Resource-constrained Coding for Communication and Computation Applications

by

Chin-Jen Pang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in the University of Michigan
2023

Doctoral Committee:

Associate Professor Hessam Mahdavifar, Co-Chair
Professor S. Sandeep Pradhan, Co-Chair
Associate Professor Achilleas Anastasopoulos
Associate Professor Mahdi Cheraghchi

Chin-Jen Pang

cjpang@umich.edu

ORCID iD: 0000-0002-0735-8967

To My Family.

ACKNOWLEDGEMENTS

I am deeply grateful to my advisors Prof. Hessam Mahdavifar and Prof. S. Sandeep Pradhan for all that they have provided throughout the course of my PhD. They have been of constant help and support in every way possible, and it has truly been a pleasure to work with them. I have learnt a great amount from them on numerous aspects of academic research, and I am sure that it will reflect positively on my entire career.

I would also like to thank the rest of the committee members: Prof. Achilleas Anastasopoulos and Prof. Mahdi Cheraghchi. Their mentorship and insightful comments have helped shape the dissertation tremendously.

I am thankful to all of my past and present colleagues for their ongoing friendship, and for providing a highly enjoyable working environment. It was also a great pleasure to work and interact with each of them.

Finally, I express my deepest gratitude to my parents, my brothers, and my wife for their constant care and support throughout all stages of my life, and I sincerely thank all of my friends in Ann Arbor for making my experiences truly memorable.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

FIGURE

LIST OF APPENDICES

ABSTRACT

Coding for data transmission has been extensively studied since the publication of Shannon's seminal work [126] in 1948. The research in coding theory has seen significant advances, including the invention of Reed-Solomon codes, convolutional code, LDPC codes, and Polar codes, over the past seven decades. However, many modern applications from disciplines such as wireless communications, machine learning, and quantum information transmission give rise to new challenges for which the coding theory is a framework that can offer novel solutions. Driven by such challenges, in this dissertation, we consider four coding theory problems with constraints on various aspects of the codes. In the first part, we model the problem of crowdsourced labeling, where labels for target items are retrieved through queries from a crowd of workers, as a coding problem where the generator matrices must be sparse. Leveraging prior results for codes with sparse representation, we propose querying schemes with *almost* optimal number of queries, each of which involving only a constant or a *relatively small* number of labels, for the reliable and unreliable query response scenarios, respectively. We further consider clustering the items based on two correlated classification criteria.

Motivated by the utility of codes with sparse generator matrices in the first problem, in the second part, we study the design of codes with a certain constraint on the weights of all the columns in the generator matrix (GM). In particular, we propose *polar-based* coding schemes to construct capacity-achieving codes, referred to as polar-DRS and polar-ADRS codes, for BEC and BMS channels respectively. We make significant contributions by demonstrating several properties of the codes including low encoding and decoding complexity, fast decay in error rate, and state-of-the-art upper bounds on the weights of the GM columns.

Next, we consider the critical problem of finding channel codes with large minimum distance. The large-minimum distance regime is of both theoretical and practical interests, as the transmission of information in many extreme conditions requires coding schemes that are highly robust with small or moderate rate. We present two novel construction of BCH-like cyclic codes and lower bounds on the maximal size of codes in the targeted regime. We also give asymptotic upper bounds that are stricter than all prior known bounds, which is proved by combining a new approach to bound the *maximal eigenvalues* of adjacency matrix induced by a Hamming ball with harmonic analysis on the Hamming cube as a group.

We then study the *one-shot* channel coding problem over classical and classical-quantum channels, where the underlying codes are constrained to be *group codes*. In the achievability part, we introduce a new distribution that incorporates the encoding homomorphism and the underlying channel law. Using a random coding argument, we characterize the performance in terms of a single-letter relative-entropy type quantity. In the converse part, we establish bounds by leveraging a hypothesis testing-based approach.

CHAPTER 1

Introduction

1.1 Motivation and Background

The subject of coding for data transmission have been extensively studied since the publication of Shannon's seminal work [126] in 1948, which showed that, given a noisy communication channel, there is a quantity, known as the capacity of the channel, such that reliable communication can be achieved at any rate below the channel capacity, if proper encoding and decoding techniques are used. In barely more than 70 years, coding theory has seen phenomenal growth. It is now a field with broad applications in a multitude of communication and storage applications. While Shannon [126] settled the question "Do good codes exist?" in the affirmative, two other questions naturally follow: "How can we find such codes?" and "How can we decode them?". In the effort to find good codes for practical purposes, researchers have studied not only various block codes, *e.g.*, the Hamming codes, the BCH codes, Reed-Solomon codes, and the low-density-parity-heck(LDPC) codes, but also codes in other paradigms, such as convolutional codes, turbo codes, and even quantum codes. Polar codes, discovered by Erdal Arıkan in 2009 [13] using a concept called *channel polarization*, have been the subject of active research in the past decade. They are the first class of provably capacity-achieving codes with explicit construction and low complexity of encoding and decoding. While the polar codes provide a solution to the two questions about code construction and decoding, the search of codes that meet requirements other than the rate and decoding complexity remains a vibrant realm, from both the theoretical and practical points of view. Inspired by real-world applications, we study codes with constraints on the sparsity of the generator matrices and/or parity-check matrices, the minimum distance, or the underlying algebraic structure in this dissertation.

1.2 Codes for Crowdsourced Label Learning

From personal assistants on mobile devices such as Siri and Alexa, to the analysis of genome sequencing data sets [85], to spam detectors in email servers, machine learning (ML)-based systems have become ubiquitous in our daily lives. ML owes its success not only to the improvements in computing power but also to the development of big data technology. Specifically, driving these systems are ML models that must be trained on data sets labeled according to target concepts (e.g., speech labeled by their corresponding commands, genome sequences labeled by the associated diseases, emails labeled as malicious or not). Additional application include activity recognition which labels videos with activities, and sentiment annotation that labels texts with sentimental markers. Although such tasks are easy for humans, they are in general difficult for the computers. How, then, can we provide a sufficiently large number of labels for the representative objects to train these learning models? Crowdsourcing systems, e.g., Amazon's Mechanical Turk and CloudResearch, have emerged as an effective large-scale human-powered platform [144] for the labelling problem. Since most crowdsourcing tasks are labeled by amateur workers instead of domain experts and are assigned with low reward (a few cents per task), the labels are often noisy and hence unreliable. To improve the quality of labels for subsequent use, it is essential to utilize appropriate inference algorithms (e.g. Majority voting). In this dissertation, we model the crowdsourced labelling problem as a coding problem, and aim to obtain the best possible trade-off between reliability and redundancy. Specifically, we leverage the duality between source coding and channel coding problems to provide querying schemes with *almost* optimal number of queries, each of which involving only a constant number of items when the workers are perfectly reliable. A similar result, where the number of items in each query scales at most logarithmically with the the total number of items, is shown when the workers are not always responsive.

1.3 Codes Defined by Sparse Generator Matrices

Capacity-approaching error-correcting codes such as low-density parity-check (LDPC) codes [50] and polar codes [13] have been extensively studied for applications in wireless and storage systems. Besides conventional applications of codes for error correction, a surge of new applications has also emerged in the past decade including crowdsourcing [74, 139], distributed storage [34], and speeding up distributed machine learning [84, 67]. To this end, new motivations have arisen to study codes with sparsity constraints on their generator and/or parity-check matrices. For instance, the stored data in a failed server needs to be recovered by downloading data from a few servers only, due to bandwidth constraints, imposing sparsity constraints in the decoding process in a distributed storage system. In crowdsourcing applications, e.g., when workers are asked to label items in a dataset, each worker can be assigned only a few items due to capability limitations, imposing sparsity constraints in the encoding process. More specifically, codes defined by sparse generator matrices become relevant for such applications [100, 108]. In this dissertation, we focus on polar codes in order to construct a sequence of codes defined by sparse generator matrices with practical utility, such as low decoding complexity, explicit construction, sufficiently fast decay in the error probability, and the potential to approach capacity at large block-length.

1.4 Codes with Large Minimum Distance

Let $A(n, d)$ denote the maximum size of a binary code of length n and minimum Hamming distance d. Studying $A(n, d)$, including efforts to determine it as well to derive bounds on $A(n, d)$ for large n's, is one of the most fundamental subjects in coding theory. In this dissertation, we explore new lower and upper bounds on $A(n, d)$ in the large-minimum distance regime, in particular, when $d = n/2 - \Omega(\sqrt{n})$. We first provide a new construction of cyclic codes, by carefully selecting specific roots in the binary extension field for the check polynomial, with length $n = 2^m - 1$, distance $d \geqslant n/2 - 2^{c-1}\sqrt{n}$, and size $n^{c+1/2}$, for any $m \geqslant 4$ and any integer c with $0 \leqslant c \leqslant m/2 - 1$. These code parameters are slightly worse than those of the Delsarte–Goethals (DG) codes that provide the previously known best lower bound in the large-minimum distance regime. However, using a similar and extended code construction technique we show a sequence of cyclic codes that improve upon DG codes and provide the best lower bound in a narrower range of the minimum distance d, in particular, when $d = n/2 - \Omega(n^{2/3})$. Furthermore, by leveraging a Fourier-analytic view of Delsarte's linear program, upper bounds on $A(n, \lceil n/2 - \rho\sqrt{n}\,\rceil)$ with $\rho \in (0.5, 9.5)$ are obtained that scale polynomially in n. To the best of authors' knowledge, the upper bound due to Barg and Nogin [20] is the only previously known upper bound that scale polynomially in n in this regime. We numerically demonstrate that our upper bound improves upon the Barg-Nogin upper bound in the specified high-minimum distance regime.

1.5 Codes with Abelian Group Structure

The results in the early development of coding theory and discussed in most standard coding theory textbooks often assume a finite field expression for the discrete input and/or output alphabets for a channel. As a finite field exists only with a prime power size, such assumption were extended to weaker algebraic structures such as rings and groups in later works [45, 28, 88, 89, 52, 121]. The motivation for studying Abelian group codes beyond the non-existence of finite fields over arbitrary alphabet size is that the algebraic structure of the code imposes certain restrictions on the performance. Specifically, in certain problems, linear codes were shown to be suboptimal [81] compared to Abelian group codes. For example, when phase shift keying (PSK) is the modulation scheme, codes over a group whose size matches the order of the PSK may outperform binary linear codes [88]. For another example, consider a distributed source coding problem with two statistically correlated but individually uniform quaternary sources X and Y that are related via the relation $X = Y + Z$, where $+$ denotes addition modulo-4 and Z is a hidden quaternary random variable that has a non-uniform distribution and is independent of Y. The joint decoder wishes to reconstruct Z losslessly. In this problem, codes over $\mathbb{Z}_4$ perform better than linear codes over the Galois field of size 4. Hence information-theoretic characterizations of the performance of Abelian group code ensembles for various communication problems and under various decoding constraints became important.

1.6 Thesis Overview

In this thesis, we consider four different problems described in separate chapters. In Chapter 2, our work "Coding for crowdsourced classification with XOR queries" [108] is presented. In this chapter, the crowdsourced labelling problem for the binary label case is shown to be equivalent to a source coding problem, when the workers are always reliable, and a source and channel coding problem when they may be unresponsive or return faulty answers. We provided two querying schemes where the number of queries are ϵ-close to the information theoretic lower bound by exploiting the trade-off between sparsity and probability of error for LDPC and LDGM codes. In Chapter 3, we present our work "Capacity-achieving Polar-based Codes with Sparsity Constraints on the Generator Matrices" [111]. This work provides our results concerning constructions for linear block codes defined by sparse generator matrices. The novel constructions, inspired by the polar codes, also inherit many of the properties including the speed of decay in error probability, the ability to achieve capacity at large block lengths, and low encoding and decoding complexities. In Chapter 4, our work "New Bounds on the Size of Binary Codes with Large Minimum Distance" [112] (see also [110]) is presented. In this chapter, we consider the problem of bounding the maximal size of binary codes with large minimum distance. In the considered regime, the size scales sub-exponentially in the block-length and renders most known bounds on the rate for a given relative distance ineffective. We provide two families of lower bounds on the size by giving explicit cyclic code constructions. A sequence of upper bounds is also shown via a new bounding technique motivated by a Fourier-analytical proof of the MRRW bound [106]. In Chapter 5, we present our work "Abelian Group Codes for Classical and C-Q Channel Coding: one-shot rate bounds" [113]. In this work, we consider the one-shot channel coding problem for both classical and classical-quantum channels. Additionally, the code is required to be a (shifted) group code, i.e., the encoding is realized by a group homomorphism with a shift vector. Characterization of the rate are provided in terms of a relative-type quantity known as the *hypothesis testing relative entropy*.

CHAPTER 2

Coding for Crowdsourced Classification with XOR Queries

This chapter presents a model for the crowdsourced labeling/ classification problem as a *sparsely encoded* source coding problem, where each query answer, regarded as a code bit, is the XOR of a small number of labels, as source information bits. We leverage the connections between this problem and well-studied codes with sparse representations for the channel coding problem to provide querying schemes with *almost* optimal number of queries, each of which involving only a constant number of labels. We also extend this scenario to the case where some workers can be unresponsive. For this case, we propose querying schemes where each query involves only $\log n$ items, where n is the total number of items to be labeled. Furthermore, we consider classification of two correlated labeling systems and provide *two-stage* querying schemes with *almost* optimal number of queries each involving a constant number of labels.

2.1 Introduction

2.1.1 Crowdsourcing for Classification

Crowdsourcing is a human-based problem-solving mechanism that allows a large crowd to distributively handle a massive number of queries. These problems, such as image classification, video annotation, form data entry, optical character recognition, translation, recommendation, and proofreading [75, 140], typically require human involvement or suit human better than machines [136]. In crowdsourcing systems, there are usually platforms, such as Amazon Mechanical Turk, CrowdFlower, and Figure Eight, that match the taskmaster to a huge worker crowd. However, the workers may be unreliable for several reasons: the reward for each task is usually as small as a few cents, the tasks are tedious, and one can still collect his rewards even if his answer is incorrect.

Many crowdsourcing-based real-life problems have one common goal deep down: classification/labeling of the items [69]. Formally, the label learning problem can be defined as follows: suppose there are n items, and the i-th item has a label $X_i \in \{0, 1, ..., L-1\}$, for $i \in \{1, 2, ..., n\}$. The goal is to identify the labels of the items. This problem is equivalent to clustering n items into L clusters with ground truth. Using crowdsourcing, a *taskmaster* can send queries to workers (sometimes called *oracles*, *human annotators* , or *labelers*). For instance, *same cluster* queries are adopted in [16], where in each query two items u and v are sent to a worker and the worker is asked "do u and v belong to the same label cluster?" In [75] and [76], single-item queries are considered, where the worker receives an item for each query and answers a question, such as "This is a picture of a dog, true or false?". In another related work, each item is associated with a certain number of properties and workers are imperfect, and the goal is to identify an item by querying the workers the properties [138]. In general, in all these scenarios, the objective is to minimize the number of queries, for a given type of query, sent to the workers, while being able to recover the labels with certain reliability/fidelity constraints. Note that, in practice, a certain number of queries can be assigned to one worker as a task, however, in the context of this thesis, the goal is to minimize the number of queries regardless of how many workers are involved.

2.1.2 Our Contributions

In this chapter we consider the 2-cluster case, i.e., $L = 2$, together with XOR queries, similar to the model adopted in [101]. Unlike several prior works which considered queries involving only 1 or 2 items [16, 75], we consider a generalized scenario, as considered in [101], where the number of items involved in a query can be more than 2 and up to a certain threshold. Furthermore, besides the scenario where answers to queries are perfect, we consider an extension where some of the workers may not answer the queries assigned to them. We first show that, in the scenario with perfect answers, the proof of [101, Theorem 1] is incorrect, with the random ensemble adopted therein, and hence fails to support the theorem. Note that the existential claim of the theorem may still be true if an alternative proof is given. We show that, with perfect workers, there exists a querying scheme within distance ϵ from the theoretic lower bound, in terms of the number of queries normalized by n, with up to $\log \frac{1}{\epsilon}$ items in each query. A similar result is shown for the scenario with unanswered queries and with up to $\log n \log \frac{1}{\epsilon}$ items in each query.

We further extend the problem to the scenario when items need to be clustered according to two different clustering criteria. Ideas from correlated-source and channel coding allow us to recover two types of clustering with less queries than when the two types of labeling are recovered separately. We show that a querying scheme within distance 2ϵ from the theoretic lower bound, in terms of the number of queries normalized by n, is achievable with $\log \frac{1}{\epsilon}$ items per query.

2.2 Preliminaries

2.2.1 Crowdsourcing and Linear Channel Codes

From crowdsourcing's perspective, when XOR query scheme is adopted, the process is similar to a linear source coding problem. In other words, the output of a query is the XOR of binary labels, i.e., addition over the binary field $\mathbb{F}_2$, of all items involved in that query. In particular, let n be the total number of items to be labeled and $\mathbf{X} = (X_1, X_2, \ldots, X_n)^t \in \mathcal{X}^n$, where $\mathcal{X} = \{0, 1\}$, considered as a column vector, denote the true labels, unknown to the taskmaster. Let also $A \in \mathbb{F}_2^{m \times n}$ denote the *query matrix*, where ones in the i-th row of A correspond to the indices of items in the i-th query. Under the XOR model for the query answers, the correct answer to the i-th query is equal to the i-th element of $A\mathbf{X}$, where all operations are over $\mathbb{F}_2$, i.e., X_i's are regarded as elements of $\mathbb{F}_2$. Given the limited capabilities of human workers, the queries need to be designed to be sparse. More specifically, the number of items per query, which is equal to the number of ones in each row of A, must be small, e.g., bounded by a constant value.

It is common to assume the apriori distribution of the labels are known to the taskmaster [75, 102]. In particular, for the binary label model adopted in this chapter, an i.i.d. Ber(p) distribution is assumed for the labels, where p is known to the taskmaster. Note that due to the duality between source coding and channel coding problems, the parity-check matrix of a linear code $\mathcal{C}$ designed for transmission over a memoryless binary symmetric channel (BSC) with transition probability p, BSC(p), can be used to compress an i.i.d. Ber(p) source. In fact, the probability of error of the maximum-likelihood (ML) decoder would be the same when the code is used either for channel coding across BSC(p) or source coding of an i.i.d. Ber(p) source. Hence, low-density parity-check (LDPC) codes become relevant for the considered crowdsourced clustering problem with XOR query scheme and bounded number of items in each query.

2.2.2 LDPC and LDGM Codes

LDPC codes were originally introduced by Gallager [49] in 60's and were later rediscovered in 90's [92] and were shown to offer near-capacity performance under practical belief-propagation (BP) decoding algorithms. In [49], Gallager proved that *right-regular* LDPC codes, where each row of the parity-check matrix H has a constant weight Δ, can not achieve the channel capacity on BSC. He also showed that the gap to capacity diminishes exponentially fast with Δ. In [119] and [124], similar results are shown when the row weights and average row weights of H are upper bounded by Δ, respectively. In terms of code constructions, several works on regular LDPC codes have shown that rates close to the provided upper bounds are attainable [124, 23, 24]. In section 2.3.1, we leverage results from [124] and [82] and propose a scheme with sparse queries in the context of crowdsourced classification.

When considering unreliable workers in the crowdsourcing setup another class of codes with sparse representations, namely low-density generator-matrix (LDGM) codes, become relevant. In LDGM code, the generator matrix is assumed to be sparse. In general, as opposed to LDPC codes, the performance of LDGM codes has not been very well-studied in the literature. Several works have empirically shown the existence of LDGM codes with close-to-capacity performance [146, 137, 43]. It is shown in [73] that ensembles of LDGM codes are capacity achieving over BSCs when the row weights scale linearly with n. Furthermore, row weights of $O(\log n)$ suffice to achieve the capacity of binary erasure channels (BECs) [73]. Scaling exponent of such codes were studied in [95]. In our works [109, 111], capacity-achieving LDGM codes with sublinear bounds on the row weights are constructed for the transimission over BSC and general BMS channels. Furthermore, the codes allow for low complexity encoding and decoding and have the same rate of decay of the error probability as the standard polar codes.

2.2.3 Prior Work

In [101], three scenarios, namely noiseless queries and exact recovery, noiseless queries and approximate recovery, and noisy queries and approximate recovery, are discussed. Here, a query is noisy if the workers are unreliable, i.e., the query answers may be inaccurate, and is noiseless if it is always correct. A recovery is assumed to be *perfect* if the (block) error probability vanishes as the number of items grows large, and is considered to be approximate if up to a constant probability of error in recovering labels/source bits is allowed.

In [101], Mazumdar and Pal attempt to show that, under the XOR query model with noiseless answers, perfect recovery of all labels (i.e. source compression) is achievable with sparsely encoded source coding, when the number of items in each query is bounded by a constant value Δ. Note that in several other prior works, queries involving $\Delta = 1, 2, 3$ items were considered [75, 148, 76, 16, 83, 141].

LDPC codes have been considered for source compression in [26] and [99]. In particular, it is pointed out in [26] that the analogy between linear source codes and LDPC channel codes was largely neglected in the literature and it is shown that LDPC-based data compression for either memoryless sources or sources with memory are practical.

2.3 Main Results

We discuss in this section strategies for perfect recovery for the scenario where all workers are perfect, and also the case with possibly having unresponsive workers. We further study the problem of clustering the items based on two correlated classification criteria. The main results of this chapter are stated in this section. The proofs of Propositions 1 to 5 can be found in Section 2.4.

2.3.1 Noiseless Queries and Perfect Recovery

Problem Formulation: We adopt the same model as in [101, Section 3.1]. Consider the crowdsourced classification problem where there are only two label types for each item and XOR queries are adopted, as discussed in Section 2.2.1. Suppose that the labels $X_i \in \mathcal{X} = \{0, 1\}$, for $1 \leqslant i \leqslant n$, are i.i.d. $\mathrm{Ber}(p)$ random variables. Without loss of generality, we may assume $p \in (0, 0.5)$. In this subsection, we assume a nonadaptive/one-shot scenario, in which all queries are generated and sent to the workers at the same time. This is in accordance with several prior works, e.g. [75, 83].

In [101, subsection 3.1], the number of items in each query given to workers is fixed to Δ, and the worker returns an error-free XOR of the labels in the given query. Let $H_b(p)$ denote the binary entropy function. The following result is claimed in [101].

Theorem 1. *[101, Theorem 1] There exists a querying scheme with*

$$m = \frac{n(H_b(p) + o(1))}{\log_2 \frac{1}{\alpha}}$$

queries, of above type, where $\alpha \overset{\text{def}}{=} \frac{1}{2}\left[1 + (1 - 4p(1-p))^{\Delta}\right]$*, that achieves perfect recovery.*

The proof provided in [101] is based on the following. The average probability of error of a randomly chosen query scheme, consisting of m independently and uniformly selected random queries involving exactly Δ items, is analyzed and is claimed that it approaches 0 as n grows large. However, we show here that the provided proof does not hold. In particular, we show in Proposition 1 that the average probability of error over the considered ensemble is bounded away from 0.

Proposition 2. *The average probability of error, denoted by P_e, in a scheme with*

$$m = \frac{n(H_b(p) + o(1))}{log\frac{1}{\alpha}}$$

independent and uniformly distributed random queries involving Δ items does not vanish as n, the number of items, grows . More precisely,

$$P_e \geqslant (1-\epsilon)\left[\exp\left(-\frac{\Delta \cdot H_b(p)}{log\frac{1}{\alpha}}\right) - \epsilon'\right] > 0, \tag{2.1}$$

where $\epsilon, \epsilon' > 0$ can be chosen arbitrarily small as $n \to \infty$

Proof: The proof is given in Section 2.4.1. ■

Remark 1. In [10, Theorem 2], the authors consider uniformly random queries, essentially a random ensemble same as in Proposition 2, and show that $m = \Theta(n \log n)$ queries is necessary and sufficient for perfect recovery regardless of whether the workers are perfect or not, when no apriori distribution is assumed.

This does not imply that the theorem itself, [101, Theorem 1], does not hold. Note also that such results are not about specific constructions and state that the average probability of error of certain random ensemble is bounded away from 0. In fact, there are trivial cases of a query matrix providing perfect recovery with $m = n$ queries, e.g., the identity matrix. However, the question of whether less than n queries, and more specifically close to $nH_b(p)$ queries, is sufficient for perfect recovery or not is not properly answered by [101] and has not been considered in [10]. We answer this question in Proposition 3.

Proposition 3. *Suppose that workers are perfect and labels have prior distribution $Ber(p)$. Then, for $\epsilon \in (0,1)$ and sufficiently large n, there exists a querying scheme using*

$$m = n[H_b(p) + \epsilon(1 - H_b(p))] \tag{2.2}$$

queries, each involving no more than

$$\frac{K_1 - K_2 ln(\epsilon)}{1-\epsilon}(H_b(p)^{-1} - 1)$$

items, that achieves perfect recovery, where K_1 and K_2 depend only on p.

Proof: The proof is given in Section 2.4.2. ■

By Shannon's source coding theorem we need at least $nH_b(p)$ queries to achieve perfect recovery, when the query answers are binary. Hence, the compression rate, i.e., the number of queries normalized by the number of items, must be at least $H_b(p)$. In Proposition 3, it is shown that for any chosen $\epsilon \in (0,1)$, compression rate as small as $H_b(p) + \epsilon(1 - H_b(p))$ can be achieved by a query scheme, with $O(\log \frac{1}{\epsilon})$ items per query, thereby providing a scheme with *almost* optimal number of queries.

2.3.2 Two-label Perfect Recovery

In this subsection, we extend the problem discussed in Section 2.3.1 to the recovery of two different clustering based on two properties/labeling criteria associated with the same set of objects. In particular, suppose that the i-th item can be classified according to two (binary) labeling systems and is labeled X_i, Y_i, respectively. For instance, one labeling system may involve identifying the objects in pictures, e.g., whether there is a cat or dog in the picture, and the other may involve identifying the location, e.g., whether this picture is taken indoor or outdoor. Furthermore, we assume that a two-stage query scheme is adopted, where queries involving X_i's are sent in the first stage and queries involving Y_i's are sent in the second stage.

By leveraging the correlation between labels X and Y, we can recover the clustering by sending an almost optimal number of queries, under the constraint of finite items per query.

Proposition 4. *Let $(X_i, Y_i) \overset{i.i.d.}{\sim} P_{X,Y}(x,y)$, for $1 \leqslant i \leqslant n$. Then there exists a two-stage querying scheme using*

$$m = n(H(X,Y) + 2\epsilon) \tag{2.3}$$

queries, where $H(X,Y)$ is the joint entropy function, each involving no more than $O(\log \frac{1}{\epsilon})$ items, that achieves perfect recovery for sufficiently large n.

Proof: The proof is given in Section 2.4.3. ■

2.3.3 Noisy Queries and Perfect Recovery

In this section, we consider imperfection in workers' replies. Due to the monotonicity of queries, low payment for completion of queries, or the fact that workers may not be experts, two noisy scenarios for the answers to queries can emerge. In the first case, it is assumed that some queries are not replied within a certain specified response time, or not replied at all. This scenario can be modeled as follows: each query is replied (correctly) with probability $1 - r$ and is not replied with probability r, for a certain parameter r, independent from other queries. Consequently, this scenario becomes related to the channel coding problem over the binary erasure channel (BEC) with erasure probability r. In the second case, it is assumed that the workers have accuracy $1-q$, that is, the answer from a worker is correct with probability $1-q$. This is related to the channel coding problem over BSC(q).

In this chapter, we focus on the first case and leave the second case for future work. In particular, we show that concatenation of LDGM and LDPC codes achieve compression rate

$$R = [H_b(p) + \epsilon(1 - H_b(p))]/(1-r),$$

with row weights upper bounded by $\Delta = O(\log \frac{1}{\epsilon} \log n)$.

Let $\mathcal{A}_{N\times K}$ denote the set of $N \times K$ binary matrices and let $B(\mathcal{A}_{N\times K}, p)$ denote a distribution on $\mathcal{A}_{N\times K}$, where the entries of a random $A \sim B(\mathcal{A}_{N\times K}, p)$ are distributed i.i.d. with Ber(p). We utilize the following result from [73] to derive the main result of this section.

Theorem 5. *[73, Theorem 5] Consider BEC(r) and let $K = NR$, where $R < 1-r$. Suppose $A \sim B(\mathcal{A}_{N\times K}, \rho(N))$, where $\rho(N) = \Theta(\frac{\log N}{N})$. Let $p_c(A)$ denote the probability of correct decoding, under ML, using A^t as the generator matrix and assuming transmissions over BEC(r). Then*

$$\lim_{n\to\infty} \mathbb{E}_A(p_c(A)) = 1, \tag{2.4}$$

where the expected value is taken with respect to A.

Note that the generation of $A \sim B(\mathcal{A}_{N\times K}, \rho(N))$ does not guarantee that all row weights are bounded by $\log N$. We extend Theorem 5, in order to ensure that all row weights are bounded, in Proposition 6.

Proposition 6. *Let $A \in \mathcal{A}_{N\times K}$ and let A^t be the generating matrix corresponding to a code of rate $R = \frac{K}{N} < 1-r$ with transmissions over BEC(r). For any $\rho(N) = \Theta(\frac{\log N}{N})$, the expected value of $p_c(A)$ over all matrices with $\tilde{B}(\mathcal{A}_{N\times K}, \rho(N))$ distribution tends to 1 as N approaches infinity, i.e.,*

$$\lim_{N\to\infty} \mathbb{E}_{A\sim\tilde{B}}(p_c(A)) = 1, \tag{2.5}$$

where $\tilde{B}(\mathcal{A}_{N\times K}, \rho(N))$ is obtained from the distribution $B(\mathcal{A}_{N\times K}, \rho(N)))$ by removing matrices that have at least one row with Hamming weight larger than or equal to $\Theta(\log N)$.

Proof: The proof is given in Section 2.4.4. ■

Theorem 7. *Suppose that queries are answered with probability $1-r$, independent to each other. Then there exists a query scheme with*

$$m = n[H_b(p) + \epsilon(1 - H_b(p))]/(1-r) \tag{2.6}$$

queries, each involving $O(\log \frac{1}{\epsilon} \log n)$ items, that guarantees perfect recovery of the labels as n grows large.

Proof: The following statements holds for sufficiently large n. The parity-check matrix of LDPC codes is applied first for the compression. By Proposition 3 there exists an $m_H(n) \times n$ binary matrix H_n, where $m_H(n) = n[H_b(p) + \epsilon(1 - H_b(p))]$, with row weights bounded from above by $\left(H_b(p)^{-1} - 1\right) \frac{K_1 - K_2 \ln \frac{1}{\epsilon}}{1-\epsilon}$. Note that H_n can be used to compress an i.i.d. $Ber(p)$ source sequence with perfect recovery.

Then, the compressed bits are further encoded to recover from erasures caused by the BEC(r). By Proposition 6, there exists a $m_G(n) \times m_H(n)$ matrix G_n, with all row weights being $O(\log m_H(n))$, for transmission of $m_H(n)$ bits over BEC(r), while the expected probability of correct decoding approaches 1 as $n \to \infty$ for any $R_c \overset{\text{def}}{=} m_H(n)/m_G(n) < 1 - r$.

Next, the two linear operations, one for the compression and the other one for the erasure correction, are concatenated, leading to the overall query matrix. Let $G^{SC}(n) \overset{\text{def}}{=} G_n H_n$, with dimensions $m_G(n) \times n$, denote the corresponding overall encoding matrix. Note that all row weights of $G^{SC}(n)$ are bounded from above by

$$\frac{K_1 - K_2 \ln \frac{1}{\epsilon}}{1 - \epsilon} \left(H_b(p)^{-1} - 1\right) O(\log m_H(n)) = O\left(\log \frac{1}{\epsilon} \log n\right),$$

and also the number of queries is equal to $m_G(n)$, where

$$m_G(n) > \frac{m_H(n)}{1 - r} = \frac{n[H_b(p) + \epsilon(1 - H_b(p))]}{1 - r},$$

is sufficient to show that $G^{SC}(n)$ guarantees perfect recovery of the labels as discussed next. Let $\mathbf{X}$ denote the vector of n labels, $\mathbf{Y} = H_n\mathbf{X}$ denote the compressed labels, and $\mathbf{Z} = G_n\mathbf{Y} = G^{SC}(n)\mathbf{X}$ and the taskmaster collects $\mathbf{Z}$ corrupted with erasures with probability r. The taskmaster can recover $\mathbf{Y}$, with high probability, by the choice of G_n. Then, having recovered $\mathbf{Y}$, perfect recovery of $\mathbf{X}$ is possible by the choice of H_n. That completes the proof of theorem. ■

2.4 Proofs

2.4.1 Proof of Proposition 2

We analyze the average probability of error, P_e, of the querying scheme given in [101, Theorem 1], and provide a lower bound for it. Consider the following two typical sets in $\{0,1\}^n$,

$$A_\epsilon^n(X) \overset{\text{def}}{=} \{x^n : np(1 - n^{-\frac{1}{3}}) \leqslant w_H(x^n) \leqslant np(1 + n^{-\frac{1}{3}})\},$$

$$B_\epsilon^n(X) \overset{\text{def}}{=} \{x^n : np(1 - n^{-\frac{1}{3}}) + 1 \leqslant w_H(x^n) \leqslant np(1 + n^{-\frac{1}{3}}) - 1\},$$

where $w_H(x^n)$ denotes the hamming weight of vector x^n. Then, we have

$$\Pr(A_\epsilon^n(X)) \to 1 \text{ and } \Pr(B_\epsilon^n(X)) \to 1 \text{ as } n \to \infty.$$

Note that for $x^n \in B_\epsilon^n(X)$,

$$np(1 - n^{-\frac{1}{3}}) \leqslant w_H(x^n + e) \leqslant np(1 + n^{-\frac{1}{3}}),$$

for any $e \in \{0,1\}^n$ with unit weight, i.e., $w_H(e) = 1$. Hence, $x^n + e \in A_\epsilon^n(X)$, for any $x^n \in B_\epsilon^n(X)$. Then P_e can be expressed and lower bounded as follows:

$$\begin{aligned}
P_e &= \sum_{x^n \in \mathcal{X}^n} P_X^n(x^n) \Pr(\hat{x}^n \neq x^n) \\
&= \sum_{x^n \in \mathcal{X}^n} P_X^n(x^n) \Pr(\exists \tilde{x}^n \neq x^n, \tilde{x}^n \in A_\epsilon^n(X), Qx^n = Q\tilde{x}^n) \\
&\geqslant \sum_{x^n \in B_\epsilon^n(X)} P_X^n(x^n) \Pr(\exists \tilde{x}^n \neq x^n, \tilde{x}^n \in A_\epsilon^n(X), Qx^n = Q\tilde{x}^n) \\
&\geqslant \sum_{x^n \in B_\epsilon^n(X)} P_X^n(x^n) \Pr(x^n + e_1 \in A_\epsilon^n(X), Qx^n = Q(x^n + e_1)),
\end{aligned}$$

where $e_1 = (1, 0, 0..., 0, 0)$ and Q is the $m \times n$ query matrix corresponding to the query scheme described in Proposition 1.

