathbf{i}} > \gamma) = \sum_{\underline{e}\in\mathcal{E}} P_{\mathcal{S}}(\underline{e})\mathbb{1}(\sum_{\mathbf{i}:N_{\mathbf{i}}\le m,\mathbf{i}\neq\mathbf{i}_k}\mathbf{P}_{k,\mathbf{i}} > \gamma) \overset{(a)}{=} \sum_{\underline{e}\in\mathcal{E}} P_{\mathcal{S}}(\underline{e}_{\pi_{k\to k'}})\mathbb{1}(\sum_{\mathbf{i}:N_{\mathbf{i}}\le m,\mathbf{i}\neq\mathbf{i}_k}\mathbf{P}_{k,\mathbf{i}} > \gamma|\underline{e})$$

$$\overset{(b)}{=} \sum_{\underline{g}\in\mathcal{E}} P_{\mathcal{S}}(\underline{g})\mathbb{1}(\sum_{\mathbf{i}:N_{\mathbf{i}}\le m,\mathbf{i}\neq\mathbf{i}_k}\mathbf{P}_{\pi_{k\to k'}k,\pi_{k\to k'}\mathbf{i}} > \gamma|\underline{g}) = \sum_{\underline{g}\in\mathcal{E}} P_{\mathcal{S}}(\underline{g})\mathbb{1}(\sum_{\underline{l}:N_{\underline{l}}\le m,\underline{l}\neq\pi_{k\to k'}\mathbf{i}_k}\mathbf{P}_{k',\underline{l}} > \gamma|\underline{g})$$

$$= P_{\mathcal{S}}(\sum_{\mathbf{i}:N_{\mathbf{i}}\le m,\mathbf{i}\neq\mathbf{i}_{k'}}\mathbf{P}_{k',\mathbf{i}} > \gamma).$$

Where in (a) we have used property (2) 3) in Definition 5.5, and in (b) we have defined $\underline{g} \triangleq \underline{e}_{\pi_{k\to k'}}$ and used $\pi^2_{k\to k'} = 1$.

□

Using the previous proposition, it is enough to show the theorem holds for $k = 1$. For ease of notation we drop the subscript k for the rest of the proof and denote $\mathbf{P}_{1,\mathbf{i}}$ by $\mathbf{P}_{\mathbf{i}}$. By the Markov inequality, we have the following:

$$P_{\mathcal{S}}(\sum_{\mathbf{i}:N_{\mathbf{i}}\le m,\mathbf{i}\neq\mathbf{i}_1}\mathbf{P}_{\mathbf{i}} \ge \gamma) \le \frac{\sum_{\mathbf{i}:N_{\mathbf{i}}\le m,\mathbf{i}\neq\mathbf{i}_1}\mathbb{E}_{\mathcal{S}}(\mathbf{P}_{\mathbf{i}})}{\gamma}. \quad (5.12)$$

So, we need to show that $\sum_{\mathbf{i}:N_{\mathbf{i}}\le m,\mathbf{i}\neq\mathbf{i}_1}\mathbb{E}_{\mathcal{S}}(\mathbf{P}_{\mathbf{i}})$ goes to 0 for all fixed m. We first prove the following claim.

Claim 37. *Fix* $\mathbf{i}$*, such that* $N_{\mathbf{i}} \le m$*, the following holds:*

$$\mathbb{E}_{\tilde{E},X_{\mathbf{i}}}(\mathbb{E}^2_{X^n|X_{\mathbf{i}}}(\tilde{E}|X_{\mathbf{i}})) = \mathbb{E}_{X_{\mathbf{i}}}(\mathbb{E}^2_{\tilde{E},X^n|X_{\mathbf{i}}}(\tilde{E}|X_{\mathbf{i}})) + O(e^{-n\delta_X}).$$

Proof.

$$\mathbb{E}_{\tilde{E},X_{\mathbf{i}}}(\mathbb{E}^2_{X^n|X_{\mathbf{i}}}(\tilde{E}|X_{\mathbf{i}})) = \sum_{x_{\mathbf{i}},\tilde{e}} P(x_{\mathbf{i}})P(\tilde{e})(\sum_{x_{\sim\mathbf{i}}} P(x_{\sim\mathbf{i}})\tilde{e}(x^n))^2$$
$$= \sum_{x_{\mathbf{i}},\tilde{e}} P(x_{\mathbf{i}})P(\tilde{e}) \sum_{x_{\sim\mathbf{i}}} \sum_{y^n:y_{\mathbf{i}}=x_{\mathbf{i}}} P(x_{\sim\mathbf{i}})P(y_{\sim\mathbf{i}})\tilde{e}(x^n)\tilde{e}(y^n)$$
$$= \sum_{x^n} P(x^n) \sum_{y^n:y_{\mathbf{i}}=x_{\mathbf{i}}} P(y_{\sim\mathbf{i}})\mathbb{E}_{\tilde{E}}(\tilde{E}(x^n)\tilde{E}(y^n))$$
$$= \sum_{x^n} P(x^n) \sum_{y^n:y_{\mathbf{i}}=x_{\mathbf{i}},y^n\in B_n(x^n)} P(y_{\sim\mathbf{i}})\mathbb{E}_{\tilde{E}}(\tilde{E}(x^n)\tilde{E}(y^n)) + \sum_{x^n} P(x^n) \sum_{y^n:y_{\mathbf{i}}=x_{\mathbf{i}},y^n\notin B_n(x^n)} P(y_{\sim\mathbf{i}})\mathbb{E}_{\tilde{E}}(\tilde{E}(x^n)\tilde{E}(y^n))$$
$$\overset{(a)}{\leq} \sum_{x^n} P(x^n) \sum_{y^n:y_{\mathbf{i}}=x_{\mathbf{i}},y^n\in B_n(x^n)} P(y_{\sim\mathbf{i}}) + \sum_{x^n} P(x^n) \sum_{y^n:y_{\mathbf{i}}=x_{\mathbf{i}},y^n\notin B_n(x^n)} P(y_{\sim\mathbf{i}})\mathbb{E}_{\tilde{E}}(\tilde{E}(x^n)\tilde{E}(y^n))$$
$$= P(Y^n \in B_n(X^n)|Y_{\mathbf{i}} = X_{\mathbf{i}}) + \sum_{x^n} P(x^n) \sum_{y^n:y_{\mathbf{i}}=x_{\mathbf{i}},y^n\notin B_n(x^n)} P(y_{\sim\mathbf{i}})\mathbb{E}_{\tilde{E}}(\tilde{E}(x^n)\tilde{E}(y^n))$$
$$\overset{(b)}{=} O(e^{-n\delta_X}) + \sum_{x^n} P(x^n) \sum_{y^n:y_{\mathbf{i}}=x_{\mathbf{i}},y^n\notin B_n(x^n)} P(y_{\sim\mathbf{i}})\mathbb{E}_{\tilde{E}}(\tilde{E}(x^n))\mathbb{E}_{\tilde{E}}(\tilde{E}(y^n))$$
$$\leq O(e^{-n\delta_X}) + P(Y^n \in B_n(X^n)|Y_{\mathbf{i}} = X_{\mathbf{i}}) + \sum_{x_{\mathbf{i}}} P(x_{\mathbf{i}}) \sum_{x_{\sim\mathbf{i}}} \sum_{y^n:y_{\mathbf{i}}=x_{\mathbf{i}}} P(x_{\sim\mathbf{i}})P(y_{\sim\mathbf{i}})\mathbb{E}_{\tilde{E}}(\tilde{E}(x^n))\mathbb{E}_{\tilde{E}}(\tilde{E}(y^n))$$
$$= O(e^{-n\delta_X}) + \mathbb{E}_{X_{\mathbf{i}}}(\mathbb{E}^2_{\tilde{E},X^n|X_{\mathbf{i}}}(\tilde{E}|X_{\mathbf{i}})).$$

□

In (a) we use the fact that $\tilde{E} \leq 1$ by definition, in (b) follows from property 1) in Definition 5.5. Define $\bar{E}_{\mathbf{i}} = \mathbb{E}_{\tilde{E}}(\tilde{E}_{\mathbf{i}}) = \mathbb{E}_{\tilde{E}|X_{\mathbf{i}}}(\tilde{E}|X_{\mathbf{i}}) - \sum_{\mathbf{j}<\mathbf{i}} \bar{E}_{\mathbf{j}}$, and also define $\bar{P}_{\mathbf{i}} \triangleq Var(\bar{E}_{\mathbf{i}})$. Using the above claim we have:

$$P_{\mathcal{S}}(\sum_{\mathbf{i}:N_{\mathbf{i}}\leq m,\mathbf{i}\neq\mathbf{i}_1} \mathbf{P}_{\mathbf{i}} \geq \gamma) \leq \frac{\sum_{\mathbf{i}:N_{\mathbf{i}}\leq m,\mathbf{i}\neq\mathbf{i}_1} \mathbb{E}_{\mathcal{S}}(\mathbf{P}_{\mathbf{i}})}{\gamma} \leq \frac{2^m O(e^{-n\delta_X}) + \sum_{\mathbf{i}:N_{\mathbf{i}}\leq m} \mathbb{E}_{\mathcal{S}}(\bar{P}_{\mathbf{i}}) - \mathbb{E}_{\mathcal{S}}(\bar{P}_{\mathbf{i}_1})}{\gamma}. \quad (5.13)$$

Using the arguments from the proof of Proposition 16, we can see that the properties stated in that Proposition hold for $\bar{E}_{\mathbf{i}}$ as well. Using the same results as in Lemma 19 and

Corollary 20, we have that $\sum_{\mathbf{i}\in\{0,1\}^n} \bar{P}_{\mathbf{i}} = \bar{P}_{\underline{\mathbf{1}}}$. Following the calculations in (5.13):

$$\begin{aligned}
P_{\mathcal{S}}(\sum_{\mathbf{i}:N_{\mathbf{i}}\leq m, \mathbf{i}\neq \mathbf{i}_1} \mathbf{P}_{\mathbf{i}} \geq \gamma) &\leq \frac{2^m O(e^{-n\delta_X}) + \sum_{\mathbf{i}:N_{\mathbf{i}}\leq m} \mathbb{E}_{\mathcal{S}}(\bar{P}_{\mathbf{i}}) - \mathbb{E}_{\mathcal{S}}(\bar{P}_{\mathbf{i}_1})}{\gamma} \\
&\leq \frac{2^m O(e^{-n\delta_X}) + \sum_{\mathbf{i}\in\{0,1\}^n} \mathbb{E}_{\mathcal{S}}(\bar{P}_{\mathbf{i}}) - \mathbb{E}_{\mathcal{S}}(\bar{P}_{\mathbf{i}_1})}{\gamma} \\
&= \frac{2^m O(e^{-n\delta_X}) + \mathbb{E}_{\mathcal{S}}(\sum_{\mathbf{i}\in\{0,1\}^n} \bar{P}_{\mathbf{i}}) - \mathbb{E}_{\mathcal{S}}(\bar{P}_{\mathbf{i}_1})}{\gamma} \\
&= \frac{2^m O(e^{-n\delta_X}) + \mathbb{E}_{X^n}\left(\mathbb{E}^2_{\tilde{E}|X^n}(\tilde{E}(X^n)|X^n)\right) - \mathbb{E}_{\mathcal{S}}(\bar{P}_{\mathbf{i}_1})}{\gamma} \\
&\leq \frac{2^m O(e^{-n\delta_X}) + \mathbb{E}_{\mathcal{S}}(\bar{P}_{\mathbf{i}_1}) + O(\epsilon) - \mathbb{E}_{\mathcal{S}}(\bar{P}_{\mathbf{i}_1})}{\gamma} \\
&= \frac{2^m O(e^{-n\delta_X}) + O(\epsilon)}{\gamma}
\end{aligned}$$

Where in the last inequality we have used the second property in Definition 5.5. The last line goes to 0 as $n \to \infty$. This completes the proof.

The following Theorem provides a bound on the correlation preserving ability of SLCS's.

□

Theorem V.38. *Let (X, Y) be a pair of DMS's, with $P(X = Y) = 1 - \epsilon$. Also, assume that the pair of BBE's $\underline{E}, \underline{F}$ are produced using SLCS's. Define $E \triangleq E_1$, and $F \triangleq F_1$. Then,*

$$\forall \delta > 0 : P_{\mathcal{S}}\left(P_{X^n,Y^n}\left(E(X^n) \neq F(Y^n)\right) > 2\mathbf{P}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}} - 2(1-2\epsilon)\mathbf{P}^{\frac{1}{2}}_{\mathbf{i}_1}\mathbf{Q}^{\frac{1}{2}}_{\mathbf{i}_1} - \delta\right) \to 1,$$

as $n \to \infty$. Where $\mathbf{P_i} \triangleq Var(\tilde{E}_{\mathbf{i}})$, $\mathbf{Q_i} \triangleq Var(\tilde{F}_{\mathbf{i}})$, $\mathbf{P} \triangleq Var(\tilde{E})$, and $\mathbf{Q} \triangleq Var(\tilde{F})$.

Proof. From Theorem V.22, we have:

$$\mathbf{P}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}} - 2\sum_{\mathbf{i}} C_{\mathbf{i}}\mathbf{P}^{\frac{1}{2}}_{\mathbf{i}}\mathbf{Q}^{\frac{1}{2}}_{\mathbf{i}} \leq P(E(X^n) \neq F(Y^n)).$$

From Theorem V.34 we have:

$$\forall m \in \mathbb{N}, \gamma > 0, P_{\mathcal{S}}(\sum_{\mathbf{i}:N_{\mathbf{i}}\leq m, \mathbf{i}\neq\mathbf{i}_1} \mathbf{P}_{\mathbf{i}} < \gamma) \to 1, \qquad P_{\mathcal{S}}(\sum_{\mathbf{i}:N_{\mathbf{i}}\leq m, \mathbf{i}\neq\mathbf{i}_1} \mathbf{Q}_{\mathbf{i}} < \gamma) \to 1. \tag{5.14}$$

Note that:

$$\begin{aligned}
&\sum_{\mathbf{i}:N_{\mathbf{i}}\leq m, \mathbf{i}\neq\mathbf{i}_1} \mathbf{P}_{\mathbf{i}} < \gamma, \sum_{\mathbf{i}:N_{\mathbf{i}}\leq m, \mathbf{i}\neq\mathbf{i}_1} \mathbf{Q}_{\mathbf{i}} < \gamma \\
&\Rightarrow \sum_{\mathbf{i}} C_{\mathbf{i}} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}} > (1-2\epsilon)(\mathbf{P}_{\mathbf{i}_1}+\gamma)^{\frac{1}{2}}(\mathbf{Q}_{\mathbf{i}_1}+\gamma)^{\frac{1}{2}} + (1-2\epsilon)^m \mathbf{P}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}},
\end{aligned} \tag{5.15}$$

which converges to $(1-2\epsilon)\mathbf{P}_{\mathbf{i}_1}^{\frac{1}{2}}\mathbf{Q}_{\mathbf{i}_1}^{\frac{1}{2}} + (1-2\epsilon)^m \mathbf{P}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}}$ as $\gamma \to 0$. Also $C_{\mathbf{i}}$ is decreasing in $N_{\mathbf{i}}$ and goes to 0 as $N_{\mathbf{i}} \to \infty$. Choose γ small enough and m large enough such that $(1-2\epsilon)(\mathbf{P}_{\mathbf{i}_1}+\gamma)^{\frac{1}{2}}(\mathbf{Q}_{\mathbf{i}_1}+\gamma)^{\frac{1}{2}} + (1-2\epsilon)^m \mathbf{P}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}} - (1-2\epsilon)\mathbf{P}_{\mathbf{i}_1}^{\frac{1}{2}}\mathbf{Q}_{\mathbf{i}_1}^{\frac{1}{2}} < \delta$. Then Equations (5.14) and (5.15) gives

$$P_{\mathcal{S}}\left(P_{X^n,Y^n}(E(X^n) \neq F(Y^n)) < 2\mathbf{P}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}} - 2(1-2\epsilon)\mathbf{P}_{\mathbf{i}_1}^{\frac{1}{2}}\mathbf{Q}_{\mathbf{i}_1}^{\frac{1}{2}} - \delta\right) \to 0.$$

This is equivalent to the statement of the theorem. □

Remark 39. *The result in the theorem holds even if the encoders use the same encoding function (i.e. $E = F$) produced using a SLCS.*

Remark 40. *The previous theorem gives a bound on the correlation preserving properties on SLCS's. The theorem shows that in order to increase correlation in these schemes the encoder needs to put more variance on the element $\tilde{E}_{k,\mathbf{i}_k}, k \in [1,n]$. This would require more correlation between the input and output of the encoder which itself would require more rate. As an example consider the extreme case where $Var(\tilde{E}_k) = Var(\tilde{E}_{k,\mathbf{i}_k})$, which requires $E_k(X^n) = X_k$. This means that in order to achieve maximum correlation, the encoder must use uncoded transmission.*

Remark 41. *In the case when $X = Y$, there is common-information [11] available at the encoders. If the encoders use the same encoding function E, their outputs would be equal. Whereas from theorem V.38, for any non-zero ϵ, the output correlation is bounded away from 0 (except when doing uncoded transmission). So, the correlation between the outputs of SLCS's is discontinuous as a function of ϵ.*

CHAPTER VI

Application of Finite Block-length Codes to Distributed Source Coding

In the classic lossy distributed source coding problem, two distributed encoders observe the outputs of two correlated sources and communicate a compressed version of the source sequences to a joint decoder. The decoder then wishes to produce a lossy reconstruction of the two sources. The suitability of this reconstruction is gauged by the means of two separable distortion criteria, one for each source. This scenario is depicted in Figure 6.1. The goal is to characterize the optimal rate-distortion trade-off. The problem of deriving the optimal achievable rate-distortion region has remained open for several decades. The main challenge is in devising a scheme which optimally utilizes the correlation between the sources without requiring the encoders to communicate with each other.

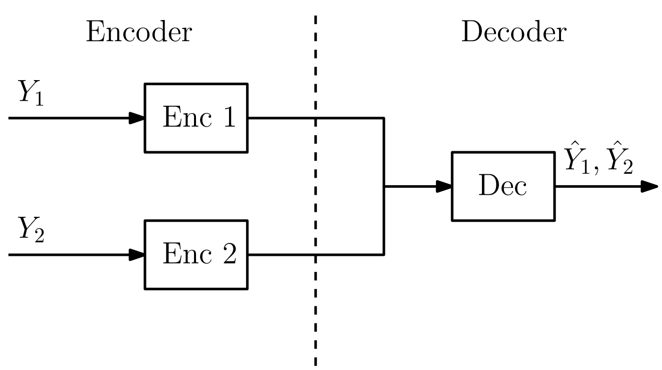

Figure 6.1: General Lossy Distributed Source Coding

Prior to this work, the best known inner bound to the optimal rate-distortion (RD) region was the Berger-Tung (BT) bound [45]. The BT bound is based on a coding strategy called quantize and bin. In this strategy the two sources are quantized using two independent, random vector quantizers. In this scheme the length of the quantizers approaches infinity. The outputs of these quantizers are binned to reduce the transmission rate. The independent quantization approach leads to the so-called long Markov chain. The Markov chain implies that conditioned on the sources, the single-letter distribution of the quantized versions of the sources decompose into the product of conditional marginal distributions. In [50] it was pointed out that in the presence of common components, further correlation can be induced between the quantized versions, in other words the Markov chain can be relaxed using the common component (CC). Based on this observation the authors in [50] propose a coding scheme which outperforms the BT strategy in the presence of common components, but reduces to the latter in their absence. The CC achievable RD region shrinks discontinuously in source probability distribution as common components are replaced with highly correlated components. Using the continuity of the optimal RD region, it was proved in [50] that the CC scheme is also sub-optimal since it is discontinuous. Hence the optimal RD region strictly contains the CC region. However, it was not clear how to achieve points outside of the CC rate-distortion region.

In the first part of the chapter, we consider the one-help-one problem. The binary one-help-one problem is a distributed source coding problem, where the decoder is to reconstruct a compressed version of the input to one of the encoders. The second encoder - called the helper- receives a noisy version of the source and facilitates the transmission of the compressed version to the decoder. In this setup, the two sources which are available to the two encoders have highly correlated components but no common components. We notice that in the absence of exact common components, the scheme presented in [50] reduces to the BT strategy. Even though sources have highly correlated components, the correlation is lost in the quantization step as shown in the previous chapter. In this simple case, we are able to provide a more compact argument than the one in the last chapter, and show that no matter how highly correlated the two components are, the quantization noises of the two sources approach two independent random vectors. However, if the blocklength is kept finite, the quantization noises remain correlated. This allows the encoders to bin their outputs more efficiently. Based on this, we introduce a new scheme which utilizes finite length quantizers in the first stage (quantizing the highly correlated variables) and then uses large blocklength quantizers in the second stage. An interesting implication is that in order to get gains in terms of the achievable rate, the length of the first quantization stage cannot be too small or too high; meaning that for some finite length, the scheme achieves its best performance.

Since we use finite-length quantizers, a characterization for the finite-length performance of codes is needed. The exact characterization of the rate-distortion region as a function of the quantization blocklength is unknown even in the binary case; however, several upper bounds are provided in the literature for finite-length quantization rate with a constant distortion [19]. Using these results, we show that the method presented in this chapter achieves a better rate-distortion region than other known results.

The main difficulty in analyzing finite blocklength coding strategies is that in the absence of simplifying theorems such as laws of large numbers, the resulting characterizations of achievable inner bounds are in terms of multi-letter probability distributions. This makes the computation of such inner bounds very complex. In the second part, we first present a coding scheme for the distributed source coding problem in the general discrete, memoryless setting. The scheme utilizes both small-length codes and codes with blocklength approaching infinity. A multi-letter characterization of the rate-distortion region achievable using this scheme is given, and it is shown that there is an approximating single-letter characterization for the inner bound. We prove that the resulting inner bound outperforms the previous known coding schemes for this communication setting.

6.1 Preliminaries

In this section, we present a formal statement of the lossy distributed source coding problem, then we restrict this to the binary one-help-one problem which is the main example discussed in this chapter. For the rest of the chapter, a sequence of length n is denoted by $x(1:n)$, its i^{th} element is denoted by $x(i)$, and the subsequence consisting of its i^{th} element to its j^{th} element is shown by $x(i:j)$. A two dimensional matrix of size $m \times n$ is denoted by $x(1:m, 1:n)$. Random variables are shown by capital letters and their realizations are denoted by small letters. Let $\{Y_1(i)\}$ and $\{Y_2(i)\}$ be two source sequences from the alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ for the sources shown in Figure 6.1.

Let the sources be $i.i.d$ samples of a joint PMF on $\mathcal{Y}_1 \times \mathcal{Y}_2$ given by $P_{Y_1,Y_2}(y_1, y_2)$. Let the functions $d_i : \mathcal{Y}_i \times \mathcal{Y}_i \to R_{\geq 0}, i = 1, 2$ be the distortion criteria for the sources. Without loss of generality, the reconstruction alphabets are assumed to be the same as source alphabets. A $(2^{nR_1}, 2^{nR_2}, n)$ code consists of: 1) Two encoding functions $m_i : \mathcal{Y}_i^n \to [1 : 2^{nR_i}], i = 1, 2$, and 2) A decoding function $r : [1 : 2^{nR_1}] \times [1 : 2^{nR_2}] \to \mathcal{Y}_1^n \times \mathcal{Y}_2^n$. Let $\hat{Y}_i(1 : n) = r(m_1(Y_1(1 : n)), m_2(Y_2(1 : n)))$ be the reconstruction of the two sources. A quadruple (R_1, R_2, D_1, D_2) is said to be achievable if there exists a sequence of codes $(2^{nR_1}, 2^{nR_2}, n)$ such that $\limsup_{n\to\infty} E(d(Y_i, \hat{Y}_i)) \leq D_i, i = 1, 2$.

6.2 Binary One Help One Example

The example is shown in Figure 6.2. X, Z and E are Bernoulli random variables with $P(X = 1) = 0.5, P(Z = 1) = p, P(E = 1) = \epsilon$ where $p, \epsilon \in [0, 0.5)$. All random variables are assumed to be independent. As shown in the figure $Y_1 = X + E$ and $Y_2 = (X, Z)$, also $d_1(y_1, \hat{y}_1) = 0, y_1, \hat{y}_1 \in \{0, 1\}$ and $d_2(y_2, \hat{y}_2) = d_H(\hat{x} + \hat{z}, x + z)$ where $\hat{y}_2 = (\hat{x}, \hat{z}), y_2 = (x, z)$ and d_H represents the Hamming distance. Since the distortion function for Y_1 is 0, encoder 1 is only helping encoder 2 transmit its compression. The following gives an achievable RD vector for this problem.

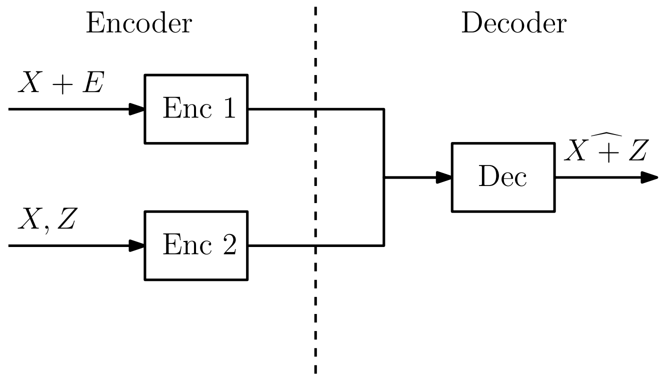

Figure 6.2: The binary one help one example

Lemma 1 ([50]). *For $\epsilon = 0$, the following RD quadruples are achievable using the CC scheme.However, the RD vector is not achievable using the BT scheme. scheme in [45]:*

$$(r_1, r_2, d_1, d_2) = (1 - h_b(\delta), h_b(p * \delta) - h_b(\delta_1), 0, \delta_1) \tag{6.1}$$

When $\epsilon \neq 0$, it was shown that (6.1) is not achievable. In this case, we prove that our finite-length scheme achieves a larger RD region the previous strategies.

6.2.1 Finite Length Quantizer Scheme

In this section, we introduce a new finite block length coding scheme. The rest of this section is dedicated to proving the following theorem:

Theorem VI.2. *The following rate-distortion region is achievable for any positive integer n.*

$$R_1 \geq 1 - h_b(\delta) + \theta_n \tag{6.2}$$

$$R_2 \geq h_b(p * \delta) - h_b(\delta_1) \tag{6.3}$$

$$D_2 \leq \delta_1 * ((1 - (1 - \epsilon)^n)(\delta + \frac{\epsilon}{(1 - (1 - \epsilon)^n)} * \delta)) \tag{6.4}$$

where $p * \delta \geq \delta_1$, $a * b = a(b-1) + b(a-1)$, and $\{\theta_n\}$ is defined in Section 4.

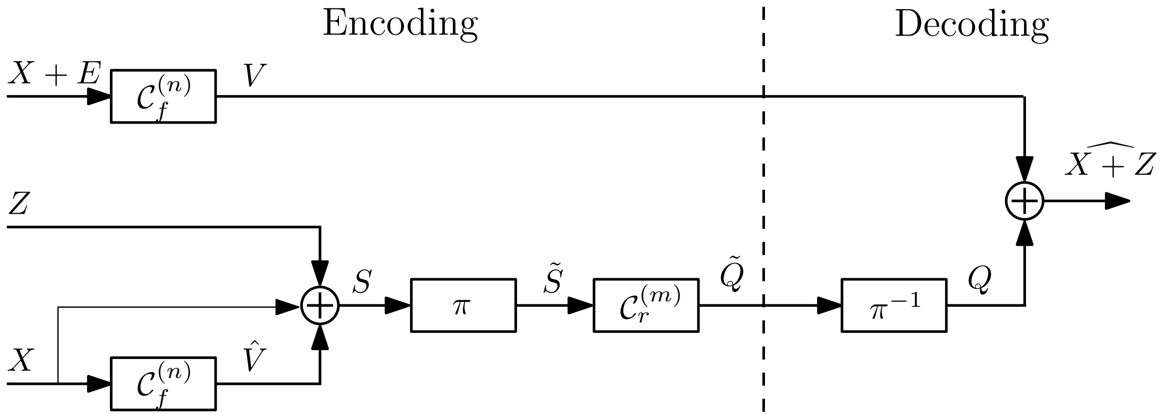

Figure 6.3: A block-diagram of the scheme

Remark 3. *The above bound is continuous in ϵ and simplifies to the one given in [50] when $\epsilon = 0$. This proves that it strictly contains the BT rate-distortion region.*

Proof.

Lemma 4 ([19])**.** *Consider the PtP problem of quantizing a BSC to Hamming distortion δ using an n-length quantizer. The following rate is achievable:*

$$R(n, \delta) = R(\delta) + \frac{1}{2}\frac{\log n}{n} + O(\frac{1}{n})$$

where $R(\delta) = 1 - h_b(\delta)$ is the binary rate distortion function. Define $\theta_n = \frac{1}{2}\frac{\log n}{n} + O(\frac{1}{n})$. Note that $\{\theta_n\}$ is a sequence of positive numbers converging to 0.

To achieve (6.4) for some fixed n, we use two quantization codes. The first code $C_f^{(n)}$ is a finite length quantizer for a binary symmetric source (BSS) with codewords of length n and average distortion δ and rate $R_f^{(n)} = 1 - h_b(\delta) + \theta_n$. The existence of such codes and bounds on θ_n are discussed in [19]. The second code $C_r^{(m)}$has codewords of length m, and it is chosen from a family of codes suitable for quantization of a Bernoulli source with parameter $p*\delta$ with average distortion δ_1. The rate of the code converges to $h_b(p*\delta) - h_b(\delta_1)$ as m tends to infinity. The existence of such codes is given by Shannon's rate distortion theorem in [39]. Let Π be the set of permutations on the set $[1:n]$. We choose permutations

$\pi_i, i \in [1:m]$ randomly and uniformly from Π. These permutations are also made available to the second encoder and the decoder beforehand.

