ALGEBRAIC STRUCTURES FOR MULTI-TERMINAL COMMUNICATION SYSTEMS

by

Dinesh Krithivasan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2010

Doctoral Committee:

Associate Professor Sandeep P. Sadanandarao, Chair
Professor Robert L. Griess, Jr.
Professor David L. Neuhoff
Associate Professor Achilleas Anastasopoulos

To my family.

ACKNOWLEDGEMENTS

It is a pleasure to thank everyone who made this thesis possible.

I am greatly indebted to my advisor Prof. Sandeep Pradhan for his constant encouragement and his expert guidance. His patience and his belief in my research abilities were strong motivations for me at times when I was frustrated with the slow progress of my research. I have learned from him the importance of keeping the big picture in mind and not getting lost in the minutiae. His high standards of research and disciplined work ethic are things I hope to emulate in my future career. I am especially indebted to him for letting me work remotely on several occasions as I was juggling my PhD with my long distance marriage.

It has been a pleasure to have Prof. David Neuhoff and Prof. Achilleas Anastasopoulos in my thesis committee. Their interest in my research and their knowledge have led to a much improved thesis. I would also like to thank my cognate committee member, Prof. Robert Griess for his interest in my research and for sharing his perspective on it from a different field. For someone with little formal training in group theory such as I, Prof. Griess's vast knowledge of the field and his patience in answering questions helped immensely.

I have had the good fortune of being taught by some excellent instructors during my time in Michigan. I would like to thank my advisor Prof. Pradhan for his multi-terminal information theory course that helped demystify my research area and gave me much needed confidence to tackle my research. The introductory probability

course I took under Prof. Neuhoff back in Fall 2003 is still fresh in my mind and I would like to thank him for giving me a sound fundamental understanding of what is surely the most important subject in my research area. The excellent coding theory course I took under Prof. Achilleas was very useful during my internship at Qualcomm, Inc. I would also like to thank Prof. Demosthenes Teneketzis for teaching me stochastic control and for the many interesting discussions, research related and otherwise, I have had with him over the years. I would also like to thank Prof. Alfred Hero and Prof. Kim Winick for teaching me adaptive signal processing and quantum information theory respectively. It was a great learning experience and a real pleasure to have been a TA for Prof. Petar Momcilovic and Prof. Wayne Stark.

I have had more than a fair share of administrative issues and a big thanks to Becky Turanski, Nancy Goings, Ann Pace, Beth Lawson and all the other administrative staff at the EECS department for guiding me through them. I would also like to thank the wonderful people at the international center, especially Athena Trentin and Emily Jenkins, for their help in my OPT/CPT applications. I had a great time interning twice at Qualcomm Inc., Santa Clara and learned a lot during my time there. I would like to thank all my colleagues there, especially Subra and Andrew Sendonaris for making my internships very enjoyable.

Ann Arbor has been a home away from home for me and I will leave this city with some wonderful memories thanks to a great group of friends I have had here over the years. A big thanks to Shiva, Bala, Sarad, Shyam, Swapnaa, Divya, Manix, Aarthi, Arvind Rao, Soks, Paidi, Aditya Ramamurthy, Preethi, Aditya Mahajan, Vijay, Nitin, Shakthivel, Sampa and many others who have shared my good times and bad and were always there for me when I needed them. Ramji and Ali have been great office-mates who have endured several of my mock presentations before conferences.

Quizclub gave me something to look forward to every week. My undergraduate days were fun-filled thanks to the 6th wing and I would especially like to thank Ashwin, Harish and Sumanth for their help with my job search. To Bharadwaj, Harish, Pradeepan and Krishnaswamy, my best friends from my school days - thanks for all the fun times and memories.

My cousin Mahesh and his wife Vani have been a wonderful neighborly presence who alleviated my home-sickness and made my initial transition to life in the United states so much easier. My brother Rajesh, my sister-in-law Shobana and my niece Nanditha have been a source of encouragement and happiness throughout my PhD. My brother's constant support has played a big role in bringing me to where I am today. My wonderful grandparents, uncle and aunt have been my earliest and best teachers. My parents' unconditional love and support have helped me through several hard times in my life. I often look back and wonder at the many sacrifices they cheerfully made for my and my brother's sake. No amount of thanks could repay them for their kindness but my heartfelt thanks to them all the same. I also wish to thank my in-laws for their patience and understanding. I met my wife Suchitra while doing my PhD and that alone is reason enough to remember my PhD days warmly. She has been a pillar of love and support and the completion of this thesis has in large part been due to her. I look forward to spending the rest of my life with her.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

We study a distributed source coding problem with multiple encoders, a central decoder and a joint distortion criterion. The encoders do not communicate with each other. The encoders observe correlated sources which they quantize and communicate noiselessly to a central decoder which is interested in minimizing a joint distortion criterion that depends on the sources and the reconstruction. We are interested in characterizing an inner bound to the optimal rate-distortion region.

We first consider a special case where the sources are jointly Gaussian and the decoder wants to reconstruct a linear function of the sources under mean square error distortion. We demonstrate a coding scheme involving nested lattice codes that reconstructs the linear function by encoding in such a fashion that the decoder is able to reconstruct the function directly. For certain source distributions, this approach yields a larger rate-distortion region compared to when the decoder reconstructs lossy versions of the sources first and then estimates the function from them. We then extend this approach to the case of reconstructing a linear function of an arbitrary number of jointly Gaussian sources.

Next, we consider the general distributed source coding problem with discrete sources. This formulation includes as a special case many famous distributed source coding problems. We present a new achievable rate-distortion region for this problem based on "good" structured nested random codes built over abelian groups. We demonstrate rate gains for this problem over traditional coding schemes using unstructured random codes. For certain sources and distortion functions, the new rate region is strictly bigger than the Berger-Tung rate region, which has been the best

known achievable rate region for the problem till now. Further, there is no known way of achieving these rate gains without exploiting the structure of the coding scheme. Achievable performance limits for single-user source coding using abelian group codes are also obtained as corollaries of the main coding theorem. Our results also imply that nested linear codes achieve the Shannon rate-distortion bound in the single-user setting. Finally, we conclude by outlining some future research directions.

CHAPTER 1

Introduction

In this thesis, we consider a general distributed source coding problem involving multiple sources, a central decoder and a joint distortion criterion. We first study a special case of the problem when the sources are jointly Gaussian and the decoder is interested in reconstructing a linear function of the sources under mean square distortion criterion. We then consider the general problem for the case of discrete sources and an arbitrary memoryless joint distortion criterion. Our approach for both these problems involves the use of structured random codes which offer rate gains otherwise unattainable using unstructured random codes.

In the following section, we explain the distributed source coding problem that we study.

1.1 Distributed Source Coding

Since its inception in 1973 by Slepian and Wolf [1], the problem of distributed source coding has been a source of inspiration for information/communication/data-compression theory community because of its formidable nature (in its full generality) and its wide scope of practical applications. In this problem, a collection of K correlated information sources, with ith source having an alphabet $\mathcal{X}_i$, is observed separately by K encoders. Each encoder maps its observations into a finite-valued

set. The indices from these sets are transmitted over K noiseless but rate-constrained channels to a joint decoder. The decoder is interested in obtaining L reconstructions with L fidelity criteria (one for each). The ith reconstruction has an alphabet $\hat{\mathcal{Y}}_i$, and the ith fidelity criterion is a mapping from the product of alphabets of a subset of the sources and $\hat{\mathcal{Y}}_i$ to the set of nonnegative real numbers. The goal is to find a computable performance limit for this communication problem. The performance limit, also referred to as the optimal rate-distortion region, is expressed as the set of all $(K + L)$-tuples of rates of the K indices transmitted by the encoders and distortions of the L reconstructions of the decoder that can be achieved in the usual Shannon sense. This problem is graphically illustrated in Figure 1.1.

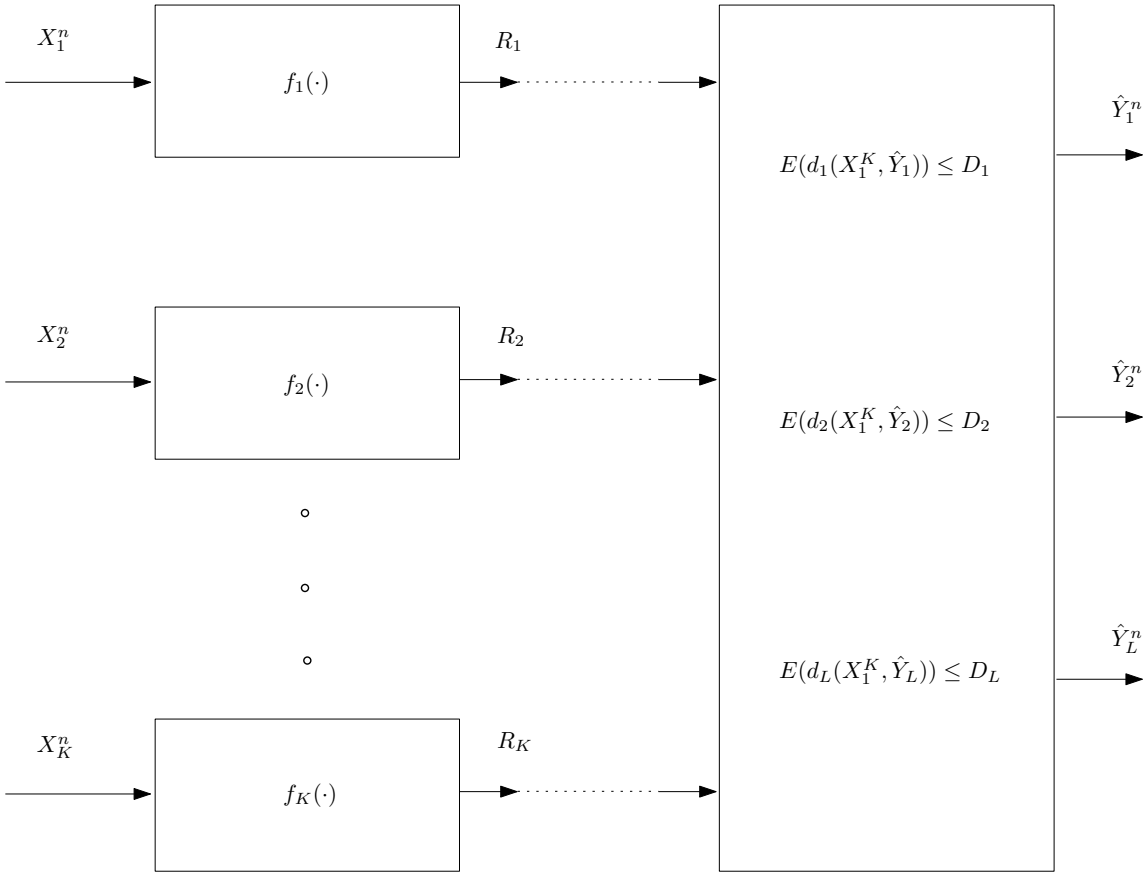

Figure 1.1: A general distributed source coding problem

One of the important motivating forces behind the study of distributed source coding is the problem of information transmission in sensor networks. In a typical application, a group of sensors observe an underlying stochastic field (such as temperature in a locality) and transmit their observations to a central decoder. Since transmission of this information costs battery power which in turn limits the lifetime of the sensors, it is imperative that the sensors encode their observations to minimize the rate of transmission while still meeting certain fidelity requirements at the decoder. If we assume that the link between the sensors and the decoder is noiseless, this problem is exactly modeled by Figure 1.1.

Toward the goal of obtaining the optimal rate-distortion region of the general distributed source coding problem, progress has been made in a number of directions. In the following we restrict our attention to the case of the collection of stationary memoryless sources. In [1], a solution to the problem was given for the case when the decoder wishes to reconstruct all the sources losslessly. In [3, 4], the case of lossless "one-help-one" problem was resolved. Here the decoder wishes to reconstruct only one of the sources[1] losslessly ($K = L + 1 = 2$). In [5], the case of lossy "one-help-one" problem was resolved for the case when the rate of the helper is greater than its entropy (also referred to as the Wyner-Ziv problem). In [6, 7], an achievable inner bound, and a converse outer bound (also known as the Berger-Tung inner and outer bounds respectively) to the performance limit are given for the case where (a) $K = L = 2$ and (b) the fidelity criterion of each source does not depend on the other source (also referred to as independent fidelity criteria). In [8], an inner bound to the performance limit is given for the case of lossy "one-help-one" problem. In

[1]The source which does not enter into any of the fidelity criteria is referred to as a helper. When the rate at which the helper is transmitted is greater than its entropy, the helper is also referred to as side information.

[11], an inner bound to the performance limit is given for the case when the decoder wishes to reconstruct a function of K sources losslessly. It was also shown that this is optimal for the case when the sources are conditionally independent given the function. In [12], the performance limit is given for reconstructing losslessly the modulo-2 sum of two binary correlated sources, and was shown to be tight for the symmetric case. This has been extended to several cases in [14] (see Problem 23 on page 400) and [16]. An improved inner bound was provided for this case in [17]. The key point to note is that the performance limits given in [12, 16, 17] are outside the inner bound given in [11]. In [18], the performance limit is given for the case where (a) $K = L = 2$, (b) one of the sources is reconstructed losslessly and the other with an independent fidelity criterion. In [20] (also see [13, 19, 21, 22, 23, 46]), an inner bound to the performance limit of the CEO problem [2] was given. The CEO problem for the quadratic Gaussian case essentially boils down to reconstructing a certain linear function of the sources with mean squared error fidelity criterion. It was shown that this inner bound is tight for some cases in [27, 32]. For the vector Gaussian CEO problem, inner and outer bounds were derived in [29, 30] and the bounds were shown to be tight under some conditions. In [33], the performance limit is given for the case of lossless reconstruction of a function of two sources with the rate of one of the sources being greater than or equal to its entropy. The lossy version is addressed in [34, 35]. Regarding the Berger-Tung inner bound, it was shown that this is tight for (a) the high-resolution case with independent fidelity criteria in [46], (b) the jointly Gaussian case $K = 2$, $L = 1$ and independent squared error fidelity

[2]This is a variant of the general distributed source coding problem mentioned above. This is closely related to another class of distributed source coding problems known as remote source coding problems. Here the encoders observe a noisy version of the sources. However it can be shown using the techniques of [24, 25] that the remote source coding problems are equivalent to a class of general distributed source coding problems mentioned above.

criterion in [26], and (c) the jointly Gaussian case with $K = 2, L = 2$ and independent squared error criteria in [37]. In [37], it was also shown that a Berger-Tung based coding scheme is optimal for the case of reconstruction of certain linear functions of two jointly Gaussian sources with squared error criterion. A general outer bound to the performance limit of the general distributed source coding problem was given in [31]. In [36], the performance limit was given for the lossy "one-help-many" problem with independent fidelity criteria and the sources being conditionally independent given the helper which is transmitted at a rate greater than its entropy. In [28], the performance limit was given for the quadratic jointly Gaussian lossy "many-help-one" problem with the condition that the helpers are conditionally independent given the source. In [38], the performance limits were obtained for the case of quadratic Gaussian "many-help-one" problem where the sources satisfy a "tree-structure". In [39], the performance limit is given for the case where one of the sources needs to be reconstructed with an independent fidelity criterion and the rest of the sources need to be reconstructed losslessly. In [40], infinite order descriptions (which consist of mutual information terms between two infinite sets of random variables and are thus not computable) were provided for the performance limits of the general case of two terminal source coding problem ($K = 2$) with independent distortion criteria. This was extended to the case of more than two sources in [41].

With regard to above set of results, we would like to make the following observations.

1. Most of the above approaches, except that of [12] and its extensions in [14, 16, 17], use random vector quantization followed by independent random binning (see Chapter 14 of [15]) of the quantizer indices.

2. The four exceptions, which consider only lossless source coding problems, devi-

ate from this norm, and instead use structured random binning based on linear codes on finite fields. Further, the binning operation of the quantizers of the sources are "correlated". This incorporation of structure in binning appears to give improvements in the rates especially for those cases that involve reconstruction of a function of the sources. Moreover, it is still not known whether it is possible to approach this performance without explicitly exploiting the structure of the codebooks.

3. For some distributed source coding problems, whose performance limits were derived using random coding and random binning, it is well-known that these limits can also be approached using structured codes. For example structured codes were considered for (a) the Slepian-Wolf problem in [42], (b) the Wyner-Ziv problem for the binary case with Hamming distortion and for the quadratic Gaussian case in [47], (c) the Berger-Tung inner bound for the two terminal quadratic Gaussian problem with independent fidelity criteria in [47] and (d) high-resolution distributed source coding problem with independent fidelity criteria in [46].

1.2 Contributions

Motivated by the rate gain offered by structured codes over unstructured codes[3]for certain problems, we adopt a similar approach to that of [12] for the general problem of distributed source coding. In particular, we demonstrate the existence of good nested structured codes whose components are "good" codes for source and channel coding for certain appropriately defined notions of "goodness". We consider two

[3]Generally speaking, structured codes are a subset of unstructured codes. However, for the purposes of this thesis, unstructured codes will be taken to mean codes which explicitly lack the structure present in structured codes.

problems below - (a) reconstructing a linear function of jointly Gaussian sources under mean square error distortion (Section 1.2.1), (b) discrete sources with a joint distortion criterion (Section 1.2.2).

1.2.1 Linear Function of Gaussian Sources

We consider a lossy distributed source coding problem with K jointly Gaussian sources with one reconstruction, i.e., $L = 1$. The fidelity criterion has the additional structure that is given by the following. The decoder wishes to reconstruct a linear function of the sources with squared error as the fidelity criterion. We consider a coding scheme with the following structure: sources are quantized using structured vector quantizers followed by "correlated" structured binning. That is, the binning operations of the quantizers of the sources are not performed "independently". The structure used in this process is given by lattice codes using which we provide an inner bound to the optimal rate-distortion region. We show that the proposed inner bound is better for certain parameter values than an inner bound that can be obtained by using a coding scheme that uses random vector quantizers following by independent random binning. For this purpose we use the machinery developed by [43, 44, 47, 48, 49] for the Wyner-Ziv problem in the quadratic Gaussian case.

In Chapter 2, we first consider the case of two jointly Gaussian sources and a decoder interested in reconstructing a linear combination of these sources to within a certain mean squared error distortion. We provide the rate region of our lattice based coding scheme for this case first and then generalize it to the case of arbitrary number of jointly Gaussian sources. For comparison, we also present another inner bound achieved by a scheme that first obtains a lossy reconstruction of the sources, and then obtains a reconstruction of the linear function. The latter scheme is based on the Berger-Tung inner bound. An overall achievable rate region can be obtained

by combining these two schemes. An outer bound is also presented for the two source case through which it is shown that for certain source distributions, the rate region of the lattice based coding scheme is within 1 bit of the optimal rate distortion region. We also provide motivation and intuition about the proposed lattice based coding scheme in this section. We also demonstrate how the general solution simplifies in certain special cases. We then provide a set of numerical results for the two-source case that demonstrate the conditions under which the lattice based scheme performs better than the Berger-Tung based scheme.

1.2.2 Discrete Sources with a Joint Distortion Criterion

In Chapter 3, we consider the distributed source coding problem of Figure 1.1 for the case of discrete sources and arbitrary memoryless distortion criteria. We focus on the case of two sources and one joint distortion criterion. The ideas presented are easily generalizable for the case of any finite number of arbitrary memoryless distortion criteria. For the two user case with one joint distortion criterion, we present an approach based on structured random codes which is very similar in spirit to the coding scheme of Korner and Marton [12] and the lattice based coding scheme of Chapter 2. Our approach relies on the use of nested group codes for encoding. The binning operation of the encoders is done in a "correlated" manner as dictated by these structured codes. This use of "structured quantization followed by correlated binning" is in contrast to the more prevalent "quantization using random codes followed by independent binning" in distributed source coding. Our approach unifies all the known results in distributed source coding such as the Slepian-Wolf problem [1], Korner-Marton problem [12], Wyner-Ahlswede-Korner problem [3, 4], Wyner-Ziv problem [5], Yeung-Berger problem [18] and Berger-Tung problem [7], under a single framework while recovering their respective rate regions. Moreover,

this approach performs strictly better than the standard Berger-Tung based approach for certain source distributions and distortion criteria.

We first present known results for the problem based on the Berger-Tung inner bound. We then motivate our coding scheme which involves the use of nested group codes. We present an overview of the properties of abelian groups in general and cyclic groups in particular that shall be exploited in the proofs. We then present our coding scheme and present an achievable rate region for the problem of distributed source coding involving discrete sources, a central decoder and a joint distortion criterion. We then present various corollaries of our coding theorem. These include achievable rates for lossless and lossy source coding while using abelian group codes. As a further corollary, we show that nested linear codes (built over Galois fields of prime order) can be used to approach the Shannon rate-distortion bound for arbitrary discrete sources and arbitrary distortion measures. This is the first known completely linear encoding scheme that achieves the Shannon bound. We also present achievable rates using group codes for the problem of function reconstruction and present numerical examples for the lossless reconstruction of a linear function of quaternary sources and the lossy reconstruction of the modulo-2 sum of binary sources. By interpreting the problem of function reconstruction of a pair of sources as a 3-user source coding problem with a joint distortion criterion, our results imply that the Berger-Tung inner bound is not tight for the general distributed source coding problem.

1.3 Conclusions and Future Work

In Chapter 4, we summarize the contributions of the thesis and outline the proposed future work. Most of the proofs are given in the appendices.

CHAPTER 2

Linear Function of Jointly Gaussian Sources

The problem of distributed source coding with multiple encoders, a central decoder and a joint distortion criterion was described in the previous chapter motivated by applications relating to sensor networks. In this chapter we consider a special case of this general problem where the sources are jointly Gaussian and the distortion criterion is such that the decoder is interested in reconstructing a linear function of the sources to within a mean-square distortion of D.

The rest of the chapter is organized as follows. In Section 2.1, we give a concise overview of the asymptotic properties of high-dimensional lattices that are known in the literature and which are exploited in the coding theorem and its proof. In Section 2.2, we define the problem formally for the case of two sources and present an inner bound to the optimal rate-distortion region given by a coding structure involving structured quantizers followed by "correlated" structured binning. Further, we also present another inner bound achieved by a scheme that first obtains a lossy reconstruction of the sources, and then obtains a reconstruction of the linear function. The latter scheme is based on the Berger-Tung inner bound. An overall achievable rate region can be obtained by combining these two schemes. Then we present our lattice based coding scheme and prove achievability of the inner bound. We also

provide motivation and intuition about the proposed coding scheme in this section. Finally, we provide an outer bound to the optimal rate distortion region for the two-user case and compare it to our inner bound. In Section 2.3, we consider a generalization of the problem that involves reconstruction of a linear function of an arbitrary finite number of sources. We also demonstrate how the general solution simplifies in certain special cases. In Section 2.3.5, we compare the rate regions of the Berger-Tung based coding scheme and the lattice based coding scheme for low distortions and demonstrate conditions (on the source statistics and the linear function being reconstructed) when the lattice based coding scheme outperforms the Berger-Tung based scheme in this regime. Finally, in Section 2.4, we numerically compare the rate regions of the Berger-Tung based coding scheme and the lattice based coding scheme.

A word about the notation used in this chapter is in order. Let $f(\cdot)$ be an arbitrary function that takes as input a scalar. Then the function $f^n(\cdot)$ takes an n-length vector as input and operates component-wise on the components of that vector. This notation generalizes to functions of more than one variable as well. Variables with superscript n denote an n-length random vector whose components are mutually independent. However, random vectors whose components are not independent are denoted without the use of the superscript. The dimension of such random vectors will be clear from the context.

2.1 Preliminaries on high-dimensional Lattices

2.1.1 Overview of Lattice Codes

Lattice codes [57] play the same role in Euclidean space that linear codes play in Hamming space. Introduction to lattices and to coding schemes that employ lattice

codes can be found in [44, 47, 48, 55, 58]. Lattice codes have been used in other related multiterminal source coding problems in the literature [59, 60, 61, 62, 63]. In the rest of this section, we will briefly review some properties of lattice codes that are relevant to our coding scheme. We start by defining various quantities of interest associated with lattices. We use the same notation as in [47] for these quantities.

An n-dimensional lattice Λ is composed of all integer combinations of the columns of an $n \times n$ matrix G called the generator matrix of the lattice.

$$\Lambda = \{l \in \mathbb{R}^n : \ l = G \cdot i \text{ for some } i \in \mathbb{Z}^n\} \tag{2.1}$$

Associated with every lattice Λ is a natural quantizer namely one that associates with every point in $\mathbb{R}^n$ its nearest lattice point. This quantizer can be described by the function

$$Q_\Lambda(x) \triangleq l \in \Lambda \text{ where } \ \| \ x - l \ \| \leq \| \ x - \hat{l} \ \| \quad \text{ for all } \hat{l} \in \Lambda. \tag{2.2}$$

The quantization error associated with the quantizer $Q_\Lambda(\cdot)$ is defined by

$$x \bmod \Lambda = x - Q_\Lambda(x). \tag{2.3}$$

The basic Voronoi region of a lattice Λ is the set of all points closer to the origin than to any other lattice point, i.e.,

$$\mathcal{V}_0(\Lambda) = \{x \in \mathbb{R}^n \ : \ Q_\Lambda(x) = 0^n\} \tag{2.4}$$

where 0^n is the origin of $\mathbb{R}^n$. The second moment of a lattice Λ is the expected value per dimension of the norm of a random vector uniformly distributed over $\mathcal{V}_0(\Lambda)$ and is given by

$$\sigma^2(\Lambda) = \frac{1}{n} \frac{\int_{\mathcal{V}_0(\Lambda)} \| \ x \ \|^2 \, \mathrm{dx}}{\int_{\mathcal{V}_0(\Lambda)} \mathrm{dx}} \tag{2.5}$$

Let the normalized second moment be give by

$$G(\Lambda) = \frac{\sigma^2(\Lambda)}{V^{2/n}(\Lambda)} \tag{2.6}$$

where $V(\Lambda) = \int_{\mathcal{V}_0(\Lambda)} \mathrm{dx}$. When used as a channel code over an unconstrained AWGN channel with noise Z having variance σ_Z^2 [64], let the probability of decoding error be denoted by

$$P_e(\Lambda, \sigma_Z^2) = Pr(Z^n \notin \mathcal{V}_0) \tag{2.7}$$

where Z^n is the random noise vector of length n.

The mod operation defined in equation (2.3) satisfies the following useful distributive property.

$$((x \bmod \Lambda) + y) \bmod \Lambda = (x + y) \bmod \Lambda \quad \forall\, x, y. \tag{2.8}$$

It is known (see [44] [48]) that the quantization error of a lattice quantizer Λ can be assumed to have a nearly uniform distribution over the fundamental Voronoi region $\mathcal{V}_0$ of the quantizer. This assumption is completely accurate in the case of subtractive dithered quantization where a vector uniformly distributed over $\mathcal{V}_0$ (called the dither) is added at the encoder before quantization and subtracted at the decoder. It has been shown in [44] that for an optimal lattice quantizer, this noise is wide-sense stationary and white. Further, as the lattice dimension $n \to \infty$, for optimal lattice quantizers, the quantization noise approaches a white Gaussian noise process in the Kullback-Leibler divergence sense.

Lattices have been studied extensively for efficient packing and covering. A systematic study of lattice packings was initiated by Minkowski in [51], where existence of good lattice packings was shown. In low dimensions, the maximum lattice packing density have also been studied using Hermite constants (see [57], Chap. 1, page

20). A formal study of lattice covering appears to have been initiated by Kershner in [53]. See [54] for a thorough review of existence of efficient lattice packings and coverings. Lattice codes have been employed in the point-to-point setting for quantization of Gaussian sources with squared error fidelity criterion and also in coding for the AWGN channel with power constraint. In [47], the existence of high dimensional lattices that are "good" for quantization and for coding is discussed. The criteria used therein to define goodness are as follows:

- A sequence of lattices $\Lambda^{(n)}$ (indexed by the dimension n) is said to be a good channel σ_Z^2-code sequence if $\forall \epsilon > 0$, there exists $N(\epsilon)$ such that for all $n > N(\epsilon)$ the following conditions are satisfied:

$$V(\Lambda^{(n)}) < 2^{n(\frac{1}{2}\log(2\pi e\sigma_Z^2)+\epsilon)}, \tag{2.9}$$

$$P_e(\Lambda^{(n)}, \sigma_Z^2) < 2^{-nE(\epsilon)} \tag{2.10}$$

  for some $E(\epsilon) > 0$. The shape of the Voronoi regions of such a good channel lattice code approaches that of an n-dimensional sphere of radius $\sqrt{n\sigma_Z^2}$ as $n \to \infty$. This along with the error criterion implies that such codes achieve the capacity per unit volume of the AWGN channel with additive noise Z [64].

- A sequence of lattices $\Lambda^{(n)}$ (indexed by the dimension n) is said to be a good source D-code sequence if $\forall \epsilon > 0$, there exists $N(\epsilon)$ such that for all $n > N(\epsilon)$ the following conditions are satisfied:

$$\log(2\pi e G(\Lambda^{(n)})) < \epsilon \tag{2.11}$$

$$\sigma^2(\Lambda^{(n)}) = D. \tag{2.12}$$

  Such codes approach the rate-distortion function $R(D)$ of the Gaussian source under mean square error distortion criterion. The shape of the Voronoi regions

of these codes approaches that of an n-dimensional sphere of radius $\sqrt{nD}$ as $n \to \infty$.

2.1.2 Nested Lattice Codes

For lossy coding problems involving side-information at the encoder/decoder, it is natural to consider nested codes. Wyner proposed an algebraic binning approach involving linear codes for the Slepian-Wolf problem [2]. Adapting this scheme to the case of lossy coding, nested codes for the Wyner-Ziv problem were proposed in [45]. We review the properties of nested lattice codes briefly here. Further details can be found in [47].

A pair of n-dimensional lattices (Λ_1, Λ_2) is nested, i.e., $\Lambda_2 \subset \Lambda_1$, if their corresponding generating matrices G_1, G_2 satisfy

$$G_2 = G_1 \cdot J \tag{2.13}$$

where J is an $n \times n$ integer matrix with determinant greater than one. Λ_1 is referred to as the fine lattice while Λ_2 is the coarse lattice. The points of the set

$$\{\Lambda_1 \bmod \Lambda_2\} \triangleq \{\Lambda_1 \cap \mathcal{V}_{0,2}\} \tag{2.14}$$

are called the coset leaders of Λ_2 relative to Λ_1. The nesting ratio of this nested lattice is defined as $\sqrt[n]{V_2/V_1}$ where $V_i = V(\Lambda_i)$ is the volume of the Voronoi region of lattice Λ_i, $i = 1, 2$.

In many applications of nested lattice codes, we require the lattices involved to be a good source code and/or a good channel code. We term a nested lattice (Λ_1, Λ_2) good if (a) the fine lattice Λ_1 is both a good δ_1-source code and a good δ_1-channel code and (b) the coarse lattice Λ_2 is both a good δ_2-source code and a δ_2-channel code. For such a nested lattice code (Λ_1, Λ_2), the number of coset leaders of Λ_2 relative to Λ_1 is

about $(\delta_2/\delta_1)^{n/2}$. A code employing the coset leaders as codewords would thus have a rate of $\frac{1}{2}\log(\delta_2/\delta_1)$. Equivalently, the rate of such a code is the logarithm of the nesting ratio of the nested lattice (Λ_1, Λ_2). A typical encoding operation using such a nested lattice would be as follows: first the source is quantized using the quantizer $Q_{\Lambda_1}(\cdot)$ to a fine lattice point in Λ_1 and then, the coset leader of the quantizer output relative to the coarse lattice Λ_2 is transmitted to the decoder.

The existence of good lattice codes and good nested lattice codes (for various notions of goodness) has been studied in [48, 49] which use the random coding method of [52, 55]. In [49], it was shown that there exists lattices which are simultaneously good in both the source and channel coding senses described above. In [48], the existence of nested lattices where the coarse lattice is simultaneously good as a source and channel code and the fine lattice is a good channel code was proved. In Section 2.2.2, we will describe the notions of goodness that the nested lattice codes used in our coding scheme need to satisfy. We prove the existence of such good nested lattice codes in Appendix A.2.

2.2 Distributed source coding for the two-source case

2.2.1 Problem Statement and Main Result

In this section we consider a distributed source coding problem for the case of two sources X_1 and X_2. The function to be reconstructed at the decoder is assumed to be the linear function $Z \triangleq F(X_1, X_2) = X_1 - cX_2$ unless otherwise specified. Consideration of this function is enough to infer the behavior of any linear function $c_1X_1 + c_2X_2$ and has the advantage of fewer variables. We consider the more general case of $F(X_1, \ldots, X_K) = \sum_{i=1}^{K} c_iX_i$ in Section 2.3.

We define the coding problem formally below. Consider a pair of correlated

jointly Gaussian sources (X_1, X_2) with a given joint distribution $p_{X_1X_2}(x_1, x_2)$. The source sequence (X_1^n, X_2^n) is independent over time and has the product distribution $\prod_{i=1}^n p_{X_1X_2}(x_{1i}, x_{2i})$. Consider the following average squared error as the fidelity criterion: $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ given by

$$d(x^n, y^n) = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2. \tag{2.15}$$

Definition 2.1. Given such a jointly Gaussian distribution $p_{X_1X_2}$ and a distortion function $d(\cdot, \cdot)$ a transmission system with parameters $(n, \theta_1, \theta_2, \Delta)$ is defined as the set of mappings

$$f_i : \mathbb{R}^n \to \{1, 2, \ldots, \theta_i\} \quad \text{for } i = 1, 2 \tag{2.16}$$

$$g : \{1, 2, \ldots, \theta_1\} \times \{1, 2, \ldots, \theta_2\} \to \mathbb{R}^n \tag{2.17}$$

such that the following constraint is satisfied

$$\mathbb{E}\left(d(F^n(X_1^n, X_2^n), g(f_1(X_1^n), f_2(X_2^n)))\right) \leq \Delta. \tag{2.18}$$

Here, $f_i(\cdot)$ represent the source encoders that take as inputs n-length vectors from $\mathbb{R}^n$ and compresses them to an index in the finite set $\{1, \ldots, \theta_i\}$ for $i = 1, 2$. The rates of the encoders are given by $\frac{1}{n}\log\theta_i$. $g(\cdot)$ represents the decoder mapping that takes as input the indices from the two encoders and produces an estimate of the function of the sources as the output. The expected distortion of this reconstruction averaged over the source distribution is given by the LHS of equation (2.18). We say that a tuple (R_1, R_2, D) is achievable if $\forall\epsilon > 0$, $\exists$ for all sufficiently large n, a transmission system with parameters $(n, \theta_1, \theta_2, \Delta)$ such that

$$\frac{1}{n}\log\theta_i \leq R_i + \epsilon \quad \text{for } i = 1, 2$$

$$\Delta \leq D + \epsilon.$$

The performance limit is given by the rate-distortion region which is defined as the set of all achievable tuples (R_1, R_2, D).

Without loss of generality, the sources can be assumed to have unit variance and let the correlation coefficient $\rho > 0$. For the rest of this section, these assumptions are made unless otherwise stated.

One possible coding scheme for this problem would be the following. The decoder reconstructs lossy versions (W_1, W_2) of the sources (X_1, X_2) and uses the best estimate of Z given (W_1, W_2) as the reconstruction $\hat{Z}$. The rate region for such a scheme can be derived using the Berger-Tung inner bound [6, 7]. From here on, this rate region will be referred to as the Berger-Tung based rate region and the associated coding scheme that achieves this rate region will be called the Berger-Tung based coding scheme. The Berger-Tung based rate region is presented in Theorem 2.

The main result in this chapter is to show that for certain parameter values, there exists a better coding scheme that enables the decoder to reconstruct $\hat{Z}$ directly without resorting to reconstructions (W_1, W_2). This coding scheme involves the use of lattice codes and shall be called the lattice based coding scheme from here on. We present the rate region of this scheme below in Theorem 1.

Theorem 1. *The set of all tuples of rates and distortion (R_1, R_2, D) that satisfy*

$$2^{-2R_1} + 2^{-2R_2} \leq \left(\frac{\sigma_Z^2}{D}\right)^{-1} \tag{2.19}$$

are achievable. Here, $\sigma_Z^2 = 1 + c^2 - 2\rho c$ is the variance of the function Z to be reconstructed.

Proof: See Section 2.2.2.

We also present another achievable rate region based on ideas similar to the Berger-Tung coding scheme [6] [7]. From here on, we shall refer to this rate re-

gion as the Berger-Tung based rate region and the scheme that achieves this as the Berger-Tung based coding scheme.

Theorem 2. *Let the region $\mathcal{RD}_{in}$ be defined as follows.*

$$\begin{aligned}(2.20)\quad \mathcal{RD}_{in} = \bigcup_{(q_1,q_2)\in\mathbb{R}_+^2} \Bigg\{ (R_1,R_2,D)\colon\ & R_1 \geq \frac{1}{2}\log\frac{(1+q_1)(1+q_2)-\rho^2}{q_1(1+q_2)}, \\ & R_2 \geq \frac{1}{2}\log\frac{(1+q_1)(1+q_2)-\rho^2}{q_2(1+q_1)}\ , R_1+R_2 \geq \frac{1}{2}\log\frac{(1+q_1)(1+q_2)-\rho^2}{q_1q_2}, \\ & D \geq \frac{q_1\alpha+q_2c^2\alpha+q_1q_2\sigma_Z^2}{(1+q_1)(1+q_2)-\rho^2}\Bigg\}.\end{aligned}$$

where $\alpha \triangleq 1-\rho^2$ and $\mathbb{R}_+$ is the set of positive reals. Then the rate distortion tuples (R_1,R_2,D) which belong to $\mathcal{RD}_{in}^*$ are achievable where * denotes convex closure.

Proof: Follows directly from the application of Berger-Tung inner bound with the auxiliary random variables involved being Gaussian.

In many distributed source coding problems involving jointly Gaussian sources ([27, 32, 37]), the use of Gaussian auxiliary random variables results in the optimal or largest known rate region. It was conjectured in [6, 7] that choosing the auxiliary random variables to be Gaussian indeed results in the optimal rate distortion region for the problem of reconstructing both sources with independent distortion criteria. This was shown to be true in [37]. With this as motivation, we have used Gaussian auxiliary random variables in Theorem 2 above to derive an inner bound to the performance limit of this problem based on the Berger-Tung coding scheme.

We have the following lemma that gives the minimum sum rate of the Berger-Tung based coding scheme which will be used in later sections for comparing the performance limits given by Theorems 1 and 2.

Lemma 2.2. *For a given distortion D, the minimum sum rate $R_{sum} \triangleq R_1+R_2$ that lies in the region $\mathcal{RD}_{in}^*$ of Theorem 2 is given by the lower convex envelope of the*

following region.

$$R_{sum} \geq \frac{1}{2}\log\frac{4c(\alpha c-\rho D)}{D^2} \quad D \leq \min\left\{\frac{2\alpha c}{\rho+c}, \frac{2\alpha c^2}{1+\rho c}\right\} \tag{2.21}$$

$$R_{sum} \geq \frac{1}{2}\log\left(\frac{(1-\rho c)^2}{D-\alpha c^2}\right) \quad \sigma_Z^2 > D > \frac{2\alpha c^2}{1+\rho c},\ c \leq 1 \tag{2.22}$$

$$R_{sum} \geq \frac{1}{2}\log\left(\frac{(c-\rho)^2}{D-\alpha}\right) \quad \sigma_Z^2 > D > \frac{2\alpha c}{\rho+c},\ c > 1 \tag{2.23}$$

$$R_{sum} = 0 \quad D \geq \sigma_Z^2 \tag{2.24}$$

Proof: This derivation is detailed in Appendix A.1.

For certain values of ρ, c and D, the sum-rate given by Theorem 1 is better than that given in Theorem 2. This implies that each rate region contains rate points which are not contained in the other. Thus, an overall achievable rate region for the coding problem can be obtained as the convex closure of the union of all rate distortion tuples (R_1, R_2, D) given in Theorems 1 and 2. A further comparison of the two schemes is presented in Section 2.4. Note that for $c < 0$, it has been shown in [37] that the rate region given in Theorem 2 is tight.

2.2.2 The Coding Scheme

In this section, we present a lattice based coding scheme for the problem of reconstructing the above linear function of two jointly Gaussian sources whose performance approaches the inner bound given in Theorem 1. In what follows, a nested lattice code is taken to mean a sequence of nested lattice codes indexed by the lattice dimension n.

We will require nested lattice codes $(\Lambda_{11}, \Lambda_{12}, \Lambda_2)$ where $\Lambda_2 \subset \Lambda_{11}$ and $\Lambda_2 \subset \Lambda_{12}$. We need the fine lattices Λ_{11} and Λ_{12} to be good source codes (of appropriate second

moment) and the coarse lattice Λ_2 to be a good channel code. The proof of the existence of such nested lattices is detailed in Appendix A.2 where we show the existence of a nested lattice $(\Lambda_{11}, \Lambda_{12}, \Lambda_2)$ such that $\Lambda_2 \subset \Lambda_{11} \subset \Lambda_{12}$ or $\Lambda_2 \subset \Lambda_{12} \subset \Lambda_{11}$ and all three lattices are good source and channel codes simultaneously. The parameters of the nested lattice are chosen to be

$$\sigma^2(\Lambda_{11}) = q_1 \tag{2.25}$$

$$\sigma^2(\Lambda_{12}) = \frac{D\sigma_Z^2}{\sigma_Z^2 - D} - q_1. \tag{2.26}$$

$$\sigma^2(\Lambda_2) = \frac{\sigma_Z^4}{\sigma_Z^2 - D} \tag{2.27}$$

where $0 < q_1 < D\sigma_Z^2/(\sigma_Z^2 - D)$. The coding problem is non-trivial only for $D < \sigma_Z^2$ and in this range, $D\sigma_Z^2/(\sigma_Z^2 - D) < \sigma^2(\Lambda_2)$ and therefore $\Lambda_2 \subset \Lambda_{11}$ and $\Lambda_2 \subset \Lambda_{12}$ indeed. Note that the order of nesting between the lattices Λ_{11} and Λ_{12} depends on whether $q_1 > D\sigma_Z^2/2(\sigma_Z^2 - D)$ or not. However, this is irrelevant for the proof which only requires $\Lambda_2 \subset \Lambda_{11}$ and $\Lambda_2 \subset \Lambda_{12}$.

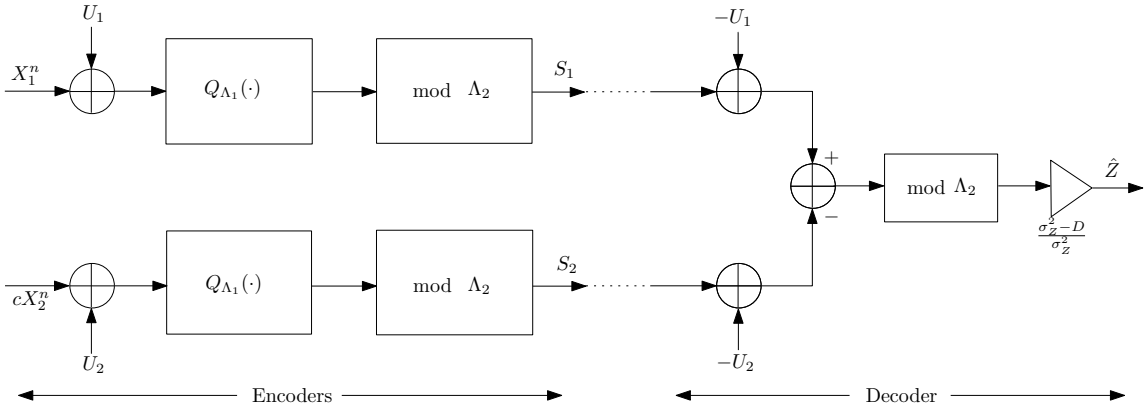

Figure 2.1: Distributed coding using lattice codes to reconstruct $Z = X_1 - cX_2$

Let U_1 and U_2 be random vectors (dithers) that are independent of each other and of the source pair (X_1, X_2). Let U_i be uniformly distributed over the basic Voronoi region $\mathcal{V}_{0,1i}$ of the fine lattices Λ_{1i} for $i = 1, 2$. The decoder is assumed to share

this randomness with the encoders. The source encoders use these nested lattices to quantize X_1 and cX_2 respectively according to equation

$$S_1 = (Q_{\Lambda_{11}}(X_1^n + U_1)) \mod \Lambda_2, \tag{2.28}$$

$$S_2 = (Q_{\Lambda_{12}}(cX_2^n + U_2)) \mod \Lambda_2. \tag{2.29}$$

Note that the second encoder scales the source X_2 before encoding it. The decoder receives the indices S_1 and S_2 and reconstructs

$$\hat{Z} = \left(\frac{\sigma_Z^2 - D}{\sigma_Z^2}\right) ([(S_1 - U_1) - (S_2 - U_2)] \mod \Lambda_2). \tag{2.30}$$

The decoder reconstruction can be intuitively understood as follows. In the low distortion limit as $D \to 0$, the quantization of the fine lattices can be ignored and $S_1 \approx X_1 + U_1, S_2 \approx cX_2 + U_2$. Plugging these approximations (and $D \approx 0$) into equation (2.30) gives us $\hat{Z} = Z \mod \Lambda_2$. Correct decoding occurs if $(Z \mod \Lambda_2) = Z$ which happens with high probability since $\sigma^2(\Lambda_2) > \sigma_Z^2$. A decoding error occurs otherwise. Thus, with high probability, $\hat{Z} = Z$ in the low distortion regime. We present the analysis for the more general case of arbitrary distortion D below.