For $x^n \in B_\epsilon^n(X)$, we always have $x^n + e_1 \in A_\epsilon^n(X)$. Note that $Qx^n = Q(x^n + e_1)$ if and only if $Q(e_1) = 0$, which is also equivalent to the first label not being queried by any of the m queries. The probability that a query does not use the first label is $\frac{\binom{n-1}{\Delta}}{\binom{n}{\Delta}} = 1 - \frac{\Delta}{n}$. Thus,

$$\Pr\Big(Q(e_1) = 0\Big) = \left(1 - \frac{\Delta}{n}\right)^m,$$

where $m = \frac{n(H_b(p)+o(1))}{log\frac{1}{\alpha}}$ as in [101, Theorem 1]. Then we have

$$\Pr\Big(Q(e_1) = 0\Big) = \left[(1 - \frac{\Delta}{n})^n\right]^{\frac{H_b(p)+o(1)}{log\frac{1}{\alpha}}} \xrightarrow[n\to\infty]{} \exp\left(-\frac{\Delta \cdot H_b(p)}{log\frac{1}{\alpha}}\right) > 0.$$

Therefore,

$$P_e \geqslant \sum_{x^n \in B_\epsilon^n(X)} P_X^n(x^n) \left(1 - \frac{\Delta}{n}\right)^m \geqslant (1 - \epsilon)\left[\exp\left(-\frac{\Delta \cdot H_b(p)}{log\frac{1}{\alpha}}\right) - \epsilon'\right] \neq 0,$$

for n sufficiently large, where ϵ, ϵ' can be chosen arbitrarily small as $n \to \infty$.

2.4.2 Proof of Proposition 3

We leverage a result from [124], which establishes a connection between sparsity of the parity-check matrix and reliability performance of LDPC codes. First, the *density* of parity-check matrix of a linear code is defined as follows.

Definition 1. Given an $m \times n$ parity-check matrix H, the *density* of H, denoted by $\rho = \rho(H)$, is the number of ones in H normalized by n, i.e., the total number of ones in H is $(n-m)\rho$.

Theorem 8. *[124, Theorem 2.2] For any BSC or BEC, there exists a sequence of ensembles of regular LDPC codes which achieves, under ML decoding, a fraction $1-\epsilon$ of the channel capacity with vanishing block error probability, and the asymptotic density of their parity-check matrices satisfies*

$$\lim_{n\to\infty} \rho_n \leqslant \frac{K_1 - K_2 ln(\epsilon)}{1-\epsilon}, \tag{2.7}$$

where K_1 and K_2 depend only on the channel.

In particular, for $BSC(p)$ with capacity $1 - H_b(p)$, there exists an ensemble of regular $(n, n-m)$ LDPC codes achieving channel rate

$$R = \frac{n-m}{n} = (1-\epsilon)C = (1-\epsilon)(1-H_b(p)).$$

The parity check matrices from this ensemble have $m = n[H_b(p) + \epsilon(1-H_b(p))]$ rows, each with weight

$$\begin{aligned}(n-m)\rho_n/m &= \left([H_b(p) + \epsilon(1-H_b(p))]^{-1} - 1\right)\rho_n \\ &\leqslant (H_b(p)^{-1} - 1)\rho_n \\ &\leqslant (H_b(p)^{-1} - 1)\frac{K_1 - K_2 ln(\epsilon)}{1-\epsilon},\end{aligned} \tag{2.8}$$

for sufficiently large n.

As mentioned in Section 2.2.1, we note that the parity-check matrix of a linear code $\mathcal{C}$ designed for transmission over a BSC(p), can be used to compress an i.i.d. Ber(p) source with the same block error probability P_e under ML decoding.

2.4.3 Proof of Proposition 4

We use a two-stage querying scheme. Let $X_i \overset{i.i.d.}{\sim} Ber(p)$, $q = \Pr(Y_i = 1|X_i = 1)$ and $r = \Pr(Y_i = 1|X_i = 0)$. First, we use

$$m_1 = n\left[H_b(p) + \epsilon(1 - H_b(p))\right]$$

queries to retrieve X_i's. From Proposition 3, $(H_b(p)^{-1} - 1)\frac{K_1 - K_2 ln(\epsilon)}{1-\epsilon}$ items per query are sufficient for perfect recovery of the X_i's.

Second, we consider the conditional distribution of the Y_i labels. Suppose that X_i's are recovered correctly. Let $n_1 = \sum_{i=1}^{n} X_i$ and $n_2 = n - n_1$ denote the number of items with labels $X_i = 1$ and $X_i = 0$, respectively. By law of large numbers, for any given ϵ', we have $n_1 \leqslant np(1+\epsilon')$ and $n_2 \leqslant n(1-p)(1+\epsilon')$, with high probability, for sufficiently large n. By Proposition 3, there exists a querying scheme with

$$m_2 = n_1[H_b(q) + \epsilon(1 - H_b(q))]$$

queries, each involving no more than $(H_b(q)^{-1} - 1)\frac{K_1' - K_2' ln(\epsilon)}{1-\epsilon}$ items that recovers the Y_i labels for the items labeled $X_i = 1$. Also, there exists a querying scheme with

$$m_3 = n_2[H_b(r) + \epsilon(1 - H_b(r))]$$

queries, each involving no more than $(H_b(r)^{-1} - 1)\frac{K_1'' - K_2'' ln(\epsilon)}{1-\epsilon}$ items that recovers the Y_i labels for the items labeled $X_i = 0$. The total number of queries is then

$$\begin{aligned} m_1 + m_2 + m_3 &\leqslant n[H_b(p) + pH_b(q) + (1-p)H_b(r)] + n\epsilon(1 + p + (1-p)) \\ &\quad - n[\epsilon H_b(p) + p\ \epsilon H_b(q) + (1-p)\epsilon H_b(r)) - \epsilon'] \\ &\leqslant nH(X,Y) + 2n\epsilon. \end{aligned}$$

Note that, given p, q, r, the number of items involved in a query is related to the gap to capacity as $O(\log(\frac{1}{\epsilon}))$.

2.4.4 Proof of Proposition 6

The following lemma is a direct consequence of Chernoff bound and the proof is omitted here:

Lemma 9. *Let $X_1, X_2, ..., X_N$ be N independent random variables with $X_i \sim Ber(p_i)$. Let*

$\mu = \sum_{i=1}^{N} p_i$. *Then*

$$\Pr\left(\sum\nolimits_{i=1}^{N} X_i \geqslant (1+\delta)\mu\right) \leqslant e^{-\frac{\delta^2}{2+\delta}\mu}, \quad \forall \delta > 0. \tag{2.9}$$

For any $\rho(N) = \Theta(\frac{\log N}{N})$, there exist $M > 1, m > 0$ such that $m\frac{\log N}{N} \leqslant \rho(N) \leqslant M\frac{\log N}{N}$ for N sufficiently large. If $p = p_1 = p_2 = ... = p_N = \rho(N)$, we have

$$m \log N \leqslant \mu = Np \leqslant M \log N.$$

Then,

$$\Pr\left(\sum_{i=1}^{N} X_i \geqslant \delta(m) M \log N\right) \leqslant \Pr\left(\sum_{i=1}^{N} X_i \geqslant \delta(m)\mu\right) \leqslant e^{-2\log N} = N^{-2},$$

where $\delta(m) > 1$ is chosen such that $\frac{(\delta(m)-1)^2}{2+\delta(m)-1} m > 2$.

Next, we discuss the probability that a matrix $A \sim B(\mathcal{A}_{N\times K}, \rho(N))$ has *heavy* rows, where a heavy row is a row with weight larger or equal to $\delta(m)M\log N$.

$$\begin{aligned}
&\Pr(\text{each row of } A \text{ has weight less than } \delta(m)M\log N)\\
&= 1 - \Pr\left(\bigcup_{i=1}^{N} \left\{i^{th} \text{ row has weight } \geqslant \delta(m)M\log N\right\}\right)\\
&\geqslant 1 - \sum_{i=1}^{N} \Pr\left(i^{th} \text{ row has weight } \geqslant \delta(m)M\log N\right)\\
&= 1 - N \cdot \Pr(1^{st} \text{ row has weight } \geqslant \delta(m)M\log N)\\
&\geqslant 1 - N \cdot N^{-2},
\end{aligned}$$

which goes to 1 as N grows large. Let $p_h(N)$ denote the probability that A has at least one heavy row. Then we have $p_h(N) \to 0$. Note that

$$\mathbb{E}_A(p_c(A)) = p_h(N)\mathbb{E}_A(p_c(A)) + (1 - p_h(N))\mathbb{E}_{A\sim\tilde{B}}(p_c(A)),$$

where the expectations are taken with $B(\mathcal{A}_{N\times K}, \rho(N))$ distribution, $B(\mathcal{A}_{N\times K}, \rho(N))$ distribution and heavy-row condition, and $\tilde{B}(\mathcal{A}_{N\times K}, \rho(N))$ distribution.

From [73, theorem 5],

$$
\begin{aligned}
1 &= \lim_{N\to\infty} \mathbb{E}_{A\in\mathcal{A}_{N\times K}}(p_c(A)) \\
&= \lim_{N\to\infty} [p_h(N)\mathbb{E}_A(p_c(A)) + (1-p_h(N))\mathbb{E}_{A\sim\tilde{B}}(p_c(A))] \\
&= \lim_{N\to\infty} \mathbb{E}_{A\sim\tilde{B}}(p_c(A))
\end{aligned}
$$

2.5 Conclusion and Outlook

In this chapter, crowdsourced classification problems with XOR querying schemes involving a limited number of items in each query sent to the crowd workers are considered. The goal is to perform the classification efficiently, that is, to minimize the number of queries sent to workers. We discuss the scenario where all workers are perfect, and then extend to the case with possibly having unresponsive workers. We further consider clustering the items based on two correlated classification criteria. In all of the above cases, we provide querying schemes with almost optimal number of queries each with limited number of items.

There are several directions for future work. In this chapter, we focus on binary labels and generalizing the results to classification problems with more than two possible labels is an interesting direction. To this end, one needs to consider non-binary codes with sparse representations. When considering noisy queries, only the scenario with unresponsive workers are studied. As stated in Section 2.3.3, it is possible that workers do not provide the correct answer. In the binary labeling case, such scenario corresponds to coding over BSC. Hence, designs of *good* LDGM codes for transmission over BSCs are needed for such classification problems with unreliable workers, which is another direction for future work. Furthermore, alternative methods, such as a joint source-channel coding design, can be utilized and may lead to more efficient querying schemes involving unreliable/unresponsive workers.

CHAPTER 3

Capacity-achieving Polar-based Codes with Sparsity Constraints on the Generator Matrices

3.1 Introduction

In general, the generator matrix sparsity is a critical factor in determining the encoding complexity of a linear code. Further, certain applications, e.g., distributed crowdsourcing schemes utilizing linear codes, require most or even all the columns of the generator matrix to have some degree of sparsity. In this chapter, we leverage polar codes and the well-established channel polarization to design capacity-achieving codes with a certain constraint on the weights of all the columns in the generator matrix (GM) while having a low-complexity decoding algorithm. We first show that given a binary-input memoryless symmetric (BMS) channel W and a constant $s \in (0, 1]$, there exists a polarization kernel such that the corresponding polar code is capacity-achieving with the *rate of polarization* $s/2$, and the GM column weights being bounded from above by N^s. To improve the sparsity versus error rate trade-off, we devise a column-splitting algorithm and two coding schemes for BEC and then for general BMS channels. The *polar-based* codes generated by the two schemes inherit several fundamental properties of polar codes with the original 2×2 kernel including the decay in error probability, decoding complexity, and the capacity-achieving property. Furthermore, they demonstrate the additional property that their GM column weights are bounded from above sublinearly in N, while the original polar codes have some column weights that are linear in N. In particular, for any BEC and $\beta < 0.5$, the existence of a sequence of capacity-achieving polar-based codes where all the GM column weights are bounded from above by N^λ with $\lambda \approx 0.585$, and with the error probability bounded by $\mathcal{O}(2^{-N^{\beta}})$ under a decoder with complexity $\mathcal{O}(N \log N)$, is shown. The existence of similar capacity-achieving polar-based codes with the same decoding complexity is shown for any BMS channel and $\beta < 0.5$ with $\lambda \approx 0.631$.

3.1.1 Polar Codes

Channel polarization, introduced by Arıkan [13, 12], is one of the most recent breakthroughs in coding theory. Polar codes are a class of provably capacity-achieving channel codes with explicit construction for general BMS channels, and have attracted significant attention due to their error correction performance, as well as their low-complexity decoding algorithms. Within the ongoing fifth generation wireless systems (5G) standardization process, polar codes have been adopted for uplink and downlink control information for the enhanced mobile broadband (eMBB) communication service. Furthermore, polar codes and polarization phenomenon have been successfully applied to a wide range of problems including data compression [14, 5], broadcast channels [104, 55], multiple access channels [30, 97], physical layer security [98, 11], and coded modulations [96].

3.1.2 LDGM and Related Works

A related line of work on studying linear codes with sparsity constraints on their generator matrices is by associating them with sparse graph representations [27]. In this context, they are referred to as low-density generator matrix (LDGM) codes, also regarded as the counterpart of LDPC codes. The sparsity of the generator matrices of LDGM codes leads to a low encoding complexity, and has been adopted in applications such as lossy source compression [58] and multiple description coding [145]. In [92, 93] it was pointed out that certain constructions of LDGM codes are not asymptotically *good*, a behavior which is also studied using an error floor analysis in [147, 51].

In terms of the sparsity of the GM, the authors of [73] showed the existence of capacity-achieving codes over binary symmetric channels (BSC) using random linear coding arguments when the column weights of the GM are upper bounded by ϵN, for any $\epsilon > 0$, where N is the code block length. Also, it is conjectured in [73] that column weight upper bounds that scale sublinearly in N suffice to achieve the capacity. For binary erasure channels (BEC), bounds that scale as $\mathcal{O}(\log N)$ suffice for achieving the capacity, again using random linear coding arguments [73]. Furthermore, the scaling exponent of such random linear codes are studied in [95]. Later, in [86], the existence of capacity-achieving systematic LDGM ensembles over any BMS channel with the expected value of the weight of the entire GM bounded by ϵN^2, for any $\epsilon > 0$, is shown. While the (ensemble-averaged) block-error probability for the codes goes to zero as the block-length grows large, the speed of decay in the error probability is not provided in [73, 86].

In [100], the problem of label learning through queries from a crowd of workers was formulated as a coding theory problem. Due to practical constraints in such crowdsourcing scenarios, each query can only contain a small number of items. In [108], we considered the same setting as in [100] with the additional consideration that some workers may not respond to queries, a scenario that resembles a binary erasure channel. Then we showed that a combination of LDPC codes and LDGM codes gives a query scheme where the number of queries approaches the information-theoretic lower bound [108].

In the realm of quantum error correction, quantum low-density-generator-matrix (QLDGM) codes, quantum low-density-parity-check (QLDPC) codes, and other sparse-graph-based schemes have been extensively studied due to the small numbers of quantum interactions per qubit during the encoding and/or error correction procedure, avoiding additional quantum gate errors and facilitating fault-tolerant decoding. Amongst these schemes, the error correction performance of the LDGM-based codes proposed in [47] was shown to outperform all other Calderbank-Steane-Shor (CSS) and non-CSS codes of similar complexity.

In the realm of machine learning, gradient-based methods, such as the gradient descent (GD) algorithm, are one of the most commonly used algorithms to fit the machine learning models over the training data. Using LDGM codes, authors of [68] proposed a distributed implementation of a stochastic GD scheme, which allowed the master node to recover a "high-quality unbiased" estimate of the gradient at low computational cost and provided overall performance improvement over the GD scheme with gradient coding.

In all three applications highlighted above, the benefit of the LDGM codes follows from a certain upper bound on the column weights of the GM, ensuring the columns are relatively low weight. Motivated by these applications, the main goal of this chapter is to construct sequences of codes where all of the columns of the GM are *low weight*, where certain upper bounds on the weight will be specified later.

3.1.3 Our Contribution

In this chapter, we study capacity-achieving polar and polar-based codes over BMS channels with sparsity constraints on generator matrix column weights. Leveraging polar codes based on general kernels, with rates of polarization studied in [78], we show that capacity-achieving polar codes with column weights bounded from above by N^s exist for any given $s > 0$, where N is the code block length. This verifies the conjecture given in [73]. There is, however, a trade-off between the sparsity parameter s and the rate of polarization, given by $\frac{s}{2}$.

For the case when the speed of decay for block-error probability and the GM sparsity are both constrained, we propose two new code constructions with sparse GM columns, which provide a better trade-off for $s > 0.585$. We first consider BEC, and propose a splitting algorithm termed *decoder-respecting splitting (DRS)* algorithm, which, roughly speaking, splits *heavy* columns in the GM into several *light* columns. Note that if one splits the heavy columns in an arbitrary manner to form a new GM, the code defined by the new GM may be substantially different from the original one in terms of the error probability and/or having a low-complexity decoder. Leveraging the fact that the polarization transform of a BEC leads to BECs, the DRS algorithm converts the encoder of a standard polar code into an encoder defined by a sparse GM without hurting the reliability of the bit-channels observed by the source bits. Furthermore, the specific structure of DRS enables a low-complexity successive cancellation decoder in a recursive fashion inheriting that of original polar codes. In particular, we show a sequence of codes defined by GMs with column weights upper bounded by N^{λ}, for any $\lambda > \lambda^{*} \approx 0.585$ and the existence of a decoder with computation complexity $\mathcal{O}(N \log N)$ under which the block-error probability is bounded by $2^{-N^{\beta}}$ for any $\beta < 0.5$.

Next, for general BMS channels, we propose an enhancement of the DRS-based encoding scheme, referred to as *augmented-DRS (ADRS)* scheme, which requires additional channel uses and decoding complexity. In spite of these limitations, we show that there exists a sequence of capacity-achieving codes, referred to as the *polar-ADRS* codes. The sequence of codes is defined by GMs with column weights upper bounded by N^{λ}, for any $\lambda > \lambda^{\dagger} \approx 0.631$, and can be decoded with complexity $\mathcal{O}(N \log N)$.

The rest of this chapter is organized as follows. In Section 3.2, we introduce basic notations and definitions for channel polarization and polar codes. Section 3.3 provides a sparsity result for polar codes with general kernels. In Section 3.4, we introduce the notion of splitting algorithms and define the DRS algorithm. Section 3.5 provides the corresponding code construction over the BEC. The successive cancellation (SC) decoder is also described and shown to be of low computation complexity. Another code construction scheme aimed for transmission over general BMS channels, denoted as the ADRS scheme, is introduced in Section 3.6, where a corresponding SC decoder with low computation complexity is also discussed. Finally, Section 3.7 concludes this chapter. The proofs for the results in Sections 3.3, 3.5, and 3.6 are included in Appendix A.

3.2 Preliminaries

Let $h_b(\cdot)$ denote the binary entropy function, $\exp_2(x)$ denotes the function 2^x, $\ln(\cdot)$ be the logarithmic function with base e, and $\log(\cdot)$ be the logarithmic function with base 2. $Z(W)$ denote the Bhattacharyya parameter of a channel W. We recall definitions for the

BMS channel and capacity-achieving codes.

Definition 2. A binary memoryless symmetric channel (BMS) $W : \mathcal{X} \to \mathcal{Y}$ is a noisy memoryless channel with binary input alphabet $\mathcal{X}$, and channel output alphabet $\mathcal{Y}$, (we use $\mathcal{X} = \{0, 1\}$, and assume $\mathcal{Y}$ is finite, in this chapter.) such that $\Pr[Y = y|X = 0] = \Pr[Y = \phi(y)|X = 1]$ for all $y \in \mathcal{Y}$ for some involution ϕ on $\mathcal{Y}$.

Definition 3. A type of code is said to be capacity achieving over a BMS channel W with capacity $C = I(W) > 0$ if, for any given constant $R < C$, there exists a sequence of codes with rate R and the block-error probability vanishes as the block length N grows large. The block-error probability is evaluated under the maximum likelihood (ML) decoder, unless a different decoding scheme is specified.

3.2.1 Channel Polarization and Polar Codes

The *channel polarization* phenomenon was discovered by Arıkan [13] and is based on a 2×2 polarization transform as the building block. Let $A \otimes B$ denote the Kronecker product of matrices A and B, and $A^{\otimes n}$ denote the n-fold Kronecker product of A, i.e., $A^{\otimes n} = A \otimes A^{\otimes (n-1)}$ for $n \geqslant 2$ and $A^{\otimes 1} = A$. Let $\mathcal{W}$ denote the class of all BMS channels. Consider a channel transform $W \mapsto (W^-, W^+)$ that maps $\mathcal{W}$ to $\mathcal{W}^2$. Suppose the transform operates on an input channel $W : \mathcal{X} \to \mathcal{Y}$ to generate the channels $W^- : \mathcal{X} \to \mathcal{Y}^2$ and $W^+ : \mathcal{X} \to \mathcal{Y}^2 \times \mathcal{X}$ with transition probabilities

$$\begin{aligned} W^-(y_1, y_2|x_1) &= \sum_{x_2 \in \mathcal{X}} \frac{1}{2} W(y_1|x_1 \oplus x_2) W(y_2|x_2), \\ W^+(y_1, y_2, x_1|x_2) &= \frac{1}{2} W(y_1|x_1 \oplus x_2) W(y_2|x_2), \end{aligned}$$

where $\oplus$ denotes mod-2 addition. This transformation $W \mapsto (W^-, W^+)$ is referred to as a polarization recursion. Then a channel $W^{s_1, s_2, \dots, s_n}$ is defined for $s_i \in \{-, +\}, i = 1, 2, \dots, n$, resulting from applying the channel transform n times recursively as

$$W^{s_1, s_2, \dots, s_n} = \begin{cases} (W^{s_1, s_2, \dots, s_{n-1}})^- & \text{if } s_n = -, \\ (W^{s_1, s_2, \dots, s_{n-1}})^+ & \text{if } s_n = +. \end{cases}$$

For $N = 2^n$, the polarization transform is obtained from the $N \times N$ matrix $G_2^{\otimes n}$, where $G_2 = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$ [13], and $A^{\otimes n}$ denote the n-fold Kronecker product of A. A polar code of length N is constructed by selecting certain rows of $G_2^{\otimes n}$ as its generator matrix. More specifically, let K denote the code dimension. Then all the N bit-channels in the set $\{W^{s_1, s_2, \dots, s_n} : s_i \in$

$\{-,+\}$ for $i = 1, 2, \ldots, n\}$, resulting from the polarization transform, are sorted with respect to an associated parameter, e.g., their probability of error (or Bhattacharyya parameter), the best K of them with the lowest probability of error are selected, and then the corresponding rows from $G_2^{\otimes n}$ are selected to form the GM. Hence, the GM of an (N, K) polar code is a $K \times N$ sub-matrix of $G_2^{\otimes n}$. Then the probability of error of this code, under successive cancellation decoding, is upper bounded by the sum of probabilities of error of the selected K best bit-channels [13].

3.2.2 General Kernels and Error Exponent

It is shown in [78] that if G_2 is replaced by an $l \times l$ matrix G_l, then polarization still occurs if and only if G_l is an invertible matrix in $\mathbb{F}_2$ and it is not upper triangular under any column permutation, in which case the matrix G_l is called a *polarization kernel*. Furthermore, the authors of [78] provided a general formula for the speed of the error rate decay of polar codes constructed based on an arbitrary $l \times l$ polarization kernel G_l. More specifically, let $N = l^n$ denote the block length and C denote the capacity of the channel. For any fixed $\beta < E(G_l)$ and fixed code rate $R < C$, where $E(G_l)$ denotes the rate of polarization (see [78, Definition 7]), there is a sequence of polar codes based on G_l with probability of error P_e under SC decoding bounded by $P_e(n) \leqslant 2^{-N^\beta}$, for all sufficiently large n. The rate of polarization $E(G_l)$ is given by $E(G_l) = \frac{1}{l}\sum_{i=1}^{l} \log_l D_i$, where $\{D_i\}_{i=1}^{l}$ are the *partial distances* of G_l. More specifically, for $G_l = [g_1^T, g_2^T, \ldots, g_l^T]^T$, the partial distances D_i are given by $D_i \triangleq d_H(g_i, \text{span}(g_{i+1}, \ldots, g_l))$ for $i = 1, 2, \ldots, l$, where $d_H(a, b)$ is the Hamming distance between two vectors a and b, and $d_H(a, U)$ is the minimum distance between a vector a and a subspace U, i.e., $d_H(a, U) = \min_{u \in U} d_H(a, u)$.

3.3 Sparse Polar Code Constructions based on Large Kernels

In this section we first show the existence of capacity-achieving polar codes with generator matrices for which all column weights scale at most polynomially with arbitrarily small degree in the block length N, hence validating the conjecture in [73]. Second, we show that, for any polar code of rate 1, *almost* all of the column weights of the GM are polynomial in N.

Theorem 10. *For any fixed $s \in (0, 1)$ and any BMS channel, there are capacity-achieving polar codes under SC decoding, with generator matrices having column weights bounded by N^s, where N denotes the block length of the code.*

Proof: Consider an $l \times l$ polarizing matrix $G_l = \begin{bmatrix} I_{\frac{l}{2}} & 0_{\frac{l}{2}} \\ I_{\frac{l}{2}} & I_{\frac{l}{2}} \end{bmatrix}$, where l is an even integer such that $l \geqslant 2^{\frac{1}{s}}$. The partial distances are $D_i = 1$ for $1 \leqslant i \leqslant \frac{l}{2}$ and $D_i = 2$ for $\frac{l}{2}+1 \leqslant i \leqslant l$. Hence, the rate of polarization $E(G_l) = \frac{1}{2}\log_l 2 > 0$, and there is a sequence of capacity-achieving polar codes constructed using G_l as the polarizing kernel. Note that in G_l, each column has weight at most 2 and, hence, the column weights of $G_l^{\otimes n}$ are upper bounded by 2^n. By the specific choice of l, we have $2^n \leqslant (l^s)^n = (l^n)^s = N^s$, where $N = l^s$ is the block length of the code. This completes the proof. ■

Remark 2. While Theorem 10 provides a theoretical guarantee on the existence of capacity-achieving polar codes with sparse generator matrices, the sparsity comes at a cost. Specifically, the rate of polarization $E(G_l) = \frac{1}{2}\log_l 2 \leqslant \frac{s}{2}$ is smaller than that associated with the kernel G_2, given by $E(G_2) = 0.5$. On the other hand, while the SC decoding complexity for polar codes defined by general $l \times l$ kernels behaves as $\mathcal{O}(\frac{2^l}{l} N \log_l N)$[78], in this case, the complexity scales as $\mathcal{O}(N \log_l N)$ by considering the following viewpoint. Interleave $(l/2)^n$ copies of the polar code with block length 2^n based the standard G_2 kernel, to form a code with block length $N = l^n$ with an n-stage recursive encoder structure. By decoding each copy with complexity $\mathcal{O}(2^n \log 2^n) = \mathcal{O}(n2^n)$ under the SC decoder, the entire code can be decoded with complexity $\mathcal{O}((l/2)^n \cdot n2^n) = \mathcal{O}(N \log_l N)$.

Since we can construct capacity-achieving codes with column weights upper bounded by N^s with any fixed $s > 0$, by using polar codes, the question now is whether it is possible to further improve the sparsity of polar code GMs. For instance, we know it is possible to have an upper bound of $\mathcal{O}(\log N)$ on all the GM column weights of capacity-achieving codes, over the BEC, by utilizing random linear ensembles [73]. For rate-1 polar codes, the proposition below answers the inquiry in the negative, by showing that almost all the GM columns have weights lower bounded by a polynomial in N.

Proposition 11. *Given any $l \geqslant 2$, $l \times l$ polarizing kernel G_l, and $\frac{1}{l} > r > 0$, the fraction of columns in $G_l^{\otimes n}$ with $\mathcal{O}(N^{r \log_l 2})$ Hamming weight vanishes as n grows large, where $N = l^n$.*

Proof: The proof is given in Appendix Section A.1. ■

The trade-off highlighted in Remark 2 suggests that off-the-shelf polar code constructions with large kernels may not be the ideal option when the speed of decay of the error probability is a concern. However, the heaviest column in the polar code with kernel G_2 scales as $\Theta(N)$ for any code rate. To construct codes with sparse GM and suitable decay of the error probability, in the next section, we propose a *splitting* algorithm for the generator matrix and investigate the resulting codes in terms of the error probability, GM column sparsity, and the decoding complexity.

3.4 Splitting Algorithm

When all columns of a matrix G are required to be sparse, that is, have low Hamming weights, a *splitting algorithm* is applied. Given a column weight threshold $w_{u.b.}$, a splitting algorithm splits any column in G with weight exceeding $w_{u.b}$ into columns that sum to the original column both in $\mathbb{F}_2$ and in $\mathbb{R}$, and that have weights no larger than $w_{u.b.}$.

Note that a column of G is left intact by the splitting algorithm as long as its Hamming weight does not only exceed $w_{u.b.}$. Thus a splitting algorithm would be described as an algorithm which takes as input a column vector $\mathbf{v}$ and a weight threshold $w_{u.b.}$, and returns a set of column vectors whose lengths are equal to the length of $\mathbf{v}$. Given a matrix A with m columns and a threshold $w_{u.b.}$, with a slight abuse of notation, the matrix *generated* by a splitting algorithm is defined as the matrix whose column vectors are those from the m sets, which are respectively the outputs of the algorithm for each column of A. For example, consider a 4×2 matrix $A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}^T = [\mathbf{a_1}, \mathbf{a_2}]$, a threshold $w_{u.b.} = 2$, and a splitting algorithm $\mathcal{S}$. Let $\mathcal{S}(\mathbf{a_i}, w_{u.b.})$, $i = 1, 2$, be the sets of vectors returned by $\mathcal{S}$, given by $\mathcal{S}(\mathbf{a_1}, w_{u.b.}) = \left\{[1,0,1,0]^T, [0,0,0,1]^T\right\}, \mathcal{S}(\mathbf{a_2}, w_{u.b.}) = \left\{[1,0,1,0]^T, [0,1,0,0]^T\right\}$. The matrix generated by $\mathcal{S}$ for A is then a 4×4 matrix of the form $[[1,0,1,0]^T, [0,0,0,1]^T, [1,0,1,0]^T, [0,1,0,0]^T]$, or a column permutation of it.

Let an (N, K) polar code $\mathcal{C}$ have a $K \times N$ submatrix of $G = G_2^{\otimes n}$ as the generator matrix, and G' denote the $N \times N(1+\gamma)$ matrix generated by the splitting algorithm, where $N = 2^n$. A new code *based on* G' selects the same K rows of G' as the polar code $\mathcal{C}$ to form the generator matrix, where all the column weights are bounded by $w_{u.b.}$. Such a code is referred to as a *polar-based code corresponding to* G', or a PB(G') code, in this dissertation.

Note that more detailed description is needed to uniquely specify a splitting algorithm, which then determines the term γ and the performance of the PB(G') code. Specifically, the channel polarization phenomenon and the recursive encoding and decoding structure may be invalid when the GM is modified by the splitting algorithm. These changes also imply that the codes with the split GM may suffer from drawbacks such as weaker bounds on error probability and larger decoding complexity, as well as the rate loss with a multiplicative factor of $1 + \gamma$, when compared to the polar codes. In this section, we introduce a *splitting* algorithms, referred to as the decoder-respecting splitting (DRS) algorithm. When the threshold $w_{u.b.}$ is chosen appropriately, we show in Section 3.4.1 that the term γ goes to 0 exponentially fast in n, when the DRS algorithm is applied to columns of the matrix $G = G_2^{\otimes n}$.

The DRS algorithm is crucial to two encoding schemes, described in subsequent sections, that are effective in avoiding the drawbacks that may arise with an arbitrary splitting algorithm. These schemes enable low-complexity SC decoders based on likelihood ratios that can be calculated with a recursive algorithm. The encoding of the resulting PB(G') codes can be realized by a encoding scheme which inherits the recursive structure of the original polar codes, except only at locations that corresponds to a split of a column of G, as dictated by the DRS algorithm. At these locations, the exclusive-OR operations are removed and additional copies of the underlying channel are used. The PB(G') codes suffer only a negligible $1+\gamma$ multiplicative factor of rate loss compared to the original polar codes for large n. For BEC, this sequence of codes is capacity-achieving with an error exponent of $\frac{1}{2}$, under a new SC decoding scheme (see Theorem 15 in Section 3.5). For general BMS channels, another encoding scheme, referred to as the *ADRS* scheme, is proposed in Section 3.6. This scheme introduces additional 'noise' nodes and requires even more copies of the underlying channel when the DRS algorithm requires a split. For codes generated by this scheme, results similar to that in Theorem 15 are available with a slightly stricter condition on the choice of $w_{u.b.}$.

3.4.1 Decoder-Respecting Splitting Algorithm

The main idea of the DRS algorithm is to construct a generator matrix that can be realized with an encoding pattern similar to conventional polar codes such that the column weights of the matrix associated with the diagram are at most $w_{u.b.}$. The pseudo code for the algorithm is provided in Algorithm 1.

The core of the algorithm is the DRS-Split function. When the weight of the input the vector $\mathbf{x}$ is larger than the threshold, it splits the vector in half into vectors $\mathbf{x}_h$ and $\mathbf{x}_t$, and recursively finds two sets, Y_h and Y_t, composed of vectors with the length halved compared to the length of $\mathbf{x}$. The vectors are then appended to the length of $\mathbf{x}$, which collectively form the output of the function. For a vector $\mathbf{u} \in \{0,1\}^{m\times 1}$, let $|\mathbf{u}| = m$ denote its length, and $w_H(\mathbf{u})$ its Hamming weight. We note that the weights of vectors in Y_h and Y_t are respectively upper bounded by the weights of $\mathbf{x}_h$ and $\mathbf{x}_t$, both of which are bounded by $k = |\mathbf{x}_h| = |\mathbf{x}_t|$, and that the value of k is halved each iteration. Hence, the function is guaranteed to terminate as long as the threshold is a positive integer.

We use a simple example to illustrate the algorithm. Let $n=3$, $v = [0,0,0,0,1,1,1,1]^T$ and $w_{u.b.} = 2$. Since the weight of v exceeds the threshold, it is first split into $\mathbf{x}_h = [0,0,0,0]^T$ and $\mathbf{x}_t = [1,1,1,1]^T$. Since $\mathbf{x}_h$ is an all-zero vector, Y_h is an empty set according to line 14 to 15. To compute $Y_t =$DRS-Split$(2,[1,1,1,1]^T)$, the function splits the input into half again, thereby obtaining $\mathbf{x}'_h = [1,1]^T$ and $\mathbf{x}'_t = [1,1]^T$. The corresponding Y'_h and Y'_t are then both $\left\{[1,1]^T\right\}$ and, hence, we have $Y_t = \left\{[0,0,1,1]^T\right\} \cup \left\{[1,1,0,0]^T\right\} =$

Algorithm 1 DRS algorithm

```
Input: weight threshold $w_{u.b.} \in \mathbb{N}$, a column vector $\mathbf{v} \in \{0,1\}^{2^n \times 1}$
Output: the set of vectors with length $2^n$ returned by DRS-Split($w_{u.b.}, \mathbf{v}$)
 1: function DRS-Split($w_{u.b.}, \mathbf{x}$)
 2:     if $w_H(\mathbf{x}) > w_{u.b.}$ then
 3:         $k \leftarrow \text{length}(\mathbf{x})/2$
 4:         $\mathbf{x}_h \leftarrow (x_1, \dots, x_k)^T$, $\mathbf{x}_t \leftarrow (x_{k+1}, \dots, x_{2k})^T$
 5:         $Y_h \leftarrow$ DRS-Split($w_{u.b.}, \mathbf{x}_h$)
 6:         $Y_t \leftarrow$ DRS-Split($w_{u.b.}, \mathbf{x}_t$)
 7:         if $\mathbf{x}_h = \mathbf{0}_{k\times 1}$ then
 8:             return $\bigcup_{y \in Y_t} \{(\mathbf{0}_{1\times k}, y^T)^T\}$
 9:         else if $\mathbf{x}_t = \mathbf{0}_{k\times 1}$ then
10:             return $\bigcup_{y \in Y_h} \{(y^T, \mathbf{0}_{1\times k})^T\}$
11:         else
12:             return $\bigcup_{y \in Y_t} \{(\mathbf{0}_{1\times k}, y^T)^T\} \cup \bigcup_{y \in Y_h} \{(y^T, \mathbf{0}_{1\times k})^T\}$
13:         end if
14:     else if $w_H(\mathbf{x}) = 0$ then
15:         return {}
16:     else
17:         return {$\mathbf{x}$}
18:     end if
19: end function
```

$\{[0,0,1,1]^T, [1,1,0,0]^T\}$. Since $\mathbf{x}_h = \mathbf{0}_{4\times 1}$, the function proceeds to lines 7 and 8, and returns $\{[0,0,0,0,0,0,1,1]^T, [0,0,0,0,1,1,0,0]^T\}$.