First we give a summary of the scheme and then we present the formal scheme and a proof of achievability. As shown in Figure 6.3, the first encoder uses $C_f^{(n)}$ to quantize a block of length n of its input source $X + E$. It then transmits the quantized version to the decoder. The second encoder guesses the quantized codeword sent by the first encoder by quantizing the block of length n of source X into $\hat{V}$. The estimation is precise with high probability if $\epsilon \ll \frac{1}{n}$. This is true since the expected number of bits where E is 1 in one block of length n is $n\epsilon$ which is small under this condition. Using $\hat{V}$ the encoder calculates the quantization noise $X + \hat{V}$. The quantization noise is correlated with $X + E + V$, the quantization noise in the other encoder. Note that in the case when $\epsilon = 0$, $\hat{V} = V$. So, $X + \hat{V}$ completely captures the uncertainty at the decoder. The second encoder sends a quantized version of $X + \hat{V} + Z$ to the decoder to refine the description it received from the first encoder. When $\epsilon \neq 0$ we make the second encoder use $X + \hat{V}$ as an approximation to $X + V$ and add it to Z. Let us denote the resulting random variable as S. This process is repeated for m blocks each of length n. The encoder first uses the permutations $\pi_i, i \in [1 : m]$ described above to transform S into an i.i.d Bernoulli source $\tilde{S}$ with parameter $p * \delta$. This will be explained in more detail in the next paragraphs. Finally, this new i.i.d source is quantized to distortion δ_1 and sent to the decoder. Let the quantized version of $\tilde{S}$ be denoted by $\tilde{Q}$. The decoder having received $\tilde{Q}$, calculates Q which is a quantized version of S. Finally the decoder declares $Q + V$ as the reconstruction of $X + Z$.

Now we proceed to formally present the scheme and prove the theorem. The first encoder receives a string of nm bits of $Y_1 = X + E$. The encoder breaks this vector into m blocks of length n. We denote each bit in this string by $X(i, j) + E(i, j), i \in [1:m], j \in [1:n]$ where i indicates the block containing the bit and j indicates the index of the bit in the block. The encoder uses $C_f^{(n)}$ to quantize each block of length n. So, for $i = 1, ..., m$, it finds

$V(i, 1:n)$ such that:

$$V(i, 1:n) = argmin_{v(1:n)\in C_f^{(n)}}\{d_H(Y_1(i, 1:n), v(1:n))\}.$$

The index of $V(i, 1 : n)$ is then transmitted to the decoder. The rate of transmission for this encoder is $R_1 = R_f^{(n)} = 1 - h_b(\delta) + \theta_n$.

Next we explain the mechanism used in the second encoder. The encoder receives nm bits from sources X and Z. It divides them into m blocks of length n as is the previous case. It quantizes each block of $X(i, 1 : n)$ in the same manner as in the first encoder. Let $\hat{V}(i, 1 : n)$ be the quantized codeword corresponding to $X(i, 1 : n)$. The encoder computes $S(1 : m, 1 : n) = X(1 : m, 1 : n) + \hat{V}(1 : m, 1 : n) + Z(1 : m, 1 : n)$. Let $\tilde{S}(i, j) = S(i, \pi_i(j)), i \in [1:m], j \in [1:n]$, a permuted version of S. The next lemma proves that the result is an i.i.d source.

Lemma 5. *$\tilde{S}(1:m, j)$ is a string of i.i.d Bernoulli random variables with parameter $p * \delta$.*

The proof of the lemma is given in the appendix. The encoder quantizes each $\tilde{S}(1:m, j)$ using the code $C_r^{(m)}$. Let $\tilde{Q}(1 : m, j)$, be the quantized version of $\tilde{S}(1 : m, j)$. The encoder transmits the index of $\tilde{Q}(1 : m, j)$ in $C_r^{(m)}$ to the decoder. Let $\tilde{T}(1 : m, 1 : n) = \tilde{Q}(1 : m, 1 : n)+\tilde{S}(1:m, 1:n)$ be the quantization noise. We know that this noise becomes *i.i.d.* Bernoulli with parameter δ_1 and $\tilde{T}(i, j)$ is independent of $\tilde{S}(i, j)$ as m tends to infinity. Also, define $T(i, j) = \tilde{T}(i, \pi_i^{-1}(j)), i \in [1 : m], j \in [1 : n]$. The rate of transmission in the second encoder is $R_2 = R_r^{(m)}$ which approaches $h_b(p * \delta) - h_b(\delta_1)$ as m goes to infinity.

The decoder computes $Q(i, j) = \tilde{Q}(i, \pi_i^{-1}(j))$, that is the decoder undoes the permutation. Note that $E(d_H(Q(i, j), S(i, j))) = E(d_H(\tilde{Q}(i, j), \tilde{S}(i, j))) = E(w_H(T(i, j))) = \delta_1$. The decoder declares $Q(1 : m, 1 : n) + V(1 : m, 1 : n)$ as the reconstruction of the source $X + Z$. The resulting average distortion is:

$$D = \frac{1}{mn}E\{d_H((X + Z)(1:m, 1:n), (Q + V)(1:m, 1:n))\}.$$

This can be computed as follows:

$$E\{d_H((X+Z)(1:m,1:n),(Q+V)(1:m,1:n))\} = E\{w_H((X+Z+S+T+V)(1:m,1:n))\}$$
$$= E\{w_H((\hat{V}+V+T)(1:m,1:n))\} \stackrel{a}{=} mn(\delta_1 * \frac{1}{mn}E\{w_H((\hat{V}+V)(1:m,1:n))\}).$$

Now we calculate $E\{w_H((\hat{V}+V)(1:m,1:n))\}$:

$$\sum_{i=1}^{m} E\{w_H((\hat{V}+V)(i,1:n))\}$$
$$\stackrel{b}{=} m(E\{w_H((\hat{V}+V)(1:n)|E(1:n)=0)P(E(1:n)=0)\}$$
$$+ E\{w_H((\hat{V}+V)(1:n)|E(1:n)\neq 0)P(E(1:n)\neq 0))\})$$
$$\stackrel{c}{=} m(E\{w_H((\hat{V}+V)(1:n)|E(1:n)\neq 0)P(E(1:n)\neq 0))$$
$$= m(1-(1-\epsilon)^n)E\{w_H((\hat{V}+V)(1:n))|E(1:n)\neq 0\}$$
$$= m(1-(1-\epsilon)^n)E\{w_H((X+\hat{V}+X+V)(1:n))|E(1:n)\neq 0\}$$
$$\leq m(1-(1-\epsilon)^n)(E\{w_H((X+\hat{V})(1:n))|E(1:n)\neq 0\}$$
$$+ E\{w_H((X+E+V+E)(1:n))|E(1:n)\neq 0\})$$
$$\stackrel{d}{=} m(1-(1-\epsilon)^n)(E\{w_H((X+\hat{V})(1:n))\}$$
$$+ \frac{\epsilon}{(1-(1-\epsilon)^n)} * E\{w_H((X+E+V)(1:n))\})$$
$$\stackrel{e}{=} mn(1-(1-\epsilon)^n)(\delta + \frac{\epsilon}{(1-(1-\epsilon)^n)} * \delta)$$
$$\to D \leq \delta_1 * ((1-(1-\epsilon)^n)(\delta + \frac{\epsilon}{(1-(1-\epsilon)^n)} * \delta)).$$

(a) holds since T becomes independent of all the other variables as m tends to infinity; (b) holds since each block is quantized identically and hence the expected value is equal for all blocks; (c) is correct since if $E(1:n)=0$ then $V(1:N)=\hat{V}(1:N)$ since they are both quantized versions of $X(1:n)$; (d) holds since $(X+E+V)(1:n)$ is a function of $(X+E)(1:n)$ also $(X+E)(1:n)$ is independent of $E(1:n)$ since X is Bernoulli with parameter 0.5, and finally (e) holds since the average distortion of the finite length quantizer was assumed to be δ. This completes the proof of theorem VI.2. $\square$

6.2.2 Comparison with the Common Component Scheme

Proposition 6. *The scheme presented here achieves a larger rate distortion region than the one presented in [50].*

Proof. We shall prove there exists p and ϵ such that the rate-distortion region in theorem VI.2 strictly contains the rate-distortion region in [50]. It was shown in [50] that when $\epsilon = 0$ the Berger-Tung bound does not include the set of quadruples $(r_1, r_2, d_1, d_2) = (1 - h_b(\delta), h_b(p * \delta) - h_b(\delta_1), 0, \delta_1)$ when $\delta \in (0, 0.5)$ and $\delta_1 < p * \delta$. Also it is stated that the rate region in [50] reduces to the standard Berger-Tung bound for $\epsilon \neq 0$ since there is no common component between Y_1 and Y_2 in that case. Since the Berger-Tung scheme must perform worse when $\epsilon \neq 0$ as compared to the case when $\epsilon = 0$, we infer that it cannot achieve $(1 - h_b(\delta), h_b(p * \delta) - h_b(\delta_1), 0, \delta_1)$ when $\epsilon \neq 0$. This means that for a given δ and δ_1 there exists a radius $\gamma > 0$ for which no quadruple in the set $B((r_1, r_2, d_1, d_2), \gamma) = \{(R_1, R_2, 0, D_2) : d_E((R_1, R_2, D_2), (r_1, r_2, d_2)) \leq \gamma\}$ is achievable by the scheme in [50]. Note that d_E is just the Euclidean distance in the three dimensional space. Note that for a given ϵ and n we showed that $(r'_1, r'_2, 0, d'_2) = (1 - h_b(\delta) + \theta_n, h_b(p * \delta) - h_b(\delta_1), 0, \delta_1 * ((1 - (1 - \epsilon)^n)(\delta + \frac{\epsilon}{(1-(1-\epsilon)^n)} * \delta))$ is achievable by our scheme. We have:

$$d_E((r'_1, r'_2, d'_2), (r_1, r_2, d_2)) = \sqrt{\theta_n^2 + (\delta_1 * ((1 - (1 - \epsilon)^n)(\delta + \frac{\epsilon}{(1 - (1 - \epsilon)^n)}\delta)) - \delta_1)^2} \tag{6.5}$$

Since θ_n is converging to 0, one can take n to be large enough so that θ_n is less than $\frac{\gamma}{2}$. Since (6.5) is a continuous function of ϵ which is less than $\frac{\gamma}{2}$ as ϵ goes to 0, there exists non-negative ϵ for which $(r'_1, r'_2, 0, d'_2) \in B((r_1, r_2, 0, d_2)$ for n described as above. Hence the point $(r'_1, r'_2, 0, d'_2)$ is achievable by the scheme purposed here while it is not achievable by [50]. This shows that the rate-distortion region in theorem VI.2 strictly contains the one in [50] for non-zero ϵ. □

6.3 Large blocklength Quantization of Binary Variables

In this section, we show that minimum distance quantization of two highly correlated BSS's using linear codes results in quantization noises that behave similar to independent random variables. We prove the result for linear codes since the congruence of the Voronoi regions of a linear code facilitates our analysis. This is a special case of the results proved in the previous chapter. Consider two strings of binary random variables $\{x_i\}$ and $\{x_i + e_i\}$ generated by sources X and $X + E$ described in section 4. Choose a family of linear codes $\{C_{\mathcal{G}}^{(n)}\}$ randomly and uniformly with blocklength n and rate $\frac{1}{n}\lfloor(1 - h_b(\delta))\rfloor$. It is well known that as n approaches infinity the average distortion of a this linear coding scheme approaches δ. We show that as n goes to infinity for any fixed ϵ, the average Hamming distance between the quantization noises approaches the ϵ-vicinity of $\delta * \delta * \epsilon$. Let $Q(x(1:n)) = argmin_{c \in C_{\mathcal{G}}^{n}}(d_H(x(1:n), c))$ be the quantized version of $x(1:n)$ and let $s(1:n) = Q(x(1:n)) + x(1:n)$ and $t(1:n) = Q((x+e)(1:n)) + (x+e)(1:n)$ be the quantization noises. Let the Voronoi region for the 0 codeword be P_0. It is relatively straightforward to show that:

$$
\begin{aligned}
&p(S(1:n) = s(1:n), T(1:n) = t(1:n)) \rightarrow \\
&\frac{1}{2^{nh_b(\delta)}} P((s + t + E)(1:n) \in C_{\mathcal{G}}^{(n)}), (s,t)(1:n) \in P_0 \times P_0,
\end{aligned}
\tag{6.6}
$$

as n goes to infinity. The proof is omitted because of space limitations. Then we have:

$$
E(w_H((S+T)(1:n))) \rightarrow= \frac{1}{2^{nh_b(\delta)}} \sum_{(s,t) \in P_0 \times P_0, c \in C_{\mathcal{G}}^{(n)}} w_H((s+t)(1:n)) P(E(1:n) = (s+t+c)(1:n)) \tag{6.7}
$$

Define $u = t+c$, then $Q(u(1:n)) = c(1:n)$ and $(u + Q(u))(1:n) = t(1:n)$. (6.7) becomes:

$$\frac{1}{2^{nh_b(\delta)}} \sum_{t\in P_0, u(1:n)\in\{0,1\}^n} w_H((s+u)(1:n) + Q(u(1:n)))P(E(1:n) = (s+u)(1:n)).$$

It can be shown that Equations (6.6) and (6.7) yield:

$$\Big|\frac{1}{2^{nh_b(\delta)}} \sum_{t\in P_0, u(1:n)\in\{0,1\}^n} w_H(Q((s+u)(1:n)))P(E(1:n)=u(1:n)) - E(w_H((S+T)(1:n)))\Big| \leq n\epsilon \tag{6.8}$$

Using (6.6) and (6.8) we get:

$$\Big|E(w_H(Q((S+E)(1:n))))-E(w_H((S+T)(1:n)))\Big| \leq n\epsilon.$$

$(S+E)(1:n)$ has average weight $n(\delta * \epsilon)$ as n goes to infinity. The probability distribution of $Q((S+E)(1:n))$ is given by $P(Q(s+e)(1:n) = q(1:n)) \to \frac{1}{2^{nh_b(\delta)}}, q(1:n) \in B(s+e, n\delta)$ as $n \to \infty$. Taking the average weight of $Q((S+E)(1:n))$ gives $\delta * \delta * \epsilon$. Therefor, the scheme in [50] would give quantization noise in ϵ vicinity of $\delta_1 * \delta * \delta * \epsilon$ which is worse than what finite length quantizers would achieve. Also, the average Hamming distance between the quantization noises is not continuous in ϵ. If ϵ is 0, the distance would be 0, since both quantizers are quantizing the exact same sequence. However, if $\epsilon \neq 0$ then the distance is bigger than $\delta * \delta$.

6.4 Simulations for Hamming codes

In this section, we present our results for the case where the first encoder uses a Hamming code as its finite blocklength quantizer. Hamming codes are perfect codes of blocklength $2^r - 1$ and rate $1 - \frac{r}{2^r-1}$. They have minimum distance of 3. Using (6.4), one can compare the performance of the scheme presented here for $\epsilon \neq 0$ with (6.1). As stated

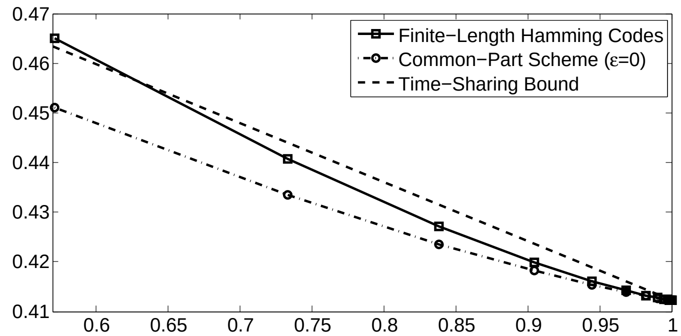

Figure 6.4: Comparison of the new scheme using Hamming codes and [50], $\delta_1 = 0.1, p = 0.3, \epsilon = 10^{-10}$

before, (6.1) contains the rate-distortion region of the binary one-help-one problem for any ϵ. Figure 6.4 shows the two bounds along with the time-sharing bound which is described next. One strategy in this setting is for the first encoder to transmit $X + E$ losslessly and the other encoder to send a quantized version of Z and for the decoder to add them together. Another strategy is for the second encoder to quantize $X + Z$ and transmit it while the first encoder does not send anything. The third bound in Figure 6.4 illustrates the bound resulting from time-sharing between these two strategies. This time-sharing strategy seems to be a good strategy for the Berger-Tung approach since if we use independent quantization for two encoders simultaneously the quantizations noises will add to each other and we will get a worse distortion than the time-sharing strategy. It can be seen from the plot that when we use finite blocklength Hamming codes for quantization, we achieve better results than the time-sharing bound.

6.5 The New Coding Strategy

In this section, we derive a new coding strategy for the general DSC based on finite blocklength codes. Let X_1 and X_2 be two correlated DMS's. Assume there exist functions $f_1 : \mathcal{X}_1 \to \mathcal{Z}$ and $f_2 : \mathcal{X}_2 \to \mathcal{Z}$, such that $P(f_1(X_1) \neq f_2(X_2)) \leq \epsilon', \epsilon' \in (0, 0.5)$. Here $\mathcal{X}_i$ and $\mathcal{Z}$ are the underlying alphabets for X_i and Z. Let $\epsilon = 1 - (1 - \epsilon')^n$. Also define $S_i = f_i(X_i), i \in \{1, 2\}$. The next theorem presents the main result of this chapter:

Theorem VI.7. *The following RD vectors are achievable:*

$$R_1 \geq I(X_1; U_1|U_2W) + E_{n,\epsilon} + 2|\mathcal{X}_1||\mathcal{U}_1| \log(\frac{p_1}{p_1 - \epsilon}),$$

$$R_2 \geq I(X_2; U_2|U_1W) + 2E_{n,\epsilon} + 2|\mathcal{X}_2||\mathcal{U}_2| \log(\frac{p_2}{p_2 - \epsilon}),$$

$$R_1 + R_2 \geq I(X_1X_2; U_1U_2W) + 3E_{n,\epsilon} + 2|\mathcal{X}_1||\mathcal{U}_1| \log(\frac{p_1}{p_1 - \epsilon}) + \theta_n,$$

$$D_i \geq E\{d_i(h_i(U_1, U_2, W), X_i)\}.$$

For every distribution P_{X_1,X_2,W,U_1,U_2} satisfying the following constraints:

$$1) U_1 \leftrightarrow (X_1, W) \leftrightarrow (X_2, W) \leftrightarrow U_2,$$

$$2) W \leftrightarrow S_1 \leftrightarrow (X_1, X_2),$$

$$3) p_i > \epsilon, i = 1, 2,$$

Also p_i and $E_{n,\epsilon}$ are defined as follows:

$$p_1 = \min_{x_1,w,u_1,u_2} (\{P_{U_1|X_1,W,U_2}\}, \{P_{U_1|W,U_2}\}),$$

$$p_2 = \min_{x_2,w,u_1,u_2} (\{P_{U_2|X_2,W,U_1}\}, \{P_{U_2|W,U_1}\}),$$

$$E_{n,\epsilon} = \frac{h(\epsilon)}{n} + \epsilon \log |\mathcal{W}|.$$

In the above formulas, θ_n is a sequence approaching 0 which depends on $P_{S,W}$. Also $\mathcal{U}_i$ and $\mathcal{W}$ are the alphabets for U_i and W.

Remark 8. *The above RD region reduces to the CC region when $\epsilon = 0$ (i.e when $S_1 = S_2$). Also if S_1 and S_2 are taken to be trivial, the bound reduces to the BT region.*

Remark 9. *As n becomes larger, ϵ increases, which in turn causes $E_{n,\epsilon}$ to increase. On the other hand, θ_n is a decreasing function of n. This illustrates the trade-off between rate-loss due to application of small blocklength codes θ_n, and the gains from preservation of* *correlation between the sources $E_{n,\epsilon}$.*

Remark 10. *Finding the achievable rate-distortion region involves sweeping over all possible choices of f_1 and f_2. However $\log(\frac{p_1}{p_1-\epsilon})$ increases as f_1 and f_2 become less correlated (i.e. as ϵ increases). This suggests choosing highly correlated functions gives larger achievable regions.*

Remark 11. *The above inner bound is not symmetric with respect to the two encoders, one can symmetrize the region by swapping the indices for encoders* 1 *and* 2 *in the theorem and taking the union of the two resulting regions.*

To prove theorem VI.7, an achievable RD region using finite-length coding schemes is derived, then it is shown that the region contains the inner bound in the theorem.

Let W_1 be a random variable with alphabet $\mathcal{W}_1$. Take an arbitrary probability distribution P_{S_1,W_1}. Also let I be a random variable uniformly distributed on $\{1, 2, 3, ..., n\}$. Consider Q_n, an n-length quantizer which quantizes S_1^n to W_1^n such that:

$$P_{S_1(I),W_1(I)} = \frac{1}{n} \sum_{i\in[1:n]} P_{S_1(i),W_1(i)} = P_{S_1,W_1}.$$

There exists a Q_n with rate $R_n = I_{P_{S_1,W_1}}(S_1; W_1) + \theta_n$, where θ_n can be bounded given the distribution P_{S_1,W_1} and approaches 0 as n goes to infinity [19]. Let $W_2^n = Q_n(S_2^n)$. W_2^n can be perceived as the second encoder's "estimate" of W_1^n. We have:

$$P_{X_1^n,X_2^n,W_1^n,W_2^n} = \sum_{s_1^n,s_2^n} P_{X_1^n,X_2^n|S_1^n,S_2^n} P_{S_1^n,S_2^n,W_1^n,W_2^n}.$$

Define $P_{X_1,X_2,W_1,W_2} = P_{X_1(I),X_2(I),W_1(I),W_2(I)}$. Also define $\mathcal{P}$ as the set of all probability distributions on $X_1, X_2, W_1, W_2, U_1, U_2$ such that P_{X_1,X_2,W_1,W_2} is produced by the above process and U_1 and U_2 satisfy $U_1 \leftrightarrow (X_1, W_1) \leftrightarrow (X_2, W_2) \leftrightarrow U_2$. The ensuing theorem states the n-letter achievable bound for the new coding strategy.

Theorem VI.12. *RD vectors satisfying the following bounds are achievable:*

$$R_1 \geq I(X_1; U_1|U_2 W_1 W_2) + E_{n,\epsilon},$$

$$R_2 \geq I(X_2; U_2|U_1 W_1 W_2) + E_{n,\epsilon},$$

$$R_1 + R_2 \geq I(W_1; S_1) + I(X_1; U_1|W_1, W_2) + I(X_2; U_2|W_1, W_2) - I(U_1; U_2|W_1 W_2) + \theta_n + E_{n,\epsilon},$$

$$D_i \geq E\{d_i(h_i(U_1, U_2, W_1, W_2), X_i)\}.$$

For every probability distribution $P_{X_1,X_2,W_1,W_2,U_1,U_2}$ chosen form $\mathcal{P}$. Here $h_i : \mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}_i$ are the reconstruction functions at the decoder.

Remark 13. *Q_n completely determines P_{X_1,X_2,W_1,W_2}, also from the Markov chain $P_{U_1|X_1,W_1}$, $P_{U_2|X_2,W_2}$ and Q_n fix the induced joint probability distribution $P_{X_1,X_2,W_1,W_2,U_1,U_2}$. Hence, determining the RD region given in theorem 2 involves taking the union of RD vectors satisfying the above bounds for some given f_i, h_i, Q_n, $P_{U_1|X_1,W_1}$ and $P_{U_2|X_2,W_2}$.*

Proof. First we present a summary of the proof. Fix f_i, h_i, Q_n, $P_{U_1|X_1,W}$ and $P_{U_2|X_2,W_2}$. Using Q_n the encoders quantize S_i^n to W_i^n. If multiple realizations of W_1^n's are available at the decoder, encoder 2 can transmit the corresponding sequence of W_2^n's using rate at most $\frac{1}{n}H(W_2^n|W_1^n)$. So the encoders transmit the sequences of W_i's with sum-rate less than $R_n + \frac{1}{n}max\{H(W_2^n|W_1^n), H(W_1^n|W_2^n)\}$. Since S_i's are highly correlated, the vectors S_i^n are almost always equal. Consequently the quantizations W_i^n are almost always equal, and using this the term $\frac{1}{n}max\{H(W_2^n|W_1^n), H(W_1^n|W_2^n)\}$ can be bounded. In the next step (W_i, X_i) is transformed into a DMS by applying the interleaving method explained in [7]. The rest of the problem can be viewed as distributed source coding with sources (W_i, X_i) and side-information (W_1, W_2) available at the decoder. A more rigorous proof is presented next.

Codebook Generation: Three codebooks are used for the quantization. C^n is the underlying codebook for Q_n. The other two codebooks C_i^m, are constructed by choosing each of their elements from $\mathcal{U}_i$ based on distribution P_{U_i}. C_i^m have rates $I(X_i, W_i; U_i) + \lambda_m$ where $\lambda_m \to 0$. Let $Q_{i,m}$ be the quantizers associated with these codebooks. Each of the codebooks C_i^m are randomly binned at rate $I(X_i; U_i|W_1, W_2) - r_i$, where $r_1 + r_2 = I(U_1; U_2|W_1, W_2)$. Given $t \in [0, 1]$, randomly and uniformly bin the space of all vectors $W^{tnm} \in \mathcal{W}^{tnm}$ with rate $E_{n,\epsilon}$, and define B_1 as the binning function. Also, bin the space of all vectors $W^{(1-t)nm}$ using the same rate, let B_2 be the binning function. Finally, choose m permutations π_j, $j \in [1 : m]$ randomly and uniformly from the set of all n-length permutations S_n.

Encoding: Communication is carried out over blocks of length mn. Denote the source sequence in one block as the matrix $X_i(1:m, 1:n)$. The ith encoder calculates $W_i(j, 1{:}n) = Q_n(S_i(j, 1 : n))$ for all j. The encoder wishes to utilize the codebooks C_i^m for quantizing (X_i, W_i), however the source (X_i, W_i) is not a DMS since W_i is produced by a finite-length quantizer. To overcome this difficulty we use the method explained in [7]. Let $\tilde{X}_i(j, 1 : n) = \pi_j(X(j, 1 : n))$, define $\tilde{W}_i$ in the same manner. As shown in [7], $(\tilde{X}_i, \tilde{W}_i)(1 : m, l)$ would behave like a DMS with probability distribution P_{X_i,W_i}. Each encoder calculates $\tilde{U}_i(1:m, l) = Q_{i,m}((\tilde{X}_i, \tilde{W}_i)(1:m, l))$. For rows $(1:tm)$, the first encoder transmits $W_1(j, 1:n)$ while the second encoder sends the bin index $B_1(W_2(1:tm, 1:n))$. For the rest of the rows encoder 1 sends the bin index $B_2(W_1(tm + 1:m, 1:n))$ while encoder 2 sends $W_2(j, 1:n)$. In other words, the encoders time-share between two strategies. In the first strategy encoder 1 transmits W_1 while encoder 2 only sends the bin number for the sequence of W_2, in the second strategy the encoders reverse roles. For every column l, the i_{th} encoder also sends the bin index of $\tilde{U}_i(1:m, l)$ in C_i^m. The resulting rates are:

$$R_1 = tR_n + (1 - t)E_{n,\epsilon} + I(X_1; U_1|W_1 W_2) - r_1$$

$$R_2 = (1 - t)R_n + tE_{n,\epsilon} + I(X_2; U_2|W_1 W_2) - r_2.$$

The achievability of these rates would complete the proof, since they include both corner points of the region in theorem VI.12.

Decoding: The decoder first decodes $W_i(1:m, 1:n)$. For elements $(1:tm, 1:n)$, W_1^{tmn} are available, while only the bin index of W_2^{tmn} is available. Since m is going to infinity, W_2^{tmn}

is losslessly recovered as long the binning rate is more than $\frac{1}{n}H(W_2^n|W_1^n)$. By the following lemma, we have $\frac{1}{n}H(W_2^n|W_1^n) \leq E_{n,\epsilon}$.

Lemma 14. *Let S_1 and S_2 be two DMS's such that $P(S_1 \neq S_2) \leq \epsilon'$ for some $\epsilon' > 0$. Also Let $W_i^n = f_i(S_i^n)$ be n-letter functions of S_i^n to alphabet $\mathcal{W}^n$. Let $\epsilon = 1-(1-\epsilon')^n$. Then the following are true:*

1) $P(W_1^n \neq W_2^n) \leq \epsilon$,

2) $\frac{1}{n}H(W_2^n|W_1^n) \leq \frac{h(\epsilon)}{n} + \epsilon \log|\mathcal{W}_2|$

Proof. In the appendix. □

By the same argument (W_1, W_2) are recovered losslessly for the rest of the rows.

The bin size for each vector $\tilde{U}_1(1\!:\!m, l)$ is:

$$\begin{aligned} I(X_1, W_1; U_1) - I(X_1; U_1|W_1, W_2) + r_1 &= I(X_1, W_1, W_2; U_1) - I(X_1; U_1|W_1, W_2) + r_1 \\ &= I(W_1, W_2; U_1) + r_1 \end{aligned}$$

The long Markov chain is used in the second equation. By the same calculations the bin size for $\tilde{U}_2$ is $I(W_1, W_2; U_2)+r_2$. Using typicality arguments for these bin sizes, there is a unique pair $(U_1, U_2)(1:m, l)$, jointly typical with $(W_1, W_2)(1:m, l)$. We present a summary of the proof. The decoder first creates two ambiguity sets $\mathcal{L}_i$ from the sequences of $U_i(1\!:\!m, l)$'s in the corresponding bins. Each of these sets contains all sequences $U_i(1 : m, l)$ in the bin, which are typical with $(W_1, W_2)(1:m, l)$. There is roughly one such sequence in each $2^{mI(W_1,W_2;U_i)}$ vectors. So the size of $\mathcal{L}_i$ is close to 2^{mr_i}. The decoder finds a pair of vectors in the two ambiguity sets which are typical with each other. Since all these vectors are typical with W_1 and W_2, as long as $r_1 + r_2 \leq I(U_1; U_2|W_1, W_2)$ there is no more than one pair $(U_1, U_2)(1\!:\!m, l)$ typical with respect to $P_{U_1,U_2|W_1,W_2}$. This completes the proof. □

The calculation of the RD region in the theorem requires taking union over all possible n-length quantizers. This renders the characterization practically incomputable. The next proof shows that the RD region in theorem VI.7 is contained in the one in theorem VI.12.

Proof. In the next step, W_2 is removed from the mutual information terms:

$$I(X_1; U_1|U_2, W_1, W_2) = H(U_1|U_2, W_1, W_2) - H(U_1|X_1, W_1, U_2) \leq I(U_1; X_1|W_1, U_2).$$

Also,

$$\begin{aligned} I(X_2; U_2|U_1, W_1, W_2) \leq I(X_2; W_2, U_2|W_1, U_1) &\leq I(X_2; U_2|W_1, U_1) + H(W_2|W_1) \\ &\leq I(X_2; U_2|W_1, U_1) + E_{n,\epsilon}. \end{aligned}$$

W_2 in the terms $I(X_1; U_1|W_1, W_2)$ and $I(X_2; U_2|W_1, W_2)$ in the sum-rate bound can be removed using the same method. For $I(U_1; U_2|W_1, W_2)$ an upper-bound is necessary:

$$I(U_1; U_2|W_1, W_2) \geq I(U_1; U_2|W_1) - H(W_2|W_1) \geq I(U_1; U_2|W_1) - E_{n,\epsilon}.$$

Also, $I(W_1; S_1) = I(W_1; X_1)$. This gives the following inner bound:

$$\begin{aligned} R_1 &\geq I(X_1; U_1|W_1, U_2) + E_{n,\epsilon}, \\ R_2 &\geq I(X_2; U_2|W_1, U_1) + 2E_{n,\epsilon}, \\ R_1 + R_2 &\geq I(W_1; X_1) + 3E_{n,\epsilon} + I(X_1; U_1|W_1), \\ &\quad + I(X_2; U_2|W_1) - I(U_1; U_2|W_1) + \theta_n, \\ D_i &\geq E\{d_i(h_i(U_1, U_2, W_1), X_i)\}. \end{aligned}$$

Q_n is still playing a role in the calculation of the RD region by determining P_{X_1,X_2,W_1,W_2}, which in turn affects $P_{U_2|X_1,X_2,W_1}$. The following lemma provides the means to eliminate this dependency on Q_n.