This coding scheme is illustrated in Fig. 2.1. The rates of the two encoders are given by the logarithm of the nesting ratio of the nested lattices $(\Lambda_{11}, \Lambda_2)$ and $(\Lambda_{12}, \Lambda_2)$. From equations (2.25)-(2.27), it follows that

$$R_1 = \frac{1}{2} \log \frac{\sigma_Z^4}{q_1(\sigma_Z^2 - D)} \tag{2.31}$$

$$R_2 = \frac{1}{2} \log \frac{\sigma_Z^4}{D\sigma_Z^2 - q_1(\sigma_Z^2 - D)} \tag{2.32}$$

Clearly, for a fixed choice of q_1 all rates greater than those given in equations (2.31) and (2.32) are achievable. The union of all achievable rate-distortion tuples (R_1, R_2, D) over all choices of q_1 gives us an achievable region. Eliminating q_1

between the two rate equations gives us

$$2^{2R_2} \geq \frac{1}{\frac{D}{\sigma_Z^2} - 2^{-2R_1}} \tag{2.33}$$

which is the rate region claimed in Theorem 1. It remains to show that this scheme indeed reconstructs the function Z to within a distortion D. We show this in the following.

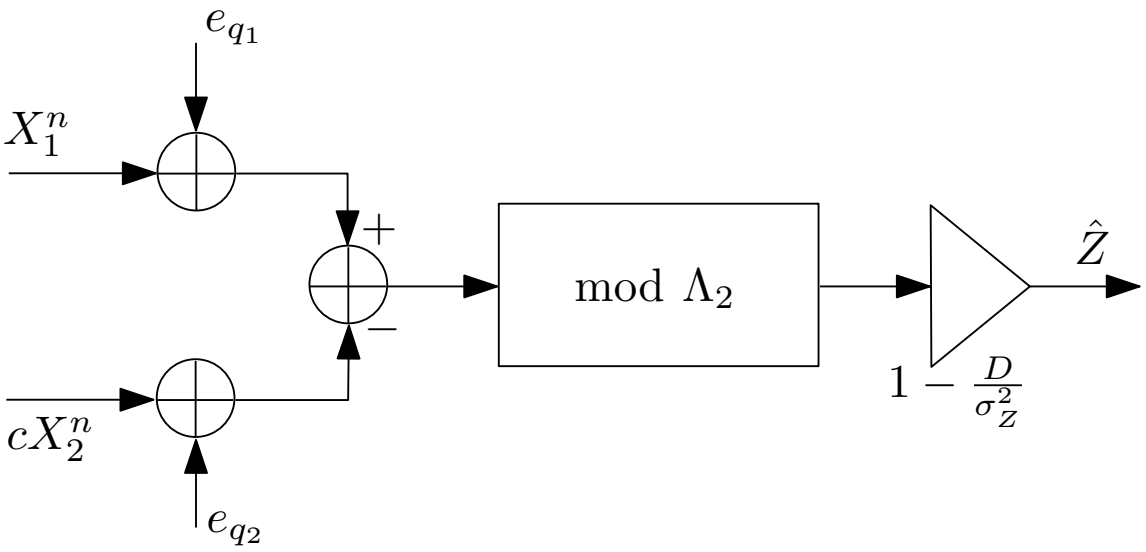

Figure 2.2: Equivalent representation of Fig. 2.1

Using the distributive property of lattices described in equation (2.8), we can reduce the coding scheme to a simpler equivalent scheme by eliminating the first mod-Λ_2 operation in both the signal paths. This results in an equivalent representation of the coding scheme as shown in Fig. 2.2. The decoder can now be described by the equation

$$\hat{Z} = \left(\frac{\sigma_Z^2 - D}{\sigma_Z^2}\right) ([(X_1^n + e_{q_1}) - (cX_2^n + e_{q_2})] \bmod \Lambda_2) \tag{2.34}$$

$$= \left(\frac{\sigma_Z^2 - D}{\sigma_Z^2}\right) ([Z^n + e_{q_1} - e_{q_2}] \bmod \Lambda_2) \tag{2.35}$$

where e_{q_1} and e_{q_2} are dithered lattice quantization noises given by

$$e_{q_1} = Q_{\Lambda_{11}}(X_1^n + U_1) - (X_1^n + U_1), \tag{2.36}$$

$$e_{q_2} = Q_{\Lambda_{12}}(cX_2^n + U_2) - (cX_2^n + U_2). \tag{2.37}$$

The subtractive dither quantization noise e_{q_i} is independent of both sources X_1 and X_2 and has the same distribution as $-U_i$ for $i = 1, 2$ [47]. Since the dithers U_1 and U_2 are independent and for a fixed choice of the nested lattice e_{q_i} is a function of U_i alone, e_{q_1} and e_{q_2} are independent as well.

Let $e_q = e_{q_1} - e_{q_2}$ be the effective dither quantization noise. The decoder reconstruction in equation (2.35) can be simplified as

$$\hat{Z} = \left(\frac{\sigma_Z^2 - D}{\sigma_Z^2}\right)([Z^n + e_q] \bmod \Lambda_2) \tag{2.38}$$

$$\stackrel{c.d}{=} \left(\frac{\sigma_Z^2 - D}{\sigma_Z^2}\right)(Z^n + e_q) \tag{2.39}$$

$$= Z^n + \left(\left(\frac{\sigma_Z^2 - D}{\sigma_Z^2}\right)e_q - \frac{D}{\sigma_Z^2}Z^n\right) \tag{2.40}$$

$$\triangleq Z^n + N. \tag{2.41}$$

We declare a decoding error if the equality in equation (2.39) does not hold. The $\stackrel{\text{c.d}}{=}$ in equation (2.39) stands for equality under the assumption of correct decoding. We show below that this definition of correct decoding is equivalent to the decoder reconstruction $\hat{Z}$ being within mean square error distortion D of $Z = X_1 - cX_2$. Let P_e be the probability of decoding error. Assuming correct decoding, the distortion achieved by this scheme is the second moment per dimension[1] of the random vector N in equation (2.41). This can be expressed as

$$\frac{\mathbb{E} \parallel N \parallel^2}{n} = \left(\frac{\sigma_Z^2 - D}{\sigma_Z^2}\right)^2 \frac{\mathbb{E} \parallel e_q \parallel^2}{n} + \left(\frac{D}{\sigma_Z^2}\right)^2 \frac{\mathbb{E} \parallel Z^n \parallel^2}{n} \tag{2.42}$$

[1]We refer to this quantity also as the normalized second moment of the random vector N. This should not be confused with the normalized second moment of a lattice as defined in equation (2.6).

where we have used the independence of e_{q_1} and e_{q_2} to each other and to the sources X_1 and X_2 (and therefore to $Z = X_1 - cX_2$). Since e_{q_i} has the same distribution as $-U_i$, their expected norm per dimension is just the second moment of the corresponding lattice $\sigma^2(\Lambda_{1i})$. Thus the effective distortion achieved by the scheme is

$$\frac{1}{n}\mathbb{E}\|Z^n - \hat{Z}\|^2 = \left(\frac{\sigma_Z^2 - D}{\sigma_Z^2}\right)^2 \left(\frac{D\sigma_Z^2}{\sigma_Z^2 - D}\right) + \frac{D^2\sigma_Z^2}{\sigma_Z^4}$$

$$= D. \tag{2.43}$$

Hence, the proposed scheme achieves the desired distortion provided correct decoding occurs at equation (2.39). Let us now prove that equation (2.39) indeed holds with high probability for an optimal choice of the nested lattice, i.e., there exists a nested lattice code for which $P_e \to 0$ as $n \to \infty$ where,

$$P_e = Pr\left((Z^n + e_q) \bmod \Lambda_2 \neq (Z^n + e_q)\right). \tag{2.44}$$

To this end, let us first compute the normalized second moment of $(Z^n + e_q)$.

$$\frac{\mathbb{E} \parallel Z^n + e_q \parallel^2}{n} = \frac{\mathbb{E} \parallel Z^n \parallel^2}{n} + \frac{\mathbb{E} \parallel -U_1 - U_2 \parallel^2}{n} \tag{2.45}$$

$$= \sigma_Z^2 + q_1 + \frac{\sigma_Z^2 D}{\sigma_Z^2 - D} - q_1 \tag{2.46}$$

$$= \frac{\sigma_Z^4}{\sigma_Z^2 - D} = \sigma^2(\Lambda_2). \tag{2.47}$$

It was shown in [44] that as $n \to \infty$, the quantization noises e_{q_i} tend to a white Gaussian noise for an optimal choice of the nested lattice. The following lemma states that e_q also converges in the same way.

Lemma 2.3. *If the two independent subtractive dither quantization noises e_{q_i} tend to a white Gaussian noise of the same variance as e_{q_i} in the Kullback-Leibler divergence sense, then $e_q = e_{q_1} - e_{q_2}$ also tends to a white Gaussian noise of the same variance as e_q in the divergence sense.*

Proof: The proof of convergence to Gaussianity of e_q is detailed in Appendix A.3.

We choose Λ_2 to be an exponentially good channel code in the sense defined in Section 2.1.1 (also see [47]). For such lattices, the probability of decoding error P_e in equation (2.44) goes to 0 exponentially fast if $(Z^n + e_q)$ is Gaussian. It can be shown that if $(Z^n + e_q)$ tends to a white Gaussian noise vector, the effect on P_e of the deviation from Gaussianity is sub-exponential. Hence, the overall error behavior is asymptotically the same as the behavior if $(Z^n + e_q)$ were Gaussian, i.e., $P_e \to 0$ exponentially as $n \to \infty$. The proof is similar to the one presented in [48] and is given in Appendix A.5. This implies that the reconstruction error $Z^n - \hat{Z}$ tends in probability to the random vector N defined in equation (2.41). Since all random vectors involved have finite normalized second moment, this convergence in probability implies convergence in second moment as well. Thus the normalized second moment of the reconstruction error tends to that of N which is shown to be D in equation (2.43). Averaged over the random dithers U_1 and U_2, we have shown that the appropriate distortion is achieved. Hence there must exist a pair of deterministic dithers that also achieve the given distortion. Combining equations (2.33) and (2.43), we have proved the claim of Theorem 1.

Remark: Instead of focussing on the entire rate region, if one is interested in minimizing the sum rate of the encoders, then it can be checked that the optimal choice of lattice parameters is $\sigma^2(\Lambda_{11}) = \sigma^2(\Lambda_{12}) = \frac{1}{2}\frac{D\sigma_Z^2}{\sigma_Z^2 - D}$. In this case, we require only one nested lattice (Λ_1, Λ_2) with both encoders using the same nested lattice for encoding.

2.2.3 Intuition about the Coding Scheme

In this section, we outline some arguments that justify our choice of lattice codes and the scaling constants described in the previous subsection. Our use of lattice codes is motivated by the following. Suppose there exists a centralized encoder that has access to both sources X_1 and X_2. Clearly, the optimal encoding strategy then would be to compute $Z = X_1 - cX_2$, quantize and bin it using an encoder, say $f(\cdot)$, that achieves the optimal rate distortion function of a Gaussian source of variance σ_Z^2. Such a centralized coding scheme can be adapted to a distributed setting if the binning operation $f(\cdot)$ *distributes* over the linear function $X_1 - cX_2$ in the sense described by equation (2.48). For then, from the decoder's perspective, there is no distinction between the centralized and distributed coding scheme since

$$f(X_1 - cX_2) = f(X_1) - f(cX_2). \tag{2.48}$$

A lattice code satisfies the functional form mentioned in equation (2.48) and is known to achieve the optimal rate distortion function for Gaussian sources. Hence it is an ideal candidate for use as the source encoder.

The parameters of the lattice code as given in equations (2.25) and (2.26) can be justified as below. Without loss of generality, let the second source alone be scaled by an arbitrary constant η. Let the fine lattices in the signal path of the two sources have second moments $q_i \triangleq \sigma^2(\Lambda_{i,1})$ for $i = 1, 2$. For the case of optimal lattices in high enough dimensions, one can think of quantization using the fine lattices $\Lambda_{i,1}, i = 1, 2$ as simulating an AWGN channel of noise variance q_i, i.e., the subtractive dither quantization noises approach a white Gaussian noise of variance q_i. Such a statement can be made precise by analysis similar to the one carried out in the previous subsection. Let $Q_i, i = 1, 2$ be $\mathcal{N}(0, q_i)$ random variables that are

single-letter asymptotic equivalents of the subtractive dither quantization noises e_{q_i} encountered in the previous subsection.

Referring to the equivalent coding scheme represented in Fig. 2.2, we see that it suffices to choose the coarse lattice Λ_2 to be a good AWGN channel code of second moment equal to

$$\begin{aligned}\sigma^2(\Lambda_2) &= \mathrm{Var}(X_1 + Q_1 - (\eta X_2 + Q_2)) \\ &= 1 + \eta^2 - 2\eta\rho + q_1 + q_2. \end{aligned} \tag{2.49}$$

Using the distributive property of lattices (equation (2.8)), this scheme can be converted to the one represented by Fig. 2.1.

The rates achieved by this scheme are given by

$$R_i = \frac{1}{2}\log\frac{1+\eta^2-2\eta\rho+q_1+q_2}{q_i} \quad \text{for } i = 1, 2 \tag{2.50}$$

This region can be optimized over all choices of η subject to an appropriate distortion constraint. It turns out that the scaling chosen in Section 2.2.2 is the optimal choice. The details are described (for the more general K user case) in Appendix A.4.

2.2.4 Outer Bounds

In this section, we present some outer bounds to the optimal rate distortion region as defined in Definition 2.1. A simple cut-set bound for this problem can be derived by lower bounding R_1 assuming that the decoder has full knowledge of X_2 and vice versa. Such a bound is given by

$$\mathcal{RD}_{CS} = \left\{(R_1, R_2, D)\colon R_1 \geq \frac{1}{2}\log^+\frac{1-\rho^2}{D}, R_2 \geq \frac{1}{2}\log^+\frac{c^2(1-\rho^2)}{D}\right\} \tag{2.51}$$

where $\log^+ x \triangleq \max\{\log x, 0\}$. Another outer bound was presented in [50] for the case when $\rho \leq c \leq 1$ which we reproduce below.

Fact 1. Suppose that $\rho \leq c \leq 1$ and let θ be defined as

$$\theta = \frac{1 - \rho c}{\sigma_Z^2}. \tag{2.52}$$

Then $0 \leq \theta \leq 1$ and

$$\mathcal{RD}_o = \left\{ (R_1, R_2, D) \colon \theta 2^{-2R_1} + (1-\theta) 2^{-2R_2} \leq \frac{D}{\sigma_Z^2} \right\} \tag{2.53}$$

is an outer bound to the optimal rate distortion region.

It is further established in [50] that the gap between the sum rates of the rate regions presented in Theorem 1 and Fact 1 is at most $-\frac{1}{2}\log\theta(1-\theta)$. In particular, this implies that when $c = 1$, the sum rate given by Theorem 1 is within one bit of the optimum sum rate for any distortion D.

2.3 Distributed source coding for the K source case

In this section, we consider the case of reconstructing a linear function of an arbitrary number of sources. In the case of two sources, the two strategies used in Theorems 1 and 2 were direct reconstruction of the function Z and estimating the function from noisy versions of the sources respectively. Henceforth, we shall refer to the coding scheme used to derive Theorem 1 as lattice binning and that used in Theorem 2 as random binning.

In the presence of more than two sources, a host of strategies which are a combination of these two strategies become available. For example, in the case of 3 sources, one possible strategy would be for all users to use the lattice binning while another strategy would be for users 1 and 2 to use lattice binning and user 3 to employ random binning. The union of the rate-distortion tuples achieved by all such schemes gives an achievable rate region of the problem.

When a combination of the two strategies are used among the K sources, the order of decoding at the decoder becomes important. The indices which are decoded earlier can be used as side information for the indices which are to be decoded later. Thus, the order of decoding becomes significant with the sources being encoded later having more side information available for their decoding. Also, this raises the question of how to adapt the coding schemes of lattice binning to the case when side information is present at the decoder. Consider an example when the decoder is interested in reconstructing a linear function of 3 sources, i.e., $Z = \sum_{i=1}^{3} c_i X_i$. Suppose encoders 1 and 2 use an identical coarse lattice and encoder 3 uses a different coarse lattice for encoding. If the source X_3 is decoded first, it can be used as side information for decoding $c_1 X_1 + c_2 X_2$. For ease of exposition and understanding in the following section, we first describe a lattice coding strategy for the distributed source coding problem involving two sources with the goal of reconstruction of their linear function at the decoder and, in addition, the decoder has access to some side information. We then use this to formally describe an achievable rate region for the problem of reconstructing $Z = \sum_{i=1}^{K} c_i X_i$.

2.3.1 Lattice coding in presence of decoder side information

In this section, we consider the problem of distributed encoding of correlated sources using lattices in the presence of side information at the decoder. As we will see, this can be used as a building block in reconstructing a linear function of multiple sources.

Assume that we have correlated Gaussian sources X_1 and X_2 and the decoder is interested in reconstructing a linear function $Z \triangleq \sum_{i=1}^{2} c_i X_i$. Suppose the decoder also has available to it side information Y that is correlated with the sources X_1, X_2. Y and X_1, X_2 are jointly Gaussian. Each source X_i is observed by an encoder which

maps its outcomes to a finite set. The indices produced by the encoders are transmitted to a joint decoder using two rate-constrained noiseless channels. The goal is to find the optimal rate-distortion region which is the set of all achievable tuples (R_1, R_2, D).

Note that, this reduces to the Wyner-Ziv problem when there is only one source X at the encoder. For this problem, it is known that the conditional rate-distortion bound is still achievable [5] despite the side information being available only at the decoder. Also, if the decoder is interested in reconstructing only one of the sources with mean square distortion, this problem reduces to the lossy jointly Gaussian "one-help-one" problem considered in [8].

In this subsection we provide an inner bound to the optimal rate-distortion region for this problem using a lattice-based "correlated" binning strategy. We use the notation $\hat{Z}_Y$ to denote the minimum mean-squared error (MMSE) estimate of Z given Y, namely $\mathbb{E}(Z \mid Y)$. The innovations random variable $Z - \hat{Z}_Y$ is denoted by $\eta_{Z|Y}$.

The lattice coding strategy in the presence of side information can be inferred by considering what the strategy would be in the presence of a central encoder that has access to all the sources X_1, X_2 and the side information Y. In that case, the central encoder would first compute $Z = \sum_{i=1}^{2} c_i X_i$ and then quantize and transmit only the innovations random variable $\eta_{Z|Y}$. This can be accomplished with subtractive dither lattice quantization using a nested lattice $\Lambda_2 \subset \Lambda_1$ of parameter

$$\sigma^2(\Lambda_1) = \frac{D\sigma_\eta^2}{\sigma_\eta^2 - D} \tag{2.54}$$

$$\sigma^2(\Lambda_2) = \frac{\sigma_\eta^4}{\sigma_\eta^2 - D} \tag{2.55}$$

where σ_η^2 is the variance of the innovations random variable $\eta_{Z|Y}$ and D is the

desired distortion in the reconstruction of Z. The rate incurred in this system is given by $\frac{1}{2}\log(\sigma_\eta^2/D)$. The decoder would use this quantized innovations with the side information to obtain a reconstruction that is within a distortion of D of Z.

The two assumptions in the setup above that deviate from our distributed coding problem are that all sources are available to a central encoder and that side information is available at the encoder. The first assumption can be gotten rid of by employing the distributive property (equation (2.8)) of lattice codes. The second assumption can be eliminated by using the linear nature of the forward test channel for the case of Gaussian quantization. This linear nature enables one to move the side information present at the encoder to the decoder thus obviating its necessity at the encoder. Thus, we can convert the above centralized coding strategy to our distributed setting to yield the following encoding scheme.

The source encoders are described by the equations

$$S_i = (Q_{\Lambda_{1i}}(c_i X_i^n + U_i)) \bmod \Lambda_2 \quad \text{for } i = 1, 2, \tag{2.56}$$

where U_is are independent random dithers uniformly distributed over the fundamental Voronoi region $\mathcal{V}_{0,1i}$ of the fine lattices Λ_{1i}s. As in Section 2.2, we require $\Lambda_2 \subset \Lambda_{1i}$, $i = 1, 2$, the fine lattices Λ_{1i} to be good source codes and the coarse lattice Λ_2 to be a good channel code. The second moments of the nested lattices are given by

$$\sigma^2(\Lambda_{11}) = q_1 \tag{2.57}$$

$$\sigma^2(\Lambda_{12}) = \frac{D\sigma_\eta^2}{\sigma_\eta^2 - D} - q_1 \tag{2.58}$$

$$\sigma^2(\Lambda_2) = \sigma_\eta^2 + \sigma^2(\Lambda_{11}) + \sigma^2(\Lambda_{12}) = \frac{\sigma_\eta^4}{\sigma_\eta^2 - D} \tag{2.59}$$

where q_1 is chosen such that $0 < q_1 < \frac{D\sigma_\eta^2}{\sigma_\eta^2 - D}$. This gives a quantization rate of

$$R_1 = \frac{1}{2}\log\frac{\sigma_\eta^4}{q_1(\sigma_\eta^2 - D)} \tag{2.60}$$

$$R_2 = \frac{1}{2}\log\frac{\sigma_\eta^4}{D\sigma_\eta^2 - q_1(\sigma_\eta^2 - D)} \tag{2.61}$$

Clearly, for a fixed choice of q_1 all rates beyond that given above can be achieved. Eliminating q_1 between the two rates now gives us an expression of the overall achievable region as

$$2^{-2R_1} + 2^{-2R_2} \leq \left(\frac{\sigma_\eta^2}{D}\right)^{-1} \tag{2.62}$$

The decoder is given by the equation

$$\hat{Z} = \left(1 - \frac{D}{\sigma_\eta^2}\right)\left(\left[\sum_{i=1}^{2}(S_i - U_i) - \hat{Z}_Y^n\right] \bmod \Lambda_2\right) + \hat{Z}_Y^n \tag{2.63}$$

The encoding operation given by equation (2.56) is similar to that used in Section 2.2.2. The decoding operation can be understood as follows. The decoder shifts the origin of the nested lattice code to the point $\hat{Z}_Y^n$, decodes the innovations random variable $\eta_{Z|Y}^n$ and computes the best estimate of Z given $\hat{Z}_Y^n$ and the decoded value of $\eta_{Z|Y}^n$. By mimicking the analysis of Section 2.2.2, we can show that the first part of the decoder operation, given by $([\sum_{i=1}^{2}(S_i - U_i) - \hat{Z}_Y^n] \bmod \Lambda_2)$ in equation (2.63) which corresponds to shifting the origin to $\hat{Z}_Y^n$ and decoding $\eta_{Z|Y}^n$, produces with high probability $\eta_{Z|Y}^n + N$ where N approaches a white Gaussian noise vector with each element having variance $\sigma^2(\Lambda_{11}) + \sigma^2(\Lambda_{12}) = \frac{D\sigma_\eta^2}{\sigma_\eta^2 - D}$. The decoder then obtains an estimate of the function Z based on $\eta_{Z|Y} + N$ and the side information Y. For an optimal choice of the nested lattices, in the limit as the dimension $n \to \infty$, the variables $\hat{Z}_Y^n, \eta_{Z|Y}^n + N$ and Z become jointly Gaussian and the optimal MMSE estimate of Z is a linear function of $\hat{Z}_Y^n$ and $\eta_{Z|Y}^n + N$. It can be checked that equation

(2.63) describes such an estimate and that this estimate indeed achieves the desired distortion D. Thus, we have an achievable rate-distortion tuple given by equation (2.62) for reconstructing a linear function in the presence of any side information. The rationale for choosing the lattice parameters and scaling constants is very similar to that given in Section 2.2.3.

2.3.2 Reconstructing a linear function of K sources

Previously, we considered the problem of reconstructing a linear function of two sources. In this section, we generalize the problem to an arbitrary number of sources. Let the sources be given by $X_1, X_2, \ldots, X_K$ which are jointly Gaussian. The encoder of X_i maps its outcome to a finite set. The output of the encoder is transmitted over a noiseless but rate-constrained channel to a joint decoder. The rate of channel i is given by R_i. The decoder wishes to reconstruct a linear function given by $Z = \sum_{i=1}^{K} c_i X_i$ with squared error fidelity criterion. The performance limit $\mathcal{RD}$ is given by the set of all rate-distortion tuples $(R_1, R_2, \ldots, R_K, D)$ that are achievable in the sense defined in Section 2.2. In this section we provide an inner bound based on "correlated" lattice-structured binning.

Note that, if the decoder is interested in reconstructing only one of the sources with mean square distortion, this problem reduces to the lossy jointly Gaussian "many-help-one" problem similar to the one studied in [28]. As indicated earlier, there are several possible coding schemes based on each user's choice of coding strategy and also the choice of order of decoding. Before, we describe these coding schemes, we introduce some relevant notation.

For any set $A \subset \{1, \ldots, K\}$, let X_A denote those sources whose indices are in A, i.e., $X_A \triangleq \{X_i : i \in A\}$. Let Z_A be defined as $\sum_{i \in A} c_i X_i$. Let Θ be a partition of $\{1, \ldots, K\}$ with $\theta = |\Theta|$. Let $\pi_\Theta : \Theta \to \{1, \ldots, \theta\}$ be a permutation. One can think

of π_Θ as ordering the elements of Θ. Each set of sources $X_A, A \in \Theta$ are decoded simultaneously at the decoder with the objective of reconstructing Z_A. The order of decoding is given by $\pi_\Theta(A)$ with the lower ranked sets of sources decoded earlier. Let $\mathcal{Q} = (q_1, \ldots, q_K) \in \mathbb{R}_+^K$ be a tuple of positive reals. Let $\mathbb{E}(\cdot)$ denote the expectation operator.

For any partition Θ and ordering π_Θ, let us define recursively a positive-valued function $\sigma^2_\Theta : \Theta \to \mathbb{R}^+$ as follows:

$$\sigma^2_\Theta(A) = \mathbb{E}\left[(Z_A - f_A(S_A))^2\right], \tag{2.64}$$

where

$$f_A(S_A) = \mathbb{E}(Z_A|S_A) \tag{2.65}$$

$$S_A = \{Z_B + Q_B : B \in \Theta, \pi_\Theta(B) < \pi_\Theta(A)\} \tag{2.66}$$

and $\{Q_A : A \in \Theta\}$ is a collection of $|\Theta|$ independent zero-mean Gaussian random variables with variances given by $q_A = \mathrm{Var}(Q_A) \triangleq \sum_{i \in A} q_i$, and this collection is independent of the sources. As will be seen later, Q_A can be thought of as approximating the sum of the subtractive dither lattice quantization noises that result from the encoding of the sources X_A. Let

$$f(\{Z_A + Q_A : A \in \Theta\}) \triangleq \mathbb{E}\left(Z|\{Z_A + Q_A : A \in \Theta\}\right). \tag{2.67}$$

Theorem 3. *For a given tuple of sources $X_1, \ldots, X_K$ and tuple of real numbers $(c_1, c_2, \ldots, c_K)$, we have $\mathcal{RD}^*_{in} \subset \mathcal{RD}$, where*

$$\mathcal{RD}_{in} = \bigcup_{\Theta, \pi_\Theta, \mathcal{Q}} \Bigg\{ (R_1, \ldots, R_K, D) : R_i \geq \frac{1}{2} \log \frac{\sigma^2_\Theta(A) + q_A}{q_i} \text{ for } i \in A$$

$$D \geq \mathbb{E}\left[\left(Z - f\left(\{Z_A + Q_A : A \in \Theta\}\right)\right)^2\right] \quad \Bigg\}, \tag{2.68}$$

and * denotes convex closure.

Proof: We give a description of a lattice-based coding scheme that achieves the inner bound. Fix Θ, π_Θ and $\mathcal{Q}$. For each $A \in \Theta$, construct a family of good nested lattices Λ_{1i}^A and Λ_2^A such that $\Lambda_2^A \subset \Lambda_{1i}^A$ for $i \in A$. Existence of such good nested lattices has been shown in Appendix A.2. The second moment of the fine lattice Λ_{1i}^A is chosen to be q_i. The second moment of the coarse lattice is chosen based on the amount of side information available to the decoder at the time of decoding the set of sources X_A which in turn depends on $\pi_\Theta(A)$. The function σ_Θ^2 governs this choice. More precisely, for $i \in A$ and $A \in \Theta$, the second moments of the lattices are given by

$$\sigma^2(\Lambda_{1i}^A) = q_i \tag{2.69}$$

$$\sigma^2(\Lambda_2^A) = \sigma_\Theta^2(A) + q_A \tag{2.70}$$

Here, $\sigma_\Theta^2(A)$ plays a role analogous to σ_η^2 in equations (2.57)-(2.59) and approximates the variance of the innovations process when estimating Z_A given the side information S_A.

Encoder: For each $A \in \Theta$, the source X_i, $i \in A$ is encoded using the nested lattice $\Lambda_2^A \subset \Lambda_{1i}^A$. The encoders can be described by the equations

$$T_i = (Q_{\Lambda_{1i}^A}(c_i X_i^n + U_i)) \bmod \Lambda_2^A \quad \text{for } i \in A \tag{2.71}$$

where U_i are independent random dithers uniformly distributed over the fundamental Voronoi region $\mathcal{V}_{0,1i}^A$ of the fine lattice Λ_{1i}^A. This would give an encoding rate of

$$R_i = \frac{1}{2} \log \frac{\sigma_\Theta^2(A) + q_A}{q_i} \quad \text{for } i \in A \tag{2.72}$$

Decoder: For $A \in \Theta$, in order to decode Z_A, the decoder has access to some side information and its operation can be recursively described similar to equations (2.30)

and (2.63) as

$$\hat{Z}_A = \left(\left[\sum_{i \in A} (T_i - U_i) - f_A^n(\hat{S}_A) \right] \bmod \Lambda_2^A \right) + f_A^n(\hat{S}_A) \tag{2.73}$$

where

$$\hat{S}_A = \{ \hat{Z}_B : B \in \Theta, \pi_\Theta(B) < \pi_\Theta(A) \}. \tag{2.74}$$

After decoding $\hat{Z}_A$ for all $A \in \Theta$, the decoder obtains the reconstruction as a linear function of $\{\hat{Z}_A : A \in \Theta\}$ as

$$\hat{Z} = f^n(\{\hat{Z}_A : A \in \Theta\}). \tag{2.75}$$

We now show that the above system achieves the inner bound given in Theorem 3. From equation (2.72), it is clear that this scheme achieves the rate tuple claimed in Theorem 3. It remains to prove that the claimed distortion is achieved. The crucial observation is that while S_A in equation (2.66) denotes the side information available to decode Z_A in test channels, $\hat{S}_A$ in equation (2.74) denotes the side information available to decode $\hat{Z}_A$ in the actual coding system. If we were to assume $\hat{S}_A$ to be Gaussian, then by definition of the functions $f_A(\cdot)$ (equation (2.65)) and $f(\cdot)$ (equation (2.67)), it is easy to see that the distortion given in Theorem 3 is achieved. However such an assumption is not true for $\hat{S}_A$ for any finite lattice dimension n.

Fortunately, loosely speaking, we can show that even though the assumption of Gaussianity of $\hat{Z}_A$ is not strictly true, it becomes increasingly valid as the lattice dimension $n \to \infty$. By analysis similar to that in Section 2.2.2, we can show that the subtractive dither quantization noises tend to a white Gaussian of the same variance (in the K-L divergence sense). This implies that as the lattice dimension $n \to \infty$, for an optimal choice of nested lattices, $\hat{Z}_A$ tends to $Z_A^n + Q_A^n$ and hence $\hat{S}_A$ tends to S_A^n (in the K-L divergence sense). By virtue of the "goodness" of the nested lattices,

this then implies that the probability of incorrect decoding goes to 0 exponentially in the lattice dimension. Thus the reconstruction error $(Z^n - \hat{Z})$ tends in probability (and hence in normalized second moment) to N where N approaches a Gaussian random vector with each component having variance D. Thus, the proposed lattice scheme indeed achieves the claimed rate-distortion tuples and Theorem 3 is proved.

To show this formally using induction, we need some more notation. For each $A \in \Theta$ and for each $i \in A$, let

$$e_i = Q_{\Lambda_{1i}^A}(c_i X_i^n + U_i) - c_i X_i^n - U_i, \tag{2.76}$$

and

$$e_A \triangleq \sum_{i \in A} e_i. \tag{2.77}$$

For each $A \in \Theta$, let the linear function $f_A(\cdot)$ be given by

$$f_A(S_A) = \sum_{B:\pi_\Theta(B)<\pi_\Theta(A)} \alpha_A(B)(Z_B + Q_B). \tag{2.78}$$

By noting that e_i are independent for $i \in \{1, 2, \ldots, K\}$, we note that for all $A \in \Theta$,

$$\frac{1}{n}\mathbb{E}\|e_A\|^2 = q_A. \tag{2.79}$$

Let $E \in \Theta$ be such that $\pi_\Theta(E) = 1$. Thus $\hat{S}_E = \phi$. Hence using the distributive property, and noting the normalized second moments of e_i for $i \in E$, we have with high probability (i.e., under correct decoding)

$$\hat{Z}_E = Z_E^n + e_E. \tag{2.80}$$

For any $1 \leq j < K$, we assume correct decoding with high probability at the jth stage and show correct decoding with high probability at the $(j+1)$th stage. Let

$C \in \Theta$ be such that $\pi_\Theta(C) = j+1$. Under the above assumption, we have, with high probability, for all $B \in \Theta$ with $\pi_\Theta \leq j$

$$\hat{Z}_B = Z_B^n + e_B. \tag{2.81}$$

Using this we have

$$\hat{Z}_C = \left(Z_C^n + e_C - \sum_{B:\pi_\Theta(B)\leq j} \alpha_C(B)\hat{Z}_B \right) \bmod \Lambda_2^C + \sum_{B:\pi_\Theta(B)\leq j} \alpha_C(B)\hat{Z}_B \tag{2.82}$$

$$\stackrel{c.d}{=} \left(Z_C^n + e_C - \sum_{B:\pi_\Theta(B)\leq j} \alpha_C(B)\hat{Z}_B \right) + \sum_{B:\pi_\Theta(B)\leq j} \alpha_C(B)\hat{Z}_B \tag{2.83}$$

$$= Z_C^n + e_C, \tag{2.84}$$

where the second equality holds with high probability (correct decoding) because of the following reasons. (a) The normalized second moment of the term inside the mod operation satisfies the following equalities:

$$\frac{1}{n}\mathbb{E}\left\| Z_C^n + e_C - \sum_{B:\pi_\Theta(B)\leq j} \alpha_C(B)\hat{Z}_B \right\|^2 = \tag{2.85}$$

$$= \frac{1}{n}\mathbb{E}\left\| Z_C^n - \sum_{B:\pi_\Theta(B)\leq j} \alpha_C(B) Z_B^n \right\|^2 + q_C + \sum_{B:\pi_\Theta(B)\leq j} \alpha_C^2(B) q_B \tag{2.86}$$

$$= q_C + \mathbb{E}\left(Z_C - \sum_{B:\pi_\Theta(B)\leq j} \alpha_C(B)(Z_B + Q_B) \right)^2 \tag{2.87}$$

$$= \sigma_\Theta^2(C) + q_C \tag{2.88}$$

$$= \sigma^2(\Lambda_2^C). \tag{2.89}$$

(b) Using the arguments of Section 2.2.2 (see Appendix A.3),

$$\lim_{n\to\infty} h\left(Z_C^n + e_C - \sum_{B:\pi_\Theta(B)\leq j} \alpha_C(B)\hat{Z}_B \right) = \frac{n}{2}\log 2\pi e\sigma^2(\Lambda_2^C). \tag{2.90}$$

where $h(\cdot)$ denotes differential entropy. Hence we have for all $A \in \Theta$, with high probability,

$$\hat{Z}_A = Z_A^n + e_A. \tag{2.91}$$

Now regarding the final estimation, an argument similar to the above can be given that shows that a distortion given in the theorem is achieved asymptotically. The rationale for the specific choice of scaling constants is explained in detail in Appendix A.4.

Remark: An important point worth noting before proceeding further is that the nesting relations we need the lattices to satisfy is $\Lambda_2^A \subset \Lambda_{1i}^A$ for $i \in A$. But, for $A, B \in \Theta$, we don't need the lattice families $(\Lambda_{1i}^A, \Lambda_2^A)$ and $(\Lambda_{1j}^B, \Lambda_2^B)$ to be related in any way for $A \neq B$. Also, just as in the two user case, if we are interested only in minimizing the sum rate of this encoding scheme, then for all encoders in a given set $A \in \Theta$, the second moment of their respective fine lattices are equal. This means that all encoders in a given set $A \in \Theta$ can use the same nested lattice $\Lambda_2^A \subset \Lambda_1^A$ for encoding.

2.3.3 An illustration of Theorem 3

For clarity, an illustration of the coding scheme of Theorem 3 for the case of 6 users and specific choices of Θ and π_Θ is described below. Let us choose $\Theta = \{\{1,2,3\},\{4,5\},\{6\}\}$. Let π_Θ be the identity permutation so that $\pi_\Theta(\{1,2,3\}) = 1, \pi_\Theta(\{4,5\}) = 2, \pi_\Theta(\{6\}) = 3$. This means that the decoder decodes $Z_{\{1,2,3\}} = \sum_{i=1}^3 c_i X_i$ first which is then used as side information for decoding $Z_{\{4,5\}}$ and so on. Let us also fix $\mathcal{Q} = \{q_1, \ldots, q_6\}$ where q_i are all positive. We use A, B, C to denote the sets $\{1,2,3\}, \{4,5\}$ and $\{6\}$ respectively.

The fine lattice of the encoder of source X_i has second moment q_i as given in equation (2.69). Encoders for the sources X_1, X_2, X_3 use nested lattices where the second moment of the coarse lattices are given by equation (2.70). The decoder decodes $\hat{Z}_A$ according to equation (2.73). To decode $\hat{Z}_A$, the decoder does not have access to any side information. Encoders for X_4, X_5 use nested lattices whose parameters depend on the function $\sigma^2_\Theta(B)$ which in turn is determined by the fact that $\hat{Z}_A$ has been decoded earlier. The decoder then decodes $\hat{Z}_B$ from T_4, T_5 and the functional value $f^n_B(\cdot)$ of the side information $\hat{S}_B = \hat{Z}_A$. Similarly, to decode $\hat{Z}_C$, the decoder has side information $\hat{S}_C = \{\hat{Z}_A, \hat{Z}_B\}$ along with the index T_6. After having decoded $\hat{Z}_A, \hat{Z}_B, \hat{Z}_C$, the decoder uses the function $f^n(\cdot)$ of equation (2.67) to estimate Z. This is illustrated in Fig. 2.3. Notice the correspondence between this coding strategy and the schematic of the general distributed source coding problem as shown in Figure 1.1.

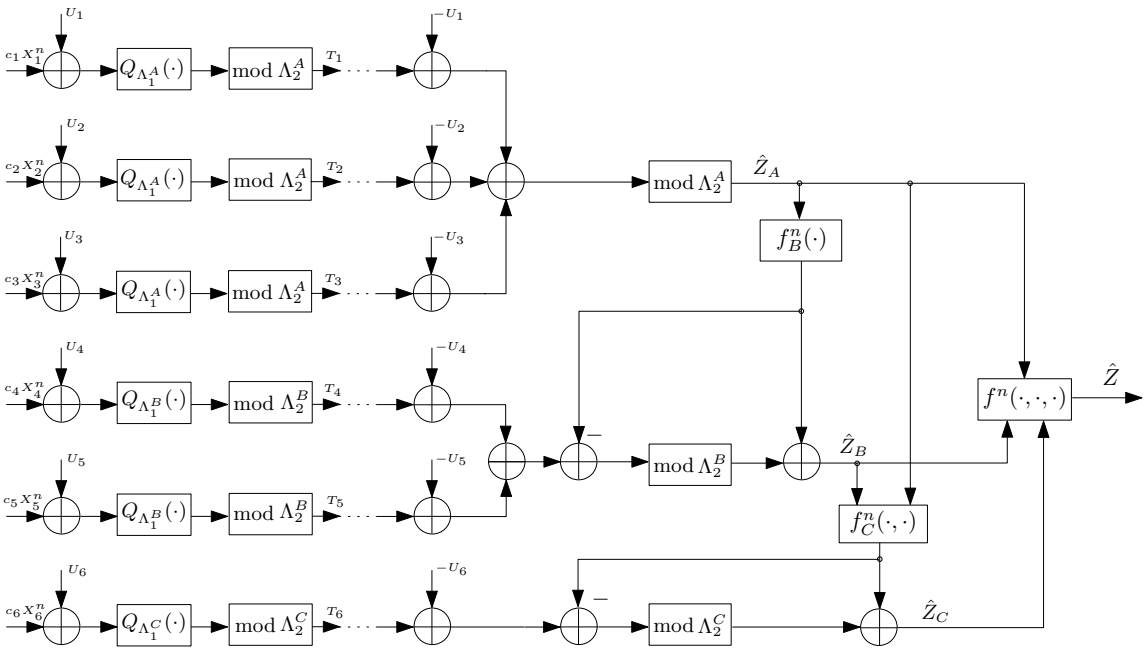

Figure 2.3: Illustration of the coding scheme of Theorem 3

2.3.4 A Few Special Cases

In this section, we consider a few special cases of the general coding problem treated above. In particular, we examine the rate distortion region derived above for specific choices of the partition Θ. First, we demonstrate that we can recover the two user rate region of Theorems 1 and 2 from the more general K-user rate region described above. Then, we illustrate a scheme for the case where the decoder estimates the function directly, i.e., $\Theta = \{\{1, 2, \ldots, K\}\}$.

Berger Tung coding for the two user case

In this section, we rederive the result of Theorem 2 using the more general framework of Theorem 3. Let the function to be reconstructed be $Z = X_1 - cX_2$ as in Section 2.2. Individual reconstruction of the sources corresponds to the partition $\Theta = \{\{1\}, \{2\}\}$. There are two possible choices of π_Θ corresponding to which source is decoded first. Let us choose π_Θ to be the identity permutation. Thus $Z_{\{1\}} = X_1$ is decoded first and used as side information to decode $Z_{\{2\}} = -cX_2$.