In order to analyze the effect of the DRS algorithm on the matrix $G_2^{\otimes n}$, we show that the size of the algorithm output does not depend on the order of a sequence of Kronecker product operations, where the *size* of a set of vectors stands for the number of vectors in the set. Suppose that the Kronecker product operations with the vector $[1,\ 1]^T$ for n_1 times and with the vector $[0,\ 1]^T$ for n_2 times are applied on a vector v, where $n = n_1 + n_2$ and the order of the operations is specified by a sequence $(s_1, s_2, \dots, s_n) \in \{-,+\}^n$ with $|\{i : s_i = -\}| = n_1$ and $|\{i : s_i = +\}| = n_2$. Also, let $v^{(i)}$ denote the output of applying the first i Kronecker product operations on v. It is defined by the following recursive relation:

$$v^{(i)} = \begin{cases} v^{(i-1)} \otimes [1,1]^T, \text{ if } s_i = -, \\ v^{(i-1)} \otimes [0,1]^T, \text{ if } s_i = +, \end{cases} \tag{3.1}$$

for $i \geqslant 1$ and the initial condition $v^{(0)} = v$. We use $v^{(s_1, s_2, \dots, s_i)}$ instead of $v^{(i)}$ when the sequence is needed for clarity. The following lemma shows that any two vectors of the form $v^{(s_1, s_2, \dots s_n)}$ will be split into the same number of columns under the DRS algorithm as long as the sequences associated with them contain the same number of $-$ and $+$ signs.

Lemma 12. *Let $n = n_1 + n_2$ and $(s_1, \ldots, s_n) \in \{-, +\}^n$ be a sequence with n_1 minus signs and n_2 plus signs. Let $v^{(n)}$ be the vector defined by a vector v and the sequence $(s_1, \ldots, s_n)$ through equation* (3.1)*. Then the size of the DRS algorithm output for $v^{(n)}$ depends only on the values n_1 and n_2.*

Proof: The proof is given in Appendix Section A.2.1. ■

Let a $K \times N$ matrix $M = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N]$ and a threshold $w_{u.b.}$ be given. Suppose that the DRS algorithm is applied to each column in M and the sum of the sizes of the output sets is $N(1+\gamma)$. Then $\mathrm{DRS}(M)$ is defined as the $K \times N(1+\gamma)$ matrix consisting of all the vectors in the output sets (with repetition).

We study the effect of the DRS algorithm in terms of the multiplicative rate loss, i.e., $1+\gamma$. Since all the columns of $G_2^{\otimes n}$ are in the form of $v^{(s_1, s_2, \ldots, s_n)}$ with $v = [0, 1]^T$ or $v = [1, 1]^T$, Lemma 12 substantially simplifies the analysis for γ. In particular, the following proposition shows an appropriate choice of $w_{u.b.}$ guarantees the existence of a sparse polar-based GM with vanishing γ.

Proposition 13. *Let the columns of $G_2^{\otimes n}$ be the inputs for the DRS algorithm and $\mathrm{DRS}(G_2^{\otimes n})$ be the $N \times N(1+\gamma)$ matrix generated by the DRS algorithm for $G_2^{\otimes n}$. The term γ vanishes exponentially fast as n goes to infinity for any $w_{u.b.} = 2^{n\lambda}$ with $\lambda > \lambda^* \triangleq h_b(\frac{2}{3}) - \frac{1}{3} \approx 0.585$.*

Proof: The proof is given in Appendix Section A.2.2. ■

For the effect of the DRS algorithm on $G_2^{\otimes n}$ with finite n, we compute values of γ for various combinations of n and λ, as shown in Figure 3.1. The numerical results with $6 \leqslant n \leqslant 26$ indicate that, for $0.5 \leqslant \lambda < 0.6$, the multiplicative rate loss γ is larger with larger n, and for $\lambda \geqslant 0.65$, γ is smaller with larger n. The fact that the $n = 26$ does not provide the smallest γ for λ close to λ^* should not be considered a contradiction to Proposition 13. Instead, the closer $\lambda > \lambda^*$ is, the larger n it takes for the exponential decay of γ to dominate.

3.5 Low-complexity Decoder for Polar-based Codes: BEC

In this section, we show two results for the polar-based code corresponding to $\mathrm{DRS}(G_2^{\otimes n})$ over the BEC. Such codes are referred to as the *polar-DRS* codes in this dissertation. First, we propose a low-complexity suboptimal decoder for the polar-DRS codes. Second, with the

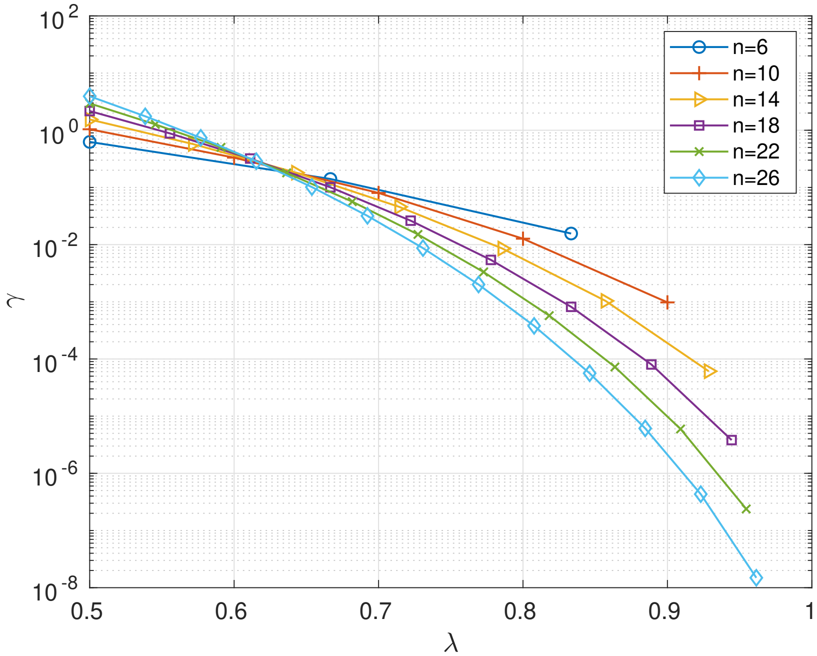


Figure 3.1: Multiplicative rate loss factor γ versus λ

low-complexity suboptimal decoder, the polar-DRS codes are capacity-achieving for suitable column weight threshold.

It is known that when the channel transformation with kernel G_2 is applied to two BECs, the two new bit-channels are also BECs. Specifically, for two binary erasure channels W_1 and W_2 with erasure probabilities ϵ_1 and ϵ_2, respectively, the polarized bit-channels $W^-(W_1, W_2)$ and $W^+(W_1, W_2)$ are binary erasure channels with erasure probabilities $\epsilon_1 + \epsilon_2 - \epsilon_1\epsilon_2$ and $\epsilon_1\epsilon_2$, respectively.

The mutual information $I(\cdot)$ and Bhattacharyya parameter $Z(\cdot)$ of a BEC W with erasure probability ϵ are given by: $I(W) = 1-\epsilon, Z(W) = \epsilon$. For a sequence $(s_1, s_2, \ldots, s_n) \in \{-,+\}^n$, the function $\mathsf{Bi2De}(s_1, s_2, \ldots, s_n)$ returns the decimal value of the binary string in which a minus sign for s_i is regarded as a 0 and a plus sign as a 1, e.g., $\mathsf{Bi2De}(-,+,+) = (011)_2 = 3$. Let G denote $G_2^{\otimes n}$ and G' denote $\mathrm{DRS}(G_2^{\otimes n})$, and let $Z_G^{(s_1 s_2 \ldots s_n)}$ denote the Bhattacharyya parameter of the bit-channel $W^{s_1 s_2 \ldots s_n}$, which is equal to $W_N^{(\mathsf{Bi2De}(s_1,s_2,\ldots,s_n)+1)}$ in [13, page 3]. The term $Z_{G'}^{(s_1 s_2 \ldots s_n)}$ denotes the Bhattacharyya parameter of the bit-channel observed by the source bit of the same index corresponding to G'.

The following lemma shows that the bit-channel observed by each source bit is better in terms of the Bhattacharyya parameter when G' is the generator matrix instead of G.

Lemma 14. *Let $w_{u.b.}$ and n be given, and let G denote $G_2^{\otimes n}$ and G' denote $\mathrm{DRS}(G_2^{\otimes n})$. The* *following is true for any* $(s_1, s_2, \ldots, s_n) \in \{-,+\}^n$*:*

$$Z_{G'}^{(s_1 s_2 \ldots s_n)} \leqslant Z_{G}^{(s_1 s_2 \ldots s_n)}.$$

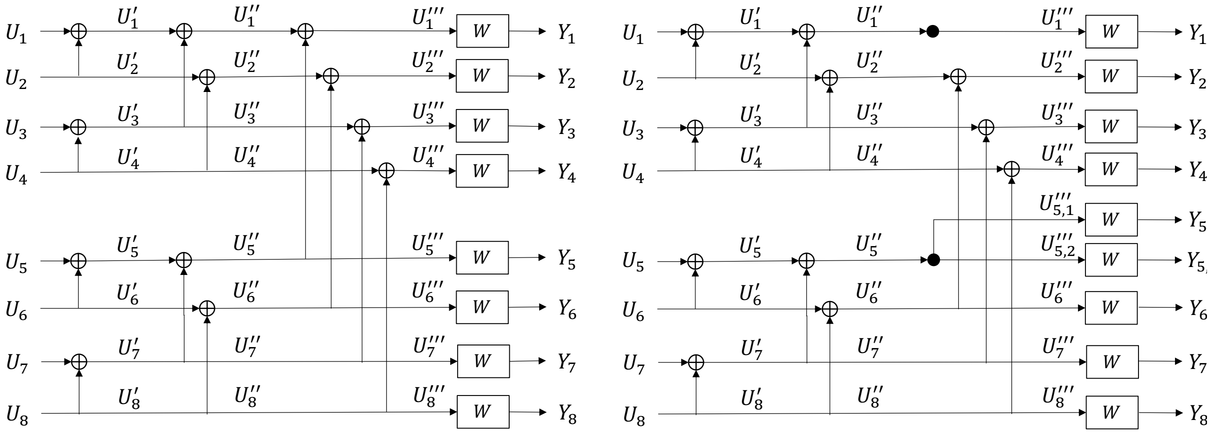
Proof: The proof is given in Appendix A.3. ■

Remark 3. A key to the proof of Lemma 14 is a recursive encoding scheme for the relationship $\mathbf{x} = \mathbf{u}G'$, where $\mathbf{u}$ and $\mathbf{x}$ are row vectors of lengths $N = 2^n$ and $N(1+\gamma)$, respectively. The encoding scheme is most easily understood by considering the low-complexity encoding structure for the standard polar code, as seen in [15], and replacing the exclusive-OR (XOR) operations at locations that correspond to the splitting operations dictated by the DRS algorithm. Specifically, when a split is required on a column of G, the corresponding XOR node is removed, the input bit for the 'worse' channel remains untouched, and two copies of the input bit for the 'good' channel are transmited through the underlying channel. For example, consider $G = G_2^{\otimes 3}$. The encoding diagram for codes defined by G is shown in Figure 3.2a. We have

$$G = \begin{bmatrix} 1&0&0&0&0&0&0&0 \\ 1&1&0&0&0&0&0&0 \\ 1&0&1&0&0&0&0&0 \\ 1&1&1&1&0&0&0&0 \\ 1&0&0&0&1&0&0&0 \\ 1&1&0&0&1&1&0&0 \\ 1&0&1&0&1&0&1&0 \\ 1&1&1&1&1&1&1&1 \end{bmatrix}, G' = \begin{bmatrix} 1&0&0&0&0&0&0&0&0 \\ 1&0&1&0&0&0&0&0&0 \\ 1&0&0&1&0&0&0&0&0 \\ 1&0&1&1&1&0&0&0&0 \\ 0&1&0&0&0&1&0&0&0 \\ 0&1&1&0&0&1&1&0&0 \\ 0&1&0&1&0&1&0&1&0 \\ 0&1&1&1&1&1&1&1&1 \end{bmatrix},$$

where G' is the matrix $\mathrm{DRS}(G_2^{\otimes 3})$ when $w_{u.b.} = 4$. The encoding structure for G' is shown in Figure 3.2b. Since the first column is the only column of G split by the DRS algorithm when $w_{u.b.} = 4$, we remove the XOR node that performs $U_1''' = U_1'' + U_5''$, and assigns $U_1''' = U_1''$, $U_{5,1}''' = U_5''$ and $U_{5,2}''' = U_5''$. Two solid circles, representing transparent nodes where the output variable(s) are identical to the input variable, are used to indicate the location of the removed XOR node. For the case when $w_{u.b.} = 2$, the DRS algorithm would split the first column of G into three vectors, and the second, third, and fifth column once each. The corresponding encoding diagram is shown in Figure 3.3, where we color the solid circles associated with splits on the first, second, third, and fifth columns by black, green, orange, and blue, respectively.

We are ready to show the existence of a sequence of capacity-achieving codes over the BEC with GMs where the column weights are bounded by a polynomial in the blocklength, and that the block error probability under a low complexity decoder vanishes as n grows large.



(a) Encoding block for generator matrix $G_2^{\otimes 3}$ (b) Encoding block for generator matrix G'

Figure 3.2: Encoding structure change due to the DRS algorithm with $N = 8, w_{u.b.} = 4$

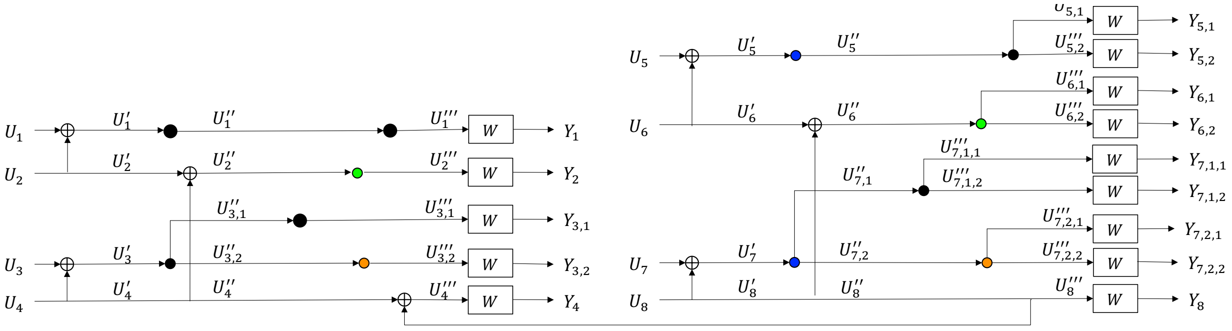


Figure 3.3: Encoding block for generator matrix $\mathrm{DRS}(G_2^{\otimes 3})$ when $w_{u.b.} = 2$

Theorem 15. *Let $\beta < E(G_2) = 0.5$, $\lambda > \lambda^* = h_b(\frac{2}{3}) - \frac{1}{3} \approx 0.585$, and a BEC W with capacity C be given. There exists a sequence of polar-based codes corresponding to* DRS*$(G_2^{\otimes n})$ with the following properties for all sufficiently large n: (1) The error probability under a SC decoder is upper bounded by 2^{-N^β}, where $N = 2^n$, (2) The Hamming weight of each column of the GM is upper bounded by N^λ, (3) The rate approaches C as n grows large, and (4) The codes can be decoded by a SC decoding scheme with complexity $\mathcal{O}(N \log N)$.*

Proof: Let the threshold for DRS algorithm be $w_{u.b.} = 2^{n\lambda}$, G denote $G_2^{\otimes n}$, and G' denote $\mathrm{DRS}(G_2^{\otimes n})$. We prove the four claims in order. First, Lemma 14 shows that for a given n and any $t > 0$, the following is true:

$$\{\mathbf{s} \in \{-,+\}^n : Z_G^{\mathbf{s}} \leqslant t\} \subseteq \{\mathbf{s} \in \{-,+\}^n : Z_{G'}^{\mathbf{s}} \leqslant t\}. \tag{3.2}$$

Using [13, Theorem 2], for any $\beta < \frac{1}{2}$, we have

$$\liminf_{n\to\infty} \frac{1}{N} \left| \left\{ \mathbf{s} \in \{-,+\}^n : Z_G^{\mathbf{s}} \leqslant 2^{-N^\beta} \right\} \right| = I(W) = C \tag{3.3}$$

Let $\boldsymbol{S_G}$ and $\boldsymbol{S_{G'}}$ denote the sets of the sequences $\mathbf{s} \in \{-,+\}^n$ that satisfy $Z_G^{\mathbf{s}} \leqslant 2^{-N^\beta}$ and $Z_{G'}^{\mathbf{s}} \leqslant 2^{-N^\beta}$, respectively. Equation (3.2) guarantees that $\boldsymbol{S_G}$ is a subset of $\boldsymbol{S_{G'}}$. Assume the code corresponding to G freezes the input bits observing bit-channels $W^{s_1 s_2 \dots s_n}$ for all $(s_1, s_2, \dots, s_n) \notin \boldsymbol{S_G}$. For the code corresponding to G', we use the bit-channels with the same index as the code corresponding to G, for transmission of information bits, and leave the rest as frozen. The probability of block error under SC decoding, which is described in the last part of this proof, for the code corresponding to G', $P_{e,G'}$, can be bounded above, as in [13], by the sum of the Bhattacharyya parameters of the bit-channels for the source bits (that are not frozen), that is,

$$P_{e,G'} \leqslant \sum_{\mathbf{s}\in\boldsymbol{S_G}} Z_{G'}^{\mathbf{s}} \leqslant \sum_{\mathbf{s}\in\boldsymbol{S_G}} 2^{-N^\beta} = |\boldsymbol{S_G}|\, 2^{-N^\beta},$$

where the second inequality follows because, for $\mathbf{s} \in \boldsymbol{S_G}$, we must have $\mathbf{s} \in \boldsymbol{S_{G'}}$ and thus $Z_{G'}^{\mathbf{s}} \leqslant 2^{-N^\beta}$. From (3.3), for all sufficiently large n, we have $P_{e,G'} \leqslant NC2^{-N^\beta}$. With some calculus, one may show that, for any $\beta' < \frac{1}{2}$, $P_{e,G'} \leqslant 2^{-N^{\beta'}}$ for all sufficiently large n.

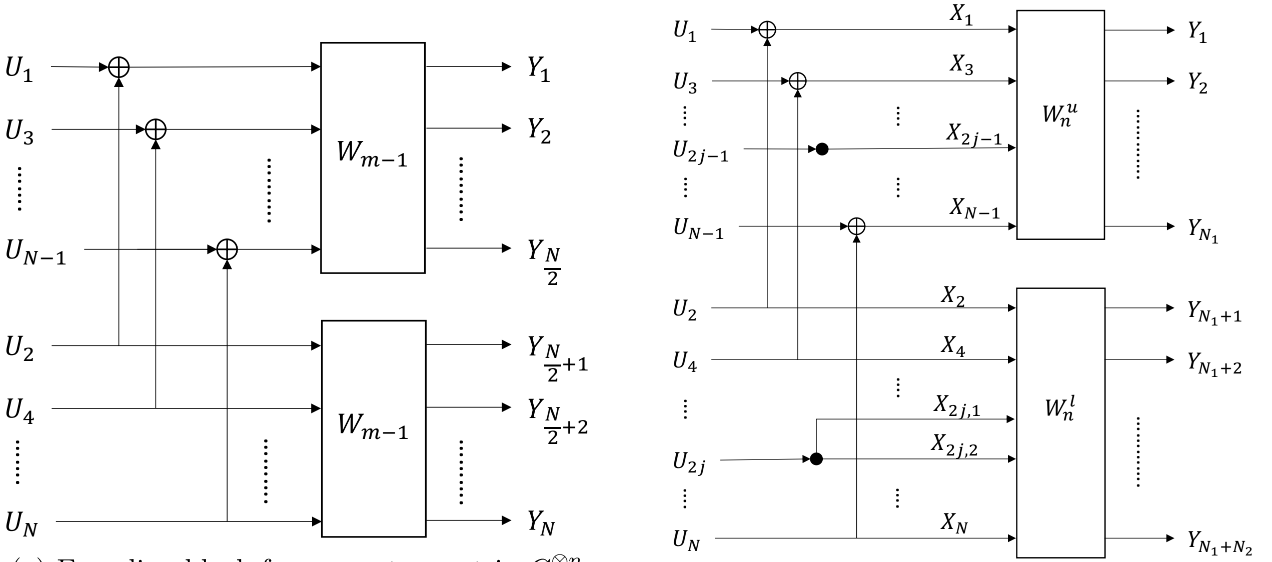
The second claim follows from the fact that the GM for the code corresponding to G' is a submatrix of G', and the Hamming weight of each column of G' is upper bounded by $w_{u.b.} = 2^{n\lambda} = N^\lambda$.

The third claim is a consequence of Proposition 13 and Lemma 14. The number of information bits of the code corresponding to G' is given by $|\boldsymbol{S_G}|$, and the length of the code is $N(1+\gamma)$. Hence the code rate is $\frac{|\boldsymbol{S_G}|}{N(1+\gamma)}$. Since the term γ vanishes as n grows large, we have

$$\liminf_{n\to\infty} \frac{|\boldsymbol{S_G}|}{N(1+\gamma)} = \liminf_{n\to\infty} \frac{|\boldsymbol{S_G}|}{N} = I(W) = C. \tag{3.4}$$

Finally, we prove the claim for the existence of a low-complexity decoder. Just like that of the SC decoder for conventional polar codes with kernel G_2, the decoding algorithm proceeds in a recursive manner. Let $U_1, \dots, U_N$ be the inputs, and $Y_1, \dots, Y_{N_1}, Y_{N_1+1}, \dots, Y_{N_1+N_2}$ the outputs, where $N_1 + N_2 = N(1+\gamma)$, as shown in Figure 3.4b. However, while the polar code based on $G_2^{\otimes n}$, as shown in Figure 3.4a, is recursive in the encoder structure, i.e., the two encoding sub-blocks corresponding to $G_2^{\otimes n-1}$ are identical, the code based on $(G_2^{\otimes n})'$ is not, as the blocks W_n^u and W_n^l are not necessarily equal. In fact, when there is a split in the GM due to the DRS algorithm, i.e., when one or more of the XOR operations shown in

Figure 3.4b is replaced by two solid black circle, the number of inputs of the block W_n^l will be larger than that of W_n^u.



(a) Encoding block for generator matrix $G_2^{\otimes n}$

(b) Encoding block for generator matrix $(G_2^{\otimes n})'$

Figure 3.4: Encoding/decoding diagrams for standard and DRS-modified polar codes

Let $\mathcal{F} \subseteq \{1, \dots, N\}$ be the set of the indices of the frozen bits. The decoder declares estimates $\hat{U}_i$ of the inputs, for $1 \leqslant i \leqslant N$, sequentially by:

$$\hat{U}_i = \begin{cases} u_i, & \text{if } i \in \mathcal{F}, \\ \psi_i(Y_1^{N_1+N_2}, \hat{U}_1^{i-1}, W_n) & \text{if } i \notin \mathcal{F}, \end{cases} \tag{3.5}$$

where $\psi_i(Y_1^{N_1+N_2}, \hat{U}_1^{i-1}, W_n)$ can be found in following four cases, and W_n denotes the encoding block shown in Figure 3.4b. Let the symbol e denotes an erasure, and assume $e \oplus b = e$ for $b \in \{0, 1, e\}$. We write ψ_i in place of $\psi_i(Y_1^{N_1+N_2}, \hat{U}_1^{i-1}, W_n)$ for the sake of space in the following.

- If i is odd and $X_i = U_i \oplus U_{i+1}$, which corresponds to an unsplit XOR operation observed by U_i,

$$\psi_i \triangleq \begin{cases} \hat{X}_i \oplus \hat{X}_{i+1} & \text{if } \hat{X}_i \neq e, \hat{X}_{i+1} \neq e, \\ e, & \text{otherwise.} \end{cases}$$

- If i is odd and $X_i = U_i$, which corresponds to a split XOR operation, $\psi_i \triangleq \hat{X}_i$.

- If i is even and $X_{i-1} = U_i \oplus U_{i-1}$, which corresponds to an unsplit XOR operation,

$$\psi_i \triangleq \begin{cases} \hat{X}_i, & \text{if } \hat{X}_i \neq e, \hat{X}_{i-1} \neq e, \hat{X}_i = \hat{X}_{i-1} \oplus \hat{U}_{i-1} \\ & \text{or } \hat{X}_i \neq e, \hat{X}_{i-1} = e, , \\ \hat{X}_{i-1} \oplus \hat{U}_{i-1}, & \text{if } \hat{X}_i = e, \hat{X}_{i-1} \neq e, \hat{U}_{i-1} \neq e \\ e, & \text{otherwise.} \end{cases}$$

- If i is even and $X_{i,1} = X_{i,2} = U_i$, which corresponds to a split XOR operation,

$$\psi_i \triangleq \begin{cases} \hat{X}_{i,1}, & \text{if } \hat{X}_{i,1} \neq e, \hat{X}_{i,2} = e, \\ & \text{or } \hat{X}_{i,1} = \hat{X}_{i,2} \neq e, \\ \hat{X}_{i,2}, & \text{if } \hat{X}_{i,1} = e, \hat{X}_{i,2} \neq e, \\ e, & \text{otherwise.} \end{cases}$$

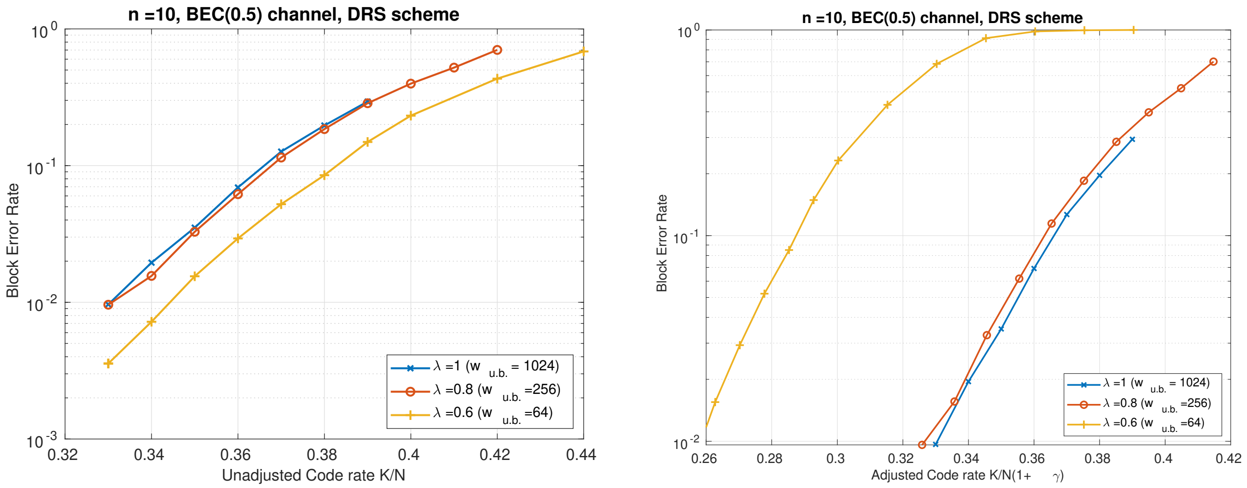
The estimates $\hat{X}_1, \hat{X}_3, \ldots, \hat{X}_{N-1}$ and $\hat{X}_2, \hat{X}_4, \ldots,$ $\hat{X}_{2j,1}$, $\hat{X}_{2j,2}, \ldots, \hat{X}_N$ are found in a similar approach using the blocks W_n^u and W_n^l along with the outputs $Y_1, \ldots, Y_{N_1}$ and $Y_{N_1+1}, \ldots, Y_{N_1+N_2}$, respectively.

For the right-most variables, the blocks they observe are identical copies of the BEC W. Hence the estimates of the variables, denoted as $\hat{X}_1^{(n)}, \hat{X}_2^{(n)}, \ldots, \hat{X}_{N_1+N_2}^{(n)}$ are naturally defined by the outputs of the channels, i.e., $\hat{X}_i^{(n)} = Y_i$ for $i = 1, 2, \ldots, N_1 + N_2$.

At each stage there are at most $N_1 + N_2 = N(1+\gamma) = \mathcal{O}(N)$ estimates to make, and the recursion ends in $\log(N)$ steps. Since each estimate is obtained with constant complexity, the total decoding complexity for the code based on $\text{DRS}(G_2^{\otimes n})$ is bounded by $\mathcal{O}(N \log N)$.

■

We evaluate the performance of the polar-DRS codes with $n = 10$ and $\lambda = 0.6, 0.8, 1.0$, under the SC decoding scheme described in the proof of Theorem 15, over the BEC with erasure probability $\epsilon = 0.5$. In Figure 3.5a, the block error probabilities for the curves with smaller λ are smaller, due to the improvement of some of the Bhattacharyya parameters observed by the information bits. That is, there are sequences $(s_1, s_2, \ldots, s_n) \in \{-,+\}^n$ for which the inequality in Lemma 14 is strict. However, after factoring in multiplicative rate loss γ, we may observe in Figure 3.5b that the performance of the codes with $\lambda = 0.6$ are substantially worse than the original polar code ($\lambda = 1$ curve), and those with $\lambda = 0.8$ deliver trade-off between code rate and error probability comparable to the original polar code, while guaranteeing the threshold $w_{u.b.}$ is one-fourth of the latter.



(a) Unadjusted code rate

(b) Adjusted code rate

Figure 3.5: Error probability for polar-DRS codes with $n = 10$

Remark 4. We note that for general BMS channels, Lemma 14 may fail. One key part in the proof (see Appendix A.3) is the fact that the Bhattacharyya parameter for the bit-channel observed by U_i is a non-decreasing function of those of $W(X_1), \ldots, W(X_{f(m)})$ for $i \leqslant 2^m$, and of $W(X_{f(m)+1}), \ldots, W(X_{2f(m)})$ for $i > 2^m$, when all the channels are BECs. We now provide an example where we see the argument for Lemma 14 fail for BMS channels. Let a, a', b, b' be four distinct elements and $\mathcal{Y} = \{a, a', b, b'\}$. Let two BMS channels $W_1, W_2 : \{0,1\} \to \mathcal{Y}$ be given, and that $W_1(y|0) = W_1(\phi(y)|1)$ and $W_2(y|0) = W_2(\phi(y)|1)$ for all $y \in \mathcal{Y}$ where the involution ϕ maps $a \mapsto a', b \mapsto b'$. Assume the channel transition probabilities are $W_1(a|0) = 6/9, W_1(b|0) = 1/9$, $W_1(b'|0) = 1/9, W_1(a'|0) = 1/9$ and $W_2(a|0) = 5/11$, $W_2(b|0) = 4/11$, $W_2(b'|0) = 1/11$, $W_2(a'|0) = 1/11$. The Bhattacharyya parameters for W_1, W_2 are respectively 0.7666 and 0.7702. If $m = 1$ and B_m is simply the kernel G_2, the symbols X_1, X_2 are functions of U_1, U_2 given by $X_1 = U_1 + U_2$ and $X_2 = U_2$.

We now consider two possible cases for the pair $(W(X_1), W(X_2))$. If $(W(X_1), W(X_2)) = (W_1, W_2)$, the Bhattacharyya parameters for the bit-channels observed by U_1, U_2 are respectively 0.9147 and 0.5904. If $(W(X_1), W(X_2)) = (W_2, W_2)$, the Bhattacharyya parameters for the bit-channels observed by U_1, U_2 are respectively 0.9137 and 0.5932. We note that while the Bhattacharyya parameters for $W(X_1)$, $W(X_2)$ in the second case are no less than in the first case, the Bhattacharyya parameter $Z(U_1)$ in the second case is smaller than in the first case. With the above observation, one can not claim the validity of Theorem 15 for general BMS channels. This motivates a new code construction for general BMS channels.

3.6 Low-complexity Decoder for Polar-based Codes: BMS

This section introduces a capacity-achieving polar-based coding scheme with low-complexity decoder for general BMS channels. For general BMS channels, the Bhattacharyya parameter of the bit-channel W^- cannot be expressed only in terms of parameters of the channel W. This implies that Lemma 14 and Theorem 15 are not applicable for channels other than BEC, as pointed out in Remark 4. A procedure that augments the generator matrix corresponding to G', the output of the DRS algorithm for the matrix $G_2^{\otimes n}$, may be used to construct a capacity-achieving linear code over any BMS channel W.

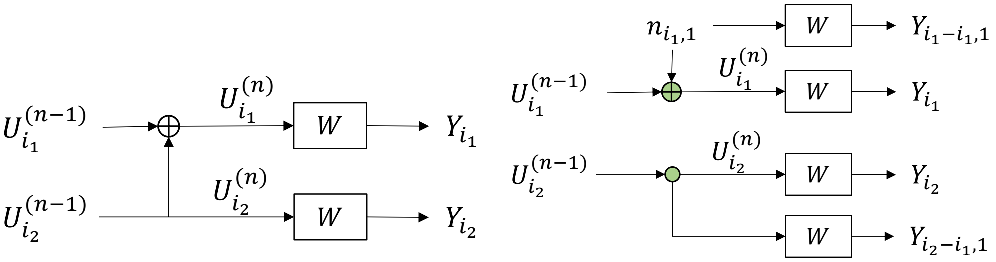
3.6.1 ADRS Scheme

The encoding scheme, termed augmented-DRS (ADRS) scheme, avoids heavy columns in the GM and, at the same time, guarantees that the bit-channels observed by the source bits U_i have the same statistical characteristics as when they are encoded with the generator matrix $G_2^{\otimes n}$. Specifically, the ADRS scheme modifies the encoder for $G_2^{\otimes n}$ starting from the split XOR operations associated with the first polarization recursion, then the second recursion, and proceed all the way to the n-th recursion, where a XOR operation is split if and only if it is split in an encoder with generator matrix $\mathrm{DRS}(G_2^{\otimes n})$.