Lemma 15. *Consider a probability distribution P_{X_1,X_2,W_1,U_1,U_2} satisfying the Markov chains*

$U_1 \leftrightarrow (X_1, W_1) \leftrightarrow (X_2, W_1) \leftrightarrow U_2$ and $W_1 \leftrightarrow S_1 \leftrightarrow (X_1, X_2)$, where S_1 and S_2 are as defined previously. let P_{S_1,W_1} be the marginal distribution of (S_1, W_1), take a quantizer Q_n from $\mathcal{Q}_{P_{S,W}}$. Assume P'_{X_1,X_2,W_1,W_2} is the probability distribution induced by Q_n. Let $P'_{U_1|X_1,W_1} = P_{U_1|X_1,W_1}$ and $P'_{U_2|X_2,W_2} = P_{U_2|X_2,W_1}$. Define:

$$P'_{X_1,X_2,W_1,W_2,U_1,U_2} = P'_{X_1,X_2,W_1,W_2} P'_{U_1|X_1,W_1} P'_{U_2|X_2,W_2}.$$

The following hold:

$$1) P'_{X_1,X_2,W_1,U_1,U_2} \doteq P_{X_1,X_2,W_1,U_1,U_2} \pm \epsilon$$
$$2) I_P(X_1; U_1|W_1, U_2) \doteq I_{P'}(X_1; U_1|W_1, U_2) \pm 2|\mathcal{X}_1||\mathcal{U}_1| \log\left(\frac{p_1}{p_1 - \epsilon}\right)$$
$$3) I_P(X_2; U_2|W_2, U_1) \doteq I_{P'}(X_2; U_2|W_2, U_1) \pm 2|\mathcal{X}_2||\mathcal{U}_2| \log\left(\frac{p_2}{p_2 - \epsilon}\right).$$

In the above equations, $a \doteq b \pm \epsilon$ means $a \in [b - \epsilon, b + \epsilon]$. Also, $I_P(A; B|C)$ denotes the mutual information with respect to P.

Proof. In the appendix. □

The lemma shows that for every probability distribution in theorem VI.7, there is a probability distribution in theorem VI.12, for which the above bounds are well-approximated when calculated using one of the distributions instead of the other. Hence we can provide an inner bound for the above rates by considering the distributions from the first theorem and bounding the estimation error using lemma 15. Applying the estimation bounds gives the region presented in theorem VI.7.

□

Finally, we show that the RD region in theorem VI.12 strictly contains the CC rate-distortion region.

Theorem VI.16. *For the BOHO problem in [7], the RD region in theorem VI.12 achieves points outside of the CC rate-distortion region.*

Proof. Take $U_1 = \phi$, $U_2 = Z + X + W + N_{\delta_0}$ and $W = X + N_\epsilon + N_\delta$, where N_δ is $Be(\delta)$,N_{δ_0} is $Be(\delta_0)$, and the quantization noises are independent of the sources and each other. The reconstruction function is $U_2 + W = X + Z + N_{\delta_0}$. Consider the corner point where encoder 1 is transmitting W by itself and encoder 2 is binning its correlated quantization at rate $E_{n,\epsilon}$. The resulting RD vector approaches $(R_1, R_2, D) = (1 - h_b(\delta), h_b(p * \delta) - h_b(\delta_0), \delta_0)$ as $\epsilon' \to 0$ and $n \to \infty$. In [50] it was shown that these RD vectors are not achievable by the CC scheme when $\epsilon' \neq 0$. By the same argument as in [7], it can be proved that there exist n and ϵ' for which the resulting rates are not achievable by the CC scheme. □

There can be highly correlated components between the sources given W_1 and W_2. In this case, there must be several finite-length codebooks super-imposed on each other, one for each of the highly correlated components. The inner bound presented here can be extended to include these new layers.

APPENDICES

APPENDIX A

Proofs for Chapter II

A.1 Proofs for Section 2.4

A.1.1 Proof of Lemma 17

Proof. Let $(R_i, D_N)_{i\in\mathsf{L},N\in\mathcal{L}} \in \mathcal{RD}_{SSC}(P_{\mathbf{U},X})$ and $(R'_i, D'_N)_{i\in\mathsf{L},N\in\mathcal{L}} \in \mathcal{RD}_{SSC}(P'_{\mathbf{U},X})$. Without loss of generality, assume $\mathsf{U}_{\{1\},\{2\},\{3\}} = \mathsf{U}'_{\{1\},\{2\},\{3\}}$. Let $\tilde{U}_{\{1\},\{2\},\{3\}}$ be defined on $\mathsf{U}_{\{1\},\{2\},\{3\}} \times \{0,1\}$. Also let $\tilde{U}_{\mathcal{M}} = U_{\mathcal{M}}$ if $\mathcal{M} \neq \{\{1\},\{2\},\{3\}\lambda\}$. For $\lambda \in [0,1]$, define a new distribution $\tilde{P}_{\tilde{\mathbf{U}},X}$ as follows:

$$\tilde{P}_{\tilde{\mathbf{U}},X}(\tilde{\mathbf{u}}, x) = \begin{cases} \lambda P_{\mathbf{U},X}(\mathbf{u}, x) & \tilde{u}_{\{1\},\{2\},\{3\}} = (u_{\{1\},\{2\},\{3\}}, 0) \\ (1-\lambda) P'_{\mathbf{U},X}(\mathbf{u}, x) & \tilde{u}_{\{1\},\{2\},\{3\}} = (u_{\{1\},\{2\},\{3\}}, 1) \end{cases}$$

Then it is straightforward to check that $\lambda(R_i, D_N)_{i\in\mathsf{L},N\in\mathcal{L}} + (1-\lambda)(R'_i, D'_N)_{i\in\mathsf{L},N\in\mathcal{L}} \in \mathcal{RD}_{SSC}(\tilde{P}_{\tilde{\mathbf{U}},X})$.

□

A.1.2 Proof of Lemma 19

Proof. We provide an outline of the proof. Fix $\mathsf{M}' \in \mathcal{L}$. Consider a new scheme where the reconstruction function at decoder M' is defined as $f_{\mathsf{M}'} : \prod_{\mathcal{M}\in\mathbf{M}_{\mathsf{M}'}} U_{\mathcal{M}} \to \mathsf{X}$ with the rest of the reconstruction functions defined as in Theorem II.16. Let the RD vector $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},\mathsf{N}\in\mathcal{L}}$ be achievable in the new scheme using the distribution $P_{U_{\mathsf{S}_\mathsf{L}},X}$ and reconstruction functions $f_{\mathsf{M}'}, g_{\mathsf{M}}, \mathsf{M} \in \mathcal{L}\backslash\{\mathsf{N}\}$. We provide a new probability distribution $P_{U'_{\mathsf{S}_\mathsf{L}},X}$and reconstruction functions $g'_{\mathsf{M}} : U_{\mathsf{M}} \to \mathsf{X}, \mathsf{M} \in \mathcal{L}$ to shows that the RD region given in Theorem II.16 contains $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},\mathsf{N}\in\mathcal{L}}$. To construct the probability distribution define $U'_{\mathcal{M}} = U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathsf{N}}\backslash\{\{\mathsf{M}'\}\}$, and $U'_{\mathsf{M}'} = (U_{\mathsf{M}'}, f_{\mathsf{M}'}(U_{\mathbf{M}_{\mathsf{M}'}}))$. As for the reconstruction functions define $g'_{\mathsf{M}}(U_{\mathsf{M}}) = g_{\mathsf{M}}(U_{\mathsf{M}}), \mathsf{M} \in \mathcal{L}\backslash\{\mathsf{N}\}$ and $g'_{\mathsf{M}'}(U'_{\mathsf{M}'}) = f_{\mathsf{M}'}(U_{\mathbf{M}_{\mathsf{M}'}})$. It is straightforward to check that with these parameters, the RD region in theorem II.16 contains $(R_i, D_{\mathsf{N}})_{i\in\mathsf{L},\mathsf{N}\in\mathcal{L}}$. Intuitively, since the reconstruction functions are the same, the same distortion is achieved by both schemes. As for the rates, in the first scheme, wherever $U_{\mathsf{M}'}$ is decoded, all of the random variables $U_{\mathbf{M}_{\mathsf{M}'}}$ are also decoded. So, adding a function of these random variables to $U_{\mathsf{M}'}$ does not require additional rate. □

A.1.3 Proof of Lemma 26

Proof. Let $U_{\{1\}} = \hat{W}$, $U_{\{1,2\},\{3\}} = W$, $U_{\{1,2\}} = \hat{X}_1$, $U_{\{3\}} = \hat{X}_2$, where $\hat{X}_i$ are the reconstructions at decoder $\{i\}$ in the two user problem in Example 14. Then it is straightforward to check that the RD vector is achievable from Theorem II.16. Next, assuming the codebook $C_{\{1,2\},\{3\}}$ is empty, we consider all of the remaining 16 codebooks in the SSC scheme and show that the RD vector is not achievable.

Step 1: In this step, we argue that the only non-trivial codebooks are $C_{\{1\}}$, $C_{\{3\}}$, $C_{\{1\},\{3\}}$,$C_{\{1,2\}}$ and $C_{\{2\},\{3\}}$. Due to the structure of the problem, a number of the codebooks are functionally equivalent, meaning they are decoded at exactly the same decoders. So we can merge these codebooks without any loss. For example, description $\{2\}$ is only received by decoders $\{1,2\}$ and $\{1,2,3\}$, hence we can merge $C_{\{2\}}$ into $C_{\{1,2\}}$ without any loss. $C_{\{1,3\},\{23\}}$, $C_{\{1,3\}}$, $C_{\{2,3\}}$

and $C_{\{1,2,3\}}$ are only decoded at decoder $\{1,2,3\}$ so they are redundant from the results in [51]. $C_{\{1\},\{2\}}$ can be merged into $C_{\{1\}}$ since decoder $\{2\}$ is not present. $C_{\{1\},\{2\},\{3\}}$ is equivalent to $C_{\{1\},\{3\}}$ and can be eliminated. $C_{\{1,2\},\{1,3\},\{2,3\}}$, $C_{\{1,2\},\{1,3\}}$ and $C_{\{1,2\},\{2,3\}}$ can be merged into $C_{\{1,2\}}$. Finally $C_{\{2,3\},\{1\}}$ can be merged with $C_{\{1\}}$. Also $C_{\{1,2\}}$ can be merged with $C_{\{1,2,3\}}$ and is eliminated. So we are left with four codebooks $C_{\{1\}}$, $C_{\{3\}}$, $C_{\{1\},\{3\}}$,$C_{\{1,2\}}$ and $C_{\{2\},\{3\}}$.

Step 2: In this step, we show that if we set $U_{\{1\},\{3\}} = \hat{W}$ and $U_{\{1\}} = \phi$, there would be no loss in terms of RD function. The codebooks $C_{\{1\}}$ and $C_{\{1\},\{3\}}$ are decodable using description 1. Since decoder $\{1\}$ is at PtP optimality, these codebooks only carry $\hat{W}$. To be more precise there is a Markov chain $(U_{\{1\}}, U_{\{1\},\{3\}}) \leftrightarrow \hat{W} \leftrightarrow X$, which we prove in the following lemma.

Lemma 1. *In a PtP setup assume the decoder is at optimal PtP RD. It receives variables $U_{\mathbf{M}}$, and the reconstruction function is $f(U_{\mathbf{M}})$. Then the following Markov chain holds $U_{\mathbf{M}} \leftrightarrow f(U_{\mathbf{M}}) \leftrightarrow X$.*

Proof.

$$
\begin{aligned}
&R \geq I(U_{\mathbf{M}}; X) \overset{(a)}{=} I(f(U_{\mathbf{M}}), U_{\mathbf{M}}; X) = I(f(U_{\mathbf{M}}); X) + I(U_{\mathbf{M}}; X | f(U_{\mathbf{M}})) \\
&\overset{(b)}{\geq} R + I(U_{\mathbf{M}}; X | f(U_{\mathbf{M}})) \Rightarrow I(U_{\mathbf{M}}; X | f(U_{\mathbf{M}})) = 0,
\end{aligned}
$$

where in (a) we used the fact that $f(U_{\mathbf{M}})$ is a function of $U_{\mathbf{M}}$ and in (b) we used the PtP optimality. □

Since $\hat{W}$ is decoded both at decoder $\{1\}$ and $\{3\}$, if we replace $U_{\{1\},\{3\}}$ with $(U_{\{1\},\{3\}}, \hat{W})$, the decoders decode the same random variables as before, so no extra rate is required. Also, from the lemma $(U_{\{1\}}, U_{\{1\},\{3\}}) \leftrightarrow \hat{W} \leftrightarrow X$. Hence, we conclude that we can set $U_{\{1\},\{3\}} = \hat{W}$ and $U_{\{1\}} = \phi$ without any loss in terms of distortion.

Step 3: Assume there are random variables $U_{\{1\},\{3\}}$ and $U_{\{2\},\{3\}} = (W, U'_{\{2\},\{3\}})$ such that the RD vector is achievable in the SSC scheme. From the Markov chain $\hat{W} \leftrightarrow W \leftrightarrow X$, description 1 is not used in the reconstruction in decoders $\{1,2\}$, $\{3\}$ and $\{1,2,3\}$. If we set

$U_{\{1\}} = \phi$, the distortions constraint in decoders $\{1,2\}$, $\{3\}$ and $\{1,2,3\}$ are satisfied. So we have constructed a scheme to send the descriptions at a lower rate (by setting $U_{\{1\}} = \phi$) without any loss in terms of distortion in these three decoders. This contradicts optimality of the random variables chosen for the two user scheme.

A.1.4 Proof of Lemma 24 for l>3

We have proved that if $C_{\{12\},\{3\}} = \phi$, the RD vector is not achievable but if the constraint is lifted the scheme can achieve this RD vector, so the codebook is non-redundant. For the general l-descriptions problem, we provide an outline of the non-redundancy proof for $C_{\mathcal{H}}, \mathcal{H} \in \mathbf{S}_{\mathsf{L}}$. Let $\{a_{1,i}, a_{2,i}, \ldots, a_{n_i,i}\}, i \in [1,k]$ be the elements of $\mathcal{H}$. Then to construct an example where $C_{\mathcal{H}}$ is non-redundant, first consider a set up where for any i, each set of three decoders $\{a_{1,i}, a_{2,i}, \ldots, a_{n_i,i}\}$ and $\{a_{1,i+1}, a_{2,i+1}, \ldots, a_{n_{i+1},i+1}\}$ and $\{a_{1,i}, a_{2,i}, \ldots, a_{n_i,i}, a_{1,i+1}, a_{2,i}, \ldots, a_{n_{i+1},i+1}\}$ are as in the two user setup in Example 14. Then there should be a common component between each two of the descriptions. It is straightforward to show that the common components must be the same for all of the decoders, otherwise since the codebooks are independent there would be a rate-loss as explained in the previous section. We ensure that the common component can be decoded only when all descriptions $a_{1,i}a_{2,i}, \ldots a_{n_i,i}a_{1,i+1}a_{2,i} \ldots a_{n_{i+1},i+1}$ are received and not when a subset of the descriptions is received. This is done by adding decoders $\{a_{1,i}\}$, $\{a_{1,i}, a_{2,i}\}$ through $\{a_{1,i}, a_{2,i}, \ldots a_{n_i,i}, a_{1,i+1}, a_{2,i+1}, \ldots a_{n_{i+1},i+1}\}$ such that each of them would be at PtP optimality by receiving a refined version of W (i.e $\{a_{1,i}\}$ would receive $\hat{W}$ and $\{a_{1,i}, a_{2,i}\}$ would receive a refinement of $\hat{W}$ and so on). In this way the only codebook that can carry W without rate-loss is $C_{\mathcal{H}}$. □

A.1.5 Proof of Lemma 28

Proof. Let $\rho_{\{1,2\},\{1,3\},2} + \rho_{\{1,2\},\{1,3\},3} > 0$, description 1 carries $\hat{W}$ to decoder $\{1\}$ with rate $I(\hat{W}; X)$. Descriptions 2 and 3 send W to decoders $\{1,2\}$ and $\{1,3\}$ by sending a refinement on $C_{\{1,2\},\{1,3\}}$. In other words $U_{\{1\}} = \hat{W}$, $U_{\{1,2\},\{1,3\}} = W$, $U_{\{1,2\}} = \hat{X}_1$ and $U_{\{1,2\}} = \hat{X}_2$ similar to the proof of Lemma 26. Then one can check that the RD vector is achievable using the SSC scheme. Next, assume $\rho_{\{1,2\},\{1,3\},2} + \rho_{\{1,2\},\{1,3\},3} = 0$, then $\rho_{\{1,2\},\{1,3\},i} = 0, i \in \{2,3\}$. As in the previous section, we begin by eliminating the redundant codebooks for this communications setting.

Step 1: In this step we argue that only the codebooks $C_{\{1,3\}}$, $C_{\{1,2\}}$ $C_{\{1\}}$ and $C_{\{1,2\},\{1,3\}}$ are non-trivial. Due to the structure of this communications setting many of the codebooks are functionally the same and can be merged together. The codebooks $C_{\{1\},\{2\},\{3\}}$, $C_{\{1\},\{3\}}$, $C_{\{1\},\{2\}}$, $C_{\{2,3\},\{1\}}$ are decoded at all four of the decoders and can be merged with $C_{\{1\}}$. $C_{\{1,3\},\{2,3\}}$ can be merged with $C_{\{1,3\}}$ since decoder $\{2,3\}$ is not present, by the same argument $C_{\{1,2\},\{1,3\},\{2,3\}}$ is concatenated with $C_{\{1,2\},\{1,3\}}$, also $C_{\{1,2\},\{2,3\}}$ and $C_{\{1,2\},\{3\}}$ are merged with $C_{\{1,2\}}$. $C_{\{1,3\},\{2\}}$ and $C_{\{2\},\{3\}}$ are combined with $C_{\{1,2\},\{2,3\}}$. Lastly since decoders 2 and 3 are not present, $C_{\{2\}}$ and $C_{\{3\}}$ can be merged into $C_{\{1,2\}}$ and $C_{\{1,3\}}$, respectively. So only the four codebooks $C_{\{1,3\}}$, $C_{\{1,2\}}$ $C_{\{1\}}$ and $C_{\{1,2\},\{1,3\}}$ remain.

Step 2: By the same arguments as in step 2 of Lemma 26, we can set $U_{\{1\}} = \hat{W}$.

Step 3: By assumption, the codebook $C_{\{1,2\},\{1,3\}}$ is only carried by the first description. However, the codebook is not decoded at decoder $\{1\}$. Since the decoder is at PtP optimality, $C_{\{1,2\},\{1,3\}}$ can't be sent through the first description either (i.e $\rho_{\{1,2\},\{1,3\},1} = 0$ and $C_{\{1,2\},\{1,3\}}$ can be eliminated.).

Step 4: After Fourier-Motzkin elimination, the covering and packing bounds for the remaining three codebooks give the following inequality,

$$R_1 + R_2 + R_3 \geq I(U_{\{1,2\}}, U_{\{1,3\}}; X|\hat{W}) + I(U_{\{1,3\}}; U_{\{1,2\}}|\hat{W}) + I(\hat{W}; X) \tag{A.1}$$

By the definition of $\hat{W}$ we have $\hat{W} \leftrightarrow W \leftrightarrow X$ and $I(\hat{W}; X) < I(W; X)$, so the bound above is strictly larger than the case when $\hat{W}$ is replaced by W (i.e. when $U_{\{1,2\},\{1,3\}} = W$.). This concludes the proof.

□

A.2 Proofs for Section 2.5

A.2.1 Proof of Lemma 31

Proof.

$$\begin{aligned}\frac{1}{n}E(d_H(\hat{X}^n \oplus_2 \hat{Z}^n, X^n \oplus_2 Z^n)) &= \frac{1}{n}E(w_H(\hat{X}^n \oplus_2 \hat{Z}^n \oplus_2 X^n \oplus_2 Z^n)) \\ &= \frac{1}{n}E(w_H(X^n \oplus_2 \hat{X}^n \oplus_2 Z^n \oplus_2 \hat{Z}^n)) \\ &= \frac{1}{n}E(d_H(X^n \oplus_2 \hat{X}^n, Z^n \oplus_2 \hat{Z}^n)).\end{aligned}$$

Note that $X^n \oplus_2 \hat{X}^n$ is the quantization noise of quantizing X^n and $Z^n \oplus_2 \hat{Z}^n$ is the quantization noise of quantizing Z^n. Since the source vectors are independent, the noise vectors are also independent and the summation converges to $\delta * \delta$ (The arguments are similar to the ones given in [7].).

□

A.2.2 Proof of Lemma 32

Proof. We assume that there exists a probability distribution P on X and $U_{\mathbf{S}_L}$ for which the RD vector is achievable using the SSC scheme and arrive at a contradiction. Since all of the decoders are present in this setup, we need to consider the SSC with all the codebooks present, so the proof is more involved than the proofs in the previous section.

Step 1: In this step we show that description i, where $i = 1, 2$, does not carry any bin indices for codewords from codebook $C_{\mathcal{M}}$ if $\mathcal{M} \notin \mathbf{M}_{\{i\}}$. Descriptions 1 and 2 only carry indices which are used in the reconstruction at decoders {1} and {2}, respectively. This is true since these two decoders are receiving information at optimal PtP rate-distortion. Note that this does not mean the corresponding codebooks are empty, we can only conclude that no bin indices for the codewords are sent through these descriptions. For example if $\mathcal{M} = \{\{2\},\{1,3\}\}$ and $i = 1$, then $\rho_{\mathcal{M},i} = 0$.

Lemma 2. *For $i \in \{1,2\}$, and $\mathcal{M}$ such that $\{i\} \notin \mathcal{M}$, $\rho_{\mathcal{M},i} = 0$.*

Proof. From optimality at decoder $\{1\}$ we have the following equality:

$$R_i = I(U_{\mathbf{M}_{\{i\}}}; X, Z) \tag{A.2}$$

Consider the following covering bound on the random variables $U_{\mathbf{M}_{\{i\}}}$:

$$H(U_{\mathbf{M}_{\{i\}}}|X,Z) \geq \sum_{\mathcal{M}\in\mathbf{M}_{\{i\}}} (H(U_{\mathcal{M}}) - r_{\mathcal{M}}), \tag{A.3}$$

also we have the following packing bound at decoder $\{i\}$:

$$H(U_{\mathbf{M}_{\{i\}}}) \leq \sum_{\mathcal{M}\in\mathbf{M}_{\{i\}}} (H(U_{\mathcal{M}}) + \rho_{\mathcal{M},i} - r_{\mathcal{M}}), \tag{A.4}$$

adding (A.3) and (A.4) we get:

$$\sum_{\mathcal{M}\in\mathbf{M}_{\{i\}}} \rho_{\mathcal{M},i} \geq I(U_{\mathbf{M}_{\{i\}}}; X, Z), \tag{A.5}$$

$R_i = \sum_{\mathcal{M}\in\mathbf{S}_\mathrm{L}} \rho_{\mathcal{M},i}$, comparing this equality with (A.2) completes the proof. □

Step 2: In this step, we show that there are no common codebooks decoded at decoders $\{1\}$ and $\{2\}$. Since decoder $\{1,2\}$ receives descriptions 1 and 2 at optimal RD from a PtP perspective, the random variables decoded at decoder $\{1\}$ must be independent of those decoded at decoder $\{2\}$. From the next lemma we have that if $\mathcal{M} \in \mathbf{M}_{\{1\}} \bigcap \mathbf{M}_{\{2\}}$ then $r_{\mathcal{M}} = 0$.

Lemma 3. *Consider the setup in Figure 2.1, let $(R_1, R_2, D_1, D_2, D_{\{1,2\}})$ be such that $R_1+R_2 = RD_{d_{\{1,2\}}}(D_{\{1,2\}})$, where $RD_d(D)$ is Shannon's optimal PtP RD function for distortion function*

d at point D. For any distribution $P_{U_{\{1\}},U_{\{2\}},U_{\{1,2\}},U_{\{1\},\{2\}}}$ which achieves this RD vector, the following conditions must hold: 1)$U_{\{1\}} \perp\!\!\!\perp U_{\{2\}}$ and $C_{\{1\},\{2\}} = \phi$

2)If in addition $R_i = RD_{d_{\{i\}}}(D_i), i \in \{1,2\}$ then, $U_{\{1,2\}} \leftrightarrow (U_{\{1\}}, U_{\{2\}}) \leftrightarrow X$.

Proof. Consider the following packing bounds:

$$\textbf{Dec}\,\{1\} : H(U_{\{1\},\{2\}}, U_{\{1\}}) \leq H(U_{\{1\},\{2\}})+H(U_{\{1\}}) +\rho_{\{1\},\{2\},1}+\rho_{\{1\},1}-r_{\{1\},\{2\}}-r_{\{1\}} \tag{A.6}$$

$$\textbf{Dec}\,\{2\} : H(U_{\{1\},\{2\}}, U_{\{2\}}) \leq H(U_{\{1\},\{2\}})+H(U_{\{2\}}) +\rho_{\{1\},\{2\},2}+\rho_{\{2\},2}-r_{\{1\},\{2\}}-r_{\{2\}} \tag{A.7}$$

$$\textbf{Dec}\,\{1,2\} : H(U_{\{1,2\}}|U_{\{1\},\{2\}}, U_{\{1\}}, U_{\{2\}}) \leq H(U_{\{1,2\}}) +\rho_{\{1,2\},1}+\rho_{\{1,2\},2}-r_{\{1,2\}} \tag{A.8}$$

Also the mutual covering bound:

$$H(U_{\{1,2\}}, U_{\{1\}}, U_{\{2\}}, U_{\{1\},\{2\}}|X) \geq \tag{A.9}$$

$$H(U_{\{1,2\}}) + H(U_{\{1\}}) + H(U_{\{2\}}) + H(U_{\{1\},\{2\}}) - r_{\{1,2\}} - r_{\{1\}} - r_{\{2\}} - r_{\{1\},\{2\}} \tag{A.10}$$

Now we add inequalities (A.6-A.8) and subtract (A.10), we get:

$$I(U_{\{1,2\}}, U_{\{1\}}, U_{\{2\}}, U_{\{1\},\{2\}}; X) + I(U_{\{1\}}; U_{\{2\}}|U_{\{1\},\{2\}}) \leq R_1 + R_2 - r_{\{1\},\{2\}}$$

Using the condition $R_1 + R_2 = RD_{d_{12}}(D_{\{1,2\}})$ we conclude:

$$I(U_{\{1\}}; U_{\{2\}}|U_{\{1\},\{2\}}) + r_{\{1\},\{2\}} \leq 0 \tag{A.11}$$

From (A.11) one may deduce $C_{\{1\},\{2\}} = \phi$ and $U_{\{1\}} \perp\!\!\!\perp U_{\{2\}}$. Furthermore we get:

$$R_1 + R_2 = I(U_{\{1,2\}}, U_{\{1\}}, U_{\{2\}}; X) = I(U_{\{1\}}; X)+I(U_{\{2\}}; X)+I(U_{\{1\}}; U_{\{2\}}|X) +I(U_{\{1,2\}}; X|U_{\{1\}}, U_{\{2\}}),$$

where the right-hand side of the second equality is the sum-rate of the two-descriptions problem. Using the conditions $R_i = RD_{d_i}(D_i), i \in \{1,2\}$, we have:

$$I(U_{\{1\}}; U_{\{2\}}|X)+I(U_{\{1,2\}}; X|U_{\{1\}}, U_{\{2\}}) = 0.$$

So $I(U_{\{1,2\}}; X|U_{\{1\}}, U_{\{2\}}) = 0$, which gives the desired Markov chain in (2). □

Assuming the original scheme achieves the RD vector in the theorem, we give a new scheme which also achieves the RD vector. We propose that the encoder operates as before, but decoder $\{1,2\}$ decodes $U_{\mathcal{M}}$ only if $\mathcal{M} \in \mathbf{M}_{\{1\}}$ or $\mathcal{M} \in \mathbf{M}_{\{2\}}$. It needs to be shown that the RD vector is the same. First we consider the resulting rates. The covering bounds are not changed. The packing bounds are the same at all decoders other than decoder $\{1,2\}$ since the same variables are being decoded at those decoders. $\mathbf{M}_{\{\mathbf{1}\}} \cap \mathbf{M}_{\{\mathbf{2}\}} = \phi$. Let $\tilde{\mathbf{M}}_{\{1\}}$ and $\tilde{\mathbf{M}}_{\{2\}}$ be subsets of $\mathbf{M}_{\{1\}}$ and $\mathbf{M}_{\{2\}}$. We need to show that the following packing bound is satisfied:

$$H(U_{\mathbf{M}_{\{1\}}}, U_{\mathbf{M}_{\{2\}}}|U_{\tilde{\mathbf{M}}_{\{1\}}}, U_{\tilde{\mathbf{M}}_{\{2\}}}) \leq \sum_{\mathcal{M}\in\mathbf{M}_{\{1\}}\cup\mathbf{M}_{\{2\}}\setminus\tilde{\mathbf{M}}_{\{1\}}\cup\tilde{\mathbf{M}}_{\{1\}}} (H(U_{\mathcal{M}}) + \rho_{\mathcal{M},1} + \rho_{\mathcal{M},2} - r_{\mathcal{M}}) \quad \text{(A.12)}$$

We have the following two packing bounds from decoders $\{1\}$ and $\{2\}$:

$$H(U_{\mathbf{M}_{\{1\}}}|U_{\tilde{\mathbf{M}}_{\{1\}}}) \leq \sum_{\mathcal{M}\in\mathbf{M}_{\{1\}}\setminus\tilde{\mathbf{M}}_1} (H(U_{\mathcal{M}}) + \rho_{\mathcal{M},1} - r_{\mathcal{M}}) \quad \text{(A.13)}$$

$$H(U_{\mathbf{M}_{\{2\}}}|U_{\tilde{\mathbf{M}}_{\{2\}}}) \leq \sum_{\mathcal{M}\in\mathbf{M}_{\{2\}}\setminus\tilde{\mathbf{M}}_1} (H(U_{\mathcal{M}}) + \rho_{\mathcal{M},2} - r_{\mathcal{M}}) \quad \text{(A.14)}$$

Note that from arguments in Lemma 3, $U_{\mathbf{M}_{\{1\}}}$ is independent of $U_{\mathbf{M}_{\{2\}}}$. Hence adding (A.13) and (A.14), we get (A.12). This proves that the packing bounds are also the same.