Let $\mathcal{Q} = (q_1, q_2)$ where q_i are positive for $i = 1, 2$. For ease of notation, we drop the set notation in the subscripts below. In what follows, S_1 is taken to mean $S_{\{1\}}$ and so on. Equations (2.64) to (2.66) simplify in this case to

$$S_1 = \phi \tag{2.92}$$

$$f_1(S_1) = \mathbb{E}(Z_1) = 0 \tag{2.93}$$

$$\sigma^2_\Theta(\{1\}) = \mathbb{E}(Z_1^2) = 1 \tag{2.94}$$

$$S_2 = \{X_1 + Q_1\} \tag{2.95}$$

$$f_2(S_2) = \mathbb{E}(Z_2 \mid S_2) = \mathbb{E}(-cX_2 \mid X_1 + Q_1) = \frac{-\rho c}{1+q_1} S_2 \tag{2.96}$$

$$\sigma^2_{\Theta}(\{2\}) = \mathbb{E}\left(Z_2 + \frac{\rho c}{1+q_1} S_2\right)^2 = c^2 + q_2 - \frac{\rho^2 c^2}{1+q_1}. \tag{2.97}$$

Since the random variables $Z, Z_1 + Q_1, Z_2 + Q_2$ are jointly Gaussian, the optimal MMSE estimator of Z given Z_1+Q_1 and Z_2+Q_2 is a linear function of Z_1+Q_1, Z_2+Q_2 and is given by $f(Z_1 + Q_1, Z_2 + Q_2) = a(Z_1 + Q_1) + b(Z_2 + Q_2)$ where the constants a, b are given by

$$\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} \frac{\alpha c^2 + q_2(1-\rho c)}{(1+q_1)(c^2+q_2)-\rho^2 c^2} & \frac{c(\alpha c + q_1(c-\rho))}{(1+q_1)(c^2+q_2)-\rho^2 c^2} \end{bmatrix} \tag{2.98}$$

where $\alpha \triangleq 1 - \rho^2$.

As stated in Theorem 3, q_i have to satisfy the distortion constraint of equation (2.68) which in this case simplifies to

$$D \geq \frac{q_1 c^2 \alpha + q_2 c^2 \alpha + q_1 q_2 \sigma_Z^2}{(1+q_1)(c^2+q_2) - \rho^2 c^2} \tag{2.99}$$

The parameters of the nested lattices are given by equations (2.69) and (2.70) to be

$$\sigma^2(\Lambda_1^{\{1\}}) = q_1 \tag{2.100}$$

$$\sigma^2(\Lambda_2^{\{1\}}) = 1 + q_1 \tag{2.101}$$

$$\sigma^2(\Lambda_1^{\{2\}}) = q_2 \tag{2.102}$$

$$\sigma^2(\Lambda_2^{\{2\}}) = c^2 + q_2 - \frac{\rho^2 c^2}{1+q_1}. \tag{2.103}$$

This gives the following rates.

$$R_1 = \frac{1}{2} \log \frac{1+q_1}{q_1} \tag{2.104}$$

$$R_2 = \frac{1}{2} \log \frac{(c^2+q_2)(1+q_1) - \rho^2 c^2}{q_2(1+q_1)} \tag{2.105}$$

where $\mathcal{Q} = (q_1, q_2)$ is subject to the distortion constraint of equation (2.99). It can be checked that these equations parameterize one of the corner points of the rate region of Theorem 2. Reversing the roles of the two sources (equivalently, choosing $\pi_\Theta(\{1\}) = 2, \pi_\Theta(\{2\}) = 1$), we can achieve the other end point of the rate region. Time sharing between these two points achieves the entire rate region of Theorem 2.

Note that the inner bound of Theorem 2 is derived using the Berger-Tung inner bound [6, 7] which employs random quantization followed by random binning. Here, we have rederived this result using lattice quantization followed by lattice-structured binning.

Lattice coding for the K user case

In this section, we derive an achievable rate region for the K user case when all the users encode in such a way that the decoder estimates the function directly without reconstructing any intermediate variables. This corresponds to the case where $\Theta = \{\{1, \ldots, K\}\}$. π_Θ is trivial in this case. Let $\mathcal{Q} = \{q_1, \ldots, q_K\} \in \mathbb{R}_+^K$. Let A denote the set $\{1, \ldots, K\}$. Then $q_A = \sum_{i=1}^K q_i$

Equations (2.64) to (2.66) simplify in this case to

$$S_A = \phi \tag{2.106}$$

$$f_A(S_A) = \mathbb{E}(Z) = 0 \tag{2.107}$$

$$\sigma_\Theta^2(A) = \mathbb{E}(Z^2) = \sigma_Z^2. \tag{2.108}$$

The function $f(\cdot)$ of equation (2.67) is given by

$$\begin{aligned} f(Z+Q) &= \mathbb{E}(Z \mid Z+Q) \\ &= \frac{\sigma_Z^2}{\sigma_Z^2 + q_A}(Z+Q) \end{aligned} \tag{2.109}$$

and thus distortion constraint of equation (2.68) fixes the value of q_A to be $\frac{\sigma_Z^2 D}{\sigma_Z^2 - D}$.

The encoders use the nested lattices $(\Lambda_{1i}, \Lambda_2), i = 1, \ldots, K$ for encoding. The parameters of the nested lattices are given by

$$\sigma^2(\Lambda_{1i}) = q_i \tag{2.110}$$

$$\sigma^2(\Lambda_2) = \sigma_Z^2 + q_A = \frac{\sigma_Z^4}{\sigma_Z^2 - D} \tag{2.111}$$

This gives an encoding rate of

$$R_i = \frac{1}{2} \log \frac{\sigma_Z^4}{q_i(\sigma_Z^2 - D)} \tag{2.112}$$

This corresponds to the rate region

$$\sum_{i=1}^{K} 2^{-2R_i} \leq \left(\frac{\sigma_Z^2}{D} \right)^{-1} \tag{2.113}$$

For $K = 2$, this recovers the rate region of Theorem 1.

2.3.5 Comparison of the Sum Rates for Low Distortions

In this section, we compare the Berger-Tung based coding scheme and the lattice based coding scheme for the general K-user case. Specifically, we compare the sum rates of the following encoding schemes in the low distortion regime - (a) all encoders use the same coarse lattice and encode in such a way that the decoder reconstructs the function directly ($\Theta = \{\{1, \ldots, K\}\}$) and (b) the encoders use different coarse lattices and the decoder estimates the function from lossy reconstruction of the sources ($\Theta = \{\{1\}, \ldots, \{K\}\}$). While the minimum sum rate required by the lattice based coding scheme to achieve a distortion D is easily derived from equation (2.113) for any D, a similar analysis is analytically intractable for the Berger-Tung based coding scheme except for low values of distortion D.

Let the decoder be interested in reconstructing the function $Z = \sum_{i=1}^{K} c_i X_i$ to within a mean square distortion of D. Let the covariance matrix of the jointly

Gaussian random variables $X_1, \ldots, X_K$ be the $K \times K$ matrix Σ. Let the column vector be defined as $\bar{c} \triangleq [c_1 \ldots c_K]^T$. It is easy to see that the minimum sum rate $\sum_{i=1}^{K} R_i$ is achieved in equation (2.112) when $q_1 = \cdots = q_K = \frac{\sigma_Z^2 D}{K(\sigma_Z^2 - D)}$ and the minimum sum rate is given by

$$\begin{aligned} R_{latsum} &\triangleq \frac{K}{2} \log \frac{K\sigma_Z^2}{D} \\ &= \frac{K}{2} \log \frac{K}{D} \bar{c}^T \Sigma \, \bar{c} \end{aligned} \tag{2.114}$$

An approximate expression for the sum rate R_{BTsum} of the Berger-Tung based coding scheme can be derived in the low distortion regime as shown below. An achievable sum rate-distortion region for this problem can be derived using the Berger-Tung based coding scheme with the auxiliary random variables being Gaussian.

Lemma 2.4. *Define the region RD_{BTsum} as*

(2.115)

$$RD_{BTsum} \triangleq \bigcup_{q_1,\ldots,q_K \in \mathbb{R}_+^K} \left\{ R_{BTsum} \geq \frac{1}{2} \log \frac{|\Sigma_U|}{\prod_{i=1}^{K} q_i},\ D \geq \sigma_Z^2 - (\Sigma\bar{c})^T \Sigma_U^{-1} (\Sigma\bar{c}) \right\}$$

where $\Sigma_U \triangleq \Sigma + \Lambda_Q$ and Λ_Q is a $K \times K$ diagonal matrix with diagonal entries $q_1, \ldots, q_K$. $|\Sigma_U|$ is the determinant of the matrix Σ_U. Then, there exists an achievable rate-distortion tuple $(R_1, \ldots, R_K, D)$ such that $(\sum_{i=1}^{K} R_i, D) \in RD_{BTsum}^*$ where * denotes convex closure.

Proof: Follows directly from the application of Berger-Tung inner bound for the K user case with the auxiliary random variables involved being Gaussian.

If the function Z is not directly related to a source X_i, i.e., if $c_i = 0$ for some $1 \leq i \leq K$, then the optimal strategy that minimizes the sum rate for a given distortion would involve not transmitting that source at all. Thus, we can assume without loss of generality that $c_i \neq 0$ for all $i = 1, \ldots, K$. This assumption is made

throughout this section. When the distortion $D \to 0$, it follows that $\Sigma_U \to \Sigma$, i.e., $q_i \to 0$ for $i = 1, \ldots, K$. Under these conditions, the expressions for sum rate and distortion in Lemma 2.4 can be considerably simplified by expanding them in a Taylor series and retaining only the first order terms. We detail this approximation below.

We define some quantities that we use in the derivations below. Let $\Lambda_Q^{(i)}$ be the diagonal matrix with the jth diagonal entry equal to q_j for $j \neq i$ and the ith diagonal entry set to 0. Let $\Sigma_U^{(i)} \triangleq \Sigma + \Lambda_Q^{(i)}$. Let e_i denote the K-length column vector with a 1 in the ith position and 0 elsewhere.

In the limit as $D \to 0$, we can write

$$D = D_{\{q_1=0,\ldots,q_K=0\}} + \sum_{i=1}^{K} q_i \left(\frac{\partial D}{\partial q_i}\right)_{\{q_1=0,\ldots,q_K=0\}} + O(q_i^2) \tag{2.116}$$

$$= \sum_{i=1}^{K} q_i \left(\frac{\partial D}{\partial q_i}\right)_{\{q_1=0,\ldots,q_K=0\}} + O(q_i^2) \tag{2.117}$$

The partial derivatives can be evaluated as follows. From the Sherman-Morrison formula, it follows that for an invertible matrix Σ and the product uv^T of two column vectors u, v, we have

$$(\Sigma + uv^T)^{-1} = \Sigma^{-1} - \frac{\Sigma^{-1}uv^T\Sigma^{-1}}{1 + v^T\Sigma^{-1}u} \tag{2.118}$$

$$|\Sigma + uv^T| = \left(1 + v^T\Sigma^{-1}u\right)|\Sigma| \tag{2.119}$$

In order to evaluate the partial derivative of the distortion D with respect to q_i,

set $u = q_i e_i$ and $v = e_i$ and $q_j = 0$ for $j \neq i$. Then,

$$D = \sigma_Z^2 - \bar{c}^T \Sigma^T (\Sigma_U^{(i)} + u_i v_i^T)^{-1} \Sigma \bar{c} \tag{2.120}$$

$$= \sigma_Z^2 - \bar{c}^T \Sigma^T \left(\Sigma_U^{(i)^{-1}} - \frac{\Sigma_U^{(i)^{-1}} q_i e_i e_i^T \Sigma_U^{(i)^{-1}}}{1 + q_i e_i^T \Sigma_U^{(i)^{-1}} e_i} \right) \Sigma \bar{c} \tag{2.121}$$

$$= \sigma_Z^2 - \bar{c}^T \Sigma^T \Sigma_U^{(i)^{-1}} \Sigma \bar{c} + q_i \frac{\bar{c}^T \Sigma^T \Sigma_U^{(i)^{-1}} e_i e_i^T \Sigma_U^{(i)^{-1}} \Sigma \bar{c}}{1 + q_i e_i^T \Sigma_U^{(i)^{-1}} e_i} \tag{2.122}$$

$$= K + \frac{q_i \alpha}{1 + q_i \beta} \tag{2.123}$$

where K, α, β are independent of q_i. Taking the partial derivative with respect to q_i and setting $q_i = 0$ for $1 \leq i \leq K$, we get

$$\left(\frac{\partial D}{\partial q_i} \right)_{\{q_1 = 0, \ldots, q_K = 0\}} = (\alpha)_{\{q_1 = 0, \ldots, q_K = 0\}} \tag{2.124}$$

$$= c_i^2 \tag{2.125}$$

Therefore, in the low distortion regime, we have

$$D = \sum_{i=1}^{K} c_i^2 q_i. \tag{2.126}$$

The sum rate R_{BTsum} can be written as

$$R_{BTsum} = \frac{1}{2} \frac{|\Sigma|}{\prod_{i=1}^{K} q_i} + \frac{1}{2} \log \frac{|\Sigma_U|}{|\Sigma|} \tag{2.127}$$

A good approximation can be obtained for the sum rate in the low distortion regime by expanding $T \triangleq \frac{1}{2} \log \frac{|\Sigma_U|}{|\Sigma|}$ using Taylor series and retaining only the first order terms. We can write T as

$$T = \frac{1}{2} \log \frac{|\Sigma_U|}{|\Sigma|} = \frac{1}{2} \log \frac{|\Sigma_U^{(i)}|}{|\Sigma|} + \frac{1}{2} \log(1 + q_i e_i^T \Sigma_U^{(i)^{-1}} e_i) \tag{2.128}$$

Taking the partial derivative of T with respect to q_i and setting $q_i = 0$ for $1 \leq i \leq K$, we get

$$\left(\frac{\partial T}{\partial q_i} \right)_{\{q_1 = 0, \ldots, q_K = 0\}} = \frac{e_i^T \Sigma^{-1} e_i}{2} = \frac{\Sigma_{ii}^{-1}}{2} \tag{2.129}$$

where Σ_{ii}^{-1} is the ith diagonal element of Σ^{-1}. Therefore, the sum rate R_{BTsum} can be approximated as

$$R_{BTsum} = \frac{1}{2}\log\frac{|\Sigma|}{\prod_{i=1}^{K} q_i} + \frac{1}{2}\sum_{i=1}^{K} q_i \Sigma_{ii}^{-1} \tag{2.130}$$

Since, in the low distortion regime, $q_i \to 0$, the sum rate R_{BTsum} is dominated by the first term and we get

$$R_{BTsum} = \frac{1}{2}\log\frac{|\Sigma|}{\prod_{i=1}^{K} q_i}. \tag{2.131}$$

From equations (2.126) and (2.131), it is easy to see that for a given distortion D, the sum rate is minimized when $q_i = \frac{D}{Kc_i^2}$ for $i = 1, \ldots, K$ and the minimum sum rate is

$$R_{minBTsum} = \frac{K}{2}\log\frac{K}{D}\left(|\Sigma|\prod_{i=1}^{K} c_i^2\right)^{1/K} \tag{2.132}$$

Comparing equation (2.132) to equation (2.114) we find that in the low distortion regime,

$$R_{minBTsum} - R_{latsum} = \frac{K}{2}\log\frac{K}{D}\left(|\Sigma|\prod_{i=1}^{K} c_i^2\right)^{1/K} - \frac{K}{2}\log\frac{K}{D}\,\bar{c}^T\Sigma\bar{c} \tag{2.133}$$

$$= \frac{K}{2}\log|\Sigma|^{1/K}\frac{\left(\prod_{i=1}^{K} c_i^2\right)^{1/K}}{\bar{c}^T\Sigma\bar{c}} \tag{2.134}$$

For the symmetric two user case considered in Section 2.2, the difference in minimum sum rates as given by equation (2.134) can be evaluated exactly. When the function to be reconstructed is $Z = X_1 - cX_2$, i.e., $\bar{c} = [1 - c]^T$, it evaluates to

$$R_{minBTsum} - R_{latsum} = \log\frac{|c|\sqrt{1-\rho^2}}{1 + c^2 - 2\rho c} \tag{2.135}$$

It is easy to verify that this difference in sum rate is always negative for any $\rho > 0$ if $c < 0$. Thus, in this regime, the Berger-Tung based coding scheme always

outperforms the lattice based coding scheme. Indeed, it has been shown in [37] that the Berger-Tung based coding scheme is optimal in this regime.

Also, the difference in sum rates is maximum when $c = 1$, i.e., $Z = X_1 - X_2$ and in this case, the difference in minimum sum rate is

$$R_{minBTsum} - R_{latsum} = \log \frac{1}{2}\sqrt{\frac{1+\rho}{1-\rho}} \tag{2.136}$$

which tends to ∞ as $\rho \to 1$. Thus, the lattice based coding scheme gives arbitrarily large rate gains over the Berger-Tung based coding scheme in this regime.

For the general K-user case, given a $K \times K$ covariance matrix Σ, a natural question to ask is which choice of the vector $\bar{c}$ does the lattice coding scheme offer maximum rate gains for? Observe that the difference in minimum sum rates given by equation (2.134) can be bounded as follows.

$$R_{minBTsum} - R_{latsum} = \frac{K}{2} \log |\Sigma|^{1/K} \frac{\left(\prod_{i=1}^{K} c_i^2\right)^{1/K}}{c^T \Sigma c} \tag{2.137}$$

$$\leq \frac{K}{2} \log \frac{|\Sigma|^{1/K}}{K\lambda_{min}} \frac{\left(\prod_{i=1}^{K} c_i^2\right)^{1/K}}{c^T c/K} \tag{2.138}$$

$$\leq \frac{K}{2} \log \frac{|\Sigma|^{1/K}}{K\lambda_{min}} \tag{2.139}$$

$$= \frac{K}{2} \log \frac{1}{K} \left(\prod_{i=1}^{K} \frac{\lambda_i}{\lambda_{min}}\right)^{1/K} \tag{2.140}$$

where the first inequality follows from the well known inequality that $c^T \Sigma c \geq \lambda_{min} c^T c$ and the second from the arithmetic-geometric mean inequality. Here λ_{min} is the smallest eigenvalue of the covariance matrix Σ. Equality is achieved if $\bar{c} = \nu_{min}$, the eigenvector corresponding to the smallest eigenvalue λ_{min} and all components of ν_{min} have equal magnitude. This is the case for the two user symmetric case considered in Section 2.2 where the covariance matrix Σ has eigenvalues $(1+\rho)$ and $(1-\rho)$ with the corresponding eigenvectors $[1,\ 1]^T$ and $[1,\ -1]^T$ respectively. Thus, in this case, the

lattice coding scheme offers maximum rate gains over the Berger-Tung based scheme when the function to be reconstructed is $Z = X_1 - X_2$ whereas the Berger-Tung based scheme outperforms the lattice coding scheme when $Z = X_1 + X_2$.

For the general case of an arbitrary $K \times K$ covariance matrix Σ, the rate gains offered by the lattice based coding scheme over the Berger-Tung based scheme increases as the function vector c becomes more closely aligned with the eigenvector corresponding to the smallest eigenvalue of A. Equation (2.140) also offers some necessary conditions for the lattice coding scheme to outperform the Berger-Tung based scheme. One such condition is that

$$\left(\prod_{i=1}^{K} \frac{\lambda_i}{\lambda_{min}}\right)^{1/K} \geq K \tag{2.141}$$

For the symmetric two user case of Section 2.2, this implies that a necessary condition for lattice coding scheme to outperform the Berger-Tung based coding scheme is $\rho \geq 0.6$. We shall see that this is indeed the case in Section 2.4.

2.4 Comparison of the Rate Regions

In this section, we compare the rate regions of the lattice based coding scheme given in Theorem 1 and the Berger-Tung based coding scheme given in Theorem 2 for the case of two users. The function under consideration is $Z = X_1 - cX_2$. We would like to emphasize that we have assumed that the sources have unit variance and that $\rho > 0$. To demonstrate the performance of the lattice binning scheme, we choose the sum rate of the two encoders as the performance metric.

Fig. 2.4 shows the sum rate of the lattice based scheme for different values of c and distortion D. In Fig. 2.5 and Fig. 2.6, we compare the sum-rates of the two schemes for varying values of ρ while fixing $c = 1$. From these figures, it can be seen that as $\rho \to 1$, the rate gain offered by the lattice based coding scheme increases especially

in the low distortion regime. This agrees with the analysis of Section 2.3.5. These figures also demonstrate that the rate region of Theorem 1 contains points outside that of the rate region of Theorem 2. The opposite is also true since for $D = \sigma_Z^2$, the region in Theorem 2 contains the rate point $(0, 0)$ while the one in Theorem 1 does not.

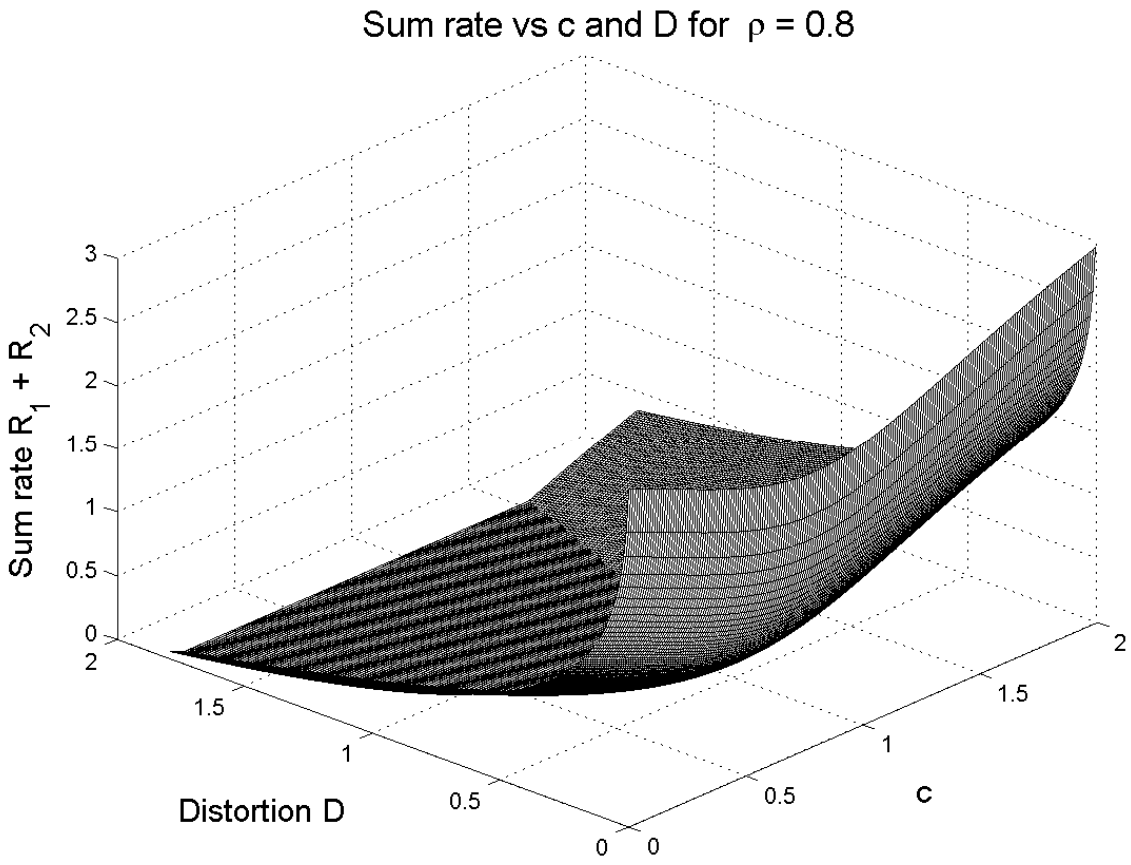

Figure 2.4: Lattice based scheme's sum-rate vs c and distortion D for $\rho = 0.8$

We observe that the lattice based scheme performs better than the Berger-Tung based scheme for small distortions provided ρ is sufficiently high and c lies in a certain interval. Fig. 2.7 and 2.8 are contour plots that illustrate this phenomenon in detail. The contour labeled R encloses that region in which the pair (ρ, c) should lie for the lattice binning scheme to achieve a sum rate that is at least R units less

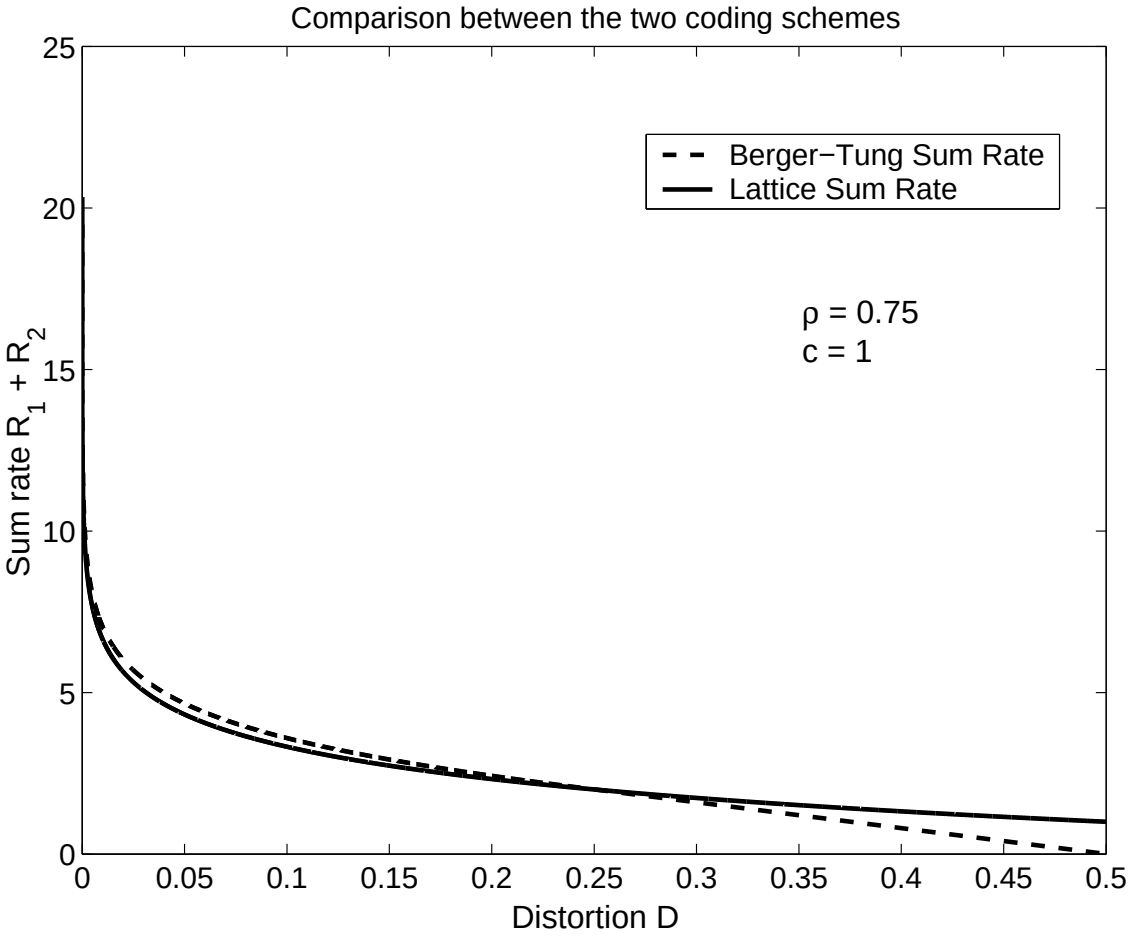

Figure 2.5: Comparison of sum rates when ρ is small and $c = 1$

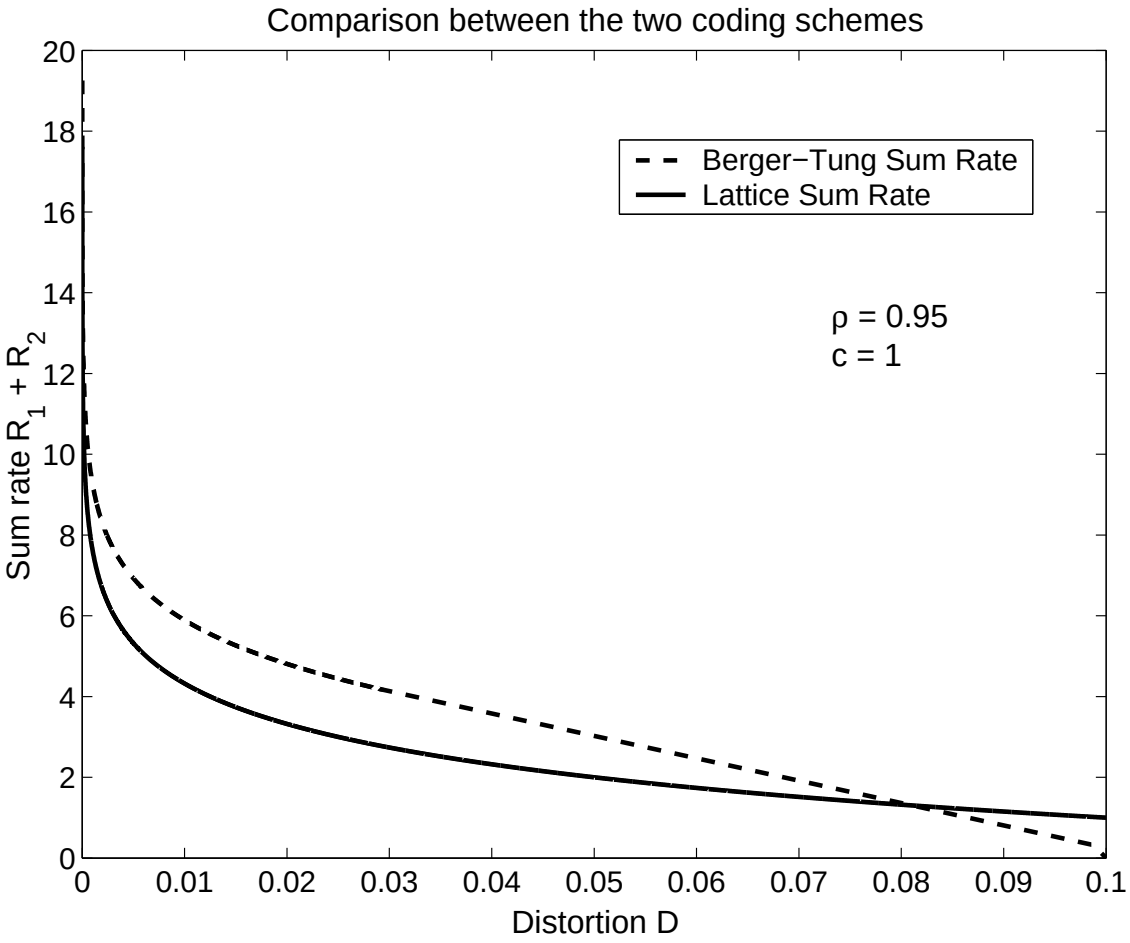

Figure 2.6: Comparison of sum rates when ρ is large and $c = 1$

than the sum rate of the Berger-Tung scheme at the distortion D. Observe that we get improvements in the limit as $D \to 0$ only when $\rho > 0.6$ as predicted by equation (2.141). In Fig. 2.7, the contour labeled R encloses those values of (ρ, c) for which the RHS of equation (2.135) exceeds R and can be analytically calculated. Also, the region where (ρ, c) can lie shrinks as the target distortion D increases suggesting that the rate gains offered by the lattice coding scheme decreases as the distortion D increases.

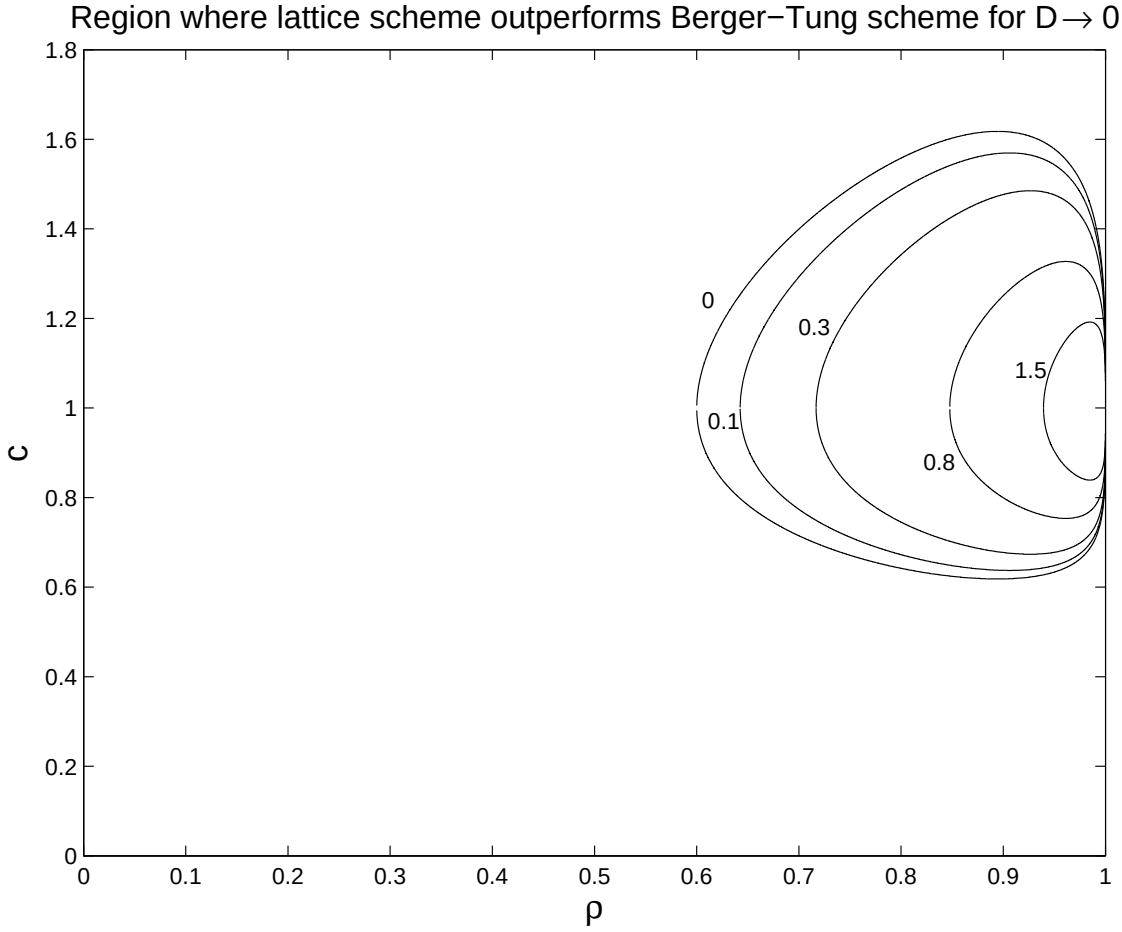

Figure 2.7: Range of (ρ, c) where the lattice scheme performs better than the Berger Tung scheme for $D \to 0$

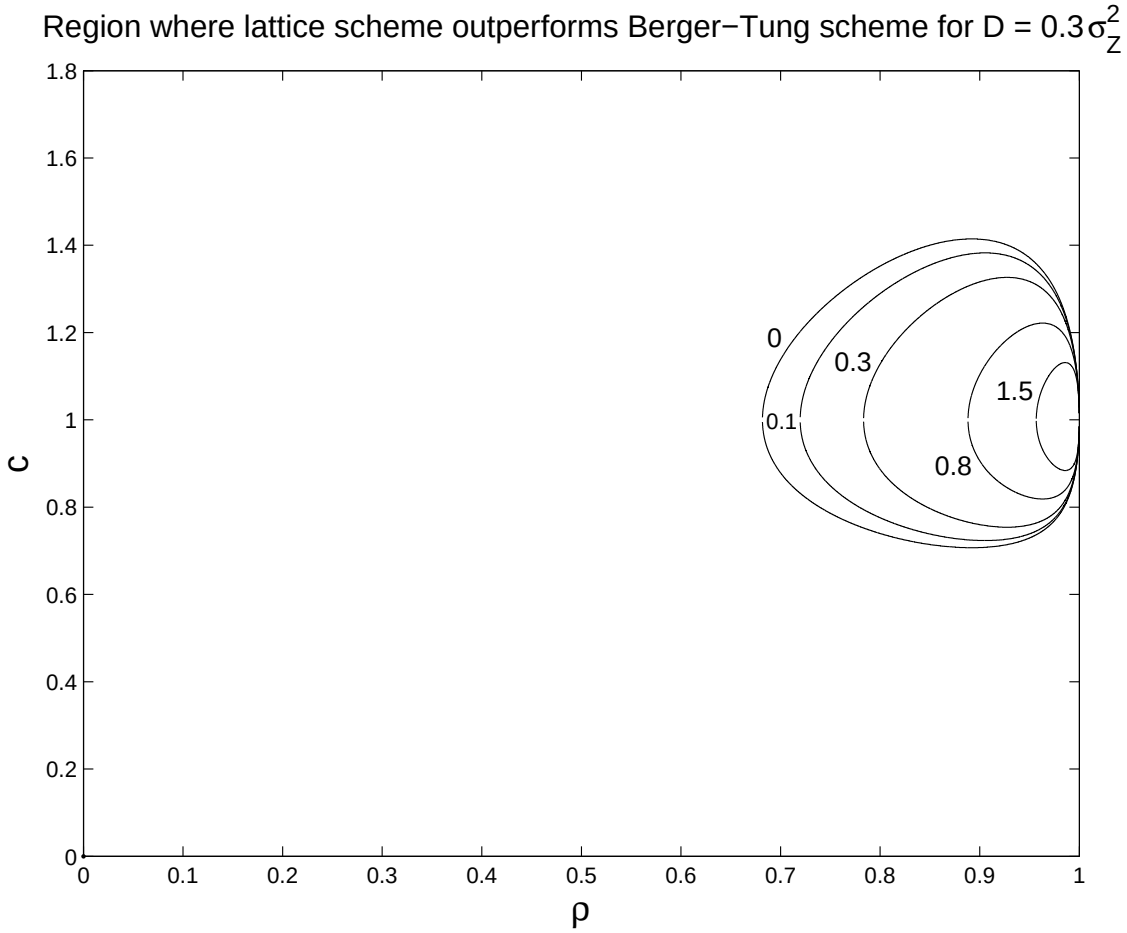

Figure 2.8: Range of (ρ, c) where the lattice scheme performs better than the Berger Tung scheme for $\frac{D}{\sigma_Z^2} = 0.3$

CHAPTER 3

Distributed Source Coding with Abelian Group Codes

3.1 Introduction

In Chapter 2, we studied a distributed source coding problem when the sources are jointly Gaussian. In this chapter, we turn our attention to the case of arbitrary discrete valued sources when the decoder is interested in minimizing a joint distortion criterion including the sources and the reconstruction. We develop a structured coding framework for this problem along the same lines as the lattice coding solution of Chapter 2. The role played by nested lattice codes there will here by played by nested group codes built over abelian groups.

This approach is developed using the following two new ideas. First, we use the fact that any abelian group is isomorphic to the direct sum of primary cyclic groups to enable the decomposition of the source into its constituent "digits" which are then encoded sequentially. Second, we show that, although group codes may not approach the Shannon rate-distortion function in a single source point-to-point setting, it is possible to construct non-trivial group codes which contain a code that approaches it. Using these two ideas, we provide an all-group-code solution to the problem and characterize an inner bound to the performance limit using single-letter information quantities. We also demonstrate the superiority of this approach over

the conventional coding approach based on unstructured random codes using some examples.

3.2 Survey of Group Codes Literature

We now present a brief survey of known results in group codes. Good codes over groups have been studied extensively in the literature when the order (size) of the group is a prime which enables the group to have a field structure. Such codes over Galois fields have been studied for the purpose of packing and covering (see [72, 57] and the references therein). Two kinds of packing problems have received attention in the literature: a) combinatorial rigid packing where the spheres are not allowed to intersect with each other at all and b) probabilistic soft packing where the spheres can have intersections of infinitesimally small measure with one another. Probabilistic soft packing is equivalent to the problem of achieving the capacity of symmetric channels. Similarly, covering problems have been studied in two ways: a) combinatorial complete covering where the entire space needs to be completely covered and (b) probabilistic almost covering where a space of infinitesimally small measure can be left uncovered. Probabilistic soft covering is equivalent to the problem of achieving the rate-distortion function of symmetric sources with Hamming distortion. Some of the salient features of these two approaches have been studied in [55]. In the following we give a sample of works in the direction of probabilistic packing and covering. Elias [65] showed that linear codes can achieve the capacity of binary symmetric channels. A reformulation of this result can be used to show [12] that linear codes can be used to losslessly compress any discrete source down to its entropy. Dobrushin [67] showed that linear codes achieve the random coding error exponent while Forney and Barg [9] showed that linear codes also achieve the

expurgated error exponent. Further, these results have been shown to be true for almost all linear codes. Gallager [68] shows that binary linear codes succeeded by a nonlinear mapping can approach the capacity of any discrete memoryless channel. It follows from Goblick's work [66, 73, 74] on the covering radius of linear codes that linear codes can be used to achieve the rate distortion bound for binary sources with Hamming distortion. Blinovskii [75] derived upper and lower bounds on the covering radius of linear codes and also showed that almost all linear codes (satisfying rate constraints) are good source codes for binary sources with Hamming distortion. If the size of the finite field is sufficiently large, it was shown that in [76] that linear codes followed by a nonlinear mapping can achieve the rate distortion bound of a discrete memoryless source with arbitrary distortion measure. Wyner [2] derived an algebraic binning approach to provide a simple derivation of the Slepian-Wolf [1] rate region for the case of correlated binary sources. Csiszar [42] showed the existence of universal linear encoders which attain the best known error exponents for the Slepian-Wolf problem derived earlier using nonlinear codes. In [45, 47], nested linear codes were used for approaching the Wyner-Ziv rate-distortion function for the case of doubly symmetric binary source and side information with Hamming distortion. Random structured codes have been used in other related multiterminal communication problems [10, 77, 102] to get performance that is superior to that obtained by random unstructured codes. In [71], a coding scheme based on sparse matrices and ML decoding was presented that achieves the known rate regions for the Slepian-Wolf problem, Wyner-Ziv problem and the problem of lossless source coding with partial side information.

Codes over general cyclic groups were first studied by Slepian [78] in the context of signal sets for the Gaussian channel. Forney [79] formalized the concept of geo-

metrically uniform codes and showed that many known classes of good signal space codes were geometrically uniform. Biglieri and Elia [80] addressed the problem of existence of group codes for the Gaussian channel as defined by Slepian. Forney and Loeliger [81, 82] studied the state space representation of group codes and derived trellis representations which were used to build convolutional codes over abelian groups. An efficient algorithm for building such minimal trellises was presented in [83]. Loeliger [84] extended the concept of the M-PSK signal set matched to the M-ary cyclic group to the case of matching general signal sets with arbitrary groups. Building codes over abelian groups with good error correcting properties was studied in [85]. The distance properties of group codes have also been extensively studied. In [86, 87, 88], bounds were derived on the minimum distance of group codes and it was also shown that codes built over nonabelian groups have asymptotically bad minimum distance behavior. Group codes have also been used to build LDPC codes with good distance properties [89]. The information theoretic performance limits of group codes when used as channel codes over symmetric channels was studied in [90]. Similar analysis for the case of turbo codes and geometrically uniform constellations was carried out in [91]. In [92], Ahlswede established the achievable capacity using group codes for several classes of channels and showed that in general, group codes do not achieve the capacity of a general discrete memoryless channel. Sharper results were obtained for the group codes capacity and their upper bounds in [93, 94].

3.3 Problem Definition and Known Results

Consider a pair of discrete random variables (X, Y) with joint distribution $p_{XY}(\cdot, \cdot)$. Let the alphabets of the random variables X and Y be $\mathcal{X}$ and $\mathcal{Y}$ respectively. The source sequence (X^n, Y^n) is independent over time and has the product distribution

$Pr((X^n, Y^n) = (x^n, y^n)) = \prod_{i=1}^{n} p_{XY}(x_i, y_i)$. We consider the following distributed source coding problem. The two components of the source are observed by two encoders which do not communicate with each other. Each encoder communicates a compressed version of its input through a noiseless channel to a joint decoder. The decoder is interested in reconstructing the sources with respect to a general fidelity criterion. Let $\hat{\mathcal{Z}}$ denote the reconstruction alphabet, and the fidelity criterion is characterized by a mapping: $d : \mathcal{X} \times \mathcal{Y} \times \hat{\mathcal{Z}} \to \mathbb{R}^+$. We restrict our attention to additive distortion measures, i.e., the distortion among three n-length sequences x^n, y^n and $\hat{z}^n$ is given by

$$\hat{d}(x^n, y^n, \hat{z}^n) \triangleq \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i, \hat{z}_i). \tag{3.1}$$

In this chapter, we will concentrate on the above distributed source coding problem (with one distortion constraint), and provide an information-theoretic inner bound to the optimal rate-distortion region.