Assume an XOR operation with operands $U_{i_1}^{(n-j)}$ and $U_{i_2}^{(n-j)}$ and the output $U_{i_1}^{(n-j+1)}$, where $i_1 = \mathsf{Bi2De}(s_1, \ldots, s_{j-1}, s_j = -, s_{j+1}, \ldots, s_n) + 1$ and $i_2 = \mathsf{Bi2De}(s_1, \ldots, s_{j-1}, s_j = +, s_{j+1}, \ldots, s_n) + 1 = i_1 + 2^{n-j}$, is to be split (see Section 3.5 for the function $\mathsf{Bi2De}(\cdot)$). If $j = 1$, before modification, the variables $U_{i_1}^{(n)}$ and $U_{i_2}^{(n)}$ are transmitted through two copies of W, and the bit-channels observed by $U_{i_1}^{(n-1)}$ and $U_{i_2}^{(n-1)}$ are W^- and W^+, respectively, as shown in Figure 3.6a. If the XOR operation is split according to $\mathrm{DRS}(G_2^{\otimes n})$, ADRS scheme replaces the structure by that given in Figure 3.6b, where $n_{i_1,1}$ is a Bernoulli(0.5) random variable independent of all the other variables.

If $j \geqslant 2$, assume that the ADRS modification for the split operations for the first $(j-1)$ recursions are completed. Let $n_{i_1,j}$ be a Bernoulli(0.5) random variable independent of all the other given variables. The part of encoding diagram to the right of $U_{i_1}^{(n-j+1)}$ is replicated, where $n_{i_1,j}$ takes the place of $U_{i_1}^{(n-j+1)}$ in the replica. And then we let $U_{i_1}^{(n-j+1)} = U_{i_1}^{(n-j)} \oplus n_{i_1,j}$. In addition, the part of encoding diagram to the right of $U_{i_2}^{(n-j+1)}$ is replicated, and a copy of $U_{i_2}^{(n-j)}$ is transmitted through the replica. The variable $U_{i_2}^{(n-j+1)}$ remains $U_{i_2}^{(n-j+1)} = U_{i_2}^{(n-j)}$.

We demonstrate the procedure described above through the following example. Assume $n = 3$, $N = 8$, and $w_{u.b.} = 2$. The encoding diagram for $G_2^{\otimes 3}$ is shown in Figure 3.7a, and the



(a) Before Modification

(b) ADRS modification

Figure 3.6: ADRS scheme for a split XOR of first iteration of polarization

XOR operations that are split in $\mathrm{DRS}(G_2^{\otimes 3})$ are marked in green and blue, which indicate the operations are due to the first and the second polarization recursions (i.e., s_1 and s_2), respectively. The notations U_i', U_i'', U_i''' are used to represent $U_i^{(1)}, U_i^{(2)}, U_i^{(3)}$. Replacing the XOR operations marked in green as described for the case of $j = 1$, the encoding diagram is now shown in Figure 3.7b. For the XOR operations marked in blue, we proceed by using the step for $j \geqslant 2$ and obtain the diagram shown in Figure 3.7c.

It can be noted that the bit-channels observed by each of $U_i^{(j)}$, for $i = 1, 2, \ldots, N$ and $j = 0, 1, 2, \ldots, n$, in the ADRS encoder are the same as those in the standard encoder for the generator matrix $G_2^{\otimes n}$ (The variable $U_i^{(0)}$ are given by U_i for $1 \leqslant i \leqslant N$). When an XOR operation associated with the j-th recursion, with operands $U_{i_1}^{(n-j)}$ and $U_{i_2}^{(n-j)}$ and the output $U_{i_1}^{(n-j+1)}$, is split and modified under the ADRS scheme, the complexity of computing the likelihood or log-likelihood for $U_{i_1}^{(n-j)}$ and $U_{i_2}^{(n-j)}$ can be upper bounded by $2(2^1 + 2^2 + \ldots + 2^j)c = 2(2^{j+1} - 2)c$, for some constant $c > 0$.

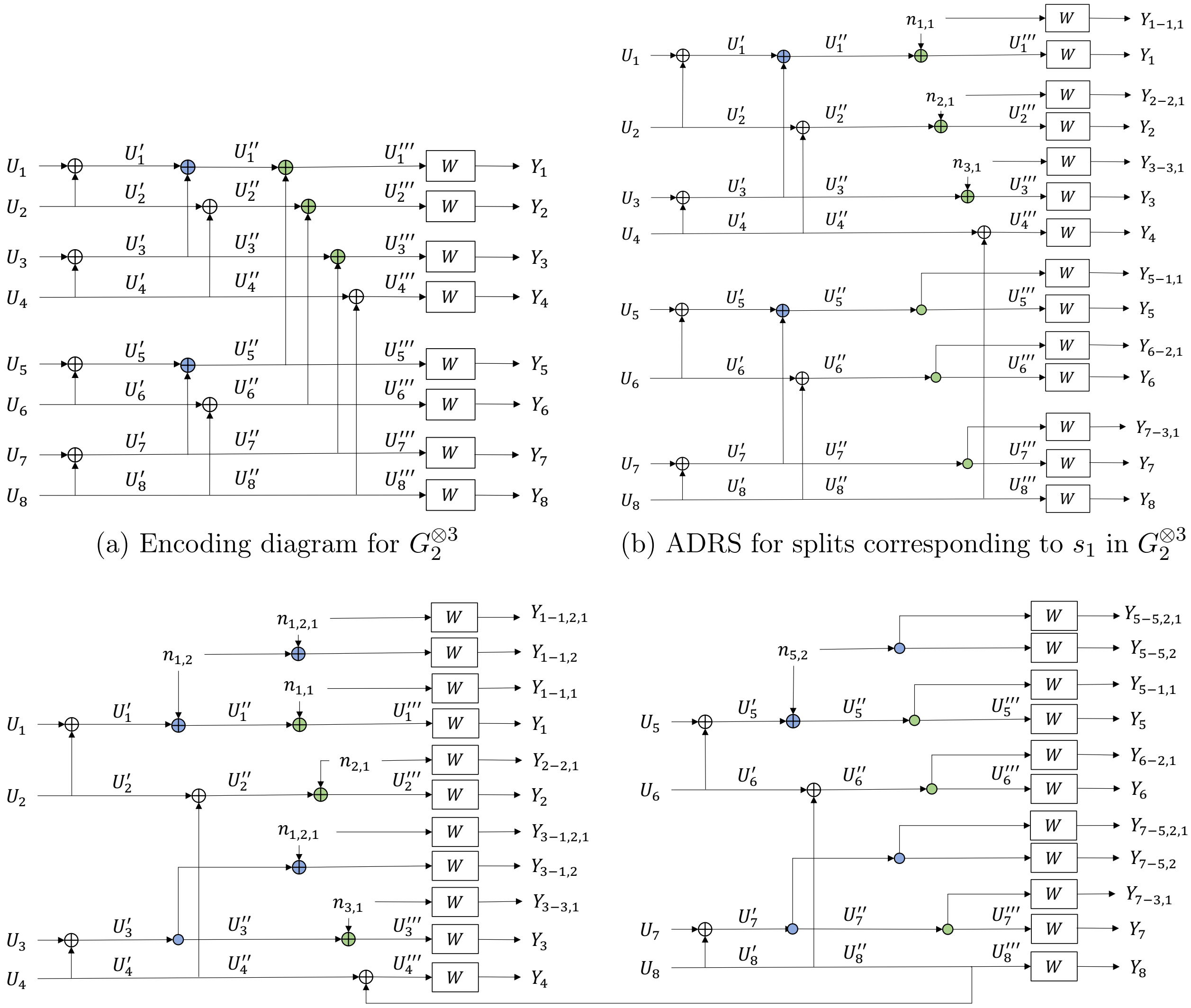
3.6.2 Polar-ADRS Code Performance

We are now ready to show the performance of the polar-based code whose encoding structure is given by the ADRS scheme, referred to as the *polar-ADRS* code. First we show the existence of a low-complexity decoder.

Proposition 16. *Let a constant* $\lambda > \lambda^\dagger \triangleq (\log_2 3)^{-1} \approx 0.631$ *be given. The decoding complexity for a SC decoder for the polar-ADRS code is bounded by* $\mathcal{O}(N \log N)$ *for all sufficiently large* n *if the threshold for the DRS algorithm is* $w_{u.b.} = 2^{n\lambda}$.

Proof: The proof is provided in Appendix A.4.1. ■

Second, it can be observed that the number of additional copies of channels due to the modification for an XOR operation at the j-th polarization recursion is 2^j. We find the total number of extra channel uses and the ratio γ of that to the number $N = 2^n$ of channel



(a) Encoding diagram for $G_2^{\otimes 3}$

(b) ADRS for splits corresponding to s_1 in $G_2^{\otimes 3}$

(c) ADRS Encoding Diagram for $G_2^{\otimes 3}$ with $w_{u.b.} = 2$

Figure 3.7: ADRS example with $N = 8$ and $w_{u.b.} = 2$

uses for the code corresponding to $G_2^{\otimes n}$ in the following. Assume that the column weight threshold of the DRS algorithm is given by $w_{u.b.} = 2^{n\lambda}$.

Proposition 17. *Let $N(1+\gamma)$ be the number of channel uses of the encoder for the ADRS scheme based on* DRS*($G_2^{\otimes n}$) with $w_{u.b.} = 2^{n\lambda}$. Then the term γ goes to* 0 *as n grows large, if we have $\lambda > \lambda^{\dagger}$.*

Proof: The proof is provided in Appendix A.4.2. ■

We are ready to show the existence of a sequence of capacity-achieving codes over general BMS channels with GMs where the column weights are bounded by a polynomial in the blocklength, and that the block error probability under a low complexity decoder vanishes as n grows large. Note that while this result is also applicable when the underlying channel is a BEC, the constraint on λ is stricter than that in Theorem 15, due to the difference in the encoding and decoding schemes.

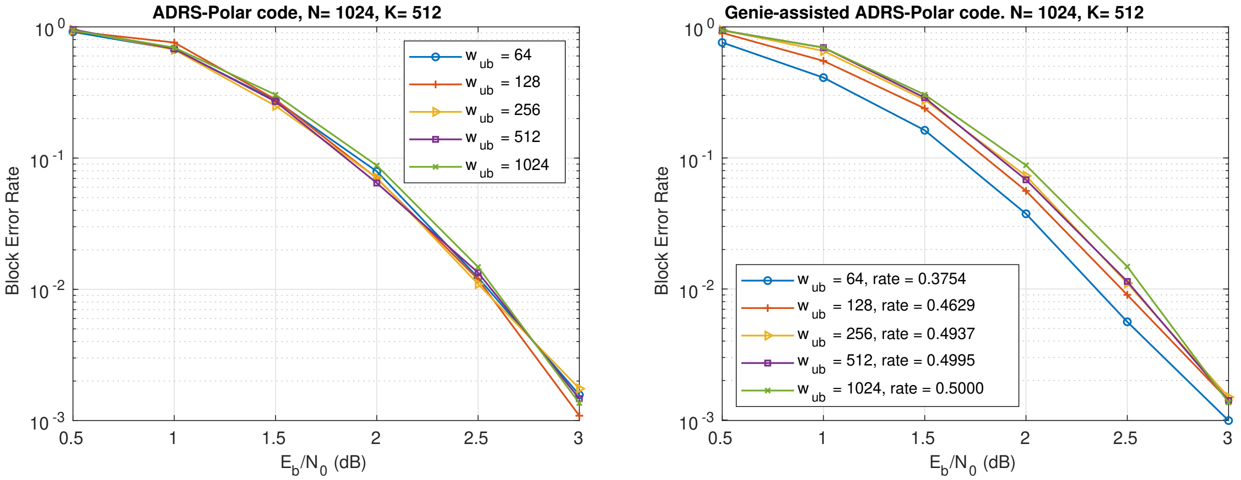
Theorem 18. *Let $\beta < E(G_2) = 0.5$, $\lambda > \lambda^\dagger$, and a BMS channel W with capacity C be given. There exists a sequence of codes with the following properties for all sufficiently large n: (1) The error probability under SC decoding is upper bounded by 2^{-N^β}, where $N = 2^n$, (2) The Hamming weight of each column of the GM is upper bounded by N^λ, (3) The rate approaches C as n grows large, and (4) The codes can be decoded by a SC decoding scheme with complexity $\mathcal{O}(N \log N)$.*

Proof: We prove the four properties in order as follows. First, similar to the proof of Theorem 15, for $i = 1, 2, \ldots, N$, the bit U_i is frozen in the polar-ADRS code with rate $R < C$ if and only if it is frozen in the polar code with kernel G_2, blocklength $N = 2^n$, and the rate R. Hence, the probability of error of the polar-ADRS code can be bounded in the same way as its polar-code counterpart, since the bit-channels observed by the source bits U_i, and the corresponding Bhattacharyya parameters, are identical to those when they are encoded with the standard polar code.

Second, when the ADRS scheme is based on $\mathrm{DRS}(G_2^{\otimes n})$ with $w_{u.b.} = 2^{n\lambda}$, the generator matrix for the polar-ADRS code is a submatrix of $\mathrm{DRS}(G_2^{\otimes n})$. The column weights of the GM for the polar-ADRS code are thus upper bounded by $w_{u.b.} = 2^{n\lambda} = N^\lambda$. The third claim holds by using an argument similar to the one used in the proof of Theorem 15. This is because the term γ vanishes as n grows large according to Proposition 17. Finally, note that the fourth claim is equivalent to Proposition 16. ■

We evaluate the performance of the polar-ADRS codes with $N = 1024, K = 512$ and weight thresholds $w_{u.b.} \in \{64, 128, 256, 512, 1024\}$, under the SC decoding scheme, over the BI-AWGN channels whose SNR (E_b/N_0) ranges from 0.5 to 3 dB. The indices of frozen bits are designed using a Monte-Carlo scheme for the BI-AWGN channel with $E_b/N_0 = 2$ dB. The auxiliary noise variables are generated with independent Bernoulli(0.5) distribution. The stopping criterion for each data point is ITERATION $= 10^5$ or BLOCK ERROR COUNT $= 100$, whichever is reached first. In Figure 3.8a, the block error probabilities are very close for codes with different weight thresholds over the range of simulated SNR. This is because the bit-channels observed by each information bit are exactly the same with or without the ADRS scheme, and that the block error probability of the codes are determined by the characteristics of the information bit channels.

Note that the use of the auxiliary noise variables are of theoretical reasons and is required for the proof of Theorem 18. In practice, we may freeze the auxiliary noise variables to



(a) Random auxiliary noise nodes

(b) Frozen auxiliary noise nodes

Figure 3.8: Error probability for polar-ADRS codes with $N = 1024, K = 512$

avoid the channel uses for the transmission of them. In particular, when the auxiliary noise variables are frozen to fixed values known to both the encoder and decoder, the decoder can be considered as a genie-assisted decoder and its performance is better or equivalent to that of a decoder without the information of the auxiliary noise variables. We simulate the performance of the ADRS-polar code with $N = 1024, K = 512$ and weight thresholds $w_{u.b.} \in \{64, 128, 256, 512, 1024\}$, where the values of the auxiliary noise variables are fixed to 0's, over the BI-AWGN channels with $0.5 \leqslant E_b/N_0 \leqslant 3$ (dB), and obtain Figure 3.8b. The error rates of codes with small thresholds are lower than those with large or no thresholds, a trend which is similar to that observed in Figure 3.5a.

3.7 Conclusion and Outlook

This chapter provided three constructions for capacity-achieving linear codes, based on polar coding, where all the GM column weights are upper bounded sublinearly in the block length. The first construction is a sequence of polar codes based on general polarization kernels where the GM column weights are upper bounded by N^s for any fixed $s > 0$, and allows the codes to be decoded by a SC decoder. In order to attain a better trade-off between the GM sparsity and the fall in error probability, we then proposed a column-splitting algorithm for the GM, termed the DRS algorithm. With the DRS algorithm, we designed two encoding schemes which yield two polar-based codes, referred to as polar-DRS codes and polar-ADRS codes, that are decodable with low-complexity decoders for the BECs and general BMS channels, respectively. The polar-based codes preserve several fundamental properties of the standard polar code with G_2 kernel including the asymptotic error rate upper bound and decoding complexity. Further, the GM column weights of the polar-DRS and polar-ADRS codes are bounded from above by N^{λ}, for $\lambda \approx 0.585$ and $\lambda \approx 0.631$, respectively, while the best bound for the standard polar codes scales linearly in N. The proposed constructions are also distinct from known constructions for codes with constraints on the GM sparsity by having analytical error probability upper bounds scaling as $\mathcal{O}(2^{-N^t})$ under SC decoders, for any $t < 0.5$. A future direction is to design splitting algorithm and/or encoding schemes that preserve key properties of the polar codes based on general polarization kernels, and show that the corresponding polar-based codes exhibit better sparsity versus error rate trade-off.

CHAPTER 4

New Bounds on the Size of Binary Codes with Large Minimum Distance

4.1 Introduction

An error-correcting code C of length n and minimum distance d over a finite field $\mathbb{F}_q$ is a subset of the vector space $\mathbb{F}_q^n$ with $d = \min d_H(x,y)$, over all distinct $x, y \in C$. Here, $d_H(x,y) = \sum_{i=1}^{n} \mathbb{1}_{\{x_i \neq y_i\}}$ is the Hamming distance between x and y. The code C is said to be linear if C is a subspace of the vector space $\mathbb{F}_q^n$. The capabilities and limitations of error-correcting codes are, in general, closely related to their minimum distance. For instance, the maximum number of errors a code can correct in the Hamming space is upper bounded by half of its minimum distance. This has led to a vast range of studies spanning several decades to answer one of the most fundamental and classical problems in coding theory, which is to determine (or to derive bounds on) the maximum size $A_q(n,d)$ of an error-correcting code C of length n over $\mathbb{F}_q$ and with minimum distance d [94, 59, 134]. Several of the most well-known results in the literature focus on the regime where $n \to \infty$ and d is proportional to n, namely $d = \delta n$, for some $0 < \delta < 1$. The question then is to find the asymptotic maximal rate $R(\delta)$ of an error-correcting code with relative distance δ, where we define $R(\delta) \overset{\text{def}}{=} \limsup_{n\to\infty} \frac{1}{n} \log_q A_q(n, \lfloor n\delta \rfloor)$.

Lower bounds on $A_q(n,d)$ are often obtained by constructions, either explicitly or implicitly, i.e., via existence arguments. One of the most well-known lower bounds on $A_q(n,d)$ is the Gilbert–Varshamov (GV) bound [53, 135]:

$$A_q(n,d) \geqslant \frac{q^n}{V_q(n,d-1)},$$

where $V_q(n,d) = \sum_{i=0}^{d} \binom{n}{i} (q-1)^i$ is size of a Hamming ball of radius d in $\mathbb{F}_q^n$. Several improvements have been proposed to strengthen the GV bound, see, e.g., [42, 17, 107, 71, 130, 143]. Among them, the most notable improvement in the binary case is due to Jiang and Vardy [71], who improved the GV lower bound on $A_2(n, d)$ for $\delta < 0.499$ by a multiplicative factor of $c \log_2 V_2(n, d)$, for some constant c, via studying the independence number of the sparse Gilbert graph on $\mathbb{F}_2^n$. For prime powers $q = p^{2k}$ with $q \geqslant 49$, explicit constructions of q-ary linear codes, obtained through algebraic-geometric codes, that surpass the GV bound are known [132]. For $q = 2$, a well-known conjecture asserts that the binary version of the GV bound is asymptotically tight, when expressed as a lower bound on $R(\delta)$. For a survey on the known bounds with finite n and d, the reader is referred to [87] and the websites [6, 22]. For asymptotic lower bounds and an overview of known results the reader is referred to [71, 48].

The best asymptotic upper bounds currently known are due to McEliece, Rodemich, Ramsey and Welch (MRRW) [103]. Built upon Delsarte's linear program (LP) approach [32], these bounds are established by showing valid solutions to the dual LPs, and are often called the first and the second linear programming bounds. In addition to the original proof in [94], which utilizes Delsarte's LP and properties of Krawtchouk polynomials, the first LP bound has also been proved using various techniques including harmonic analysis of Boolean functions [46, 106, 122], spectral analysis [20], and functional and linear-algebraic approaches [19]. There is substantial empirical evidence [18] indicating that the bounds in [103] asymptotically give the exact answer in the asymptotic Delsarte problem. Consequently, several works introduce new hierarchies of LPs, which include Delsarte's LP as the weakest member of this family [29, 90, 91]. Among them, the concurrent works [29] and [90] study similar families of LPs applicable to linear codes only, and empirically show significant improvement compared to Delsarte's. In [91], new hierarchies of LPs for linear and general codes along with the first dual feasible solutions to the LP's that recover the first LP bound, are developed.

4.1.1 Motivation

Low-rate codes are becoming increasingly important with the emergence of low-capacity scenarios, including the Internet of Things (IoT) and satellite communications. For instance, in IoT network, the devices need to operate under extreme power constraints and often need to communicate at very low signal-to-noise ratio [116]. In the standard, legacy Turbo codes or convolutional codes at moderate rates together with many repetitions are adopted to support communication at low rates. It is expected, however, that repeating a moderate-rate code to enable low-rate communication will result in rate loss and suboptimal performance. As a result, studying low-rate error-correcting codes for reliable communications in such low-capacity regimes has become a subject of extensive recent works [127, 44, 1, 38, 2, 40, 4, 70, 39, 133, 3].

Motivated by the need to revisit various aspects of channel coding in the low-rate regime, from efficient code design to reliable decoding algorithms, in this dissertation we focus on the minimum distance properties of binary codes in the low-rate regime, which can be also described as the large minimum distance regime, to be specified later. Let C be a binary code of length n, size M, and minimum distance $d = (n-j)/2$, referred to as an (n, M, d) code (for the sake of simplifying the equations, we reserve the parameter j to denote $n - 2d$ in various places throughout the chapter). With a slight abuse of terminology, the *dimension* of C, including for non-linear codes, is denoted by $k = \log_2 M$. Also, let $R = k/n$ denote the code rate. In this chapter, we focus on studying bounds on $A_2(n, d)$ in the large-minimum distance regime, in particular, when $d = n/2 - \Omega(\sqrt{n})$, i.e., $j = \Omega(\sqrt{n})$. For ease of notation, we use $A(n, d)$ to denote $A_2(n, d)$ throughout the chapter keeping in mind that the focus is on studying binary codes.

4.1.2 Related Works

For $j = n - 2d \leqslant 0$, provided that a sufficient number of Hadamard matrices exist, a widely accepted conjecture, Plotkin and Levenshtein (see [94, Chapter 2, Theorem 8]) have essentially settled the problem and showed that $A(2d, d) = 4d$, $A(n, d) = 2\lfloor d/(2d-n) \rfloor$ for even $d > n/2$, and $A(n, d) = 2\left\lfloor \frac{d+1}{2d+1-n} \right\rfloor$ for odd $d > (n-1)/2$.

In what follows, we consider the scenario with $j > 0$. When j scales linearly with n, asymptotic results can be found in [59, 71]. In particular, the conjecture is that there does not exist any binary code exceeding the GV lower bound (Theorem 19). There are a limited number of studies in the literature targeting the regime where j is sub-linear in n. In 1973, McEliece (see [94, Chapter 17, Theorem 38]), utilizing the LP approach, established the following bound that is valid for $j = o(\sqrt{n})$:

$$A(n, d) \lesssim n(j+2). \tag{4.1}$$

For $j \approx n^{1/3}$, codes have been constructed [128] to meet McEliece's upper bound, and hence, showing the tightness of this bound in this regime. A few improvements [131, 79] have been derived in the literature in the regime $j = o(n^{1/3})$. For $j = \Omega(\sqrt{n})$, the Delsarte–Goethals (DG) codes, first introduced as a generalization of Reed-Muller codes [77, 33, 57, 56, 63], are a class of nonlinear code and are known to be the best known codes, in terms of the minimum distance given a code size, in this regime. When $j = \Theta(\sqrt{n})$, a sequence of DG codes with sizes scaling polynomially in n can be constructed.

While no explicit upper bounds on $A(n, d)$ are derived in the literature targeting the specific regime $j = \Omega(\sqrt{n})$, the results by Barg and Nogin [20] can be tailored to provide a sequence of bounds on $A(n, d)$ scaling polynomially in n for $j = \Theta(\sqrt{n})$. To the best of our knowledge, this is the only existing result in the literature leading to upper bounds that scale polynomially with n in this regime.

4.1.3 Our Contribution

We study the cardinality $A(n, d)$ of binary codes in the large minimum distance regime where $j = n - 2d$ scales as $j = \Omega(\sqrt{n}) \cap o(n)$. In particular, we show the following results:

- Two code constructions with sizes scaling polynomially and quasi-polynomially in n are presented for cases with $d = n/2 - \Omega(\sqrt{n})$ and $d = n/2 - \Omega(n^{2/3})$, respectively, by demonstrating explicit and carefully designed BCH-like cyclic linear codes. Specifically, for $c \in \mathbb{N}$, the first construction has size $n^{c+\frac{1}{2}}$ and $d \geqslant n/2 - 2^{c-1}\sqrt{n}$, and the second construction has size $n^{\frac{\log n}{6}+\frac{3c^2}{2}+\frac{5c}{2}+\frac{5}{6}}$ and $d \geqslant n/2 - 2^{c-1}n^{2/3} - 2^{2c-1}n^{1/3}$.

- Compared with the state-of-the-art lower bounds on $A(n, d)$ based on the Delsarte–Goethals codes, the first cyclic construction is inferior by a multiplicative factor of $\Theta(n^{3/2})$ in the regime $j = \Theta(\sqrt{n})$. In the regime $j = \Theta(n^{2/3})$, the second construction is superior to the DG codes by a multiplicative factor of $\Theta(n^{\frac{3}{2}c^2+\frac{3}{2}c-1})$ and provides the best lower bound in this regime.

- Asymptotic upper bounds for $A(n, n/2 - \rho\sqrt{n})$, based on an improved bounding technique inspired by [106] and a new method to bound the maximal eigenvalues of adjacency matrix induced by a Hamming ball $B_r \in \{0, 1\}^n$ with finite r, are shown.

- The asymptotic scaling behaviour of the proposed Fourier-analytical based upper bounds for $A(n, n/2 - \rho\sqrt{n})$ and the spectral-based bounds derived from [20], both of which are polynomial in n, are plotted for $\rho \in (0.5, 9.5)$, where the former are slightly stronger.

The rest of this chapter is organized as follows. In Section 4.2 we review several well-known bounds on $A(n, d)$ and examine their scaling behaviour when $j = \Theta(\sqrt{n})$. Results in a prior literature by Barg and Nogin [20] that can be used to provide upper bounds on $A(n, d)$ when $j = \Theta(\sqrt{n})$ are discussed in Section 4.2.3. In Section 4.3.1, a BCH-like cyclic code construction, with j scaling from $\Theta(\sqrt{n})$ to $\Theta(n)$, is presented. Section 4.3.2 describes another construction with better performance in the regime $j = \Omega(n^{\frac{2}{3}})$. In Section 4.4.1, we review an alternative proof of the first linear programming bound on $A(n, d)$ (formally decribed in Section 4.2.2) through a covering argument using Fourier analysis on the group $\mathbb{F}_2^n$. An asymptotic upper bound on $A(n,d)$ with $d \geqslant n/2 - \sqrt{n}$ that are strictly tighter than all prior results is derived in Section 4.4.2. The upper bounding technique is extended in Section 4.4.3 and yields a family of bounds on $A(n,d)$ with $d \geqslant n/2 - \rho\sqrt{n}$ for $\rho \in (0.5, 9.5)$. The bounds are compared with a sequence of bounds derived from [20] for the same range of d in Section 4.4.4. Finally, the chapter is concluded in Section 4.5.

4.2 Preliminaries

Notation. Let f and g be two real-valued functions of $n \in \mathbb{N}$. We write $f(n) \lesssim g(n)$ if $f(n) \leqslant (1+o(1))\, g(n)$, write $f(n) \gtrsim g(n)$ if $f(n) \geqslant (1+o(1))\, g(n)$, and write $f(n) \sim g(n)$ if $\lim_{n \to \infty} f(n)/g(n) = 1$. Let $H_2(\cdot)$ denote the binary entropy function. For positive integers $r, n \in \mathbb{N}$ with $n \geqslant r$, let $B_r(\mathbf{0}, n) \in \{0,1\}^n$ denote the Hamming ball of radius r centered at $\mathbf{0} = (0,0,\ldots,0)$, and its volume by $\mathrm{Vol}(r,n) \overset{\text{def}}{=} |B_r(\mathbf{0},n)| = \sum_{i=0}^{r} \binom{n}{i}$. When n is clear from the context, we write B_r and $B_r(\mathbf{0},n)$ interchangeably. We recall the following bounds for $r \leqslant n/2$

1) $\mathrm{Vol}(r,n) \leqslant 2^{H_2(r/n)n}$; and

2) $\mathrm{Vol}(r,n) \geqslant 2^{H_2(r/n)n - o(n)}$ for sufficiently large n.

We study asymptotic lower and upper bounds on $A(n,d)$ in this section, and evaluate them in the large minimum distance regime $j = \Omega(\sqrt{n}) \cap o(n)$.

4.2.1 Lower Bounds

We review some asymptotic lower bounds on $A(n,d)$ in this section. The first one is the well-known GV lower bound. Note that there is an improvement to the GV bound by Jiang and Vardy [71] that is not considered here because the constraint on the relative distance $0 \leqslant \delta < 0.499$ in [71] does not hold for large n when $j = \Omega(\sqrt{n}) \cap o(n)$.

Theorem 19 (GV lower bound, [53, 135])**.** *Let positive integers n and $d \leqslant n/2$ be given. Then*

$$A(n,d) \geqslant \frac{2^n}{\mathrm{Vol}(d-1,n)}. \tag{4.2}$$

Asymptotically, suppose $0 \leqslant \delta < 1/2$, then there exists an infinite sequence of (n, M, d) binary linear codes with $d/n > \delta$ and rate $R = k/n$ satisfying $R \geqslant 1 - H_2(d/n)$. To evaluate Theorem 19 when $j = \Theta(\sqrt{n})$, consider $j = 2a\sqrt{n}$. The central limit theorem, coupled with the Berry–Esseen theorem, provides an upper bound $\mathrm{Vol}(d-1,n) = 2^n\left[Q(2a) + O(1/\sqrt{n})\right]$, where $Q(\cdot)$ denotes the tail distribution function of the standard normal distribution. Hence we have

$$A(n, n/2 - a\sqrt{n}) \geqslant \left[Q(2a) + O(1/\sqrt{n})\right]^{-1}, \tag{4.3}$$

which is loose compared with the Plokin-Levenshtein bound $A(2d, d) = 4d$.

The Delsarte–Goethals (DG) codes are a class of nonlinear codes that are associated with the Reed-Muller codes and are the best known codes for their parameters.

Theorem 20 (Delsarte–Goethals code [77, 33, 57, 56, 63])**.** *Let $m \geqslant 4$ be an even integer and $0 \leqslant r \leqslant m/2 - 1$ be an integer. The Delsarte–Goethals code DG(m, r) is a binary code of block length $n = 2^m$, size 2^k, where $k = r(m-1) + 2m$, and minimum distance $2^{m-1} - 2^{m/2+r-1}$. For $r = m/2 - 1$, DG(m, r) = RM$(m, 2)$, the second order Reed-Muller code. For $0 \leqslant r \leqslant m/2 - 2$, DG$(m, r)$ is a nonlinear subcode of RM$(m, 2)$.*

For the case when $j = \Theta(\sqrt{n})$, one may consider DG(m, r) codes with a finite r, and show that $A(n, d) \geqslant 2^{-r} n^{r+2}$ for n an even power of 2 and $d = n/2 - 2^{r-1}\sqrt{n}$. For the case when $j = \Theta(n^{2/3})$, considering DG(m, r) codes with m a multiple of 6 and $r = m/6 + c$ for some finite c, one may show that $A(n, d) \geqslant n^2 (n/2)^{\frac{\log n}{6} + c}$, where $n = 2^{6\ell}$ for some $\ell \in \mathbb{N}$ and $d = n/2 - 2^{c-1} n^{2/3}$.

4.2.2 Asymptotic Upper Bounds

The following upper bounds on the size of binary codes can be found in standard coding theory textbooks, e.g. [94],[59]. Bounds for the regime $j = n - 2d = \Theta(\sqrt{n})$ are derived and given following the general bounds, e.g. inequalities (4.5), (4.6), (4.7), and (4.8). When the scaling behaviour of j matters, we choose $j = 2a\sqrt{n}$, i.e., $d = n/2 - a\sqrt{n}$, for ease of comparison between bounds.

Theorem 21 (Hamming Bound)**.** *For every (n, M, d) code $C \subset \{0, 1\}^n$,*

$$M \leqslant 2^n / \operatorname{Vol}(e, n), \tag{4.4}$$

where $e = \lfloor (d-1)/2 \rfloor$.

In the asymptotics, Theorem 21 bounds the rate from above, in terms of the relative distance δ, by $R \lesssim 1 - H_2(\delta/2)$. For $j = \Theta(\sqrt{n})$, the term $e = n/4 - \Theta(\sqrt{n})$, and $\operatorname{Vol}(e, n) \geqslant 2^{H_2(1/4)n - o(n)}$. Hence Theorem 21 becomes

$$M \leqslant 2^{(1 - H_2(1/4))n + o(n)} \lesssim 2^{0.189n}, \tag{4.5}$$

for all sufficiently large n.

Theorem 22 (Singleton Bound)**.** *Let $C \subset \{0,1\}^n$ be a binary code with distance d and dimension k, then $k \leqslant n - d + 1$.*

For $j = \Theta(\sqrt{n})$, Theorem 22 yields $M \leqslant 2^{n/2+O(\sqrt{n})}$, which is weak compared to (4.5).

Theorem 23 (Plotkin Bound, [115])**.** *The following holds for any code $C \subset \{0,1\}^n$ with distance d.*

1) If $d = n/2$, $|C| \leqslant 2n$.

2) If $d > n/2$, $|C| < 2\left\lceil \frac{d}{2d-n} \right\rceil$.

One may use a combinatorial argument and Theorem 23 to derive the following corollary.

Corollary 24. *If a (n, M, d) binary code C has distance $d < n/2$, then the size $M \leqslant d \cdot 2^{n-2d+2}$.*

Using Corollary 24, one may bound the size of any code with $d = (n-j)/2 < n/2$ by

$$M \leqslant d \cdot 2^{j+2} < 2n \cdot 2^j. \tag{4.6}$$

When j scales as $j = \Theta(\sqrt{n})$, the size M is bounded sub-exponentially in n. In particular, set $j = 2a\sqrt{n}$, i.e. $d = \lceil n/2 - a\sqrt{n}\,\rceil$, (4.6) becomes

$$M \leqslant 2n \cdot 2^{2a\sqrt{n}}. \tag{4.7}$$

Theorem 25 (Elias-Bassalygo Bound)**.** *For sufficiently large n, every code $C \subset \{0,1\}^n$ with relative distance $\delta \leqslant 1/2$ and rate R satisfies the following: $R \lesssim 1 - H_2(J_2(\delta))$, where $J_2(\delta) \overset{\text{def}}{=} \frac{1}{2}(1 - \sqrt{1-2\delta})$.*

Assuming $d = \lceil n/2 - a\sqrt{n}\,\rceil$, one may adopt steps similar to the proof of Theorem 25 as in [59, p.147] to show an upper bound:

$$M \leqslant n^3 \cdot 2^{\frac{a}{\ln 2}\sqrt{n}+O(1)}. \tag{4.8}$$

The last upper bound we introduce is known as the *first linear programming bound* or the *MRRW bound* on binary error correcting codes, or, alternatively, on optimal packing of Hamming balls in a Hamming cube. The bound was originally proved by McEliece, Rodemich, Rumsey, and Welch [103], following Delsarte's linear programming approach [32], and is the best known asymptotic upper bound on the cardinality of a code with a given minimal distance scaling linearly in n, for a significant range of the relative distance.