From lemma 3, we have $U_{\mathbf{M}_{\{1,2\}}} \leftrightarrow U_{\mathbf{M}_{\{1\}}}, U_{\mathbf{M}_{\{2\}}} \leftrightarrow X, Z$. Lemma 4 shows that the new scheme achieves the same distortions as the previous one.

Lemma 4. *Let the random variables U, V, X be such that $U \leftrightarrow V \leftrightarrow X$. Then for an arbitrary distortion function $f : \mathsf{X} \times \hat{\mathsf{X}} \to \mathsf{R}^{+}$, there is an optimal reconstruction of X using*

U and V which is a only function of V.

Proof. We know that the optimal reconstruction function for X given U and V is given by:

$$g(u,v) = \arg\min_{\hat{x}\in\hat{\mathsf{X}}} \mathsf{E}(f(\hat{x},X)|u,v) = \arg\min_{\hat{x}\in\hat{\mathsf{X}}} \mathsf{E}(f(\hat{x},X)|v),$$

which is only a function of V. $\square$

By these arguments, codebook $U_{\mathcal{M}}$ is eliminated if $\mathcal{M} \in \mathbf{M}_{\{1,2\}} \backslash \widetilde{\mathbf{M}}_{\{1,2\}}$. Also in the new scheme, $U_{\{1,2\},\{1,3\},\{2,3\}}$ and $U_{\{1,3\},\{2,3\}}$ are functionally similar since by the same arguments as in this step $U_{\{1,2\},\{1,3\},\{2,3\}}$ is not used in the reconstruction in decoder $\{1,2\}$, so we can eliminate $C_{\{1,2\},\{1,3\},\{2,3\}}$. In summary, thus far we have eliminated 7 codebooks.

Step 3: We have the following lemma:

Lemma 5. *From optimality of rate and distortion at decoders* $\{1,3\},\{2,3\}$ *we have:*

$$\rho_{\{2,3\},3} = \rho_{\{1,3\},3} = \rho_{\{2,3\},\{1\},3} = \rho_{\{1,3\},\{2\},3} = 0$$

Proof. First we argue that $\rho_{\{2,3\},3} = 0$. If this is not true, it contradicts optimality at decoder $\{1,3\}$. $U_{\{2,3\}}$ is not decoded at decoder $\{1,3\}$, but its bin index is carried through description 3. So if the bin index is non-zero, one could reduce R_3 by setting the bin index equal to 0 without increasing distortion at decoder $\{1,3\}$, this contradicts optimality at that decoder. By the same arguments $\rho_{\{1,3\},3} = 0$. Now assume $\rho_{\{2,3\},\{1\},3} \neq 0$. We show that this contradicts optimality at decoder $\{1,3\}$. $U_{\{2,3\},\{1\}}$ is decodable using description 1 (since it is decodable at decoder $\{1\}$). Hence, if we set $\rho_{\{2,3\},\{1\}\},3}$ to 0 (i.e. do not send the bin index on description 3), then decoder $\{1,3\}$ can still decode $U_{23,1}$ using description 1. So the distortion is the same at this decoder, but the rate R_3 is reduced which contradicts optimality. By the same arguments, $\rho_{\{13,2\},3} = 0$. $\square$

Step 4: We proceed by showing that $r_{\{1,3\}} = r_{\{2,3\}} = 0$. So far we have shown that none of the descriptions carry the bin indices for these codebooks.Consider the following packing bounds in decoders $\{1\}$, $\{2,3\}$ and $\{1,3\}$:

$$H(U_{\{1\}}U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}) \leq H(U_{\{1\}}) + H(U_{\{1\},\{3\}}) + H(U_{\{2,3\},\{1\}}) + R_1 - r_{\{1\}} - r_{\{1\},\{3\}} - r_{\{2,3\}\{1\}}$$

$$H(U_{\{2\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{2,3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\}\{2,3\}}) \leq H(U_{\{2\}}) + H(U_{\{3\}}) + H(U_{\{1\},\{3\}})+$$

$$H(U_{\{2\}\{3\}}) + H(U_{\{2,3\}}) + H(U_{\{1,3\},\{2\}}) + H(U_{\{2,3\},\{1\}}) + H(U_{\{1,3\},\{2,3\}}) + R_2 + R_3 - r_{\{2\}}$$

$$- r_{\{3\}} - r_{\{1\}\{3\}} - r_{\{2\},\{3\}} - r_{\{2,3\}} - r_{\{1,3\}\{2\}} - r_{\{2,3\}\{1\}} - r_{\{1,3\}\{2,3\}}$$

$$H(U_{\{1,3\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\}\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\}\{2,3\}}) \leq H(U_{\{1,3\}}) - r_{\{1,3\}}$$

We add the above inequalities and subtract the mutual covering bound on all RV's, we get:

$$H(U_{\{1\}}U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}) + H(U_{\{2\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{2,3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\}\{2,3\}})$$

$$+ H(U_{\{1,3\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\}\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\}\{2,3\}})$$

$$- H(U_{\{1\}}, U_{\{2\}}, U_{\{3\}}, U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, U_{\{1,3\},\{2\}}, U_{\{2,3\},\{1\}}, U_{\{2,3\}}, U_{\{1,3\},\{2,3\}}, U_{\{1,3\}}|X, Z)$$

$$\leq H(U_{\{1,3\}}) + H(U_{\{2,3\},\{1\}}) - r_{\{1,3\}} - r_{\{2,3\},\{1\}} + R_1 + R_2 + R_3$$

$$\Rightarrow I(U_{\{1,3\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\}\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\}\{2,3\}}; X, Z) + I(U_{\{1\}}, U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}; X, Z)+$$

$$I(U_{\{1,3\}}; X, Z|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\}\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\}\{2,3\}}) \leq R_1 + R_2 + R_3$$

$$\Rightarrow I(X, Z; U_{\{1,3\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\}\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\}\{2,3\}}) = 0.$$

This imposes the Markov chain $U_{\{1,3\}} \leftrightarrow U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}\ U_{\{2\},\{3\}}U_{\{1,3\}\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\}\{2,3\}} \leftrightarrow X, Z$. Hence by the same arguments as in step 2, we can eliminate $C_{\{1,3\}}$. Also by the same arguments $C_{\{2,3\}}$ can be eliminated.

Step 5: In this step we eliminate $C_{\{1\},\{3\}}$ and $C_{\{2\},\{3\}}$.

Lemma 6. *The following equality holds:*

$$\rho_{\{1\},\{3\},1} = \rho_{\{1\},\{3\},3} = \rho_{\{2\},\{3\},2} = \rho_{\{2\},\{3\},1} = 0$$

Proof. Assume $\rho_{\{1\},\{3\},1} > 0$. We claim this contradicts optimality at decoder $\{1,3\}$, since

$U_{\{1\},\{3\}}$ can readily be decoded from the bin number carried by description 3, so setting $\rho_{\{1\},\{3\},1}$ to 0 would decease rate without increasing distortion. The rest of the proof follows by the same argument. $\square$

Now consider the following packing bounds at decoders $\{1\}$, $\{3\}$ and $\{1,3\}$ and the mutual covering bound:

$$H(U_{\{1\}}U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}) \leq H(U_{\{1\}}) + H(U_{\{1\},\{3\}}) + H(U_{\{2,3\},\{1\}}) + R_1 - r_{\{1\}} - r_{\{1\},\{3\}} - r_{\{2,3\}\{1\}}$$

$$H(U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}) \leq H(U_{\{3\}}) + H(U_{\{1\},\{3\}}) + H(U_{\{2\}\{3\}}) + R_3 - \rho_{\{1,3\}\{2,3\},3} - r_{\{3\}} - r_{\{1\}\{3\}} - r_{\{2\},\{3\}}$$

$$H(U_{\{1,3\},\{2,3\}}, U_{\{1,3\}\{2\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{2,3\},\{1\}}) \leq$$

$$H(U_{\{1,3\},\{2,3\}}) + H(U_{\{1,3\},\{2\}}) + \rho_{\{1,3\}\{2,3\},3} - r_{\{1,3\},\{2,3\}}$$

$$H(U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\}\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}|X,Z) \geq$$

$$H(U_{\{1\}}) + H(U_{\{3\}}) + H(U_{\{1\},\{3\}}) + H(U_{\{2\},\{3\}}) + H(U_{\{1,3\}\{2\}}) + H(U_{\{2,3\},\{1\}}) + H(U_{\{1,3\},\{2,3\}})$$

$$- r_{\{1\}} + r_{\{3\}} - r_{\{1\},\{3\}} - r_{\{2\},\{3\}} - r_{\{1,3\},\{,2\}} - r_{\{2,3\},\{1\}} - r_{\{1,3\},\{2,3\}}.$$

Adding the above packing bounds and subtracting the mutual covering bound we get:

$$\begin{aligned}
&H(U_{\{1\}}U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}) + H(U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}) \\
&+ H(U_{\{1,3\},\{2,3\}}, U_{\{1,3\}\{2\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{2,3\},\{1\}}) \\
&- H(U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\}\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}|X,Z) \leq R_1 + R_3 + H(U_{\{1\}\{3\}} - r_{\{1\},\{3\}} \\
&\Rightarrow I(U_{\{1\}}U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}; U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}})+ \\
&I(U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\}\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}|X,Z) - H(U_{\{1\},\{3\}}) \leq R_1 + R_3 - r_{\{1\},\{3\}} \\
&\Rightarrow I(U_{\{1\}}U_{\{2,3\},\{1\}}; U_{\{3\}}U_{\{2\},\{3\}}|U_{\{1\},\{3\}}) + r_{\{1\},\{3\}} \leq 0.
\end{aligned}$$

Particularly $r_{\{1\},\{3\}} = 0$, by the same arguments $r_{\{2\},\{3\}} = 0$.

Step 6: So far we have shown that only $C_{\{1\}}, C_{\{2\}}, C_{\{3\}}, C_{\{1,3\},\{2\}}, C_{\{2,3\},\{1\}}$ and $C_{\{1,3\}\{2,3\}}$ can be non-trivial. From optimality at decoders $\{1\}$ and $\{1,3\}$ we have the following equalities:

$$R_1 = I(U_{\{1\}}, U_{\{2,3\},\{1\}}; X, Z), R_1 + R_3 = I(U_{\{1\}}, U_{\{2,3\},\{1\}}, U_{\{3\}}, U_{\{1,3\},\{2\}}, U_{\{1,3\},\{2,3\}}; X, Z) \quad \text{(A.15)}$$

Hence we have:

$$R_3 = I(U_{\{3\}}, U_{\{1,3\},\{2\}}, U_{\{1,3\},\{2,3\}}; X, Z|U_{\{1\}}, U_{\{2,3\},\{1\}}) \quad \text{(A.16)}$$

Define the following:

$$N_\delta^1 \triangleq X + h_{\{1\}}(U_{\{1\}}, U_{\{2,3\},\{1\}}) \quad \text{(A.17)}$$

$$N_{\delta*\delta}^3 \triangleq X + Z + h_{\{3\}}(U_{\{3\}}) \quad \text{(A.18)}$$

$$N_\delta^{1,3} \triangleq Z + h_{\{1,3\}}(U_{\{1\}}, U_{\{2,3\},\{1\}}, U_{\{3\}}, U_{\{1,3\},\{2\}}, U_{\{1,3\},\{2,3\}}), \quad \text{(A.19)}$$

where $h_{\{1\}}$ is the reconstruction of X at decoder $\{1\}$, $h_{\{3\}}$ is the reconstruction of $X + Z$ at decoder $\{3\}$, and $h_{\{1,3\}}$ is the reconstruction of Z at decoder $\{1, 3\}$. Then from (A.16):

$$\begin{aligned}
&R_3 \geq I(h_{\{1,3\}}(.), h_{\{3\}}(U_{\{3\}}); X, Z|U_{\{1\}}, U_{\{2,3\},\{1\}} h_{\{1\}}(.)) \\
&\Rightarrow R_3 \geq I(Z + N_\delta^{1,3}, X + Z + N_{\delta*\delta}^3; X, Z|U_{\{1\}}, U_{\{2,3\},\{1\}}, X + N_\delta^{\{1\}})) \\
&\Rightarrow R_3 \geq H(Z|U_{\{1\}}, U_{\{2,3\},\{1\}}, X + N_\delta^{\{1\}}) \\
&- H(Z|Z + N_\delta^{1,3}, X + Z + N_{\delta*\delta}^3, U_{\{1\}}, U_{\{2,3\},\{1\}}, X + N_\delta^{\{1\}}) \\
&\Rightarrow R_3 \overset{(a)}{\geq} 1 - H(Z|Z + N_\delta^{1,3}, X + Z + N_{\delta*\delta}^3, U_{\{1\}}, U_{\{2,3\},\{1\}}, X + N_\delta^{\{1\}}) \\
&\Rightarrow R_3 \geq 1 - H(Z|Z + N_\delta^{1,3}) \\
&\Rightarrow R_3 \geq 1 - H(N_\delta^{1,3}|Z + N_\delta^{1,3}) \\
&\overset{(b)}{\Rightarrow} R_3 \geq 1 - h_b(\delta)
\end{aligned}$$

All the above inequalities must be equality. In particular we have:

$$(a) \Rightarrow Z \leftrightarrow Z + N_{\delta}^{1,3} \leftrightarrow X + Z + N_{\delta*\delta}^{3}, X + N_{\delta}^{\{1\}}$$
$$\Rightarrow N_{\delta}^{1,3} \leftrightarrow Z + N_{\delta}^{1,3} \leftrightarrow Z + N_{\delta*\delta}^{3} + N_{\delta}^{\{1\}}$$
$$\Rightarrow N_{\delta}^{1,3} \leftrightarrow Z + N_{\delta}^{1,3} \leftrightarrow N_{\delta}^{1,3} + N_{\delta*\delta}^{3} + N_{\delta}^{\{1\}}$$

Note that from (b), we can conclude that Z is independent of $N_{\delta}^{1,3}$, we have $N_{\delta}^{1,3}$ and $N_{\delta}^{1,3} + N_{\delta*\delta}^{3} + N_{\delta}^{\{1\}}$ are independent. Define $N' \triangleq N_{\delta*\delta}^{3} + N_{\delta}^{\{1\}}$. We have:

$$P(N_{\delta}^{\{1,3\}} + N' = 0) \stackrel{(a)}{=} P(N_{\delta}^{\{1,3\}} + N' = 0|N_{\delta}^{\{1,3\}} = 0) = P(N' = 0|N_{\delta}^{\{1,3\}} = 0)$$
$$\stackrel{(b)}{\Rightarrow} P(N' = 0, N_{\delta}^{\{1,3\}} = 0) = (1-\delta)(P(N' = 0, N_{\delta}^{\{1,3\}} = 0) + P(N' = 1, N_{\delta}^{\{1,3\}} = 1))$$
$$\Rightarrow P(N' = 0, N_{\delta}^{\{1,3\}} = 0) = \frac{1-\delta}{\delta} P(N' = 1, N_{\delta}^{\{1,3\}} = 1)$$

(a) holds since $N_{\delta}^{1,3}$ and $N_{\delta}^{1,3} + N_{\delta*\delta}^{3} + N_{\delta}^{\{1\}}$ are independent. In (b) we have replaced $P(N_{\delta}^{\{1,3\}} + N' = 0)$ by $P(N' = 0, N_{\delta}^{\{1,3\}} = 0) + P(N' = 1, N_{\delta}^{\{1,3\}} = 1)$.

Define $a \triangleq P(N' = 1, N_{\delta}^{\{1,3\}} = 1)$, then by the same calculations $P(N' = 1, N_{\delta}^{\{1,3\}} = 0) = (1-\delta)(1 - \frac{1}{\delta}a)$, so $P(N' = 1) = 1 - \delta + \frac{2\delta-1}{\delta}a$. Note $a = P(N' = 1, N_{\delta}^{\{1,3\}} = 1) \leq P(N_{\delta}^{\{1,3\}} = 1) = \delta$, hence using $P(N' = 1) = 1 - \delta + \frac{2\delta-1}{\delta}a$, we get $P(N' = 1) \leq \delta$ with equality if and only if $a = \delta$. Also note that $Z + N'$ is available at decoder $\{1, 3\}$ so $P(N' = 1) = \delta$ and $a = \delta$, otherwise there is a contradiction with optimality of $h_{\{1,3\}}$. If $a = \delta$, then $N_{\delta}^{\{1,3\}}$ is equal to N'. So by the same arguments we have:

$$N_{\delta*\delta}^{\{3\}} = N_{\delta}^{\{1,3\}} + N_{\delta}^{\{1\}} = N_{\delta}^{\{2,3\}} + N_{\delta}^{\{2\}},$$

where

$$N_\delta^2 \triangleq Z + h_{\{2\}}(U_{\{2\}}, U_{\{1,3\},\{2\}}) \tag{A.20}$$

$$N_\delta^{2,3} \triangleq Z + h_{\{2,3\}}(U_{\{2\}}, U_{\{1,3\},\{2\}}, U_{\{3\}}, U_{\{2,3\},\{1\}}, U_{\{1,3\},\{2,3\}}) \tag{A.21}$$

Since $N_\delta^{\{1\}} \perp\!\!\!\perp N_\delta^{\{2\}}$, $N^{\{1\}} \perp\!\!\!\perp N^{\{1,3\}}$ and $N^{\{2\}} \perp\!\!\!\perp N^{\{2,3\}}$, we have:

$$N_\delta^{\{1,3\}} = N_\delta^2, N_\delta^{\{2,3\}} = N_\delta^{\{1\}}, N_{\delta*\delta}^{\{3\}} = N_\delta^1 + N_\delta^2$$

We argue that $C_{\{1,3\},\{2\}}, C_{\{2,3\},\{1\}}$ and $C_{\{1,3\}\{2,3\}}$ can be taken eliminated without any loss in RD. To prove this assume we have a scheme with $P_{U_{\{1,3\},\{2\}},U_{\{2,3\},\{1\}},U_{\{1,3\}\{2,3\}},U_{\{1\}},U_{\{2\}},U_{\{3\}}}$. Construct new random variables $\tilde{U}_{\{1\}} = X + N_\delta^{\{1\}}, U_{\{1\}}, U_{\{1\},\{2,3\}}$, $\tilde{U}_{\{2\}} = Z + N_\delta^{\{2\}}, U_{\{2\}}, U_{\{2\},\{1,3\}}$ and $\tilde{U}_3 = U_{\{3\}}$ and eliminate the rest of the codebooks. From the independence relations above, the packing bounds would stay the same. Since we have merged codebooks, the covering bounds would loosen, and it is straightforward to see that the reconstructions at each decoder are still the same. We are left with four codebooks, $C_{\{1\}}, C_{\{2\}}$ and $C_{\{3\}}$. Note that since decoder $\{1\}$ is only decoding $C_{\{1\}}$ we must have $\rho_{\{1\},1} = r_{\{1\}} = R_1$. This is deduced from the packing bound in decoder $\{1\}$:

$$H(U_{\{1\}}) \leq H(U_{\{1\}}) + \rho_{\{1\},1} - r_{\{1\}} \Rightarrow r_{\{1\}} \leq \rho_{\{1\},1}$$

But $\rho_{\{1\},1} \leq r_{\{1\}}$ so they are equal. The same argument gives $\rho_{\{2\},2} = r_{\{2\}} = R_2$, and $\rho_{\{3\},3} = r_{\{3\}} = R_3$. Also, from optimality at the joint decoders and lemma 3, we have $U_i \perp\!\!\!\perp U_j, \forall i \neq j$.

$$\begin{aligned}
&H(U_{\{1\}}, U_{\{2\}}, U_{\{3\}}|X, Z) \geq H(U_{\{1\}} + H(U_{\{2\}} + H(U_{\{3\}} - R_1 - R_2 - R - 3 \\
&\Rightarrow I(U_{\{1\}}, U_{\{2\}}, U_{\{3\}}; X, Z) + I(U_{\{3\}}; X, Z, U_{\{1\}}, U_{\{2\}}) \leq R_1 + R_2 + R_3 \\
&\Rightarrow I(U_{\{3\}}; X, Z, U_{\{1\}}, U_{\{2\}}) \leq R_3
\end{aligned} \tag{A.22}$$

Note that $R_1 + R_3 = I(U_{\{1\}}, U_{\{3\}}; X, Z)$ and $R_1 = I(U_{\{1\}}; X)$ from optimality at decoders $\{1\}$ and $\{1,3\}$. So $R_3 = I(U_{\{3\}}; X, Z|U_{\{1\}})$. Replacing R_3 into (A.22), we get $I(U_{\{3\}}; U_{\{2\}}|U_{\{1\}}, X, Z) = 0$. So we have the Markov chain $U_{\{3\}} \leftrightarrow U_{\{1\}}, X, Z \leftrightarrow U_{\{2\}}$. By the same arguments we can derive the Markov chain $U_{\{3\}} \leftrightarrow U_{\{2\}}, X, Z \leftrightarrow U_{\{1\}}$. Using lemma 7 and the previous two Markov chains we get $U_{\{3\}} \leftrightarrow X, Z \leftrightarrow U_{\{1\}}, U_{\{2\}}$. Take the Markov chain $U_{\{3\}} \leftrightarrow X, Z \leftrightarrow U_{\{1\}}$, along with $Z \perp\!\!\!\perp X, U_{\{1\}}$ we get $U_{\{3\}}, Z \leftrightarrow X \leftrightarrow U_{\{1\}}$. Also from the optimality of the reconstruction of X at decoders $\{1\}$ and $\{1,3\}$, we have:

$$I(U_{\{1\}}; X) = I(U_{\{1\}}, U_{\{3\}}; X) \Rightarrow I(U_{\{3\}}; X|U_{\{1\}}) = 0.$$

From the above and $Z \perp\!\!\!\perp X, U_{\{1\}}$, we conclude $U_{\{3\}}, Z \leftrightarrow U_{\{1\}} \leftrightarrow X$. Applying Lemma 7 we get $Z, U_{\{3\}} \perp\!\!\!\perp X, U_{\{1\}}$.

Lemma 7. *Let A,B,C and D be RV's such that $A \leftrightarrow B, C \leftrightarrow D$ and $A \leftrightarrow B, D \leftrightarrow C$, and also assume there is no $b \in \mathcal{B}$ for which given $B = b$ there are non-constant functions $f_b(C)$ and $g_b(D)$ with $f_b(C) = g_b(D)$ with probability 1. Then $A \leftrightarrow B \leftrightarrow C, D$.*

Proof. This lemma is a generalization of the one in [50]. We need to show that $p(A = a|B = b, C = c, D = d) = p(A = a|B = b, C = c', D = d')$ for any a, b, c, c', d, d'. Note since functions f_b and g_b do not exist, it is straightforward to show that there is a finite sequence of pairs (c_i, d_i) such that $(c_1, D_{\{1\}}) = (c, d)$ and $(c_n, d_n) = (c', d')$ with the property that either $c_i = c_{i+1}$ or $d_i = d_{i+1}$ and that $p(B = b, C = c_i, D = d_i) \neq 0$. Then from the first Markov chain if $d_i = d_{i+1}$, we have $p(A = a|B = b, C = c_i, D = d_i) = p(A = a|B = b, C = c_{i+1}, D = d_{i+1})$, also if $c_i = c_{i+1}$ the second Markov chain gives this result. So $p(A = a|B = b, C = c_i, D = d_i)$ is constant on all of the sequence particularly $p(A = a|B = b, C = c, D = d) = p(A = a|B = b, C = c', D = d')$. □

Let $g(U_{\{1\}}, U_{\{3\}})$ be the reconstruction of Z at decoder $\{1,3\}$. We have:

$$\sum_{z,u_{\{1\}},u_{\{3\}}} p(z,u_{\{1\}},u_{\{3\}})d_H(g(u_{\{1\}},u_{\{3\}}),z) \leq \delta \Rightarrow \sum_{u_{\{1\}}} p(u_{\{1\}}) \sum_{z,u_{\{3\}}} p(z,u_{\{3\}})d_H(g(u_{\{1\}},u_{\{3\}}),z) \leq \delta$$

So there is at least one $u_{\{1\}} \in \mathsf{U}_{\{1\}}$ such that $\sum_{z,u_{\{3\}}} p(z,u_{\{3\}})d_H(g(u_{\{1\}},u_{\{3\}},z) \leq \delta$. Let $g_{u_{\{1\}}}(U_{\{3\}}) = g(u_{\{1\}}, U_{\{3\}})$ be the reconstruction of Z using $U_{\{3\}}$. By the same argument we can find a reconstruction of X using $U_{\{3\}}$, then $I(U_{\{3\}}; X, Z) \geq 2(1 - h_b(\delta))$ from a PtP perspective which is a contradiction. □

A.2.3 Proof of Lemma 35

Proof. We provide an outline of the proof here, the arguments are similar to the ones in the previous proofs.

Step 1: I Any codebook which is not decoded at decoders $\{1\}$, $\{1,2\}$, $\{2,3\}$, $\{3,4\}$ and $\{4\}$ is redundant. This implies that there are at most only 17 codebooks which are non-redundant. These codebooks are $C_{\{1\}}$, $C_{\{1\},\{2,3\}}$, $C_{\{1\},\{3,4\}}$, $C_{\{1\},\{4\}}$,$C_{\{1\},\{2,3\},\{3,4\}}$, $C_{\{1\},\{4\},\{2,3\}}$, $C_{\{4\}}$, $C_{\{4\},\{2,3\}}$, $C_{\{4\},\{1,2\}}$, $C_{\{4\},\{2,3\},\{1,2\}}$,$C_{\{1,2\}}$,$C_{\{2,3\}}$, $C_{\{3,4\}}$, $C_{\{1,2\},\{2,3\}}$,$C_{\{1,2\},\{3,4\}}$, $C_{\{2,3\},\{3,4\}}$ and $C_{\{1,2\},\{2,3\},\{3,4\}}$.

Step 2: In this step we prove that the only non-trivial codebook decoded at decoder $\{i\}$ is $C_{\{i\}}$ for $i = 1,4$. All possible codebooks decoded at decoder $\{1\}$ are $C_{\{1\}}$, $C_{\{1\},\{2,3\}}$, $C_{\{1\},\{3,4\}}$,$C_{\{1\},\{4\}}$, $C_{\{1\},\{2,3\},\{3,4\}}$ and $C_{\{1\},\{2,3\},\{4\}}$. From optimality at decoder $\{1,2\}$, $C_{\{1\},\{2,3\}}$ is redundant. The reason is $\rho_{\{1\},\{2,3\},2} = 0$ otherwise we can set it to zero without any loss in distortion at decoder $\{1,2\}$ which contradicts optimality, also any random variable that description $\{3\}$ carries must be used in reconstructing Z at decoder $\{3,4\}$ because that decoder is at optimality, which means $\rho_{\{1\},\{2,3\},3} = 0$ so the codebook is decoded at decoder $\{2,3\}$ but not sent through either description $\{2\}$ or $\{3\}$, from similar arguments as before the codebook is redundant. Same arguments can be provided to deduce redundancy of $C_{\{1\},\{3,4\}}$, $C_{\{2,3\}}$, $C_{\{1\},\{2,3\},\{3,4\}}$ and $C_{\{1\},\{2,3\},\{4\}}$. This implies that only $C_{\{1\}}$ is decoded at decoder $\{1\}$ and $C_{\{4\}}$ at decoder $\{4\}$.

Step 3: We proceed with eliminating $C_{\{1,2\},\{3,4\}}$ and $C_{\{1,2\},\{2,3\},\{3,4\}}$. Using the PtP optimality of decoder $\{1,2\}$ we have:

$$
\begin{aligned}
&I(U_{\{1\}}, U_{\{1,2\}}, U_{\{1,2\},\{2,3\}}, U_{\{1,2\},\{3,4\}}, U_{\{1,2\},\{2,3\},\{3,4\}}; X) = \\
&R_1 + R_2 \overset{(a)}{\geq} I(U_{\{1\}}, U_{\{1,2\}}, U_{\{1,2\},\{2,3\}}, U_{\{1,2\},\{3,4\}}, U_{\{1,2\},\{2,3\},\{3,4\}}; X, Z)
\end{aligned}
$$

where (a) follows from the usual PtP source coding results. Comparing the LHS with the RHS we conclude the Markov chain $U_{\{1\}}, U_{\{1,2\}}, U_{\{1,2\},\{2,3\}}, U_{\{1,2\},\{3,4\}}, U_{\{1,2\},\{2,3\},\{3,4\}} \leftrightarrow X \leftrightarrow Z$. In particular we are interested in $U_{\{1,2\},\{3,4\}}, U_{\{1,2\},\{2,3\},\{3,4\}} \leftrightarrow X \leftrightarrow Z$. By the same arguments and using the optimality at decoder $\{3,4\}$, we get $U_{\{1,2\},\{3,4\}}, U_{\{1,2\},\{2,3\},\{3,4\}} \leftrightarrow Z \leftrightarrow X$. These two Markov chains along with lemma 7 prove $U_{\{1,2\},\{3,4\}}, U_{\{1,2\},\{2,3\},\{3,4\}} \perp\!\!\!\perp X, Z$. So these two variables are not used in reconstructing the source and the corresponding codebooks are eliminated.

Step 4: The only remaining codebooks are $C_{\{1\}}$, $C_{\{4\}}$, $C_{\{1,2\}}$, $C_{\{3,4\}}$, $C_{\{1,2\},\{2,3\}}$ and $C_{\{2,3\},\{3,4\}}$. From optimality at decoders $\{1,2\}$ and $\{3,4\}$ we must have $U_{\{1\}}, U_{\{1,2\}}, U_{\{1,2\},\{2,3\}} \leftrightarrow (X,Z) \leftrightarrow U_{\{4\}}, U_{\{3,4\}}, U_{\{2,3\},\{3,4\}}$, also $U_{\{1\}}, U_{\{1,2\}}, U_{\{1,2\},\{2,3\}} \leftrightarrow X \leftrightarrow Z$ and $X \leftrightarrow Z \leftrightarrow U_{\{4\}}, U_{\{3,4\}}, U_{\{2,3\},\{3,4\}}$. From lemma 8, we get $U_{\{1\}}, U_{\{1,2\}}, U_{\{1,2\},\{2,3\}} \leftrightarrow X \leftrightarrow Z \leftrightarrow U_{\{4\}}, U_{\{3,4\}}, U_{\{2,3\},\{3,4\}}$.