Definition 3.1. Given a discrete source with joint distribution $p_{XY}(x, y)$ and a distortion function $d(\cdot, \cdot, \cdot)$, a transmission system with parameters $(n, \theta_1, \theta_2, \Delta)$ is defined by the set of mappings

$$f_1 \colon \mathcal{X}^n \to \{1, \ldots, \theta_1\}, \qquad f_2 \colon \mathcal{Y}^n \to \{1, \ldots, \theta_2\} \tag{3.2}$$

$$g \colon \{1, \ldots, \theta_1\} \times \{1, \ldots, \theta_2\} \to \hat{\mathcal{Z}}^n \tag{3.3}$$

such that the following constraint is satisfied.

$$\mathbb{E}(\hat{d}(X^n, Y^n, g(f_1(X^n), f_2(Y^n)))) \leq \Delta. \tag{3.4}$$

Here $f_i(\cdot)$ are the encoders and $g(\cdot)$ is the decoder mapping. $\hat{d}(X^n, Y^n, g(f_1(X^n), f_2(Y^n)))$ is the distortion incurred in the reconstruction.

Definition 3.2. We say that a tuple (R_1, R_2, D) is achievable if $\forall \epsilon > 0$, $\exists$ for all sufficiently large n a transmission system with parameters $(n, \theta_1, \theta_2, \Delta)$ such that the encoder rates $\frac{1}{n} \log \theta_i$ satisfy

$$\frac{1}{n} \log \theta_i \leq R_i + \epsilon \quad \text{for } i = 1, 2 \quad \Delta \leq D + \epsilon. \tag{3.5}$$

The performance limit is given by the optimal rate-distortion region $\mathcal{RD}$ which is defined as the set of all achievable tuples (R_1, R_2, D).

Note the similarities between these definitions and Definition 2.1. While the latter definition presumes Gaussian sources, the above definitions are suitable for arbitrary discrete valued sources. We remark that this problem formulation is very general. For example, defining the joint distortion measure $d(X, Y, \hat{Z})$ as $d_1(F(X, Y), \hat{Z})$ enables us to consider the problem of lossy reconstruction of a function of the sources as a special case. Though we only consider a single distortion measure in this chapter, it will become apparent that the results presented here are easily generalizable to the case of multiple distortion criteria. This implies that the problem of reconstructing the sources subject to two independent distortion criteria (the Berger-Tung problem [7]) can be subsumed in this formulation with multiple distortion criteria. The Slepian-Wolf [1] problem, the Wyner-Ziv problem [5], the Yeung-Berger problem [18] and the problem of coding with partial side information [3, 4] can also be subsumed by this formulation since they all are special cases of the Berger-Tung problem. The problem of remote distributed source coding [19, 23], where the encoders observe the sources through noisy channels, can also be subsumed in this formulation using the techniques of [24, 25]. We shall see that our coding theorem has implications on the tightness of the Berger-Tung inner bound [7]. The two-user function computation problem of lossy reconstruction of $Z = F(X, Y)$ can also be viewed as a special case

of three-user Berger-Tung problem of encoding the correlated sources (X, Y, Z) with three independent distortion criteria, where the rate of the third encoder is set to zero and the distortions of the first two sources are set to their maximum values. We shall see in Section 3.9.2 that for this problem, our rate region indeed yields points outside the Berger-Tung rate region thus demonstrating that the Berger-Tung inner bound is not tight for the case of three or more sources.

An achievable rate region for the problem defined in Definitions 3.1 and 3.2 can be obtained based on the Berger-Tung coding scheme [7] as follows. Let $\mathcal{P}$ denote the family of pair of conditional probabilities $(P_{U|X}, P_{V|Y})$ defined on $\mathcal{X} \times \mathcal{U}$ and $\mathcal{Y} \times \mathcal{V}$, where U and V are finite sets. For any $(P_{U|X}, P_{V|Y}) \in \mathcal{P}$, let the induced joint distribution be $P_{XYUV} = P_{XY} P_{U|X} P_{V|Y}$. U, V play the role of auxiliary random variables. Define $G \colon \mathcal{U} \times \mathcal{V} \to \hat{\mathcal{Z}}$ as that function of U, V that gives the optimal reconstruction $\hat{Z}$ with respect to the distortion measure $d(\cdot, \cdot, \cdot)$. With these definitions, an achievable rate region for this problem is presented below.

Fact 2. For a given source (X, Y) and distortion $d(\cdot, \cdot, \cdot)$ define the region $\mathcal{RD}_{BT}$ as

$$\mathcal{RD}_{BT} \triangleq \bigcup_{(P_{U|X}, P_{V|Y}) \in \mathcal{P}} \Big\{ R_1 \geq I(X;U|V),\ R_2 \geq I(Y;V|U),\ R_1 + R_2 \geq I(XY;UV), \tag{3.6}$$
$$D \geq \mathbb{E} d(X, Y, G(U, V)) \Big\}$$

Then any $(R_1, R_2, D) \in \mathcal{RD}_{BT}^*$ is achievable where * denotes convex closure[1].

Proof: Follows from the analysis of the Berger-Tung problem [7] in a straightforward way. □

This rate region is entirely analogous to the rate region presented in Theorem 2 for the case of reconstructing a linear function of jointly Gaussian sources.

---

[1]The cardinalities of U and V can be bounded using Caratheodary theorem [14].

3.4 Groups - An Introduction

In this section, we present an overview of some properties of groups that are used later. We refer the reader to [96] for more details. It is assumed that the reader has some basic familiarity with the concept of groups. We shall deal exclusively with abelian groups and hence the additive notation will be used for the group operation. The group operation of the group G is denoted by $+_G$. Similarly, the identity element of group G is denoted by e_G. The additive inverse of $a \in G$ is denoted by $-a$. The subscripts are omitted when the group in question is clear from the context. A subset H of a group G is called a subgroup if H is a group by itself under the same group operation $+_G$. This is denoted by $H < G$. The direct sum of two groups G_1 and G_2 is denoted by $G_1 \oplus G_2$. The direct sum of a group G with itself n times is denoted by G^n.

An important tool in studying the structure of groups is the concept of group homomorphisms.

Definition 3.3. Let G, H be groups. A function $\phi\colon G \to H$ is called a homomorphism if for any $a, b \in G$

$$\phi(a +_G b) = \phi(a) +_H \phi(b). \tag{3.7}$$

A bijective homomorphism is called an isomorphism. If G and H are isomorphic, it is denoted as $G \cong H$.

A homomorphism $\phi(\cdot)$ has the following properties: $\phi(e_G) = e_H$ and $\phi(-a) = -\phi(a)$. The kernel $\ker(\phi)$ of a homomorphism is defined as $\ker(\phi) \triangleq \{x \in G\colon \phi(x) = e_H\}$. An important property of homomorphisms is that they preserve the subgroup structure. Let $\phi\colon G \to H$ be a homomorphism. Let $A < G$ and $B < H$. Then $\phi^{-1}(B) < G$ and $\phi(A) < H$. In particular, taking $B = \{e_H\}$, we get that $\ker(\phi) < G$.

One can define a congruence result analogous to number theory using subgroups of a group. Let $H < G$ and let $a \in G$. Consider the set $H + a = \{h + a : h \in H\}$. The members of this set form an equivalence class under the equivalence relation $a \sim b$ if $a - b \in H$. This equivalence class is called the right coset of H in G with a as the coset leader. The left coset of H in G is similarly defined. Since we deal exclusively with abelian groups, we shall not distinguish cosets as being left or right. All cosets are of the same size as H and two different cosets are either distinct or identical. Thus, the set of all distinct cosets of H in G form a partition of G. These properties shall be used in our coding scheme.

It is known that a finite cyclic group of order n is isomorphic to the group $\mathbb{Z}_n$ which is the set of integers $\{0, \ldots, n-1\}$ with the group operation as addition modulo-n. A cyclic group whose order is the power of a prime is called a primary cyclic group. The following fact demonstrates the role of primary cyclic groups as the building blocks of all finite abelian groups.

Fact 3. Let G be a finite abelian group of order $n > 1$ and let the unique factorization of n into distinct prime powers be $n = \prod_{i=1}^{k} p_i^{e_i}$. Then,

$$G \cong A_1 \oplus A_2 \cdots \oplus A_k \quad \text{where } |A_i| = p_i^{e_i} \tag{3.8}$$

Further, for each $A_i, 1 \leq i \leq k$ with $|A_i| = p_i^{e_i}$, we have

$$A_i \cong \mathbb{Z}_{p_i^{h_1}} \oplus \mathbb{Z}_{p_i^{h_2}} \cdots \oplus \mathbb{Z}_{p_i^{h_t}} \tag{3.9}$$

where $h_1 \geq h_2 \cdots \geq h_t > 0$ are integers determined by A_i and $\sum_{j=1}^{t} h_j = e_i$. This decomposition of A_i into direct sum of primary cyclic groups is called the invariant factor decomposition of A_i. Putting equations (3.8) and (3.9) together, we get a decomposition of an arbitrary abelian group G into a direct sum of possibly repeated primary cyclic groups. Further, this two step decomposition of G into A_is and then

the decomposition of A_is into $\mathbb{Z}_{p_i}^{h_i}$s is unique, i.e., if $G \cong B_1 \oplus \cdots \oplus B_k$ with $|B_i| = p_i^{e_i}$ for all i, then $B_i \cong A_i$ and B_i and A_i have the same invariant factors.

Proof: See [96], Section 5.2, Theorem 5. □

For example, Fact 3 implies that a given abelian group G of order 8 is isomorphic to either $\mathbb{Z}_8$ or $\mathbb{Z}_4 \oplus \mathbb{Z}_2$ or to $\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$ where $\oplus$ denotes the direct sum of groups. Thus, we first consider the coding theorems only for the primary cyclic groups $\mathbb{Z}_{p^r}$. Results obtained for such groups are then extended to hold for arbitrary abelian groups through this decomposition. Suppose G has a decomposition $G \cong \mathbb{Z}_{p_1^{e_1}} \oplus \cdots \oplus \mathbb{Z}_{p_r^{e_r}}$ where $p_1 \geq \cdots \geq p_r$ are primes. A random variable X taking values in G can be thought of as a vector valued random variable $X = (X_1, \ldots, X_r)$ with X_i taking values in the cyclic group $\mathbb{Z}_{p_i^{e_i}}, 1 \leq i \leq r$. X_i are called the digits of X.

We now present some properties of primary cyclic groups that we shall use in our proofs. The group $\mathbb{Z}_m$ is a commutative ring with the addition operation being addition modulo-m and the multiplication operation being multiplication modulo-m. This multiplicative structure is also exploited in the proofs. The group operation in $\mathbb{Z}_m^n$ is denoted by $u_1^n + u_2^n$. Addition of u_1^n with itself k times is denoted by ku_1^n. The multiplication operation between elements x and y of the underlying ring $\mathbb{Z}_m$ is denoted by xy. We shall say that $x \in \mathbb{Z}_m$ is invertible if there exists $y \in \mathbb{Z}_m$ such that $xy = 1$ where 1 is the multiplicative identity of $\mathbb{Z}_m$. The multiplicative inverse of $x \in \mathbb{Z}_m$, if it exists, is denoted by x^{-1}. The additive inverse of $u_1^n \in \mathbb{Z}_m^n$ which always exists is denoted by $-u_1^n$. The group operation in the group $\mathbb{Z}_m$ is often explicitly denoted by $\oplus_m$.

We shall build our codebooks as kernels of homomorphisms from $\mathbb{Z}_{p^r}^n$ to $\mathbb{Z}_{p^r}^k$, i.e., every sequence in $\mathbb{Z}_{p^r}^n$ that gets mapped to the identity element of $\mathbb{Z}_{p^r}^k$ under

a given homomorphism $\phi(\cdot)$ is considered a codeword and the collection of all such codewords is defined as the codebook corresponding to that homomorphism $\phi(\cdot)$. Justification for restricting the domain of our homomorphisms to $\mathbb{Z}_{p^r}^n$ comes from the decomposition result of Fact 3. The reason for restricting the image of the homomorphisms to $\mathbb{Z}_{p^r}^k$ shall be made clear later on (see the proof of Lemma B.1). We need the following lemma on the structure of homomorphisms from $\mathbb{Z}_{p^r}^n$ to $\mathbb{Z}_{p^r}^k$.

Fact 4. Let $\mathrm{Hom}(\mathbb{Z}_{p^r}^n, \mathbb{Z}_{p^r}^k)$ be the set of all homomorphisms from the group $\mathbb{Z}_{p^r}^n$ to $\mathbb{Z}_{p^r}^k$ and $M(k, n, \mathbb{Z}_{p^r})$ be the set of all $k \times n$ matrices whose elements take values from the group $\mathbb{Z}_{p^r}$. Then, there exists a bijection between $\mathrm{Hom}(\mathbb{Z}_{p^r}^n, \mathbb{Z}_{p^r}^k)$ and $M(k, n, \mathbb{Z}_{p^r})$ given by the invertible mapping $f\colon \mathrm{Hom}(\mathbb{Z}_{p^r}^n, \mathbb{Z}_{p^r}^k) \to M(k, n, \mathbb{Z}_{p^r})$ defined as $f(\phi) = \Phi$ such that $\phi(x^n) = \Phi \cdot x^n$ for all $x^n \in \mathbb{Z}_{p^r}^n$. Here, the multiplication and addition operations involved in the matrix multiplication are carried out modulo-p^r.

Proof: See [97], Section VI. □

3.5 Motivation of the Coding Scheme

In this section, we present a sketch of the ideas involved in our coding scheme by demonstrating them for the simple case when the sources are binary. The emphasis in this section is on providing an overview of the main ideas and the exposition is kept informal. Formal definitions and theorems follow in subsequent sections. We first review the linear coding strategy of [12] to reconstruct losslessly the modulo-2 sum of $Z = X \oplus_2 Y$ of the binary sources X and Y. We then demonstrate that the Slepian-Wolf problem can be solved by a similar coding strategy. We generalize this coding strategy for the case when the fidelity criterion is such that the decoder needs to losslessly reconstruct a function $F(X, Y)$ of the sources. This shall motivate the problem of building "good" channel codes over abelian groups. We then turn

our attention to the lossy version of the problem where the sources X and Y are quantized to U and V respectively first. For this purpose, we need to build "good" source codes over abelian groups. Then, encoding is done in such a way that the decoder can reconstruct $G(U,V)$ which is the optimal reconstruction of the sources with respect to the fidelity criterion $d(\cdot,\cdot,\cdot)$ given U,V. This shall necessitate the need for "good" nested group codes where the coarse code is a good channel code and the fine code is a good source code. These concepts shall be made precise later on in Sections 3.6 and 3.7.

3.5.1 Lossless Reconstruction of the Modulo-2 Sum of the Sources

This problem was studied in [12] where an ingenious coding scheme involving linear codes was presented. This coding scheme can be understood as follows. It is well known [2] that linear codes can be used to losslessly compress a source down to its entropy. Formally, for any binary memoryless source Z with distribution $p_Z(z)$ and any $\epsilon > 0$, there exists a $k \times n$ binary matrix A with $\frac{k}{n} \leq H(Z)+\epsilon$ and a function ψ such that

$$P(\psi(Az^n) \neq z^n) < \epsilon \tag{3.10}$$

for all sufficiently large n. This binary matrix A is the parity check matrix of a linear code that achieves the symmetric channel capacity of a additive noise channel with the noise being independent of channel input and having distribution $p_Z(z)$. Since the encoder transmits the k bit sequence Az^n, the rate of the lossless source is $\frac{k}{n} \leq H(Z) + \epsilon$.

Now, let $Z = X \oplus_2 Y$ be the modulo-2 sum of the binary sources X and Y. Let the matrix A satisfy equation (3.10). The encoders of X and Y transmit $s_1 = Ax^n$ and $s_2 = Ay^n$ respectively at rates $(H(Z), H(Z))$. The decoder, upon receiving s_1

and s_2, computes $\psi(s_1 \oplus_2 s_2) = \psi(Ax^n \oplus_2 Ay^n) = \psi(Az^n)$. Since the A matrix was chosen in accordance with equation (3.10), the decoder output equals z^n with high probability. Thus, the rate pair $(H(Z), H(Z))$ is achievable. If the source statistics is such that $H(Z) > H(X)$, then clearly it is better to compress X at a rate $H(X)$. Thus, the Korner-Marton coding scheme achieves the rate pair (R_1, R_2) with $R_1 \geq \min\{H(X), H(Z)\}$ and $R_2 \geq \min\{H(Y), H(Z)\}$. This coding strategy shall be referred to as the Korner-Marton coding scheme from here on.

The crucial part played by linear codes in this coding scheme is noteworthy. Had there been a centralized encoder with access to x^n and y^n, the coding scheme would be to compute $z^n = x^n \oplus_2 y^n$ first and then compress it using any method known to achieve the entropy bound. Because the encoding is linear, it enables the decoder to use the *distributive* nature of the linear code over the modulo-2 operation to compute $s_1 \oplus_2 s_2 = Az^n$. Thus, from the decoder's perspective, there is no distinction between this distributed coding scheme and a centralized scheme involving a linear code. Also, in contrast to the usual norm in information theory, there is no other known coding scheme that approaches the performance of this linear coding scheme. This critical role played by linear codes in this example is completely analogous to the role played by lattice codes in Chapter 2 and indeed the intuition for using linear codes described above is the same as that described in Section 2.2.3 for using lattice codes.

More generally, in the case of a prime q, a sum rate of $2H(X \oplus_q Y)$ can be achieved [16] for the reconstruction of the sum of the two q-ary sources $Z = X \oplus_q Y$ in any prime field $\mathbb{Z}_q$. Abstractly, the Korner-Marton scheme can be thought of as a structured coding scheme with codes built over groups that enable the decoder to reconstruct the group operation losslessly. It turns out that extending the scheme would involve building "good" channel codes over arbitrary abelian groups. It is

known (see Fact 3) that primary cyclic groups $\mathbb{Z}_{p^r}$ are the building blocks of all abelian groups and hence it suffices to build "good" channel codes over the cyclic groups $\mathbb{Z}_{p^r}$.

3.5.2 Lossless Reconstruction of the Sources

The classical result of Slepian and Wolf [1] states that it is possible to reconstruct the sources X and Y noiselessly at the decoder with a sum rate of $R_1+R_2 = H(X,Y)$. As was shown in [42], the Slepian-Wolf bound is achievable using linear codes. Here, we present an interpretation of this linear coding scheme and connect it to the one in the previous subsection. We begin by making the observation that reconstructing the function $Z = (X,Y)$ for binary sources can be thought of as reconstructing a linear function in the field $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. This equivalence is demonstrated below. Let the elements of $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ be $\{00, 01, 10, 11\}$. Denote the addition operation of $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ by $\oplus_K$ [2].

Define the mappings

$$\tilde{X} = \begin{cases} 00 & \text{if } X = 0 \\ 01 & \text{if } X = 1 \end{cases} \tag{3.11}$$

$$\tilde{Y} = \begin{cases} 00 & \text{if } Y = 0 \\ 10 & \text{if } Y = 1 \end{cases} \tag{3.12}$$

Clearly, reconstructing (X,Y) losslessly is equivalent to reconstructing the function $\tilde{Z} = \tilde{X} \oplus_K \tilde{Y}$ losslessly. The next observation is that elements in $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ can be represented as two dimensional vectors whose components are in $\mathbb{Z}_2$. Further, addition in $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ is simply vector addition with the components of the vector added in $\mathbb{Z}_2$. Let the first and second bits of $\tilde{X}$ be denoted by $\tilde{X}_1$ and $\tilde{X}_2$ respectively. The

[2] The subscript K derives from the $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ group being also known as the Klein-4 group

same notation holds for $\tilde{Y}$ and $\tilde{Z}$ as well. Then, we have the decomposition of the vector function $\tilde{Z}$ as $\tilde{Z}_i = \tilde{X}_i \oplus_2 \tilde{Y}_i$ for $i = 1, 2$.

Encoding the vector function $\tilde{Z}$ directly using the Korner-Marton coding scheme would entail a sum rate of $R_1+R_2 = \min\{H(X,Y), H(X)\}+\min\{H(X,Y), H(Y)\} = H(X)+H(Y)$ which is more than the sum rate dictated by the Slepian-Wolf bound. Instead, we encode the scalar components of the function $\tilde{Z}$ sequentially using the Korner-Marton scheme. Suppose the first digit $\tilde{Z}_1$ is encoded first. Assuming that it gets decoded correctly at the decoder, it is available as side information for the encoding of the second digit $\tilde{Z}_2$. Clearly, the Korner-Marton scheme can be used to encode the first digit $\tilde{Z}_1$. The rate pair (R_{11}, R_{21}) achieved by the scheme is given by

$$R_{11} \geq \min\{H(\tilde{Z}_1), H(\tilde{X}_1)\} = H(\tilde{X}_1) = 0 \tag{3.13}$$

$$R_{21} \geq \min\{H(\tilde{Z}_1), H(\tilde{Y}_1)\} = H(\tilde{Z}_1) \tag{3.14}$$

It is straightforward to extend the Korner-Marton coding scheme to the case where decoder has available to it some side information. Since $\tilde{Z}_1$ is available as side information at the decoder, the rates needed to encode the second digit $\tilde{Z}_2$ are

$$R_{12} \geq \min\{H(\tilde{Z}_2 \mid \tilde{Z}_1), H(\tilde{X}_2 \mid \tilde{Z}_1)\} = H(\tilde{Z}_2 \mid \tilde{Z}_1) \tag{3.15}$$

$$R_{22} \geq \min\{H(\tilde{Z}_2 \mid \tilde{Z}_1), H(\tilde{Y}_2 \mid \tilde{Z}_1)\} = H(\tilde{Y}_2 \mid \tilde{Z}_1) = 0 \tag{3.16}$$

Thus, the overall rate pair needed to reconstruct the sources losslessly is

$$R_1 = R_{11} + R_{12} \geq H(\tilde{Z}_2 \mid \tilde{Z}_1) = H(\tilde{X}_2 \mid \tilde{Y}_1) \tag{3.17}$$

$$R_2 = R_{21} + R_{22} \geq H(\tilde{Z}_1) = H(\tilde{Y}_1). \tag{3.18}$$

The sum rate for this scheme is $R_1 + R_2 = H(\tilde{X}_2, \tilde{Y}_1) = H(X, Y)$ thus equaling the Slepian-Wolf bound.

3.5.3 Lossless Reconstruction of an Arbitrary Function $F(X, Y)$

While there are more straightforward ways of achieving the Slepian-Wolf bound than the method outlined in Section 3.5.2, our encoding scheme has the advantage of putting the Korner-Marton coding scheme and the Slepian-Wolf coding scheme under the same framework. The ideas used in these two examples can be abstracted and generalized for the problem when the decoder needs to losslessly reconstruct some function $F(X, Y)$ in order to satisfy the fidelity criterion.

Let us assume that the cardinality of X and Y are respectively α and β. The steps involved in such an encoding scheme can be described as follows. We first represent the function as equivalent to the group operation in some abelian group A. This is referred to as "embedding" the function in A. This abelian group is then decomposed into its constituent cyclic groups and the embedded function is sequentially encoded using the Korner-Marton scheme outlined in Section 3.5.1. Encoding is done keeping in mind that, to decode a digit, the decoder has as available side information all previously decoded digits.

It suffices to restrict attention to abelian groups A such that $|\mathcal{F}| \leq |A| \leq \alpha\beta$ where $\mathcal{F}$ is the alphabet over which the output of the function $F(\cdot, \cdot)$ takes values. Clearly, if the function $F_1(X, Y) \triangleq (X, Y)$ can be embedded in a certain abelian group, then any function $F(X, Y)$ can be reconstructed in that abelian group. This is because the decoder can proceed by reconstructing the sources (X, Y) and then computing the function $F(X, Y)$. It can be shown (see Appendix B.6) that the function $F_1(X, Y) \triangleq (X, Y)$ can be reconstructed in the group $\mathbb{Z}_\alpha \oplus \mathbb{Z}_\beta$ which is of size $\alpha\beta$. Clearly, $|A| \geq |\mathcal{Z}|$ is a necessary condition for the reconstruction of $Z = F(X, Y)$.

3.5.4 Lossy Reconstruction

We now turn our attention to the case when the decoder wishes to obtain a reconstruction $\hat{Z}$ with respect to a fidelity criterion. The coding strategy is as follows: Quantize the sources X and Y to auxiliary variables U and V. Given the quantized sources U and V, let $G(U,V)$ be the optimal reconstruction with respect to the distortion measure $d(\cdot,\cdot,\cdot)$. Reconstruct the function $G(U,V)$ losslessly using the coding scheme outlined in Section 3.5.3.

Just like we used nested lattice codes in Chapter 2, we shall use nested group codes to effect this quantization. Nested group codes arise naturally in the area of distributed source coding and require that the fine code be a "good" source code and the coarse code be a "good" channel code for appropriate notions of goodness. We have already seen that to effect lossless compression, the channel code operates at the digit level. It follows then that we must use a series of nested group codes, one for each digit, over appropriate cyclic groups. For instance, if the first digit of $G(U,V)$ is over the cyclic group $\mathbb{Z}_{p_1^{e_1}}$, then we need nested group codes over $\mathbb{Z}_{p_1^{e_1}}$ that encode the sources X and Y to $\tilde{U}_1$ and $\tilde{V}_1$ respectively. The quantization operation is also carried out sequentially, i.e., the digits $\tilde{U}_2$ and $\tilde{V}_2$ are encoded given the knowledge that either $\tilde{Z}_1$ or $(\tilde{U}_1,\tilde{V}_1)$ is available at the decoder and so on. The existence of "good" nested group codes over arbitrary cyclic groups is shown later.

The steps involved in the overall coding scheme can be detailed as follows:

- Let U,V be discrete random variables over the alphabet $\mathcal{U},\mathcal{V}$ respectively. Further suppose that $|\mathcal{U}| = \alpha, |\mathcal{V}| = \beta$. Choose the joint density $P_{X,Y,U,V} = P_{X,Y}P_{U|X}P_{V|Y}$ satisfying the Markov chain $U - X - Y - V$.
- Let $G(U,V)$ be the optimal reconstruction function with respect to $d(\cdot,\cdot,\cdot)$ given

U, V.

- Embed the function $G(U, V)$ in an abelian group A, $|\mathcal{G}| \leq |A| \leq \alpha\beta$.
- Decompose $G(U, V)$ into its constituent digits. Fix the order in which the digits are to be sequentially encoded.
- Suppose the b^{th} digit is the cyclic group $\mathbb{Z}_{p_b^{e_b}}$. Quantize the sources (X^n, Y^n) into digits $(\tilde{U}_b, \tilde{V}_b)$ using the digits already available at the decoder as side information. The details of the quantization procedure are detailed later.
- Encode $\tilde{Z}_b = \tilde{U}_b \oplus_{p_b^{e_b}} \tilde{V}_b$ using group codes.

3.6 Definitions

When a random variable X takes value over the group $\mathbb{Z}_{p^r}$, we need to ensure that it doesn't just take values in some proper subgroup of $\mathbb{Z}_{p^r}$. This leads us to the concept of a non-redundant distribution over a group.

Definition 3.4. A random variable X with alphabet $\mathcal{X} = \mathbb{Z}_{p^r}$ and its distribution P_X are said to be non-redundant if $P_X(x) > 0$ for at least one symbol $x \in \mathbb{Z}_{p^r} \backslash p\mathbb{Z}_{p^r}$.

It follows from this definition that a sequence x^n belonging to the typical set $A_\epsilon^n(X)$ contains at least one $x \in \mathbb{Z}_{p^r} \backslash p\mathbb{Z}_{p^r}$ if X is non-redundant. Such sequences are called non-redundant sequences. A redundant random variable taking values over $\mathbb{Z}_{p^r}$ can be made non-redundant by a suitable relabeling of the symbols. Also, note that a redundant random variable over $\mathbb{Z}_{p^r}$ is non-redundant when viewed as taking values over $\mathbb{Z}_{p^{r-i}}$ for some $0 < i \leq r$. Our coding scheme involves good nested group codes for source and channel coding and the notion of embedding the optimal reconstruction function in a suitable abelian group. These concepts are made precise in the following series of definitions.

Definition 3.5. A bivariate function $G\colon \mathcal{U} \times \mathcal{V} \to \mathcal{G}$ is said to be embeddable in an abelian group A with respect to the distribution $p_{UV}(u, v)$ on $\mathcal{U} \times \mathcal{V}$ if there exists injective functions $S_U^{(A)}\colon \mathcal{U} \to A, S_V^{(A)}\colon \mathcal{V} \to A$ and a surjective function $S_G^{(A)}\colon A \to \mathcal{G}$ such that

$$S_G^{(A)}(S_U^{(A)}(u) +_A S_V^{(A)}(v)) = G(u, v) \quad \forall (u, v) \in \mathcal{U} \times \mathcal{V} \text{ with } p_{UV}(u, v) > 0 \tag{3.19}$$

If $G(U, V)$ is indeed embeddable in the abelian group A, it is denoted as $G(U, V) \subset A$ with respect to the distribution $p_{UV}(u, v)$. Define the mapped random variables $\bar{U} = S_U^{(A)}(U)$ and $\bar{V} = S_V^{(A)}(V)$. For the remainder of this chapter, the dependence of $\tilde{U}$ and $\tilde{V}$ on A is suppressed and the group in question will be clear from the context.

Suppose the function $G(U, V) \subset A$ with respect to p_{UV}. We encode the function $G(U, V)$ sequentially by treating the sources as vector valued over the cyclic groups whose direct sum is isomorphic to A. This alternative representation of the sources is made precise in the following definition.

Definition 3.6. Suppose the function $G(U, V) \subset A$ with respect to p_{UV}. Let A be isomorphic to $\oplus_{i=1}^{k} \mathbb{Z}_{p_i^{e_i}}$ where $p_1 \leq \cdots \leq p_k$ are primes and e_i are positive integers. Then, it follows from Fact 3 that there exists a bijection $S_A\colon A \to \mathbb{Z}_{p_1^{e_1}} \times \ldots \mathbb{Z}_{p_k^{e_k}}$. Let $\tilde{U} = S_A(\bar{U}), \tilde{V} = S_A(\bar{V})$. Let $\tilde{U} = (\tilde{U}_1, \ldots, \tilde{U}_k)$ be the vector representation of $\tilde{U}$. The random variables $\tilde{U}_i$ are called the digits of $\tilde{U}$. A similar decomposition holds for $\tilde{V}$. Define $\tilde{Z} = (\tilde{Z}_1, \ldots, \tilde{Z}_k)$ where $\tilde{Z}_i \triangleq \tilde{U}_i \oplus_{p_i^{e_i}} \tilde{V}_i$. It follows that $S_A^{-1}(\tilde{Z}) = \bar{U} +_A \bar{V}$.

Our coding operation proceeds thus: we reconstruct the function $G(U, V)$ by first embedding it in some abelian group A and then reconstructing $\bar{U} +_A \bar{V}$ which we accomplish sequentially by reconstructing $\tilde{U}_i \oplus_{p_i^{e_i}} \tilde{V}_i$ one digit at a time. While reconstructing the ith digit, the decoder has as side information the previously reconstructed $(i - 1)$ digits. This digit decomposition approach requires that we build

codes over the primary cyclic groups $\mathbb{Z}_{p^r}$ which are "good" for various coding purposes. We define the concepts of group codes and what it means for group codes to be "good" in the following series of definitions.

Definition 3.7. Let A be a finite abelian group. A group code $\mathcal{C}$ of blocklength n over the group A is a subset of A^n which is closed under the group addition operation, i.e., $\mathcal{C} \subset A^n$ is such that if $c_1^n, c_2^n \in \mathcal{C}$, then so does $c_1^n +_{A^n} c_2^n$.

Recall that the kernel $\ker(\phi)$ of a homomorphism $\phi\colon A^n \to A^k$ is a subgroup of A^n. We use this fact to build group codes. As mentioned earlier, we build codes over the primary cyclic group $\mathbb{Z}_{p^r}$. In this case, every group code $\mathcal{C} \subset \mathbb{Z}_{p^r}^n$ has associated with it a $k \times n$ matrix H with entries in $\mathbb{Z}_{p^r}$ which completely defines the group code as

$$\mathcal{C} \triangleq \{x^n \in \mathbb{Z}_{p^r}^n \colon Hx^n = 0^k\}. \tag{3.20}$$

Here, the multiplication and addition are carried out modulo-p^r. H is called the parity-check matrix of the code $\mathcal{C}$. We employ nested group codes in our coding scheme. In distributed source coding problems, we often need one of the components of a nested code to be a good source code while the other one to be a good channel code. We shall now define nested group codes and the notions of "goodness" used to classify a group code as a good source or channel code.

Definition 3.8. A nested group code $(\mathcal{C}_1, \mathcal{C}_2)$ is a pair of group codes such that every codeword in the codebook $\mathcal{C}_2$ is also a codeword in $\mathcal{C}_1$, i.e., $\mathcal{C}_2 < \mathcal{C}_1$. Their associated parity check matrices are the $k_1 \times n$ matrix H_1 and the $k_2 \times n$ matrix H_2. They are related to each other as $H_1 = J \cdot H_2$ for some $k_1 \times k_2$ matrix J. One way to enforce

this relation between H_1 and H_2 would be to let

$$H_2 = \begin{bmatrix} H_1 \\ \Delta H \end{bmatrix} \tag{3.21}$$

where ΔH is a $(k_2 - k_1) \times n$ matrix over $\mathbb{Z}_{p^r}$.

The code $\mathcal{C}_1$ is called the fine group code while $\mathcal{C}_2$ is called the coarse group code. When nested group codes are used in distributed source coding, typically the coset leaders of $\mathcal{C}_2$ in $\mathcal{C}_1$ are employed as codewords. In such a case, the rate of the nested group code would be $n^{-1}(k_2 - k_1) \log p^r$ bits.

We define the notion of "goodness" associated with a group code below. To be precise, these notions are defined for a family of group codes indexed by the blocklength n. However, for the sake of notational convenience, this indexing is not made explicit.

Definition 3.9. Let P_{XU} be a distribution over $\mathcal{X} \times \mathcal{U}$ such that the marginal P_U is a non-redundant distribution over $\mathbb{Z}_{p^r}$ for some prime power p^r. For a given group code $\mathcal{C}$ over $\mathcal{U}$ and a given $\epsilon > 0$, let the set $A_\epsilon(\mathcal{C})$ be defined as

$$A_\epsilon(\mathcal{C}) \triangleq \{x^n \colon \exists u^n \in \mathcal{C} \text{ such that } (x^n, u^n) \in A_\epsilon^{(n)}(X, U)\}. \tag{3.22}$$

The group code $\mathcal{C}$ over $\mathcal{U}$ is called a good source code for the triple $(\mathcal{X}, \mathcal{U}, P_{XU})$ if we have $\forall \epsilon > 0$,

$$P_X^n(A_\epsilon(\mathcal{C})) \geq 1 - \epsilon \tag{3.23}$$

for all sufficiently large n.

Note that, a group code which is a good source code in this sense may not be a good source code in the usual Shannon sense. Rather, such a group code contains

a subset which is a good source code in the Shannon sense for the source P_X with forward test channel $P_{U|X}$.

Definition 3.10. Let P_{ZS} be a distribution over $\mathcal{Z} \times \mathcal{S}$ such that the marginal P_Z is a non-redundant distribution over $\mathbb{Z}_{p^r}$ for some prime power p^r. For a given group code $\mathcal{C}$ over $\mathcal{Z}$ and a given $\epsilon > 0$, define the set $B_\epsilon(\mathcal{C})$ as follows:

$$B_\epsilon(\mathcal{C}) \triangleq \{(z^n, s^n) \colon \exists \tilde{z}^n \text{ such that } (\tilde{z}^n, s^n) \in A_\epsilon^{(n)}(Z, S) \text{ and } H\tilde{z}^n = Hz^n\}. \tag{3.24}$$

Here, H is the $k(n) \times n$ parity check matrix associated with the group code $\mathcal{C}$. The group code $\mathcal{C}$ is called a good channel code for the triple $(\mathcal{Z}, \mathcal{S}, P_{ZS})$ if we have $\forall \epsilon > 0$,

$$P_{ZS}^n(B_\epsilon(\mathcal{C})) \leq \epsilon \tag{3.25}$$

for all sufficiently large n. Associated with such a good group channel code would be a decoding function $\psi : \mathbb{Z}_{p^r}^k \times \mathcal{S}^n \to \mathbb{Z}_{p^r}^n$ such that

$$P(\psi(Hz^n, s^n) = z^n) \geq 1 - \epsilon. \tag{3.26}$$

Note that, as before, a group code which is a good channel code in this sense may not a good channel code in the usual Shannon sense. Rather, every coset of such a group code contains a subset which is a good channel code in the Shannon sense for the channel $P_{S|Z}$ with input distribution P_Z. This interpretation is valid only when S is a non-trivial random variable.

Lemma 3.11. *For any triple $(\mathcal{Z}, \mathcal{S}, P_{ZS})$ of two finite sets and a distribution, with $|\mathcal{Z}| = p^r$ a prime power and P_Z non-redundant, there exists a sequence of group codes $\mathcal{C}$ that is a good channel code for the triple $(\mathcal{Z}, \mathcal{S}, P_{ZS})$ such that the dimensions of their associated $k(n) \times n$ parity check matrices satisfy*

$$\lim_{n\to\infty} \frac{k(n)}{n} \log p^r = \max_{0 \leq i < r} \left(\frac{r}{r-i}\right) (H(Z|S) - H([Z]_i|S)) \tag{3.27}$$

where $[Z]_i$ is a random variable taking values over the set of all distinct cosets of $p^i\mathbb{Z}_{p^r}$ in $\mathbb{Z}_{p^r}$. For example, if $\mathcal{Z} = \mathbb{Z}_8$, then $[Z]_2$ is a 4-ary random variable with symbol probabilities $(p_Z(0) + p_Z(4)), (p_Z(1) + p_Z(5)), (p_Z(2) + p_Z(6))$ and $(p_Z(3) + p_Z(7))$.

Proof: See Appendix B.1. □

Note that $[Z]_0$ is a constant and $[Z]_r = Z$. When building codes over groups, each proper subgroup of the group contributes a term to the maximization in equation (3.27). Since the smaller the right hand side of equation (3.27), the better the channel code is, we incur a penalty by building codes over groups with large number of subgroups.

Lemma 3.12. *For any triple $(\mathcal{X}, \mathcal{U}, P_{XU})$ of two finite sets and a distribution, with $|\mathcal{U}| = p^r$ a prime power and P_U non-redundant, there exists a sequence of group codes $\mathcal{C}$ that is a good source code for the triple $(\mathcal{X}, \mathcal{U}, P_{XU})$ such that the dimensions of their associated $k(n) \times n$ parity check matrices satisfy*

$$\lim_{n\to\infty} \frac{k(n)}{n} \log p^r = r|H(U|X) - \log p^{r-1}|^+ \tag{3.28}$$

where $|x|^+ = \max(x, 0)$.

Proof: See Appendix B.2. □

Putting $r = 1$ in equations (3.27) and (3.28), we recover the known results of performance limits obtainable while using linear codes built over Galois fields.

Lemma 3.13. *Let X, Y, S, U, V be five random variables where U and V take value over the group $\mathbb{Z}_{p^r}$ for some prime power p^r. Let $Z = U \oplus_{p^r} V$. Let $U \to X \to Y \to V$ form a Markov chain, and let $S \to (X, Y) \to (U, V)$ form a Markov chain. From the Markov chains, it follows that $H(U|X) \leq H(Z|S), H(V|Y) \leq H(Z|S)$. Without loss*

of generality, let $H(U|X) \leq H(V|Y) \leq H(Z|S)$*. Then, there exists a pair of nested group codes* $(\mathcal{C}_{11}, \mathcal{C}_2)$ *and* $(\mathcal{C}_{12}, \mathcal{C}_2)$ *such that*

- $\mathcal{C}_{11}$ *is a good group source code for the triple* $(\mathcal{X}, \mathcal{U}, P_{XU})$ *with*

$$\lim_{n\to\infty} \frac{k_{11}(n)}{n} \log p^r = r|H(U|X) - \log p^{r-1}|^+ \tag{3.29}$$

- $\mathcal{C}_{12}$ *is a good group source code for the triple* $(\mathcal{Y}, \mathcal{V}, P_{YV})$ *with*

$$\lim_{n\to\infty} \frac{k_{12}(n)}{n} \log p^r = r|H(V|Y) - \log p^{r-1}|^+ \tag{3.30}$$

- $\mathcal{C}_2$ *is a good group channel code for the triple* $(\mathcal{Z}, \mathcal{S}, P_{ZS})$ *with*

$$\lim_{n\to\infty} \frac{k_2(n)}{n} \log p^r = \max_{0\leq i<r} \left(\frac{r}{r-i}\right) (H(Z|S) - H([Z]_i|S)) \tag{3.31}$$

Proof*:* See Appendix B.3 □

Note that while choosing the codebooks $\mathcal{C}_{11}, \mathcal{C}_{12}$ and $\mathcal{C}_2$, the perturbation parameters ϵ in Definitions 3.9 and 3.10 need to be chosen appropriately relative to each other so that the n-length sequences $(X^n, Y^n, S^n, U^n, V^n, Z^n)$ are jointly typical with high probability. Due to the Markov chains $U \to X \to Y \to V$ and $S \to (X, Y) \to (U, V)$, it follows from Markov lemma [6] that if (X^n, Y^n, S^n) is generated according to P_{XYS} and if U^n is generated jointly typical with X^n and V^n is generated jointly typical with Y^n, then $(X^n, Y^n, S^n, U^n, V^n, Z^n)$ is jointly strongly typical (for an appropriate choice of ϵ) with high probability.

3.7 The Coding Theorem

We are given discrete random variables X and Y which are jointly distributed according to P_{XY}. Let $\mathcal{P}$ denote the family of pair of conditional probabilities $(P_{U|X}, P_{V|Y})$ defined on $\mathcal{X} \times \mathcal{U}$ and $\mathcal{Y} \times \mathcal{V}$, where $\mathcal{U}$ and $\mathcal{V}$ are finite sets, $|\mathcal{U}| =$

$\alpha, |\mathcal{V}| = \beta$. For any $(P_{U|X}, P_{V|Y}) \in \mathcal{P}$, let the induced joint distribution be $P_{XYUV} = P_{XY}P_{U|X}P_{V|Y}$. U, V play the role of auxiliary random variables. Define $G\colon \mathcal{U} \times \mathcal{V} \to \hat{\mathcal{Z}}$ as that function of U, V that gives the optimal reconstruction $\hat{Z}$ with respect to the distortion measure $d(\cdot, \cdot, \cdot)$. Let $\mathcal{G}$ denote the image of $G(U, V)$. Let $\mathcal{T} = \{A\colon A \text{ is abelian}, |\mathcal{G}| \leq |A| \leq \alpha\beta, G(U, V) \subset A \text{ with respect to } P_{UV}\}$. It is shown in Appendix B.6 that the set $\mathcal{T}$ is non-empty, i.e., there always exists an abelian group $A \in \mathcal{T}$ in which any function $G(U, V)$ can be embedded. For any $A \in \mathcal{T}$, let A be isomorphic to $\oplus_{i=1}^{k} \mathbb{Z}_{p_i^{e_i}}$. Let $\tilde{U} = S_A(S_U^{(A)}(U))$ and $\tilde{V} = S_A(S_V^{(A)}(V))$ where the mappings are as defined in Definitions 3.5 and 3.6. Define $\tilde{Z} = (\tilde{Z}_1, \ldots, \tilde{Z}_k)$ where $\tilde{Z}_i = \tilde{U}_i \oplus \tilde{V}_i$ and the addition is done in the group to which the digits $\tilde{U}_i, \tilde{V}_i$ belong. Assume without loss of generality that the digits $\tilde{U}_i, \tilde{V}_i, \tilde{Z}_i, 1 \leq i \leq k$ are all non-redundant. If they are not, they can be made so by suitable relabeling of the symbols. Recall the definition of $[Z]_i$ from Lemma 3.11. The encoding operation of the X and Y encoders proceed in k steps with each step producing one digit of $\tilde{U}$ and $\tilde{V}$ respectively. Let $\pi_A\colon \{1, \ldots, k\} \to \{1, \ldots, k\}$ be a permutation. The permutation π_A can be thought of as determining the order in which the digits get encoded and decoded. Let the set $\Pi_A(b), 1 \leq b \leq k$ be defined as $\Pi_A(b) = \{l\colon \pi_A(l) < b\}$. The set $\Pi_A(b)$ contains the indices of all the digits that get encoded before the bth stage. At the bth stage, let the digits $\tilde{U}_{\pi_A(b)}, \tilde{V}_{\pi_A(b)}$ take values over the group $\mathbb{Z}_{p_b}^{r_b}$. With these definitions, an achievable rate region for the problem is presented below.