Theorem 26 (MRRW Bound, [103]). *For sufficiently large n, every code $C \subset \{0,1\}^n$ with relative distance δ and rate R satisfies the following:*

$$R \lesssim H_2\left(1/2 - \sqrt{\delta(1-\delta)}\right). \tag{4.9}$$

Remark 5. Another bound, known as the *second linear programming bound*, is also given in [103] in the form

$$R \lesssim \min_{0 \leqslant u \leqslant 1-2\delta} 1 + g(u^2) - g(u^2 + 2\delta u + 2\delta), \tag{4.10}$$

where the function $g(x) \overset{\text{def}}{=} H_2((1-\sqrt{1-x})/2)$. For $0.273 \leqslant \delta \leqslant 0.5$, the bound (4.10) simplifies to that of (4.9). For $\delta < 0.273$, the inequality (4.10) is strictly tighter than (4.9).

Plugging in $\delta = d/n$ into (4.9), we have the following bound:

$$M \leqslant 2^{nH_2\left(1/2 - \sqrt{d/n(1-d/n)}\right) + o(n)}. \tag{4.11}$$

Note that, due to the $o(n)$ term, the bound (4.11) is not tighter than (4.6) when $j = \Theta(\sqrt{n})$. This appears to the contrary of the fact the MRRW bound is tighter than all the other bounds for relative distance $0.273 < \delta < 0.5$. However, a tailored treatment of the proof technique may lead to a nontrivial bound as in the derivation of (4.8) from Theorem 25. In Section 4.4.2, one such bound is given through an alternative proof of the Theorem 26 by working with the maximal eigenfunctions of Hamming balls.

4.2.3 Spectral-Based Upper Bound

One approach to proving the first linear programming bound is the spectral-based technique in [20], which relies on the analysis of eigenvectors of some finite-dimensional operators related to the Krawtchouk polynomials. While the main goal of the work [20] is to establish the MRRW bounds from a spectral perspective, some of the analytical results in it can be used to derive upper bounds on $A(n,d)$ in the large minimum distance regime. In particular, while all the other upper bounds on $A(n,d)$ scale superpolynomially in n when $j = \Theta(\sqrt{n})$, a sequence of bounds scaling polynomially in n can be derived from [20]. A key result in [20] is the following bound on the size of a binary code with minimum distance d.

Theorem 27 ([20] Theorem 2, binary case). *Let C be an (n, M, d) binary code. Then*

$$M \leqslant \frac{4(n-k)}{n-\lambda_k}\binom{n}{k} \tag{4.12}$$

for all k such that $\lambda_{k-1} \geqslant n-2d$, where λ_k is the maximal eigenvalue of the $(k+1)\times(k+1)$ self-adjoint matrix $S=(s_{i,j})_{i,j=1}^{k+1}$ defined by $s_{i,i+1}=s_{i+1,i}=\sqrt{i(n+1-i)}$ for $i=1,2,\ldots,k$ and $s_{i,j}=0$ otherwise.

Upper and lower bounds on λ_k are also provided.

Lemma 28 ([20] Lemma 2, binary case)**.** *Let $k<n/2$. For all $s=2,\ldots,k+1$,*

$$2\sqrt{k(n-k+1)} \geqslant \lambda_k \geqslant \frac{2(s-1)}{s}\sqrt{(k-s+2)(n-k+s-1)}.$$

To establish bounds on $A(n,d)$ for the regime $d=n/2-\Theta(\sqrt{n})$, consider a finite $k\in\mathbb{N}$ and $s\in\{2,\ldots,k+1\}$. Letting $n\to\infty$, we have

$$\lambda_k \geqslant \frac{2(s-1)}{s}\sqrt{(k-s+2)(n-k+s-1)} = \frac{2(s-1)}{s}\sqrt{k-s+2}(1+o(1))\sqrt{n}.$$

Thus $\lambda_k \gtrsim \underline{\lambda}_k\sqrt{n}$, where $\underline{\lambda}_k$ is given by

$$\underline{\lambda}_k = \max_{2\leqslant s\leqslant k+1}\left\{\frac{2(s-1)}{s}\sqrt{k-s+2}\right\}. \tag{4.13}$$

Since λ_k scales as $\Theta(\sqrt{n})$ for all finite k, the bound on $A(n,d)$ is asymptotically equivalent to

$$A(n,d)=O(n^k) \text{ as long as } d\geqslant \frac{n}{2}-\frac{\underline{\lambda}_{k-1}}{2}\sqrt{n}. \tag{4.14}$$

By solving (4.13) for $k=1,2,3$, we obtain $\underline{\lambda}_1=1$, $\underline{\lambda}_2=\sqrt{2}$, $\underline{\lambda}_3=\frac{4}{3}\sqrt{2}$, which, via (4.14), lead to $A(n,d_1)=O(n^2)$, $A(n,d_2)=O(n^3)$, $A(n,d_3)=O(n^4)$ as long as $d_1\geqslant\frac{n}{2}-\frac{1}{2}\sqrt{n}$, $d_2\geqslant\frac{n}{2}-\frac{\sqrt{2}}{2}\sqrt{n}$, $d_3\geqslant\frac{n}{2}-\frac{2\sqrt{2}}{3}\sqrt{n}$, respectively. Many more bounds for the large minimum distance regime can be obtained by choosing other $k\in\mathbb{N}$. These bounds and the new upper bounds shown in Section 4.4.3 are both polynomial in n and are tighter than all other known bounds, as discussed in Section 4.2.2. In Section 4.4.4, the asymptotic behavior of the two types of bounds when $d=n/2-\rho\sqrt{n}$ are plotted for $\rho\in(0.5,9.5)$.

4.3 Main Results - Lower Bounds

Two polynomial-based cyclic code constructions are given in this section. The first construction, described in Section 4.3.1, leads to a family of codes where the term j ranges from $\Theta(\sqrt{n})$ to $\Theta(n)$. The second construction, described in Section 4.3.2, applies to a smaller range of j, between $\Theta(n^{2/3})$ and $\Theta(n)$, but are tighter than the first construction over this range. Note that the results in this section are not asymptotic and hold for finite values of n, i.e., the first construction only requires $n \geqslant 15$ and the second one requires $n \geqslant 63$.

4.3.1 Cyclic Codes with $j = \Omega(\sqrt{n})$

We construct a binary cyclic code C with high minimum distance as follows.

Theorem 29. *Let $n = 2^m - 1$, and $m \in \mathbb{N}$ be an even integer with $m \geqslant 4$. Let c be an integer with $0 \leqslant c \leqslant m/2 - 1$. There exists a binary cyclic code C of length n, dimension $(c + 1/2)m$, and minimum distance*

$$d \geqslant 2^{m-1} - 2^{m/2+c-1} \geqslant n/2 - 2^{c-1}\sqrt{n}. \tag{4.15}$$

Proof: Consider the finite field $F = \mathbb{F}_{2^m}$ and the subfield $K = \mathbb{F}_2 < F$. Let α be a primitive root of unity in F, and set $\alpha_i = \alpha^{1+2^{m/2+i}}$ for $i = 0, 1, 2, \ldots, c$. Consider the binary cyclic code with the generator polynomial

$$g(x) = \frac{x^n - 1}{\prod_{i=0}^{c} M_{\alpha_i}(x)},$$

where $M_\beta(\cdot)$ is the minimal polynomial of β over K. Note that the α_i's belong to different conjugacy classes, i.e,

$$A_i \overset{\text{def}}{=} \left\{ \alpha_i^{2^j} \mid j = 0, 1, 2, \ldots, m-1 \right\} = \left\{ \alpha^{2^j + 2^{m/2+i+j}} \mid j = 0, 1, 2, \ldots, m-1 \right\}$$

are disjoint subsets of $F \setminus \{0\}$, and $|A_0| = m/2$, $|A_i| = m$ for $i \neq 0$. This is ensured by the particular choice of α_i's. More specifically, let

$$P_i \equiv \{2^j + 2^{m/2+i+j} \bmod 2^m - 1 \mid j = 0, 1, \ldots, m-1\}$$

be the set of the exponents of α for elements in A_i. Each P_i is a cyclotomic coset mod 2 in F and the length-m binary representation for each p, p' in P_i are cyclic shifts of each other. Let $p_i = 1 + 2^{m/2+i}$ be the coset representative of P_i. The claim on the size of $|A_i|$ holds by noting that $|A_i| = |P_i| = m$ for $i \neq 0$, and $|A_0| = |P_0| = m/2$. To claim that A_i's are disjoint, it suffices to show that the cyclotomic cosets P_i's are disjoint. First note that for two cyclotomic cosets P_i and P_k, they are either disjoint or identical. Assume for some $i \neq k$, cosets P_i and P_k are identical. Then $p_i = 1 + 2^{m/2+i}$ is an element in P_k, that is, there is a $p' = 2^\ell + 2^{m/2+k+\ell} \in P_k$ for which $p_i = p'$ modulo $2^m - 1$. As both p_i and p' are sums of two powers of 2, we note that neither $m \mid \ell$ and $m \mid (\ell + k - i)$, nor $m \mid (\ell - m/2 - i)$ and

$m \mid (m/2+k+\ell)$, can happen. Hence $p_i \notin P_k$, and thus P_i and P_k are disjoint. Thus the degree of the polynomial $g(x)$ is $n-(c+1/2)m$. Hence, the dimension of the code is at least $(c+1/2)m$.

For the minimum distance, let $t = 2^{m-1} + 2^{m/2+c-1} + 1$. We show next that for $j = t, t+1, \ldots, 2^m - 1$, α^j is a root for the generator polynomial $g(x)$. In other words, $P_i \cap \{t, t+1, \ldots, 2^m-1\} = \emptyset$, for $i = 0, 1, 2, \ldots, c$. This is by noting that the elements in P_i, after taking modulo $2^m - 1$, can be written as a sum of two powers of two, i.e., $2^\ell + 2^j$, where the difference between ℓ and j is at least $m/2 - c$, and that such a number does not belong to $\{t, t+1, \ldots, 2^m - 1\}$. Hence, the minimum distance of the code d is at least $2^m - t + 1 = 2^{m-1} - 2^{m/2+c-1}$ by BCH bound [65, 21] (see also [94, 117]). ■

Note that the parameters of the codes constructed in Theorem 29 and the Delsarte–Goethals codes are both sitting between those of the first order and the second order Reed–Muller (RM) codes of length $n = 2^m$. More specifically, RM$(m,1)$ has minimum distance equal to $n/2$ and dimension equal to $m+1$, while RM$(m,2)$ has minimum distance $n/4$, and dimension $1 + m + \binom{m}{2}$. A comparison between the DG codes and our new construction in the regime $j = \Theta(\sqrt{n})$ can be made as follows. Let $c \geqslant 0$ be a finite integer. The DG(m,c) code has length $n = 2^m$, minimum distance $d = n/2 - 2^{c-1}\sqrt{n}$ and size $2^{-c}n^{c+2}$. On the other hand, the code parameters given in Theorem 29 are length $n = 2^m - 1$, minimum distance $d \geqslant n/2 - 2^{c-1}\sqrt{n}$, and size $(n+1)^{c+1/2}$. Therefore the former leads to a stronger lower bound on the size, by a multiplicative factor of $2^{-c}n^{3/2}$, in the asymptotics.

4.3.2 Cyclic Constructions for $j = \Omega(n^{2/3})$

We adopt an approach similar to that in the proof of Theorem 29 to construct a sequence of cyclic binary codes with $j = n - 2d$ scaling as $j = \Omega(n^{2/3})$ in this section. This construction is preferred over that in Section 4.3.1 for all $j = \Omega(n^{2/3})$, and yields a tighter bound on $A(n,d)$ than the DG codes does.

Theorem 30. *Let $n = 2^m - 1$, and $m \in \mathbb{N}$ be a multiple of 6. Let c be an integer with $1 \leqslant c \leqslant m/3 - 1$. There exists a binary cyclic code C of length n, dimension $k = m\left(\frac{m}{6} + \frac{3c^2}{2} + \frac{5c}{2} + \frac{5}{6}\right)$, and minimum distance*

$$d \geqslant 2^{m-1} - 2^{\frac{2m}{3}+c-1} - 2^{\frac{m}{3}+2c-1} \geqslant n/2 - 2^{c-1}n^{\frac{2}{3}} - 2^{2c-1}n^{\frac{1}{3}}. \tag{4.16}$$

Proof: Consider the finite field $F = \mathbb{F}_{2^m}$ and the subfield $K = \mathbb{F}_2 < F$. Let α be a primitive root of unity in F. Let $\ell = m/3 - c$, and define three sets consisting of triples of integers,

$$\begin{aligned}
&S_1 = \{(m/3, m/3, m/3)\}\,,\\
&S_2 = \{(d_1, d_1, d_2) \mid d_1 \geqslant \ell, d_2 \geqslant \ell, d_1 \neq d_2, 2d_1 + d_2 = m\}\,,\\
&S_3 = \{(d_1, d_2, d_3) \mid d_1 > d_2 > d_3 \geqslant \ell \text{ or } d_1 > d_3 > d_2 \geqslant \ell, d_1 + d_2 + d_3 = m\}.
\end{aligned}$$

A combinatorial argument shows that the sizes of the sets are

$$|S_1| = 1, \quad |S_2| = \left\lfloor \frac{m-3l}{2} \right\rfloor, \quad |S_3| = \frac{1}{3}\left[\binom{m-3l+2}{2} - 3\left\lfloor \frac{m-3l}{2} \right\rfloor - 1\right].$$

For $(d, e, f) \in S_1 \cup S_2 \cup S_3 \overset{\text{def}}{=} S$, define $\alpha_{d,e,f} = \alpha^{p_{d,e,f}} \in F$ where $p_{d,e,f} = 2^{m-1} + 2^{e+f-1} + 2^{f-1}$. We define a binary cyclic code with the generator polynomial

$$g(x) = \frac{x^n - 1}{\prod_{(d,e,f)\in S} M_{\alpha_{d,e,f}}(x) \cdot \prod_{i=0}^{m/6+c} M_{\alpha_i}(x)}.$$

For $i = 0, 1, 2, \ldots, c$, define the sets $P_i = \{2^j + 2^{m/2+i+j} \bmod 2^m - 1 \mid j = 0, 1, 2, \ldots, m-1\}$ and $A_i = \{\alpha^p \mid p \in P_i\}$ (as in the proof of Theorem 29). Consider sets $A_{d,e,f}, P_{d,e,f}$ as follows:

$$\begin{aligned}
&A_{d,e,f} \overset{\text{def}}{=} \left\{\alpha_{d,e,f}^{2^j} \mid j = 0, 1, 2, \ldots\right\}\\
&P_{d,e,f} \overset{\text{def}}{=} \{2^{m-1+j} + 2^{e+f-1+j} + 2^{f-1+j} \bmod 2^m - 1 \mid j = 0, 1, 2, \ldots\}.
\end{aligned}$$

The set $P_{d,e,f}$ consists of the exponents of α for elements in $A_{d,e,f}$. Note that $|P_{d,e,f}| = m/3$ for $(d, e, f) \in S_1$, $|P_{d,e,f}| = m$ for $(d, e, f) \in S_2$, and $|P_{d,e,f}| = m$ for $(d, e, f) \in S_3$. Two cyclotomic cosets $P_{d,e,f}$ and $P_{d',e',f'}$ are disjoint as long as $(d, e, f) \neq (d', e', f')$. Since each element of P_i is a sum of two powers of 2 and each element of $P_{d,e,f}$ a sum of three powers of 2, we also have $P_i \cap P_{d,e,f} = \emptyset$, for all $i \in \{0, 1, 2, \ldots, c\}$ and $(d, e, f) \in S$. Thus the degree of the polynomial $g(x)$ is

$$\begin{aligned}
\deg g(x) &= n - (m\,|S_3| + m\,|S_2| + m/3\,|S_1|) - m\,(m/6 + c + 1/2)\\
&= n - m\left(m/6 + 3c^2/2 + 5c/2 + 5/6\right).
\end{aligned}$$

Hence, the dimension of the code is at least $m(m/6 + 3c^2/2 + 5c/2 + 5/6)$.

For the minimum distance, we proceed similarly to the steps taken in the proof of Theorem 29. Let $t = 2^{m-1} + 2^{m-1-\ell} + 2^{m-1-2\ell} + 1$. We show next that for $j = t, t+1, \ldots, 2^m - 1$, α^j is a root for the generator polynomial $g(x)$. In other words, $P_i \cap \{t, t+1, \ldots, 2^m - 1\} = \emptyset$, for $i = 0, 1, 2, \ldots, m/6 + c$, and $P_{d,e,f} \cap \{t, t+1, \ldots, 2^m - 1\} = \emptyset$ for $(d, e, f) \in S$. This is by noting that elements in P_i and $P_{d,e,f}$ are sums of two or three powers of 2, and the powers differ by at least ℓ. Such a number can not be found in $\{t, t+1, \dots, 2^m - 1\}$. Hence, by the BCH bound [65, 21] (see also [94, 117]), the minimum distance d of the code is at least $2^m - t + 1 = 2^{m-1} - 2^{m-1-\ell} - 2^{m-1-2\ell} = 2^{m-1} - 2^{2m/3+c-1} - 2^{m/3+2c-1}$. ■

For the regime $j = \Theta(n^{2/3})$ we compare the performance of the DG codes and the construction in Theorem 30. Let c be a positive integer. The DG(m, r) codes with $r = m/6 + c$ has length $n = 2^m$, minimum distance $d = n/2 - 2^{c-1} n^{2/3}$, and size $M = n^2 \left(\frac{n}{2}\right)^{\frac{\log n}{6}+c}$. The cyclic code described in the proof of Theorem 30 has length $n = 2^m - 1$, minimum distance $d' \geqslant n/2 - 2^{c-1} n^{\frac{2}{3}} - 2^{2c-1} n^{\frac{1}{3}}$, and size $M' \geqslant n^{\frac{\log n}{6} + \frac{3c^2}{2} + \frac{5c}{2} + \frac{5}{6}}$. While the terms $j = n - 2d = 2^c n^{\frac{2}{3}}$ and $j' = n - 2d' \leqslant 2^c n^{\frac{2}{3}} - 2^{2c} n^{\frac{1}{3}}$ are asymptotically equivalent, i.e., $j \gtrsim j'$, the bound of the size based on the new cyclic construction is stronger. More explicitly, we have $M' \geqslant 2^c n^{\frac{3c^2}{2} + \frac{3c}{2} - 1} \cdot M$, where the degree $\frac{3c^2}{2} + \frac{3c}{2} - 1$ is positive for all $c \in \mathbb{N}$.

Remark 6. Codes with minimum distance scaling as $n/2 - \Theta(n^{\frac{2}{3}})$ can be constructed in the manner described in Theorem 29 too, by choosing $c \approx m/6$. Specifically, if one chooses $c = m/6 + r$ in Theorem 29, the code would have dimension $k = (m/6 + r + 1/2)m$ and minimum distance $d \geqslant n/2 - 2^{r-1} n^{\frac{2}{3}}$. For the same r, if one chooses $c = r - 1$ in Theorem 30, the code has dimension $k' = \left(\frac{m}{6} + \frac{3(r-1)^2}{2} + \frac{5(r-1)}{2} + \frac{5}{6}\right) m$ and minimum distance $d' \geqslant n/2 - 2^{r-2} n^{\frac{2}{3}} - 2^{2r-3} n^{\frac{1}{3}}$. For all suffciently large n and $r \geqslant 2$, the latter construction provides a better trade-off since $k' > k$ and $n/2 - 2^{r-2} n^{\frac{2}{3}} - 2^{2r-3} n^{\frac{1}{3}} > n/2 - 2^{r-1} n^{\frac{2}{3}}$.

The advantage of the second construction is even more evident when one considers the following cases. Taking $c = m/6 + s\sqrt{m}$ for $s > 0$ in Theorem 29, we have a code C with dimension $k = (m/6 + s\sqrt{m} + 1/2)m$ and minimum distance $d \geqslant n/2 - 2^{s\sqrt{m}-1} n^{\frac{2}{3}}$. Taking $c = s\sqrt{m}$ in Theorem 30, we have a code C' with dimension $k' = \left(\frac{m}{6} + \frac{3s^2 m}{2} + \frac{5s\sqrt{m}}{2} + \frac{5}{6}\right) m \geqslant \left(\frac{1+9s^2}{6} m + \frac{5}{6}\right) m$, and minimum distance $d' \geqslant n/2 - 2^{s\sqrt{m}-1} n^{\frac{2}{3}} - 2^{2s\sqrt{m}-1} n^{\frac{1}{3}}$. For large n, the bounds for the minimum distances d and d' are almost the same, and the dimension k' of the code C' is multiple times larger than k, as $k' \approx (1 + 9s^2)k$.

Remark 7. The constructions used in the proofs of Theorem 29 and Theorem 30 draw on the fact that nonzero elements in F can be generated by a primitive root of unity α. Any nonzero element $\beta \in F$ can thus be expressed as α^p for some $p \in \{0, 1, \dots, 2^m - 1\}$. Using the binary expansion, $p = b_{m-1} 2^{m-1} + b_{m-2} 2^{m-2} + \cdots + b_1 2^1 + b_0$, roots for $M_\beta(x)$ are of the form $\alpha^{p \cdot 2^j} = \alpha^{p'}$ where $p' = b_{m-1} 2^{m-1+j} + b_{m-2+j} 2^{m-2+j} + \cdots + b_1 2^{1+j} + b_0 2^j \equiv b_{m-1-j} 2^{m-1} + b_{m-2-j} 2^{m-2} + \cdots + b_1 2^{1+j} + b_0 2^j + b_{m-1} 2^{j-1} + \cdots + b_{m-j}$ after taking modulo

$2^m - 1$. The length-m binary expressions

$$\begin{aligned} p &= (b_{m-1}b_{m-2}\ldots b_1b_0)_2, \\ p' &= (b_{m-1-j}b_{m-2-j}\ldots b_{m-j})_2 \end{aligned}$$

are cyclic shifts of each other. The key idea behind the proof techniques of Theorem 29 and Theorem 30 is to leverage the BCH bound with a focus on finding length-m binary sequences $(b_{m-1}, b_{m-2}, \ldots, b_1, b_0)$ and two integers t_1, t_2 with $t_1 < t_2 \leqslant 2^m - 1$, such that cyclic shifts of the binary sequences do not correspond to values in the range $\{t_1, t_1 + 1 \ldots, t_2\}$.

Remark 8. According to the discussion in Remark 7, the BCH-like construction technique for large minimum distance codes described in Sections 4.3.1 and 4.3.2 can be tailored towards deriving lower bounds in *narrower* ranges of d. In particular, the construction in Section 4.3.1 admits all m-bit binary sequences with exactly two 1's spacing at least $\ell = m/2 - c$ bits apart (distance is evaluated in a wrap-around manner), and that in Section 4.3.2 all m-bit binary sequences with two or three 1's spacing at least $\ell = m/3 - c$ bits apart. By extending such arguments to consider all m-bit binary sequences with Hamming weight between 2 and $w \in \mathbb{N}$, such that the 1's are at least $\ell = m/w - c$ bits apart, one can construct codes with minimum distance scaling as $d = n/2 - O(n^{\frac{w-1}{w}})$.

4.4 Main Results - Upper Bounds

4.4.1 Harmonic Analysis Approach

We adopt a covering argument similar to Navon and Samorodnitsky [106] and show upper bounds on the size of any code C with length n and minimum distance d scaling as $d \geqslant n/2 - \Omega(\sqrt{n})$. The viewpoint presented in [106], providing an alternative proof to the MRRW bound, is different from the original proof found in [103] which relies on analytical properties of the Krawchouk polynomials, and instead employs Fourier analysis on the group $\mathbb{F}_2^n$ as their main tool.

In particular, the authors of [106] exploit the expediency of working with the maximal eigenfunctions of Hamming balls. One key finding was that, given any real-valued function f on $\{0,1\}^n$ with a small support $B \subset \{0,1\}^n$, such that the adjacency matrix of the Hamming cube acts on f by multiplying it pointwise by a large factor, the cardinality of error-correcting codes with minimum distance d can be upper bounded by $n\,|B|$. The applicability will depend on the value of the multiplying factor. By proposing functions f supported on Hamming balls $B = B_r(\mathbf{0}, n)$ of different radii r, one may derive a lower bound of the multiplying factor, formally called the *maximal eigenvalue of adjacency matrix of the subgraph incduced by B*. This makes possible a simple proof of the first linear programming bound.

Let us now state the definition of the *maximal eigenvalue* of a graph. Let $G = (V, E)$ be a (finite, undirected, simple) graph. Let $A_G = (A_{ij})$ denote the $|V| \times |V|$ adjacency matrix of G, defined by $A_{ij} = 1$ if $(i, j) \in E$ and $A_{ij} = 0$ otherwise for vertices $i, j \in V$. Note that A_G is symmetric, so its eigenvalues are real, and can be ordered as $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$. For any function f on $\mathbb{F}_2^n$, the function Af sums at each point of $\{0,1\}^n$ the values of f at its neighbours. That is, the value taken by the function Af at a vertex $x \in \mathbb{F}_2^n$, denoted by $(Af)(x)$ or $Af(x)$, is given by $Af(x) = \sum_{y \in \mathbb{F}_2^n : w_H(x,y)=1} f(y)$. When B is a subset of the cube $\mathbb{F}_2^n$, set

$$\lambda_B \overset{\text{def}}{=} \max \left\{ \frac{\langle Af, f\rangle}{\langle f, f\rangle} \,\middle|\, f : \mathbb{F}_2^n \to \mathbb{R},\ \operatorname{supp}(f) \subseteq B \right\}, \tag{4.17}$$

where $\langle f, g\rangle \overset{\text{def}}{=} \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} f(x)g(x)$ for real-valued functions f, g on $\mathbb{F}_2^n$. That is, λ_B is the maximal eigenvalue of adjacency matrix of the subgraph of $\{0,1\}^n$ induced by B.

Two lemmas were shown in [106] to show (4.9).

Lemma 31 ([106] Prop 1.1)**.** *Let C be a code with block length n and minimal distance d. Let B be a subset of $\{0,1\}^n$ with $\lambda_B \geqslant n - 2d + 1$. Then $|C| = M \leqslant n\,|B|$.*

Lemma 32 ([106] Lemma 1.4)**.** *Let $B = B_r(\mathbf{0}, n) \subseteq \{0,1\}^n$. The maximal eigenvalue associated with B is $\lambda_B \geqslant 2\sqrt{r(n-r)} - o(n)$.*

To prove (4.9), we note that Lemma 32 implies that a radius $r^* = n/2 - \sqrt{d(n-d)} + o(n)$ exists such that $\lambda_{B_{r^*}} \geqslant n - 2d + 1$. Lemma 31 in turn shows that any code of length n and minimal distance d has at most $n\,|B_{r^*}| = n \cdot \operatorname{Vol}(r^*, n)$ codewords. The cardinality of a Hamming ball of radius r is $\operatorname{Vol}(r, n) = 2^{H_2(r/n)n + o(n)}$. Equation (4.11) follows the above argument, hence yielding equation (4.9).

We note that the above argument can not be used directly to show an upper bound when $d = n/2 - \Theta(\sqrt{n})$. In particular, the $o(n)$ term in Lemma 32 renders the search for a meaningful r^* impossible, as we would ideally require a subset B with λ_B close to $n - 2d + 1 = \Theta(\sqrt{n})$.

4.4.2 Bounds for $d \geqslant n/2 - \sqrt{n}$

We show in this section an approach to lower bound λ_B for the Hamming ball $B = B_3(\mathbf{0}, n)$, which, when coupled with a new proposition stronger than Lemma 31, leads to an upper bound scaling as $A(n, d) = O(n^{3.5})$ for $d \geqslant n/2 - \sqrt{n}$.

First we provide a proposition in place of Lemma 32 that does not require an $o(n)$ term.

Proposition 33. *Let $B = B_3(\mathbf{0}, n) \subseteq \{0,1\}^n$ be the Hamming ball of radius* 3. *The maximal eigenvalue associated with B is $\lambda_B \geqslant \left(\sqrt{3+\sqrt{6}} + o(1)\right)\sqrt{n} \gtrsim 2.334\sqrt{n}$.*

Proof: Recall the definition of the maximal eigenvalue in (4.17). We prove the proposition by constructing a function f with support in B, and for which $\langle Af, f\rangle / \langle f, f\rangle \approx \sqrt{3+\sqrt{6}}\sqrt{n}$. The function f will be symmetric, namely its value at a point will depend only on the Hamming weight of the point. With a slight abuse of notation, such a function is fully defined by its values $f(0), f(1), \ldots, f(n)$ at Hamming weights $0, 1, \ldots, n$.

Set $f(0) = 1$, $f(j) = 0$ for $j \geqslant 4$, and let

$$\lambda f(i) = Af(i) = if(i-1) + (n-i)f(i+1) \tag{4.18}$$

for $i = 0, 1, 2$ (assuming $f(-1) = 0$), where $\lambda = t\sqrt{n}$. We have

$$\begin{aligned} f(1) &= \frac{\lambda f(0)}{n} = \frac{t}{\sqrt{n}}, \quad f(2) = \frac{\lambda f(1) - 1f(0)}{n-1} = \frac{t^2-1}{n-1}, \\ f(3) &= \frac{\lambda f(2) - 2f(1)}{n-2} = \frac{1}{n-2}\left(\frac{t^2-1}{n-1}t\sqrt{n} - 2\frac{t}{\sqrt{n}}\right). \end{aligned}$$

We may use the values $f(i)$ and calculate

$$\begin{aligned} 2^n\langle Af, f\rangle &= 2t\sqrt{n} + t(t^2-1)^2\frac{n\sqrt{n}}{n-1} = \left(2t + t(t^2-1)^2 + o(1)\right)\sqrt{n}, \\ 2^n\langle f, f\rangle &= 1 + t^2 + \frac{1}{2}\frac{n}{n-1}(t^2-1)^2 + \frac{1}{6}\frac{n-1}{n-2}t^2\left[\frac{n}{n-1}(t^2-1) - 2\right]^2 \\ &= 1 + t^2 + (t^2-1)^2/2 + t^2(t^2-3)^2/6 + o(1). \end{aligned}$$

We are now ready to optimize the value

$$\frac{\langle Af, f\rangle}{\langle f, f\rangle} = \left[\frac{2t + t(t^2-1)^2}{(t^6 - 3t^4 + 9t^2 + 9)/6} + o(1)\right]\sqrt{n} \tag{4.19}$$

over $t > 0$. Taking $t = \sqrt{3+\sqrt{6}}$, the square bracket term in (4.19) achieves its maximum $\sqrt{3+\sqrt{6}} + o(1)$. ■

In order to provide a bound as tight as possible, we improve upon Lemma 31 and show the following proposition.

Proposition 34. *Let C be a code with block length n and minimal distance d. Let B be a subset of $\{0,1\}^n$ with $\lambda_B > n - 2d$. Then $|C| = M \leqslant \frac{n}{\lambda_B - (n-2d)}|B|$.*

The proof can be shown using a similar argument as in the proof of Lemma 31 in [106], and is provided in Appendix B.2 for reference.

With Propositions 33 and 34, we are ready to state the upper bound on $A(n, \lceil n/2 - \sqrt{n}\,\rceil)$.

Theorem 35. *If a (n, M, d) binary code C has minimum distance $d \geqslant n/2 - \sqrt{n}$, then*

$$M \leqslant \frac{\sqrt{n}}{\sqrt{3+\sqrt{6}} - 2 + o(1)} \operatorname{Vol}(3, n) = O(n^{3.5}).$$

Proof: Let $B = B_3(\mathbf{0}, n)$ be the radius-3 Hamming ball. The maximal eigenvalue induced by B is $\lambda_B \gtrsim \left(\sqrt{3+\sqrt{6}} + o(1)\right)\sqrt{n}$ according to Proposition 33. Since $n - 2d \leqslant 2\sqrt{n} \lesssim \lambda_B$, the cardinality of C can be upper bounded using Proposition 34 as

$$M \leqslant \frac{n}{\lambda_B - (n-2d)} |B| \leqslant \frac{\sqrt{n}}{\sqrt{3+\sqrt{6}} - 2 + o(1)} \operatorname{Vol}(3, n). \qquad \blacksquare$$

Remark 9. We note that the argument above can upper bound the size as $M = O(n^{3.5})$ as long as $(n-2d)/\sqrt{n}$ is strictly smaller than $\sqrt{3+\sqrt{6}}$. That is, for any $d \gtrsim n/2 - \rho\sqrt{n}$, for some constant $\rho < \sqrt{3+\sqrt{6}}/2 \approx 1.167$, we have $A(n, d) = O(n^{3.5})$.

4.4.3 Bounds for $d \geqslant n/2 - \Theta(\sqrt{n})$

In general, it is possible to generalize the approach in Section 4.4.2 that lower bounds λ_B for $B = B_r(\mathbf{0}, n)$ from $r = 3$ to any given $r \in \mathbb{N}$. Specifically, setting $\lambda = t\sqrt{n}$, we would need to apply the recurrence equation (4.18) iteratively to find $f(i)$ as a function of both t and n, for $i = 1, \ldots, r$, compute inner products $\langle Af, f\rangle$ and $\langle f, f\rangle$, and solve the optimization problem that maximizes the quotient as in (4.19). The procedure could be almost intractable for large (but finite) r.

A more feasible approach is given in this section to lower bound the maximal eigenvalue associated with $B = B_r(\mathbf{0}, n)$. The approach is comprised of four parts. The first part constructs a symmetric function g on $\{0,1\}^n$ based on a recursive relation involving $\lambda = t\sqrt{n}$, and a scaled version of g denoted by $\tilde{g}$, both of which are defined independent of r and have support on the entire domain $\{0,1\}^n$. In the second part, we define a function f which is identical to g on B and 0 elsewhere, evaluate the quotient seen in (4.17), i.e., $\langle Af, f\rangle/\langle f, f\rangle$, and show that it can be expressed concisely as the difference of λ and another term involving $\tilde{g}(r)$ and $\tilde{g}(r+1)$, where $\tilde{g}(i)$ depends on i, n and t. The quotient can thus be lower bounded by λ which guarantees that either $\tilde{g}(r)$ or $\tilde{g}(r+1)$ is 0. (That is, for a given r, one may choose t and n appropriately so that $\tilde{g}(r) = 0$, or $\tilde{g}(r+1) = 0$, and $t\sqrt{n}$ would be a lower bound of λ_B.) In the third part, we introduce two other functions h and $\tilde{h}$ which, broadly speaking, act as the respective proxies of g and $\tilde{g}$. In particular, as n grows large, the maximal root of $\tilde{g}(k)$ (when viewed as a function of t) converges to that of $\tilde{h}(k)$, denoted by $t_h(k)$, a value independent of n. The fourth part concludes the argument by showing that the quantity $\lambda_B/\sqrt{n}$ is lower bounded by the maximal root of $\tilde{h}(r+1)$, when viewed as a function of t, for sufficiently large n. Finally we leverage Proposition 34 to show $A(n,d) = O(n^{r+1})$ as long as $(n-2d)/\sqrt{n} < t_h(r+1) - s$ for some $s > 0$. Throughout this section, we assume $\lambda = t\sqrt{n}$ for some constant $t > 0$.