Lemma 8. *For random variables A,B,C,D, the three short Markov chains $A \leftrightarrow (B,C) \leftrightarrow D$, $A \leftrightarrow B \leftrightarrow C$ and $B \leftrightarrow C \leftrightarrow D$ are equivalent to the long Markov chain $A \leftrightarrow B \leftrightarrow C \leftrightarrow D$.*

Proof. We only need to show that $A \leftrightarrow B \leftrightarrow D$, the rest of the implications of the long Markov chain are either direct results of the three short Markov chains or follow by symmetry. For arbitrary a, b, d we have:

$$P(D = d|B = b, A = a) = \sum_{c \in C} P(C = c|B = b, A = a)P(D = d|A = a, B = b, C = c)$$
$$= \sum_{c \in C} P(C = c|B = b)P(D = d|B = b, C = c) = P(D = d|B = b)$$

□

We get an inner bound for $R_2 + R_3$ at decoder $\{2, 3\}$:

$$R_2 + R_3 \geq \min I(U, V; X, Z) = H(X, Z) = 1 + h_b(p),$$

where the minimum is taken over all $P_{U,V|X,Z}$ for which the long Markov chain $U \leftrightarrow X \leftrightarrow Z \leftrightarrow V$ is satisfied and (U, V) produce a lossless reconstruction of $X + Z$. This resembles the distributed source coding problem in [18]. So the RD vector can't be achieved using random codes. □

A.2.4 Proof of Lemma 39

Proof. In this proof we use bold letters to denote vectors and matrices. Fix integers n, l, l' and k. Choose the elements of the matrices $\mathbf{\Delta G}_{l \times n}$, $\mathbf{\Delta G}'_{k' \times n}$ and $\mathbf{G}_{k \times n}$ and vectors $\mathbf{B}^n$ and $\mathbf{B}'^n$ randomly and uniformly from $\mathbb{F}_q$. The codebooks C_o^n and $C_o'^n$ are defined as follows:

$$C_o = \{\mathbf{aG} + \mathbf{m\Delta G} + \mathbf{B} | \mathbf{a} \in \mathbb{F}_q^k, \mathbf{m} \in \mathbb{F}_q^l\}$$
$$C'_o = \{\mathbf{bG} + \mathbf{m}'\mathbf{\Delta G}' + \mathbf{B}' | \mathbf{b} \in \mathbb{F}_q^k, \mathbf{m}' \in \mathbb{F}_q^{l'}\}$$

For a typical sequence $\mathbf{x}$ with respect to P_X, we define $\theta(\mathbf{x})$ as the function which counts the number of codewords in C_o and C'_o jointly typical with respect to P_{XUV}:

$$\begin{aligned}\theta(\mathbf{x}) &= \sum_{\mathbf{u}\in C'_o, \mathbf{v}\in C_o} \mathbb{I}\{(\mathbf{u},\mathbf{v}) \in A_\epsilon^n(U,V|\mathbf{x})\} \\ &= \sum_{\mathbf{m},\mathbf{m}'} \sum_{\mathbf{a},\mathbf{b}\in\mathbb{F}_q^k} \sum_{(\mathbf{u},\mathbf{v})\in A_\epsilon^n(U,V|\mathbf{x})} \mathbb{I}\{\mathbf{aG} + \mathbf{m\Delta G} + \mathbf{B} = \mathbf{u}, \mathbf{bG} + \mathbf{m}'\mathbf{\Delta G}' + \mathbf{B}' = \mathbf{v}\}\end{aligned}$$

Our goal is to find bounds on n, l, l' and k such that $P(\theta(\mathbf{x}) = 0) \to 0$ as $n \to \infty$.

For $\mathbf{a} \in \mathbb{F}_q^k$ and $\mathbf{m} \in \mathbb{F}_q^l$, we denote the corresponding codeword as $g(\mathbf{a},\mathbf{m}) := \mathbf{aG} + \mathbf{m\Delta G} + \mathbf{B}$. Similarly define $g'(\mathbf{b},\mathbf{m}') := \mathbf{bG} + \mathbf{m}'\mathbf{\Delta G}' + \mathbf{B}'$ for any $\mathbf{b} \in \mathbb{F}_q^k$ and $\mathbf{m}' \in \mathbb{F}_q^{l'}$. The following lemma proves several results on the pairwise independence of the codewords.

Lemma 9. *The following hold:*

1. $g(\mathbf{a},\mathbf{m})$ *and* $g'(\mathbf{b},\mathbf{m}')$ *are distributed uniformly uniform over* $\mathbb{F}_q^n$.
2. *If* $\mathbf{a} \neq \tilde{\mathbf{a}}$, *then* $g(\mathbf{a},\mathbf{m})$ *is independent of* $g(\tilde{\mathbf{a}},\mathbf{m})$.
3. *If* $\mathbf{b} \neq \tilde{\mathbf{b}}$, *then* $g'(\mathbf{b},\mathbf{m}')$ *is independent of* $g'(\tilde{\mathbf{b}},\mathbf{m}')$.
4. *If* $\mathbf{B}$ *and* $\mathbf{B}'$ *are chosen independently and uniformly over* $\mathbb{F}_q^n$, *then* $g(\mathbf{b},\mathbf{m}')$ *and* $g'(\mathbf{a},\mathbf{m})$ *are independent.*

Proof. Follows from [29], and the fact that $\mathbf{B}, \mathbf{B}'$ are independent and uniform. □

We intend to use Chebyshev's inequality to obtain:

$$P\{\theta(\mathbf{X}) = 0\} \leq \frac{4var\{\theta(\mathbf{X})\}}{\mathsf{E}\{\theta(\mathbf{X})\}^2} \to 0$$

Lemma 10. *For* $\mathbf{X} \in A_\epsilon^{(n)}(X)$ *we have the following bound on* $\frac{var\{\theta(\mathbf{X})\}}{\mathsf{E}\{\theta(\mathbf{X})\}^2}$*:*

$$\begin{aligned}\frac{var\{\theta(X)\}}{\mathsf{E}\{\theta(X)\}^2} \leq & \frac{q^{2n}}{q^{l+l'}q^{2k}}2^{-n(H(U,V|X)} + \frac{q^n}{q^{l+l'}q^k}2^{-n(H(U|X))} \\ & + \frac{q^n}{q^{l+l'}q^k}2^{-n(H(V|X))} + \frac{q^n}{q^{l+l'}q^k}2^{-n(H(U,V|X)-\max_{i\neq 0} H(U,V|X,V+iU))} \\ & + \frac{q^n}{q^l q^k}2^{-n(H(U|X))} + \frac{q^n}{q^{l'}q^k}2^{-n(H(V|X))} + \frac{1}{q^l} + \frac{1}{q^{l'}} + \frac{1}{q^{l+l'}}. + \frac{1}{q^{l+k}} + \frac{1}{q^{l'+k}}\end{aligned}$$

Proof. We calculate the expected value of $\theta(\mathbf{X})$ for any $\mathbf{X} \in A_\epsilon^{(n)}(X)$:

$$\begin{aligned}\mathsf{E}\{\theta(\mathbf{X})\} &= \sum_{\mathbf{x}\in A_\epsilon^n(X)} \sum_{\substack{\mathbf{m}\in\mathbb{F}_q^l \\ \mathbf{m}'\in\mathbb{F}_q^{l'}}} \sum_{\mathbf{a}\neq\mathbf{b}} \sum_{(\mathbf{u},\mathbf{v})\in A_\epsilon^n(U,V|\mathbf{x})} P(\mathbf{x}) \quad P\{g(\mathbf{a},\mathbf{m}) = \mathbf{u}, g'(\mathbf{b},\mathbf{m}') = \mathbf{v}\} \\ &= \sum_{\mathbf{m},\mathbf{m}'} \sum_{\mathbf{x}\in A(X)} \sum_{\mathbf{a}\neq\mathbf{b}} |A_\epsilon^n(U,V|\mathbf{x})| P(\mathbf{x}) \frac{1}{q^{2n}} = \frac{q^{l+l'}q^{2k}}{q^{2n}} 2^{n(H(U,V|X)+O(\epsilon))}\end{aligned}$$

Also:

$$\mathsf{E}\{\theta(\mathbf{X})^2\} = \sum_{\substack{\mathbf{m},\tilde{\mathbf{m}} \\ \mathbf{m}',\tilde{\mathbf{m}}'}} \sum_{\mathbf{a},\tilde{\mathbf{a}}} \sum_{\mathbf{b},\tilde{\mathbf{b}}} \sum_{(\mathbf{u},\mathbf{v})} \sum_{(\tilde{\mathbf{u}},\tilde{\mathbf{v}})\in A_\epsilon^n(U,V|\mathbf{x})} P\{g(\mathbf{a},\mathbf{m}) = \mathbf{u}, g(\tilde{\mathbf{a}},\tilde{\mathbf{m}}) = \tilde{\mathbf{u}}, g'(\mathbf{b},\mathbf{m}') = \mathbf{v}, g'(\tilde{\mathbf{b}},\tilde{\mathbf{m}}') = \tilde{\mathbf{v}}\} \tag{A.23}$$

Using Lemma 9:

$$\begin{aligned}P_S &\triangleq P\{g(\mathbf{a},\mathbf{m}) = \mathbf{u}, g(\tilde{\mathbf{a}},\tilde{\mathbf{m}}) = \tilde{u}, g(\mathbf{b},\mathbf{m}') = \mathbf{v}, g(\tilde{\mathbf{b}},\tilde{\mathbf{m}'}) = \tilde{\mathbf{v}}\} \\ &= \frac{1}{q^{2n}} \times P\{g_0(\mathbf{a}-\tilde{\mathbf{a}},\mathbf{m}-\tilde{\mathbf{m}}) = \mathbf{u}-\tilde{\mathbf{u}}, g_0'(\mathbf{b}-\tilde{\mathbf{b}},\mathbf{m}'-\tilde{\mathbf{m}}') = \mathbf{v}-\tilde{\mathbf{v}}\}\end{aligned}$$

At this point we have to consider several different cases for the values of $\mathbf{a}, \tilde{\mathbf{a}}, \mathbf{b}, \tilde{\mathbf{b}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{m}', \tilde{\mathbf{m}}'$.

1) $\mathbf{m} = \tilde{\mathbf{m}}, \mathbf{m}' = \tilde{\mathbf{m}}'$

1.1: $\mathbf{a} = \tilde{\mathbf{a}}, \mathbf{b} = \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{2n}}\delta(\mathbf{u}-\tilde{\mathbf{u}})\delta(\mathbf{v}-\tilde{\mathbf{v}})$

1.2: $\mathbf{a} = \tilde{\mathbf{a}}, \mathbf{b} \neq \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{3n}}\delta(\mathbf{u}-\tilde{\mathbf{u}})$

1.3: $\mathbf{a} \neq \tilde{\mathbf{a}}, \mathbf{b} = \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{3n}}\delta(\mathbf{v}-\tilde{\mathbf{v}})$

1.4: $\mathbf{a} \neq \tilde{\mathbf{a}}, \mathbf{b} \neq \tilde{\mathbf{b}} \Rightarrow P_s = \sum_{\alpha \in \mathbb{F}_q} \frac{1}{q^{3n}} \delta(\mathbf{u} - \tilde{\mathbf{u}} - \alpha(\mathbf{v} - \tilde{\mathbf{v}})) + \frac{1}{q^{4n}}(1 - \sum_{\alpha \in \mathbb{F}_q} \delta(\mathbf{u} - \tilde{\mathbf{u}} - \alpha(\mathbf{v} - \tilde{\mathbf{v}})))$

2) $\mathbf{m} \neq \tilde{\mathbf{m}}, \mathbf{m}' = \tilde{\mathbf{m}}'$

2.1: $\mathbf{a} = \tilde{\mathbf{a}}, \mathbf{b} = \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{3n}} \delta(\mathbf{v} - \tilde{\mathbf{v}})$

2.2: $\mathbf{a} = \tilde{\mathbf{a}}, \mathbf{b} \neq \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{4n}}$

2.3: $\mathbf{a} \neq \tilde{\mathbf{a}}, \mathbf{b} = \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{3n}} \delta(\mathbf{v} - \tilde{\mathbf{v}})$

2.4: $\mathbf{a} \neq \tilde{\mathbf{a}}, \mathbf{b} \neq \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{4n}}$

Cases when $\mathbf{m} = \tilde{\mathbf{m}}, \mathbf{m}' \neq \tilde{\mathbf{m}}'$ and $\mathbf{m} \neq \tilde{\mathbf{m}}, \mathbf{m}' \neq \tilde{\mathbf{m}}'$ are similarly considered but the derivations are omitted for brevity. Considering cases 1.1−4:

$$\mathsf{E}\{\theta(\mathbf{x})^2 | \mathbf{m} = \tilde{\mathbf{m}}, \mathbf{m}' = \tilde{\mathbf{m}}'\} = \sum_{\mathbf{m},\mathbf{m}'} \Bigg[\sum_{\mathbf{a}=\tilde{\mathbf{a}}} \sum_{\mathbf{b}=\tilde{\mathbf{b}}} \sum_{(\mathbf{u},\mathbf{v}) \in A_\epsilon^n(U,V|\mathbf{x})} \frac{1}{q^{2n}} + \sum_{\mathbf{a}=\tilde{\mathbf{a}}} \sum_{\mathbf{b}\neq\tilde{\mathbf{b}}} \sum_{(\mathbf{u},\mathbf{v}),(\mathbf{u},\tilde{\mathbf{v}})} \frac{1}{q^{3n}} \tag{A.24}$$

$$+ \sum_{\mathbf{a}\neq\tilde{\mathbf{a}}} \sum_{\mathbf{b}=\tilde{\mathbf{b}}} \sum_{(\mathbf{u},\mathbf{v}),(\tilde{\mathbf{u}},\mathbf{v})} \frac{1}{q^{3n}} + \sum_{\alpha \in \mathbb{F}_q \setminus \{0\}} \sum_{\mathbf{a}\neq\tilde{\mathbf{a}}} \sum_{\substack{\mathbf{b}\neq\tilde{\mathbf{b}} \\ \mathbf{b}-\tilde{\mathbf{b}}=\alpha(\mathbf{a}-\tilde{\mathbf{a}}) \\ \alpha\in\mathbb{F}_q\setminus 0}} \sum_{\substack{(\mathbf{u},\mathbf{v}),(\tilde{\mathbf{u}},\tilde{\mathbf{v}}) \\ \mathbf{v}-\tilde{\mathbf{v}}=\alpha(\mathbf{u}-\tilde{\mathbf{u}})}} \frac{1}{q^{3n}} + \sum_{\alpha \in \mathbb{F}_q - \{0\}} \sum_{\mathbf{a}\neq\tilde{\mathbf{a}}} \sum_{\substack{\mathbf{b}\neq\tilde{\mathbf{b}} \\ \mathbf{b}-\tilde{\mathbf{b}}\neq\alpha(\mathbf{a}-\tilde{\mathbf{a}}) \\ \alpha\in\mathbb{F}_q\setminus 0}} \sum_{\substack{(\mathbf{u},\mathbf{v}),(\tilde{\mathbf{u}},\tilde{\mathbf{v}}) \\ \mathbf{v}-\tilde{\mathbf{v}}\neq\alpha(\mathbf{u}-\tilde{\mathbf{u}})}} \frac{1}{q^{4n}} \Bigg] \tag{A.25}$$

Consequently:

$$\begin{aligned} &\mathsf{E}\{\theta(\mathbf{X})^2 | \mathbf{m}' = \tilde{\mathbf{m}}, \mathbf{m}' = \tilde{\mathbf{m}}'\} \\ &\quad \leq \frac{q^{l+l'} q^{2k}}{q^{2n}} 2^{n(H(U,V|X))} + \frac{q^{l+l'} q^{3k}}{q^{3n}} 2^{n(H(U,V|X)+H(V|X,U))} + \frac{q^{l+l'} q^{3k}}{q^{3n}} 2^{n(H(U,V|X)+H(U|X,V))} + \\ &\quad \frac{q^{l+l'} q^{3k}}{q^{3n}} 2^{n(H(U,V|X)+\max_{\alpha\neq 0} H(U,V|X,V+\alpha U))} + \frac{q^{l+l'} q^{4k}}{q^{4n}} 2^{2n(H(U,V|X))}, \end{aligned}$$

where we have used Lemma 8 in [47] to get the fourth term. After considering all the cases, the only non-redundant bounds are the ones mentioned in the lemma. □

So, the following bounds need to be satisfied:

$$\begin{aligned}
r_o + r'_o &\geq 2\log q - H(U, V|X)\\
r_o + r'_o - r_i &\geq \log q - \min\{H(U|X), H(V|Z)\}\\
r_o + r'_o - r_i &\geq \log q - H(U, V|X) + \max_{\alpha\neq 0} H(U, V|X, V + \alpha U)\\
r_o &\geq \log q - H(U|X))\\
r'_o &\geq \log q - H(V|X))\\
\min\{r_o, r'_o\} &\geq r_i
\end{aligned}$$

Observe that

$$H(U, V|X, V+\alpha U) = H(U, V, V+\alpha U|X) - H(V+\alpha U|X) = H(U, V|X) - H(V + \alpha U|X)$$

$\square$

A.2.5 Proof of Lemma 42

Proof. The proof follows the same arguments as that of Lemma 39. We provide an outline of the proof. Define the probability of error P_e as follows:

$$P_e = P(\{(\mathbf{x}, \mathbf{u}, \mathbf{v}) \in \mathsf{X}^n \times C_1 \times C_2 | \exists(\mathbf{u}', \mathbf{v}') \in A_\epsilon^n(U, V) \cap B_2(\mathbf{u}) \times B_2(\mathbf{v})\})$$

We define a new conditional probability of error for any triple $\mathbf{x}, \mathbf{u}, \mathbf{v} \in A_\epsilon(X, U, V)$:

$$P_{e|\mathbf{x},\mathbf{u},\mathbf{v}} = P(\exists(\mathbf{u}', \mathbf{v}') \in A_\epsilon^n(U, V) \cap B_2(\mathbf{u}) \times B_2(\mathbf{v}) | \mathbf{X} = \mathbf{x}, (\mathbf{u}, \mathbf{v}) \in C_1 \times C_2)$$

Clearly if $P_{e|\mathbf{x},\mathbf{u},\mathbf{v}}$ goes to 0 for all $\mathbf{x},\mathbf{u},\mathbf{v} \in A_\epsilon(X,U,V)$ as $n \to \infty$, then P_e goes to 0. Also define: $P_{\mathbf{x},\mathbf{u},\mathbf{v}} = P((\mathbf{x},\mathbf{u},\mathbf{v}) \in \mathsf{X}^n \times C_1 \times C_2)$, and $P_{e,\mathbf{x},\mathbf{u},\mathbf{v}} = P_{e|\mathbf{x},\mathbf{u},\mathbf{v}} P_{\mathbf{x},\mathbf{u},\mathbf{v}}$. We have:

$$\begin{aligned}
P_{\mathbf{x},\mathbf{u},\mathbf{v}} &= \sum_{\mathbf{x}\in A_\epsilon^n(X)} \sum_{\substack{\mathbf{m}\in\mathbb{F}_q^l \\ \mathbf{m}'\in\mathbb{F}_q^{l'}}} \sum_{\mathbf{a}\neq\mathbf{b}} \sum_{(\mathbf{u},\mathbf{v})\in A_\epsilon^n(U,V|\mathbf{x})} P(\mathbf{x}) P\{g(\mathbf{a},\mathbf{m}) = \mathbf{u}, g'(\mathbf{b},\mathbf{m}') = \mathbf{v}\} \\
&= \sum_{\mathbf{m},\mathbf{m}'} \sum_{\mathbf{x}\in A(X)} \sum_{\mathbf{a}\neq\mathbf{b}} |A_\epsilon^n(U,V|\mathbf{x})| P(\mathbf{x}) \frac{1}{q^{2n}} = \frac{q^{l+l'} q^{2k}}{q^{2n}} 2^{n(H(U,V|X)+O(\epsilon))}
\end{aligned}$$

$$P_{e,\mathbf{x},\mathbf{u},\mathbf{v}} = \sum_{\substack{\mathbf{m},\tilde{\mathbf{m}} \\ \mathbf{m}',\tilde{\mathbf{m}}'}} \sum_{\mathbf{x}} \sum_{\mathbf{a},\tilde{\mathbf{a}}} \sum_{\mathbf{b},\tilde{\mathbf{b}}} \sum_{\substack{(\mathbf{u},\mathbf{v})\in \\ A_\epsilon^n(U,V|\mathbf{x})}} \sum_{\substack{(\tilde{\mathbf{u}},\tilde{\mathbf{v}})\in \\ A_\epsilon^n(U,V)}} \sum_{b_1\in[1,2^{n\rho_1}]} \sum_{b_2\in[1,2^{n\rho_2}]} P(\mathbf{x})$$

$$P\{g(\mathbf{a},\mathbf{m}) = \mathbf{u}, g(\tilde{\mathbf{a}},\tilde{\mathbf{m}}) = \tilde{\mathbf{u}}, g'(\mathbf{b},\mathbf{m}') = \mathbf{v}, g'(\tilde{\mathbf{b}},\tilde{\mathbf{m}}') = \tilde{\mathbf{v}}\} P\{B_1(\mathbf{u}) = B_1(\tilde{\mathbf{u}}) = b_1, B_2(\mathbf{u}) = B_2(\tilde{\mathbf{u}}) = b_2\}$$

Note that the binning is done independently and uniformly, so $P\{B_1(\mathbf{u}) = B_1(\tilde{\mathbf{u}}) = b_1, B_2(\mathbf{u}) = B_2(\tilde{\mathbf{u}}) = b_2\} = 2^{-2(\rho_1+\rho_2)}$. The rest of the summations are the ones which were present in the proof of Lemma 39. Again we have to do a case by case investigation of the summation. The only new bond comes from the case when $\mathbf{m} = \tilde{\mathbf{m}}$ and $\mathbf{m}' = \tilde{\mathbf{m}}'$, $a \neq \tilde{a}, b \neq \tilde{b}$ and $a - \tilde{a} = i(b - \tilde{b})$. We have:

$$\begin{aligned}
A &= \sum_{\mathbf{m},\tilde{\mathbf{m}}} \sum_{\substack{\mathbf{a},\tilde{\mathbf{a}} \\ \mathbf{a}\neq\tilde{\mathbf{a}}}} \sum_{\substack{\mathbf{b},\tilde{\mathbf{b}} \\ \mathbf{a}-\tilde{\mathbf{a}}=i(\mathbf{b}-\tilde{\mathbf{b}})}} \sum_{\substack{(\mathbf{u},\mathbf{v})\in \\ A_\epsilon^n(U,V|\mathbf{x})}} \sum_{\substack{(\tilde{\mathbf{u}},\tilde{\mathbf{v}})\in \\ A_\epsilon^n(U,V) \\ \mathbf{u}-\tilde{\mathbf{u}}=i(\mathbf{v}-\tilde{\mathbf{v}})}} q^{-3n} 2^{-n(\rho_1+\rho_2)} \\
&= \frac{q^{l+l'}}{q^{3n}} q^{3k} 2^{nH(U,V|X)} 2^{nH(U,V|U+iV)} 2^{-n(\rho_1+\rho_2)}
\end{aligned}$$

Dividing this last term by $P_{\mathbf{x},\mathbf{u},\mathbf{v}}$:

$$\frac{A}{P_{\mathbf{x},\mathbf{u},\mathbf{v}}} = \frac{q^k}{q^n} 2^{nH(U,V|U+iV)} 2^{-n(\rho_1+\rho_2)}$$

which goes to 0 if the following is satisfied:

$$r_i - \rho_1 - \rho_2 \leq \log q - H(U, V|U + iV) \tag{A.26}$$

However as shown in the next lemma the new bound in (A.26) is redundant.

Lemma 11. *The inequality (*A.*26) in Lemma 42 is redundant.*

Proof. Assume there is a distribution $P_{X,U,V}$ for which (A.26) is violated, we show that either (2.17) or (2.20) is also violated. Conversely, as long as (2.17) and (2.20) are satisfied, (A.26) is also satisfied. Assume we have:

$$\begin{aligned}
&(r_o - \rho_1) + (r'_o - \rho_2) \leq 2 \log q - H(U, V) \\
&r_i - \rho_1 - \rho_2 > \log q - H(U, V|U + iV), \forall i \in \mathbb{F}_q.
\end{aligned}$$

Adding the two bounds we get:

$$\begin{aligned}
r_o + r'_o - r_i &< \log q - H(U, V) + H(U, V|U + iV) \\
&= \log q - H(U + iV) \leq \log q - H(U + iV|X)
\end{aligned}$$

which contradicts (2.17).

□

□

A.2.6 Proof of Lemma 43

Proof. The proof follows the same arguments as in the previous two examples. First we assume there exists a joint distribution $P_{\mathbf{U}X}$ such that the SSC scheme achieves the RD vector, then we arrive at a contradiction by eliminating all codebooks. First note that from our definition of $P_{X,V_{\{1\}},V_{\{2\}}}$, direct calculation shows that $R_1 + R_2 = I(V_{\{1\}}, V_{\{2\}}; X) = 1 -$

$h_b(D_0)$. This means that decoder $\{1,2\}$ is at PtP optimality. Also by the definition of the distortion function $D_{\{3\}}$, decoder $\{3\}$ is at optimal RD.

Step 1: From the optimality of decoder $\{1,2\}$ and Lemma 3, there can't be any codebook common between decoders $\{1\}$ and $\{2\}$. So $C_{\{1\},\{2\}}$ and $C_{\{1\},\{2\},\{3\}}$ are eliminated.

Step 2: From optimality of decoder $\{3\}$, description 3 can't carry the bin number of any codebook which is not decoded at that decoder. Also description 1 and 2 can't carry the bin numbers of codebooks which are not decoded at $\{1,2\}$ because of optimality at this decoder. So codebooks $C_{\{1,3\},\{2,3\}}$, $C_{\{1,3\}}$ and $C_{\{2,3\}}$ are not sent on any description and are redundant.

Step 3: The codebook $C_{\{1\},\{2,3\}}$ is not binned by description 2 or 3. Description 3 can't bin the codebook since it is not decoded at decoder $\{3\}$, and that decoder is at PtP optimality. Note $C_{\{1\},\{2,3\}}$ can be decoded using description 1, so any bin information for this codebook that is carried by description 2 is not used at decoder $\{1,2\}$, since decoder $\{1,2\}$ is at PtP optimality we must have $\rho_{\{1\},\{2,3\},2} = 0$. The codebook is not sent on description 2 or 3, so by the same arguments as in the previous proofs it can't help in the reconstruction at decoder $\{2,3\}$ and is redundant. By the same arguments $C_{\{2\},\{1,3\}}$ is redundant.

Step 4: In this step we show that there is no refinement codebook decoded at decoder $\{1,2\}$. This would eliminate $C_{\{1,2\}}, C_{\{1,2\},\{3\}}, C_{\{1,2\},\{1,3\}}, C_{\{1,2\},\{2,3\}}$ and $C_{\{1,2\},\{1,3\},\{2,3\}}$. More precisely we show that the reconstruction at decoder $\{1,2\}$ is a function of the reconstructions at decoders $\{1\}$ and $\{2\}$. This means that sending a refinement codebook to decoder $\{1,2\}$ will not help in the reconstruction, so the codebook is redundant.

To prove this claim we consider the two user example depicted in Figure [54]. Here all distortions are Hamming distortions. We are interested in achieving the rate distortion vector $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$ given in (4.14). Let $P_{X,U_{\{1,2\}},U_{\{1\}},U_{\{2\}},U_{\{1\},\{2\}}}$ be a distribution on the random variables in the two user SSC achieving this RD vector. Define $\hat{X}_1, \hat{X}_2$ and $\hat{X}_{12}$ as the reconstructions at the corresponding codebooks.