Theorem 4. *For a given source (X, Y), define the region $\mathcal{RD}_{in}$ as*

$$\mathcal{RD}_{in} \triangleq \bigcup_{\substack{(P_{U|X}, P_{V|Y}) \in \mathcal{P} \\ A \in \mathcal{T}, \pi_A}} \Bigg\{ (R_1, R_2, D) \colon R_1 \geq \sum_{b=1}^{k} \min\left(R_{1b}^{(1)}, R_{1b}^{(2)}\right), \\ R_2 \geq \sum_{b=1}^{k} \min\left(R_{2b}^{(1)}, R_{2b}^{(2)}\right), \\ D \geq \mathbb{E}d(X, Y, G(U, V)) \Bigg\} \tag{3.32}$$

where

$$R_{1b}^{(1)} > \left[\max_{0 \leq i < r_b} \left(\frac{r_b}{r_b - i}\right) \left(H(\tilde{Z}_{\pi_A(b)} \mid \tilde{Z}_{\Pi_A(b)}) - H([\tilde{Z}_{\pi_A(b)}]_i | \tilde{Z}_{\Pi_A(b)})\right)\right] \\ - r_b(|H(\tilde{U}_{\pi_A(b)} \mid X, \tilde{U}_{\Pi_A(b)}) - \log p_b^{r_b - 1}|^+) \tag{3.33}$$

and

$$R_{1b}^{(2)} > \left[\max_{0 \leq i < r_b} \left(\frac{r_b}{r_b - i}\right) \left(H(\tilde{U}_{\pi_A(b)} \mid \tilde{Z}_{\Pi_A(b)}) - H([\tilde{U}_{\pi_A(b)}]_i \mid \tilde{Z}_{\Pi_A(b)})\right)\right] \\ - r_b(|H(\tilde{U}_{\pi_A(b)} \mid X, \tilde{U}_{\Pi_A(b)}) - \log p_b^{r_b - 1}|^+) \tag{3.34}$$

The quantities $R_{2b}^{(1)}$ and $R_{2b}^{(2)}$ are similarly defined with (X, U) replaced by (Y, V). Then any $(R_1, R_2, D) \in \mathcal{RD}_{in}^*$ is achievable where * denotes convex closure.

***Proof**:* Since the encoders don't communicate with each other, we impose the Markov chain $V - Y - X - U$ on the joint distribution P_{XYUV}. The family $\mathcal{P}$ contains all distributions that satisfy this Markov chain. Fix such a joint distribution. Fix $A \in \mathcal{T}$ and the permutation $\pi_A \colon \{1, \ldots, k\} \to \{1, \ldots, k\}$. The encoding proceeds in k stages with the bth stage encoding the digits $\tilde{U}_{\pi_A(b)}, \tilde{V}_{\pi_A(b)}$ in order to produce the digit $\tilde{Z}_{\pi_A(b)}$. For this, the decoder has side information $\tilde{Z}_{\Pi_A(b)}$.

Let $\tilde{U}_{\pi_A(b)}, \tilde{V}_{\pi_A(b)}$ take values over the group $\mathbb{Z}_{p^b}^{r_b}$. The encoders have two encoding options available at the bth stage. They can either encode the digits $\tilde{U}_{\pi_A(b)}$ and

$\tilde{V}_{\pi_A(b)}$ directly or encode in such a way that the decoder is able to reconstruct $\tilde{Z}_{\pi_A(b)}$ directly. We present a coding scheme to achieve the latter first.

We shall use a pair of nested group codes $(\mathcal{C}_{11b}, \mathcal{C}_{2b})$ and $(\mathcal{C}_{12b}, \mathcal{C}_{2b})$ to encode $\tilde{Z}_{\pi_A(b)}$. Let the corresponding parity check matrices of these codes be H_{11b}, H_{12b} and H_{2b} respectively. Let the dimensionality of these matrices be $k_{11b} \times n$, $k_{12b} \times n$ and $k_{2b} \times n$ respectively. These codebooks are all over the group $\mathbb{Z}_{p^b}^{r_b}$. We need C_{11b} to be a good source code for the triple $(\mathcal{X} \times \tilde{\mathcal{U}}_{\Pi_A(b)}, \tilde{\mathcal{U}}_{\pi_A(b)}, P_{X\tilde{U}_{\Pi_A(b)}\tilde{U}_{\pi_A(b)}})$, $\mathcal{C}_{12b}$ to be a good source code for the triple $(\mathcal{Y} \times \tilde{\mathcal{V}}_{\Pi_A(b)}, \tilde{\mathcal{V}}_{\pi_A(b)}, P_{Y\tilde{V}_{\Pi_A(b)}\tilde{V}_{\pi_A(b)}})$ and $\mathcal{C}_{2b}$ to be a good channel code for the triple $(\tilde{\mathcal{Z}}_{\pi_A(b)}, \tilde{\mathcal{Z}}_{\Pi_A(b)}, P_{\tilde{Z}_{\pi_A(b)}\tilde{Z}_{\Pi_A(b)}})$.

The encoding scheme used by the X-encoder to encode the bth digit, $1 \leq b \leq k$ is detailed below. The X-encoder looks for a typical sequence $\tilde{U}^n_{\pi_A(b)} \in \mathcal{C}_{11b}$ such that it is jointly typical with the source sequence X^n and the previous encoder output digits $\tilde{U}^n_{\Pi_A(b)}$. If it finds at least one such sequence, it chooses one of these sequences and transmits the syndrome $Sx_b \triangleq H_{2b}\tilde{U}^n_{\pi_A(b)}$ to the decoder. If it finds no such sequence, it declares an encoding error. The operation of the Y-encoder is similar.

Let $\psi_b(\cdot,\cdot)$ be the decoder corresponding to the good channel code $\mathcal{C}_{2b}$. The decoder action is described by the following series of equations. The decoder receives the syndromes Sx_b and Sy_b.

$$\begin{aligned}
\hat{\tilde{Z}}_{\pi_A(b)} &= \psi_b\left(Sx_b \oplus_{{p_b}^{r_b}} Sy_b, \tilde{Z}^n_{\Pi_A(b)}\right) \\
&= \psi_b\left(H_{2b}\tilde{U}^n_{\pi_A(b)} \oplus_{{p_b}^{r_b}} H_{2b}\tilde{V}^n_{\pi_A(b)}, \tilde{Z}^n_{\Pi_A(b)}\right) \\
&= \psi_b\left(H_{2b}\left(\tilde{U}^n_{\pi_A(b)} \oplus_{{p_b}^{r_b}} \tilde{V}^n_{\pi_A(b)}\right), \tilde{Z}^n_{\Pi_A(b)}\right) \\
&= \psi_b\left(H_{2b}\tilde{Z}^n_{\pi_A(b)}, \tilde{Z}^n_{\Pi_A(b)}\right) \\
&\overset{(a)}{=} \tilde{Z}^n_{\pi_A(b)} \quad \text{with high probability}
\end{aligned} \tag{3.35}$$

where (a) follows from the fact that $\mathcal{C}_{2b}$ is a good channel code for the triple

$(\tilde{\mathcal{Z}}_{\pi_A(b)}, \tilde{\mathcal{Z}}_{\Pi_A(b)}, P_{\tilde{Z}_{\pi_A(b)}\tilde{Z}_{\Pi_A(b)}})$.

The rate expended by the X-encoder at the bth stage can be calculated as follows. Since $\mathcal{C}_{11b}$ is a good source code for the triple $(\mathcal{X} \times \tilde{\mathcal{U}}_{\Pi_A(b)}, \tilde{\mathcal{U}}_{\pi_A(b)}, P_{X\tilde{U}_{\Pi_A(b)}\tilde{U}_{\pi_A(b)}})$, we have from equation (3.28) that the dimensions of the parity check matrix H_{11b} satisfy

$$\frac{k_{11b}}{n} \log p_b^{r_b} \leq r_b(|H(\tilde{U}_{\pi_A(b)} \mid X, \tilde{U}_{\Pi_A(b)}) - \log p_b^{r_b-1}|^+) - \epsilon_1 \tag{3.36}$$

Since $\mathcal{C}_{2b}$ is a good channel code for the triple $(\tilde{\mathcal{Z}}_{\pi_A(b)}, \tilde{\mathcal{Z}}_{\Pi_A(b)}, P_{\tilde{Z}_{\pi_A(b)}\tilde{Z}_{\Pi_A(b)}})$, the dimensions of the parity check matrix H_{2b} satisfy

$$\frac{k_{2b}}{n} \log p_b^{r_b} \geq \max_{0 \leq i < r_b} \left(\frac{r_b}{r_b - i}\right) \Big(H(\tilde{Z}_{\pi_A(b)} \mid \tilde{Z}_{\Pi_A(b)}) - H([\tilde{Z}_{\pi_A(b)}]_i | \tilde{Z}_{\Pi_A(b)})\Big) + \epsilon_2 \tag{3.37}$$

The rate of the nested group code in bits would be $R_1 = n^{-1}(k_{2b} - k_{11b}) \log p_b^{r_b}$. Therefore,

$$\begin{aligned} R_{1b}^{(1)} \geq & \left[\max_{0 \leq i < r_b} \left(\frac{r_b}{r_b - i}\right) \Big(H(\tilde{Z}_{\pi_A(b)} \mid \tilde{Z}_{\Pi_A(b)}) - H([\tilde{Z}_{\pi_A(b)}]_i | \tilde{Z}_{\Pi_A(b)})\Big)\right] \\ & - r_b(|H(\tilde{U}_{\pi_A(b)} \mid X, \tilde{U}_{\Pi_A(b)}) - \log p_b^{r_b-1}|^+) + \epsilon_1 + \epsilon_2 \end{aligned} \tag{3.38}$$

The other option that the encoders have is to directly encode the digits $\tilde{U}_{\pi_A(b)}$ and $\tilde{V}_{\pi_A(b)}$. This can also be accomplished using nested group codes as follows. The X encoder uses the nested group code $(\mathcal{C}_{11b}, \mathcal{C}_{21b})$ such that the fine group code $\mathcal{C}_{11b}$ is a good source code for the triple $(\mathcal{X} \times \tilde{\mathcal{U}}_{\Pi_A(b)}, \tilde{\mathcal{U}}_{\pi_A(b)}, P_{X\tilde{U}_{\Pi_A(b)}\tilde{U}_{\pi_A(b)}})$ and $\mathcal{C}_{21b}$ is a good channel code for the triple $(\tilde{\mathcal{U}}_{\pi_A(b)}, \tilde{\mathcal{Z}}_{\Pi_A(b)}, P_{\tilde{U}_{\pi_A(b)}\tilde{Z}_{\Pi_A(b)}})$. The Y encoder uses the nested group code $(\mathcal{C}_{12b}, \mathcal{C}_{22b})$ such that the fine group code $\mathcal{C}_{12b}$ is a good source code for the triple $(\mathcal{Y} \times \tilde{\mathcal{V}}_{\Pi_A(b)}, \tilde{\mathcal{V}}_{\pi_A(b)}, P_{Y\tilde{V}_{\Pi_A(b)}\tilde{V}_{\pi_A(b)}})$ and $\mathcal{C}_{22b}$ is a good channel code for the triple $(\tilde{\mathcal{V}}_{\pi_A(b)}, \tilde{\mathcal{Z}}_{\Pi_A(b)}, P_{\tilde{V}_{\pi_A(b)}\tilde{Z}_{\Pi_A(b)}})$. The encoding operation is similar to that described earlier and it is easy to verify its correctness.

The rate of this nested group code in bits would be $R_1 = n^{-1}(k_{2b} - k_{11b}) \log p_b^{r_b}$.

Therefore,

$$R_{1b}^{(2)} \geq \left[\max_{0\leq i<r_b}\left(\frac{r_b}{r_b-i}\right)\left(H(\tilde{U}_{\pi_A(b)} \mid \tilde{Z}_{\Pi_A(b)}) - H([\tilde{U}_{\pi_A(b)}]_i \mid \tilde{Z}_{\Pi_A(b)})\right)\right]$$

$$- r_b(|H(\tilde{U}_{\pi_A(b)} \mid X, \tilde{U}_{\Pi_A(b)}) - \log p_b^{r_b-1}|^+) + \epsilon_1 + \epsilon_2 \tag{3.39}$$

Combining equations (3.38) and (3.39), we have proved Theorem 4. □

Remark 1: The design of the channel code used in the above derivation assumes that the side information available to the decoder at the bth stage is $\tilde{Z}_{\Pi_A(b)}$. However, it is possible that at some stage $1 \leq i \leq k$, the encoding was done in such a way that the decoder could decode $(\tilde{U}_{\pi_A(i)}, \tilde{V}_{\pi_A(i)})$ and not just $\tilde{Z}_{\pi_A(b)}$. Taking such considerations into account while designing the channel code for the bth stage would lead to a possible improvement of the rate region in Theorem 4.

Remark 2: In the above derivation, if the encoders choose to encode the sources $\tilde{U}_{\pi_A(b)}, \tilde{V}_{\pi_A(b)}$ directly instead of encoding the function $\tilde{Z}_{\pi_A(b)}$, further rate gains are possible when one encoder encodes its source conditional on the other source in addition to the side information already available at the decoder. Such improvements are omitted for the sake of clarity of the expressions constituting the definition of the achievable rate region.

Remark 3: The above coding theorem can be extended to the case of multiple distortion constraints in a straightforward fashion.

3.8 Special cases

In this section, we consider the various special cases of the rate region presented in Theorem 4.

3.8.1 Lossless Source Coding using Group Codes

We start by demonstrating the achievable rates using codes over groups for the problem of lossless source coding with one encoder and one decoder. A good group channel code $\mathcal{C}$ for the triple $(\mathcal{X}, 0, P_X)$ as defined in Definition 3.10 can be used to achieve lossless source coding of the source X. The source encoder outputs Hx^n where H is the $k \times n$ parity check matrix of $\mathcal{C}$. The decoder uses the associated decoding function $\psi(\cdot, \cdot)$ to recover $\psi(Hx^n, 0) = x^n$ with high probability. From equation (3.27), it follows that the dimensions of the parity check matrix satisfy

$$\frac{k}{n} \log p^r \geq \max_{0 \leq i < r} \left(\frac{r}{r-i} \right) (H(X) - H([X]_i)) \tag{3.40}$$

Recognizing the term in the left as the rate of the coding scheme, we get that there exists a group based coding scheme that achieves a rate equalling the RHS of equation (3.40). A sufficient condition for the existence of group codes that attain the entropy bound is that

$$H([X]_i) \geq \frac{i}{r} H(X) \quad \text{for } 0 < i < r \tag{3.41}$$

In Appendix B.4, we show that given a random variable X taking values in $\mathcal{X} = \mathbb{Z}_{p^r}$, it is always possible to relabel the symbols in $\mathcal{X}$ such that the sufficient condition of equation (3.41) is met. Thus, we get the following corollary to Theorem 4.

Corollary 1. Suppose X is a non redundant random variable over the group $\mathbb{Z}_{p^r}$ and the decoder wants to reconstruct X losslessly. Then, there exists a group based coding scheme (possibly involving relabeling of the elements of $\mathcal{X}$) that can encode the source X at rates arbitrarily close to $H(X)$, the entropy of X.

3.8.2 Lossy Source Coding using Group Codes

We next consider the case of lossy point to point source coding using codes built over the group $\mathbb{Z}_{p^r}$. Consider a memoryless source X with distribution P_X. The decoder attempts to reconstruct U that is within distortion D of X as specified by some additive distortion measure $d\colon \mathcal{X} \times \mathcal{U} \to \mathbb{R}^+$. Suppose U takes its values from the group $\mathbb{Z}_{p^r}$. A good group source code $\mathcal{C}$ for the triple $(\mathcal{X},\mathcal{U}, P_{XU})$ as defined in Definition 3.9 can be used to achieve lossy coding of the source X provided the joint distribution P_{XU} is such that $\mathbb{E}(d(X,U)) \leq D$ and U is non-redundant. The source encoder outputs $u^n \in \mathcal{C}$ that is jointly typical with the source sequence x^n. An encoding error is declared if no such u^n is found. The decoder uses u^n as its reconstruction of the source x^n. From equation (3.28), it follows that the dimensions of the parity check matrix associated with $\mathcal{C}$ satisfy

$$\frac{k}{n}\log p^r \leq r|H(U|X) - \log p^{r-1}|^+ \tag{3.42}$$

The rate of this encoding scheme is $R = \left(1 - \frac{k}{n}\right)\log p^r$. Thus, we get the following corollary to Theorem 4.

Corollary 2. Let X be a discrete memoryless source and $\mathcal{U}$ be the reconstruction alphabet. Suppose $\mathcal{U} = \mathbb{Z}_{p^r}$ and the decoder wants to reconstruct the source to within distortion D as measured by the fidelity criterion $d(\cdot,\cdot)$. Without loss of generality, assume that U is non-redundant. Then, there exists a group based coding scheme that achieves the rate

$$R \geq \min_{\substack{P_{U|X} \\ \mathbb{E}d(X,U)\leq D}} \log p^r - r|H(U|X) - \log p^{r-1}|^+. \tag{3.43}$$

If U takes values in a general abelian group of order n that is not necessarily a primary cyclic group, then a decomposition based approach similar to the one used

in the proof of Theorem 4 can be used. When $H(U|X) < \log p^{r-1}$, equation (3.43) suggests that there are no good group codes in the ensemble that we have considered. When $H(U|X) > \log p^{r-1}$, the equation (3.43) can be simplified to read

$$R \geq \min_{\substack{P_{U|X} \\ \mathbb{E}d(X,U) \leq D}} r\left(\log p^r - H(U|X)\right) \tag{3.44}$$

When $r = 1$, this can be viewed as providing an achievable rate-distortion pair for lossy source coding using linear codes built over Galois fields. Note that it is possible to construct codebooks with rate $R = H(U) - H(U|X)$ by choosing codewords independently and uniformly from the set $A_\epsilon^n(U)$. By imposing the group structure on the codebook, we incur a rate loss of $(\log p - H(U))$ bits per sample. This rate loss is strictly positive unless the random variable U is uniformly distributed over $\mathcal{U} = \mathbb{Z}_p$.

When $r > 1$, the multiplicative factor of r in equation (3.44) implies that the rate loss incurred by using group codes over $\mathbb{Z}_{p^r}$ increases as the number of subgroups of the underlying group over which the code is built increases. Unlike the case of lossless source coding where group codes can be used to achieve the Shannon entropy bound, group codes always incur a strictly positive rate loss (except in the trivial case when $(H(U|X) = \log |\mathcal{U}|)$ compared to the Shannon rate-distortion bound.

3.8.3 Nested Linear Codes

We specialize the rate region of Theorem 4 to the case when the nested group codes are built over cyclic groups of prime order, i.e., over Galois fields of prime order. In this case, group codes over $\mathbb{Z}_{p^r}$ reduce to the well known linear codes over prime fields. It was already shown in Sections 3.8.1 and 3.8.2 that Lemmas 3.11 and 3.12 imply that linear codes achieve the entropy bound and incur a rate loss while used in lossy source coding. In this section, we demonstrate the implications

of Theorem 4 when specialized to the case of nested linear codes, i.e., when r is set to 1.

Shannon Rate-Distortion Function

We remark that Theorem 4 shows the existence of nested linear codes that can be used to approach the rate-distortion bound in the single-user setting for arbitrary discrete sources and arbitrary distortion measures.

Corollary 3. Let X be a discrete memoryless source with distribution P_X and let $\hat{\mathcal{X}}$ be the reconstruction alphabet. Let the fidelity criterion be given by $d\colon \mathcal{X}\times\hat{\mathcal{X}} \to \mathbb{R}^+$. Then, there exists a nested linear code $(\mathcal{C}_1, \mathcal{C}_2)$ that achieves the rate-distortion bound

$$R(D) = \min_{\substack{P_{\hat{X}|X} \\ \mathbb{E}d(X,\hat{X})\leq D}} I(X;\hat{X}) \tag{3.45}$$

Proof: Let the optimal forward test channel that achieves the bound be given by $P_{\hat{X}|X}$. Suppose q is a prime such that $\hat{\mathcal{X}} \subset \mathbb{Z}_q$ and $\hat{X}$ is non-redundant[3]. The rate bound, given by $I(X;\hat{X})$ can be approached using a nested linear code $(\mathcal{C}_1, \mathcal{C}_2)$ built over the group $\mathbb{Z}_q$. Here $\mathcal{C}_1$ is a good source code for the triple $(\mathcal{X}, \hat{\mathcal{X}}, P_{X,\hat{X}})$ and $\mathcal{C}_2$ is a good channel code for the triple $(\hat{\mathcal{X}}, \mathcal{S}, P_{\hat{X}S})$ where $\mathcal{S} = \{0\}$ and S is a degenerate random variable with $P_S(0) = 1$. It follows from Lemmas 3.12 and 3.11 that the dimensions of the parity check matrices associated with $\mathcal{C}_1$ and $\mathcal{C}_2$ satisfy

$$\lim_{n\to\infty} \frac{k_1(n)}{n} \log q = H(\hat{X}|X) \tag{3.46}$$

$$\lim_{n\to\infty} \frac{k_2(n)}{n} \log q = H(\hat{X}) \tag{3.47}$$

Thus, the rate achieved by this scheme is given by $n^{-1}(k_2(n)-k_1(n)) \log q = I(X;\hat{X})$.

□

[3]Here, we assume without loss of generality that $\hat{\mathcal{X}} \subset \mathbb{Z}_q$. For a random variable $\hat{X}$ that takes values in an arbitrary set $\hat{\mathcal{X}}$, we can justify this assumption by choosing a prime $q > |\hat{\mathcal{X}}|$ and defining a one-to-one mapping $\pi\colon \hat{\mathcal{X}} \to \mathbb{Z}_q$ such that the random variable $\pi(\hat{X})$ is non-redundant.

This can be intuitively interpreted as follows. For a code to approach the optimal rate-distortion function, the "Voronoi" region (under an appropriate encoding rule) of most of the codewords should have a certain shape (say, shape A), and a high-probability set of codewords should be bounded in a region that has a certain shape (say, shape B). We choose $\mathcal{C}_1$ such that the "Voronoi" region (under the joint typicality encoding operation with respect to $p_{\hat{X},X}$) of each codeword has shape A. $\mathcal{C}_2$ is chosen such that its "Voronoi" region has shape B. Hence the set of "coset leaders" of $\mathcal{C}_1$ in $\mathcal{C}_2$ forms a code that can approach the optimal rate-distortion function. This reminds us of a similar phenomenon first observed in the case of Gaussian sources with mean squared error criterion in [101], where the performance of a quantizer is measured by so-called granular gain and boundary gain. Granular gain measures how closely the Voronoi regions of the codewords approach a sphere, and boundary gain measures how closely the boundary region approaches a sphere.

Berger-Tung Rate Region

We now show that Theorem 4 implies that nested linear codes built over prime fields can achieve the rate region of the Berger-Tung based coding scheme presented in Lemma 2.

Corollary 4. Suppose we have a pair of correlated discrete sources (X, Y) and the decoder is interested in reconstructing $\hat{Z}$ to within distortion D as measured by a fidelity criterion $d\colon \mathcal{X} \times \mathcal{Y} \times \hat{\mathcal{Z}} \to \mathbb{R}^+$. For this problem, an achievable rate region using nested linear codes is given by

$$\mathcal{RD}_{BT} = \bigcup_{(P_{U|X}, P_{V|Y}) \in \mathcal{P}} \{(R_1, R_2)\colon R_1 \geq I(X;U|Y),$$

$$R_2 \geq I(Y;V|X), R_1 + R_2 \geq I(X;U) + I(Y;V) - I(U;V)\} \tag{3.48}$$

where $\mathcal{P}$ is the family of all joint distributions P_{XYUV} that satisfy the Markov chain $U-X-Y-V$ such that the distortion criterion $\mathbb{E}d(X,Y,\hat{Z}(U,V)) \leq D$ is met. Here $\hat{Z}(U,V)$ is the optimal reconstruction of $\hat{Z}$ with respect to the distortion criterion given U and V.

Proof: We proceed by first reconstructing the function $G(U,V) = (U,V)$ at the decoder and then computing the function $\hat{Z}(U,V)$. For ease of exposition, assume that $\mathcal{U} = \mathcal{V} = \mathbb{Z}_q$ for some prime q. If they are not, a decomposition based approach can be used and the proof is similar to the one presented below. Clearly, $G(U,V)$ can be embedded in the abelian group $A \triangleq \mathbb{Z}_q \oplus \mathbb{Z}_q$. The associated mappings are given by $\tilde{U} = (U,0)$ and $\tilde{V} = (0,V)$ where 0 is the identity element in $\mathbb{Z}_q$. Thus, $\tilde{Z}_1 = U + 0 = U$ and $\tilde{Z}_2 = 0 + V = V$. Encoding is done in two stages. Let the permutation $\pi_A(\cdot)$ be the identity permutation. Substituting this into equations (3.38) and (3.39) gives us

$$
\begin{aligned}
R_{11} &\geq \min\{H(\tilde{Z}_1), H(\tilde{U}_1)\} - H(\tilde{U}_1|X) = I(X;\tilde{U}_1) = I(X;U), \\
R_{21} &\geq \min\{H(\tilde{Z}_1), H(\tilde{V}_1)\} - H(\tilde{V}_1|Y) = 0, \\
R_{12} &\geq \min\{H(\tilde{Z}_2 \mid \tilde{Z}_1), H(\tilde{U}_2 \mid \tilde{U}_1)\} - H(\tilde{U}_2 \mid X, \tilde{U}_1) = 0, \\
R_{22} &\geq \min\{H(\tilde{Z}_2 \mid \tilde{Z}_1), H(\tilde{V}_2 \mid \tilde{V}_1)\} - H(\tilde{V}_2 \mid Y, \tilde{V}_1) \\
&= H(\tilde{Z}_2 \mid \tilde{Z}_1) - H(\tilde{V}_2 \mid Y, \tilde{V}_1) = H(V|U) - H(V|Y) \\
&= I(Y;V|U) \qquad (3.49)
\end{aligned}
$$

This is one of the corner points of the rate region given in equation (3.48). Choosing the permutation $\pi_A(\cdot)$ to be the derangement gives us the other corner point and time sharing between the two points yields the entire rate region of equation (3.48). The rate needed to reconstruct U, V at the decoder coincides with the Berger-Tung rate region [6, 7]. □

We note that this implies that our theorem recovers the rate regions of the problems considered by Wyner and Ziv [5], Ahlswede-Korner-Wyner [4, 3], Berger and Yeung [18] and Slepian and Wolf [1] since the Berger-Tung problem encompasses all these problems as special cases.

3.8.4 Lossless Reconstruction of Modulo-2 Sum of Binary Sources

In this section, we show that Theorem 4 recovers the rate region derived by Korner and Marton [12] for the reconstruction of the modulo-2 sum of two binary sources. Let X, Y be correlated binary sources. Let the decoder be interested in reconstructing the function $F(X, Y) = X \oplus_2 Y$ losslessly. In this case, the auxiliary random variables can be chosen as $U = X, V = Y$. Clearly, this function can be embedded in the groups $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_4$ and $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. For embedding in $\mathbb{Z}_2$, the rate region of Theorem 4 reduces to

$$R_1 \geq \min(H(X), H(X \oplus_2 Y)), \quad R_2 \geq \min(H(Y), H(X \oplus_2 Y)) \tag{3.50}$$

It can be verified that embedding in $\mathbb{Z}_3$ or $\mathbb{Z}_4$ always gives a worse rate than embedding in $\mathbb{Z}_2$. Embedding in $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ results in the Slepian-Wolf rate region. Combining these rate regions, we see that a sum rate of $R_1 + R_2 = \min(2H(X \oplus_2 Y), H(X, Y))$ is achievable using our coding scheme. This recovers the Korner-Marton rate region for this problem [12, 14]. Moreover, one can also show that this approach can recover the Ahlswede-Han rate region [17] for this problem, which is an improvement over the Korner-Marton region.

3.9 Examples

In this section, we consider examples of the coding theorem (Theorem 4). First we consider the problem of losslessly reconstructing a function of correlated quaternary

sources. We then derive an achievable rate region for the case when the decoder is interested in the modulo-2 sum of two binary sources to within a Hamming distortion of D.

3.9.1 Lossless Encoding of a Quaternary Function

Consider the following distributed source coding problem. Let (X, Y) be correlated random variables both taking values in $\mathbb{Z}_4$. Let X, Z be independent random variables taking values in $\mathbb{Z}_4$ according to the distributions P_X and P_Z respectively. Define $p_i \triangleq P_X(i), q_i \triangleq P_Z(i)$ for $i = 0, \ldots, 3$. Assume further that the random variable Z is non-redundant, i.e., $q_1 + q_3 > 0$. Define the random variable Y as $Y = X \oplus_4 Z$. Suppose X and Y are observed by two separate encoders which communicate their quantized observations to a central decoder. The decoder is interested in reconstructing the function $Z = (X - Y) \mod 4$ losslessly.

Since we are interested in lossless reconstruction, we can choose the auxiliary random variables U, V to be $U = X, V = Y$. The function $G(U, V)$ then reduces to $F(X, Y) \triangleq (X - Y) \mod 4$. This function can be embedded in several groups with order less than or equal to 16. We claim that this function $F(X, Y)$ can be embedded in the groups $\mathbb{Z}_4, \mathbb{Z}_7, \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$ and $\mathbb{Z}_4 \oplus \mathbb{Z}_4$ with nontrivial performance. For each of these groups, we compute the achievable rate as given by Theorem 4 below. For simplicity, we restrict ourselves to the rate regions given by equation (3.38) alone.

Lets consider the group $\mathbb{Z}_4$ first. Define the mappings $\tilde{x} \triangleq S_X^{(\mathbb{Z}_4)}(x) = x$ for all $x \in \mathbb{Z}_4, \tilde{y} \triangleq S_Y^{(\mathbb{Z}_4)}(y) = -y$ for all $y \in \mathbb{Z}_4$ and $S_F^{(\mathbb{Z}_4)}(z) = z$ for all $z \in \mathbb{Z}_4$. With these mappings, it follows from Definition 3.5 that $F(X, Y)$ is embeddable in $\mathbb{Z}_4$ with respect to the distribution P_{XY}. From Theorem 4, it follows that an achievable rate

region using this embedding is given by

$$R_1 \geq \max\{H(Z), 2(H(Z) - H([Z]_1))\}$$

$$= \max\{h(q_0, q_1, q_2, q_3), 2(h(q_0, q_1, q_2, q_3) - h(q_0 + q_2, q_1 + q_3))\} \tag{3.51}$$

$$R_2 \geq \max\{H(Z), 2(H(Z) - H([Z]_1))\}$$

$$= \max\{h(q_0, q_1, q_2, q_3), 2(h(q_0, q_1, q_2, q_3) - h(q_0 + q_2, q_1 + q_3))\} \tag{3.52}$$

giving a sum rate of

$$R_{\mathbb{Z}_4} \triangleq R_1 + R_2 \geq 2\max\{h(q_0, q_1, q_2, q_3), 2(h(q_0, q_1, q_2, q_3) - h(q_0 + q_2, q_1 + q_3))\} \tag{3.53}$$

It can be verified that $F(X, Y)$ can't be embedded in $\mathbb{Z}_5$ or $\mathbb{Z}_6$. It can be embedded in $\mathbb{Z}_7$ with the following mappings. Define $\tilde{x} \triangleq S_X^{(\mathbb{Z}_7)}(x) = x$ for all $x \in \mathbb{Z}_4$, $\tilde{y} \triangleq S_Y^{(\mathbb{Z}_7)}(y) = -y$ for all $y \in \mathbb{Z}_4$ where $-y$ is the additive inverse of y in $\mathbb{Z}_7$ and $S_F^{(\mathbb{Z}_7)}(0) = 0, S_F^{(\mathbb{Z}_7)}(1) = S_F^{(\mathbb{Z}_7)}(4) = 1, S_F^{(\mathbb{Z}_7)}(2) = S_F^{(\mathbb{Z}_7)}(5) = 2, S_F^{(\mathbb{Z}_7)}(3) = S_F^{(\mathbb{Z}_7)}(6) = 3$. Let $Z = \tilde{X} \oplus_7 \tilde{Y}$. From Theorem 4, it follows that an achievable rate region using this embedding is given by

$$R_1 \geq H(Z) = h(q_0, (1 - p_0)q_3, (1 - p_0 - p_1)q_2, p_3q_1, p_0q_3, (p_0 + p_1)q_2, (1 - p_3)q_1) \tag{3.54}$$

$$R_2 \geq H(Z) = h(q_0, (1 - p_0)q_3, (1 - p_0 - p_1)q_2, p_3q_1, p_0q_3, (p_0 + p_1)q_2, (1 - p_3)q_1) \tag{3.55}$$

giving a sum rate of

$$R_{\mathbb{Z}_7} \triangleq R_1 + R_2 \geq 2h(q_0, (1 - p_0)q_3, (1 - p_0 - p_1)q_2, p_3q_1, p_0q_3, (p_0 + p_1)q_2, (1 - p_3)q_1) \tag{3.56}$$

Of the three abelian groups of order 8, it can be verified that embedding $F(X, Y)$ in $\mathbb{Z}_8$ results in the same rate region as given by equations (3.54) and (3.55) and

(a) $S_X(\cdot)$

X	$\tilde{X}$
0	000
1	001
2	100
3	101

(b) $S_Y(\cdot)$

Y	$\tilde{Y}$
0	000
1	010
2	100
3	110

(c) $S_F(\cdot)$

z	$S_F(z)$
000,011	0
001,110	1
100,111	2
010,101	3

Table 3.1: Mappings for embedding $F(X,Y)$ in $\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$

embedding $F(X,Y)$ in $\mathbb{Z}_2 \oplus \mathbb{Z}_4$ results in the same rate region as given by equations (3.51) and (3.52). So, we consider embedding $F(X,Y)$ in $\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$. Recall that elements of the abelian group $\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$ can be treated as 3 bit vectors over $\mathbb{Z}_2$. The mappings $S_X^{(\mathbb{Z}_2\oplus\mathbb{Z}_2\oplus\mathbb{Z}_2)}(\cdot), S_Y^{(\mathbb{Z}_2\oplus\mathbb{Z}_2\oplus\mathbb{Z}_2)}(\cdot)$ and $S_F^{(\mathbb{Z}_2\oplus\mathbb{Z}_2\oplus\mathbb{Z}_2)}(\cdot)$ are as given in Table 3.1.

Define the random variable $\tilde{Z} = \tilde{U} \oplus \tilde{V}$ where $\oplus$ is addition in $\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$. With these mappings, an achievable rate region can be derived using Theorem 4 as below. Choose the permutation $\pi_{\mathbb{Z}_2\oplus\mathbb{Z}_2\oplus\mathbb{Z}_2}(\cdot)$ as $\pi(1) = 2, \pi(2) = 3, \pi(3) = 1$. Encoding is carried out in 3 stages with the corresponding rates being

$$R_{11} = 0, R_{21} = H(\tilde{Z}_2) \tag{3.57}$$

$$R_{12} = H(\tilde{Z}_3 \mid \tilde{Z}_2), R_{22} = 0 \tag{3.58}$$

$$R_{13} = H(\tilde{Z}_1 \mid \tilde{Z}_2, \tilde{Z}_3), R_{23} = H(\tilde{Z}_1 \mid \tilde{Z}_2, \tilde{Z}_3). \tag{3.59}$$

Summing over the 3 stages of encoding, we get an achievable sum rate of $R_1 + R_2 \geq H(Z) + H(\tilde{Z}_1 \mid \tilde{Z}_2, \tilde{Z}_3) = 2H(Z) - H(\tilde{Z}_2, \tilde{Z}_3)$. In terms of p_i, q_i, this sum rate can be expressed as

$$\begin{aligned} R_{\mathbb{Z}_2\oplus\mathbb{Z}_2\oplus\mathbb{Z}_2} \triangleq R_1 + R_2 \geq\ & 2h(p_{02}q_0, p_{13}q_3, p_{02}q_1, p_{13}q_0, p_{02}q_2, p_{13}q_1, p_{02}q_3, p_{13}q_2) \\ & - h(p_{02}q_{02}, p_{13}q_{13}, p_{02}q_{13}, p_{13}q_{02}) \end{aligned} \tag{3.60}$$

where $p_{02} \triangleq p_0 + p_2, p_{13} \triangleq p_1 + p_3, q_{02} \triangleq q_0 + q_2$ and $q_{13} \triangleq q_1 + q_3$.

Embedding $F(X,Y)$ in groups of order 9 to 15 result in rate regions which are

P_X	P_Z	$R_{\mathbb{Z}_4}$	$R_{\mathbb{Z}_7}$	$R_{\mathbb{Z}_2\oplus\mathbb{Z}_2\oplus\mathbb{Z}_2}$	$R_{\mathbb{Z}_4\oplus\mathbb{Z}_4}$
$[\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$	$[\frac{1}{2}\ 0\ \frac{1}{4}\ \frac{1}{4}]$	3	3.9056	3.1887	3.5
$[\frac{3}{10}\ \frac{6}{10}\ \frac{1}{10}\ 0]$	$[0\ \frac{4}{5}\ \frac{1}{20}\ \frac{3}{20}]$	2.3911	2.0797	2.4529	2.1796
$[\frac{1}{3}\ \frac{1}{10}\ \frac{1}{2}\ \frac{1}{15}]$	$[\frac{3}{7}\ \frac{1}{7}\ \frac{1}{7}\ \frac{2}{7}]$	3.6847	4.5925	3.3495	3.4633
$[\frac{9}{10}\ \frac{1}{30}\ \frac{1}{30}\ \frac{1}{30}]$	$[\frac{3}{20}\ \frac{3}{4}\ \frac{1}{20}\ \frac{1}{20}]$	2.308	2.7065	1.9395	1.7815

Table 3.2: Example distributions for which embedding in a given group gives the lowest sum rate.

worse than the ones already derived. We next present an achievable rate region when $F(X,Y)$ is embedded in $\mathbb{Z}_4 \oplus \mathbb{Z}_4$. We use the mappings $S_X^{(\mathbb{Z}_4\oplus\mathbb{Z}_4)}(x) = x0$ for all $x \in \mathbb{Z}_4$, $S_Y^{(\mathbb{Z}_4\oplus\mathbb{Z}_4)}(y) = 0y$ for all $y \in \mathbb{Z}_4$ and $S_F^{(\mathbb{Z}_4\oplus\mathbb{Z}_4)}(xy) = (x,y)$ for all $(x,y) \in \mathbb{Z}_4^2$. This embedding corresponds to reconstructing the sources X and Y losslessly and the rate region coincides with the Slepian-Wolf rate region.

(3.61)

$$R_{\mathbb{Z}_4\oplus\mathbb{Z}_4} \triangleq R_1 + R_2 \geq H(X,Y) = H(X) + H(Z) = h(p_0,p_1,p_2,p_3) + h(q_0,q_1,q_2,q_3)$$

Combining equations (3.53), (3.56), (3.60) and (3.61) gives us an achievable rate region for this problem. Each of these achievable rate regions outperform the others for certain values of P_X and P_Z. This is illustrated in Table 3.2.

3.9.2 Lossy Reconstruction of the Modulo-2 Sum of Binary Sources

This example concerns the reconstruction of the binary XOR function with the Hamming distortion criterion. The rate region of Theorem 4 is very cumbersome to calculate analytically in the general case. So, we restrict our attention to the case of symmetric source distribution and additive test channels in the derivation below where the intention is to demonstrate the analytical evaluation of the rate region of Theorem 4. We then present plots where the entire sum rate-distortion region is computed without any restrictive assumptions.

Consider a binary correlated source (X,Y) with symmetric joint distribution

$P_{XY}(0,0) = P_{XY}(1,1) = q/2$ and $P_{XY}(1,0) = P_{XY}(0,1) = p/2$. Suppose we are interested in reconstructing $F(X,Y) = X \oplus_2 Y$ within Hamming distortion D. We present an achievable rate pair for this problem based on Theorem 4 and compare it to the achievable rate region presented in Lemma 2. It was shown in [99] that it suffices to restrict the cardinalities of the auxiliary random variables U and V to the cardinalities of their respective source alphabets in order to compute the Berger-Tung rate region. Since the scheme presented in Lemma 2 is based on the Berger-Tung coding scheme, the rate region $\mathcal{RD}_{BT}$ for this problem can be computed by using binary auxiliary random variables.

Let us now evaluate the rate region provided by Theorem 4 for this problem. The auxiliary random variables U and V are binary and suppose the test channel $P_{XY}P_{U|X}P_{V|Y}$ is fixed. The function $G(U,V)$ which is the optimal reconstruction of $X \oplus_2 Y$ given U and V can then be computed. In general, this function can take any of the 16 possible values depending upon the test channel $P_{XY}P_{U|X}P_{V|Y}$.

Let us choose the auxiliary random variables U and V to be binary and for ease of exposition, let them be defined as $U = X \oplus_2 Q_1$ and $V = Y \oplus_2 Q_2$. Here Q_1, Q_2 are independent binary random variables with $P(Q_i = 0) = q_i, i = 1,2$. Let $p_i = 1 - q_i, i = 1,2$. Define $\alpha = q_1q_2 + p_1p_2$ and $\beta = 1 - \alpha$. Once the test channel $P_{XY}P_{U|X}P_{V|Y}$ is thus fixed, the optimal reconstruction function $G(U,V)$ that minimizes the probability $P(F(X,Y) \neq G(U,V))$ can be computed. It can be showed that

$$G(U,V) = \begin{cases} 0 & \alpha > p, \alpha < q \\ U \oplus_2 V & \alpha > p, \alpha > q \\ \overline{U \oplus_2 V} & \alpha < p, \alpha < q \\ 1 & \alpha < p, \alpha > q \end{cases} \tag{3.62}$$

where $\overline{a}$ denotes the complement of the bit a. The corresponding distortion for these reconstructions can be calculated as

$$D(\alpha) = \begin{cases} p & \alpha > p, \alpha < q \\ \beta & \alpha > p, \alpha > q \\ \alpha & \alpha < p, \alpha < q \\ q & \alpha < p, \alpha > q \end{cases} \tag{3.63}$$

Clearly, no rate need be expended if the function to be reconstructed is $G(U,V) = 0$ or $G(U,V) = 1$. It is also easy to see that the rates needed would be the same for both $G(U,V) = U \oplus_2 V$ and $G(U,V) = \overline{U \oplus_2 V}$. Let us therefore consider only reconstructing $G(U,V) = U \oplus_2 V$. It can be shown that this function is embeddable in the groups $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_4$ and $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. Let us consider the group $A \triangleq \mathbb{Z}_2$. The associated mappings $S_U^{(A)}(\cdot), S_V^{(A)}(\cdot)$ and $S_G^{(A)}(\cdot)$ are all identity mappings. In this case, we have only one digit to encode. Further, note that $P(Z_1 = 0) = P(U_1 \oplus_2 V_1 = 0) = q\alpha + p\beta$.