Part 1

We first consider a symmetric function $g : \mathbb{F}_2^n \to \mathbb{R}$, and with a slight abuse of notation, write $g(\mathbf{x}) = g(w_H(\mathbf{x}))$ for $\mathbf{x} \in \mathbb{F}_2^n$. Define g by the initial condition $g(0) = 1$, and the recurrence relations

$$\lambda g(i) = Ag(i) = ig(i-1) + (n-i)g(i+1) \tag{4.20}$$

for $i = 0, 1, 2, \dots, n-1$, assuming $g(-1) = 0$. For example, we have

$$g(1) = \frac{\lambda}{n}, \quad g(2) = \frac{1}{n-1}\left(\frac{\lambda^2}{n} - 1\right), \quad g(3) = \frac{1}{n-2}\left(\frac{\lambda^3}{n(n-1)} - \frac{2\lambda}{n} - \frac{\lambda}{n-1}\right).$$

Define a real-valued function $\tilde{g}$ by $\tilde{g}(i) = n^{i/2}g(i)$ for $i = 0, 1, \dots, n$. The values $\tilde{g}(i)$ for $i = 0, 1, 2, 3$ are $\tilde{g}(0) = 1$, $\tilde{g}(1) = t$, $\tilde{g}(2) = \frac{n}{n-1}(t^2 - 1)$, $\tilde{g}(3) = \frac{n}{n-2}\left(\frac{t^3 n}{n-1} - 2t - \frac{tn}{n-1}\right)$.

Remark 10. For each $i = 0, 1, 2, \dots$, both $g(i)$ and $\tilde{g}(i)$ are functions of t and n. For a fixed t, $\tilde{g}(i)$ scales with n as $O(1)$, and that $g(i) = O(n^{-i/2})$.

Remark 11. For any finite $k \in \mathbb{N}$, it can be shown that $\tilde{g}(k+1) = 0$ only when $\tilde{g}(k) \neq 0$ and $\tilde{g}(k-1) \neq 0$, and that $\tilde{g}(k)$ is a degree-k polynomial in t with leading coefficient $1 + O(n^{-1})$. Specifically, for an even $k \in \mathbb{N}$,

$$\tilde{g}(k) = t^k(1 + O(n^{-1})) - g_{k,k-2}t^{k-2}(1 + O(n^{-1})) + \ldots + (-1)^{k/2}g_{k,0}(1 + O(n^{-1}))\ ;$$

for an odd $k \in \mathbb{N}$,

$$\tilde{g}(k) = t^k(1 + O(n^{-1})) - g_{k,k-2}t^{k-2}(1 + O(n^{-1})) + \ldots + (-1)^{(k-1)/2}g_{k,1}(1 + O(n^{-1})),$$

where the coefficients $g_{k,k-2\ell}, 1 \leqslant \ell \leqslant \left\lfloor \frac{k}{2} \right\rfloor$ are positive integers independent of n and t.

Part 2

Let $B = B_r(\mathbf{0}, n)$ for some finite r. Consider a symmetric function f supported on B, defined by $f(i) = g(i)$ for all $i = 0, 1, \dots, r$, and $f(i) = 0$ for all $i > r$. First, we have

$$2^n\langle f, f\rangle = f(0)^2 + \binom{n}{1} f(1)^2 + \ldots + \binom{n}{r} f(r)^2 = \sum_{i=0}^{r} \binom{n}{i} f(i)^2 = \sum_{i=0}^{r} \binom{n}{i} g(i)^2$$

Using the observation $g(k) = O(n^{-k/2})$ and that $f(0)^2 = 1$, the sum scales as $2^n\langle f, f\rangle = \Theta(1)$. Note that $Af(i) = Ag(i) = \lambda g(i)$ for $i = 0, 1, \dots, r-1$ and $Af(r) = rf(r-1) + (n-r)f(r+1) = rg(r-1)$. Hence,

$$\begin{aligned}
&2^n\langle Af, f\rangle \\
&= Af(0)f(0) + \binom{n}{1} Af(1)f(1) + \ldots + \binom{n}{r} Af(r)f(r) \\
&= \lambda g(0)^2 + \binom{n}{1}\lambda g(1)^2 + \ldots + \binom{n}{r-1}\lambda g(r-1)^2 + \binom{n}{r} rg(r-1)g(r) \\
&= \lambda 2^n\langle f, f\rangle + \binom{n}{r}\left(rg(r-1)g(r) - \lambda g(r)^2\right) \\
&= \lambda 2^n\langle f, f\rangle - \binom{n}{r} g(r)(n-r)g(r+1) \\
&= \lambda 2^n\langle f, f\rangle - \left(n^{-r}\binom{n}{r}\right)\left(1 - \frac{r}{n}\right)\left(n^{r/2}g(r)\right)\left(n^{\frac{r+1}{2}}g(r+1)\right)\sqrt{n} \\
&= \lambda 2^n\langle f, f\rangle - \left(n^{-r}\binom{n}{r}\right)\left(1 - \frac{r}{n}\right)\tilde{g}(r)\tilde{g}(r+1)\sqrt{n},
\end{aligned}$$

where the fourth equality holds by evaluating the recursion relation (4.20) with $i = r$.

The ratio between $\langle Af, f\rangle$ and $\langle f, f\rangle$ is thus

$$\frac{\langle Af, f\rangle}{\langle f, f\rangle} = \lambda - \frac{1}{2^n\langle f, f\rangle}\left(n^{-r}\binom{n}{r}\right)\left(1 - \frac{r}{n}\right)\tilde{g}(r)\tilde{g}(r+1)\sqrt{n}, \tag{4.21}$$

which scales as $\Theta(\sqrt{n})$ since $2^n\langle f, f\rangle = \Theta(1)$. Denote by $t_g(i) = t_g(i, n)$ the maximal root of $\tilde{g}(i) = \tilde{g}(i, n, \lambda = t\sqrt{n})$ when viewed as a function of t, for $i = 1, 2, \ldots$. For example, $t_g(1) = 0$ since $\tilde{g}(1) = t$, $t_g(2) = 1$ since $\tilde{g}(2) = \frac{n}{n-1}(t^2 - 1)$, and $t_g(3)$ is the maximal root of the polynomial $t^3 n - 3tn + 2t = 0$. Then $\lambda_B \geqslant \max\left(t_g(r)\sqrt{n}, t_g(r+1)\sqrt{n}\right)$ because the second term in (4.21) is 0 when λ is either $t_g(r)\sqrt{n}$ or $t_g(r+1)\sqrt{n}$.

The first two steps successfully simplify the problem for finding a lower bound on λ_B to the following. First solve $g(r)$ and $g(r+1)$ by recursively applying (4.20), and then find the maximal roots of $\tilde{g}(r) = 2^{r/2}g(r)$ and $\tilde{g}(r+1) = 2^{(r+1)/2}g(r+1)$, which are viewed as functions of t. The recursive steps, however, still pose a great challenge when the radius r for $B = B_r(\mathbf{0}, n)$ is large. For example, $g(3) = \frac{1}{n-2}\left[\frac{t\sqrt{n}}{n-1}(t^2-1) - \frac{2t}{\sqrt{n}}\right]$ and $g(4) = \frac{1}{n-3}\left(\frac{t\sqrt{n}}{n-2}\left[\frac{t\sqrt{n}}{n-1}(t^2-1) - \frac{2t}{\sqrt{n}}\right] - \frac{3}{n-1}(t^2-1)\right)$. Solving roots of $\tilde{g}(r)$ and $\tilde{g}(r+1)$, which in general are complicated functions in both n and t, is an even greater challenge. The following third step shows that $t_g(i)$ converges to the maximal root of a simpler polynomial in t for all finite i when $n \to \infty$.

Part 3

Define a function $h : \{0, 1, \ldots, n\} \to \mathbb{R}$ by setting $h(0) = 1$ and the recurrence relation (assuming $h(-1) = 0$)

$$\lambda h(i) = ih(i-1) + nh(i+1) \text{ for } i = 0, 1, 2, \ldots, n-1, \tag{4.22}$$

which differs from (4.20) only in the coefficient of $h(i+1)$. It can be shown that $h(k+1) = 0$ only when $h(k) \neq 0$ and $h(k-1) \neq 0$, and that $h(k)$ is either 0 or scales as $\Theta(n^{-k/2})$ for each finite k. The functions h and g coincide asymptotically for all finite k. We state precisely a bound on the ratio between the two in the following lemma.

Lemma 36. *For any given $k \in \mathbb{N} \cup \{0\}$, if $h(i) \neq 0$ for all $i \leqslant k-1$, then*

$$g(k) = \begin{cases} h(k)(1 + O(n^{-1})) & \text{if } h(k) \neq 0, \\ O(n^{-(k+2)/2}) & \text{if } h(k) = 0. \end{cases}$$

Proof: First note that $g(0) = h(0) = 1$, $g(1) = h(1) = \frac{\lambda}{n}$ and $g(2) = \frac{1}{n-1}\left(\frac{\lambda^2}{n} - 1\right) = \frac{1}{n}(1 + \frac{1}{n-1})\left(\frac{\lambda^2}{n} - 1\right) = (1 + O(n^{-1}))\frac{1}{n}(\lambda h(1) - 1h(0)) = h(2)(1 + O(n^{-1}))$. Also, $h(2) = 0$ if and only if $g(2) = 0$. Hence the lemma holds for $k = 0, 1, 2$. Assume, for some $k \geqslant 2$, $h(i) \neq 0$ and $g(i) = h(i)(1 + O(n^{-1}))$ for $i = 0, 1, \ldots, k$. Then $g(i)$ and $h(i)$ both scale as $\Theta(n^{-i/2})$ for $i = 0, 1, \ldots, k$. Equation (4.22) yields

$$h(k+1) = n^{-1}(\lambda h(k) - kh(k-1))$$

and (4.20) yields

$$\begin{aligned} g(k+1) &= (n-k)^{-1}(\lambda g(k) - kg(k-1)) \\ &= \left(1 + O(n^{-1})\right) n^{-1}[\lambda h(k)(1 + O(n^{-1})) - kh(k-1)\left(1 + O(n^{-1})\right))], \end{aligned}$$

which implies $g(k+1) = h(k+1)(1+O(n^{-1}))$ when $h(k+1) \neq 0$ since both $\lambda h(k)$ and $kh(k-1)$ scale as $\Theta(n^{-(k-1)/2})$. When $h(k+1) = 0$, $g(k+1) = O(n^{-1}n^{-(k-1)/2}n^{-1}) = O(n^{-(k+3)/2})$. Hence the lemma holds by the principle of mathematical induction. ■

Consider a real-valued function $\tilde{h} : \{0, 1, \ldots, n\} \to \mathbb{R}$ defined by $\tilde{h}(i) = n^{i/2}h(i)$. The values $\tilde{h}(i)$ for $i = 0, 1, 2, 3, 4$ are $\tilde{h}(0) = 1, \tilde{h}(1) = t, \tilde{h}(2) = t^2 - 1, \tilde{h}(3) = t^3 - 3t, \tilde{h}(4) = t^4 - 6t^2 + 3$.

Remark 12. The function $\tilde{h}$ satisfies the recurrence relation

$$t\tilde{h}(i) = i\tilde{h}(i-1) + \tilde{h}(i+1), \tag{4.23}$$

which follows from the recurrence relation (4.22) and that $\tilde{h}(i) = n^{i/2}h(i)$.

Remark 13. For k a finite positive integer, $h(k)$ depends on both n and t, whereas $\tilde{h}(k)$ is independent of n and is a degree-k monic polynomial of t. The polynomial $\tilde{h}(k)$ can be expressed as follows:

$$\tilde{h}(k) = \begin{cases} t^k - g_{k,k-2}t^{k-2} + \ldots + (-1)^{k/2}g_{k,0} & \text{for even } k, \\ t^k - g_{k,k-2}t^{k-2} + \ldots + (-1)^{(k-1)/2}g_{k,1} & \text{for odd } k, \end{cases}$$

where the coefficients $g_{k,k-2\ell}, 1 \leqslant \ell \leqslant \left\lfloor \frac{k}{2} \right\rfloor$ are the same as those in Remark 11.

Denote by $t_h(i)$ the maximal root of $\tilde{h}(i)$ when viewed as a function of t, for $i \geqslant 1$. For example, $t_h(1) = 0$, $t_h(2) = 1$, and $t_h(3) = \sqrt{3}$. We now show that, for all finite k, the maximal roots $t_h(i)$ and $t_g(i)$ are equal when $n \to \infty$.

Lemma 37. *Let $k \in \mathbb{N}$ be finite. Then $\lim_{n\to\infty} t_g(k,n) = t_h(k)$.*

Proof: First we prove that for any finite $k \in \mathbb{N}$, $(t_h(1), t_h(2), \ldots, t_h(k))$ is a strictly increasing sequence, using the principle of mathematical induction. For $k = 3$, the sequence is $(0, 1, \sqrt{3})$, which verifies the claim. Assume for a finite k, the sequence $(t_h(1), t_h(2), \ldots, t_h(k))$ is strictly increasing. Note that $\tilde{h}(k+1) = t\tilde{h}(k) - k\tilde{h}(k-1)$ is negative when $t = t_h(k)$ since $\tilde{h}(k) = 0$ and $\tilde{h}(k-1) > 0$. Since the leading term in $\tilde{h}(k+1)$ is t^{k+1} and thus grows to positive infinity for t large enough, we must have $t_h(k+1) > t_h(k)$.

We now show that, the sequence of roots $t_g(k,n)$ of $\tilde{g}(k)$, for $n = 1, 2, \ldots$, converges to $t_h(k)$. Let $\delta \in (0, t_h(k) - t_h(k-1))$ be a small constant such that $\tilde{h}(k,t) > 0$ for all $t \in (t_h(k), t_h(k) + \delta]$, and $\tilde{h}(k,t) < 0$ for all $t \in [t_h(k) - \delta, t_h(k))$. Let $\epsilon > 0$ be the minimum $\epsilon = \min\left\{-\tilde{h}(k, t_h(k) - \delta), \tilde{h}(k, t_h(k) + \delta)\right\}$. Leveraging Lemma 36, $\tilde{g}(k, t_h(k) + \delta) \geqslant \epsilon(1 + O(n^{-1}))$ and $\tilde{g}(k, t_h(k) - \delta) \leqslant -\epsilon(1 + O(n^{-1}))$. This implies the existence of a root of $\tilde{g}(k)$

when viewed as a function of t in the interval $(t_h(k)-\delta, t_h(k)+\delta)$, for all sufficiently large n. Note also that, for $t > t_h(k)+\delta$, $\tilde{g}(k,t)$ is strictly positive for all sufficiently large n, because $\tilde{h}(k,t) \geqslant \epsilon$. Since $\delta > 0$ can be chosen arbitrarily small, the root $t_g(k,n)$ converges to $t_h(k)$ as n grows to infinity. ■

Remark 14. It is known that the roots of a polynomial (counting multiplicities and only up to permutation) depend continuously on the coefficients of the polynomial [61, 64]. Hence, Theorem 36 also follows as a direct consequence of Remarks 11 and 13.

Part 4

We are now ready to state the main result in this section, which admits a practical approach to lower bound λ_B for fixed r and sufficiently large n.

Theorem 38. *Let $B = B_r(\mathbf{0}, n)$ and λ_B be the maximal eigenvalue of adjacency matrix of the subgraph of $\{0,1\}^n$ induced by B. Then λ_B is lower bounded by $(t_h(r+1)+o(1))\sqrt{n}$.*

Proof: (4.21) guarantees that $\lambda_B \geqslant \max\{t_g(r,n)\sqrt{n},\ t_g(r+1,n)\sqrt{n}\}$. Since $t_h(k)$ is strictly increasing in k, Lemma 37 implies that for large n, the bound reduces to $\lambda_B \geqslant (t_h(r+1)+o(1))\sqrt{n}$. ■

Theorem 38 generalizes Proposition 33 by establishing lower bounds on λ_{B_r} for fixed r other than the special case $r = 3$. This in turn yields a sequence of bounds on $|C|$ whose applicability depends on the scaling behaviour of the minimum distance in terms of the blocklength n.

Corollary 39. *If a (n, M, d) binary code C has distance $d > \frac{1}{2}\left[n - (t_h(r+1)-s)\sqrt{n}\,\right]$ for some $r \in \mathbb{N}$ and $s > 0$, and n is sufficiently large, then*

$$M \leqslant \frac{\sqrt{n}}{s+o(1)} \operatorname{Vol}(r,n) = O(n^{r+\frac{1}{2}}).$$

Proof: Let $B = B_r(\mathbf{0}, n)$. We have $n-2d < (t_h(r+1)-s)\sqrt{n} \lesssim (t_h(r+1)+o(1))\sqrt{n} \leqslant \lambda_B$ due to Theorem 38. The size of the code C can thus be bounded using Proposition 34 as

$$M \leqslant \frac{n}{\lambda_B - (n-2d)}\,|B| \leqslant \frac{\sqrt{n}}{(t_h(r+1)+o(1)) - (t_h(r+1)-s)}\,|B| = \frac{\sqrt{n}}{s+o(1)} \operatorname{Vol}(r,n). \quad ■$$

4.4.4 Bounds for $d \geqslant n/2 - \Theta(\sqrt{n})$ - Numerical Results

Using similar steps as in the proof of Proposition 33, one may show lower bounds of the maximal eigenvalues associated with Hamming balls of different radii, which could be a

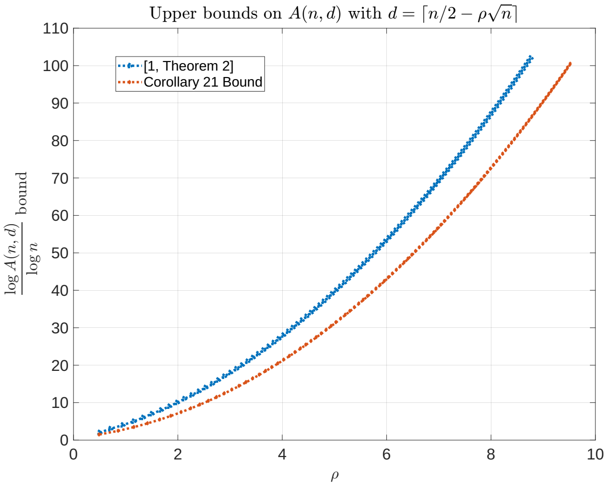


Figure 4.1: Upper bounds on $\frac{\log A(n,d)}{\log n}$ for $d = \lceil n/2 - \rho\sqrt{n}\,\rceil$

daunting procedure even for a radius as small as 5. Alternatively, results from Section 4.4.3 suggest a more feasible approach. For $k = 1, 2, \ldots$, we first find $\tilde{h}(k)$ by solving the recursive relations 4.23 with the initial conditions $\tilde{h}(0) = 1, \tilde{h}(-1) = 0$, and solve the maximal roots $t_h(k)$ of the polynomials $\tilde{h}(k)$. Theorem 38 then yields a bound $\lambda_{B_k} \gtrsim t_h(k+1)\sqrt{n}$. For example, we have $\tilde{h}(1) = t$, $\tilde{h}(2) = t^2 - 1$, $\tilde{h}(3) = t^3 - 3t$, $\tilde{h}(4) = t^4 - 6t^2 + 3$, $\tilde{h}(5) = t^5 - 10t^3 + 15t$. The corresponding maximal roots are $t_h(1) = 0$, $t_h(2) = 1$, $t_h(3) = \sqrt{3}$, $t_h(4) = \sqrt{3 + \sqrt{6}} \approx 2.334$, $t_h(5) = \sqrt{5 + \sqrt{10}} \approx 2.857$. Applying Theorem 38 to $r = 3$ shows that $\lambda_{B_3} \gtrsim \sqrt{3 + \sqrt{6}}\sqrt{n}$, which recovers Proposition 33 in Section 4.4.2. With computer program assistance, we are able to solve maximal roots $t_h(k)$ for k as large as 101.

Consider now a (n, M, d) binary code C with minimum distance $d = \lceil n/2 - \rho\sqrt{n}\,\rceil$, i.e., $j = 2\rho\sqrt{n}$. Corollary 39 entails the following: If $\rho < t_h(r+1)/2$ for some $r \in \mathbb{N}$, then $M = O(n^{r+0.5})$. For example, with $\rho = 1$, the smallest integer r for the inequality to hold is $r = 3$ since $1 < t_h(4)/2 \approx 1.167$, and thus $M = O(n^{3.5})$. We plot in Figure 4.1 the the exponent $r+0.5$ in the asymptotic bound for $A(n,d)$ for $0.5 = t_h(2)/2 < \rho < t_h(101)/2 \approx 9.5$, based on values of $t_h(k)$ for $k = 2, 3, \ldots, 101$. For example, values of $t_h(3)$ and $t_h(5)$ lead to the bounds $A(n, n/2 - \lceil \rho_1\sqrt{n}\,\rceil) = O(n^{2.5})$, $A(n, \lceil n/2 - \rho_2\sqrt{n}\,\rceil) = O(n^{4.5})$, for all $\rho_1 < \sqrt{3}/2 \approx 0.866$ and $\rho_2 < 1.428$. One another case, when $r = 7$, the point $(2.072, 7.5)$ guarantees that a code with minimum distance $d \geqslant n/2 - 2\sqrt{n}$ must have $M \lesssim \frac{\sqrt{n}}{4.14-4}\,\mathrm{Vol}(7, n) = O(n^{7.5})$.

As discussed in Section 4.2.3, the spectral-based bounds for $A(n,d)$ [20] in the large minimum distance regime can be stated as follows:

$$A(n,d) = O(n^k) \text{ as long as } d \geqslant n/2 - \underline{\lambda}_{k-1}\sqrt{n}/2,$$

where $\underline{\lambda}_k$ is defined in equation (4.13). We compute numerically $\underline{\lambda}_k$ for $k = 1, 2, \ldots, 100$ and show the associated bounds in Figure 4.1. We can see that the two families of bounds scale in a similar fashion, while our newly derived upper bounds, due to Corollary 39, are slightly tighter.

Remark 15. When the function f is constrained to be symmetric and the support of f constrained to be the Hamming ball $B_r(\mathbf{0}, n)$, the value λ_B (see (4.17)) reduces to the maximal eigenvalues λ_r in [20]. However, due to the distinct proof techniques, our upper bounds, based on harmonic analysis on the Hamming space, only require knowledge of λ_r, while bounds in [20] require that of both λ_r and λ_{r-1}.

4.5 Conclusion and Outlook

In this chapter, we study bounds on the cardinality of codes with specified minimum distance d targeting the regime with $d = n/2 - \Omega(\sqrt{n})$. The codes in this regime have vanishing rate, which renders known bounds that dictate the tradeoff between the code rate and relative distance ineffective. We obtain two families of codes based on specifically crafted BCH-like constructions, and a sequence of upper bounds for $d \geqslant n/2 - \rho\sqrt{n}$ for $\rho \in (0.5, 9.5)$.

The proposed cyclic code constructions are targeted at the regimes $d \geqslant n/2 - \Omega(\sqrt{n})$ and $d \geqslant n/2 - \Omega(n^{2/3})$, and have sizes that are polynomial and quasi-polynomial in n, respectively. The proof of the upper bound makes extensive use of Fourier analysis on the Hamming cube as a group, and a new bounding technique for the maximal eigenvalue associated with Hamming balls of finite radii.

An interesting problem for future work is to study the potential of the Fourier-analytical approach to upper bound the sizes of constant weight codes with large minimum distance. This problem has been studied in the regime $d = \delta n$ with $\delta \in (0, 0.5)$ and the best known result is the second linear programming bound. Another interesting problem, as pointed out by authors of [20], is the underlying similarities between the spectral-based and the Fourier-analytical approaches. For example, if one considers only the Hamming balls $B = B_r(\mathbf{0}, n)$ among all subsets of $\{0, 1\}^n$, and requires functions f to be symmetric, the maximal eigenvalue λ_{B_r} appears to be equivalent to the value λ_r in [20]. This implies that our bounds on λ_{B_r} improves on the bounds on λ_r for finite r, and that for r scaling sub-linearly in n, the latter may yield new bounds on $A(n, d)$ in the regime $d = n/2 - \Theta(n^s)$ for $s \in (0.5, 1)$.

CHAPTER 5

Abelian Group Codes for Classical and C-Q Channel Coding: one-shot rate bounds

5.1 Introduction

Information Theory is the mathematical theory of information–processing tasks such as storage and transmission of information. An information source produces some outputs (or signals) more frequently than others. Due to this redundancy, one can reduce the amount of space needed for its storage by identifying and using structures that exist in the data while guaranteeing a certain degree of recovery of its content. Such data compression is achieved by a suitable *encoding* of the output of the source, hence it is usually known as the *source coding* problem. In the transmission of information through a channel, it is often advantageous to add redundancy to a message, in order to combat the effects of noise. Specifically, a *channel* models a physical device that takes an input and generates an output. Reliable and efficient transmission is typically realized with *coding*, where an *encoder* incorporates some degree of redundancy into the source information so as to form a codeword and where a *decoder* reconstructs the data from the noisy channel output. The *channel coding* problem investigates the minimal amount of redundancy which needs to be added, as well as efficient implementation of the encoding and decoding.

Classical Information Theory studies the storage and transmission of information when the signals, including the source output, channel input, and channel output, are all classical. Quantum Information Theory studies how such tasks can be accomplished using quantum mechanical systems. It deals with how the quantum–mechanical properties of physical systems can be exploited to achieve efficient storage and transmission of information. The underlying quantum mechanics leads to important differences between Quantum and Classical Information theory.

We study in this chapter both classical channel coding and *classical-quantum channel coding*, to be specified later. In both problems, the data to be transmitted reliably are classical, but the channel output of the former must have a classical alphabet. Classical-quantum channel coding has been studied extensively in a scenario where the channel can be used arbitrarily many times. The channel coding theorem for stationary memoryless classical-quantum channels, established by Holevo [66] and Schumacher and Westmoreland [125], provides an explicit formula for the rate at which data can be transmitted under the assumption that each use of the channel is independent of the previous uses. More general channel coding theorems that do not rely on this independence assumption have been developed in later work by Hayashi and Nagaoka [62] and by Kretschmann and Werner [80]. These results are asymptotic, i.e., they refer to a limit where the number of channel uses tends to infinity while the probability of error is required to approach zero.

5.1.1 Abelian Group Codes

Approaching information theoretic performance limits of communication using structured codes has been of great interest for the last several decades [8, 9, 54, 36, 35]. The earlier attempts dictate the use of finite fields in the coding schemes. In the channel coding problem [94], the channel input alphabets are replaced with finite fields and encoders are replaced with matrices. Later these coding approaches were extended to weaker algebraic structures such as rings and groups [45, 28, 88, 89, 52]. The motivation for this are two fold: a) finite fields exist only for alphabets with a prime power size, and b) for communication under certain constraints, codes with weaker algebraic structures have better properties. For example, when communicating over an additive white Gaussian noise channel with 8-PSK constellation, codes over $\mathbb{Z}_8$, the cyclic group of size 8, are more desirable over binary linear codes because the structure of the code is matched to the structure of the signal set [88], and hence the former have superior error correcting properties. As another example, construction of polar codes over alphabets of size p^r, for $r > 1$ and p prime, is simpler with a module structure rather than a vector space structure [120, 123, 114].

Group codes were first studied by Slepian [129] for the Gaussian channel. In [7], the capacity of group codes for certain classes of channels has been computed. Further results on the capacity of group codes were established in [8, 9]. The capacity of group codes over a class of channels exhibiting symmetries with respect to the action of a finite Abelian group has been investigated in [28]. In [121], the asymptotic performance of Abelian group codes for the lossy source coding problem for arbitrary discrete memoryless sources, as well as the channel coding problem for arbitrary discrete memoryless channels is studied. Specifically, for the channel coding problem, the *group capacity* is characterized in a single-letter information-theoretic form.

5.1.2 One-shot Approach

All the aforementioned capacities are evaluated under the asymptotic assumption, and many assume additionally that the channels are memoryless and stationary. However, in many real-world scenarios, we encounter channels which are neither stationary nor memoryless. Therefore, it is of fundamental importance to think of coding schemes for the channels which fail to satisfy these assumptions. The independent channel uses are relaxed in [60, 62] and general channels with memory are studied in [31, 37], albeit these results are derived in the form of a limit such that the error probability vanishes as the number of channel uses goes to infinity. Later researchers considered single-serving scenarios where a given channel is used only once. This approach gives rise to a high level of generality that no assumptions are made on the structure of the channel and the associated capacity is usually referred to as *one-shot* capacity. The one-shot capacity of a classical channel was characterized in terms of min- and max-entropies in [118]. The one-shot classical capacity of a quantum channel is addressed by a hypothesis testing approach in [105] and [142], yielding expressions in terms of the generalized (Rényi) relative entropies and a smooth relative entropy quantity, respectively.

In this work, we consider a scenario where a given classical or classical-quantum channel is used only once and derive tight bounds on the number of classical bits that can be transmitted with a given average error probability ϵ. This one-shot approach provides a high level of generality, as nothing needs to be assumed about the structure of the channel (Note that any situation in which a channel is used repeatedly can be equivalently described as one single use of a larger channel.) Our derivation is based on the idea of relating the problem of channel coding to hypothesis testing. Here, we use a relative-entropy-type quantity defined in [142] known as hypothesis testing relative entropy, denoted $D_{\mathrm{H}}^{\epsilon}(\cdot\|\cdot)$.

The rest of this chapter is organized as follows: In Section 5.2, we briefly introduce some definitions for the channel models, achievability conditions, and the hypothesis testing relative entropy, and basic facts about algebraic groups. In Section 5.3, we introduce the ensemble of Abelian group codes and a few useful results from [121]. In Section 5.4, achievability and converse bounds are established for the one-shot classical coding problem in terms of $D_{\mathrm{H}}^{\epsilon}(\cdot\|\cdot)$ Similar results for the one-shot classical-quantum coding problem are shown in Section 5.5.

5.2 Preliminaries

Classical Channel Model

We consider discrete memoryless channels used without feedback. We associate two finite sets $\mathcal{X}$ and $\mathcal{Y}$ with the channel as the channel input and output alphabets. The input-output relation of the channel is characterized by a conditional probability law $W_{Y|X}(y|x)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The channel is specified by $(\mathcal{X}, \mathcal{Y}, W_{Y|X})$.

Classical-Quantum Channel Model

We also study the case of *classical-quantum channel coding*, where the data to be transmitted reliably are *classical.* No assumptions are made about the channel that is used to achieve this task, i.e., the inputs and outputs may be arbitrary quantum states. However, since the quantum-mechanical structure of the input space is irrelevant for the encoding of classical data, it can be represented by a (classical) set $\mathcal{X}$. For any input $x \in \mathcal{X}$, the channel produces an output, specified by a density operator ρ_x on a Hilbert space $\mathbb{B}$. For our purposes, it is therefore sufficient to characterize a channel by a mapping $\mathcal{N} : x \mapsto \rho_x$ from a set $\mathcal{X}$ to a set of density operators.

Groups

All groups referred to in this paper are *Abelian groups.* Given a group $(G, +)$, a subset H of G is called a *subgroup* of G if it is closed under the group operation. In this case, $(H, +)$ is a group in its own right. This is denoted by $H \leqslant G$. A *coset* C of a subgroup H is a shift of H by an arbitrary element $a \in G$ (i.e. $C = a + H$ for some $a \in G$). For a subgroup H of G, the number of cosets of H in G is called the *index* of H in G and is denoted by $|G : H|$. The index of H in G is equal to $|G|/|H|$ where $|G|$ and $|H|$ are the cardinality or size of G and H respectively. For a prime p dividing the cardinality of G, the *Sylow-p* subgroup of G is the largest subgroup of G whose cardinality is a power of p. Group isomorphism is denoted by $\cong$.

Group Codes

Given a group G, a group code $\mathbb{C}$ over G with block length n is any subgroup of G^n. A shifted group code over G, $\mathbb{C} + V$ is a translation of a group code $\mathbb{C}$ by a fixed vector $V \in G^n$. Group codes generalize the notion of linear codes over fields to sources with reconstruction alphabets (and channels with input alphabets) having composite sizes.

Achievability for Classical Channel Coding

For a group G, a group transmission system with parameters (n, Θ, τ) for reliable communication over a given channel $(\mathcal{X} = G, \mathcal{Y}, W_{Y|X})$ consists of a codebook, an encoding mapping and a decoding mapping. The codebook $\mathbb{C}$ is a shifted subgroup of G^n whose size is equal to Θ and the mappings are defined as

$$\begin{aligned} \text{Enc} &: \{1, 2, \cdots, \Theta\} \to \mathbb{C} \\ \text{Dec} &: \mathcal{Y}^n \to \{1, 2, \cdots, \Theta\} \end{aligned}$$

such that

$$\sum_{m=1}^{\Theta} \frac{1}{\Theta} \sum_{x \in \mathcal{X}^n} \mathbb{1}_{\{x=\text{Enc}(m)\}} \sum_{y \in \mathcal{Y}^n} \mathbb{1}_{\{m \neq \text{Dec}(y)\}} W^n(y|x) \leqslant \tau$$

We say an (M, τ) coding scheme exists for a channel $(\mathcal{X} = G, \mathcal{Y}, W_{Y|X})$ if a group transmission system with parameters $(1, M, \tau)$ does. Given a channel $(\mathcal{X} = G, \mathcal{Y}, W_{Y|X})$, the rate R is said to be achievable using group codes if for all $\epsilon > 0$ and for all sufficiently large n, there exists a group transmission system for reliable communication with parameters (n, Θ, τ) such that

$$\frac{1}{n} \log \Theta \geqslant R - \epsilon, \qquad \tau \leqslant \epsilon$$

The group capacity of the channel C is defined as the supremum of the set of all achievable rates using group codes.

Achievability for Classical-Quantum Channel Coding

Given a classical-quantum channel $\mathcal{N} = \{\rho_x\}_{x \in \mathcal{X}}$ from the classical alphabet $\mathcal{X}$ to the quantum system $\mathbb{B}$, where $\mathcal{X} = G$ is an Abelian group, a group transmission system with parameters (n, Θ, τ) for reliable communication over $\mathcal{N}$ consists of a codebook, an encoding mapping and a decoding positive operator-valued measure (POVM). The codebook $\mathbb{C}$ is a shifted subgroup of G^n whose size is equal to Θ. The encoding mapping is defined as

$$\text{Enc} : \{1, 2, \cdots, \Theta\} \to \mathbb{C}.$$

The decoding POVM is a set $\{\Lambda_m\}_{m=1}^{\Theta}$ of operators such that $\Lambda_m \geqslant 0, \forall m$ and $\sum_m \Lambda_m = I$. The probability of obtaining outcome j is $\text{tr}(\Lambda_j \rho)$ if the state is in a mixed state described by some density operator ρ. The group transmission system with parameters (n, Θ, τ) over

$\mathcal{N}$ exists if

$$\sum_{m=1}^{\Theta} \frac{1}{\Theta} \sum_{x \in \mathcal{X}^n} \mathbb{1}_{\{x=\mathrm{Enc}(m)\}}[1 - \mathrm{tr}(\Lambda_m \rho_x)] \leqslant \tau$$

We say an (M, τ) coding scheme exists for $\mathcal{N}$ if a group transmission system with parameters $(1, M, \tau)$ does. Given a channel $\mathcal{N}$, the rate R is said to be achievable using group codes if for all $\epsilon > 0$ and for all sufficiently large n, there exists a group transmission system for reliable communication with parameters (n, Θ, τ) such that

$$\frac{1}{n} \log \Theta \geqslant R - \epsilon, \qquad \tau \leqslant \epsilon$$

The group capacity of the channel $C = C(\mathcal{N})$ is defined as the supremum of the set of all achievable rates using group codes.