Lemma 12. *There are only two choices for the joint distribution $P_{X,\hat{X}_1,\hat{X}_2,\hat{X}_{12}}$, furthermore in* *both choices, $\hat{X}_{12}$ is a function of $\hat{X}_1$ and $\hat{X}_2$.*

Proof. As in step 1, from optimality of decoder $\{1,2\}$, $C_{\{1,2\}}$ is redundant. Also $U_{\{1\}}$ and $U_{\{2\}}$ are independent from Lemma 3. Note that $\hat{X}_1$ is a function of $U_{\{1\}}$ and $\hat{X}_2$ is a function of $U_{\{2\}}$, so $\hat{X}_1 \perp\!\!\!\perp \hat{X}_2$. We proceed by characterizing $P_{X,\hat{X}_{12}}$. Note that decoder $\{1,2\}$ is at PtP optimality. It is well-known result that when quantizing a BSS to Hamming distortion D_0 with rate $1 - h_b(D_0)$, the reconstruction is uniquely given by $\hat{X}_{12} = X + \mathsf{N}_0, \mathsf{N}_0 \sim Be(D_0)$ where $\mathsf{N}_0 \perp\!\!\!\perp X$. $\hat{X}_1, \hat{X}_2$ and $\hat{X}_{12}$ are available at decoder $\{1,2\}$, from optimality at this decoder we must have:

$$1 - h_b(D_0) = I(\hat{X}_1, \hat{X}_2, \hat{X}_{1,2}; X) \geq I(\hat{X}_{12}, X) = 1 - h_b(D_0).$$

So the inequality must be equality, which means $I(\hat{X}_1, \hat{X}_2; X|X_{12}) = 0$. In other words the Markov chain $\hat{X}_1, \hat{X}_2 \leftrightarrow \hat{X}_{12} \leftrightarrow X$ must hold. Using the three facts 1) $\hat{X}_{12} = X \oplus_2 \mathsf{N}_0$, 2) $\hat{X}_1 \perp\!\!\!\perp \hat{X}_2$ and 3) $\hat{X}_1, \hat{X}_2 \leftrightarrow \hat{X}_{12} \leftrightarrow X$, we can characterize all possible distributions on $P_{X,\hat{X}_{12},\hat{X}_1,\hat{X}_2}$. Let $\hat{X}_1 \sim Be(a_1)$ and $\hat{X}_2 \sim Be(a_2)$. Then from $\hat{X}_1 \perp\!\!\!\perp \hat{X}_2$, $P_{\hat{X}_1,\hat{X}_2}$ is fixed. Assume the distribution $P_{\hat{X}_{12},\hat{X}_1,\hat{X}_2}$ is as given below: As shown on the table there are 5 independent

$\hat{X}_{12}$ \ $\hat{X}_1, \hat{X}_2$	00	01	10	11	Sum
0	P_{000}	P_{001}	P_{010}	P_{011}	$\frac{1}{2}$
1	P_{100}	P_{101}	P_{110}	P_{111}	$\frac{1}{2}$
Sum	$(1-a_1)(1-a_2)$	$(1-a_1)a_2$	$a_1(1-a_2)$	$a_1 a_2$	

Table A.1: Joint probability distribution of $P_{\hat{X}_{12},\hat{X}_1,\hat{X}_2}$.

linear constraints on P_{ijk}'s. We have:

$$P_{011} = \frac{1}{2} - P_{000} - P_{001} - P_{010}, \quad P_{100} = (1-a_1)(1-a_2) - P_{000}, \quad P_{101} = (1-a_1)a_2 - P_{001},$$
$$P_{110} = (1-a_1)a_2 - P_{010}, \quad P_{111} = a_1a_2 - \frac{1}{2} + P_{000} + P_{001} + P_{010}$$
$$a_1 \in [0,1], a_2 \in [0,1], P_{000} \in [0,(1-a_1)(1-a_2)], P_{001} \in [0,(1-a_1)a_2], P_{010} \in [0,a_1(1-a_2)]$$
$$P_{000} + P_{001} + P_{010} \in [\frac{1}{2} - a_1a_2, \frac{1}{2}]$$

Using the Markov chain $\hat{X}_1, \hat{X}_2 \leftrightarrow \hat{X}_{12} \leftrightarrow X$, we have $P_{X,\hat{X}_1,\hat{X}_2} = \sum_{\hat{x}_{12}} P_{X|\hat{X}_{12}} P_{\hat{X}_1,\hat{X}_2,\hat{X}_{12}}$. So $P_{X,\hat{X}_1,\hat{X}_2}$ is as follows: We can minimize the resulting distortion at decoders 1 and 2 by

X \ $\hat{X}_1, \hat{X}_2$	00	01	10	11
0	$(1-D_0)P_{000} + D_0((1-a_1)(1-a_2) - P_{000})$	$(1-D_0)P_{001} + D_0((1-a_1)a_2 - P_{001})$	$(1-D_0)P_{010} + D_0(a_1(1-a_2) - P_{010})$	$(1-D_0)(\frac{1}{2} - P_{000} - P_{001} - P_{010}) + D_0(a_1a_2 - \frac{1}{2} + P_{000} + P_{001} + P_{010})$
1	$D_0P_{000} + (1-D_0)((1-a_1)(1-a_2) - P_{000})$	$D_0P_{001} + (1-D_0)((1-a_1)a_2 - P_{001})$	$D_0P_{010} + (1-D_0)(a_1(1-a_2) - P_{010})$	$D_0(\frac{1}{2} - P_{000} - P_{001} - P_{010}) + (1-D_0)(a_1a_2 - \frac{1}{2} + P_{000} + P_{001} + P_{010})$

Table A.2: Joint probability distribution of $P_{X,\hat{X}_1,\hat{X}_2}$

choosing P_{000}, P_{001} and P_{010} optimally. Let $P^*_{X,\hat{X}_1,\hat{X}_2}$ be the optimal joint distribution, we will show that there are two choices for $P^*_{X,\hat{X}_1,\hat{X}_2}$. We have:

$$\mathsf{E}(d_H(\hat{X}_1, X)) + \mathsf{E}(d_H(\hat{X}_2, X)) = P(\hat{X}_1 \neq X) + P(\hat{X}_2 \neq X)$$

$$= (P_{X,\hat{X}_1,\hat{X}_2}(0,0,1) + P_{X,\hat{X}_1,\hat{X}_2}(1,0,1)) + (P_{X,\hat{X}_1,\hat{X}_2}(0,1,0)+$$

$$P_{X,\hat{X}_1,\hat{X}_2}(1,1,0)) + 2(P_{X,\hat{X}_1,\hat{X}_2}(0,1,1) + P_{X,\hat{X}_1,\hat{X}_2}(1,0,0))$$

$$= P_{001} + (1 - a_1)a_2 - P_{001} + P_{010} + (1 - a_2)a_1 - P_{010}+$$

$$2D_0(P_{000} + a_1a_2 - \frac{1}{2} + P_{000} + P_{001} + P_{010})$$

$$+ 2(1 - D_0)(\frac{1}{2} - P_{000} - P_{001} - P_{010} + (1 - a_1)(1 - a_2) - P_{000})$$

$$= (2D_0 - 1)a_1 + (2D_0 - 1)a_2 + 4(2D_0 - 1)P_{000} + 2(2D_0 - 1)P_{001} + 2(2D_0 - 1)P_{010} - 4D_0 + 3.$$

This is an optimization problem on $a_1, a_2, P_{000}, P_{001}, P_{010}$ with respect to the constraints:

$$a_1 \in [0, 1], a_2 \in [0, 1], P_{000} \in [0, (1 - a_1)(1 - a_2)], P_{001} \in [0, (1 - a_1)a_2], P_{010} \in [0, a_1(1 - a_2)]$$

$$P_{000} + P_{001} + P_{010} \in [\frac{1}{2} - a_1a_2, \frac{1}{2}].$$

Also note that for fixed a_1 and a_2 the problem becomes a linear optimization problem (otherwise the constraints are not linear). So we fix a_1 and a_2 and optimize P_{000}, P_{001} and P_{010} for each value of a_1 and a_2. In this case the simplex algorithm provides a straightforward solution. We investigate the solution in several different cases:

Case 1: $(1 - a_1)(1 - a_2) \geq \frac{1}{2}$: Note that in the simplex algorithm, the variable with smallest (most negative) coefficient takes its maximum possible value first.Since $D_0 < \frac{1}{2}$, $(2D_0 - 1) < 0$, so the algorithm would first maximize the value of P_{000}. Since $(1 - a_1)(1 - a_2) \geq \frac{1}{2}$, we have $P^*_{000} = \frac{1}{2}$. This along with constraint $P_{000} + P_{001} + P_{010} \in [\frac{1}{2} - a_1a_2, \frac{1}{2}]$ sets

$P^*_{001} = 0$ and $P^*_{010} = 0$. So in this case:

$$\mathsf{E}(d_H(\hat{X}_1, X)) + \mathsf{E}(d_H(\hat{X}_2, X)) = (2D_0 - 1)a_1 + (2D_0 - 1)a_2 + 2(2D_0 - 1) - 4D_0 + 3$$
$$= 1 + (2D_0 - 1)(a_1 + a_2).$$

Now we optimize on a_1, a_2 such that $(1 - a_1)(1 - a_2) \geq \frac{1}{2}$. Increasing a_1 or a_2 decreases the distortion so the optimal value is achieved when $(1 - a_1)(1 - a_2) = \frac{1}{2}$, so $a_2 = 1 - \frac{1}{2(1-a_1)}$. We have:

$$\mathsf{E}(d_H(\hat{X}_1, X)) + \mathsf{E}(d_H(\hat{X}_2, X)) = 1 + (2D_0 - 1)(a_1 + 1 - \frac{1}{2(1 - a_1)})$$

Optimizing the value of a_1, we get $a_1^* = a_2^* = 1 - \frac{\sqrt{2}}{2}$. These values give $P_{X,\hat{X}_1,\hat{X}_2} = P_{X,V_{\{1\}},V_{\{2\}}}$. Also replacing the values in $P_{\hat{X}_{12},\hat{X}_1,\hat{X}_2}$, we get: which shows that $\hat{X}_{12}$ is a function of $\hat{X}_1$

$\hat{X}_{12}$ \ $\hat{X}_1, \hat{X}_2$	00	01	10	11
0	$\frac{1}{2}$	0	0	0
1	0	$\frac{\sqrt{2}-1}{2}$	$\frac{\sqrt{2}-1}{2}$	$\frac{3-2\sqrt{2}}{2}$

Table A.3: The optimal joint distribution

and $\hat{X}_2$. **Case 2:** $(1 - a_1)(1 - a_2) < \frac{1}{2}$, $a_1 \leq \frac{1}{2}$: In this case the simplex method yields the following set of optimal distributions:

$$P^*_{000} = (1 - a_1)(1 - a_2), P^*_{001} = \alpha, P^*_{010} = \frac{1}{2} - (1 - a_1)(1 - a_2) - \alpha, P^*_{011} = 0$$
$$P^*_{100} = 0, P^*_{101} = (1 - a_1)a_2 - \alpha, P^*_{010} = (1 - a_2)a_1 - \frac{1}{2} + (1 - a_1)(1 - a_2) + \alpha, P^*_{111} = a_1 a_2.$$

Where $\alpha \in [a_2 - \frac{1}{2}, \frac{1}{2} - (1 - a_1)(1 - a_2)]$ is an auxiliary variable that does not play a role in the distortion since the coefficients of P^*_{001} and P^*_{010} are equal in the distortion formula. We get:

$$\mathsf{E}(d_H(\hat{X}_1, X)) + \mathsf{E}(d_H(\hat{X}_2, X)) = 1 + (2D_0 - 1)((1 - a_1)(1 - a_2) + a_1 a_2).$$

Note that since $a_1 < \frac{1}{2}$, the term $(1 - a_1)(1 - a_2) + a_1 a_2$ is decreasing with a_2, so the distortion is increasing with a_2 and the optimal values are $a_2^* = max(0, 1 - \frac{1}{2(1-a_1)})$, since $a_1 \leq \frac{1}{2}$, $a_2^* = 1 - \frac{1}{2(1-a_1)}$, replacing a_2^* we have:

$$\mathsf{E}(d_H(\hat{X}_1, X)) + \mathsf{E}(d_H(\hat{X}_2, X)) = 1 + (2D_{\{1\}} - 1)(\frac{1}{2} + a_1(1 - \frac{1}{2(1 - a_1)})).$$

Solving for a_1 we get $a_1 = 1 - \frac{1}{\sqrt{2}}$ and in tun $a_2 = 1 - \frac{1}{\sqrt{2}}$ as in the previous case.

Case 3: $(1 - a_1)(1 - a_2) < \frac{1}{2}, a_1 > \frac{1}{2}, a_1 a_2 < \frac{1}{2}$: The probabilities are as in the last case with $\alpha \in [0, \frac{1}{2} - (1 - a_1)(1 - a_2)]$. The distortion is similar to the last case. Since $a_1 > \frac{1}{2}$, the distortion is decreasing in a_2. So $a_2^* = \frac{1}{2a_1}$. Which yields:

$$\mathsf{E}(d_H(\hat{X}_1, X)) + \mathsf{E}(d_H(\hat{X}_2, X)) = 1 + (2D_{\{1\}} - 1)((1 - a_1)(1 - \frac{1}{2a_1}) + \frac{1}{2}).$$

This would have no solution for optimizing a_1 at the given range.

Case 4: $a_1 a_2 > \frac{1}{2}$: By the same arguments the optimal solution is

$$\begin{aligned}
&P^*_{000} = (1 - a_1)(1 - a_2), P^*_{001} = (1 - a_1)a_2, P^*_{010} = (1 - a_2)a_1, P^*_{011} = 0 \\
&P^*_{100} = 0, P^*_{101} = 0, P^*_{010} = 0, P^*_{111} = \frac{1}{2}.
\end{aligned}$$

Then $P^*_{\hat{X}_{12}, \hat{X}_1, \hat{X}_2}$ is: which is the second choice for the optimal joint distribution. Note that again $\hat{X}_{12}$ is a function of $\hat{X}_1$ and $\hat{X}_2$. □

Step 5: We are left with $C_{\{1\},\{3\}}$, $C_{\{2\},\{3\}}$, $C_{\{1\}}$,$C_{\{2\}}$ and $C_{\{3\}}$. Let X_i be the reconstruction at decoder $\{i\}$ for $i \in \{1, 2, 3\}$.

$\hat{X}_{12}$ \ $\hat{X}_1, \hat{X}_2$	00	01	10	11
0	$\frac{3-2\sqrt{2}}{2}$	$\frac{\sqrt{2}-1}{2}$	$\frac{\sqrt{2}-1}{2}$	0
1	0	0	0	$\frac{1}{2}$

Table A.4: The optimal joint distribution

Lemma 13. *The following Markov chains hold:*

$$U_{\{1\},\{3\}}, U_{\{1\}}, X_1 \perp\!\!\!\perp U_{\{2\},\{3\}}, U_{\{2\}}, X_2 \tag{A.27}$$

$$U_{\{1\}}, U_{\{2\}}, U_{\{1\},\{3\}}, U_{\{2\},\{3\}} \leftrightarrow X_1, X_2 \leftrightarrow X \tag{A.28}$$

$$U_{\{1\},\{3\}}, U_{\{1\}} \leftrightarrow X_1 \leftrightarrow X, U_{\{2\},\{3\}}, U_{\{2\}} \tag{A.29}$$

$$U_{\{2\},\{3\}}, U_{\{2\}} \leftrightarrow X_2 \leftrightarrow X, U_{\{1\},\{3\}}, U_{\{1\}} \tag{A.30}$$

$$U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, U_{\{3\}} \leftrightarrow X_3 \leftrightarrow X \tag{A.31}$$

$$X_1, X_2, U_{\{1\}}, U_{\{2\}} \leftrightarrow U_{\{1\},\{3\}} U_{\{2\},\{3\}}, X \leftrightarrow U_{\{3\}}, X_3 \tag{A.32}$$

$$U_{\{1\}} \leftrightarrow U_{\{1\},\{3\}} U_{\{2\},\{3\}}, X_1, U_{\{3\}} \leftrightarrow X \tag{A.33}$$

$$U_{\{2\}} \leftrightarrow U_{\{1\},\{3\}} U_{\{2\},\{3\}}, X_2, U_{\{3\}} \leftrightarrow X \tag{A.34}$$

$$U_{\{3\}} \leftrightarrow U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3, X_1 \leftrightarrow X \tag{A.35}$$

$$U_{\{3\}} \leftrightarrow U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3, X_2 \leftrightarrow X \tag{A.36}$$

Proof. (A.27) holds from Lemma 3. From the optimality at decoder $\{1, 2\}$ and step 4 we have:

$$I(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, U_{\{1\}}, U_{\{2\}}, X_1, X_2; X) = I(X_1, X_2; X) = 1 - h_b(D_0),$$

which proves (A.28). Next we prove (A.29):

$$P(U_{\{1\},\{3\}}, U_{\{1\}}, U_{\{2\},\{3\}}, U_{\{2\}}, X_1, X) \overset{(a)}{=} \sum_{X_2} P(U_{\{1\},\{3\}}, U_{\{1\}}, X_1) P(X_2, U_{\{2\},\{3\}}, U_{\{2\}}) P(X|X_1, X_2)$$

$$= P(U_{\{1\},\{3\}}, U_{\{1\}}, X_1) P(U_{\{2\},\{3\}}, U_{\{2\}}) \sum_{X_2} P(X_2|U_{\{2\}}, U_{\{2\},\{3\}}) P(X|X_1, X_2)$$

$$= P(U_{\{1\},\{3\}}, U_{\{1\}}, X_1) P(U_{\{2\},\{3\}}, U_{\{2\}}) P(X|X_1, U_{\{2\}}, U_{\{2\},\{3\}})$$

$$\overset{b}{=} P(U_{\{1\},\{3\}}, U_{\{1\}}, X_1) P(U_{\{2\}}, U_{\{2\},\{3\}}, X|X_1)$$

In (a) we have used (A.27) and the Markov chain (A.28), in (b), we have used (A.27). (A.30) follows by symmetry. (A.31) can be proved using optimality at decoder $\{3\}$ and the argument given in the proof of (A.27). We proceed with the proof of (A.32). Consider the following packing bounds at decoder $\{1, 2\}$ and $\{3\}$:

$$H(U_{\{1\}}, U_{\{2\}}, U_{\{1\},\{3\}}, U_{\{2\},\{3\}}) \leq H(U_{\{1\}}) + H(U_{\{2\}}) + H(U_{\{1\},\{3\}}) + H(U_{\{2\},\{3\}})$$

$$- r_1 - r_2 - r_{1,3} - r_{2,3} + R_1 + R_2$$

$$H(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, U_{\{3\}}) \leq H(U_{\{1\},\{3\}}) + H(U_{\{2\},\{3\}}) + H(U_{\{3\}}) - r_{1,3} - r_{2,3} - r_3 + R_3$$

And the following covering bounds:

$$H(U_{\{1\}}, U_{\{2\}}, U_{\{3\}}, U_{\{1\},\{3\}}, U_{\{2\},\{3\}}|X) \geq H(U_{\{1\}}) + H(U_{\{2\}}) + H(U_{\{3\}})$$

$$+ H(U_{\{1\},\{3\}}) + H(U_{\{2\},\{3\}}) - r_1 - r_2 - r_3 - r_{1,3} - r_{2,3}$$

$$H(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}|X) \geq H(U_{\{1\},\{3\}}) + H(U_{\{2\},\{3\}}) - r_{1,3} - r_{2,3}$$

Adding all the bounds and simplifying we get:

$$R_1 + R_2 + R_3 \geq I(U_{\{1\}}, U_{\{2\}}, U_{\{1\},\{3\}}, U_{\{2\},\{3\}}; X)$$
$$+ I(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, U_{\{3\}}; X) + I(U_{\{1\}}, U_{\{2\}}; U_{\{3\}}|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X)$$

This resembles the two user sum-rate bound when the first user is sending descriptions 1 and 2 while the second user transmits description 3. From optimality at decoder {1}2, $R_1 + R_2 = I(U_{\{1\}}, U_{\{2\}}, U_{\{1\},\{3\}}, U_{\{2\},\{3\}}; X)$ and optimality at decoder {3} yields $R_3 = I(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, U_{\{3\}}; X)$. So $I(U_{\{1\}}, U_{\{2\}}; U_{\{3\}}|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X) = 0$. This proves (A.32). We have:

$$P(U_{\{1\},\{3\}}, U_{\{1\}}, X_1, U_{\{2\},\{3\}}, U_{\{3\}}, X)$$
$$=P(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1)P(U_1|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1)$$
$$P(X|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1, U_{\{1\}})P(U_{\{3\}}|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1, U_{\{1\}}, X)$$
$$\overset{(a)}{=} P(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1)P(U_1|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1)P(X|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1)$$
$$P(U_{\{3\}}|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1, U_{\{1\}}, X)$$
$$\overset{(b)}{=} P(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1)P(U_1|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1)P(X|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1)P(U_{\{3\}}|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X)$$

where (*a*) follows from (A.29) and Lemma 14 given below. (*b*) follows from (A.32). So we have shown that $U_{\{1\}} \leftrightarrow U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1 \leftrightarrow X, U_{\{3\}}$, using Lemma 14 we conclude (A.33). (A.34) follows by symmetry. Lastly we prove (A.35):

$$P(X, X_1, U_{\{3\}}|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3)$$
$$=P(X|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3)P(U_{\{3\}}|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3, X)P(X_1|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X, U_{\{3\}}, X_3)$$
$$\overset{a}{=} P(X|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3)P(U_{\{3\}}|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3, X)P(X_1|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X, X_3)$$
$$=P(U_{\{3\}}|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3, X)P(X, X_1|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3)$$
$$\overset{b}{=} P(U_{\{3\}}|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3)P(X, X_1|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_3)$$

where (a) follows form A.32. (b) holds because of (A.31). (A.35) follows from lemma 14.

Lemma 14. *For random variables A, B, C and D if we have $A, B \leftrightarrow C \leftrightarrow D$ then $A \leftrightarrow B, C \leftrightarrow D$.*

Proof. We have:

$$P(A, D|B, C) = \frac{P(A, B, C, D)}{P(B, C)} = \frac{P(C)P(A, B|C)P(D|C)}{P(C)P(B|C)} = P(A|BC)P(D|C) = P(A|BC)P(D|BC)$$

□

□

Next we argue that if we set $U_{\{1\}}$ to be equal to X_1 there would be no change in distortion and the rate does not increase. First consider decoder $\{1, 3\}$. The optimal reconstruction function is given by
$argmax_x(P_{X|U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1\}}U_{\{3\}}}(x|u_{1,3}, u_{23}, u_{\{1\}}, u_{\{3\}}))$. We have:

$$\begin{aligned}
&argmax_x(P_{X|U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1\}}U_{\{3\}}}(x|u_{1,3}, u_{23}, u_{\{1\}}, u_{\{3\}}))\\
&\overset{(a)}{=} argmax_x(P_{X|U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1\}}U_{\{3\}}X_1}(x|u_{1,3}, u_{23}, u_{\{1\}}, u_{\{3\}}, x_1))\\
&\overset{(b)}{=} argmax_x(P_{X|U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{3\}}X_1}(x|u_{1,3}, u_{23}, u_{\{3\}}, x_1))
\end{aligned}$$

where in (a) we used the fact that x_1 is a function of $U_{\{1\}}, U_{\{1\},\{3\}}$ and in (b) we use (A.33). So the distortion won't change at decoder $\{1, 3\}$. Also the reconstruction at decoder $\{1\}$ is X_1 so setting $U_{\{1\}} = X_1$ won't change the reconstruction at this decoder. At decoder $\{1, 2\}$ we showed in step 4 that X_{12} is a function of X_1, X_2 where X_2 is a function of $U_{\{2\},\{3\}}, U_{\{2\}}$, so setting $U_{\{1\}} = X_1$ does not change the distortion at this decoder either. The rest of the decoders do not receive $U_{\{1\}}$. As for the rate, note that X_1 was reconstructed at all decoders reconstructing $U_{\{1\}}$. So replacing $U_{\{1\}}$ with X_1 does not require sending any extra information. So we set $U_{\{1\}} = X_1$ without any loss in distortion and with a potential gain in rate. The same argument combined with the Markov chains (A.34) sets $U_{\{2\}} = X_2$, also using Markov chains (A.35) and (A.36) we set $U_{\{3\}} = X_3$.

Lemma 15. *The following constraints hold:*

$$P_{XX_1X_2} \textit{ is fixed and equal to } P_{X,V_{\{1\}},V_{\{2\}}} \textit{ in the previous step.} \tag{A.37}$$

$$P_{XX_3} \textit{ is fixed and equal to } P_{X,V_3} \textit{ which is the optimizing distribution for decoder } \{3\}. \tag{A.38}$$

$$U_{\{1\},\{3\}} \leftrightarrow X_1 \leftrightarrow U_{\{2\},\{3\}}, X, X_2 \tag{A.39}$$

$$U_{\{2\},\{3\}} \leftrightarrow X_2 \leftrightarrow U_{\{1\},\{3\}}, X, X_1 \tag{A.40}$$

$$U_{\{1\},\{3\}}, U_{\{2\},\{3\}} \leftrightarrow X_3 \leftrightarrow X \tag{A.41}$$

$$X_1, X_2 \leftrightarrow X, U_{\{1\},\{3\}}, U_{\{2\},\{3\}} \leftrightarrow X_3 \tag{A.42}$$

Proof. (A.37) was proved in the step 4. (A.38) follows from PtP optimality at decoder {3}. (A.39) follows from (A.29), (A.40) follows from (A.30). (A.41) follows from (A.31). (A.42) follows from (A.32). □

We proceed by bounding the cardinality of $\mathsf{U}_{\{1\},\{3\}}$ and $\mathsf{U}_{\{2\},\{3\}}$. Using Lemma 15, the joint distribution between the random variables is given as follows:

$$\begin{aligned}
&P(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1, X_2, X_3, X) = P(U_{\{1\},\{3\}}, X_1)P(U_{\{2\},\{3\}}, X_2)P(X|X_1, X_2)P(X_3|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X) \\
&= P(U_{\{1\},\{3\}}, X_1)P(U_{\{2\},\{3\}}, X_2)P(X|X_1, X_2)\frac{P(U_{\{1\},\{3\}}U_{\{2\},\{3\}}X_3X)}{P(U_{\{1\},\{3\}}U_{\{2\},\{3\}}X)} \\
&= P(U_{\{1\},\{3\}}, X_1)P(U_{\{2\},\{3\}}, X_2)P(X|X_1, X_2)\frac{P(U_{\{1\},\{3\}}U_{\{2\},\{3\}}|X_3)P(X_3X)}{\sum_{X_1,X_2} P(U_{\{1\},\{3\}}, X_1)P(U_{\{2\},\{3\}}, X_2)P(X|X_1, X_2)}
\end{aligned} \tag{A.43}$$

Also note that we have the following equality:

$$P(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X) = \sum_{X_3} P(X, X_3) P(U_{\{1\},\{3\}} U_{\{2\},\{3\}} | X_3)$$
$$= \sum_{X_1, X_2} P(U_{\{1\},\{3\}} | X_1) P(U_{\{2\},\{3\}} | X_2) P(X X_1, X_2)$$

Denote $P(X, X_1, X_2) = P_{xx_1x_2}$ and $P(U_{\{1\},\{3\}} = \theta | X_1 = i) = \alpha_i(\theta), \theta \in \mathsf{U}_{1,3}, i \in \{0, 1\}$ and $P(U_{\{2\},\{3\}} = \gamma | X_2 = i) = \beta_i(\gamma), \gamma \in \mathsf{U}_{2,3}, i \in \{0, 1\}$. We have:

$$P_{U_{\{1\},\{3\}} U_{\{2\},\{3\}} | X_3}(\theta, \gamma | 0) P_{X_3, X}(0, 0) + P_{U_{\{1\},\{3\}} U_{\{2\},\{3\}} | X_3}(\theta, \gamma | 1) P_{X_3, X}(1, 0)$$
$$= \alpha_0(\theta)\beta_0(\gamma)P_{000} + \alpha_0(\theta)\beta_1(\gamma)P_{001} + \alpha_1(\theta)\beta_0(\gamma)P_{010} + \alpha_1(\theta)\beta_1(\gamma)P_{011}.$$

$$P_{U_{\{1\},\{3\}} U_{\{2\},\{3\}} | X_3}(\theta, \gamma | 0) P_{X_3, X}(0, 1) + P_{U_{\{1\},\{3\}} U_{\{2\},\{3\}} | X_3}(\theta, \gamma | 1) P_{X_3, X}(1, 1)$$
$$= \alpha_0(\theta)\beta_0(\gamma)P_{100} + \alpha_0(\theta)\beta_1(\gamma)P_{101} + \alpha_1(\theta)\beta_0(\gamma)P_{110} + \alpha_1(\theta)\beta_1(\gamma)P_{111}.$$

Using the values given in Table (A.3), we solve the system of equations:

$$P_{U_{\{1\},\{3\}} U_{\{2\},\{3\}} | X_3}(\theta, \gamma | 0) = \alpha_0(\theta)\beta_0(\gamma),$$
$$P_{U_{\{1\},\{3\}} U_{\{2\},\{3\}} | X_3}(\theta, \gamma | 1) = \frac{\sqrt{2} - 1}{2}(\alpha_1(\theta)\beta_1(\gamma) - \alpha_0(\theta)\beta_0(\gamma)) + \frac{1}{2}(\alpha_0(\theta)\beta_1(\gamma) + \alpha_1(\theta)\beta_0(\gamma)).$$

Hence the distribution in A.43 is completely determined by α_i and $\beta_i, i \in \{0, 1\}$.

Lemma 16. *Assume there exists α_i and β_i, such that $D_{\{1,3\}} \leq D_0$, then $I(U_{\{1\},\{3\}} U_{\{2\},\{3\}} X_1 X_3; X) \geq 1 - h_b(D_0)$.*

Proof. The proof follows from Shannon's rate distortion function for PtP source coding.