The rates of the encoders are given by equations (3.38) and (3.39) to be

$$\begin{aligned} R_{11} &= \min\{H(U_1), H(Z_1)\} - H(U_1 \mid X) \\ &= \min\{1, h(q\alpha + p\beta)\} - h(q_1) \\ &= h(q\alpha + p\beta) - h(q_1) \end{aligned} \tag{3.64}$$

$$\begin{aligned} R_{21} &= \min\{H(V_1), H(Z_1)\} - H(V_1 \mid Y) \\ &= \min\{1, h(q\alpha + p\beta)\} - h(q_2) \\ &= h(q\alpha + p\beta) - h(q_2) \end{aligned} \tag{3.65}$$

where $h(\cdot)$ is the binary entropy function. Thus, an achievable rate region for this

problem is

(3.66)

$$\mathcal{R} = \bigcup_{0 \le q_1, q_2 \le 1} \{(R_1, R_2, D) \colon R_1 \ge h(q\alpha + p\beta) - h(q_1), R_2 \ge h(q\alpha + p\beta) - h(q_2), \\ D \ge D(\alpha)\}$$

where $D(\alpha)$ is given in equation (3.63). Rate points achieved by embedding the function in the abelian groups $\mathbb{Z}_3, \mathbb{Z}_4$ are strictly worse than that achieved by embedding the function in $\mathbb{Z}_2$ while embedding in $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ gives the Slepian-Wolf rate region for the lossless reconstruction of (X, Y).

We now plot the entire sum rate-distortion region for the case of a general source distribution and general test channels $P_{U|X}, P_{V|Y}$ and compare it with the Berger-Tung rate region $\mathcal{RD}_{BT}$ of Fact 2.

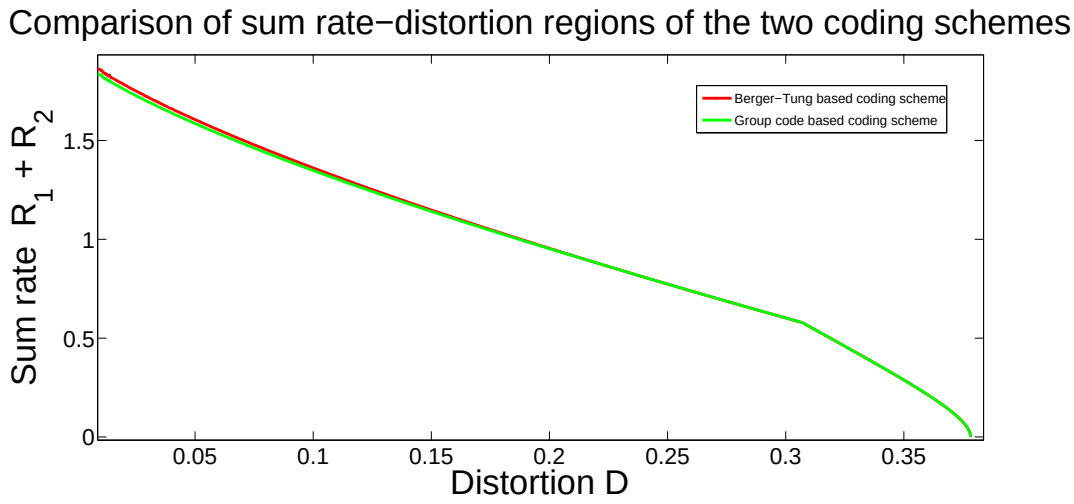

Figure 3.1: Sum rate-distortion region for the distribution given in Table 3.3

Figures 3.1 and 3.2 demonstrate that the sum rate-distortion regions of Theorem 4 and Fact 2. Theorem 4 offers improvements over the rate region of Fact 2 for low distortions as shown more clearly in Figure 3.3. The joint distribution of the sources used in this example is given in Table 3.3.

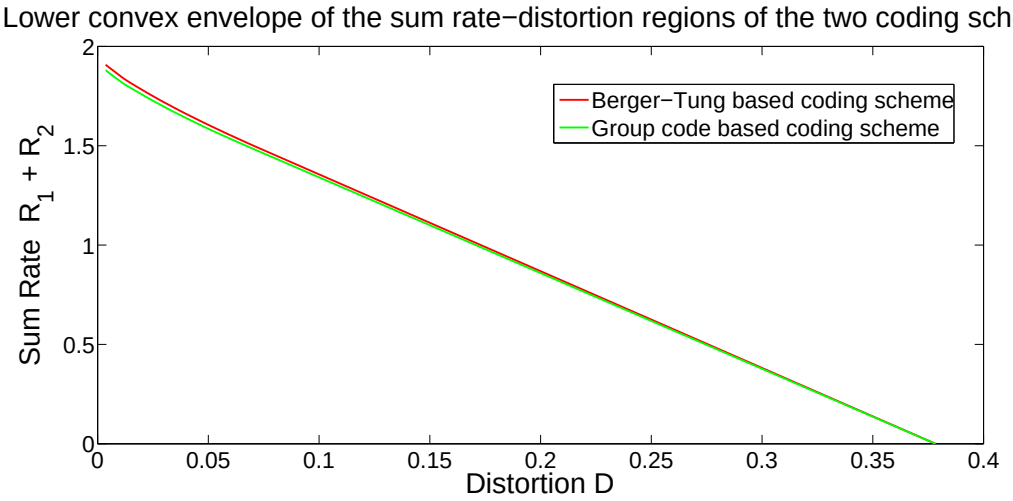

Figure 3.2: Lower Convex envelope of the sum rate-distortion region

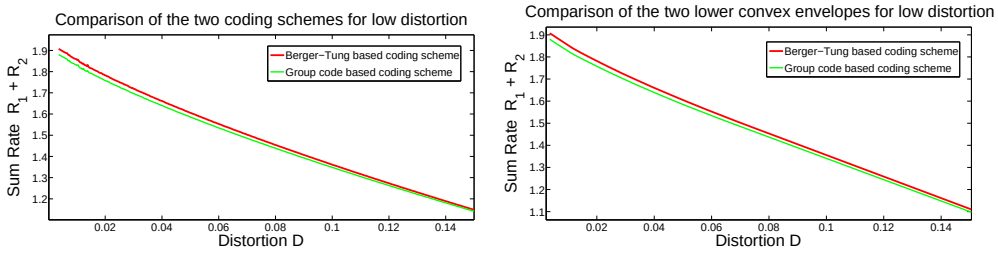

Figure 3.3: Zoomed versions of Figures 3.1 and 3.2

P_{XY}	0	1
0	0.3381	0.1494
1	0.2291	0.2834

Table 3.3: Joint distribution used for example in Figures 3.1 and 3.2

Motivation of choosing this example is as follows. Evaluation of the Berger-tung rate region is a computationally intensive operation since it involves solving a nonconvex optimization problem. The only procedure that we are aware of for this is using linear programming followed by quantizing the probability space and searching for optimum values [99]. The computational complexity increases dramatically as the size of the alphabet of the sources goes beyond two. Hence we chose the simplest nontrivial lossy example to make the point. This forces us to operate with abelian groups of order less than or equal to 4 of which there are only 3. One of the three groups corresponds to the Berger-Tung bound. We would like to remark that even for this simple example, the Berger-Tung bound is not tight. We expect the gains afforded by Theorem 4 over the rate region of Lemma 2 would increase as we increase the cardinality of the source alphabets and more abelian groups become available for embedding.

CHAPTER 4

Conclusions and Future Work

4.1 Summary

In this thesis, a fairly general distributed source coding problem with multiple encoders, a central decoder and a joint distortion criterion is studied. Two variants of the problem are studied - (a) jointly Gaussian sources and a decoder interested in reconstructing a linear function of the sources and (b) arbitrary discrete valued sources and a decoder interested in minimizing a joint distortion criterion.

In Chapter 2, a formal definition of the problem where multiple encoders observe components of a jointly Gaussian source is presented. A central decoder is interested in reconstructing a linear function of the sources to within a certain mean square error distortion. Two coding strategies for a special case of this problem involving two encoders are presented - one which involved lossy reconstruction of the sources at the decoder first and estimation of the linear function later and our approach which involved direct reconstruction of the function at the decoder. The use of nested lattice codes in our coding scheme is motivated and justified. An inner bound to the optimal rate-distortion region is obtained using both schemes. An outer bound is alo presented and its gap from the inner bound is investigated. It is shown that for certain source statistics, our inner bound is within 1 bit of the optimal rate-distortion

region. A general coding scheme for an arbitrary number of sources is then presented which is a combination of both coding strategies discussed above. Certain special cases of this general case are discussed. The two different coding strategies are then analytically compared in the low distortion regime which yielded insights into the scenarios when one scheme outperforms another. To conclude the chapter, some numerical calculations were presented that corroborated the analysis of the previous sections. Appendix A contains many of the proofs for this chapter including a proof of the existence of "good" nested lattice codes for the notions of goodness needed in our coding scheme.

In Chapter 3, the problem of distributed source coding with discrete sources and a joint distortion criterion is discussed. A survey of known results most of which followed the common paradigm of "independent quantization followed by independent binning" is first presented. Just as nested lattice codes were used in the coding problem of Chapter 2, the need for nested group codes over abelian groups is motivated and justified. A quick survey of the properties of abelian groups and their associated homomorphisms that are relevant to our coding scheme is then given. An overview of the coding scheme is then given through illustrative examples. Existence results for "good" group source and channel codes along with the associated notions of goodness are then given before the main coding theorem is presented. As a special case of this coding theorem, several important corollaries are derived including the achievable rates for lossless and lossy source coding using group codes and the achievability of the Shannon rate-distortion bound using nested linear codes. Finally, two examples are presented that demonstrate the use of the main coding theorem in lossless and lossy distributed source coding.

4.2 Future Work

In the course of this research, we have encountered several interesting problems that merit further study. Some of them are listed below.

- **Use of structured codes in multi-terminal channel coding** The focus of this thesis has been on the application of structured codes (lattice/group codes) for the distributed source coding problem. As has been observed, structured codes offer performance gains for many problems in this domain. The duality between source and channel coding in information theory suggests that there must exist problems in multi-terminal channel coding where existing capacity results can be improved using structured codes. In particular, broadcast channels are a direct dual to the distributed source coding problem and their capacity region is not fully known. It is an area where structured codes might offer improvement over the capacity regions of existing coding schemes.

- **Nested lattice codes for arbitrary continuous sources** In Chapter 2, we used nested lattice codes for the case when the sources were jointly Gaussian. However, the coding scheme presented in Chapter 3 while being similar in spirit to the lattice based coding scheme of Chapter 2 works for arbitrary discrete sources and arbitrary additive distortion measures. This strongly suggests that the theory of nested lattice codes is powerful enough to deal with arbitrary continuous sources rather than just Gaussian sources. In this general case, lattice quantization will no longer be based on the "nearest neighbor" rule but rather on joint typicality. The notions of goodness presented in Section 2.1.1 are tailored towards Gaussian sources and channels and need to be suitably generalized.

- **Practical lattice/group code construction** A significant advantage that structured codes offer over unstructured random codes is their ease of implementation. Over the past decade, great strides have been made towards implementation of capacity achieving codes which have efficient encoding and decoding operations. In view of our result that nested linear codes can achieve the known rate regions for many distributed coding problems, it is an interesting and practically relevant problem to build practical codes that approach their theoretical counterparts in performance. The machinery of low density parity check (LDPC) codes and low density generator matrix (LDGM) codes can be used for this purpose.

- **Good group codes over non-abelian groups** In Chapter 3, we demonstrated the existence of good codes over abelian groups. A natural extension of this problem is to build good codes over non-abelian groups. Apart from being an interesting problem in its own right, such non-abelian group codes, if they exist, have the potential to further improve the rate gains structured codes offer for the distributed source coding problem. While abelian groups have a reasonably simple classification as the direct sum of primary cyclic groups (a fact we used in our proofs), non-abelian groups have no such classification. Even if such a classification were to exist, it would likely be of little practical value. A promising strategy for building codes over non-abelian groups would be to restrict attention to a well-studied and well-understood class of non-abelian groups (such as nilpotent groups) rather than attempting to build codes over a general non-abelian group. Mimicking the strategy of Chapter 3 and allowing the codebooks to be kernels of homomorphisms from G^n to G^k turns out to be too restrictive when G is non-abelian. This is because such kernels are always

normal subgroups of G^n and the ensemble of normal subgroups of G^n does not contain good codes when G is non-abelian. Analytically tractable ways of dealing with ensembles of subgroups of G^n need to be developed and this would likely involve more sophisticated tools from group theory than what was needed for the case of constructing good codes over abelian groups.

APPENDICES

APPENDIX A

Proofs for Chapter 2

A.1 Derivation of Berger-Tung based scheme's sum rate

In this section, we derive the sum-rate of the Berger-Tung based scheme given in equations (2.21)-(2.23). The sum-rate of the Berger-Tung based coding scheme is given by

$$R_1 + R_2 \geq \frac{1}{2} \log \frac{(1+q_1)(1+q_2) - \rho^2}{q_1 q_2} \tag{A.1}$$

where $(q_1, q_2) \in \mathbb{R}_+^2$ should satisfy the distortion constraint

$$D \geq \frac{q_1\alpha + q_2 c^2 \alpha + q_1 q_2 \sigma_Z^2}{(1+q_1)(1+q_2) - \rho^2} \tag{A.2}$$

where $\mathbb{R}_+$ is the set of positive reals and $\alpha = 1 - \rho^2$.

To minimize the sum-rate, we need to minimize the quantity given by equation (A.1). Using the fact that the log function is monotone and that (q_1, q_2) must satisfy the distortion constraint in equation (A.2), the minimization problem is equivalent to minimizing

$$\frac{(1+q_1)(1+q_2) - \rho^2}{q_1 q_2} = \frac{q_1\alpha + q_2 c^2 \alpha + q_1 q_2 \sigma_Z^2}{D q_1 q_2} \tag{A.3}$$

and this is equivalent to minimizing

$$\frac{1}{q_2} + \frac{c^2}{q_1} \tag{A.4}$$

subject to the constraint in equation (A.2).

Assuming that (q_1, q_2) satisfy the distortion constraint with equality, one can solve for q_2 in terms of q_1 to give

$$q_2 = \frac{\alpha D - q_1(\alpha - D)}{(c^2\alpha - D) + q_1(\sigma_Z^2 - D)}. \tag{A.5}$$

Substituting this in equation (A.4) gives the function to be minimized as a function of q_1 alone. The optimal choice of q_1 is then

$$q_1^* = \text{argmin } \frac{q_1^2(\sigma_Z^2 - D) + q_1 D(c^2 - 1) + \alpha D c^2}{-q_1^2(\alpha - D) + \alpha D q_1}. \tag{A.6}$$

Differentiating with respect to q_1 and setting the derivative to 0 gives us a quadratic in q_1 whose roots are

$$q_1^* = \frac{\alpha c}{\rho - c} \quad \text{or} \quad \frac{\alpha c D}{2\alpha c - (\rho + c)D} \tag{A.7}$$

The second root given above is where the minima occurs. The q_2 value corresponding to this value of q_1 is

$$q_2^* = \frac{\alpha D}{2\alpha c^2 - (1 + \rho c)D}. \tag{A.8}$$

Note that these optimal values of q_1 and q_2 are positive only for distortions in the range

$$D \leq \min\left\{ \frac{2\alpha c}{\rho + c}, \frac{2\alpha c^2}{1 + \rho c} \right\}. \tag{A.9}$$

For values of D outside this range, the optimal strategy is to let q_1 or q_2 go to ∞ which effectively means that we encode and transmit only one source.

For D in the range given in equation (A.9), the sum rate $R_{\text{sum}} = R_1 + R_2$ is found by substituting q_1^* and q_2^* in equation (A.1) to get

$$R_{\text{sum}} \geq \frac{1}{2} \log \frac{4c(\alpha c - \rho D)}{D^2} \quad D \leq \min\left\{ \frac{2\alpha c}{\rho + c}, \frac{2\alpha c^2}{1 + \rho c} \right\}. \tag{A.10}$$

For D outside the range given in equation (A.9), the minimum sum rate is attained by setting either q_1 or q_2 as ∞. Which quantity goes to ∞ depends on which argument of the min function in equation (A.9) is smaller; equivalently on whether $c > 1$ or not. It is easy to see that if $c < 1$, $q_2 = \infty$ and

$$R_{\text{sum}} = \frac{1}{2}\log\frac{(1-\rho c)^2}{D-\alpha c^2} \quad \text{for } D > \frac{2\alpha c^2}{1+\rho c}, \tag{A.11}$$

and if $c > 1$, $q_1 = \infty$ and

$$R_{\text{sum}} = \frac{1}{2}\log\frac{(c-\rho)^2}{D-\alpha} \quad \text{for } D > \frac{2\alpha c}{\rho+c}. \tag{A.12}$$

Combining equations (A.10), (A.11) and (A.12) and taking the convex closure of the resulting region, the complete rate region for the Berger-Tung based scheme can be found.

A.2 Existence of good nested lattices

We show the existence of a sequence of nested lattices $(\Lambda_1^{(n)}, \Lambda^{(n)})$ with $\Lambda^{(n)} \subset \Lambda_1^{(n)}$ such that both lattices are "good" for appropriately defined notions of goodness. The sequence is indexed by the lattice dimension n. The goodness notions used are Rogers-goodness (for source coding) and Poltyrev-goodness (for channel coding). These notions are defined precisely below. The existence of a sequence of lattices $\Lambda^{(n)}$ which are good in both senses has been shown earlier [49]. Also, the existence of nested lattices where the coarse lattice is good in both senses and the fine lattice is Poltyrev-good has also been shown [48]. We show that the same construction as used in [48] results in a fine lattice that in addition to being Poltyrev-good is also Rogers-good. Our proof is essentially identical to the one given in [49]. For a more complete version of the proof, we refer the reader to [95].

We describe the construction of the nested lattice first. We start with a coarse lattice Λ (the superscript is dropped from here on) which is both Rogers and Poltyrev-good. Let $\mathcal{V}$ be the Voronoi region of Λ and $\sigma^2(\mathcal{V})$ be the second moment per dimension of Λ [47]. Let the generator matrix of Λ be G_Λ, i.e., $\Lambda = G_\Lambda \cdot \mathbb{Z}^n$. Formally, Λ satisfies

- (Rogers-good) Let R_u and R_l be the covering and effective radius of the lattice Λ. Λ (more precisely, a sequence of such lattices) is called Rogers-good if its covering efficiency $\rho_{\text{cov}}(\Lambda) \to 1$.
- (Poltyrev-good) For any $\sigma^2 < \sigma^2(\mathcal{V})$, let $\mathbf{N}$ be a Gaussian random vector whose components are i.i.d $\mathcal{N}(0, \sigma^2)$. Then, Λ (more precisely, a sequence of such lattices) is called Poltyrev-good if

  $$Pr(\mathbf{N} \notin \mathcal{V}) < \exp\{-n[E_p(\mu) - o_n(1)]\} \tag{A.13}$$

  where $\mu = \sigma^2(\mathcal{V})/\sigma^2$ is the VNR (volume to noise ratio) of the lattice Λ relative to $\mathcal{N}(0, \sigma^2)$ and $E_p(\mu)$ is the Poltyrev exponent [49].

We now construct the fine lattice Λ_1 using Loeliger's type-A construction [55]. Let k, n, p be integers such that $k \leq n$ and p is prime. Their precise magnitudes are described later. Let G be a $k \times n$ generating matrix with its elements chosen uniformly from $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$. The construction of the fine lattice is now described by the following steps:

1. Define the discrete codebook $\mathcal{C} = \{x : x = y \cdot G \text{ for some } y \in \mathbb{Z}_p^k\}$
2. Lift $\mathcal{C}$ to $\mathbb{R}^n$ to form $\Lambda_1' = p^{-1}\mathcal{C} + \mathbb{Z}^n$, (c) $\Lambda_1 \triangleq G_\Lambda \cdot \Lambda_1'$ is the fine lattice.

Note that, by construction, $\Lambda \subset \Lambda_1$. We now show that a randomly chosen member from this ensemble of nested lattices is such that Λ_1 is both Rogers and Poltyrev-good. The fact that such random selection results in a fine lattice which is with high

probability Poltyrev-good has already been shown [48]. We now show that a similar selection results in Rogers-good fine lattices as well. By union bound then, we will have proved our claim.

To show Rogers-goodness, we show that a random fine lattice (with high probability) covers all the points inside the Voronoi region $\mathcal{V}$ of the coarse lattice with a covering efficiency that asymptotically reaches unity. We do this by first showing that almost every point in $\mathcal{V}$ is covered with high probability by a subset of the fine lattice points. We then show that increasing the number of points in the fine lattice decreases the number of uncovered points at a certain rate till no points remain uncovered. We then show that the covering efficiency of this construction asymptotically approaches unity.

Number the points of the fine lattice Λ_1 that lie inside $\mathcal{V}$. Let $\Lambda_1(i)$ be the ith such point for $i = 0, 1, \ldots, p^k - 1$. Since the whole space is tiled by regions congruent to $\mathcal{V}$, we restrict attention to only $\mathcal{V}$. Let A^* then denote $A \mod \mathcal{V}$ for any set A. It can be shown that (see [56]) the random ensemble described above satisfies the following properties:

1. $\Lambda_1(0) = \mathbf{0}$ deterministically
2. $\Lambda_1(i)$ is equally likely to be any of the points in $p^{-1}\Lambda \cap \mathcal{V}$
3. For any $i \neq j$, $(\Lambda_1(i) - \Lambda_1(j))^*$ is uniformly distributed over $p^{-1}\Lambda \cap \mathcal{V}$.

If we use the lattice points $\Lambda_1 \cap \mathcal{V}$ as codewords, then the effective rate of such a code would be $R = \frac{k}{n} \log p$. In what follows, we will be interested in keeping this code rate fixed as $n \to \infty$. Thus $p^k \to \infty$ as $n \to \infty$. We also remark that the following proof works for any $R > 0$.

Part I: Almost complete covering

Fix an $r > 0$ to be chosen later. Fix an arbitrary $x \in \mathcal{V}$. Let $S_1(x)$ be the

set of all points in $p^{-1}\Lambda \cap \mathcal{V}$ that are within a distance $(r-d)$ of x, i.e., $S_1(x) = (p^{-1}\Lambda \cap (x + (r-d)\mathcal{B}))^*$. Here, $\mathcal{B}$ denotes a ball of unit radius and d is the covering radius of Voronoi region of the lattice $p^{-1}\Lambda$. The probability that x is covered by the ith point of the fine lattice Λ_1 is given by

$$Pr(x \in (\Lambda_1(i) + (r-d)\mathcal{B})^*) = \frac{|S_1(x)|}{p^n} \tag{A.14}$$

It can be shown that the above probability can be lower bounded as (see [56] for details)

$$Pr(x \in (\Lambda_1(i) + (r-d)\mathcal{B})^*) \geq \frac{V_{\mathcal{B}}(r-2d)}{|\mathcal{V}|} \quad \text{for } i = 1, \ldots, p^k - 1 \tag{A.15}$$

Note that we exclude $i = 0$ from consideration since $\Lambda_1(0) = \mathbf{0}$ deterministically. Let η_i be the indicator random variable that indicates whether x is covered by $\Lambda_1(i)$ for $i = 1, \ldots, p^k - 1$. Let χ be the total number of points in $\Lambda_1 \cap \mathcal{V}$ that cover x. Then it can be shown that

$$\mathbb{E}(\chi) \geq c_n \frac{V_{\mathcal{B}}(r-2d)}{|\mathcal{V}_1|} = c_n \left(\frac{r-2d}{r_{\Lambda_1}}\right)^n, \qquad \mathrm{Var}(\chi) \leq \mathbb{E}(\chi) \tag{A.16}$$

where r_{Λ_1} is the effective radius of the Voronoi region $\mathcal{V}_1$ of the fine lattice Λ_1 and $c_n = 1 - e^{-nR} \to 1$. Let $\mu(\nu) \triangleq \mathbb{E}(\chi) - 2^{\nu}\sqrt{\mathbb{E}(\chi)}$. From Chebyshev's inequality, for any $\nu > 0$, we have $Pr(\chi < \mu(\nu)) \leq 4^{-\nu}$. If $\mu(\nu) > 1$, then $4^{-\nu}$ also bounds the probability that none of the points of $p^{-1}\Lambda$ cover x.

Call $x \in \mathcal{V}$ remote from a set A if none of the points in A are within distance $(r-d)$ from x. Then, $\chi(x) < 1$ is the same as saying x is remote from Λ_1. Let Q be the set of points $x \in \mathcal{V}$ that are remote from Λ_1 and let $q \triangleq |Q|/|\mathcal{V}|$. Then, $|Q| = \int_{\mathcal{V}} \mathbf{1}_{(\chi(x)<1)} dx \leq \int_{\mathcal{V}} \mathbf{1}_{(\chi(x)<\mu(\nu))} dx$ if $\mu(\nu) \geq 1$. Using the previously obtained bound, we then have $\mathbb{E}(q) \leq 4^{-\nu}$. From Markov's inequality, it then follows that $Pr(q > 2^{\nu}\mathbb{E}(q)) < 2^{-\nu}$ and thus

$$Pr(q > 2^{-\nu}) < 2^{-\nu} \tag{A.17}$$

If we let $\nu \to \infty$ while still keeping $\mu(\nu) \geq 1$, we can let this probability decay to 0. This can be achieved by letting $\nu = o(\log n)$ and $\mathbb{E}(\chi) > n^{\lambda}$ for some $\lambda > 0$. But, we have $\mathbb{E}(\chi) \geq (p^k - 1)V_{\mathcal{B}}(r - 2d)/|\mathcal{V}|$. Thus, it is enough to choose r such that

$$\log\left(\frac{r-2d}{r_{\Lambda_1}}\right) \geq \frac{\lambda}{n}\log n \tag{A.18}$$

With such a choice of parameters, for most lattices in the ensemble, almost all points of the region $\mathcal{V}$ are $(r-d)$ covered by points of the randomly chosen lattice Λ_1 with high probability. Note that, it suffices to choose $k = 1$ even to reach this conclusion (in which case, p needs to grow exponentially). In what follows, we will restrict attention to covering only the points of the grid $p^{-1}\Lambda \cap \mathcal{V}$. We note that the bound obtained in equation (A.17) holds when q is interpreted as the fraction of uncovered points in $p^{-1}\Lambda \cap \mathcal{V}$ as well.

Part II: Complete covering

We now extend the analysis to provide complete covering of $\mathcal{V}$. The main idea is as follows. Any point $x \in \mathcal{V}$ is within a distance d from a point in $p^{-1}\Lambda \cap \mathcal{V}$. This simply follows from the definition of d as the covering radius of $p^{-1}\Lambda$. Thus, an $(r-d)$ covering of the points of $p^{-1}\Lambda$ will automatically result in an r covering of $\mathcal{V}$. Thus, we restrict our attention to the lattice $p^{-1}\Lambda \cap \mathcal{V}$ and attempt to cover only these lattice points in what follows. Correspondingly, we define $Q(A)$ to be the set of all lattice points $p^{-1}\Lambda \cap \mathcal{V}$ that are remote from the set A. Also, let x_i, $i = 0, \ldots, p^n - 1$ denote the ith point of the constellation $p^{-1}\Lambda \cap \mathcal{V}$.

Let $\Lambda_1[k_1]$ be the fine lattice obtained using the Loeliger construction while using only the first k_1 rows of the random matrix G. We saw in the previous section that any such k_1 would suffice to get an almost complete covering of $\mathcal{V}$. We will now demonstrate that the fraction of uncovered points squares when we go from $\Lambda_1[k_1]$ to $\Lambda_1[k_1 + 1]$ and thus when sufficient number of rows are added, the fraction of

uncovered points becomes less than p^{-n} with high probability. Since there are only p^n points in $p^{-1}\Lambda$, this means that every point is covered.

Fix k_1 which grows faster than $(\log n)^2$. Let x_j be the jth lattice point. Again, we exclude $j = 0$ from consideration. Let Q_i be the set of lattice points that remain uncovered by the lattice $\Lambda_1[k_1 + i]$, $i = 0, 1, \ldots, k_2 = k - k_1$. Correspondingly, define $q_i = |Q_i|/p^n$. Consider the set $S = (\Lambda_1[k_1] \cup (\Lambda_1[k_1] + p^{-1}\mathbf{g}_{k_1+1}))^*$ where $\mathbf{g}_i$ is the ith row of the random matrix G. Note that $S \subset \Lambda_1[k_1 + 1]$. This implies that $Q(\Lambda_1[k_1+1]) \subset Q(S)$ and $q_1 \leq |Q(S)|/p^n$. Since $\Lambda_1[k_1]+p^{-1}\mathbf{g}_{k_1+1}$ is an independent shift of $\Lambda_1[k_1]$, the probability that x_j is remote from $\Lambda_1[k_1] + p^{-1}\mathbf{g}_{k_1+1}$ is the same as the probability that x_j is remote from $\Lambda_1[k_1]$. Also note that, given a $\Lambda_1[k_1]$, q_0 is a deterministic function of $\Lambda_1[k_1]$. These considerations give us the following.

$$\mathbb{E}\left(\frac{|Q(S)|}{p^n}\,\middle|\,\Lambda_1[k_1]\right) = \frac{q_0}{p^n}\sum_{j=1}^{p^n-1}\mathbf{1}(x_j \in Q(\Lambda_1[k_1]) \mid \Lambda_1[k_1]) = q_0^2 \tag{A.19}$$

where the last equality follows from the definition of q_0. Since q_0 is a deterministic function of $\Lambda_1[k_1]$, we have

$$\mathbb{E}\left(\frac{|Q(S)|}{p^n}\,\middle|\,q_0\right) = q_0^2 \tag{A.20}$$

This in turn implies that $\mathbb{E}(q_1 \mid q_0) \leq q_0^2$. Appealing to Markov inequality gives us (for any $\gamma > 0$)

$$Pr(q_1 > 2^{\gamma}\mathbb{E}(q_1 \mid q_0) \mid q_0) \leq 2^{-\gamma} \tag{A.21}$$

Combining this with the bound on $\mathbb{E}(q_1 \mid q_0)$, we get $Pr(q_1 \leq 2^{\gamma-2\nu} \mid q_0 \leq 2^{-\nu}) \geq 1 - 2^{-\gamma}$. By Bayes' rule, we finally arrive at $Pr(q_1 \leq 2^{\gamma-2\nu}) \geq (1 - 2^{-\gamma})(1 - 2^{-\nu})$. Iterating this procedure k_2 times gives us

$$Pr(q_{k_2} \leq 2^{2^{k_2}(\gamma-\nu)-\gamma}) \geq (1 - 2^{-\nu})(1 - 2^{-\gamma})^{k_2} \tag{A.22}$$

It can be verified that this probability can be made to go to 1 by choosing the following rates of growth for the different quantities: k grows as fast as $(\log n)^2$, k_2 grows at least as fast as $\lceil \log n + \log\log p \rceil$ and ν is chosen to be $\nu = 2\log(\log n + \log\log p)$ and $\gamma = \nu - 1$. From standard random coding arguments, it then follows that there exists a deterministic nested lattice (Λ, Λ_1) such that the lattice points Λ_1 r-cover $\mathbb{R}^n$ for the following choices of the parameters.

The covering efficiency of the fine lattice can now be bounded as

$$\frac{r}{r_{\Lambda_1}} = \sqrt[n]{\frac{V_{\mathcal{B}}(r)}{V_{\mathcal{B}}(r-2d)} n^{\lambda} p^{k_2}} \tag{A.23}$$

$$\leq \left(\frac{r}{r-2d}\right) \cdot n^{\frac{\lambda}{n}} \cdot 2^{(\log p \log n + \log p \log\log p + \log p)/n} \tag{A.24}$$

As $n \to \infty$, the right hand side should go to 1. It is easy to verify that the last 2 terms do indeed tend to 1. To show that the first term goes to 1, we need to show that $d \to 0$ as $n \to \infty$ for our choice of parameters. Since Λ is Rogers-good (which implies $p^{-1}\Lambda$ is Rogers-good as well), it has a covering efficiency asymptotically approaching 1. Thus the covering radius d of $p^{-1}\Lambda$ approaches $p^{-1}r_{\Lambda}$ as the lattice dimension $n \to \infty$. From the nesting ratio, we get

$$\frac{|\mathcal{V}|}{|\mathcal{V}_1|} = \left(\frac{r_{\Lambda}}{r_{\Lambda_1}}\right)^n = p^k = 2^{nR} \tag{A.25}$$

and hence d approaches $p^{-1}2^R r_{\Lambda_1}$. We know that (since k grows as $\log n + \log\log p$ and $p^k = 2^{nR}$) p grows as $o(n/\log n)$ and thus to ensure $d \to 0$, we need r_{Λ_1} to go to ∞ slower than p. Once could even take r_{Λ_1} to be constant in the above proof. Thus, we have shown that Λ_1 is an efficient covering lattice.

A lattice that is good for covering is necessarily good for quantization. This can be inferred from the following relation. For any lattice Λ

$$G(\Lambda) \leq G_n^* \cdot \frac{n+2}{n} \cdot (\rho_{\text{cov}}(\Lambda))^2 \tag{A.26}$$

where $G(\Lambda)$ is the normalized second moment of the lattice Λ, G_n^* is the normalized second moment of the n-dimensional sphere and $\rho_{\text{cov}}(\Lambda)$ is the covering efficiency of Λ. Since, we have shown that $\rho_{\text{cov}}(\Lambda_1) \to 1$ as $n \to \infty$ with high probability, it also follows that the fine lattice is good for MSE quantization with high probability.

Thus, we have proved the existence of nested lattices (Λ_1, Λ), $\Lambda \subset \Lambda_1$, such that both lattices both Rogers and Poltyrev-good. By iterating this construction process, we can show the existence of good nested lattices with any finite level of nesting. More precisely, for any finite $m > 0$, one can show the existence of a nested lattice $(\Lambda_1, \Lambda_2, \ldots, \Lambda_m)$, $\Lambda_m \subset \cdots \subset \Lambda_1$ such that all the lattices $\Lambda_i, i = 1, \ldots, m$ are both Rogers-good and Poltyrev-good. Further, such nested lattices exist for any choice of the nesting ratios. By virtue of being Rogers-good, such lattices are also good for MSE quantization.

A.3 Proof of convergence to Gaussianity of e_q

In this section, we prove the claim that $e_q = e_{q_1} - e_{q_2}$ tends to a white Gaussian noise in the Kullback-Leibler divergence sense where $e_{q_i}, i = 1, 2$ are two independent subtractive dither quantization noises. Note that the lattices $\Lambda_{1i}, i = 1, 2$ associated with $e_{q_i}, i = 1, 2$ are good source codes.

We use the following properties of subtractive dither quantization noise and the associated optimal lattice quantizers [44].

- The subtractive dither quantization noise e_{q_i} is uniformly distributed over the basic Voronoi region $\mathcal{V}_{0,1i}$ of the fine lattice Λ_{1i} for $i = 1, 2$. It follows from equation (2.5) that

$$\mathbb{E} \parallel e_{q_i} \parallel^2 = n\sigma^2(\Lambda_{1i}) \quad \text{for } i = 1, 2. \tag{A.27}$$

- For optimal lattice quantizers, the components of $e_{q_i}, i = 1, 2$ are uncorrelated and have the same power,i.e., their correlation matrices $\Sigma_{e_{q_i}}$ can be written as

$$\Sigma_{e_{q_i}} = \sigma^2(\Lambda_{1i})\mathbf{I}_{n\times n} \quad \text{for } i = 1, 2. \tag{A.28}$$

- For optimal lattice quantizers, as the lattice dimension $n \to \infty$, the distribution of $e_{q_i}, i = 1, 2$ tends to a white Gaussian vector of same covariance in the Kullback-Leibler divergence sense. Taking into account equation (A.27), this can be written as

$$\frac{1}{n} D\left(e_{q_i} \parallel \mathcal{N}(0, \sigma^2(\Lambda_{1i})\mathbf{I}_{n\times n})\right) \to 0 \quad \text{for } i = 1, 2 \tag{A.29}$$

in terms of the Kullback-Leibler divergence $D(. \parallel .)$ or equivalently,

$$h(e_{q_i}) \to \frac{n}{2} \log 2\pi e\sigma^2(\Lambda_{1i}) \quad \text{for } i = 1, 2 \tag{A.30}$$

in terms of differential entropy $h(\cdot)$.

To show the convergence of e_q to a white Gaussian random vector, we use the entropy power inequality and the fact that for a given covariance matrix, the Gaussian distribution maximizes differential entropy.

The entropy power inequality [15] states that for two independent n-dimensional random vectors X and Y (having densities),

$$2^{\frac{2}{n}h(X+Y)} \geq 2^{\frac{2}{n}h(X)} + 2^{\frac{2}{n}h(Y)}. \tag{A.31}$$

This inequality applied to the subtractive dither quantization noises gives

$$2^{\frac{2}{n}h(e_{q_1}-e_{q_2})} \geq 2^{\frac{2}{n}h(e_{q_1})} + 2^{\frac{2}{n}h(e_{q_2})}. \tag{A.32}$$

As $n \to \infty$, by equation (A.30), the right hand side of equation (A.32) tends to $2\pi e(\sigma^2(\Lambda_{11}) + \sigma^2(\Lambda_{12}))$. So, we have the following lower bound on the limit of the

differential entropy of e_q.

$$\lim_{n\to\infty} h(e_q) \geq \frac{n}{2}\log 2\pi e(\sigma^2(\Lambda_{11}) + \sigma^2(\Lambda_{12})). \tag{A.33}$$

To prove the inequality in the other direction, note that equation (A.28) implies that the covariance matrix of e_q is $(\sigma^2(\Lambda_{11}) + \sigma^2(\Lambda_{12}))\mathbf{I}_{n\times n}$. Since the Gaussian distribution maximizes differential entropy for a given covariance matrix, we have

$$h(e_q) \leq \frac{n}{2}\log 2\pi e(\sigma^2(\Lambda_{11}) + \sigma^2(\Lambda_{12})) \tag{A.34}$$

Combining equations (A.33) and (A.34), we have the desired result that (if optimal lattice quantizers are used)

$$\lim_{n\to\infty} h(e_q) = \frac{n}{2}\log 2\pi e(\sigma^2(\Lambda_{11}) + \sigma^2(\Lambda_{12})). \tag{A.35}$$

In words, e_q tends in the Kullback-Leibler divergence sense to a white Gaussian random vector with covariance matrix $(\sigma^2(\Lambda_{11}) + \sigma^2(\Lambda_{12}))\mathbf{I}_{n\times n}$.

A.4 Derivation of optimal Lattice parameters

In the coding schemes of both Section 2.2 and Section 2.3, we scale the sources before encoding them. Here, we briefly outline a justification for the specific scaling constants used. We restrict ourselves to the case where all the K users encode their sources using lattice binning. In the notation of Section 2.3.2, this corresponds to $\Theta = \{1, \ldots, K\}$.

Let the function to be reconstructed be $Z = \sum_{i=1}^{K} c_i X_i = cX^n$. Here c is a row vector with its ith component as c_i and X^n is a column vector of the sources X_i. Σ is the covariance matrix of the random vector X^n. Let the ith encoder scale its input by an arbitrary constant η_i. Let $\eta \triangleq [\eta_1, \ldots, \eta_K]$. Choose a tuple $\mathcal{Q} = (q_1, \ldots, q_K) \in \mathbb{R}_+^K$ just as in Section 2.3.4.

It can be shown from analysis similar to the ones in Section 2.2.2 and 2.3.2 that the decoder can, with high probability, reconstruct the function $\eta X^n + Q$ where Q approaches a white Gaussian noise of variance $q = \sum_{i=1}^{K} q_i$. From equation (2.67), it follows that the function f used for decoding is

$$\hat{Z} = \left(\frac{c\Sigma\eta^T}{\eta\Sigma\eta^T + q} \right) (\eta X^n + Q) \tag{A.36}$$

and the corresponding distortion is

$$D = \sigma_Z^2 - \frac{(c\Sigma\eta^T)^2}{\eta\Sigma\eta^T + q}. \tag{A.37}$$

This fixes the value of q. The second moment of the channel code used is $\sigma^2(\Lambda_2) = \mathrm{Var}(\sum_i \eta_i X_i + q_i) = \eta\Sigma\eta^T + q$. This gives us the rate tuple

$$R_i = \frac{1}{2} \log \frac{\eta\Sigma\eta^T + q}{q_i} \quad \text{for } i = 1, \ldots, K \tag{A.38}$$

Eliminating q_i using $q = \sum_i q_i$ gives us the rate region

$$\sum_{i=1}^{K} 2^{-2R_i} \leq 1 - (\sigma_Z^2 - D) \frac{\eta\Sigma\eta^T}{(c\Sigma\eta^T)^2}. \tag{A.39}$$

This rate region is largest when the RHS is maximum. Maximizing the RHS as a function of η results in $\eta = \xi \cdot c$ as the only solutions for any value of the constant ξ. However, all constants ξ result in the same rate region.

A.5 Proof of Error Probability

We outline a proof that the probability of error of the lattice based coding scheme indeed approaches zero for an optimal choice of lattices as the lattice dimension $n \to \infty$. Recall that the probability of error is given by

$$\begin{aligned} P_e &= Pr((Z^n + e_q) \bmod \Lambda_2 \neq (Z^n + e_q)) \\ &= Pr((Z^n + e_q) \notin \mathcal{V}_2) \end{aligned} \tag{A.40}$$

where $e_q = e_{q_1} - e_{q_2}$ and e_{q_i} are the subtractive dither quantization noises which are uniformly distributed over the respective Voronoi regions $\mathcal{V}_{1i}$ for $i = 1, 2$. This notion of decoding and decoding error probability is closely related to the notion of decoding in the presence of "self-noise" for the AWGN channel described in [47, 48].

We proceed as follows. We first demonstrate how e_{q_i} can be well-approximated by a Gaussian random variable N_i. This approximation becomes progressively more exact as the lattice dimension increases. Thus, $Z^n + e_q$ can be well approximated by W^n where $W \triangleq Z + N_1 - N_2$. It will then be shown that $\sigma_W^2 \leq \sigma^2(\Lambda_2) + \epsilon$ where $\epsilon \to 0$ as $n \to \infty$. Since Λ_2 is a good channel $\sigma^2(\Lambda_2)$-code, it will then follow that the probability of error goes to zero exponentially with the exponent given by the Poltyrev bound.

Lemma A.1. *Suppose the fine lattices Λ_{1i} are both Rogers-good with effective and covering radius R_{li} and R_{ui} respectively. The subtractive dithered quantization noise e_{q_i} can be well-approximated by a Gaussian noise $N_i \sim \mathcal{N}(0, \sigma_i^2 \mathbf{I}^n)$ in the sense that for any $\epsilon > 0$, we have for sufficiently large n*

$$\frac{1}{n} \log \frac{f_{e_{q_i}}(\mathbf{x})}{f_{N_i}(\mathbf{x})} \leq \epsilon \qquad \forall\, \mathbf{x} \in \mathcal{V}_{1i},\ i = 1, 2 \tag{A.41}$$

The variance σ_i^2 is related to the second moment of the corresponding lattices through the inequalities

$$\frac{n}{n+2} \sigma^2(\Lambda_{1i}) \leq \sigma_i^2 \leq \left(\frac{R_{ui}}{R_{li}} \right)^2 \sigma^2(\Lambda_{1i}) \qquad i = 1, 2 \tag{A.42}$$

Proof: We prove the lemma for the dithered quantization noise e_{q_1}. For notational convenience, the subscript is omitted from the quantities R_{u1}, R_{l1} and σ_1^2. It is known that $e_{q_1} \sim \text{Unif}(\mathcal{V}_{11})$. The approximation proceeds in two stages - (1) a random vector uniformly distributed over the Voronoi region is approximated by a random vector

uniformly distributed over a sphere in n-dimensions, (2) the random vector uniformly distributed over the sphere is approximated by a Gaussian random vector.