Hypothesis Testing Relative Entropy

The hypothesis testing relative entropy, denoted as $D_H^{\epsilon}(\rho||\sigma)$, is an entropy measure derived from the error probabilities arising from a quantum measurement that attempts to distinguish between the states ρ and σ (a quantum hypothesis test). The most general measurement that one could use in such a test is a two-outcome POVM $\{Q, I - Q\}$ where $0 \leqslant Q \leqslant I$. The outcome Q corresponds to deciding that the state is ρ and the outcome $I - Q$ corresponds to deciding that the state is σ. Thus, the probability of guessing correctly when the state is ρ is equal to $\mathrm{tr}\{Q\rho\}$, and the probability of guessing incorrectly when the state is σ is equal to $\mathrm{tr}\{Q\sigma\}$. In an asymmetric quantum hypothesis test, we try to find a POVM that guesses ρ correctly with high probability, so that

$$\mathrm{tr}\{Q\rho\} \geqslant 1 - \epsilon, \tag{5.1}$$

for some small, fixed $\epsilon \geqslant 0$, while minimizing the probability that we guess σ incorrectly. This naturally leads to a semidefinite optimization program, specified by the following quantity:

$$\beta_{\epsilon}(\rho, \sigma) \equiv \min_{Q} \{\mathrm{tr}\{Q\sigma\} : 0 \leqslant Q \leqslant I, \ \mathrm{tr}\{Q\rho\} \geqslant 1 - \epsilon\}. \tag{5.2}$$

By taking the negative logarithm of $\beta_{\epsilon}(\rho, \sigma)$, we arrive at the hypothesis testing relative entropy defined in [25, 142]:

$$D_H^{\epsilon}(\rho||\sigma) \equiv -\log \beta_{\epsilon}(\rho, \sigma). \tag{5.3}$$

One can derive other entropic measures based on the hypothesis testing relative entropy that have various natural properties [41]. Also, one may define, for two distributions P, Q on a classical alphabet $\mathcal{X}$, the hypothesis testing relative entropy $D_H^\epsilon(P\|Q)$ by

$$D_H^\epsilon(P\|Q) = -\log_2 \inf_{A:P(A)\geqslant 1-\epsilon} Q(A),$$

where $A \subset \mathcal{X}$ is sometimes called the *decision region.* This corresponds to the case when the two states ρ, σ are simultaneously diagonalizable with respect some basis $\{|i\rangle\}$, and $P(i), Q(i)$ are the eigenvalues of ρ, σ for the eigenvector $|i\rangle$, respectively.

Notation

In our notation, $O(\epsilon)$ is any function of ϵ such that $\lim_{\epsilon\to 0^+} O(\epsilon) = 0$, $\mathbb{P}$ is the set of all primes, $\mathbb{Z}^+$ is the set of positive integers and $\mathbb{R}^+$ is the set of non-negative reals. Since we deal with summations over several groups in this paper, when not clear from the context, we indicate the underlying group in each summation; e.g. summation over the group G is denoted by $\overbrace{\sum}^{(G)}$. Direct sum of groups is denoted by $\bigoplus$ and direct product of sets is denoted by $\bigotimes$.

5.3 Abelian Group Code Ensemble

In this section, we use a standard characterization of Abelian groups and introduce the ensemble of Abelian group codes used in [121] and this chapter.

5.3.1 Abelian Groups

For an Abelian group G, let $\mathcal{P}(G)$ denote the set of all distinct primes which divide $|G|$ and for a prime $p \in \mathcal{P}(G)$ let $S_p(G)$ be the corresponding Sylow subgroup of G. It is known that any Abelian group G can be decomposed as a direct sum of its Sylow subgroups in the following manner

$$G = \bigoplus_{p\in\mathcal{P}(G)} S_p(G) \tag{5.4}$$

Furthermore, each Sylow subgroup $S_p(G)$ can be decomposed into $\mathbb{Z}_{p^r}$ groups as follows: $S_p(G) \cong \bigoplus_{r\in\mathcal{R}_p(G)} \mathbb{Z}_{p^r}^{M_{p,r}}$.where $\mathcal{R}_p(G) \subseteq \mathbb{Z}^+$ and for $r \in \mathcal{R}_p(G)$, $M_{p,r}$ is a positive integer. Note that $\mathbb{Z}_{p^r}^{M_{p,r}}$ is defined as the direct sum of the ring $\mathbb{Z}_{p^r}$ with itself for $M_{p,r}$ times.

Rewriting Equation (5.4), we can represent any Abelian group as follows:

$$G \cong \bigoplus_{p \in \mathcal{P}(G)} \bigoplus_{r \in \mathcal{R}_p(G)} \mathbb{Z}_{p^r}^{M_{p,r}} = \bigoplus_{p \in \mathcal{P}(G)} \bigoplus_{r \in \mathcal{R}_p(G)} \bigoplus_{m=1}^{M_{p,r}} \mathbb{Z}_{p^r}^{(m)} \tag{5.5}$$

where $\mathbb{Z}_{p^r}^{(m)}$ is called the m^{th} $\mathbb{Z}_{p^r}$ ring of G or the $(p,r,m)^{\text{th}}$ ring of G. Equivalently, this can be written as follows

$$G \cong \bigoplus_{(p,r) \in \mathcal{Q}(G)} \mathbb{Z}_{p^r}^{M_{p,r}} = \bigoplus_{(p,r,m) \in \mathcal{G}(G)} \mathbb{Z}_{p^r}^{(m)}, \tag{5.6}$$

where $\mathcal{Q}(G) \subseteq \mathbb{P} \times \mathbb{Z}^+$ is defined as $\mathcal{Q}(G) = \{(p,r) \mid p \in \mathcal{P}(G), r \in \mathcal{R}_p(G)\}$ and $\mathcal{G}(G) \subseteq \mathbb{P} \times \mathbb{Z}^+ \times \mathbb{Z}^+$ is defined as $\mathcal{G}(G) = \{(p,r,m) \mid (p,r) \in \mathcal{Q}(G), m \in \{1,2,\ldots,M_{p,r}\}\}$. This means any element a of the Abelian group can be regarded as a vector whose components are indexed by $(p,r,m) \in \mathcal{G}(G)$ and whose $(p,r,m)^{\text{th}}$ component $a_{p,r,m}$ takes values from the ring $\mathbb{Z}_{p^r}$. With a slight abuse of notation, we represent an element a of G as $a = \bigoplus_{(p,r,m) \in \mathcal{G}(G)} a_{p,r,m}$. Furthermore, for two elements $a, b \in G$, we have $a + b = \bigoplus_{(p,r,m) \in \mathcal{G}(G)} a_{p,r,m} +_{p^r} b_{p,r,m}$, where $+$ denotes the group operation and $+_{p^r}$ denotes addition mod-p^r. This can equivalently be written as $[a+b]_{p,r,m} = a_{p,r,m} +_{p^r} b_{p,r,m}$, where $[\cdot]_{p,r,m}$ denotes the $(p,r,m)^{\text{th}}$ component of it's argument.

Example 1. Let $G = \mathbb{Z}_4 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_9^2$. Then we have $\mathcal{P}(G) = \{2,3\}$, $S_2(G) = \mathbb{Z}_4$ and $S_3(G) = \mathbb{Z}_3 \oplus \mathbb{Z}_9^2$, $\mathcal{R}_2(G) = \{2\}$, $\mathcal{R}_3(G) = \{1,2\}$, $M_{2,2} = 1$, $M_{3,1} = 1$, $M_{3,2} = 2$ and

$$\mathcal{G}(G) = \{(2,2,1),(3,1,1),(3,2,1),(3,2,2)\}$$

Each element a of G can be represented by a quadruple $(a_{2,2,1}, a_{3,1,1}, a_{3,2,1}, a_{3,2,2})$ where $a_{2,2,1} \in \mathbb{Z}_4$, $a_{3,1,1} \in \mathbb{Z}_3$ and $a_{3,2,1}, a_{3,2,2} \in \mathbb{Z}_9$.

In the following section, we introduce the ensemble of Abelian group codes which we use in the chapter.

5.3.2 The Image Ensemble

Recall that for a positive integer n, an Abelian group code of length n over the group G is a coset of a subgroup of G^n. Our ensemble of codes consists of all Abelian group codes over G; i.e., we consider all cosets of subgroups of G^n. The following lemma ([121, Lemma 1]) effectively characterizes all subgroups of G^n:

Lemma 40. *For an Abelian group $\tilde{G}$, let $\phi : J \to \tilde{G}$ be a homomorphism from some Abelian group J to $\tilde{G}$. Then $\phi(J) \leqslant \tilde{G}$; i.e. the image of the homomorphism is a subgroup of $\tilde{G}$. Moreover, for any subgroup $\tilde{H}$ of $\tilde{G}$ there exists a corresponding Abelian group J and a homomorphism $\phi : J \to \tilde{G}$ such that $\tilde{H} = \phi(J)$.*

Definition 4. Let G be an Abelian group. For $p \in \mathcal{P}(G)$, define $r_p = \max \mathcal{R}_p(G)$, and $\mathcal{S}(G) = \{(p,s) \mid p \in \mathcal{P}(G), 1 \leqslant s \leqslant r_p\}$.

It is shown in [121] that we only need to consider homomorphisms from an Abelian group J to $\tilde{G}$ such that

$$\mathcal{P}(J) \subseteq \mathcal{P}(\tilde{G}), \text{ and } s \leqslant r_q = \max \mathcal{R}_q(\tilde{G}) \text{ for all } (q,s,l) \in \mathcal{G}(J).$$

To construct Abelian group codes of length n over G, let $\tilde{G} = G^n$. We have

$$G^n \cong \bigoplus_{p \in \mathcal{P}(G)} \bigoplus_{r \in \mathcal{R}_p} \mathbb{Z}_{p^r}^{nM_{p,r}} = \bigoplus_{p \in \mathcal{P}(G)} \bigoplus_{r \in \mathcal{R}_p} \bigoplus_{m=1}^{nM_{p,r}} \mathbb{Z}_{p^r}^{(m)} = \bigoplus_{(p,r,m) \in \mathcal{G}(G^n)} \mathbb{Z}_{p^r}^{(m)} \tag{5.7}$$

Define J as

$$J = \bigoplus_{q \in \mathcal{P}(G)} \bigoplus_{s=1}^{r_q} \mathbb{Z}_{q^s}^{k_{q,s}} = \bigoplus_{q \in \mathcal{P}(G)} \bigoplus_{s=1}^{r_q} \bigoplus_{l=1}^{k_{q,s}} \mathbb{Z}_{q^s}^{(l)} = \bigoplus_{(q,s,l) \in \mathcal{G}(J)} \mathbb{Z}_{q^s}^{(l)} \tag{5.8}$$

for some positive integers $k_{q,s}$. Define $k = \sum_{q \in \mathcal{P}(G)} \sum_{s=1}^{r_q} k_{q,s}$ and $w_{q,s} = \frac{k_{q,s}}{k}$ for $q \in \mathcal{P}(G)$ and $s = 1, \cdots, r_q$, so that we can write

$$J = \bigoplus_{q \in \mathcal{P}(G)} \bigoplus_{s=1}^{r_q} \bigoplus_{l=1}^{kw_{q,s}} \mathbb{Z}_{q^s}^{(l)} \tag{5.9}$$

for some constants $w_{q,s}$ adding up to one.

Definition 5. The ensemble of Abelian group encoders consists of all mappings $\phi : J \to G^n$ of the form

$$\phi(a) = \bigoplus_{(p,r,m) \in \mathcal{G}(G^n)} \overbrace{\sum_{(q,s,l) \in \mathcal{G}(J)}}^{(\mathbb{Z}_{p^r})} a_{q,s,l} g_{(q,s,l) \to (p,r,m)} \tag{5.10}$$

for $a \in J$ where $a_{q,s,l} g_{(q,s,l) \to (p,r,m)}$ is the short-hand notation for the mod-p^r addition of

$g_{(q,s,l)\to(p,r,m)}$ to itself for $a_{q,s,l}$ times, and

$$g_{(q,s,l)\to(p,r,m)} \begin{cases} = 0 & \text{if } p \neq q \\ \sim \text{Unif}(\mathbb{Z}_{p^r}) & \text{if } p = q, r \leqslant s \\ \sim \text{Unif}(p^{r-s}\mathbb{Z}_{p^r}) & \text{if } p = q, r \geqslant s \end{cases}$$

The corresponding shifted group code is defined by

$$\mathbb{C} = \{\phi(a) + V | a \in J\}, \tag{5.11}$$

where V is a uniform random variable over G^n.

The rate of this code is given by

$$R = \frac{1}{n} \log |J| = \frac{k}{n} \sum_{q \in \mathcal{P}(G)} \sum_{s=1}^{r_q} s w_{q,s} \log q. \tag{5.12}$$

5.3.3 The $H_{\hat{\theta}}$ coset

A characterization of the joint distribution of the codewords corresponding to two messages is given in this section. This result, shown in [121], requires the following definitions of the *coset index* $\hat{\theta}$ and the corresponding subgroup $H_{\hat{\theta}}$ of G. In addition, two definitions are provided to set the notations for the conditional distributions of the channel input and output given the coset information of the codeword, for both classical and classical-quantum channels.

For an Abelian group G defined in (5.5), denote a vector $\hat{\theta}$ whose components are indexed by $(p, s) \in \mathcal{S}(G)$ by $(\hat{\theta}_{p,s})_{(p,s)\in\mathcal{S}(G)}$, where $0 \leqslant \hat{\theta}_{p,s} \leqslant s, \hat{\theta}_{p,s} \in \mathbb{Z}$. For $\hat{\theta} = (\hat{\theta}_{p,s})_{(p,s)\in\mathcal{S}(G)}$, define

$$\boldsymbol{\theta}(\hat{\theta}) = \left(\min_{\substack{(p,s)\in\mathcal{S}(G) \\ w_{p,s}\neq 0}} |r-s|^+ + \hat{\theta}_{p,s} \right)_{(p,r)\in\mathcal{Q}(G)}.$$

Let $H_{\hat{\theta}}$ be a subgroup of G defined as

$$H_{\hat{\theta}} = \bigoplus_{(p,r,m)\in\mathcal{G}(G)} p^{\boldsymbol{\theta}(\hat{\theta})_{p,r}} \mathbb{Z}_{p^r}^{(m)}. \tag{5.13}$$

For $a \in J$ and $\hat{\theta} = (\hat{\theta}_{p,s})_{(p,s)\in\mathcal{S}(G)}$, let

$$T_{\hat{\theta}}(a) = \{\tilde{a} \in J \mid \tilde{a}_{p,s} - a_{p,s} \in p^{\hat{\theta}_{p,s}}\mathbb{Z}_{p^s}^{k_{p,s}} \backslash p^{\hat{\theta}_{p,s}+1}\mathbb{Z}_{p^s}^{k_{p,s}}, \forall (p,s) \in \mathcal{S}(G)\}.$$

Then we have $|T_{\hat{\theta}}(a)| = \prod_{(p,s)\in\mathcal{S}(G)} p^{(s-\hat{\theta}_{p,s})k_{p,s}}$ for all $a \in J$. Therefore, we may write $|T_{\hat{\theta}}(a)| = |T_{\hat{\theta}}|$ without any ambiguity. Let $\omega_{\hat{\theta}}$ be defined by

$$\omega_{\hat{\theta}} = \frac{\sum_{(p,s)\in\mathcal{S}(G)} \hat{\theta}_{p,s} w_{p,s} \log p}{\sum_{(p,s)\in\mathcal{S}(G)} s w_{p,s} \log p}, \tag{5.14}$$

we have

$$\begin{aligned} \log |T_{\hat{\theta}}| &= \sum_{(p,s)\in\mathcal{S}(G)} (s - \hat{\theta}_{p,s}) k_{p,s} \log p = k \cdot \left[\frac{\sum_{(p,s)\in\mathcal{S}(G)} (s - \hat{\theta}_{p,s}) \omega_{p,s} \log p}{\sum_{(p,s)\in\mathcal{S}(G)} s \omega_{p,s} \log p} \right] \sum_{(p,s)\in\mathcal{S}(G)} s \omega_{p,s} \log p \\ &= (1 - \omega_{\hat{\theta}}) k \sum_{(p,s)\in\mathcal{S}(G)} s \omega_{p,s} \log p = (1 - \omega_{\hat{\theta}}) nR. \end{aligned} \tag{5.15}$$

For any $a \in J$, $\{T_{\hat{\theta}}(a)\}_{\hat{\theta}}$ is a union of disjoint sets whose union is $\cup_{\hat{\theta}} T_{\hat{\theta}}(a) = J$. Hence $\sum_{\hat{\theta}} |T_{\hat{\theta}}| = \sum_{\hat{\theta}} |T_{\hat{\theta}}(a)| = |J|$. Exploiting equation (5.15), we have that $\sum_{\hat{\theta}} 2^{(1-\omega_{\hat{\theta}})nR} = |J|$, or equivalently, $\sum_{\hat{\theta}} 2^{(1-\omega_{\hat{\theta}})} = 1$.

We are now ready to state the result from [121].

Lemma 41. *For $a, \tilde{a} \in J$, $x, \tilde{x} \in G^n$ and for $(p,s) \in \mathcal{Q}(J) = \mathcal{S}(G)$, let $\hat{\theta}_{p,s} \in \{0, 1, \cdots, s\}$ be such that $\tilde{a}_{p,s} - a_{p,s} \in p^{\hat{\theta}_{p,s}}\mathbb{Z}_{p^s}^{k_{p,s}} \backslash p^{\hat{\theta}_{p,s}+1}\mathbb{Z}_{p^s}^{k_{p,s}}$, i.e., $\tilde{a} \in T_{\hat{\theta}}(a)$. Consider a random homomorphism ϕ and a dither V with distribution specified as in Definition 5. Then,*

$$\Pr(\phi(a) + V = x, \phi(\tilde{a}) + V = \tilde{x}) = \begin{cases} \frac{1}{|G|^n} \frac{1}{|H_{\hat{\theta}}|^n} & \textit{if } \tilde{x} - x \in H_{\hat{\theta}}^n \\ 0 & \textit{otherwise} \end{cases}$$

We use the following notations for the conditional distributions of the codeword and channel output given the *coset* information.

Definition 6. Let $H = H_{\hat{\theta}}$ be a subgroup of G and $x_r \in G$. Let X be distributed according to $P_X \equiv \text{Unif}(\mathcal{X})$, the uniform distribution over $\mathcal{X} = G$, and $W = (\mathcal{X}, \mathcal{Y}, W_{Y|X})$ be a classical channel. Then, for a coset $[x_r]_{\hat{\theta}} = x_r + H$ of H in G, define

$$P_{[X_r]}([x_r]) \triangleq \Pr(X \in [x_r]) = \frac{|H|}{|G|},$$

$$P_{X|[X_r]}(x \,|\, [x_r]) \triangleq \Pr(X = x \mid X \in [x_r]) = \begin{cases} \frac{1}{|H|} & \text{if } x \in [x_r], \\ 0 & \text{otherwise,} \end{cases}$$

$$\begin{aligned} P_{Y|[X_r]}(y \,|\, [x_r]) \triangleq \Pr(Y = y \mid X \in [x_r]) &= \sum_{x \in [x_r]} P_{X|[X_r]}(x|[x_r]) W_{Y|X}(y|x) \\ &= \sum_{x \in [x_r]} \frac{1}{|H|} W_{Y|X}(y|x), \end{aligned}$$

where we write $[x_r]$ and $[x_r]_{\hat{\theta}}$ interchangeably when the dependency of $\hat{\theta}$ is clear from the context. For $x_r^n = (x_{r,1}, x_{r,2}, \ldots, x_{r,n}) \in G^n$, $[x_r^n]$ denotes the coset $x_r^n + H^n$ in G^n, and the product conditional distribution $P^n_{Y|[X_r]}$ is defined as

$$\begin{aligned} P^n_{Y|[X_r]}(y^n \,|\, [x_r^n]) \triangleq \prod_{i=1}^{n} P_{Y|[X_r]}(y_i \,|\, [x_{r,i}]) &= \sum_{x_1 \in [x_{r,1}]} \cdots \sum_{x_n \in [x_{r,n}]} \frac{1}{|H|^n} W_{Y|X}(y_1|x_1) \cdots W_{Y|X}(y_n|x_n) \\ &= \sum_{x^n \in [x_r^n]} P^n_{X|[X_r]}(x^n|[x_r^n]) W^n_{Y|X}(y^n|x^n). \end{aligned}$$

For the transmission over a classical-quantum channel, given the coset information, we use the following notation for the conditional distribution of (classical) codeword and the expected output (quantum) state.

Definition 7. Let $H = H_{\hat{\theta}}$ be a subgroup of G and $x_r \in G$. Let X be distributed according to $P_X \equiv \text{Unif}(\mathcal{X})$, the uniform distribution over $\mathcal{X} = G$, and $\mathcal{N} = \{\rho_x\}_{x \in \mathcal{X}}$ be a classical-quantum channel from the classical alphabet $\mathcal{X}$ to the quantum system $\mathbb{B}$. Then, for a coset $[x_r]_{\hat{\theta}} = x_r + H$ of H in G, define $P_{[X_r]}([x_r]) \triangleq \Pr(X \in [x_r]) = \frac{|H|}{|G|}$, $P_{X|[X_r]}(x \,|\, [x_r]) \triangleq \Pr(X = x \mid X \in [x_r])$, which is $\frac{1}{|H|}$ for $x \in [x_r]$ and 0 otherwise. Also, define

$$\rho^B_{[x_r]} \triangleq \sum_{x \in [x_r]} P_{X|[X_r]}(x|[x_r]) \rho^B_x = \sum_{x \in [x_r]} \frac{1}{|H|} \rho^B_x.$$

For $x_r^n = (x_{r,1}, x_{r,2}, \ldots, x_{r,n}) \in G^n$, $[x_r^n]$ denotes the coset $x_r^n + H^n$ in G^n, and the conditional expected state $\rho^{B^n}_{[x^n_r]}$ on $\mathbb{B}^{\otimes n}$ is defined as

$$\rho^{B^n}_{[x^n_r]} \triangleq \rho^{B_1}_{[x_{r,1}]} \otimes \rho^{B_2}_{[x_{r,2}]} \otimes \cdots \rho^{B_n}_{[x_{r,n}]} = \sum_{x^n \in [x^n_r]} P^n_{X|[X_r]}(x^n|[x^n_r])\rho^{B^n}_{x^n}.$$

5.4 One-shot Classical Group Coding

In this section, we consider the one-shot coding problem for a classical channel W where the code is a group code in the form of (5.11), i.e., $\mathbb{C} = \{\phi(a)+V|a \in J\}$. Two approaches are given in the achievability part. The main result from the first approach resembles the group capacity results from [121], while the second approach is more comprehensive and implies the result of the first approach. A converse bound is shown in Section 5.4.2 by considering subcodes of a given code $\mathbb{C}$.

Let Abelian groups G and J be given, as in equations (5.6) and (5.8) respectively, by

$$G \cong \bigoplus_{p\in\mathcal{P}(G)} \bigoplus_{r\in\mathcal{R}_p(G)} \mathbb{Z}^{M_{p,r}}_{p^r} = \bigoplus_{(p,r,m)\in\mathcal{G}(G)} \mathbb{Z}^{(m)}_{p^r}. \tag{5.16}$$

$$J = \bigoplus_{q\in\mathcal{P}(G)} \bigoplus_{s=1}^{r_q} \mathbb{Z}^{k_{q,s}}_{q^s} = \bigoplus_{q\in\mathcal{P}(G)} \bigoplus_{s=1}^{r_q} \bigoplus_{l=1}^{k_{q,s}} \mathbb{Z}^{(l)}_{q^s} = \bigoplus_{(q,s,l)\in\mathcal{G}(J)} \mathbb{Z}^{(l)}_{q^s}. \tag{5.17}$$

For two distributions P, Q on a classical alphabet, recall the hypothesis testing relative entropy $D^\epsilon_H(P\|Q)$ is given by

$$D^\epsilon_H(P\|Q) = -\log_2 \inf_{A:P(A)\geqslant 1-\epsilon} Q(A).$$

Given a channel $W = (\mathcal{X} = G, \mathcal{Y}, W_{Y|X})$, let the joint distribution P_{XY} be $P_{XY} = P_X \cdot W_{Y|X}$ with P_X being the uniform distribution over $\mathcal{X}$, and P_Y be the marginal distribution of P_{XY} over $\mathcal{Y}$. Let $H_{\hat{\theta}}$ be a subgroup of G defined in (5.13). When $[X] = [X]_{\hat{\theta}} = X + H_{\hat{\theta}}$ for some $\hat{\theta}$, we have

$$\begin{aligned} & D^\epsilon_H(P_{XY}\|P_{[X]}P_{X|[X]}P_{Y|[X]}) \\ &= -\log_2 \inf_{\substack{A\subset\mathcal{X}\times\mathcal{Y}\\ P_{XY}(A)\geqslant 1-\epsilon}} \sum_{[x_r]} \frac{|H_{\hat{\theta}}|}{|G|} \sum_{x\in[x_r]} P_{X|[X_r]}(x\,|\,[x_r]) \sum_{y:(x,y)\in A} P_{Y|[X_r]}(y\,|\,[x_r]) \qquad (5.18) \\ &= -\log_2 \inf_{\substack{A\subset\mathcal{X}\times\mathcal{Y}\\ P_{XY}(A)\geqslant 1-\epsilon}} \sum_{(x,y)\in A} \frac{1}{|G|} P_{Y|[X_r]}(y\,|\,[x]). \qquad (5.19) \end{aligned}$$

We can also denote this quantity as $D^{\epsilon,\hat{\theta}}_H(P_{XY}\|P_{[X]}P_{X|[X]}P_{Y|[X]})$ if we would like to make its dependence on $\hat{\theta}$ explicit.

5.4.1 Achievability

Let the ensemble of homomorphisms ϕ from J to G and the group code $\mathbb{C} = \{\phi(a)+V | a \in J\}$ be given as in Definition 5 with $n = 1$. Given a channel $W = (\mathcal{X} = G, \mathcal{Y}, W_{Y|X})$, consider a decision region $A_\epsilon \subset \mathcal{X} \times \mathcal{Y}$, which will be constructed explicitly later, such that

$$P_{XY}(A_\epsilon) = \sum_{(x,y)\in A_\epsilon} P_X(x) W_{Y|X}(y|x) \geqslant 1 - \epsilon, \tag{5.20}$$

where P_X is uniform over G.

To find an achievable rate, we use a random coding argument in which the random encoder is characterized by the random homomorphism ϕ and a random vector V uniformly distributed over G. Given a message $u \in J$, the encoder maps it to $x = \phi(u) + V$ and x is then fed to the channel. At the receiver, after receiving the channel output $y \in \mathcal{Y}$, the decoder looks for a unique $\tilde{u} \in J$ such that $(\phi(\tilde{u}) + V, y) \in A_\epsilon$. If the decoder does not find such $\tilde{u}$ or if such $\tilde{u}$ is not unique, it declares error. Thus, the error event can be characterized by the union of two events: $E(u) = E_1(u) \cup E_2(u)$ where $E_1(u)$ is the event that $(\phi(u) + V, y) \notin A_\epsilon$ and $E_2(u)$ is the event that there exists a $\tilde{u} \neq u$ such that $(\phi(\tilde{u}) + V, y) \in A_\epsilon$. We can provide an upper bound on the probability of the error event as $\Pr(E(u)) \leqslant \Pr(E_1(u)) + \Pr(E_2(u) \cap (E_1(u))^c)$.

Denote by $P_{err}(u)$ the probability of the event $E_2(u) \cap (E_1(u))^c$, averaged over the randomness of ϕ, V.

$$\begin{aligned}
P_{err}(u) &\triangleq \mathbb{E}\left[\Pr(E_2(u) \cap (E_1(u))^c)\right] \\
&= \sum_{(x,y)\in A_\epsilon} W_{Y|X}(y|x) \Pr\left(\phi(u) + V = x, \exists \tilde{u} \in J : \tilde{u} \neq u, (\phi(\tilde{u}) + V, y) \in A_\epsilon\right) \\
&\leqslant \sum_{(x,y)\in A_\epsilon} \sum_{\substack{\tilde{u}\in J \\ \tilde{u}\neq u}} \sum_{\substack{\tilde{x}\in G \\ (\tilde{x},y)\in A_\epsilon}} W_{Y|X}(y|x) \Pr\left(\phi(u) + V = x, \phi(\tilde{u}) + V = \tilde{x}\right) \\
&= \sum_{\hat{\theta}\neq \mathbf{s}} \sum_{(x,y)\in A_\epsilon} \sum_{\tilde{u}\in T_{\hat{\theta}}(u)} \sum_{\substack{\tilde{x}\in G \\ (\tilde{x},y)\in A_\epsilon}} W_{Y|X}(y|x) \Pr\left(\phi(u) + V = x, \phi(\tilde{u}) + V = \tilde{x}\right) \\
&= \sum_{\hat{\theta}\neq \mathbf{s}} P_{err}(u, \hat{\theta}),
\end{aligned} \tag{5.21}$$

where $P_{err}(u, \hat{\theta}) \triangleq \sum_{\tilde{u} \in T_{\hat{\theta}}(u)} P_{err}(u, \tilde{u})$ and for $\tilde{u} \in T_{\hat{\theta}}(u)$,

$$P_{err}(u, \tilde{u}) \triangleq \sum_{(x,y)\in A_\epsilon} \sum_{\substack{\tilde{x}\in G \\ (\tilde{x},y)\in A_\epsilon}} W_{Y|X}(y|x) \Pr\left(\phi(u) + V = x, \phi(\tilde{u}) + V = \tilde{x}\right), \tag{5.22}$$

and $\mathbf{s}$ denote the vector whose compnents satisfy $\mathbf{s}_{(p,s)} = s$ for all $(p, s) \in \mathcal{S}(G)$. The term $\Pr(\phi(u) + V = x,\, \phi(\tilde{u}) + V = \tilde{x})$ in (5.22) can be found using Lemma 41. Hence

$$\begin{aligned}
P_{err}(u, \tilde{u}) &= \sum_{(x,y)\in A_\epsilon} \sum_{\tilde{x}\in x + H_{\hat{\theta}}} W_{Y|X}(y|x) \mathbb{I}_{A_\epsilon}(\tilde{x}, y) \frac{1}{|G|} \frac{1}{|H_{\hat{\theta}}|} \\
&= \sum_{x\in\mathcal{X}} \sum_{\tilde{x}\in x+H_{\hat{\theta}}} \sum_{\substack{y:(x,y)\in A_\epsilon \\ (\tilde{x},y)\in A_\epsilon}} W_{Y|X}(y|x) \frac{1}{|G|} \frac{1}{|H_{\hat{\theta}}|} \\
&\leqslant \sum_{x\in\mathcal{X}} \sum_{\tilde{x}\in x+H_{\hat{\theta}}} \sum_{y:(\tilde{x},y)\in A_\epsilon} W_{Y|X}(y|x) \frac{1}{|G|} \frac{1}{|H_{\hat{\theta}}|} \\
&= \sum_{[x_r]} \frac{|H_{\hat{\theta}}|}{|G|} \sum_{x\in[x_r]} \sum_{\tilde{x}\in[x_r]} \sum_{y:(\tilde{x},y)\in A_\epsilon} W_{Y|X}(y|x) \frac{1}{|H_{\hat{\theta}}|^2} \\
&= \sum_{[x_r]} P_{[X_r]}([x_r]) \sum_{\tilde{x}\in[x_r]} \sum_{y:(\tilde{x},y)\in A_\epsilon} P_{X|[X_r]}(\tilde{x} \mid [x_r]) P_{Y|[X_r]}(y \mid [x_r]), \qquad (5.23)
\end{aligned}$$

where $[x_r]$ denotes a coset of $H_{\hat{\theta}}$ in G, that is, $[x_r] = x_r + H_{\hat{\theta}}$.

5.4.1.1 First Approach

Let a set of parameters $\{\epsilon_{\hat{\theta}}\}_{\hat{\theta}\neq\mathbf{s}}$ be given such that $\epsilon_{\hat{\theta}} > 0$ for each $\hat{\theta}$ and that $\sum_{\hat{\theta}} \epsilon_{\hat{\theta}} = \epsilon$. Let the set $A^*_{\epsilon_{\hat{\theta}}}$ be a maximizer of the right-hand side of (5.18) for $D_H^{\epsilon_{\hat{\theta}},\hat{\theta}}(P_{XY} \| P_{[X]} P_{X|[X]} P_{Y|[X]})$, i.e., $P_{XY}(A^*_{\epsilon_{\hat{\theta}}}) \geqslant 1 - \epsilon_{\hat{\theta}}$ and

$$\begin{aligned}
&D_H^{\epsilon_{\hat{\theta}},\hat{\theta}}(P_{XY} \| P_{[X]} P_{X|[X]} P_{Y|[X]}) \\
&\quad = -\log_2 \sum_{[x_r]} P_{[X_r]}([x_r]) \sum_{x\in[x_r]} P_{X|[X_r]}(x \,|\, [x_r]) \sum_{y:(x,y)\in A^*_{\epsilon_{\hat{\theta}}}} P_{Y|[X_r]}(y \,|\, [x_r]).
\end{aligned}$$

Now we set explicitly $A_\epsilon = \cap_{\hat{\theta}} A^*_{\epsilon_{\hat{\theta}}}$. The probability of the event $E_1(u)$ can be bounded as:

$$\begin{aligned}\Pr(E_1(u)) &= \Pr\left((\phi(u) + V, Y) \notin A_\epsilon\right) = \Pr((X,Y) \notin A_\epsilon) \\ &= \Pr\left((X,Y) \in \cup_{\hat{\theta}} (A^*_{\epsilon_{\hat{\theta}}})^C\right) \\ &\leqslant \sum_{\hat{\theta}} \Pr\left((X,Y) \in (A^*_{\epsilon_{\hat{\theta}}})^C\right) \leqslant \sum_{\hat{\theta}} \epsilon_{\hat{\theta}} = \epsilon.\end{aligned}$$

Since $A_\epsilon \subset A^*_{\epsilon_{\hat{\theta}}}$, the term in (5.22) can then be upper bounded by

$$\begin{aligned}P_{err}(u,\tilde{u}) &\leqslant \sum_{[x_r]} P_{[X_r]}([x_r]) \sum_{\tilde{x}\in[x_r]} \sum_{y:(\tilde{x},y)\in A^*_{\epsilon_{\hat{\theta}}}} P_{X|[X_r]}(\tilde{x}|\,[x_r]) P_{Y|[X_r]}(y|\,[x_r]) \\ &= \exp_2\left\{-D_H^{\epsilon_{\hat{\theta}},\hat{\theta}}(P_{XY} \| P_{[X]} P_{X|[X]} P_{Y|[X]})\right\},\end{aligned}$$

which implies the following bound on $P_{err}(u)$:

$$P_{err}(u) \leqslant \sum_{\hat{\theta}\neq \mathbf{s}} |T_{\hat{\theta}}(u)| \exp_2\left\{-D_H^{\epsilon_{\hat{\theta}}}(P_{XY} \| P_{[X]_{\hat{\theta}}} P_{X|[X]_{\hat{\theta}}} P_{Y|[X]_{\hat{\theta}}})\right\}.$$

Therefore, we have $\Pr(E(u)) \leqslant \epsilon + \sum_{\hat{\theta}\neq\mathbf{s}} |T_{\hat{\theta}}(u)| \exp_2\left\{-D_H^{\epsilon_{\hat{\theta}}}(P_{XY} \| P_{[X]_{\hat{\theta}}} P_{X|[X]_{\hat{\theta}}} P_{Y|[X]_{\hat{\theta}}})\right\}$. The average probability of error of the group transmission scheme can be upper bounded by

$$\Pr(\text{error}) = \sum_{u\in J} \frac{1}{|J|} \Pr(E(u)) \leqslant \epsilon + \sum_{\hat{\theta}\neq\mathbf{s}} |T_{\hat{\theta}}(u)| \exp_2\left\{-D_H^{\epsilon_{\hat{\theta}}}(P_{XY} \| P_{[X]_{\hat{\theta}}} P_{X|[X]_{\hat{\theta}}} P_{Y|[X]_{\hat{\theta}}})\right\}.$$

Exploiting equation (5.15), we may state the result in terms of the rate R of the code:

Theorem 42. *Let ϵ and $\{\epsilon_{\hat{\theta}}\}$ be given with $\epsilon_{\hat{\theta}} > 0$ for all $\hat{\theta}$ and $\sum_{\hat{\theta}} \epsilon_{\hat{\theta}} \leqslant \epsilon$. Then there exists a (J,ϵ')-code such that*

$$\epsilon' \leqslant \epsilon + \sum_{\hat{\theta}\neq\mathbf{s}} \exp_2\left\{(1-\omega_{\hat{\theta}})R - D_H^{\epsilon_{\hat{\theta}}}(P_{XY} \| P_{[X]_{\hat{\theta}}} P_{X|[X]_{\hat{\theta}}} P_{Y|[X]_{\hat{\theta}}})\right\},$$

where the rate $R = k\sum_{(p,s)\in\mathcal{S}(G)} s\omega_{p,s}\log p$ is given in Equation (5.12).