□

Based on the previous lemma it is enough to show that for every α_i and β_i, $I(U_{\{1\},\{3\}}U_{\{2\},\{3\}}X_1X_3;X) < 1-h_b(D_0)$, in that case we have a contradiction. We need to maximize $I(U_{\{1\},\{3\}}U_{\{2\},\{3\}}X_1X_3;X)$ as a function of α_i and β_i. We use the following lemma:

Lemma 17. *[14] Let X be a finite set and U be an arbitrary set. Let $\mathcal{P}(\mathsf{X})$ be a set of pmfs on X and $p(x|u)$ be a collection of pmfs on X for every $u \in \mathsf{U}$. Let g_j, $j \in [1,d]$ be real-valued continuous functions on $\mathcal{P}(\mathsf{X})$. Then for every $U \sim F(u)$ defined on U, there exists random variable $U' \sim p(u')$ with cardinality $|\mathsf{U}'| \leq d$ and a collection of conditional pmfs $p(u'|x)$ on X for every $u' \in \mathsf{U}'$ such that for every $j \in [1,d]$:*

$$\int_{\mathsf{U}} g_j(p_{X|U}(x|u))dF(u) = \sum_{u'} g_j(p_{X|U'}(x|u'))p(u')$$

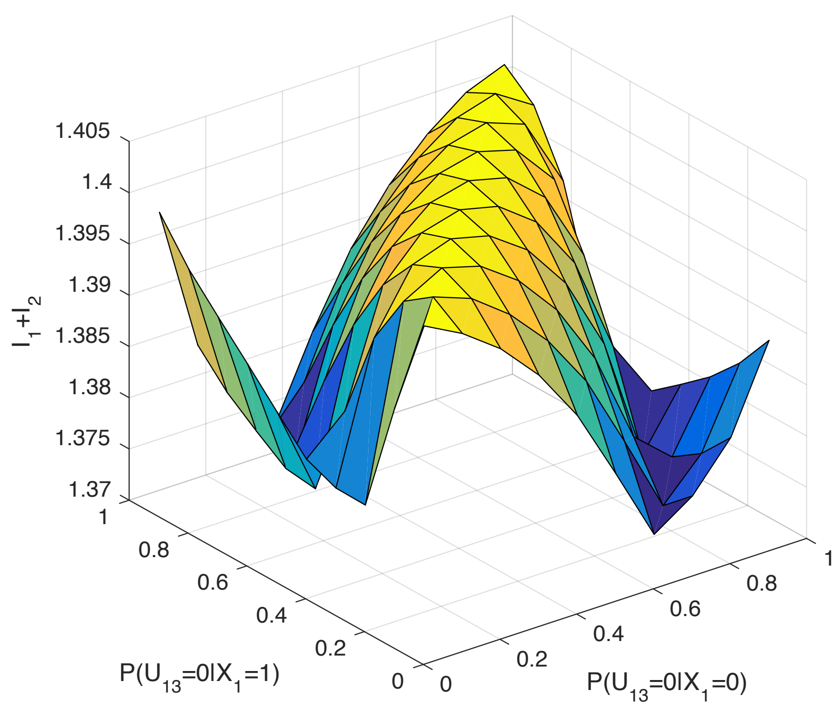

Figure A.1: Maximum value of $I(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1, X_3; X)$ $+I(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_2, X_3; X)$

We want to use the lemma to bound cardinality of $\mathsf{U}_{1,3}$. Take $g_1(p_{U_{\{2\},\{3\}},X_1,X_2,X_3,X}|U_{\{1\},\{3\}}) = p_{X_1|U_{\{1\},\{3\}}}(1|u_{13})$ and $g_2(p_{U_{\{2\},\{3\}},X_1,X_2,X_3,X}|U_{\{1\},\{3\}}) = H(X|U_{\{2\},\{3\}},X_1,X_3,X,U_{\{1\},\{3\}} = u_{1,3})$. Note that fixing the expectation on g_1 fixes the joint distribution in (A.43) and fixing the expectation of g_2 fixes the term we want to minimize. So for any $U_{\{1\},\{3\}}$ minimizing $I(U_{\{1\},\{3\}}U_{\{2\},\{3\}}X_1X_3; X)$, there exists $U'_{1,3}$ with cardinality at most 2, such that the joint distribution and $I(U_{\{1\},\{3\}}U_{\{2\},\{3\}}X_1X_3; X)$ are the same. So it is enough to search over $U_{\{1\},\{3\}}$ with cardinality 2. The same arguments hold for bounding the cardinality of $\mathsf{U}_{2,3}$. For this size of random variables, computer-assisted calculation shows that $I(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1, X_3; X) + I(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_2, X_3; X) < 1.42 < 2(1 - h_b(D_0)) = 1.58$ as shown in Figure A.1. So we have a contradiction and the SSC does not achieve the RD vector. □

A.3 Proofs for Section 2.6

A.3.1 Proof of lemma 46

Proof. Index the inequalities in the SSC from 1 to K. For every inequality in the linear coding region (LCR), there exists a unique inequality in the SSC with the same left hand side, index this inequality with the same index used in the RCR. Let $I_1 > R$ be a bound resulting from applying FME on the SSC. Assume the bound results from adding inequalities indexed $i_1, i_2, \ldots, i_k$, it is straightforward to show that adding inequalities with the same indices in the LCR gives the same bound. The reason is that by our construction, the left-hand sides would be the same. In the right-hand side, due to the FME, the terms involving r_A would be eliminated. Define $r'_A = r_A - H(U_A)$ and $r'_{o,A} = r_{o,A} - \log(q)$, eliminating r_A is equivalent to eliminating r'_A or $r'_{o,A}$. □

APPENDIX B

Proofs for Chapter III

B.1 Proofs for Chapter III

B.1.1 Proof of Lemma 13

Proof. We have shown that LCR and RCR are equal. So it is enough to show that the bounds in RCR can be written in terms of mutual informations. But that is obvious since each packing bound and covering bound in RCR can be rewritten in terms of mutual informations. let $\mathcal{L} = \{M_1, M_2, ..., M_L\}$. Then for any packing bound in RCR we have:

$$H(U_{\mathbf{M}}) - \sum_{\mathcal{M}\in\mathbf{L}} H(U_{\mathcal{M}}) \leq \sum_{\mathcal{M}\in\mathbf{M}} (\sum_{i\in[1:L]} \rho_{\mathcal{M},i} - r_{\mathcal{M}})$$

$$\to - \sum_{j\in[1,L]} I(U_{\mathcal{M}_j}; U_{\mathcal{M}_1}, U_{\mathcal{M}_2}, ..., U_{\mathcal{M}_{j-1}}) \leq$$

$$\sum_{\mathcal{M}\in\mathbf{M}} (\sum_{i\in[1:L]} \rho_{\mathcal{M},i} - r_{\mathcal{M}})$$

And for any covering bound we have:

$$
\begin{aligned}
&H(U_{\mathbf{M}}|X) - \sum_{\mathcal{M}_i \in \mathbf{M}} H(U_{\mathcal{M}_i}) \geq - \sum_{\mathcal{M}_i \in \mathbf{M}} r_{\mathcal{M}_i} \\
&\rightarrow - \sum_{j \in [1,L]} I(U_{\mathcal{M}_j}; U_{\mathcal{M}_1}, U_{\mathcal{M}_2}, ..., U_{\mathcal{M}_{j-1}}) \\
&\qquad\qquad\qquad - I(U_{\mathbf{M}}; X) \geq \sum_{\mathcal{M}_i \in \mathbf{M}} -r_{\mathcal{M}_i}
\end{aligned}
$$

Note after the FME, the $r_{\mathcal{M}}$ terms would be eliminated and only the mutual information terms would remain. □

B.1.2 Proof of Theorem III.14

Proof. For a fixed q and γ, generate a linear code for quantizing the discrete version of the source to mean squared distortion P. Use this linear construction to quantize $X_{q,\gamma}$ to U and $\mathbb{Z}_{q,\gamma}$ to V. Note that since the source is not symmetric, the linear code needs to be binned. The rate of the linear code is $r_l = log(q) - H(U|X)$ and the code is binned with rate $b_l = log(q) - H(U)$. The third description needs to bin the code in a way that $U + V$ is uniquely reconstructed at decoder 3, so it bins at rate $b_3 = log(q) - H(U + V)$. Using this scheme as $q \to \infty$ and $q\gamma \to 0$, we have:

$$
\begin{aligned}
R_1 &= R_2 = log(q) - H(U|X) - (log(q) - H(U)) \\
&= \frac{1}{2} log(2\pi e(1-p)) - \frac{1}{2} log(2\pi e(1-p)p) = \frac{1}{2} log(\frac{1}{p}) \\
R_3 &= log(q) - H(U|X) - (log(q) - H(U+V)) = \frac{1}{2} log(\frac{2}{p})
\end{aligned}
$$

One can check that the distortion constraints are also satisfied. □

$$
\begin{array}{llll}
 & C_{\{2,3\}} & & \\
C_{\{23\},\{1\}} & C_{\{1,3\}} & & \\
C_{\{13\},\{2\}} & C_{\{1,2\}} & & \\
C_{\{12\},\{3\}} & C_{\{2\},\{3\}} & C_{\{3\}} & C_{\{13\},\{23\}} \\
C_{\{1\},\{2\},\{3\}} & C_{\{1\},\{3\}} & C_{\{2\}} & C_{\{12\},\{23\}} \\
C_{\{12\},\{13\},\{23\}} & C_{\{1\},\{2\}} & C_{\{1\}} & C_{\{12\},\{13\}}
\end{array}
$$

Figure B.1: Three-Descriptions Codebook Structure

B.1.3 Proof of Theorem III.15

Proof. For convenience of notation, we denote codebook $C_{\mathcal{M}_\sigma}$ by $C_{\mathcal{M}_s}$ throughout the examples in this chapter. The proof is very similar to the one given in example A [40]. Here we give a brief summary. The codebook structure for the scheme in the three descriptions case is shown in figure B.1. Note that decoders 1,2, 12, 13 and 23 are receiving their respective descriptions at optimal PtP rate distortion. So by the same arguments as in steps one through five in [40], all the codebooks except $C_{\{1\},\{23\}}, C_{\{2\},\{13\}}, C_{\{3\},\{12\}}, C_{\{1\}}, C_{\{2\}}, C_{\{3\}}, C_{\{13\},\{23\}}$ can be eliminated. Note that since decoders 1 and 2 are operating at optimal PtP rate-distortion, descriptions 1 and 2 can't carry any bin number for $C_{\{3\},\{12\}}$, so this codebook is not used in reconstruction at decoder 12 and it can be sent on $C_{\{3\}}$. Also description 3 can't carry any bin number for $C_{\{1\},\{23\}}$ because of optimality at 13 (if the bin number is not sent, the codebook is still decodable using the bin number from description 1, so sending a bin number on description 3 would cause sub-optimality), so $C_{\{1\},\{23\}}$ can be eliminated. Same argument works for $C_{\{2\},\{13\}}$. We are left with $C_{\{1\}}, C_{\{2\}}, C_{\{3\}}, C_{\{13\},\{23\}}$. Since decoder 1 receives only $C_{\{1\}}$, it is straightforward to show that $U_{\{1\}} = X + N_1$, also $U_{\{2\}} = Z + N_2$, where N_1 and N_2 are independent Gaussian random variables. Then $\hat{Y}_{\{13\}} = Y + N_1 + N_4$ and $\hat{Y}_{\{23\}} = Y + N_2 + N_5$, also if $\hat{Y}_{\{3\}} = Y + N_3$, then this reconstruction of Y is also available at decoder $\{13\}$, since this decoder is at optimality, we must have $Y \leftrightarrow Y + N_1 + N_4 \leftrightarrow Y + N_3$ and $Y \leftrightarrow Y + N_1 + N_4 \leftrightarrow Y + N_3$. Since N_3 has power $2P$, it must equal $N_1 + N_4$ and $N_2 + N_5$, for the Markov chain to hold. This is a contradiction since in that case $U_3 = U_1 + U_2$, the rate R_3 required to send U_3 exceeds the rates given in the rate vector.

B.1.4 Proof of Lemma 18

Proof. We calculate the difference between the new bound and the previous ones and show that it can be written in terms of mutual informations.

$$H(\alpha U + \beta V|X) - H(U|X) \tag{B.1}$$

$$= H(\alpha U + \beta V|X) - H(U, V|X) + H(V|X, U) \tag{B.2}$$

$$= H(\alpha U + \beta V|X) - H(\alpha U + \beta V, V|X) + H(V|X, U) \tag{B.3}$$

$$= -H(V|X, \alpha U + \beta V) + H(V|X, U) \tag{B.4}$$

$$= I(\alpha U + \beta V; V|X) - I(U; V|X) \tag{B.5}$$

□

APPENDIX C

Proofs for Chapter IV

C.1 Proofs for chapter IV

C.1.1 Proof of Lemma 26

Assume the family $(n, \mathsf{M}_1, \mathsf{M}_2, \mathsf{M}_3, \mathbf{e}, \mathbf{d}), n \in \mathbb{N}$ achieves the rate-triple. Let $M_i, i \in \{1, 2, 3\}$ be uniform random variables defined on sets M_i. In the first step, we argue that the size of the set $\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbb{C}_3 \oplus A_\epsilon^n(N_3)$ is close to $|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus A_\epsilon^n(N_3)||\mathbb{C}_3|$. More precisely, we prove the following claim:

Claim 1. *For every $\epsilon \in \mathbb{R}^+$, there exists a sequence of numbers $\alpha_{n,\epsilon} \in \mathbb{R}^+, n \in \mathbb{N}$ such that the following inequality holds:*

$$\frac{1}{n} \log |\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbb{C}_3 \oplus A_\epsilon^n(N_3)| \geq \frac{1}{n} \log (|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus A_\epsilon^n(N_3)||\mathbb{C}_3|) - \alpha_{n,\epsilon},$$

and $\alpha_{n,\epsilon}$ goes to 0 as $n \to \infty$ and $\epsilon \to 0$.

Proof. Intuitively, if the size of $\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbb{C}_3 \oplus A_\epsilon^n(N_3)$ is much smaller than $|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus A_\epsilon^n(N_3)||\mathbb{C}_3|$, that means there exists a large number of sets of vectors $\mathbf{c_1}, \mathbf{c}_2, \mathbf{c}_3, \mathbf{n}_3$, with different $\mathbf{c}_3$'s for which the sum is equal. This causes a large error probability in decoder 3 since the decoder is unable to distinguish between these sets of vectors. More precisely, let $\mathbf{n_t}$ be a type on vectors in $A_\epsilon^n(N_3)$, and let $\mathbf{c}_3 \in \mathbb{C}_3$, define $\mathsf{B}_{\mathbf{c}_3,\mathbf{n}_t}$ as follows,

$$\mathsf{B}_{\mathbf{c}_3,\mathbf{n}_t} = \{\mathbf{c}_1 \oplus \mathbf{c}_2 \oplus \mathbf{n}_3 | \exists \mathbf{c}_1', \mathbf{c}_2', \mathbf{c}_3', \mathbf{n}_3' \in \mathbb{C}_1 \times \mathbb{C}_2 \times \mathbb{C}_3 \times \mathcal{P}_t, \text{ such that } \mathbf{c}_1' \oplus \mathbf{c}_2' \oplus \mathbf{c}_3' \oplus \mathbf{n}_3' = \mathbf{c}_1 \oplus \mathbf{c}_2 \oplus \mathbf{c}_3 \oplus \mathbf{n}_3, \mathbf{c}_3' \neq \mathbf{c}_3\},$$

where $\mathcal{P}_t$ is the set of all vectors $\mathbf{n}_3 \in A_\epsilon^n(N_3)$ with type $\mathbf{n}_t$. That is $\mathsf{B}_{\mathbf{c}_3,\mathbf{n}_t}$ is the set of $(\mathbf{c}_1, \mathbf{c}_2, \mathbf{n}_3)$'s for which the decoder has non-zero error probability for decoding $\mathbf{c}_3$ or another codeword $\mathbf{c}_3'$. From set theory, we have the following:

$$\begin{aligned}
|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbb{C}_3 \oplus A_\epsilon^n(N_3)| &\geq |\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbb{C}_3 \oplus \mathcal{P}_t| && \text{(C.1)}\\
&= |\bigcup_{\mathbf{c}_3} \mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbf{c}_3 \oplus \mathcal{P}_t| && \text{(C.2)}\\
&\geq |\bigcup_{\mathbf{c}_3} \left(\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbf{c}_3 \oplus \mathcal{P}_t - \bigcup_{\mathbf{c}_3' \neq \mathbf{c}_3} \mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbf{c}_3' \oplus \mathcal{P}_t \right)| && \text{(C.3)}\\
&= \sum_{\mathbf{c}_3} |\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbf{c}_3 \oplus \mathcal{P}_t - \bigcup_{\mathbf{c}_3' \neq \mathbf{c}_3} \mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbf{c}_3' \oplus \mathcal{P}_t| && \text{(C.4)}\\
&= \sum_{\mathbf{c}_3 \in \mathbb{C}_3} (|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathcal{P}_t| - |\mathsf{B}_{\mathbf{c}_3,\mathbf{n}_t}|) && \text{(C.5)}\\
&= |\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathcal{P}_t||\mathbb{C}_3| - \sum_{\mathbf{c}_3 \in \mathbb{C}_3} |\mathsf{B}_{\mathbf{c}_3,\mathbf{n}_t}| && \text{(C.6)}\\
&= |\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathcal{P}_t||\mathbb{C}_3| - |\mathbb{C}_3| \sum_{\mathbf{c}_3 \in \mathbb{C}_3} \frac{|\mathsf{B}_{\mathbf{c}_3,\mathbf{n}_t}|}{|\mathbb{C}_3|} && \text{(C.7)}\\
&= (|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathcal{P}_t| - E(|\mathsf{B}_{\mathbf{c}_3,\mathbf{n}_t}|))\, |\mathbb{C}_3| && \text{(C.8)}\\
&= \left(|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathcal{P}_t| (1 - \frac{E(|\mathsf{B}_{\mathbf{c}_3,\mathbf{n}_t}|)}{|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathcal{P}_t|}) \right) |\mathbb{C}_3|. && \text{(C.9)}
\end{aligned}$$

On the other hand, as $n \to \infty$, the error probability at decoder 3 goes to 0. This means that $P(\mathbf{d}_3(\mathbf{c}_1 \oplus \mathbf{c}_2 \oplus \mathbf{e}(M_3) \oplus \mathbf{n}_3) \neq M_3)$ goes to 0. Consequently, there exists a family of types type $\mathbf{n}_t$ such that $P(\mathbf{d}_3(\mathbf{c}_1 \oplus \mathbf{c}_2 \oplus \mathbf{e}(M_3) \oplus \mathbf{n}_3) \neq M_3 | \mathbf{n}_t)$ goes to 0. There exists a sequence

δ_n which approaches 0 at the limit such that:

$$\begin{aligned}
\delta_n &\geq P\left(\mathbf{d}_3(\mathbf{c}_1 \oplus \mathbf{c}_2 \oplus \mathbf{e}(M_3) \oplus \mathbf{n}_3) \neq M_3 | \mathbf{n}_t\right) \\
&\geq \frac{1}{2} P\left(\mathbf{c}_1 \oplus \mathbf{c}_2 \oplus \mathbf{n}_3 \in \mathsf{B}_{\mathbf{e}(M_3),\mathbf{n}_t} | \mathbf{n}_t\right) \\
&\geq \frac{1}{2} \sum_{\mathbf{c}_3 \in \mathbb{C}_3} \frac{|\mathsf{B}_{\mathbf{c}_3,\mathbf{n}_t}|}{|\mathbb{C}_3||\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathcal{P}_t|} \\
&= \frac{1}{2} \frac{E(|\mathsf{B}_{\mathbf{e}_3(M_3),\mathbf{n}_t}|)}{|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathcal{P}_t|}
\end{aligned}$$

Inserting this last inequality in (C.9) we have,

$$\begin{aligned}
&|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathcal{P}_t||\mathbb{C}_3| - \sum_{\mathbf{c}_3 \in \mathbb{C}_3} |\mathsf{B}_{\mathbf{c}_3}| \\
&\geq \left(|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathcal{P}_t|(1 - 2\delta_n)\right) |\mathbb{C}_3|
\end{aligned}$$

□

Observe that $\mathcal{P}_t$ and $A_\epsilon^n(N_3)$ have the same exponential rate. Note that $|\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus A_\epsilon^n(N_3)| \geq |\mathbb{C}_1 \oplus A_\epsilon^n(N_3)|$ and since decoder 1 can decode X_1 with probability of error approaching 0, we can use the same argument to show the following:

$$\begin{aligned}
&\frac{1}{n} \log |\mathbb{C}_1 \oplus A_\epsilon^n(N_3)| \rightarrow \frac{1}{n} \log |\mathbb{C}_1||A_\epsilon^n(N_3)| \rightarrow \log q - H(N_1) \oplus H(N_3) \\
&\Rightarrow \frac{1}{n} \log |\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbb{C}_3 \oplus A_\epsilon^n(N_3)| \geq \\
&\qquad (\log q - H(N_1) + H(N_3)) + (H(N_1) - H(N_3)) = \log q.
\end{aligned} \tag{C.10}$$

But we know that $\frac{1}{n} \log |\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbb{C}_3 \oplus A_\epsilon^n(N_3)| \leq \log q$. So, we should have equality at all of the inequalities. Hence, $\frac{1}{n} \log_q |\mathbb{C}_1 \oplus \mathbb{C}_2 \oplus A_\epsilon^n(N_3)| \rightarrow \frac{1}{n} \log_q |\mathbb{C}_1 \oplus A_\epsilon^n(N_3)|$. In order to have $|\mathbb{C}_1 \oplus \mathbb{C}_2|$ close to $|\mathbb{C}_1|$, we must have the properties stated in the lemma.

C.1.2 Proof of Lemma 29

We provide a coding scheme based on NQLC's which achieves the rate vector. Consider two ternary random variables V_1 and V_2 such that $H(V_1 \oplus V_2) > H(2V_1 \oplus V_2)$. We will show the achievability of the following rate-triple:

$$R_1 = \log q - H(N_1 \oplus N_2 \oplus N_3)$$
$$R_2 = (\frac{H(V_1 \oplus V_2)}{H(V_1)} - 1)(\log q - H(N_1 \oplus N_2 \oplus N_3))$$
$$R_3 = \log q - H(N_3) - \frac{H(2V_1 \oplus V_2)}{H(V_1)} (\log q - H(N_1 \oplus N_2 \oplus N_3)) .$$

Then, $R_2 + R_3 = H(N_1 \oplus N_2 \oplus N_3) - H(N_3) + \frac{(H(V_1 \oplus V_2) - H(2V_1 \oplus V_2))}{H(V_1)} (\log q - H(N_1 \oplus N_2 \oplus N_3))$. Choose random variable V_3 such that $R_3 = H(V_3 \oplus N_3) - H(N_3)$.

Codebook Generation: Construct a family of pairs NQLC's with length n and parameters $m = 1$, $k_1 = \frac{(\log q - H(N_1 \oplus N_2 \oplus N_3))}{H(V_1)} n$, $U_1 = V_1$, and $U_1' = V_2$ by choosing the dither $\mathbf{b}$ and generator matrix G_1 randomly and uniformly on $\mathbb{F}_q$. For a fixed $n \in \mathbb{N}$, Let C_1^n and C_2^n be the corresponding pair of NQLC's. Let $\Phi_i = 2^{nR_i}$ for $i \in \{1, 2\}$. Choose Φ_i of the codewords in C_i^n randomly and uniformly, and index these sequences using the indices $[1, \Phi_i]$. Also, generate an unstructured codebook C_3 randomly and uniformly with rate R_3 based on the single-letter distribution P_{V_3}. Index C_3 by $[1, 2^{nR_3}]$.

Encoding: Upon receiving message index M_i encoder i sends the sequence in C_1 which is indexed M_i for $i \in \{1, 2\}$. Let the codewords sent by encoder $i, i \in \{1, 2\}$ be denoted by $\mathbf{v}_i G_1 \oplus \mathbf{b}_i$. Encoder 3 sends the codeword in C_3 indexed by M_3. Let the codeword sent by the third decoder be denoted by $\mathbf{c}_3$.

Decoding: Decoder 1 receives $X_1^n \oplus N_1^n \oplus N_2^n \oplus N_3^n$. Using typicality decoding, the decoder can decode the message as long as $\frac{k_1}{n} H(V_1) \leq \log q - H(N_1 \oplus N_2 \oplus N_3)$. Decoder 2 receives $X_1^n \oplus X_2^n \oplus N_2^n \oplus N_3^n = (\mathbf{v}_1 \oplus \mathbf{v}_2) G_1 \oplus \mathbf{b}_1 \oplus \mathbf{b}_2 \oplus N_2^n \oplus N_3^n$. It can decode $\mathbf{v}_1, \mathbf{v}_2$ jointly as long as 1) $\frac{k_1}{n} H(V_1 \oplus V_2) < \log q - H(N_1 \oplus N_3)$, and 2) $R_1 + R_2 \leq \frac{k_1}{n} H(V_1 \oplus V_2)$. The first condition ensures that $\mathbf{v}_1 \oplus \mathbf{v}_2$ can be recovered with probability of error going to 0 as

$n \to \infty$. After recovering $\mathbf{v}_1 \oplus \mathbf{v}_2$, the decoder needs to jointly decode $\mathbf{v}_1, \mathbf{v}_2$ (for reasons explained in Lemma 27). This is a noiseless additive MAC problem and condition 2 ensures errorless decoding. Note that in condition 2, the coefficient $\frac{k_1}{n}$ is present since $\mathbf{v}_1$ is of length k_1. Also, The term $H(V_1 \oplus V_2)$ is the capacity of the MAC channel. Decoder 3 receives $X_1^n \oplus X_2^n \oplus X_3^n \oplus N_3^n = (2\mathbf{v}_1 \oplus \mathbf{v}_2)G_1 \oplus \mathbf{b}_1 \oplus \mathbf{b}_2 \oplus \mathbf{c}_3^n \oplus N_3^n$. The decoder can recover $2\mathbf{v}_1 \oplus \mathbf{v}_2$ as long as $\frac{k_1}{n} H(2V_1 \oplus V_2) < \log q - H(X_3 \oplus N_3)$. Then, the decoder subtracts $2X_1^n \oplus X_2^n$ to get $X_3^n \oplus N_3^n$. It can decode X_3 as long as $R_3 \leq H(V_3 \oplus N_3) - H(N_3)$. It is straightforward to check the rate given at the beginning satisfy all of these bounds.

APPENDIX D

Proofs for Chapter V

D.1 Proofs for Chapter V

D.1.1 Proof of Proposition 15

Proof. By definition, any element of $\mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}$ satisfies the conditions in the proposition. Conversely, we show that any function satisfying the conditions (1) and (2) is in the tensor product. Let $\tilde{f} = \sum_{\mathbf{j}} \tilde{f}_{\mathbf{j}}, \tilde{f}_{\mathbf{j}} \in \mathcal{G}_{j_1} \otimes \mathcal{G}_{j_2} \otimes \cdots \otimes \mathcal{G}_{j_n}$. Assume $i_k = 1$ for some $k \in [1, n]$. Then:

$$0 \overset{(2)}{=} \mathbb{E}_{X^n|X_{\sim i_k}}(\sum_{\mathbf{j}} \tilde{f}_{\mathbf{j}}|X_{\sim i_k}) \overset{(a)}{=} \sum_{\mathbf{j}} \mathbb{E}_{X^n|X_{\sim i_k}}(\tilde{f}_{\mathbf{j}}|X_{\sim i_k}) \overset{(1)}{=} \sum_{\mathbf{j}, j_k=0} \mathbb{E}_{X^n|X_{\sim i_k}}(\tilde{f}_{\mathbf{i}}|X_{\sim i_k}) \overset{(2)}{=} \sum_{\mathbf{j}, j_k=0} \tilde{f}_{\mathbf{j}},$$

where we have used linearity of expectation in (a), and the last two equalities use the fact that $\tilde{f}_{\mathbf{j}} \in \mathcal{G}_{j_1} \otimes \mathcal{G}_{j_2} \otimes \cdots \otimes \mathcal{G}_{j_n}$ which means it satisfies properties (1) and (2). So far we have shown that $\tilde{f} = \sum_{\mathbf{j} \geq \mathbf{i}} \tilde{f}_{\mathbf{j}}$. Now assume $i_{k'} = 1$. Then:

$$\sum_{\mathbf{j} \geq \mathbf{i}} \tilde{f}_{\mathbf{j}} = \tilde{f} \overset{(1)}{=} \mathbb{E}_{X^n|X_{\sim i_{k'}}}(\sum_{\mathbf{j} \geq \mathbf{i}} \tilde{f}_{\mathbf{j}}|X_{\sim i_{k'}}) = \sum_{\mathbf{j} \geq \mathbf{i}} \mathbb{E}_{X^n|X_{\sim i_{k'}}}(\tilde{f}_{\mathbf{j}}|X_{\sim i_{k'}}) = \sum_{\mathbf{j} \geq \mathbf{i}, j_{k'}=1} \tilde{f}_{\mathbf{j}} \Rightarrow \sum_{\mathbf{j} \geq \mathbf{i}, j_{k'}=0} \tilde{f}_{\mathbf{j}} = 0.$$

So, $\tilde{f} = \sum_{\mathbf{i} \geq \mathbf{j} \geq \mathbf{i}} \tilde{f}_{\mathbf{j}} = \tilde{f}_{\mathbf{i}}$. By assumption we have $\tilde{f}_{\mathbf{i}} \in \mathcal{G}_{i_1} \otimes \mathcal{G}_{i_2} \otimes \cdots \otimes \mathcal{G}_{i_n}$. □

D.1.2 Proof of Proposition 16

Proof. 1) We use induction. Let $\mathbf{i}_j, j \in [1, n]$ be the jth element of the standard basis. Then $\tilde{e}_{\mathbf{i}_j} = \mathbb{E}_{X^n|X_j}(\tilde{e}|X_j)$. By the smoothing property of expectation, $\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{i}_j}) = \mathbb{E}_{X^n}(\tilde{e}) = 0$. Assume that $\forall \mathbf{j} < \mathbf{i}$, $\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}) = 0$. Then,

$$\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{i}}) = \mathbb{E}_{X^n}\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}}) - \sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_{\mathbf{j}}\right) = \mathbb{E}_{X^n}(\tilde{e}) - \sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}) = 0 - \sum_{\mathbf{j}<\mathbf{i}} 0 = 0.$$

2) This statement is also proved by induction. $\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})$ is a function of $X_{\mathbf{i}}$, so by induction $\tilde{e}_{\mathbf{i}} = \mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}}) - \sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_{\mathbf{k}}$ is also a function of $X_{\mathbf{i}}$.