Denote by $\mathcal{B}(R_u)$ a ball of radius R_u and let σ^2 be the second moment per dimension of $\mathcal{B}(R_u)$. It can be shown (see [44] for instance) that

$$\sigma^2 = \frac{R_u^2}{n+2} \tag{A.43}$$

Since σ^2 is the second moment of a ball containing $\mathcal{V}_{11}$, it follows that $\sigma^2(\Lambda_{11}) < \sigma^2$. Let $\mathbf{B} \sim \text{Unif}(\mathcal{B}(R_u))$ and let R_l be such that $\text{Vol}(\mathcal{B}(R_l)) = \text{Vol}(\mathcal{V}_{11})$. Since a ball has the smallest normalized second moment of all shapes of a given volume, we have

$$\frac{1}{n}\mathbb{E} \parallel e_{q_1} \parallel^2 \stackrel{(a)}{\geq} \frac{1}{n}\mathbb{E} \parallel \mathbf{B} \cdot \frac{R_l}{R_u} \parallel^2 \tag{A.44}$$

$$= \left(\frac{R_l}{R_u}\right)^2 \sigma^2 \tag{A.45}$$

where (a) follows from the fact that $\mathbf{B} \cdot \frac{R_l}{R_u} \sim \text{Unif}(\mathcal{B}(R_l))$. From the above inequality, it follows that

$$\sigma^2 \leq \left(\frac{R_u}{R_l}\right)^2 \frac{1}{n}\mathbb{E} \parallel e_{q_1} \parallel^2 \tag{A.46}$$

$$= \left(\frac{R_u}{R_l}\right)^2 \sigma^2(\Lambda_{11}). \tag{A.47}$$

On the other hand, we have

$$\frac{1}{n}\mathbb{E} \parallel e_{q_1} \parallel^2 \leq \frac{1}{n}R_u^2 = \frac{n+2}{n}\sigma^2. \tag{A.48}$$

Putting these inequalities together, we get

$$\frac{n}{n+2}\sigma^2(\Lambda_{11}) \leq \sigma^2 \leq \left(\frac{R_u}{R_l}\right)^2 \sigma^2(\Lambda_{11}) \tag{A.49}$$

We now turn our attention to the density function of the vector $\mathbf{B}$. Since $\mathbf{B}$ and e_{q_1} are uniformly distributed over $\mathcal{B}(R_u)$ and $\mathcal{V}_{11}$ respectively, we have

$$\frac{f_{e_{q_1}}(\mathbf{x})}{f_{\mathbf{B}}(\mathbf{x})} = \frac{\text{Vol}(\mathcal{B}(R_u))}{\text{Vol}(\mathcal{V}_{11})} = \left(\frac{R_u}{R_l}\right)^n \qquad \forall\, \mathbf{x} \in \mathcal{V}_{11} \tag{A.50}$$

Let N_1 be a n-dimensional Gaussian random variable with independent components of variance σ^2. We approximate the density function of $\mathbf{B}$ with that of N_1. It is well known that

$$-\frac{1}{n}\log f_{N_1}(\mathbf{x}) = \frac{1}{2}\log 2\pi\sigma^2 + \frac{\| \mathbf{x} \|^2}{2n\sigma^2} \tag{A.51}$$

It is also easy to see that, for $\| \mathbf{x} \| \leq R_u$,

$$-\frac{1}{n}\log f_{\mathbf{B}}(\mathbf{x}) = \frac{1}{n}\log \mathrm{Vol}(\mathcal{B}(R_u)) \tag{A.52}$$

$$= \frac{1}{2}\log\frac{\sigma^2}{G_n^*} \tag{A.53}$$

$$= \frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2}\log 2\pi G_n^* \tag{A.54}$$

where G_n^* is the normalized second moment of the n-dimensional sphere. Subtracting the expressions, we get for $\| x \| \leq R_u$

$$\frac{1}{n}\log\frac{f_{\mathbf{B}}(\mathbf{x})}{f_{N_1}(\mathbf{x})} = \frac{1}{2}\log 2\pi G_n^* + \frac{\| x \|^2}{2n\sigma^2} \tag{A.55}$$

$$\overset{(a)}{\leq} \frac{1}{2}\log 2\pi G_n^* + \frac{R_u^2}{2n\sigma^2} \tag{A.56}$$

$$= \frac{1}{2}\log 2\pi G_n^* + \frac{n+2}{2n} \tag{A.57}$$

$$= \frac{1}{2}\log 2\pi e G_n^* + \frac{1}{n} \tag{A.58}$$

where (a) follows from the monotonically decreasing nature of the Gaussian density function. Combining this with equation (A.50), we get

$$\frac{1}{n}\log\frac{f_{e_{q_1}}(\mathbf{x})}{f_{N_1}(\mathbf{x})} \leq \frac{1}{2}\log 2\pi e G_n^* + \frac{1}{n} + \log\frac{R_u}{R_l} \qquad \forall\, \mathbf{x} \in \mathcal{V}_{11} \tag{A.59}$$

The right hand side of the inequality can be made arbitrarily small as $n \to \infty$ if the lattice Λ_{11} is Rogers-good. This proves the claim of the lemma. □

With a slight abuse of notation, denote by N_i the Gaussian random variable $N_i \sim \mathcal{N}(0, \sigma_i^2)$. Define the Gaussian random variable $W \triangleq Z + N_1 - N_2$. Then,

$\sigma^2(W) = \sigma^2(Z) + \sigma_1^2 + \sigma_2^2$ and this can be bounded as

$$\left(\frac{n}{n+2}\right)\frac{\sigma_Z^4}{\sigma_Z^2 - D} \leq \sigma_W^2 \leq \eta^2 \frac{\sigma_Z^4}{\sigma_Z^2 - D} \tag{A.60}$$

where $\eta \triangleq \max\{\frac{R_{u1}}{R_{l1}}, \frac{R_{u2}}{R_{l2}}\}$. Since both $\Lambda_{11}, \Lambda_{12}$ are Rogers-good, it follows that $\eta \searrow 1$ as $n \to \infty$. Let us choose $\sigma^2(\Lambda_2) = \frac{\sigma_Z^4}{\sigma_Z^2 - D}(1+\delta)$ for some fixed $\delta > 0$. It then follows that for sufficiently large n, we shall have $\sigma_W^2 < \sigma^2(\Lambda_2)$ for any $\delta > 0$.

Let us now bound the probability that $Z^n + e_q$ falls outside the Voronoi region $\mathcal{V}_2$. To this end, define the quantity

$$\varepsilon_1(\Lambda) \triangleq \log\left(\frac{R_u(\Lambda)}{R_l(\Lambda)}\right) + \frac{1}{2}\log 2\pi e G_n^* + \frac{1}{n} \tag{A.61}$$

associated with a lattice Λ. Recall that if Λ is Rogers-good, $\varepsilon_1(\Lambda) \to 0$ as $n \to \infty$. It is then clear that

$$f_{e_q(\mathbf{x})} \leq e^{n(\varepsilon_1(\Lambda_{11}) + \varepsilon_1(\Lambda_{12}))} f_{N_1 - N_2}(\mathbf{x}) \tag{A.62}$$

Therefore, we have

$$P_e = Pr(Z^n + e_q \notin \mathcal{V}_2) \tag{A.63}$$

$$\leq e^{n(\varepsilon_1(\Lambda_{11}) + \varepsilon_1(\Lambda_{12}))} Pr(W^n \notin \mathcal{V}_2) \tag{A.64}$$

$$\leq e^{-n(E_P(\mu) - (\varepsilon_1(\Lambda_{11}) + \varepsilon_1(\Lambda_{12})))} \tag{A.65}$$

where $\mu = \frac{\sigma^2(\Lambda_2)}{\sigma_W^2} \geq \frac{1+\delta}{\eta^2}$ and $E_P(\cdot)$ is the Poltyrev error exponent. To show that the error indeed decays to zero, fix $0 < \delta < 1$ and let the dimension n be sufficiently large that $\eta^2 \leq 1 + \frac{\delta}{2}$. Then, we have $\mu \geq 1 + \frac{\delta}{3}$ and it can be shown that $E_P(\mu) \geq \frac{\delta^2}{108}$. As has already been mentioned, since Λ_{11} and Λ_{12} are Rogers-good, $(\varepsilon_1(\Lambda_{11}) + \varepsilon_1(\Lambda_{12}))$ can be made arbitrarily small as the dimension $n \to \infty$. Thus, the probability of decoding error in equation (2.44) goes to 0 exponentially.

Since δ can be chosen to be arbitrarily small, it follows that all rate points (R_1, R_2, D) that satisfy

$$2^{-2R_1} + 2^{-2R_2} \leq \left(\frac{\sigma_Z^2}{D}\right)^{-1} \tag{A.66}$$

are achievable.

APPENDIX B

Proofs for Chapter 3

B.1 Good Group Channel Codes

We prove the existence of channel codes built over the space $\mathbb{Z}_{p^r}^n$ which are good for the triple $(\mathcal{Z}, \mathcal{S}, P_{ZS})$ according to Definition 3.10. Recall that the group $\mathbb{Z}_{p^r}$ has $(r-1)$ non-trivial subgroups, namely $p^i\mathbb{Z}_{p^r}, 1 \leq i \leq r-1$. Let the random variable Z take values from the group $\mathbb{Z}_{p^r}$, i.e., $\mathcal{Z} = \mathbb{Z}_{p^r}$ and further let it be non-redundant. Let $\mathrm{Hom}(\mathbb{Z}_{p^r}^n, \mathbb{Z}_{p^r}^k)$ be the set of all homomorphisms from $\mathbb{Z}_{p^r}^n$ to $\mathbb{Z}_{p^r}^k$. Let $\phi(\cdot)$ be a homomorphism picked at random with uniform probability from $\mathrm{Hom}(\mathbb{Z}_{p^r}^n, \mathbb{Z}_{p^r}^k)$.

We start by proving a couple of lemmas.

Lemma B.1. *For a homomorphism $\phi(\cdot)$ randomly chosen from Hom$(\mathbb{Z}_{p^r}^n, \mathbb{Z}_{p^r}^k)$, the probability that a given sequence z^n belongs to $\ker(\phi)$ in $\mathbb{Z}_{p^r}^n$ depends on which subgroup of $\mathbb{Z}_{p^r}$ the sequence z^n belongs to. Specifically*

$$P(\phi(z^n) = 0^k) = \begin{cases} p^{-(r-i)k} & \text{if } z^n \in p^i\mathbb{Z}_{p^r}^n \setminus p^{i+1}\mathbb{Z}_{p^r}^n, 0 \leq i < r \\ 1 & \text{if } z^n \in p^r\mathbb{Z}_{p^r}^n \end{cases} \tag{B.1}$$

Proof: Clearly, $z^n \in p^r\mathbb{Z}_{p^r}^n$ implies $z^n = 0^n$ [1]. In this case, the probability of the event $\{\phi(z^n) = 0\}$ is 1.

[1] If we consider homomorphisms from $\mathbb{Z}_{p^r}^n$ to $\mathbb{Z}_m^k$ for an arbitrary integer m, all such homomorphisms have $d\mathbb{Z}_{p^r}^n$ as their kernel where $d = (p^r, m)$ is the greatest common divisor of p^r and m. Unless $d = p^r$,

Let the $k \times n$ matrix Φ be the matrix representation of the homomorphism $\phi(\cdot)$. Let the first row of Φ be $(\alpha_1, \ldots, \alpha_n)$. Consider $\phi_1 \colon \mathbb{Z}_{p^r}^n \to \mathbb{Z}_{p^r}$, the homomorphism corresponding to the first row of Φ. The total number of possibilities for $\phi_1(\cdot)$ is $(p^r)^n$.

Let us consider the case where $z^n \in \mathbb{Z}_{p^r}^n \backslash p\mathbb{Z}_{p^r}^n$. In this case, z^n contains at least one element, say z_i which is invertible in $\mathbb{Z}_{p^r}$. Let us count the number of homomorphisms $\phi(\cdot)$ that map such a sequence z^n to a given $c \in \mathbb{Z}_{p^r}^k$. We need to choose the k homomorphisms $\phi_i(\cdot), 1 \leq i \leq k$ such that $\phi_i(z^n) = c_i$ for $1 \leq i \leq k$. Let us count the number of homomorphisms $\phi_1(\cdot)$ that map z^n to c_1. In this case, we can choose $\alpha_j, j \neq i$ to be arbitrary and fix α_i as

$$\alpha_i = z_i^{-1} \left(c_1 - \sum_{\substack{j=1 \\ j \neq i}}^{n} \alpha_j z_j \right) \tag{B.2}$$

Thus the number of favorable homomorphisms $\phi_1(\cdot)$ is $(p^r)^{(n-1)}$. Thus the probability that a randomly chosen homomorphism $\phi_1(\cdot)$ maps z^n to c_1 is p^{-r}. Since each of the k homomorphisms ϕ_i can be chosen independently, we have

$$P(\phi(z^n) = c) = p^{-rk} \quad \text{if } z^n \in \mathbb{Z}_{p^r}^n \backslash p\mathbb{Z}_{p^r}^n \tag{B.3}$$

Putting $c = 0^k$ in equation (B.3), we see that the claim in Lemma B.1 is valid for $z^n \in \mathbb{Z}_{p^r}^n \backslash p\mathbb{Z}_{p^r}^n$. Now, consider $z^n \in p^i\mathbb{Z}_{p^r}^n \backslash p^{i+1}\mathbb{Z}_{p^r}^n$ for a general $0 < i < r$. Any such z^n can be written as $p^i \tilde{z}^n$ for $\tilde{z}^n \in \mathbb{Z}_{p^r}^n \backslash p\mathbb{Z}_{p^r}^n$. Thus, the event $\{\phi(z^n) = 0\}$ will be

---

there would be exponentially many z^n for which $P(\phi(z^n) = 0) = 1$ for all $\phi \in \mathrm{Hom}(\mathbb{Z}_{p^r}^n, \mathbb{Z}_m^k)$ and this results in bad channel codes (see equation (B.38)). Thus, p^r has to divide m and all such m give identical performances as $m = p^r$.

true if and only if $\{\phi(\tilde{z}^n) = t\}$ for some $t \in p^{r-i}\mathbb{Z}_{p^r}^k$. Hence,

$$P(\phi(z^n) = 0) = P\left(\bigcup_{t \in p^{r-i}\mathbb{Z}_{p^r}^k} (\phi(\tilde{z}^n) = t)\right) \tag{B.4}$$

$$= \sum_{t \in p^{r-i}\mathbb{Z}_{p^r}^k} P(\phi(\tilde{z}^n) = t) \tag{B.5}$$

$$= |p^{r-i}\mathbb{Z}_{p^r}^k| p^{-rk} \tag{B.6}$$

$$= p^{-(r-i)k} \tag{B.7}$$

This proves the claim of Lemma B.1. □

We now estimate the size of the intersection of the conditionally typical set $A_\epsilon^n(s^n)$ with cosets of $p^i\mathbb{Z}_{p^r}^n$ in $\mathbb{Z}_{p^r}^n$.

Lemma B.2. *For a given $z^n \in A_\epsilon^n(s^n)$, consider $(z^n + p^i\mathbb{Z}_{p^r}^n)$, the coset of $p^i\mathbb{Z}_{p^r}^n$ in $\mathbb{Z}_{p^r}^n$. Define the set $S_{i,\epsilon}(z^n, s^n)$ as $S_{i,\epsilon}(z^n, s^n) \triangleq (z^n + p^i\mathbb{Z}_{p^r}^n) \cap A_\epsilon^n(s^n)$. A uniform bound on the cardinality of this set is given by*

$$\frac{1}{n}\log|S_{i,\epsilon}| \leq H(Z|S) - H([Z]_i|S) + \delta(\epsilon) \qquad \text{for } 0 \leq i \leq r \tag{B.8}$$

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. The random variable $[Z]_i$ is defined in the following manner: It takes values from the set of all distinct cosets of $p^i\mathbb{Z}_{p^r}$ in $\mathbb{Z}_{p^r}$. The probability that $[Z]_i$ takes a particular coset as its value is equal to the sum of the probabilities of the elements forming that coset.

$$P([Z]_i = a + p^i\mathbb{Z}_{p^r} \mid S = s) = \sum_{z \in a + p^i\mathbb{Z}_{p^r}} P_{Z|S}(z \mid s) \quad \forall s \in \mathcal{S}. \tag{B.9}$$

We have the nesting relation $S_{i+1,\epsilon}(z^n, s^n) \subset S_{i,\epsilon}(z^n, s^n)$ for $0 \leq i \leq r-1$. However, each nested set is exponentially smaller in size since $H([Z]_i)$ increases monotonically with i. Thus, with the same definitions as above, we also have that

(B.10)

$$\frac{1}{n}\log\left(|S_{i,\epsilon}(z^n, s^n)| - |S_{i+1,\epsilon}(z^n, s^n)|\right) \leq H(Z|S) - H([Z]_i|S) + \delta_1(\epsilon) \qquad \text{for } 0 \leq i \leq r-1$$

where $\delta_1(\epsilon) \to 0$ *as* $\epsilon \to 0$.

Proof: The set $S_{i,\epsilon}(z^n, s^n)$ can be thought of as all those sequences $\tilde{z}^n \in A_\epsilon^n(s^n)$ such that the difference $w^n \triangleq \tilde{z}^n - z^n \in p^i\mathbb{Z}_{p^r}$. Let W be a random variable taking values in $p^i\mathbb{Z}_{p^r}$ and jointly distributed with (Z, S) according to $P_{W|ZS}$. Define the random variable $\tilde{Z} \triangleq Z + W$. Let $P_{W|ZS}$ be such that $P_{\tilde{Z}S} = P_{ZS}$. Then, for a given distribution $P_{W|ZS}$, every sequence $\tilde{z}^n$ that belongs to the set of conditionally typical sequences given (z^n, s^n) will belong to the set $S_{i,\epsilon}(z^n, s^n)$. Conversely, following the type counting lemma and the continuity of entropy as a function of probability distributions [14], every sequence $\tilde{z}^n \in S_{i,\epsilon}(z^n, s^n)$ belongs to the set of conditionally typical sequences given (z^n, s^n) for some such joint distribution $P_{W|ZS}$. Thus estimating the size of the set $S_{i,\epsilon}(z^n, s^n)$ reduces to estimating the maximum of $H(\tilde{Z} \mid Z, S)$, or equivalently the maximum of $H(Z, W \mid S)$ over all joint distributions P_{ZSW} such that $P_{(Z+W),S} = P_{ZS}$.

We formulate this problem as a convex optimization problem in the following manner. Recall that the alphabet of Z is the group $\mathbb{Z}_{p^r}$. Hence, $H(Z, W \mid S)$ is a concave function of the $|\mathcal{Z}||\mathcal{S}||p^i\mathbb{Z}_{p^r}|$ variables $P_{ZSW}(Z = z, S = s, W = w), z \in \mathcal{Z}, S \in \mathcal{S}, w \in p^i\mathbb{Z}_{p^r}$ and maximizing this conditional entropy is a convex minimization problem. Since the distribution P_{ZS} is fixed, these variables satisfy the marginal constraint

$$\sum_{w \in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(Z = z, W = w \mid S = s) = P_{Z|S}(z \mid s) \quad \forall z \in \mathcal{Z}, s \in \mathcal{S} \tag{B.11}$$

The other constraint to be satisfied is that the random variable $\tilde{Z} = Z + W$ is jointly distributed with S in the same way as Z, i.e., $P_{\tilde{Z}S} = P_{ZS}$. This can be expressed as

$$\sum_{w \in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(Z = z - w, W = w \mid S = s) = P_{Z|S}(z \mid s) \quad \forall z \in \mathcal{Z}, s \in \mathcal{S}. \tag{B.12}$$

Thus the convex optimization problem can be stated as

$$\begin{aligned}
&\text{minimize } -H(Z,W \mid S) \\
&\text{subject to } \sum_{w\in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(z,w \mid s) = P_{Z|S}(z \mid s) \quad \forall z \in \mathcal{Z}, s \in \mathcal{S}, \\
&\sum_{w\in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(z-w,w \mid s) = P_{Z|S}(z \mid s) \quad \forall z \in \mathcal{Z}, s \in \mathcal{S}. \qquad \text{(B.13)}
\end{aligned}$$

Note that the objective function to be minimized is convex and the constraints of equations (B.11) and (B.12) on $P_{ZW|S}(Z=z,W=w \mid S=s)$ are affine. Thus, the Karush-Kuhn-Tucker (KKT) conditions [100] are necessary and sufficient for the points to be primal and dual optimal. We now derive the KKT conditions for this problem. We formulate the dual problem as

$$\begin{aligned}
D(P_{ZW|S}) = &-\sum_{s\in\mathcal{S}} P_S(s) \left(\sum_{z\in\mathcal{Z},w\in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(z,w \mid s) \log \frac{1}{P_{ZW|S}(z,w \mid s)} \right) \\
&+ \sum_{z\in\mathcal{Z},s\in\mathcal{S}} \lambda_{z,s} \left(\sum_{w\in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(z-w,w \mid s) - P_{Z|S}(z \mid s) \right) \\
&+ \sum_{z\in\mathcal{Z},s\in\mathcal{S}} \gamma_{z,s} \left(\sum_{w\in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(z,w \mid s) - P_{Z|S}(z \mid s) \right) \qquad \text{(B.14)}
\end{aligned}$$

where $\{\lambda_{z,s}\}, \{\gamma_{z,s}\}$ are the Lagrange multipliers. Differentiating with respect to $P_{ZW|S}(Z=z,W=w \mid S=s)$ and setting the derivative to 0, we get

$$\frac{\partial D(P_{ZW|S})}{\partial P_{ZW|S}(z,w \mid s)} = P_S(s)(1+\log P_{ZW|S}(z,w \mid s)) + \lambda_{(z+w),s} + \gamma_{z,s} = 0 \qquad \text{(B.15)}$$

$$\implies \lambda_{(z+w),s} + \gamma_{z,s} = -P_S(s)(1+\log P_{ZW|S}(z,w \mid s)) \quad \forall z \in \mathcal{Z}, s \in \mathcal{S}, w \in p^i\mathbb{Z}_{p^r}. \qquad \text{(B.16)}$$

Summing over all $z \in p^i\mathbb{Z}_{p^r}$ for a given $s \in \mathcal{S}$, we see that for all $w \in p^i\mathbb{Z}_{p^r}$, the

summation $\sum_{z\in p^i\mathbb{Z}_{p^r}} \left(\lambda_{(z+w),s} + \gamma_{z,s}\right)$ is the same. This implies that

$$\prod_{z\in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(z, w \mid s) = \text{ constant} \quad \forall w \in p^i\mathbb{Z}_{p^r},\ \forall s \in \mathcal{S}. \tag{B.17}$$

These $|\mathcal{S}|p^{r-i}$ equations form the KKT equations and any solution that satisfies equations (B.11), (B.12) and (B.17) is the optimal solution to the optimization problem (B.13). We claim that the solution to this system of equations is given by

$$P_{ZW|S}(z, w \mid s) = \frac{P_{Z|S}(z \mid s)P_{Z|S}(z + w \mid s)}{P_{Z|S}(z + p^i\mathbb{Z}_{p^r} \mid s)} \tag{B.18}$$

For this choice of $P_{ZW|S}(z, w \mid s)$, we now show that equation (B.11) is satisfied.

$$\sum_{w\in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(z, w \mid s) = \sum_{w\in p^i\mathbb{Z}_{p^r}} \frac{P_{Z|S}(z \mid s)P_{Z|S}(z + w \mid s)}{P_{Z|S}(z + p^i\mathbb{Z}_{p^r} \mid s)} \tag{B.19}$$

$$= \frac{P_{Z|S}(z \mid s)}{P_{Z|S}(z + p^i\mathbb{Z}_{p^r} \mid s)} \sum_{w\in p^i\mathbb{Z}_{p^r}} P_{Z|S}(z + w \mid s) \tag{B.20}$$

$$= P_{Z|S}(z \mid s) \quad \forall z \in \mathcal{Z}, s \in \mathcal{S}. \tag{B.21}$$

Next, lets show that the choice of $P_{ZW|S}(z, w \mid s)$ in equation (B.18) satisfies equation (B.12).

(B.22)

$$\sum_{w\in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(Z = z - w, W = w \mid S = s) = \sum_{w\in p^i\mathbb{Z}_{p^r}} \frac{P_{Z|S}(z - w \mid s)P_{Z|S}(z \mid s)}{P_{Z|S}(z - w + p^i\mathbb{Z}_{p^r} \mid s)}$$

$$= P_{Z|S}(z \mid s) \sum_{w\in p^i\mathbb{Z}_{p^r}} \frac{P_{Z|S}(z - w \mid s)}{P_Z(z + p^i\mathbb{Z}_{p^r} \mid s)} \tag{B.23}$$

$$= P_{Z|S}(z \mid s) \quad \forall z \in \mathcal{Z}, s \in \mathcal{S}. \tag{B.24}$$

Finally, we show that this choice of $P_{ZW|S}(z, w \mid s)$ satisfies the KKT conditions

given by equation (B.17).

$$\prod_{z\in p^i\mathbb{Z}_{p^r}} P_{ZW|S}(z,w \mid s) = \prod_{z\in p^i\mathbb{Z}_{p^r}} \frac{P_{Z|S}(z \mid s)P_{Z|S}(z+w \mid s)}{P_{Z|S}(z+p^i\mathbb{Z}_{p^r} \mid s)} \tag{B.25}$$

$$= \left(\frac{1}{P_{Z|S}(p^i\mathbb{Z}_{p^r} \mid s)}\right)^{p^{r-i}} \prod_{z\in p^i\mathbb{Z}_{p^r}} P^2_{Z|S}(z \mid s) \tag{B.26}$$

which is independent of w and is the same for any $w \in p^i\mathbb{Z}_{p^r}$. Thus, equation (B.18) indeed is the solution to the optimization problem described by equation (B.13). Let us now compute the maximum value that the entropy $H(W \mid Z,S)$ takes for this choice of the conditional distribution $P_{ZW|S}$.

(B.27)

$$H(W \mid Z,S) = \sum_{s\in\mathcal{S}}\sum_{z\in\mathcal{Z}} P_{ZS}(z,s)\left(\sum_{w\in p^i\mathbb{Z}_{p^r}} P_{W|ZS}(w \mid z,s)\log\frac{1}{P_{W|ZS}(w \mid z,s)}\right)$$

$$= \sum_{s\in\mathcal{S}}\sum_{z\in\mathcal{Z}} P_{ZS}(z,s)\left(\sum_{w\in p^i\mathbb{Z}_{p^r}} \frac{P_{Z|S}(z+w \mid s)}{P_{Z|S}(z+p^i\mathbb{Z}_{p^r} \mid s)}\log\frac{P_{Z|S}(z+p^i\mathbb{Z}_{p^r} \mid s)}{P_{Z|S}(z+w \mid s)}\right) \tag{B.28}$$

Let $\mathcal{DC}$ be the set of all distinct cosets of $p^i\mathbb{Z}_{p^r}$ in $\mathbb{Z}_{p^r}$ and let $\mathcal{DC}(z)$ be the unique set in $\mathcal{DC}$ that contains z. Let us evaluate the summation in the brackets of equation (B.28) first.

$$\sum_{w\in p^i\mathbb{Z}_{p^r}} \frac{P_{Z|S}(z+w \mid s)}{P_{Z|S}(z+p^i\mathbb{Z}_{p^r} \mid s)}\log\frac{P_{Z|S}(z+p^i\mathbb{Z}_{p^r} \mid s)}{P_{Z|S}(z+w \mid s)} = \tag{B.29}$$

$$= \sum_{w\in p^i\mathbb{Z}_{p^r}} \frac{P_{Z|S}(z+w \mid s)}{P_{Z|S}(\mathcal{DC}(z) \mid s)}\log\frac{P_{Z|S}(\mathcal{DC}(z) \mid s)}{P_{Z|S}(z+w \mid s)}$$

$$= \log P_{Z|S}(\mathcal{DC}(z) \mid s) + \frac{\sum_{z'\in\mathcal{DC}(z)} P_{Z|S}(z' \mid s)\log\frac{1}{P_{Z|S}(z'|s)}}{P_{Z|S}(\mathcal{DC}(z) \mid s)}$$

This sum is dependent on z only through the coset $\mathcal{DC}(z)$ to which z belongs. Thus, the sum is the same for any two z that belong to the same coset of $p^i\mathbb{Z}_{p^r}$ in $\mathbb{Z}_{p^r}$.

Thus, we have $H(W \mid Z, S)$ given by the expression

(B.30)
$$\sum_{s\in\mathcal{S}} P_S(s) \sum_{T\in\mathcal{DC}} \sum_{z\in T} P_{Z|S}(z \mid s) \left(\log P_{Z|S}(T \mid s) + \frac{\sum_{z'\in T} P_{Z|S}(z' \mid s) \log \frac{1}{P_{Z|S}(z'|s)}}{P_{Z|S}(T \mid s)} \right)$$

(B.31)
$$= \sum_{s\in\mathcal{S}} \sum_{T\in\mathcal{DC}} P_{Z|S}(T \mid s) \left(\log P_{Z|S}(T \mid s) + \frac{\sum_{z'\in T} P_{Z|S}(z' \mid s) \log \frac{1}{P_{Z|S}(z'|s)}}{P_{Z|S}(T \mid s)} \right)$$

(B.32)
$$= \sum_{s\in S} \sum_{z'\in\mathcal{Z}} P_{Z|S}(z' \mid s) \log \frac{1}{P_{Z|S}(z' \mid s)} + \sum_{s\in S} P_{Z|S}(T \mid s) \log \frac{1}{P_{Z|S}(T \mid s)}$$

(B.33)
$$= H(Z \mid S) - H([Z]_i \mid S)$$

where $[Z]_i$ is as defined in Lemma B.2. □

We are now ready to prove the existence of good group channel codes. Let Z take values in the group $\mathbb{Z}_{p^r}$ and further be non-redundant. Coding is done in blocks of length n. We show the existence of a good channel code by averaging the probability of a decoding error over all possible choices of the homomorphism $\phi(\cdot)$ from the family $\mathrm{Hom}(\mathbb{Z}_{p^r}^n, \mathbb{Z}_{p^r}^k)$. Let H be the parity check matrix and $\mathcal{C}$ be the kernel of a randomly chosen homomorphism $\phi(\cdot)$.

The probability of the set $B_\epsilon(\mathcal{C})$ can be written as

(B.34)
$$P_{ZS}(B_\epsilon(\mathcal{C})) = \sum_{(z^n, s^n)} P_{ZS}(z^n, s^n) I \left(\bigcup_{\substack{(\tilde{z}^n, s^n)\in A_\epsilon^n(Z,S) \\ \tilde{z}^n \neq z^n}} (\phi(\tilde{z}^n) = \phi(z^n)) \right)$$

where $I(E)$ is the indicator of the event E. Taking the expectation of this probability,

we get

(B.35)

$$\mathbb{E}(P_{ZS}(B_\epsilon(\mathcal{C}))) = \sum_{(z^n,s^n)} P_{ZS}(z^n,s^n) P\left(\bigcup_{\substack{(\tilde{z}^n,s^n)\in A_\epsilon^n(Z,S) \\ \tilde{z}^n \neq z^n}} (\phi(\tilde{z}^n) = \phi(z^n)) \right)$$

(B.36)
$$\leq \sum_{(z^n,s^n)\in A_\epsilon^n(Z,S)} P_{ZS}(z^n,s^n) P\left(\bigcup_{\substack{(\tilde{z}^n,s^n)\in A_\epsilon^n(Z,S) \\ \tilde{z}^n \neq z^n}} (\phi(\tilde{z}^n) = \phi(z^n)) \right) + \sum_{(z^n,s^n)\notin A_\epsilon^n(Z,S)} P_{ZS}(z^n,s^n)$$

(B.37)
$$\leq \sum_{(z^n,s^n)\in A_\epsilon^n(Z,S)} P_{ZS}(z^n,s^n) P\left(\bigcup_{\substack{(\tilde{z}^n,s^n)\in A_\epsilon^n(Z,S) \\ \tilde{z}^n \neq z^n}} (\phi(\tilde{z}^n - z^n) = 0^k) \right) + \delta_1$$

where $\delta_1 \to 0$ as $n \to \infty$.

We now derive a uniform bound for the probability that for a given $(z^n, s^n) \in A_\epsilon^n(Z,S)$, a randomly chosen homomorphism maps $\tilde{z}^n$ to the same syndrome as z^n for some $\tilde{z}^n$ such that $(\tilde{z}^n, s^n) \in A_\epsilon^n(Z,S)$. From Lemma B.1 and B.2, we see that this probability depends on which of the sets $S_{i,\epsilon}(z^n,s^n), 0 \leq i < r$ the sequence $\tilde{z}^n$ belongs to.

(B.38)
$$P\left(\bigcup_{\substack{(\tilde{z}^n,s^n)\in A_\epsilon^n(Z,S) \\ \tilde{z}^n \neq z^n}} \phi(\tilde{z}^n - z^n) = 0^k \right) \leq \sum_{\substack{(\tilde{z}^n,s^n)\in A_\epsilon^n(Z,S) \\ \tilde{z}^n \neq z^n}} P(\phi(\tilde{z}^n - z^n) = 0^k)$$

(B.39)
$$= \sum_{i=0}^{r-1} \sum_{\tilde{z}^n \in S_{i,\epsilon}(z^n,s^n) \setminus S_{i+1,\epsilon}(z^n,s^n)} P(\phi(\tilde{z}^n - z^n) = 0^k)$$

(B.40)
$$\overset{(a)}{=} \sum_{i=0}^{r-1} \frac{|S_{i,\epsilon}(z^n,s^n)| - |S_{i+1,\epsilon}(z^n,s^n)|}{p^{(r-i)k}}$$

(B.41)
$$\overset{(b)}{\leq} \sum_{i=0}^{r-1} \exp_2\left(n \left[H(Z|S) - H([Z]_i|S) - \frac{k}{n}(r-i)\log p + \delta_2(\epsilon) \right] \right)$$

where (a) follows from Lemma B.1 and (b) follows from Lemma B.2. If this summation were to go to zero with block length, it would follow from equation (B.37) that the expected probability of the set $B_\epsilon(\mathcal{C})$ also goes to zero. This implies the existence of at least one homomorphism $\phi(\cdot)$ such that the associated codebook $\mathcal{C}$ satisfies for a given $\epsilon > 0$, $P_{ZS}(B_\epsilon(\mathcal{C})) \leq \epsilon$ for sufficiently large block length.

The summation in equation (B.41) goes to zero if each of the terms goes to zero. This happens if

$$\frac{k}{n}\frac{r-i}{r}\log p^r \geq H(Z|S) - H([Z]_i|S) + \delta_2(\epsilon) \quad \text{for } i = 0, \ldots, r-1 \tag{B.42}$$

or equivalently

$$\frac{k(n)}{n}\log p^r > \max_{0 \leq i < r}\left(\frac{r}{r-i}\right)(H(Z|S) - H([Z]_i|S)) + \delta_2(\epsilon) \tag{B.43}$$

It is clear that in the limit as $n \to \infty$, good group channel codes exist such that the dimensions of the associated parity check matrices satisfy equation (3.27). When $\mathcal{C}$ is a good channel code, define the decoding function $\psi\colon \mathbb{Z}_{p^r}^k \times \mathcal{S}^n \to \mathbb{Z}_{p^r}^n$ for a given (z^n, s^n) as the unique member of the set $\{\hat{z}^n\colon H\hat{z}^n = Hz^n, (\hat{z}^n, s^n) \in A_\epsilon^n(Z,S)\}$.

B.2 Good Group Source Codes

We prove the existence of source codes built over the space $\mathbb{Z}_{p^r}^n$ which are good for the triple $(\mathcal{X}, \mathcal{U}, P_{XU})$ according to Definition 3.9. Let the random variable U take values from the group $\mathbb{Z}_{p^r}$, i.e., $\mathcal{U} = \mathbb{Z}_{p^r}$ and let U be non-redundant. Let $\phi\colon \mathbb{Z}_{p^r}^n \to \mathbb{Z}_{p^r}^k$ be a homomorphism for some k to be fixed later. The codebook $\mathcal{C}$ is the kernel $\ker(\phi)$ of this homomorphism. Note that $\ker(\phi) < \mathbb{Z}_{p^r}^n$ and hence the codebook has a group structure. We show the existence of a good code $\mathcal{C}$ by averaging the probability of error over all possible choices of $\phi(\cdot)$ from the family of all homomorphisms $\mathrm{Hom}(\mathbb{Z}_{p^r}^n, \mathbb{Z}_{p^r}^k)$.

Recall the definition of the set $A_\epsilon(\mathcal{C})$ from equation (3.22). The probability of this set can be written as

$$P(A_\epsilon(\mathcal{C})) = \sum_{x^n} P_X(x^n) I\left(\bigcup_{u^n \in \mathcal{C}} (x^n, u^n) \in A_\epsilon^n(X,U)\right) \tag{B.44}$$

The expected value of this probability is

$$\mathbb{E}(P(A_\epsilon(\mathcal{C}))) = \sum_{x^n} P_X(x^n) P\left(\bigcup_{u^n \in \mathcal{C}} (x^n, u^n) \in A_\epsilon^n(X,U)\right) \tag{B.45}$$

$$\geq \sum_{x^n \in A_\epsilon^n(X)} P_X(x^n) P\left(\bigcup_{u^n \in \mathcal{C}} (x^n, u^n) \in A_\epsilon^n(X,U)\right) \tag{B.46}$$

For a typical x^n, let us compute the probability that there exists no $u^n \in \mathcal{C}$ jointly typical with the source sequence x^n. Define the random variable $\Theta(x^n)$ as

$$\Theta(x^n) = \sum_{u^n \in A_\epsilon^n(x^n)} 1_{\{u^n \in \mathcal{C}\}}. \tag{B.47}$$

$\Theta(x^n)$ counts the number of u^n sequences in the codebook $\mathcal{C}$ that are jointly typical with x^n. The error event E given that the source sequence is x^n is equivalent to the event $\{\Theta(x^n) = 0\}$. Thus, we need to evaluate the probability of this event. Note that $\Theta(x^n)$ is a sum of indicator random variables some of which might be dependent. This dependence arises from the structural constraint on the codebook $\mathcal{C}$. For example, $u_1^n \in \mathcal{C}$ implies that $ku_1^n \in \mathcal{C}$ as well for any $k \in \mathbb{Z}_{p^r}$. We use Suen's inequality [98] to bound this probability.

In order to use Suen's inequality, we need to form the dependency graph between these indicator random variables. We do this in a series of lemmas. We first evaluate the probability that a given typical sequence belongs to the kernel of a randomly chosen homomorphism. Since U is assumed to be non-redundant, by Lemma B.1, we have

$$P(u^n \in \mathcal{C}) = p^{-rk} \tag{B.48}$$

We now turn our attention to pairwise relations between the indicator random variables. For two n-length sequences u_1^n, u_2^n, define the matrices $M_{k,l}(u_1^n, u_2^n), 1 \leq k, l \leq n$ and $k \neq l$ as

$$M_{k,l}(u_1^n, u_2^n) = \begin{bmatrix} u_{1k} & u_{1l} \\ u_{2k} & u_{2l} \end{bmatrix} \tag{B.49}$$

Let $m_{k,l}(u_1^n, u_2^n)$ be the determinant of the matrix $M_{k,l}(u_1^n, u_2^n)$. Define the set

$$M(u_1^n, u_2^n) \triangleq \{m_{k,l}(u_1^n, u_2^n) \colon u_{1k}^{-1} \text{ exists}\} \tag{B.50}$$

Note that the set $M(u_1^n, u_2^n)$ is non-empty since u_1^n is assumed to be a non-redundant sequence. Let $D(u_1^n, u_2^n)$ be the smallest subgroup of $\mathbb{Z}_{p^r}$ that contains the set $M(u_1^n, u_2^n)$. As will be shown, the probability that both u_1^n and u_2^n belong to the kernel of a randomly chosen homomorphism depends on $D(u_1^n, u_2^n)$. For ease of notation, we suppress the dependence of the various quantities on the sequences u_1^n, u_2^n in what follows.

Lemma B.3. *For two non-redundant sequences u_1^n, u_2^n, the probability that a random homomorphism $\phi \colon \mathbb{Z}_{p^r}^n \to \mathbb{Z}_{p^r}^k$ maps the sequences to 0^k is*

$$P(\phi(u_1^n) = \phi(u_2^n) = 0^k) = p^{-(2r-i)k} \quad \text{if } D(u_1^n, u_2^n) = p^i \mathbb{Z}_{p^r}, 0 \leq i \leq r \tag{B.51}$$

Proof: Let the homomorphism $\phi(\cdot)$ be decomposed as $\phi_i \colon \mathbb{Z}_{p^r}^n \to \mathbb{Z}_{p^r}, 1 \leq i \leq k$. We first count the number of homomorphisms $\phi_1(\cdot)$ that map both u_1^n and u_2^n to 0. Recall that $\phi_1(u_1^n)$ can be expressed as the linear combination $\phi_1(u_1^n) = \sum_{j=1}^n \alpha_j u_{1j}$ for $\alpha_j \in \mathbb{Z}_{p^r}, 1 \leq j \leq n$. Thus, we need to find the number of solutions $\{\alpha_j\}_{j=1}^n$ that simultaneously satisfy the equations

$$\sum_{j=1}^{n} \alpha_j u_{1j} = 0 \tag{B.52}$$

$$\sum_{j=1}^{n} \alpha_j u_{2j} = 0 \tag{B.53}$$

If $D(u_1^n, u_2^n) = p^i\mathbb{Z}_{p^r}$, then there exists some $1 \leq k \leq n$ such that u_{1k}^{-1} exists and $m_{k,j^*} \in p^i\mathbb{Z}_{p^r} \backslash p^{i+1}\mathbb{Z}_{p^r}$ for some $1 \leq j^* \leq n, j^* \neq k$. Fix such a k. Then, any solution to the equation (B.52) must be of the form $\alpha_j, j \neq k$ arbitrary, $\alpha_k = -u_{1k}^{-1} \sum_{j \neq k} \alpha_j u_{1j}$ for some k such that u_{1k}^{-1} exists. Thus, the total number of solutions to equation (B.52) is $p^{r(n-1)}$. Substituting one such solution into equation (B.53), we get

$$\sum_{j=1}^{n} \alpha_j u_{2j} = \sum_{j \neq k} \alpha_j u_{2j} - u_{1k}^{-1} u_{2k} \left(\sum_{j \neq k} \alpha_j u_{1j} \right) \tag{B.54}$$

$$= u_{1k}^{-1} \left(\sum_{j \neq k} \alpha_j (u_{1k} u_{2j} - u_{2k} u_{1j}) \right) \tag{B.55}$$

$$= u_{1k}^{-1} \sum_{j \neq k} \alpha_j m_{k,j} \tag{B.56}$$

Of the $p^{r(n-1)}$ choices for $\{\alpha_i\}_{i=1}^n$, we need to find those that satisfy $\sum_{j \neq k} \alpha_j m_{k,j} = 0$. We allow α_j to be arbitrary for $j \neq k, j^*$ and solve the equation $\alpha_{j^*} m_{k,j*} = -\sum_{j \neq k, j^*} \alpha_j m_{k,j}$. It is clear that the summation in the right hand side yields a sum that belongs to $p^i\mathbb{Z}_{p^r}$. Since k, j^* are chosen such that $m_{k,j^*} \in p^i\mathbb{Z}_{p^r} \backslash p^{i+1}\mathbb{Z}_{p^r}$, it follows from Lemma B.6 in Appendix B.5 that this equation has p^i solutions for α_{j^*} for each of the $p^{r(n-2)}$ choices of $\alpha_j, j \neq k, j^*$. Once $\alpha_j, j \neq k$ is fixed, α_k is automatically fixed at $\alpha_k = -u_{1k}^{-1} \sum_{j \neq k} \alpha_j u_{1j}$. Thus, the total number of solutions that simultaneously satisfy equations (B.52) and (B.53) is $p^i p^{r(n-2)}$.

It follows that the probability of a randomly chosen homomorphism $\phi_1(\cdot)$ mapping both u_1^n, u_2^n to 0 is given by p^i/p^{2r}. Since each of the k homomorphisms $\phi_i, 1 \leq i \leq k$ can be chosen independently, we have

$$P(\phi(u_1^n) = \phi(u_2^n) = 0) = p^{-(2r-i)k} \tag{B.57}$$

when $D(u_1^n, u_2^n) = p^i\mathbb{Z}_{p^r}$ for some $0 \leq i \leq r$. This proves the claim of Lemma B.3. □

Suppose u_1^n and u_2^n are non-redundant sequences. It follows from Lemmas B.1 and B.3 that the events $1_{\{u_1^n \in \mathcal{C}\}}$ and $1_{\{u_2^n \in \mathcal{C}\}}$ are independent when $D(u_1^n, u_2^n) = \mathbb{Z}_{p^r}$. In order to infer the dependency graph of the indicator random variables in equation (B.47), we need to count the number of sequences u_2^n for a given u_1^n such that $D(u_1^n, u_2^n) = p^i\mathbb{Z}_{p^r}$ for a given $1 \leq i \leq r$. This is the content of the next lemma.

Lemma B.4. *Let u_1^n be a non-redundant sequence. Let $D_i(u_1^n), 0 \leq i \leq r$ be the set of all u_2^n sequences such that $D(u_1^n, u_2^n) = p^i\mathbb{Z}_{p^r}$. The size of the set $D_i(u_1^n)$ is given by*

$$|D_i(u_1^n)| = \begin{cases} p^r\left(p^{(r-i)(n-1)} - p^{(r-i-1)(n-1)}\right) & 0 \leq i < r \\ p^r - 1 & i = r \end{cases} \tag{B.58}$$

Proof: We start by estimating the size of $D_r(u_1^n)$, i.e., the set of u_2^n sequences such that $D(u_1^n, u_2^n) = 0$. Since $D(u_1^n, u_2^n) = 0$, u_2^n must be such that there exists $1 \leq k \leq n$ such that u_{1k}^{-1} exists and $m_{k,j} = 0$ for all $j \neq k$. This implies that $u_{1k}u_{2j} = u_{2k}u_{1j}$ for all $j \neq k$. Define $\eta = u_{1k}^{-1}u_{2k}$. It then follows that $u_{2j} = \eta u_{1j}$ for all $1 \leq j \leq n$ which implies that $u_2^n = \eta u_1^n$ for some $\eta \in \mathbb{Z}_{p^r}$. Since it is assumed that $u_2^n \neq u_1^n$, there are $p^r - 1$ distinct values of η. Since the sequence u_1^n is non-redundant, it follows that each value of η results in a distinct value of u_2^n. Thus, $|D_r(u_1^n)| = p^r - 1$ as claimed in the Lemma.

Consider the case when $D(u_1^n, u_2^n) = p^i\mathbb{Z}_{p^r}$ for some $0 \leq i < r$. We count the number of u_2^n for a given u_1^n such that $p^i\mathbb{Z}_{p^r}$ is the smallest subgroup containing all the set $M(u_1^n, u_2^n)$. Since $D(u_1^n, u_2^n) = p^i\mathbb{Z}_{p^r}$, u_2^n must be such that there exists $1 \leq k \leq n$ such that u_{1k}^{-1} exists and $m_{k,j^*} \in p^i\mathbb{Z}_{p^r} \backslash p^{i+1}\mathbb{Z}_{p^r}$ for some $1 \leq j^* \leq n, j^* \neq k$. Consider the matrices $M_{k,l}(u_1^n, u_2^n), 1 \leq l \leq n, l \neq k$. Let $\Delta_{k,l} \in p^i\mathbb{Z}_{p^r}, 1 \leq l \leq n, l \neq k$. Fixing the values of the determinants $m_{k,l}(u_1^n, u_2^n)$ to be $\Delta_{k,l}$, we can solve for the entire sequence u_2^n. Thus, $D_i(u_1^n)$ contains the union over all permissible values of $\{\Delta_{k,l}\}_{l \neq k}$

of those sequences u_2^n such that $m_{k,l}(u_1^n, u_2^n) = \Delta_{k,l}$ for all $1 \leq l \leq n, l \neq k$.

For a given $\{\Delta_{k,l}\}_{l\neq k}$, let us investigate the number of u_2^n sequences such that $m_{k,l}(u_1^n, u_2^n) = \Delta_{k,l}$ for all $1 \leq l \leq n, l \neq k$. Consider first the equation $m_{k,l^*} = \Delta_{k,l^*}$ for some $l^* \neq k$. Since u_{1k} is invertible, there are p^r possible solutions in (u_{2k}, u_{2l^*}) for this equation. Now consider the equations $m_{k,l}, 1 \leq l \leq n, l \neq k, l^*$. Since u_{2k} is already fixed and u_{1k} is invertible, there is precisely one solution to u_{2l} in these equations. Solving these $(n-1)$ equations fixes the sequence u_2^n. Thus, the number of solutions to u_2^n for a given u_1^n and $\{\Delta_{k,l}\}_{l\neq k}$ is p^r. The number of $\Delta_{k,l}$ such that $\{\Delta_{k,l}\}_{l\neq k} \in p^i \mathbb{Z}_{p^r}^{n-1}$ is clearly $p^{(r-i)(n-1)}$. For $D(u_1^n, u_2^n) = p^i \mathbb{Z}_{p^r}$, there must exist at least one $\Delta_{k,l} \in p^i \mathbb{Z}_{p^r} \backslash p^{i+1} \mathbb{Z}_{p^r}$. The total number of such $\{\Delta_{k,l}\}_{l\neq k}$ is clearly $p^{(r-i)(n-1)} - p^{(r-i-1)(n-1)}$. Putting these arguments together, we get that the size of $D_i(u_1^n)$ is $p^r(p^{(r-i)(n-1)} - p^{(r-i-1)(n-1)})$. This proves the claim of Lemma B.4. □

We are now ready to infer the dependency graph of the indicator random variables in equation (B.47). The number of nodes in the dependency graph is $|A_\epsilon^n(x^n)|$. Let I_i be the indicator of the event $\{u_i^n \in \mathcal{C}\}$ and let I_i correspond to the ith vertex of the graph. From Lemma B.3, it follows that vertices i and j are connected (denoted by $i \sim j$) if $D(u_i^n, u_j^n) \neq \mathbb{Z}_{p^r}$. Using Lemma B.4, the degree of the ith vertex can be bounded by $p^{rn} - |D_0(u_1^n)| - 1 = p^{r+(r-1)(n-1)} - 1$. Note that this is an upper bound since not all u_2^n sequences counted in Lemma B.4 need belong to $A_\epsilon^n(x^n)$.

One version of Suen's inequality can be stated as follows. Let $I_i \in \text{Be}(p_i), i \in \mathcal{I}$ be a family of Bernoulli random variables having a dependency graph L with vertex set $\mathcal{I}$ and edge set $E(L)$. Let $X = \sum_i I_i$ and $\lambda = \mathbb{E}(X) = \sum_i p_i$. Write $i \sim j$ if $(i,j) \in E(L)$ and let $\Delta = \frac{1}{2}\sum_i \sum_{j\sim i} \mathbb{E}(I_i I_j)$ and $\delta = \max_i \sum_{k\sim i} p_k$. Then

$$P(X = 0) \leq \exp\left\{-\min\left(\frac{\lambda^2}{8\Delta}, \frac{\lambda}{2}, \frac{\lambda}{6\delta}\right)\right\} \tag{B.59}$$

Let us estimate the quantities λ, Δ and δ for our problem. It follows from equation (B.48) that $\lambda = \mathbb{E}(\Theta(x^n)) = |A_\epsilon^n(x^n)|p^{-rk}$. Uniform upper and lower bounds [14] exist for the size of the set $A_\epsilon^n(x^n)$. An upper bound to Δ can be established via Lemmas B.3 and B.4 as below.

(B.60)

$$\Delta = \frac{1}{2}\sum_i \sum_{j\sim i} \mathbb{E}(I_i I_j)$$

(B.61)

$$= \frac{1}{2}\sum_i \sum_{j\sim i} P(\phi(u_i^n) = \phi(u_j^n) = 0)$$

(B.62)

$$= \frac{1}{2}\sum_{u_i^n \in A_\epsilon^n(x^n)} \sum_{m=1}^{r} \sum_{u_j \in A_\epsilon^n(x^n)\cap D_m(u_1^n)} P(\phi(u_i^n) = \phi(u_j^n) = 0)$$

(B.63)

$$\overset{(a)}{=} \frac{1}{2}\sum_{u_i^n \in A_\epsilon^n(x^n)} \sum_{m=1}^{r} |A_\epsilon^n(x^n) \cap D_m(u_1^n)| \left(\frac{p^m}{p^{2r}}\right)^k$$

(B.64)

$$\overset{(b)}{\leq} \frac{1}{2}\sum_{u_i^n \in A_\epsilon^n(x^n)} \left((p^r - 1)\left(\frac{1}{p^r}\right)^k + \sum_{m=1}^{r-1} p^r \left(p^{(r-m)(n-1)} - p^{(r-m-1)(n-1)}\right)\left(\frac{p^m}{p^{2r}}\right)^k \right)$$

(B.65)

$$= \frac{1}{2}|A_\epsilon^n(x^n)| \left((p^r - 1)\left(\frac{1}{p^r}\right)^k + \sum_{m=1}^{r-1} p^r \left(p^{(r-m)(n-1)} - p^{(r-m-1)(n-1)}\right)\left(\frac{p^m}{p^{2r}}\right)^k \right)$$

where (a) follows from Lemma B.3 and (b) follows from Lemma B.4. This expression can be further simplified by noting that $f(m) \triangleq |D_m(u_1^n)|p^{-k(2r-m)}$ is a decreasing function of m. Thus, the summation in the parentheses of equation (B.65) can be

upper bounded by $(r-1)f(1) = (r-1)|D_1(u_1^n)|p^{-k(2r-1)}$. Thus,

$$\Delta \leq \frac{1}{2}|A_\epsilon^n(x^n)|\left(p^{r-rk} + (r-1)p^{nr+k+1-n-2rk}\left(1 - \frac{1}{p^{n-1}}\right)\right) \tag{B.66}$$

$$\leq \frac{1}{2}|A_\epsilon^n(x^n)|p^{r-rk}\left(1 + (r-1)p^{(r-1)(n-k-1)}\right) \tag{B.67}$$

We now bound the quantity δ.

$$\delta = \max_i \sum_{j\sim i} \mathbb{E}(I_j) \tag{B.68}$$

$$= \max_{u_i^n \in A_\epsilon^n(x^n)} \sum_{m=1}^{r} \sum_{u_j^n \in D_m(u_i^n)} P(\phi(u_j^n) = 0) \tag{B.69}$$

$$\overset{(a)}{\leq} \max_{u_i^n \in A_\epsilon^n(x^n)} \left(p^{r+(r-1)(n-1)} - 1\right) p^{-rk} \tag{B.70}$$

$$\leq p^{r(n-k)-(n-1)} \tag{B.71}$$

where (a) follows from equation (B.48) and the fact the $P_{U|X}$ is a non-redundant distribution. Using these bounds, we can bound the terms involved in equation (B.59).

$$\frac{\lambda^2}{8\Delta} \geq \frac{|A_\epsilon^n(x^n)|^2 p^{-2rk}}{4|A_\epsilon^n(x^n)|p^{r-rk}\left(1 + (r-1)p^{(r-1)(n-k-1)}\right)} \tag{B.72}$$

$$\geq \frac{|A_\epsilon^n(x^n)|p^{-r(k+1)}}{4(1 + rp^{(r-1)(n-k-1)})} \tag{B.73}$$

$$\geq \frac{|A_\epsilon^n(x^n)|}{8r} p^{-(n(r-1)+k+1)} \tag{B.74}$$

where the last inequality holds for sufficiently large n. The third term in the exponent in equation (B.59) can be bounded as

$$\frac{\lambda}{6\delta} \geq \frac{|A_\epsilon^n(x^n)|}{6} p^{-(n(r-1)+1)} \tag{B.75}$$

Combining equations (B.74) and (B.75), we get that the probability of the event $\{\Theta(x^n) = 0\}$ is upper bounded by

$$\exp\left\{-\min\left(\frac{|A_\epsilon^n(x^n)|}{2}p^{-rk}, \frac{|A_\epsilon^n(x^n)|}{8r}p^{-(n(r-1)+k+1)}, \frac{|A_\epsilon^n(x^n)|}{6}p^{-(n(r-1)+1)}\right)\right\} \tag{B.76}$$

As long as each of the terms in the minimizations goes to ∞ as $n \to \infty$, the probability of not finding a jointly typical sequence with x^n in the codebook $\mathcal{C}$ goes to 0. Let $x^n \in A_\epsilon^n(X)$ be a typical sequence. It is well known [14] that for sufficiently large n, the size of the set $A_\epsilon^n(x^n)$ is lower bounded as

$$|A_\epsilon^n(x^n)| \geq 2^{n(H(U|X)-\epsilon_1(\epsilon))} \tag{B.77}$$

where $\epsilon_1(\epsilon) \to 0$ as $\epsilon \to 0$. Therefore,

$$\frac{|A_\epsilon^n(x^n)|}{2} p^{-rk} \geq \frac{1}{2} \exp_2 \left(n \left[H(U|X) - \frac{rk}{n} \log p - \epsilon_1 \right] \right) \tag{B.78}$$

$$\frac{|A_\epsilon^n(x^n)|}{8r} p^{-(n(r-1)+k+1)} \geq \frac{1}{8r} \exp_2 \left(n \left[H(U|X) - \left((r-1) + \frac{k+1}{n} \right) \log p - \epsilon_1 \right] \right) \tag{B.79}$$

$$\frac{|A_\epsilon^n(x^n)|}{6} p^{-(n(r-1)+1)} \geq \frac{1}{6} \exp_2 \left(n \left[H(U|X) - \left((r-1) + \frac{1}{n} \right) \log p - \epsilon_1 \right] \right) \tag{B.80}$$

For the probability in equation (B.76) to decay to 0, we need the exponents of these three terms to be positive. Equation (B.78) gives us the condition

$$\frac{k}{n} \log p^r < H(U|X) \tag{B.81}$$

while equations (B.79) and (B.80) together give us the condition

$$0 < \frac{k}{n} \log p^r < r(H(U|X) - \log p^{r-1}) \tag{B.82}$$

Of these two bounds for $\frac{k}{n} \log p^r$, it is easy to see that the dominating bound is equation (B.82) since $H(U|X) \leq H(U) \leq \log p^r$. Thus, the dimensionality of the parity check matrix satisfies the asymptotic condition

$$\lim_{n \to \infty} \frac{k(n)}{n} \log p^r = r|H(U|X) - \log p^{r-1}|^+ \tag{B.83}$$

where $|x|^+ = \max(x, 0)$. Combining these results, we see that provided equation (B.83) is satisfied, $P(\Theta(x^n) = 0)$ goes to 0 double exponentially. We now show that there exists at least one codebook $\mathcal{C}$ such that the set $A_\epsilon(\mathcal{C})$ has high probability. We do this by calculating the ensemble average of $P(A_\epsilon(\mathcal{C}))$ over all codebooks $\mathcal{C}$. It follows from equation (B.46) that

$$\mathbb{E}(P_X(A_\epsilon(\mathcal{C}))) \geq \sum_{x^n \in A_\epsilon^n(X)} P_X(x^n) P(\Theta(x^n \neq 0)) \tag{B.84}$$

$$\geq (1 - \epsilon_2)(1 - P(\Theta(x^n) = 0)) \tag{B.85}$$

where $\epsilon_2 \to 0$ as $n \to \infty$. Thus, as long as equation (B.83) is satisfied, the expected value of $P_X^n(A_\epsilon(\mathcal{C}))$ can be made arbitrarily close to 1. This implies that there exists at least one homomorphism such that its kernel is a good source code for the triple $(\mathcal{X}, \mathcal{U}, P_{XU})$.

B.3 Good Nested Group Codes

We now show the existence of good nested group codes satisfying Lemma 3.13. As was remarked in Definition 3.8, one way to construct a nested group code is to add rows to the parity check matrix of the fine code to get the parity check matrix of the coarse code. Let the random variables X, Y, U, V, S be as given in Lemma 3.13. Let the parity check matrices of the codes $\mathcal{C}_{11}, \mathcal{C}_{12}$ and $\mathcal{C}_2$ be H_{11}, H_{12} and H_2 respectively. Let their corresponding dimensions be $k_{11} \times n, k_{12} \times n$ and $k_2 \times n$ respectively. In order to ensure nesting, impose the following structural constraints on these matrices.

$$H_{12} = \begin{bmatrix} H_{11} \\ \Delta H_1 \end{bmatrix}, \quad H_2 = \begin{bmatrix} H_{12} \\ \Delta H_2 \end{bmatrix} \tag{B.86}$$

Let the dimensions k_{11}, k_{12} and k_2 satisfy equations (3.29) - (3.31).

Generate random H_2, H_{12} matrices by constructing the matrices $H_{11}, \Delta H_1, \Delta H_2$ independently by picking entries uniformly and independently from the group $\mathbb{Z}_{p^r}$. From the proofs in Appendices B.1 and B.2, it follows that the codes $\mathcal{C}_{11}, \mathcal{C}_{12}$ and $\mathcal{C}_2$ are with high probability good source and channel codes respectively for the appropriate triples. By union bound, it follows then that there exists a choice of $H_{11}, \Delta H_1$ and ΔH_2 such that the codebook $\mathcal{C}_2$ is a good channel code and the nested codes $\mathcal{C}_{11}$ and $\mathcal{C}_{12}$ are simultaneously good source codes for their respective triples. This proves the existence of good nested group codes as claimed in Lemma 3.13.

B.4 Group Codes Achieve Shannon Entropy Bound

We prove that, when used as lossless source codes, group codes can achieve the Shannon entropy bound and thus incur no loss in first order performance. As shown in 3.8.1, a good group channel code $\mathcal{C}$ for the triple $(\mathcal{X}, 0, P_X)$ can be used to achieve a source coding rate of

$$R \geq \max_{0 \leq i < r} \left(\frac{r}{r-i} \right) \left(H(X) - H([X]_i) \right) \tag{B.87}$$

The term corresponding to $i = 0$ in the above expression equals the entropy of the source. The maximization suggests that the minimum achievable rate using group codes could be larger than $H(X)$. However, if the sufficient condition of equation (3.41) is met, group codes can attain the entropy bound. We now show that there always exists a bijection $\pi \colon \mathcal{X} \to \mathcal{X}$ such that $X^\pi \triangleq \pi(X)$ satisfies the sufficient condition of (3.41). Since lossless reconstruction of X^π is equivalent to the lossless reconstruction of X, Corollary 1 would follow.

Lemma B.5. *Given a random variable X taking values in the set $\mathcal{X}$ with $|\mathcal{X}| = p^r$,*

there exists a bijective mapping $\pi\colon \mathcal{X} \to \{0, 1, \ldots, p^r - 1\}$ *such that*

$$H([X^\pi]_i) \geq \frac{i}{r} H(X^\pi) \quad 0 \leq i \leq r \tag{B.88}$$

with $X^\pi \triangleq \pi(X)$.

Proof: We start by numbering the elements of $\mathcal{X}$ using the labels $\{0, 1, \ldots, p^r - 1\}$ in some arbitrary order. Let this numbering be denoted by the permutation $\tilde{\pi}\colon \mathcal{X} \to \{0, 1, \ldots, p^r - 1\}$ and denote by $\tilde{X}_r$ the p^r-ary random variable $\tilde{\pi}(X)$. We write down the r-digit expansion of the numbers $0, 1, \ldots, p^r - 1$ in base p. Define the p-ary random variables $(D_1, \ldots, D_r)$ as follows: D_k takes values in the set $\{0, 1, \ldots, p-1\}$ and its probability mass function is given by

$$P(D_k(i)) = P(x \in \mathcal{X}\colon k^{\text{th}} \text{ digit of } \tilde{\pi}(x) = i) \quad 0 \leq i \leq p - 1 \tag{B.89}$$

The proof proceeds in two steps. We first create from $\tilde{X}_r$ a sequence of random variables $\tilde{X}_{r-1}, \ldots, \tilde{X}_1$ where $\tilde{X}_i$ is a p^i-ary random variable. These random variables are created in such a way that they obey the inequality

$$H(\tilde{X}_i) \geq \left(\frac{i}{i+1}\right) H(\tilde{X}_{i+1}) \tag{B.90}$$

Further, the p^i-ary random variable $\tilde{X}_i$ is created by grouping the symbols of the p^{i+1}-ary random variable $\tilde{X}_{i+1}$. The second step is as follows: Once the r random variables $\tilde{X}_r, \ldots, \tilde{X}_1$ are created thus, we use them to create the permutation $\pi(\cdot)$ mentioned in Lemma B.5. The labeling $\pi(\cdot)$ is done such that the elements of $\tilde{X}_i$ are identified with the subgroup $p^i\mathbb{Z}_{p^r}$. Finally, we will show how these two steps taken together imply equation (B.88) thus completing the proof.

We start by demonstrating the creation of $\tilde{X}_{r-1}$ from $\tilde{X}_r = \tilde{X}$. To do so, we use the following inequality on the entropy rates of subsets (see [15], Section 17.6).

Suppose we have a collection of n random variables $(W_1, \ldots, W_n)$. Define for every $S \subset \{1, \ldots, n\}$, the random variable $W(S)$ as $W(S) \triangleq \{W_i \colon i \in S\}$. Let

$$h_k^{(n)} \triangleq \frac{1}{\binom{n}{k}} \sum_{S \colon |S|=k} \frac{h(W(S))}{k} \tag{B.91}$$

Then, we have $h_1^{(n)} \geq h_2^{(n)} \geq \cdots \geq h_n^{(n)}$.

Let us apply this inequality to all subsets of the random variables $(D_1, \ldots, D_r)$ of cardinality $(r-1)$. We then have

$$h_{(r-1)}^{(r)} = \frac{1}{r} \sum_{S \colon |S|=r-1} \frac{h(D(S))}{r-1} \tag{B.92}$$

The inequality $h_{(r-1)}^{(r)} \geq h_r^{(r)}$ gives us

$$\frac{1}{r} \sum_{S \colon |S|=r-1} \frac{h(D(S))}{r-1} \geq \frac{h(D_1, \ldots, D_r)}{r} \tag{B.93}$$

or equivalently

$$\sum_{S \colon |S|=r-1} h(D(S)) \geq (r-1) H(\tilde{X}_r) \tag{B.94}$$

since the collection of random variables $(D_1, \ldots, D_r)$ is the same as the random variable $\tilde{X}_r = \tilde{X}$.

This inequality implies that among the r sets $S \subset \{1, \ldots, r\}$ of cardinality $(r-1)$, there exists at least one, say S_{r-1}^* such that

$$h(D(S^*)) \geq \left(\frac{r-1}{r} \right) H(\tilde{X}_r) \tag{B.95}$$

Given the set $S_{r-1}^* \subset \{1, \ldots, r\}$ (whose cardinality is $(r-1)$), we create the random variable $\tilde{X}_{r-1}$ by grouping together all the p symbols of $\tilde{\pi}(\mathcal{X})$ whose p-ary expansion agrees in all the indices of S^*. Clearly, $\tilde{X}_{r-1}$ is a p^{r-1}-ary random variable. To make this formal, define the ith element $(0 \leq i < p^{r-1})$ of the set $\mathcal{X}_{r-1}$ as

$$\mathcal{X}_{r-1}(i) = \{x \in \mathcal{X}_r \colon \pi(\mathcal{X})(S_{r-1}^*) = c_i\} \tag{B.96}$$

where c_i is the $(r-1)$ length p-ary expansion of i and $\mathcal{X}_r$ is identified with $\mathcal{X}$. Then, $\tilde{X}_{r-1}$ takes values in the set $\{0, \ldots, p^{r-1}-1\}$ and has a probability mass function

$$P(\tilde{X}_{r-1}(i)) = P\left(\mathcal{X}_{r-1}(i)\right) \tag{B.97}$$

To create $\tilde{X}_{r-2}$, we repeat the above proof with $\tilde{X}_{r-1}$ in place of $\tilde{X}_r$. Repeated application of this gives us the sequence of random variables $\tilde{X}_r, \ldots, \tilde{X}_1$ and also their corresponding alphabets $\mathcal{X}_r, \ldots, \mathcal{X}_1$ and the optimal choice of subsets $S^*_{r-1}, \ldots, S^*_1$.

We now turn to creating the permutation $\pi(\cdot)$ that ensures that $[X^\pi]_i$ obey equation (B.88). We do this by granting each symbol $x \in \mathcal{X}$, a p-ary label of length r. To do this, we construct a p-ary tree as follows: The tree has $(r+1)$ levels with the ith level containing the p^i elements of the set $\mathcal{X}_i$ for $1 \leq i \leq r$. The root of the tree (0th level) is a singleton set containing all the elements of $\mathcal{X}$. A node $\mathcal{X}_i(j)$ at the ith level has as a child a node $\mathcal{X}_{i+1}(k)$ at the $(i+1)$th level if and only if $\mathcal{X}_{i+1}(k)$ was grouped with other elements of $\mathcal{X}_{i+1}$ to form the symbol $\mathcal{X}_i(j)$. It follows that each node has exactly p children (except the leaves of the tree). For each node at the ith level $(i < r)$, we label the edges emanating from that node to its children in the $(i+1)$th level using the labels $(0, \ldots, p-1)$ in any arbitrary order. We are now ready to define the permutation $\pi(\cdot)$. For each $x \in \mathcal{X}$, we start from its corresponding leaf node at level r and trace the (unique) path to the root of the tree reading the labels of the traversed edges along the way. The resulting p-ary label of length r is then converted to an integer in the range $\{0, \ldots, p^r - 1\}$ which is then set as the value of $\pi(x)$.

It is easy to verify that the above tree labeling procedure effectively identifies the random variables $[X^\pi]_i$ with the random variables $D(S^*_i)$ at each stage i for $1 \leq i \leq r$.

Thus it follows that the family of random variables $[X^\pi]_i$ satisfy

$$H([X^\pi]_i) \geq \left(\frac{i}{i+1}\right) H([X^\pi]_{i+1}) \quad 1 \leq i < r \tag{B.98}$$

Successive application of this inequality yields

$$\begin{aligned} H([X^\pi]_i) &\geq \left(\frac{i}{i+1}\right) H([X^\pi]_{i+1}) \\ &\geq \left(\frac{i}{i+1}\right)\left(\frac{i+1}{i+2}\right) H([X^\pi]_{i+2}) \\ &\vdots \\ &\geq \left(\frac{i}{r}\right) H([X^\pi]_r) \qquad \text{(B.99)} \\ &\geq \left(\frac{i}{r}\right) H(X) \qquad \text{(B.100)} \end{aligned}$$

thus establishing the claim of Lemma B.5. This in turn establishes the existence of good group codes that achieve the entropy bound while used for lossless source coding. □

B.5 Linear Equations in Groups

We now present a lemma on the number of solutions over the group $\mathbb{Z}_{p^r}$ for a linear equation in one variable.

Lemma B.6. *Let $a \in p^i\mathbb{Z}_{p^r} \backslash p^{i+1}\mathbb{Z}_{p^r}$ for some $0 \leq i < r$. Then, the linear equation $ax = b$ has a solution in x if and only if $b \in p^i\mathbb{Z}_{p^r}$. In that case, there are p^i distinct solutions for x over the group $\mathbb{Z}_{p^r}$.*

Proof: It is clear that the equation $ax = b$ cannot have a solution if $b \notin p^i\mathbb{Z}_{p^r}$. The rest of the proof proceeds in two stages. We first show that if there exists at least one solution to the equation $ax = b$, then there exists p^i distinct solutions. We then show that at least one solution exists for every $b \in p^i\mathbb{Z}_{p^r}$. Together, these imply Lemma B.6.

Suppose there exists at least one solution x_1 to the equation $ax = b$. Then, for any $t \in p^{r-i}\mathbb{Z}_{p^r}$, $x_1 + t$ is also a solution and all such solutions are distinct. Conversely, if x_1, x_2 are both solutions, then $x_1 - x_2 \in p^{r-i}\mathbb{Z}_{p^r}$. Thus, existence of at least one solution implies the existence of exactly p^i solutions. Now consider the number of distinct values of the set $\{ax\colon x \in \mathbb{Z}_{p^r}\}$. Since every distinct value repeats itself exactly p^i times and there are p^r elements in this set, it follows that the number of distinct values is p^{r-i}. This is exactly the size of the subgroup $p^i\mathbb{Z}_{p^r}$ which implies that $ax = b$ has exactly p^i solutions for every element $b \in p^i\mathbb{Z}_{p^r}$. □

B.6 $\mathcal{T}$ is non-empty

Recall the definition of $\mathcal{T}$ from Section 3.7 as $\mathcal{T} = \{A\colon A \text{ is abelian}, |\mathcal{G}| \leq |A| \leq \alpha\beta, G(U,V) \subset A \text{ with respect to } P_{UV}\}$. Let $|\mathcal{U}| = \alpha, |\mathcal{V}| = \beta$. We now show that the function $G(U,V)$ can always be embedded in some abelian group belonging to $\mathcal{T}$. Consider the function $G_1(U,V) = (U,V)$. Clearly, $G_1(U,V) \subset \mathbb{Z}_\alpha \oplus \mathbb{Z}_\beta$ with respect to P_{UV} for any distribution P_{UV}. Since there is an obvious surjective mapping between the functions $G_1(U,V)$ and $G(U,V)$, it follows from Definition 3.5 that $G(U,V) \subset \mathbb{Z}_\alpha \oplus \mathbb{Z}_\beta$ with respect to P_{UV}. Since $|\mathbb{Z}_\alpha \oplus \mathbb{Z}_\beta| = \alpha\beta$, it follows that this group belongs to the set $\mathcal{T}$ and hence $\mathcal{T}$ is always non-empty.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, September 1973.

[2] A. D. Wyner, "Recent results in Shannon theory," *IEEE Trans. on Inform. Theory*, vol. 20, pp. 2–10, January 1974.

[3] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. on Inform. Theory*, vol. IT-21, pp. 294–300, May 1975.

[4] R. Ahlswede and J. Korner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT- 21, pp. 629–637, November 1975.

[5] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT- 22, pp. 1–10, January 1976.

[6] T. Berger, "Multiterminal source coding," in *Lectures presented at CISM summer school on the Inform. Theory approach to communications*, July 1977.

[7] S.-Y. Tung, *Multiterminal source coding*. PhD thesis, School of Electrical Engineering, Cornell University, Ithaca, NY, May 1978.

[8] T. Berger, K. B. Housewright, J. K. Omura, S. Tung, and J. Wolfowitz, "An upper bound on the rate distortion function for source coding with partial side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT- 25, pp. 664–666, November 1979.

[9] A. Barg and G. D. Forney Jr., "Random codes: Minimum distances and error exponents", *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2568–2573, September 2002.

[10] B. A. Nazer and M. Gastpar, "Computation over Multiple-Access Channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3498-3516, Oct. 2007.

[11] S. Gelfand and M. Pinsker, "Coding of sources on the basis of observations with incomplete information," *Problemy Peredachi Informatsii*, vol. 15, pp. 45–57, Apr-June 1979.

[12] J. Korner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 219–221, March 1979.

[13] T. S. Han, "Source coding with cross observations at the encoders," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 360–361, May 1979.

[14] I. Csiszár and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press Inc. Ltd., 1981.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York:Wiley, 1991.

[16] T. S. Han and K. Kobayashi, "A dichotomy of functions F(X,Y) of correlated sources (X,Y)," *IEEE Trans. on Inform. Theory*, vol. 33, pp. 69–76, January 1987.

[17] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. on Inform. Theory*, vol. 29, pp. 396–412, May 1983.

[18] R. W. Yeung and T. Berger, "Multiterminal source coding with one distortion criterion," *IEEE Trans. on Inform. Theory*, vol. 35, pp. 228–236, March 1989.

[19] H. Viswanathan, Z. Zhang, and T. Berger, "The CEO problem," *IEEE Trans. on Inform. Theory*, vol. 42, pp. 887–902, May 1996.

[20] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. on Inform. Theory*, vol. 43, pp. 1549–1559, September 1997.

[21] P. Viswanath, "Sumrate of multiterminal Gaussian source coding ," in *DIMACS workshop on Network Information Theory*, (Piscataway, NJ), April 2002.

[22] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," *The Transactions of the IECE of Japan*, vol. E-63, pp. 700–706, October 1980.

[23] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 773–787, November 1987.

[24] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Trans. Inform. Theory*, vol. IT- 18, pp. S293–S304, 1962.

[25] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 518–521, September 1980.

[26] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. IT-43, pp. 1912–1923, November 1997.

[27] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 1057–1070, May 1998.

[28] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-51, pp. 2577–2593, July 2005.

[29] Y. Oohama, "Rate distortion region for separate coding of correlated Gaussian remote observations," in *Allerton Conference*, (Monticello, IL), September 2005.

[30] Y. Oohama, "Separate source coding of correlated Gaussian remote sources," in *Workshop on Information theory and Applications (ITA)*, (San Diego, CA), January 2006.

[31] A. B. Wagner and V. Anantharam, "An improved outer bound for the multiterminal source-coding problem," in *IEEE International Symposium on Information Theory (ISIT '05)*, (Adelaide, Australia), September 2005.

[32] V. Prabhakaran, K. Ramchandran, and D. Tse, "Rate region of the quadratic Gaussian CEO problem," in *IEEE International Symposium on Information Theory (ISIT '04)*, (Chicago, IL), p. 117, June 2004.

[33] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inform. Theory*, vol. IT-47, pp. 903–917, March 2001.

[34] H. Yamamoto, "Wyner-Ziv theory for a general function of the correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 803–807, September 1982.

[35] H. Feng, M. Effros, and S. A. Savari, "Functional source coding for networks with receiver side information," in *Proc. of the 42nd annual Allerton conference on Communication, Control and Computing*, (Monticello, IL), September 2004.

[36] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. Inform. Theory*, vol. IT-50, pp. 2762–2768, November 2004.

[37] A. B. Wagner, S. Tavildar, and P. Viswanath, "The rate-region of the quadratic Gaussian two-terminal source-coding problem," `arXiv:cs.IT/0510095`.

[38] S. Tavildar, P. Viswanath, and A. B. Wagner, "The Gaussian many-help-one distributed source coding problem," in *Proc. of the 2006 IEEE Inform. Theory Workshop (ITW '06)*, (Chengdu, China), pp. 596–600, October 2006.

[39] S. Jana and R. Blahut, "Achievable region for multiterminal source coding with lossless decoding in all sources except one," to appear in *IEEE Inform. Theory Workshop (ITW '07)*, (Lake Tahoe, CA), September 2007.

[40] S. Jana, "Unified theory of source coding: Part I - two terminal problems," `arXiv:cs/0508118v2`.

[41] S. Jana, "Unified theory of source coding: Part II - multiterminal problems," `arXiv:cs/0508119v1`

[42] I. Csiszar, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Trans. Inform. Theory*, vol. IT- 28, pp. 585–592, July 1982.

[43] R. Zamir and M. Feder, "On universal quantization by randomized uniform lattice quantizers," *IEEE Trans. Inform. Theory*, vol. IT- 38, pp. 428–436, March 1992.

[44] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 1152–1159, July 1996.

[45] S. Shamai, S. Verdu, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. on Inform. Theory*, vol. 44, pp. 564–579, March 1998.

[46] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Trans. Inform. Theory*, vol. IT-45, pp. 106–117, January 1999.

[47] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. IT-48, pp. 1250–1276, June 2002.

[48] U. Erez and R. Zamir, "Achieving 1/2 log(1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inform. Theory*, vol. IT- 50, pp. 2293–2314, October 2004.

[49] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inform. Theory*, vol. IT- 51, pp. 3401–3416, October 2005.

[50] A. B. Wagner, "An outer bound for distributed compression of linear functions," *42nd Annual Conference in Inform. Sciences and Systems, 2008*, pp. 435–440, March 2008.

[51] H. Minkowski, "Dichteste gitterförmige Lagerung kongruenter Körper," *Nachr. Ges. Wiss. Göttingen*, pp. 311–355, 1904.

[52] E. Hlawka, "Zur Geometrie der Zahlen," *Math.Z.*, vol. 49, pp. 285–312, 1944.

[53] R. Kershner, "The number of circles covering a set," *Amer. Jour. Math.*, vol. 61, pp. 665–671, 1939.

[54] C. A. Rogers, *Packing and Covering*. Cambridge University Press, Cambridge, 1964.

[55] H. A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inform. Theory*, vol. IT- 43, pp. 1767–1773, November 1997.

[56] D. Krithivasan and S.S. Pradhan, "A proof of the existence of good nested lattices," `http://www.eecs.umich.edu/techreports/systems/cspl/cspl-384.pdf`.

[57] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. Springer, 1992.

[58] A. Kirac and P. Vaidyanathan, "Results on lattice vector quantization with dithering," *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing*, vol. 43, pp. 811–826, December 1996.

[59] V. A. Vaishampayan, N. J. A. Sloane and S. D. Servetto, "Multiple-description vector quantization with lattice codebooks: design and analysis," *IEEE Trans. Inform. Theory*, vol. IT-47, pp. 1718–1734, July 2001.

[60] Y. Frank-Dayan and R. Zamir, "Dithered lattice-based quantizers for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-48, pp. 192–204, January 2002.

[61] V. K. Goyal, J. A. Kelner and J. Kovacevic, "Multiple description vector quantization with a coarse lattice," *IEEE Trans. Inform. Theory*, vol. IT-48, pp. 781–788, March 2002.

[62] S. N. Diggavi, N. J. A. Sloane and V. A. Vaishampayan, "Asymmetric multiple description lattice vector quantizers," *IEEE Trans. Inform. Theory*, vol. IT-48, pp. 174–191, January 2002.

[63] J. Ostergaard, *Multiple-description lattice vector quantization*. PhD thesis, Delft University of Technology, Netherlands, June 2007.

[64] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 409–417, March 1994.

[65] P. Elias, "Coding for noisy channels", *IRE Conv. Record*, part. 4, pp. 37-46, 1955.

[66] T. J. Goblick, Jr., "Coding for a discrete information source with a distortion measure", *Ph.D. dissertation*, Dept. Electr. Eng., MIT , Cambridge, MA, 1962

[67] R. L. Dobrushin, "Asymptotic optimality of group and systematic codes for some channels", *Theor. Probab. Appl.*, vol. 8, pp. 52–66, 1963.

[68] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley and Sons, Inc., 1968.

[69] N. Ma and P. Ishwar, "Two-terminal distributed source coding with alternating messages for function computation", *Proc. IEEE Interational Symposium on Inform. Theory*, Toronto, Canada, 2008.

[70] A. Giridhar and P. R. Kumar, "Computing and communication functions over sensor networks", *IEEE Journal on selected areas in communications*, vol. 23, no. 4, pp. 755–764, April 2005.

[71] J. Muramatsu and S. Miyake, "Hash property and coding theorems for sparse matrices and maximum-likelihood coding", *Proc. IEEE International Symposium on Inform. Theory*, Toronto, Canada, 2008.

[72] G. Cohen, I. Honkala, S. Lytsyn and A. Lobstein, *Covering Codes*. North Holland-Elsevier, 1997

[73] P. Delsarte and P. M. Piret, "Do most linear codes achieve the Goblick bound on the covering radius?", *IEEE Trnas. Inform. Theory*, vol. IT-32, no. 6, pp. 826–828, November 1986.

[74] G. D. Cohen, "A nonconstructive upper bound on covering radius", *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 352–353, May 1983.

[75] V. M.. Blinovskii, "A lower bound on the number of words of a linear code in an arbitrary sphere with given radius in $\mathbb{F}_q^n$" (in Russian), *Probl. Pered. Inform. (Prob. Inf. Transm.)*, vol. 23, no. 2, pp. 50–53, 1987.

[76] J. Chen, Da-Ke He, A. Jugmohan, "Achieving the rate-distortion bound with linear codes", *IEEE Inform. Theory Workshop 2007*, pp. 662–667, Lake Tahoe, California.

[77] T. Philosof, A. Kishty, U. Erez and R. Zamir, "Lattice Strategies for the Dirty Multiple Access Channel", *Proceedings of IEEE International Symposium on Information Theory*, July 2007, Nice, France.

[78] D. Slepian, "Group codes for the Gaussian channel", *Bell Syst. Tech. Journal*, 1968.

[79] G. D. Forney, Jr., "Geometrically uniform codes", *IEEE Trans. Inform. Theory*, vol. 37, no. 5, pp. 1241–1260, September 1991.

[80] E. Biglieri and M. Elia, "On the existence of group codes for the Gaussian channel", *IEEE Trans. Inform. Theory*, vol. 18, no. 3, pp. 399–402, May 1972.

[81] G. D. Forney, Jr. and M. D. Trott, "The dynamics of group codes: State spaces, Trellis diagrams, and Canonical encoders", *IEEE Trans. Inform. Theory*, vol. 39, no. 9, pp. 1491–1513, September 1993.

[82] H. A. Loeliger and T. Mittelholzer, "Convolutional codes over groups", *IEEE Trans. Inform. Theory*,vol. 42, no. 6, pp. 1660–1686, November 1996.

[83] V. V. Vazirani, H. Saran and B. S. Rajan, "An efficient algorithm for constructing minimal trellises for codes over finite abelian groups", *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1839–1854, November 1996.

[84] H. A. Loeliger, "Signal sets matched to groups", *IEEE Trans. Inform. Theory*, vol. 37, no. 6, pp. 1675–1682, November 1991.

[85] S. D. Berman, "On the theory of group codes", *Kibernetika*, vol. 3, no. 1, pp. 31–39, 1967.

[86] G. D. Forney, Jr., "On the Hamming distance properties of group codes", *IEEE Trans. Inform. Theory*, vol. 38, no. 6, pp. 1797–1801, November 1992.

[87] E. Biglieri and M. Elia, "Construction of linear block codes over groups", *Proc. IEEE Interational Symposium on Inform. Theory*, San Antonio, TX, 1993.

[88] J. C. Interlando, R. Palazzo and M. Elia, "Group block codes over nonabelian groups are asymptotically bad", *IEEE Trans. Inform. Theory*, vol. 42, no. 4, pp. '1277–1280, July 1996.

[89] R. M. Tanner, D. Sridhara and T. Fuja, "A class of group structured LDPC codes", *Proc. of ISCTA*, Ambleside, 2001.

[90] G. Como and F. Fagnani, "The capacity of abelian group codes over symmetric channels", *Submitted for publication.*

[91] F. Garin and F. Fagnani, "Analysis of serial turbo codes over abelian groups for geometrically uniform constellations", *Submitted for publication.*

[92] R. Ahlswede, "Group codes do not achieve Shannon's channel capacity for general discrete channels", *The Annals of Mathematical Statistics*, vol. 42, no. 1, pp. 224–240, February 1971.

[93] R. Ahlswede and J. Gemma, " Bounds on algebraic code capacities for noisy channels I", *Information and Control*, pp. 124–145, 1971.

[94] R. Ahlswede and J. Gemma, "Bounds on algebraic code capacities for noisy channels II", *Information and Control*, pp. 146–158, 1971.

[95] D. Krithivasan and S. S. Pradhan, "Lattices for distributed source coding: Jointly Gaussian Sources and Reconstruction of a linear function," *IEEE Trans. Inform. Theory*, vol. 55, pp. 5628–5651, Dec. 2009.

[96] D. S. Dummit and R. M. Foote, *Abstract Algebra*. John Wiley & sons Inc., 2004.

[97] A. G. Kurosh, *The Theory of Groups*. Chelsea publishing company, 1960.

[98] S. Janson, "New versions of Suen's correlation inequality," *Random Structures Algorithms*, vol. 13, pp, 467–483, 1998.

[99] W. Gu, S. Jana and M. Effros, "On approximating the rate regions for lossy source coding with coded and uncoded side information", *Proc. IEEE International Symposium on Inform. Theory*, Toronto, Canada, 2008.

[100] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[101] M. V. Eyuboglu and G. D. Forney, "Lattice and trellis quantization with lattice and trellis-bounded codebooks-High rate theory for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 46–59, Jan. 1993.

[102] T. Philosof and R. Zamir, "The rate loss of single-letter characterization: The "Dirty" multiple access channel," *Available at arXiv:0803.1120v3*