3) Let $\mathbf{i}_k, k \in [1, n]$ be defined as the kth element of the standard basis, and take $j, j' \in [1, n], j \neq j'$. We have:

$$\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{i}_j}\tilde{e}_{\mathbf{i}_{j'}}) = \mathbb{E}_{X^n}(\mathbb{E}_{X^n|X_j}(\tilde{e}|X_j)\mathbb{E}_{X^n|X_{j'}}(\tilde{e}|X_{j'})) \overset{(a)}{=} \mathbb{E}_{X^n}(\mathbb{E}_{X^n|X_j}(\tilde{e}|X_j))\mathbb{E}_{X^n}(\mathbb{E}_{X^n|X_{j'}}(\tilde{e}|X_{j'})) \overset{(b)}{=} \mathbb{E}^2_{X^n}(\tilde{e}) = 0,$$

where we have used the memoryless property of the source in (a) and (b) results from the smoothing property of expectation. We extend the argument by induction. Fix $\mathbf{i}, \mathbf{k}$. Assume that $\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}\tilde{e}_{\mathbf{j}'}) = \mathbb{1}(\mathbf{j} = \mathbf{j}')\mathbb{E}_{X^n}(\tilde{e}^2_{\mathbf{j}}), \forall \mathbf{j} < \mathbf{i}, \mathbf{j}' < \mathbf{k}$.

$$\begin{aligned}
\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{i}}\tilde{e}_{\mathbf{k}}) &= \mathbb{E}_{X^n}\left(\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}}) - \sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_{\mathbf{j}}\right)\left(\mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}|X_{\mathbf{k}}) - \sum_{\mathbf{j}'\leq\mathbf{k}} \tilde{e}_{\mathbf{j}'}\right)\right) \\
&= \mathbb{E}_{X_n}\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})\mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}|X_{\mathbf{k}})\right) - \sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X^n}\left(\tilde{e}_{\mathbf{j}}\mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}|X_{\mathbf{k}})\right) - \sum_{\mathbf{j}'\leq\mathbf{k}} \mathbb{E}_{X^n}\left(\tilde{e}_{\mathbf{j}'}\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})\right) \\
&+ \sum_{\mathbf{j}<\mathbf{i},\mathbf{j}'<\mathbf{k}} \mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}\tilde{e}_{\mathbf{j}'}).
\end{aligned}$$

The second and third terms in the above expression can be simplified as follows. First, note that:

$$\tilde{e}_{\mathbf{i}} = \mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}}) - \sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_{\mathbf{j}} \Rightarrow \sum_{\mathbf{j}\leq\mathbf{i}} \tilde{e}_{\mathbf{j}} = \mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}}). \tag{D.1}$$

Our goal is to simplify $\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}\mathbb{E}_{X^n|X_{\mathbf{j}'}}(\tilde{e}|X_{\mathbf{j}'}))$. We proceed by considering two different cases:

Case 1: $\mathbf{i} \nleq \mathbf{k}$ and $\mathbf{k} \nleq \mathbf{i}$:

$$\begin{aligned}
&\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}\mathbb{E}_{X^n|X_{\mathbf{j}'}}(\tilde{e}|X_{\mathbf{j}'})) \overset{(D.1)}{=} \mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}} \sum_{\underline{l}\leq\mathbf{j}'} \tilde{e}_{\mathbf{j}})) = \sum_{\underline{l}\leq\mathbf{j}'} \mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}\tilde{e}_{\underline{l}}) \\
&= \sum_{\underline{l}\leq\mathbf{j}'} \mathbb{1}(\mathbf{j} = \underline{l})\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}^2) = \mathbb{1}(\mathbf{j} \leq \mathbf{j}')\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}^2).
\end{aligned}$$

Replacing the terms in the original equality we get:

$$\begin{aligned}
\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{i}}\tilde{e}_{\mathbf{k}}) &= \mathbb{E}_{X^n}\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})\mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}|X_{\mathbf{k}})\right) - \sum_{\mathbf{j}<\mathbf{i}} \mathbb{1}(\mathbf{j} \leq \mathbf{j}')\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}^2) \\
&- \sum_{\mathbf{j}'<\mathbf{k}} \mathbb{1}(\mathbf{j}' \leq \mathbf{j})\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}'}^2) + \sum_{\mathbf{j}<\mathbf{i},\mathbf{j}'<\mathbf{k}} \mathbb{1}(\mathbf{j} = \mathbf{j}')\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}^2) \\
&= \mathbb{E}_{X^n}\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})\mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}|X_{\mathbf{k}})\right) - \sum_{\mathbf{j}\leq\mathbf{i}\wedge\mathbf{k}} \mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}^2) \\
&\overset{(a)}{=} \mathbb{E}_{X^n}(\mathbb{E}^2_{X^n|X_{\mathbf{i}\wedge\mathbf{k}}}(\tilde{e}(X^n)|X_{\mathbf{i}\wedge\mathbf{k}})) - \sum_{\mathbf{j}\leq\mathbf{i}\wedge\mathbf{k}} \mathbb{E}_{X^n}(\tilde{e}_{\mathbf{j}}^2) \\
&\overset{(b)}{=} \mathbb{E}_{X^n}(\mathbb{E}^2_{X^n|X_{\mathbf{i}\wedge\mathbf{k}}}(\tilde{e}(X^n)|X_{\mathbf{i}\wedge\mathbf{k}})) - \mathbb{E}_{X^n}\left((\sum_{\mathbf{j}\leq\mathbf{i}\wedge\mathbf{k}} \tilde{e}_{\mathbf{j}})^2 \right) \overset{(D.1)}{=} 0.
\end{aligned}$$

Where in (b) we have used that $\tilde{e}_{\mathbf{i}}$'s are uncorrelated, and (a) is proved below:

$$\begin{aligned}
&\mathbb{E}_{X^n}\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})\mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}|X_{\mathbf{k}})\right)\\
&\quad=\sum_{x_{\mathbf{i}\wedge\mathbf{k}}} P(x_{\mathbf{i}\wedge\mathbf{k}})\left(\left(\sum_{x_{|\mathbf{i}-\mathbf{k}|^+}} P(x_{|\mathbf{i}-\mathbf{k}|^+})\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})\right)\left(\sum_{x_{|\mathbf{k}-\mathbf{i}|^+}} P(x_{|\mathbf{k}-\mathbf{i}|^+})\mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}|X_{\mathbf{k}}\right)\right)\\
&\quad=\sum_{x_{\mathbf{i}\wedge\mathbf{k}}} P(x_{\mathbf{i}\wedge\mathbf{k}})\mathbb{E}^2_{X^n|X_{\mathbf{i}\wedge\mathbf{k}}}(\tilde{e}|x_{\mathbf{i}\wedge\mathbf{k}})\\
&\quad=\mathbb{E}_{X^n}(\mathbb{E}^2_{X^n|X_{\mathbf{i}\wedge\mathbf{k}}}(\tilde{e}(X^n)|X_{\mathbf{i}\wedge\mathbf{k}})).
\end{aligned}$$

Case 2: Assume $\mathbf{i} \leq \mathbf{k}$:

$$\begin{aligned}
\mathbb{E}_{X^n}(\tilde{e}_{\mathbf{i}}\tilde{e}_{\mathbf{k}}) &= \mathbb{E}_{X^n}\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})\mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}|X_{\mathbf{k}})\right) - \sum_{\mathbf{j}<\mathbf{i}} \mathbb{1}(\mathbf{j} \leq \mathbf{j}')\mathbb{E}_{X^n}(\tilde{e}^2_{\mathbf{j}})\\
&\quad- \sum_{\mathbf{j}'\leq\mathbf{k}} \mathbb{1}(\mathbf{j}' \leq \mathbf{j})\mathbb{E}_{X^n}(\tilde{e}^2_{\mathbf{j}'}) + \sum_{\mathbf{j}<\mathbf{i},\mathbf{j}'<\mathbf{k}} \mathbb{1}(\mathbf{j} = \mathbf{j}')\mathbb{E}_{X^n}(\tilde{e}^2_{\mathbf{j}})\\
&= \mathbb{E}_{X^n}(\mathbb{E}^2_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})) - \sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X^n}(\tilde{e}^2_{\mathbf{j}}) - \sum_{\mathbf{j}'\leq\mathbf{i}} \mathbb{E}_{X^n}(\tilde{e}^2_{\mathbf{j}'}) + \sum_{\mathbf{j}\leq\mathbf{i}} \mathbb{E}_{X^n}(\tilde{e}^2_{\mathbf{j}})\\
&= 0.
\end{aligned}$$

Case 3: When $\mathbf{k} \leq \mathbf{i}$ the proof is similar to case 2.

4) Clearly when $|\mathbf{i}| = 1$, the claim holds. Assume it is true for all $\mathbf{j}$ such that $|\mathbf{j}| < \mathbf{i}$. Take $\mathbf{i} \in \{0,1\}^n$ and $t \in [1,n], i_t = 1$ arbitrarily. We first prove the claim for $\mathbf{k} = \mathbf{i} - \mathbf{i}_t$:

$$\begin{aligned}
&\mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{\mathbf{i}}|X_{\mathbf{k}}) = \mathbb{E}_{X^n|X_{\mathbf{k}}}\left(\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}) - \sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_{\mathbf{j}}\right)|X_{\mathbf{k}}\right) = \mathbb{E}_{X^n|X_{\mathbf{k}}}\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})|X_{\mathbf{k}}\right) - \sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{\mathbf{j}}|X_{\mathbf{k}})\\
&\overset{(a)}{=} \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}|X_{\mathbf{k}}) - \sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{\mathbf{j}}|X_{\mathbf{k}}) \overset{(5)}{=} \sum_{\mathbf{j}\leq\mathbf{i}-\mathbf{i}_t} \tilde{e}_{\mathbf{j}} - \sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{\mathbf{j}}|X_{\mathbf{k}})\\
&\overset{(b)}{=} \sum_{\mathbf{j}\leq\mathbf{i}-\mathbf{i}_t} \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{\mathbf{j}}|X_{\mathbf{k}}) - \sum_{\mathbf{j}<\mathbf{i}} \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{\mathbf{j}}|X_{\mathbf{k}}) = \sum_{s\neq t} \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{\mathbf{i}-\mathbf{i}_s}|X_{\mathbf{k}}) \overset{(c)}{=} \sum_{s\neq t} \mathbb{E}_{X^n|X_{\mathbf{k}-\mathbf{i}_s}}(\tilde{e}_{\mathbf{i}-\mathbf{i}_s}|X_{\mathbf{k}-\mathbf{i}_s}) \overset{(d)}{=} 0.
\end{aligned}$$

Where in (a) we have used $\mathbf{i} > \mathbf{k}$, also (b) follows from $\mathbf{j} < \mathbf{k}$, (c) uses $\mathbf{k} \wedge (\mathbf{i} - \mathbf{i}_s) = \mathbf{k} - \mathbf{i}_s$, and finally, (d) uses the induction assumption. Now we extend the result to general $\mathbf{k} < \mathbf{i}$. Fix $\mathbf{k}$. Assume the claim is true for all $\mathbf{j}$ such that $\mathbf{k} < \mathbf{j} < \mathbf{i}$ (i.e $\forall \mathbf{k} < \mathbf{j} < \mathbf{i}, \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{X_{\mathbf{j}}|X_{\mathbf{k}}}) = 0$). We have:

$$
\begin{aligned}
\mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{\mathbf{i}}|X_{\mathbf{k}}) &= \mathbb{E}_{X^n|X_{\mathbf{k}}}\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}}) - \sum_{\mathbf{j}<\mathbf{i}} \tilde{e}_{\mathbf{j}} | X_{\mathbf{k}}\right) = \mathbb{E}_{X^n|X_{\mathbf{k}}}\left(\mathbb{E}_{X^n|X_{\mathbf{i}}}(\tilde{e}|X_{\mathbf{i}})|X_{\mathbf{k}}\right) - \sum_{\mathbf{j}\leq\mathbf{k}} \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}_{\mathbf{j}}|X_{\mathbf{k}}) \\
&= \mathbb{E}_{X^n|X_{\mathbf{k}}}(\tilde{e}|X_{\mathbf{k}}) - \sum_{\mathbf{j}\leq\mathbf{k}} \tilde{e}_{\mathbf{j}} \overset{(D.1)}{=} 0.
\end{aligned}
$$

$\square$

D.1.3 Proof of Lemma 25

Proof. Let the functions be given as follows:

$$
g(X) = \begin{cases} \alpha & , X = 0 \\ \beta & , X = 1. \end{cases}, \qquad h(Y) = \begin{cases} \gamma & Y = 0 \\ \delta & Y = 1. \end{cases}
$$

Also, let $P(X = 1) = p$, and $P(Y = 1) = r$. The 0 mean condition enforces the following equalities:

$$
\alpha(1-p) + \beta p = 0 \Rightarrow \beta = \frac{-(1-p)\alpha}{p}, \qquad \gamma(1-q) + \delta q = 0 \Rightarrow \delta = \frac{-(1-q)\gamma}{q}.
$$

Next, we calculate the joint distribution of P_{XY}. Let $P_{i,j} \triangleq P(X = i, Y = j), i, j \in \{0, 1\}$. We have the following:

$$
\begin{aligned}
P_{0,0} + P_{0,1} &= P(X = 0) = 1 - p, \qquad & P_{0,0} + P_{1,0} &= P(Y = 0) = 1 - q, \\
P_{0,0} + P_{1,1} &= P(X = Y) = 1 - \epsilon, \qquad & P_{0,0} + P_{0,1} + P_{1,0} + P_{1,1} &= 1.
\end{aligned}
$$

Solving the system of equations yields:

$$P_{0,0} = 1 - \frac{p+q+\epsilon}{2}, \qquad P_{0,1} = \frac{q+\epsilon-p}{2}, \qquad P_{1,0} = \frac{p+\epsilon-q}{2}, \qquad P_{1,1} = \frac{p+q-\epsilon}{2}. \tag{D.2}$$

With the following constraint on the variables:

$$p+\epsilon \geq q, \qquad p+q \geq \epsilon, \qquad q+\epsilon \geq p, \qquad p+q+\epsilon \leq 2.$$

We have:

$$\begin{aligned}
&\frac{\mathbb{E}_{X,Y}(gh)}{\mathbb{E}_X^{\frac{1}{2}}(g^2)\mathbb{E}_Y^{\frac{1}{2}}(h^2)} = \frac{\alpha\gamma\left(P_{0,0} - P_{0,1}\frac{(1-q)}{q} - P_{1,0}\frac{(1-p)}{p} + P_{1,1}\frac{(1-q)(1-p)}{pq}\right)}{\alpha\gamma\left(\left((1-p)+\frac{(1-p)^2}{p}\right)^{\frac{1}{2}}\left((1-q)+\frac{(1-q)^2}{q}\right)^{\frac{1}{2}}\right)} \\
&= \frac{P_{0,0} - P_{0,1}\frac{(1-q)}{q} - P_{1,0}\frac{(1-p)}{p} + P_{1,1}\frac{(1-q)(1-p)}{pq}}{(\frac{1-p}{p})^{\frac{1}{2}}(\frac{1-q}{q})^{\frac{1}{2}}} \\
&= \frac{P_{0,0}pq - P_{0,1}(1-q)p - P_{1,0}(1-p)q + P_{1,1}(1-q)(1-p)}{(pq(1-p)(1-q))^{\frac{1}{2}}} \\
&\overset{(D.2)}{=} \frac{(1-\frac{p+q+\epsilon}{2})pq - (\frac{q+\epsilon-p}{2})(1-q)p - (\frac{p+\epsilon-q}{2})(1-p)q + (\frac{p+q-\epsilon}{2})(1-q)(1-p)}{(pq(1-p)(1-q))^{\frac{1}{2}}} \\
&= \frac{pq + (\frac{p+q}{2})\left((1-p)(1-p) - pq\right) + (\frac{q-p}{2})\left(q(1-p) - p(1-q)\right))}{(pq(1-p)(1-q))^{\frac{1}{2}}} + \\
&\frac{\frac{\epsilon}{2}(pq + p(1-q) + q(1-p) + (1-p)(1-q))}{(pq(1-p)(1-q))^{\frac{1}{2}}} \\
&= \frac{pq + \frac{p+q}{2}(1-p-q) - \frac{p-q}{2}(q-p) - \frac{\epsilon}{2}}{(pq(1-p)(1-q))^{\frac{1}{2}}} \\
&= \frac{p+q-2pq-\epsilon}{2\left(pq(1-p)(1-q)\right)^{\frac{1}{2}}}.
\end{aligned} \tag{D.3}$$

We calculate the optimum point by taking partial derivatives:

$$\frac{\delta}{\delta p}\frac{\mathbb{E}_{X,Y}(gh)}{\mathbb{E}_X^{\frac{1}{2}}(g^2)\mathbb{E}_Y^{\frac{1}{2}}(h^2)} = 0$$
$$\Rightarrow 2(1-2q)(pq(1-p)(1-q))^{\frac{1}{2}} - \frac{(1-2p)}{\sqrt{p(1-p)}}\sqrt{q(1-q)}(p+q-2pq-\epsilon) = 0$$
$$\stackrel{(a)}{\Rightarrow} 2(1-2q)p(1-p) - (1-2p)(p+q-2pq-\epsilon) = 0$$
$$\Rightarrow 2p(1-p)(1-2q) - p(1-2p)(1-2q) - (1-2p)q + (1-2p)\epsilon = 0$$
$$\Rightarrow p(1-2q) - (1-2p)q + (1-2p)\epsilon = 0$$
$$\Rightarrow p - q + (1-2p)\epsilon = 0. \tag{D.4}$$

Where in (a) we have used $p, q \notin \{0,1\}$ to multiply by $\sqrt{pq(1-p)(1-q)}$. Taking the partial derivative with respect to q, by similar calculations we get:

$$\frac{\delta}{\delta q}\frac{\mathbb{E}_{X,Y}(gh)}{\mathbb{E}_X^{\frac{1}{2}}(g^2)\mathbb{E}_Y^{\frac{1}{2}}(h^2)} = 0 \rightarrow q - p + (1-2q)\epsilon. \tag{D.5}$$

In order for (D.4) and (D.5) to be satisfied simultaneously, we must have $\epsilon = 0, p = q$, or $\epsilon = p + q = 1$, or $p = q = \frac{1}{2}$. For $\epsilon \notin \{0,1\}$, we must have $p = q = \frac{1}{2}$ in which case the value in (D.3) is:

$$\frac{\mathbb{E}_{X,Y}(gh)}{\mathbb{E}_X^{\frac{1}{2}}(g^2)\mathbb{E}_Y^{\frac{1}{2}}(h^2)} = 1 - 2\epsilon.$$

This completes the proof. □

APPENDIX E

Proofs for Chapter VI

E.1 Proofs for Chapter VI

E.1.1 Proof of Lemma 5

First note that for $i \neq i'$ since $\tilde{S}(i, 1:n)$ is a function of $(X(i, 1:n), Z(i, 1:n))$ and $\tilde{S}(i', 1:n)$ is a function of $(X(i', 1:n), Z(i', 1:n))$, $\tilde{S}(i, 1:n)$ and $\tilde{S}(i', 1:n)$ are independent of each other. So we only need to prove that $S(i, j)$ are identically distributed for all i, j. We have:

$$
\begin{aligned}
P(\tilde{S}(i,j)=1) &= P(X(i,\pi_i(j)) + \hat{V}(i,\pi_i(j)) + Z(i,\pi_i(j)) = 1) \\
&\stackrel{a}{=} p * P(X(i,\pi_i(j)) + \hat{V}(i,\pi_i(j)) = 1) \\
&\stackrel{b}{=} p * \frac{1}{n} \sum_{j'=1}^{n} E(w_H(X(i,j') + \hat{V}(i,j'))) \\
&\stackrel{c}{=} p * \delta
\end{aligned}
$$

(a) is true because $Z(1:m, 1:n)$ is independent of $X(1:m, 1:n)$ and $X(i,\pi_i(j)) + \hat{V}(i,\pi_i(j))$ is a function of $X(1:m, 1:n)$, (b) is true since the choice of π_i is independent of the source sequences, and (c) is correct since the average distortion of $C_f^{(n)}$ is δ.

E.1.2 Proof of Lemma 14

Proof. 1)

$$P(W_1^n = W_2^n) \geq P(S_1^n = S_2^n) = 1 - \epsilon.$$

The first inequality is true since if $S_1^n = S_2^n$, then $W_1^n = W_2^n$.

2) Let 1_A be the indicator function of event A. We have:

$$H(W_2^n|W_1^n) = H(1_{\{W_1^n=W_2^n\}}, W_2^n|W_1^n) \leq H(1_{\{W_1^n=W_2^n\}}) + H(W_2^n|W_1^n, 1_{\{W_1^n=W_2^n\}})$$
$$\leq h_b(\epsilon) + P(W_1^n \neq W_2^n)H(W_2^n|W_1^n, 1_{\{W_1^n=W_2^n\}} = 0) \leq h_b(\epsilon) + P(W_1^n \neq W_2^n)H(W_2^n)$$
$$\leq h_b(\epsilon) + \epsilon n \log|\mathcal{W}_2|.$$

$\square$

E.1.3 Proof of Lemma 15

Proof. 1)

$$P'_{X_1,X_2,W_1,U_1,U_2} = \sum_{w_2} P'_{X_1,X_2,W_1,W_2,U_1,U_2} \geq \sum_{w_2} 1_{\{W_1=W_2\}}P'_{X_1,X_2,W_1,W_2,U_1,U_2}$$
$$= \sum_{w_2} 1_{\{W_1=W_2\}}P'_{X_1,X_2,W_1,W_2}P_{U_1|X_1,W_1}P_{U_2|X_2,W_1} = P_{U_1|X_1,W_1}P_{U_2|X_2,W_1}\sum_{w_2} 1_{\{W_1=W_2\}}P'_{X_1,X_2,W_1,W_2}$$
$$\overset{a}{\geq} P_{U_1|X_1,W_1}P_{U_2|X_2,W_1}(P_{X_1,X_2,W_1} - \epsilon) \geq P_{X_1,X_2,W_1,U_1,U_2} - \epsilon.$$

Where part (a) results from the following:

$$P_{X_1,X_2,W_1} = P'_{X_1,X_2,W_1} = \sum_{w_2} P'_{X_1,X_2,W_1,W_2} \leq \sum_{w_2} 1_{\{W_1=W_2\}}P'_{X_1,X_2,W_1,W_2} + P(1_{W_1\neq W_2}).$$

Now we prove the other side of the inequality:

$$P'_{X_1,X_2,W_1,U_1,U_2} = \sum_{w_2} P'_{X_1,X_2,W_1,W_2,U_1,U_2} \leq \sum_{w_2} 1_{W_1=W_2} P'_{X_1,X_2,W_1,W_2,U_1,U_2} + \epsilon$$

$$\leq P_{X_1,X_2,W_1} P_{U_1|X_1,W_1} P_{U_2|X_1,W_1} + \epsilon = P_{X_1,X_2,W_1,U_1,U_2} + \epsilon.$$

2) and 3) follow from 1 in a straightforward manner by expanding the mutual informations, the maximum difference between the terms in the mutual informations is $2\log\frac{p_i}{p_i-\epsilon}$, for the sake of brevity we omit the proofs for these parts. □

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Ahlswede, R. (1985), The rate-distortion region for multiple descriptions without excess rate, *IEEE Transactions on Information Theory*, *31*(6), 721–726, doi:10.1109/TIT.1985.1057102.

[2] Akyol, E., K. Viswanatha, and K. Rose (2012), Combinatorial message sharing and random binning for multiple description coding, in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pp. 1371–1375, doi:10.1109/ISIT.2012.6283485.

[3] Anderson, I. (1987), *Combinatorics of Finite Sets*, The Clatendon Press Oxford University, New York.

[4] Bogdanov, A., and E. Mossel (2011), On extracting common random bits from correlated sources, *IEEE Transactions on Information Theory*, *57*(10), 6351–6355, doi:10.1109/TIT.2011.2134067.

[5] C. Nair, K. X., and M. Yazdanpanah (2015), Sub-optimality of the han–kobayashi achievable region for interference channels, *arXiv preprint arXiv:1502.0258*.

[6] Cadambe, V. R., and S. A. Jafar (2008), Interference alignment and degrees of freedom of the -user interference channel, *IEEE Transactions on Information Theory*, *54*(8), 3425–3441, doi:10.1109/TIT.2008.926344.

[7] Chaharsooghi, F. S., A. G. Sahebi, and S. S. Pradhan (2013), Distributed source coding in absence of common components, in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pp. 1362–1366, doi:10.1109/ISIT.2013.6620449.

[8] Costa, M., and A. E. Gamal (1987), The capacity region of the discrete memoryless interference channel with strong interference (corresp.), *IEEE Transactions on Information Theory*, *33*(5), 710–711, doi:10.1109/TIT.1987.1057340.

[9] Csiszár, I., and J. Korner (1981), *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press Inc. Ltd.

[10] Csiszar, I., and P. Narayan (2000), Common randomness and secret key generation with a helper, *IEEE Transactions on Information Theory*, *46*(2), 344–366, doi:10.1109/18.825796.

[11] Gacs, P., and J. Körner (1972), Common information is far less than mutual information, *Problems of Control and Information Theory*, *2*(2), 119–162.

[12] Gallager, R. G. (1968), *Information Theory and Reliable Communication*, Wiley, New York.

[13] Gamal, A. E., and T. Cover (1982), Achievable rates for multiple descriptions, *IEEE Transactions on Information Theory*, *28*(6), 851–857, doi: 10.1109/TIT.1982.1056588.

[14] Gamal, A. E., and Y. H. Kim (2011), *Network Information Theory*, Cambridge University Press.

[15] Goyal, V. K. (2001), Multiple description coding: compression meets the network, *IEEE Signal Processing Magazine*, *18*(5), 74–93, doi:10.1109/79.952806.

[16] Han, T., and K. Kobayashi (1981), A new achievable rate region for the interference channel, *IEEE Transactions on Information Theory*, *27*(1), 49–60, doi: 10.1109/TIT.1981.1056307.

[17] Kleitman, D., and G. Markowsky (1975), On dedekind's problem: the number of isotone boolean functions. ii, *Trans. Amer. Math. Soc.*, *213*, 373–390.

[18] Korner, J., and K. Marton (1979), How to encode the modulo-two sum of binary sources (corresp.), *IEEE Transactions on Information Theory*, *25*(2), 219–221, doi: 10.1109/TIT.1979.1056022.

[19] Kostina, V., and S. Verdu (2012), Fixed-length lossy compression in the finite blocklength regime, *IEEE Transactions on Information Theory*, *58*(6), 3309–3338, doi: 10.1109/TIT.2012.2186786.

[20] Krithivasan, D., and S. S. Pradhan (2011), Distributed source coding using abelian group codes: A new achievable rate-distortion region, *IEEE Transactions on Information Theory*, *57*(3), 1495–1519, doi:10.1109/TIT.2010.2103852.

[21] Lalitha, V., N. Prakash, K. Vinodh, P. V. Kumar, and S. S. Pradhan (2011), A nested linear codes approach to distributed function computation over subspaces, in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pp. 1202–1209, doi:10.1109/Allerton.2011.6120304.

[22] Loeliger, H. A. (1991), Signal sets matched to groups, *IEEE Trans. Inform. Theory*, *37*(6), 1675–1682.

[23] Maddah-Ali, M. A., A. S. Motahari, and A. K. Khandani (2008), Communication over mimo x channels: Interference alignment, decomposition, and performance analysis, *IEEE Transactions on Information Theory*, *54*(8), 3457–3470, doi: 10.1109/TIT.2008.926460.

[24] Mahajan, A., A. Nayyar, and D. Teneketzis (2008), Identifying tractable decentralized control problems on the basis of information structure, in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1440–1449, doi:10.1109/ALLERTON.2008.4797732.

[25] Marton, K. (1979), A coding theorem for the discrete memoryless broadcast channel, *IEEE Transactions on Information Theory*, *25*(3), 306–311, doi: 10.1109/TIT.1979.1056046.

[26] Nazer, B., and M. Gastpar (2007), Computation over gaussian multiple-access channels, in *2007 IEEE International Symposium on Information Theory*, pp. 2391–2395, doi:10.1109/ISIT.2007.4557577.

[27] Ozarow, L. (1980), On a source-coding problem with two channels and three receivers, *The Bell System Technical Journal*, *59*(10), 1909–1921, doi:10.1002/j.1538-7305.1980.tb03344.x.

[28] Padakandla, A., and S. S. Pradhan (2013), Achievable rate region for three user discrete broadcast channel based on coset codes, in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pp. 1277–1281, doi: 10.1109/ISIT.2013.6620432.

[29] Padakandla, A., A. G. Sahebi, and S. S. Pradhan (2012), A new achievable rate region for the 3-user discrete memoryless interference channel, in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pp. 2256–2260, doi: 10.1109/ISIT.2012.6283913.

[30] Philosof, T., and R. Zamir (2009), On the loss of single-letter characterization: The dirty multiple access channel, *IEEE Transactions on Information Theory*, *55*(6), 2442–2454, doi:10.1109/TIT.2009.2018174.

[31] Philosof, T., R. Zamir, U. Erez, and A. J. Khisti (2011), Lattice strategies for the dirty multiple access channel, *IEEE Transactions on Information Theory*, *57*(8), 5006–5035, doi:10.1109/TIT.2011.2158883.

[32] Pinsker, M. S. (1964), *Information and information stability of random variables and processes*, first edition ed., Holden-Day.

[33] Pradhan, S. S., R. Puri, and K. Ramchandran (2004), n-channel symmetric multiple descriptions - part i: (n, k) source-channel erasure codes, *IEEE Transactions on Information Theory*, *50*(1), 47–61, doi:10.1109/TIT.2003.821998.

[34] Reed, M., and B. Simon (1972), *Methods of Modern Mathematical Physics, I: Functional Analysis*, Academic Press Inc. Ltd., New York.

[35] Sahebi, A. G. (2014), Group, lattice and polar codes for multi-terminal communications, Ph.D. thesis, Univeristy of Michigan, Ann Arbor, MI.

[36] Sahebi, A. G., and S. S. Pradhan (2012), Nested lattice codes for arbitrary continuous sources and channels, in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pp. 626–630, doi:10.1109/ISIT.2012.6284268.

[37] Sato, H. (1981), The capacity of the gaussian interference channel under strong interference (corresp.), *IEEE Transactions on Information Theory*, *27*(6), 786–788, doi: 10.1109/TIT.1981.1056416.

[38] Shannon, C. E. (1948), A mathematical theory of communication, *The Bell System Technical Journal*, *27*, 379–423.

[39] Shannon, C. E. (1959), Coding theorems for a discrete source with a fidelity criterion, *IRE Nat. Cov. Rec.*, *7*(4), 142–163.

[40] Shirani, F., and S. S. Pradhan (2014), An achievable rate-distortion region for the multiple descriptions problem, in *2014 IEEE International Symposium on Information Theory*, pp. 576–580, doi:10.1109/ISIT.2014.6874898.

[41] Shirani, F., and S. S. Pradhan (2016), An achievable rate-distortion region for multiple descriptions source coding based on coset codes, *http://arxiv.org/abs/1602.01911*.

[42] Shirani, F., M. Heidari, and S. S. Pradhan (2015), New lattice codes for multiple-descriptions, in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1580–1584, doi:10.1109/ISIT.2015.7282722.

[43] Sridharan, S., A. Jafarian, S. Vishwanath, S. A. Jafar, and S. Shamai (2008), A layered lattice coding scheme for a class of three user gaussian interference channels, in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pp. 531–538, doi:10.1109/ALLERTON.2008.4797604.

[44] Tian, C., and J. Chen (2010), New coding schemes for the symmetric -description problem, *IEEE Transactions on Information Theory*, *56*(10), 5344–5365, doi: 10.1109/TIT.2010.2059651.

[45] Tung, S. Y. (1978), Multiterminal source coding, Ph.D. thesis, Cornell University, Ithaca, NY.

[46] Venkataramani, R., G. Kramer, and V. K. Goyal (2003), Multiple description coding with many channels, *IEEE Transactions on Information Theory*, *49*(9), 2106–2114, doi:10.1109/TIT.2003.815767.

[47] Vinodh, K., V. Lalitha, N. Prakash, P. V. Kumar, and S. S. Pradhan (2010), On the achievable rates of sources having a group alphabet in a distributed source coding setting, in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pp. 479–486, doi:10.1109/ALLERTON.2010.5706945.

[48] Viswanatha, K., E. Akyol, and K. Rose (2011), Combinatorial message sharing for a refined multiple descriptions achievable region, in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pp. 1312–1316, doi: 10.1109/ISIT.2011.6033750.

[49] Viswanatha, K. B., E. Akyol, and K. Rose (2016), Combinatorial message sharing and a new achievable region for multiple descriptions, *IEEE Transactions on Information Theory*, *62*(2), 769–792, doi:10.1109/TIT.2015.2506586.

[50] Wagner, A. B., B. G. Kelly, and Y. Altug (2011), Distributed rate-distortion with common components, *IEEE Transactions on Information Theory*, *57*(7), 4035–4057, doi:10.1109/TIT.2011.2145570.

[51] Wang, J., J. Chen, L. Zhao, P. Cuff, and H. Permuter (2011), On the role of the refinement layer in multiple description coding and scalable coding, *IEEE Transactions on Information Theory*, *57*(3), 1443–1456, doi:10.1109/TIT.2011.2104650.

[52] Wang, Y., A. R. Reibman, and S. Lin (2005), Multiple description coding for video delivery, *Proceedings of the IEEE*, *93*(1), 57–70, doi:10.1109/JPROC.2004.839618.

[53] Wirsenhausen, H. S. (1975), On sequences of pair of dependent random variables, *SIAM Journal of Applied Mathematics*, *28*(1), 100–113.

[54] Zhang, Z., and T. Berger (1987), New results in binary multiple descriptions, *IEEE Transactions on Information Theory*, *33*(4), 502–521, doi:10.1109/TIT.1987.1057330.