Example 2. Let $J = \mathbb{Z}_4, G = \mathbb{Z}_8$. In this example, we have $\mathcal{G}(G) = \{(2,3,1)\}$ and $\mathcal{G}(J) = \{(2,2,1)\}$, $k_{2,1} = 0, k_{2,2} = 1, k_{2,3} = 0$, and the term $g_{(2,2,1)\to(2,3,1)}$ is a uniform random variable over $2\mathbb{Z}_8$. For simplicity, we write $u = u_{2,2,1} \in J$ and $g = g_{(2,2,1)\to(2,3,1)} \in \mathbb{Z}_8$. Then $r_2 = \max \mathcal{R}_2(G) = 3$ and the set $\mathcal{Q}(J) = \mathcal{S}(G) = \{(2,1),(2,2),(2,3)\}$, and $\mathbf{s}_{(2,1)} = 1, \mathbf{s}_{(2,2)} = 2, \mathbf{s}_{(2,3)} = 3$. For distinct $u, \tilde{u} \in J$, the vector $\hat{\theta} = (\hat{\theta}_{2,1}, \hat{\theta}_{2,2}, \hat{\theta}_{2,3})$ for which $\tilde{u} \in T_{\hat{\theta}}(u)$ must have $\hat{\theta}_{2,1} = 1, 0 \leqslant \hat{\theta}_{2,2} < 2, \hat{\theta}_{2,3} = 3$. Thus $P_{err}(u) \leqslant P_{err}(u,(1,0,3)) + P_{err}(u,(1,1,3))$.

The set $\mathcal{Q}(G) = \{(2,3)\}$, so $\boldsymbol{\theta}(\hat{\theta}) = \boldsymbol{\theta}(\hat{\theta})_{(2,3)}$ and

$$\boldsymbol{\theta}(\hat{\theta})_{(2,3)} = \min_{\substack{(2,s)\in\mathcal{S}(G)\\ w_{2,s}\neq 0}} \left\{|3-s|^{+} + \hat{\theta}_{2,s}\right\} = |3-2|^{+} + \hat{\theta}_{2,2} = 1 + \hat{\theta}_{2,2}.$$

Case 1: $\hat{\theta}_{2,2} = 0$, $\boldsymbol{\theta}(\hat{\theta})_{(2,3)} = 1$

For $\tilde{u} \in T_{\hat{\theta}}(u)$, $\tilde{u} - u \in Z_4\backslash 2\mathbb{Z}_4$, and $H_{\hat{\theta}} = 2\mathbb{Z}_8$. The double sum

$$\begin{aligned}
&\sum_{(x,y)\in A_\epsilon} \sum_{\substack{\tilde{x}\in G\\ (\tilde{x},y)\in A_\epsilon}} W_{Y|X}(y|x) \Pr\left(\phi(u)+V = x, \phi(\tilde{u})+V=\tilde{x}\right)\\
&= \sum_{(x,y)\in A_\epsilon} \sum_{\tilde{x}\in x+H_{\hat{\theta}}} W_{Y|X}(y|x)\mathbb{I}_{A_\epsilon}(\tilde{x},y)\frac{1}{|G|}\frac{1}{|H_{\hat{\theta}}|} \leqslant \sum_{x\in\mathcal{X}} \sum_{\tilde{x}\in x+H_{\hat{\theta}}} \sum_{y:(\tilde{x},y)\in A_\epsilon} W_{Y|X}(y|x)\frac{1}{|G|}\frac{1}{|H_{\hat{\theta}}|}\\
&= \sum_{[x_r]} \frac{|H_{\hat{\theta}}|}{|G|} \sum_{x\in[x_r]} \sum_{\tilde{x}\in[x_r]} \sum_{y:(\tilde{x},y)\in A_\epsilon} W_{Y|X}(y|x)\frac{1}{|H_{\hat{\theta}}|^2}\\
&= \sum_{[x_r]} \frac{|H_{\hat{\theta}}|}{|G|} \sum_{\tilde{x}\in[x_r]} \sum_{y:(\tilde{x},y)\in A_\epsilon} P_{X|[X_r]}(\tilde{x}|\,[x_r])P_{Y|[X_r]}(y|\,[x_r])\\
&\leqslant \sum_{[x_r]} \frac{|H_{\hat{\theta}}|}{|G|} \sum_{\tilde{x}\in[x_r]} \sum_{y:(\tilde{x},y)\in A^*_{\epsilon/2,\hat{\theta}}} P_{X|[X_r]}(\tilde{x}|\,[x_r])P_{Y|[X_r]}(y|\,[x_r])\\
&= 2^{-\log D_H^{\epsilon/2}(P_{XY}\|P_{[X]_{\hat{\theta}}}P_{X|[X]_{\hat{\theta}}}P_{Y|[X]_{\hat{\theta}}})},
\end{aligned}$$

where $A^*_{\epsilon/2,\hat{\theta}}$ is a maximizer for $D_H^{\epsilon/2}(P_{XY}\|P_{[X]_{\hat{\theta}}}P_{X|[X]_{\hat{\theta}}}P_{Y|[X]_{\hat{\theta}}})$ and $A_\epsilon \subset A^*_{\epsilon/2,\hat{\theta}}$.

Hence we have

$$\begin{aligned}
P_{err}(u,(1,0,3)) &= \sum_{\tilde{u}\in T_{\hat{\theta}}(u)} \sum_{(x,y)\in A_\epsilon} \sum_{\substack{\tilde{x}\in G\\ (\tilde{x},y)\in A_\epsilon}} W_{Y|X}(y|x)\Pr\left(\phi(u)+V=x,\phi(\tilde{u})+V=\tilde{x}\right)\\
&\leqslant \left|T_{(1,0,3)}(u)\right| \exp_2\left[-\log D_H^{\epsilon/2}(P_{XY}\|P_{[X]_{\hat{\theta}}}P_{X|[X]_{\hat{\theta}}}P_{Y|[X]_{\hat{\theta}}}\right].
\end{aligned}$$

Case 2: $\hat{\theta}_{2,2} = 1$, $\boldsymbol{\theta}(\hat{\theta})_{(2,3)} = 2$

In this case, $H_{\hat{\theta}} = 4\mathbb{Z}_8$, and we have

$$\begin{aligned}
P_{err}(u,(1,1,3)) &= \sum_{\tilde{u}\in T_{\hat{\theta}}} \sum_{(x,y)\in A_\epsilon} \sum_{\substack{\tilde{x}\in G\\ (\tilde{x},y)\in A_\epsilon}} W_{Y|X}(y|x)\Pr\left(\phi(u)+V=x,\phi(\tilde{u})+V=\tilde{x}\right)\\
&\leqslant \left|T_{(1,1,3)}(u)\right| \exp_2\left[-\log D_H^{\epsilon/2}(P_{XY}\|P_{[X]_{\hat{\theta}}}P_{X|[X]_{\hat{\theta}}}P_{Y|[X]_{\hat{\theta}}}\right].
\end{aligned}$$

Therefore the error probability for a message u is

$$\begin{aligned}\Pr(E(u)) &\leqslant \Pr(E_1(u)) + \Pr(E_2(u) \cap (E_1(u))^c) \\ &\leqslant \epsilon + \left|T_{(1,0,3)}(u)\right| \exp_2 \left[- \log D_H^{\epsilon/2}(P_{XY} \| P_{[X]_{\hat{\theta}}} P_{X|[X]_{\hat{\theta}}} P_{Y|[X]_{\hat{\theta}}}\right]_{\hat{\theta}=(1,0,3)} \\ &\quad + \left|T_{(1,1,3)}(u)\right| \exp_2 \left[- \log D_H^{\epsilon/2}(P_{XY} \| P_{[X]_{\hat{\theta}}} P_{X|[X]_{\hat{\theta}}} P_{Y|[X]_{\hat{\theta}}}\right]_{\hat{\theta}=(1,1,3)},\end{aligned}$$

where we consider the decision region $A_\epsilon = \cap_{\hat{\theta}} A^*_{\epsilon/2,\hat{\theta}}$. More generally, one may show that

$$\Pr(E(u)) \leqslant \epsilon + \sum_{\hat{\theta}} |T_{\hat{\theta}}(u)| \exp_2 \left[- \log D_H^{\epsilon_{\hat{\theta}}}(P_{XY} \| P_{[X]_{\hat{\theta}}} P_{X|[X]_{\hat{\theta}}} P_{Y|[X]_{\hat{\theta}}}\right],$$

where $\epsilon_{\hat{\theta}} > 0$ for all $\hat{\theta}$ and $\sum_{\hat{\theta}} \epsilon_{\hat{\theta}} \leqslant \epsilon$. That is, let $\{\epsilon_{\hat{\theta}}\}$ be given with $\epsilon_{\hat{\theta}} > 0$ for all $\hat{\theta}$ and $\sum_{\hat{\theta}} \epsilon_{\hat{\theta}} \leqslant \epsilon'$. Then there exists a (J, ϵ)-code such that

$$\epsilon \leqslant \epsilon' + \sum_{\hat{\theta}} |T_{\hat{\theta}}(u)| \exp_2 \left[- \log D_H^{\epsilon_{\hat{\theta}}}(P_{XY} \| P_{[X]_{\hat{\theta}}} P_{X|[X]_{\hat{\theta}}} P_{Y|[X]_{\hat{\theta}}}\right].$$

5.4.1.2 Second Approach

Let Θ be the set of vectors $\hat{\theta}$ indexed by $(p, s) \in \mathcal{S}(G)$ such that $0 \leqslant \hat{\theta}_{p,s} \leqslant s$ and $\hat{\theta} \neq \mathbf{s}$, and denote its size by $M \triangleq |\Theta|$. We may also index the vectors in the set, so that $\Theta = \left\{\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_M\right\}$. In addition, denote by $\hat{\theta}_0 = \mathbf{s}$, and denote by Θ^* the union $\Theta \cup \left\{\hat{\theta}_0\right\}$. For $\hat{\theta} \in \Theta^*$, let $[x_r]$ denote $[x_r] = [x_r]_{\hat{\theta}} = x_r + H_{\hat{\theta}}$. Define two conditional joint distributions $P_{XY|[X_r]}$ by

$$\begin{aligned}P_{XY|[X_r]}(\tilde{x}, y \mid [x_r]) &\triangleq P_{X|[X_r]}(\tilde{x} \mid [x_r]) P_{Y|[X_r]}(y \mid [x_r]), \\ P_{XY|\hat{\theta}}(\tilde{x}, y) &\triangleq \sum_{[x_r]} P_{[X_r]}([x_r]) P_{XY|[X_r]}(\tilde{x}, y \mid [x_r]),\end{aligned}$$

and a joint distribution

$$\begin{aligned}
P_{XY|J}(\tilde{x}, y) &\triangleq |J|^{-1} \sum_{\hat{\theta}\in\Theta^*} |T_{\hat{\theta}}| \, P_{XY|\hat{\theta}}(\tilde{x}, y) && (5.24)\\
&= |J|^{-1} \sum_{\hat{\theta}} |T_{\hat{\theta}}| \sum_{[x_r]} P_{[X_r]}([x_r]) P_{X|[X_r]}(\tilde{x} \mid [x_r]) P_{Y|[X_r]}(y \mid [x_r])\\
&= |J|^{-1} \sum_{\hat{\theta}} |T_{\hat{\theta}}| \frac{1}{|G|} \sum_{x\in[\tilde{x}]} \frac{1}{|H_{\hat{\theta}}|} W_{Y|X}(y|x) && (5.25)\\
&= |J|^{-1} \sum_{\hat{\theta}} \sum_{\tilde{u}\in T_{\hat{\theta}}(u)} \sum_{x\in G} W_{Y|X}(y|x) \Pr\left(\phi(u) + V = x, \phi(\tilde{u}) + V = \tilde{x}\right), && (5.26)
\end{aligned}$$

where u is an arbitrary element of J. Note that $P_X(x) = \frac{1}{|G|}$ for all $x \in \mathcal{X}$, thereby

$$\begin{aligned}
P_{XY|J}(\tilde{x}, y) &= P_X(\tilde{x}) \, |J|^{-1} \sum_{\hat{\theta}} |T_{\hat{\theta}}| \sum_{x\in[\tilde{x}]} \frac{1}{|H_{\hat{\theta}}|} W_{Y|X}(y|x) && (5.27)\\
&= P_X(\tilde{x}) \, |J|^{-1} \sum_{\hat{\theta}} |T_{\hat{\theta}}| \, P_{Y|[X_r]}(y \mid [\tilde{x}]). && (5.28)
\end{aligned}$$

Let A_ϵ be a subset of $\mathcal{X} \times \mathcal{Y}$ that achieves $D_H^\epsilon(P_{XY} \| P_{XY|J})$. That is, $P_{XY}(A_\epsilon) \geqslant 1 - \epsilon$ and

$$D_H^\epsilon(P_{XY} \| P_{XY|J}) = -\log_2 P_{XY|J}(A_\epsilon), \qquad (5.29)$$

where we use $P_{XY|J}(A)$ as a shorthand notation for $\sum_{(x,y)\in A} P_{XY|J}(x, y)$. The probability for the event $E_1(u)$ is given by

$$\begin{aligned}
\Pr(E_1(u)) &= \Pr\left((\phi(u) + V, Y) \notin A_\epsilon\right) = 1 - \Pr\left((\phi(u) + V, Y) \in A_\epsilon\right)\\
&= 1 - \sum_{x} \Pr(\phi(u) + V = x) \sum_{y:(x,y)\in A_\epsilon} W_{Y|X}(y \,|x)\\
&= 1 - \sum_{(x,y)\in A_\epsilon} P_X(x) W_{Y|X}(y \,|x) = 1 - P_{XY}(A_\epsilon) \leqslant \epsilon, && (5.30)
\end{aligned}$$

where P_X is the uniform distribution on $\mathcal{X}$.

Combining (5.21) and (5.23), we have

$$\begin{aligned}
P_{err}(u) &\leqslant \sum_{\hat{\theta}\neq \mathbf{s}} \sum_{\tilde{u}\in T_{\hat{\theta}}(u)} \sum_{[x_r]} \sum_{\tilde{x}\in[x_r]} \sum_{y:(\tilde{x},y)\in A_\epsilon} P_{[X_r]}([x_r]) P_{X|[X_r]}(\tilde{x} \mid [x_r]) P_{Y|[X_r]}(y \mid [x_r]) \\
&= \sum_{\hat{\theta}\neq \mathbf{s}} \sum_{\tilde{u}\in T_{\hat{\theta}}(u)} \sum_{[x_r]} \sum_{(\tilde{x},y)\in A_\epsilon} P_{[X_r]}([x_r]) P_{XY|[X_r]}(\tilde{x}, y \mid [x_r]) \\
&= \sum_{(\tilde{x},y)\in A_\epsilon} \sum_{\hat{\theta}\neq \mathbf{s}} \sum_{\tilde{u}\in T_{\hat{\theta}}(u)} P_{XY|\hat{\theta}}(\tilde{x}, y) \\
&= \sum_{(\tilde{x},y)\in A_\epsilon} \sum_{\hat{\theta}\neq \mathbf{s}} \left|T_{\hat{\theta}}(u)\right| P_{XY|\hat{\theta}}(\tilde{x}, y), \qquad (5.31)
\end{aligned}$$

where the first equality holds because $P_{XY|[X_r]}(\tilde{x}, y \mid [x_r])$ vanishes for $\tilde{x} \notin [x_r]$, and the last equality holds because for any x, y, the term $P_{XY|\hat{\theta}}(x, y)$ depends only on $\hat{\theta}$ and is independent of $\tilde{u}$. Finally, note that (5.31) can be further upper bounded as

$$P_{err}(u) \leqslant \sum_{(\tilde{x},y)\in A_\epsilon} \sum_{\hat{\theta}\in\Theta^*} \left|T_{\hat{\theta}}(u)\right| P_{XY|\hat{\theta}}(\tilde{x}, y) = |J| \sum_{(\tilde{x},y)\in A_\epsilon} P_{XY|J}(\tilde{x}, y) = |J|\, P_{XY|J}(A_\epsilon). \quad (5.32)$$

We now provide an upper bound for the probability of error for the coding scheme as follows.

$$\begin{aligned}
\Pr(\text{error}) &= \frac{1}{|J|} \sum_{u\in J} \Pr(E(u)) = \frac{1}{|J|} \sum_{u\in J} \left[\Pr(E_1(u)) + P_{err}(u)\right] \\
&\leqslant \epsilon + |J|\, P_{XY|J}(A_\epsilon) = \epsilon + 2^{R - D_H^\epsilon(P_{XY}\|P_{XY|J})}, \qquad (5.33)
\end{aligned}$$

where the inequality follows from (5.30) and (5.32). The last equality follows from (5.29) and the rate of the code R being $R = \log_2 |J|$, as in equation (5.12). The characterization of error probability for the coding scheme in this section implies the following theorem.

Theorem 43. *Let G be an Abelian group in the form of* (5.16)*. Then there exists an Abelian group J in the form of* (5.17) *and a $(|J|, \epsilon)$-code such that*

$$R \geqslant D_H^{\epsilon'}(P_{XY}\|P_{XY|J}) - \log_2 \frac{1}{\epsilon - \epsilon'},$$

for any $\epsilon' \in (0, \epsilon)$, where the rate $R = \log_2 |J|$.

5.4.2 Converse

According to Lemma 40, for each group code $\mathbb{C} \leqslant G$, there exists a group J and a homomorphism such that $\mathbb{C}$ is the image of the homomorphism. Assume now that a group transmission system with parameters $(1, |J|, \epsilon)$ exists over a channel $(\mathcal{X} = G, \mathcal{Y}, W_{Y|X})$, and that the group J takes the form in equation (5.17). Assume that the homomorphism ϕ for the group code $\mathbb{C}$ is a one-to-one mapping. We can express the code compactly as follows:

$$\mathbb{C} = \left\{ \bigoplus_{(p,r,m)\in\mathcal{G}(G)} \sum_{s=1}^{r_p} u_{p,s} g_{(p,s)\to(r,m)} + V : u_{p,s} \in \mathbb{Z}_{p^s}^{k_{p,s}}, \forall (p,s) \in \mathcal{S}(G) \right\}.$$

The rate of this code is given by $R = \log |J| = k \sum_{(p,s)\in\mathcal{S}(G)} s\omega_{p,s} \log p$.

Let $\hat{\theta}$ be a vector indexed by $(p,s) \in \mathcal{S}(G)$ with $0 \leqslant \hat{\theta}_{p,s} \leqslant s$. For an message $u \in J$, construct a one-to-one correspondence between $u_{p,s} \in \mathbb{Z}_{p^s}^{k_{p,s}}$ and the tuple $(\tilde{u}_{p,s}, \hat{u}_{p,s})$ where $\tilde{u}_{p,s} \in p^{\hat{\theta}_{p,s}} \mathbb{Z}_{p^s}^{k_{p,s}}$ and $\hat{u}_{p,s} \in \mathbb{Z}_{p^{\hat{\theta}_{p,s}}}^{k_{p,s}}$. Let U denote the random message of the group transmission system of the code. Let $\hat{U}$ denote the part of the random message such that $\hat{U}_{p,s} \in \mathbb{Z}_{p^{\hat{\theta}_{p,s}}}^{k_{p,s}}$, for all $(p,s) \in \mathcal{S}(G)$. Consider the subcode of $\mathbb{C}$:

$$\mathbb{C}_1(\hat{\theta}, \hat{u}) = \left\{ \bigoplus_{(p,r,m)\in\mathcal{G}(G)} \sum_{s=1}^{r_p} (\tilde{u}_{p,s} + \hat{u}_{p,s}) g_{(p,s)\to(r,m)} + V : \tilde{u}_{p,s} \in p^{\hat{\theta}_{p,s}} \mathbb{Z}_{p^s}^{k_{p,s}}, \forall (p,s) \in \mathcal{S}(G) \right\}.$$

Let $x = \phi(u) + V$ be the channel input and $H_{\hat{\theta}}$ be given as in (5.13). Then $\mathbb{C}_1(\hat{\theta}, \hat{u}) = [x]_{\hat{\theta}} = x + H_{\hat{\theta}}$. That is, there is a one-to-one correspondence between $\hat{U}$ and $[X]_{\hat{\theta}}$. Also, $|T_{\hat{\theta}}(u)| = \left|\mathbb{C}_1(\hat{\theta}, \hat{u})\right| = |H_{\hat{\theta}}|$ for all $u \in J$.

Define a one-to-one correspondence between x and the tuple $(\tilde{x}_{\hat{\theta}}, [x]_{\hat{\theta}})$ where $\tilde{x}_{\hat{\theta}} = \phi(\tilde{u})$. Consider a genie-aided receiver which gets access to $\hat{U}$ and performs maximum likelihood decoding. Equivalently, this receiver has access to the coset information $[X]_{\hat{\theta}}$ of X and can be written as $\mathcal{D}^{ga} : ([x]_{\hat{\theta}}, y) \mapsto x' \in \mathcal{X}$. Clearly the average probability of error for this decoder must be not greater than ϵ. Let $X' \in \mathcal{X}$ be the output of $\mathcal{D}^{ga}$. For every $\hat{\theta}$ with $0 \leqslant \hat{\theta}_{p,s} \leqslant s$, $\hat{\theta} \neq \mathbf{s}$, the average probability of error for this decoder is

$$\sum_{\hat{u}} \sum_{x,x'} \Pr(\hat{u}) P_{XX'|\hat{U}}(x, x' \mid \hat{u}) \mathbf{1}_{\{x' \neq x\}} = \sum_{x,x'} P_{XX'}(x, x') \mathbf{1}_{\{x' \neq x\}} \leqslant \epsilon,$$

where $P_{XX'|\hat{U}}(x, x'|\hat{u}) \triangleq P_{X|[X_r]}(x|[x]) \sum_{y:\mathcal{D}^{ga}([x],y)=x'} W(y|x)$ and $[x] = [x]_{\hat{\theta}} = x + \mathbb{C}_1(\hat{\theta}, \hat{u})$.

Consider a strategy to distinguish $P_{XX'}$ and $P_{\hat{U}}(P_{X|\hat{U}} \otimes P_{X'|\hat{U}})$ as follows. The strategy guesses $P_{XX'}$ if it sees $X = X'$, and guesses $P_{\hat{U}}(P_{X|\hat{U}} \otimes P_{X'|\hat{U}})$ otherwise. When $P_{XX'}$ is the true underlying distribution, the type-I error probability is exactly the probability that $X \neq X'$ computed from $P_{XX'}$, namely, the average probability of a decoding error, and is thus not larger than ϵ. When $P_{\hat{U}}(P_{X|\hat{U}} \otimes P_{X'|\hat{U}})$ is the true underlying distribution, the probability of type-II error (misdetection) is

$$\begin{aligned}
&\sum_{\hat{u}} P_{\hat{U}}(\hat{u}) \sum_{x,x'} P_{X|\hat{U}}(x \mid \hat{u}) P_{X'|\hat{U}}(x' \mid \hat{u}) \mathbf{1}_{\{x'=x\}} \\
&= \sum_{[x_r]} P_{[X_r]}([x_r]) \sum_{x,x'} P_{X|[X_r]}(x \mid [x_r]) P_{X'|[X_r]}(x' \mid [x_r]) \mathbf{1}_{\{x'=x\}} \qquad (5.34) \\
&= \sum_{[x_r]} \frac{|H|}{|G|} \sum_{x} P_{X|[X_r]}(x \mid [x_r]) P_{X'|[X_r]}(x \mid [x_r]) \\
&= \sum_{[x_r]} \frac{|H|}{|G|} \sum_{x \in [x_r]} \frac{1}{|H|} P_{X'|[X_r]}(x \mid [x_r]) = \frac{1}{|H|},
\end{aligned}$$

where we wrote H for $H_{\hat{\theta}}$, and (5.34) follows from the one-to-one mapping between $\hat{U}$ and $[X]_{\hat{\theta}}$. Thus,

$$\begin{aligned}
D_H^{\epsilon,\hat{\theta}}(P_{XY} \| P_{[X]} P_{X|[X]} P_{Y|[X]}) &\geqslant D_H^{\epsilon,\hat{\theta}}(P_{XX'} \| P_{[X]} P_{X|[X]} P_{X'|[X]}) = D_H^{\epsilon}(P_{XX'} \| P_{\hat{U}}(P_{X|\hat{U}} \otimes P_{X'|\hat{U}})) \\
&\geqslant -\log_2 \frac{1}{|H|} = \log_2 |H| = \log_2 |T_{\hat{\theta}}(u)| \\
&= (1-\omega_{\hat{\theta}}) k \sum_{(p,s) \in \mathcal{S}(G)} s \omega_{p,s} \log p., \qquad (5.35)
\end{aligned}$$

where the first inequality follows from the DPI and the last equality is shown in equation (5.15) in Section 5.3.3. Equivalently, we may rewrite (5.35) compactly as $D_H^{\epsilon,\hat{\theta}}(P_{XY} \| P_{[X]} P_{X|[X]} P_{Y|[X]}) \geqslant (1-\omega_{\hat{\theta}})R$. Thus we have the following theorem

Theorem 44. *Assume that a group transmission system with parameters* $(1, |J|, \epsilon)$ *exists over a channel* $(\mathcal{X} = G, \mathcal{Y}, W_{Y|X})$*, and that the group* J *takes the form as in equation* (5.17)*. Then the rate of the code,* $R = \log |J|$*, is bounded as:*

$$R \leqslant \min_{\hat{\theta} \neq \mathbf{s}} \frac{1}{1-\omega_{\hat{\theta}}} D_H^{\epsilon,\hat{\theta}}(P_{XY} \| P_{[X]} P_{X|[X]} P_{Y|[X]}).$$

Example 3. Consider the sets $J = \mathbb{Z}_4, G = \mathbb{Z}_8$ as in Example 2. We have $\mathcal{G}(G) = \{(2,3,1)\}$ and $\mathcal{G}(J) = \{(2,2,1)\}$, $\mathcal{S}(G) = \{(2,1),(2,2),(2,3)\}$, and $k_{2,1} = 0, k_{2,2} = 1, k_{2,3} = 0$. The vector $\hat{\theta}$ with $0 \leqslant \hat{\theta}_{p,s} \leqslant s$, $\hat{\theta} \neq \mathbf{s}$ must be $\hat{\theta} \in \{(1,0,3),(1,1,3)\}$.

Case 1: $\hat{\theta}_{2,2} = 0$, $\boldsymbol{\theta}(\hat{\theta})_{(2,3)} = 1$, $H_{\hat{\theta}} = 2\mathbb{Z}_8$.

In this case, $\tilde{u}_{2,2} \in \mathbb{Z}_4$, $\hat{u}_{2,2} \in \mathbb{Z}_1 \equiv \{0\}$, and $\mathbb{C}_1(\hat{\theta} = (1,0,3), \hat{u} = 0) = \{\tilde{u}_{2,2} g + V : \tilde{u}_{2,2} \in \mathbb{Z}_4\} = \mathbb{C}$. Thus, equation (5.35) yields

$$D_H^{\epsilon}(P_{XY} \| P_{[X]} P_{X|[X]} P_{Y|[X]}) = D_H^{\epsilon}(P_{XY} \| P_X P_Y) \geqslant \log_2 |H| = 2 = R,$$

where the first equality holds because $[X] = X + H_{\hat{\theta}} = \mathbb{C}_1(\hat{\theta} = (1,0,3), \hat{u} = 0) = \mathbb{C}$.

Case 2: $\hat{\theta}_{2,2} = 1$, $\boldsymbol{\theta}(\hat{\theta})_{(2,3)} = 2$, $H_{\hat{\theta}} = 4\mathbb{Z}_8$

In this case, $\tilde{u}_{2,2} \in 2\mathbb{Z}_4$, $\hat{u}_{2,2} \in \mathbb{Z}_2$, and $\mathbb{C}_1(\hat{\theta} = (1,1,3), \hat{u} = 0) = \{\tilde{u}_{2,2}g + V : \tilde{u}_{2,2} \in 2\mathbb{Z}_4\} = \{V, 2g + V\}$ and $\mathbb{C}_1(\hat{\theta} = (1,1,3), \hat{u} = 1) = \{(\tilde{u}_{2,2} + 1)g + V : \tilde{u}_{2,2} \in 2\mathbb{Z}_4\} = \{g + V, 3g + V\}$. Equation (5.35) yields

$$D_H^{\epsilon}(P_{XY} \| P_{[X]} P_{X|[X]} P_{Y|[X]}) \geqslant \log_2 |H| = 1 = R/2,$$

5.5 One-shot Classical-Quantum Group Coding

The one-shot coding problem for the transmission over a classical-quantum channel using group codes is studied in this section. The technique adopted in the achievability part is similar to the second approach in Section 5.4.1. The converse part is proved by considering subcodes of a given code $\mathbb{C}$, and the main result is given in a similar manner to that of the classical channel given in Section 5.4.2.

We recall a few properties of the hypothesis testing relative entropy $D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma)$, defined in Section 5.2, from [142].

Lemma 45. *$D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma)$ has the following properties, all of which hold for all ρ, σ and $\epsilon \in [0,1)$:*

- Positivity: $D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma) \geqslant 0$, *with equality if $\rho = \sigma$ and $\epsilon = 0$.*
- Data Processing Inequality (DPI): *for any Completely Positive Map $\mathcal{E}$, $D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma) \geqslant D_{\mathrm{H}}^{\epsilon}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma))$.*
- *Let $D(\cdot\|\cdot)$ denote the usual quantum relative entropy, then $D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma) \leqslant (D(\rho\|\sigma) + H_{\mathrm{b}}(\epsilon))/(1-\epsilon)$.*

The following lemma by Hayashi and Nagaoka [62, Lemma 2] will serve, roughly speaking, as the quantum analog to the equation $\Pr(E(u)) \leqslant \Pr(E_1(u)) + \sum_{\tilde{u} \neq u} P_{err}(u, \tilde{u})$ in Section 5.4.1.

Lemma 46. *For any positive real c and any operators $0 \leqslant S \leqslant I$ and $T \geqslant 0$, we have*

$$I - (S+T)^{-1/2} S (S+T)^{-1/2} \leqslant (1+c)(I-S) + (2+c+c^{-1})T.$$

5.5.1 Achievability

Consider Abelian groups G and J as in Section 5.3.1. Given a classical-quantum channel $\mathcal{N} = \{\rho_x\}_{x\in\mathcal{X}}$ from the classical alphabet $\mathcal{X}$ to the quantum system $\mathbb{B}$, where $\mathcal{X} = G$ is an

Abelian group. Let u, x and ρ_x be the message, the channel input and the channel output respectively. Consider the group code $\mathbb{C} = \{\phi(a) + V | a \in J\}$, where the distributions of ϕ and V are specified in Definition 5. Let π^{AB} denote the joint state of the input and output for an input chosen according to the uniform distribution $P_X \equiv \text{Unif}(\mathcal{X})$, i.e.,

$$\pi^{AB} \triangleq \sum_{x\in\mathcal{X}} P_X(x)|x\rangle\langle x|^A \otimes \rho_x^B \ ,$$

for any representation of the inputs x in terms of orthonormal vectors $|x\rangle^A$ on a Hilbert space $\mathbb{A}$, and where π^A and π^B are the corresponding marginals. Let $\hat{\theta}$ be a vector indexed by $(p,s) \in \mathcal{S}(G)$ with $0 \leqslant \hat{\theta}_{p,s} \leqslant s$. Define the joint state $\pi_{\hat{\theta}}^{AB}$ by

$$\begin{aligned}
\pi_{\hat{\theta}}^{AB} &\triangleq \sum_{x\in\mathcal{X}} P_X(x) \sum_{\tilde{x}\in x+H_{\hat{\theta}}} P_{X|[X_r]}(\tilde{x}|x+H_{\hat{\theta}})|\tilde{x}\rangle\langle\tilde{x}|^A \otimes \rho_x^B \\
&= \sum_{[x_r]} P_{[X_r]}([x_r]) \sum_{x\in[x_r]} \sum_{\tilde{x}\in[x_r]} P_{X|[X_r]}(x|[x_r]) P_{X|[X_r]}(\tilde{x}|[x_r])|\tilde{x}\rangle\langle\tilde{x}|^A \otimes \rho_x^B \\
&= \sum_{[x_r]} P_{[X_r]}([x_r]) \sum_{\tilde{x}\in[x_r]} P_{X|[X_r]}(x|[x_r])|\tilde{x}\rangle\langle\tilde{x}|^A \otimes \rho_{[x_r]}^B \ ,
\end{aligned}$$

where $[x_r] = [x_r]_{\hat{\theta}}$ is the coset $x_r + H_{\hat{\theta}}$. Let Θ be the set of vectors $\hat{\theta}$ indexed by $(p,s) \in \mathcal{S}(G)$ such that $0 \leqslant \hat{\theta}_{p,s} \leqslant s$ and $\hat{\theta} \neq \mathbf{s}$, and denote its size by $M \triangleq |\Theta|$. We may also index the vectors in the set, so that $\Theta = \left\{\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_M\right\}$. In addition, denote by $\hat{\theta}_0 = \mathbf{s}$, and denote by Θ^* the union $\Theta \cup \left\{\hat{\theta}_0\right\}$. Define a joint state π_J^{AB} by

$$\pi_J^{AB} \triangleq |J|^{-1} \sum_{\hat{\theta}\in\Theta^*} |T_{\hat{\theta}}|\, \pi_{\hat{\theta}}^{AB} = \sum_{\hat{\theta}\in\Theta^*} 2^{\left(1-\omega_{\hat{\theta}}\right)} \pi_{\hat{\theta}}^{AB}. \tag{5.36}$$

The last equality holds by noting that $\log|T_{\hat{\theta}}(a)| = (1-\omega_{\hat{\theta}})R$ according to (5.15), and that the rate of the code is $R = \log|J|$. One may also plug in the definition of $\pi_{\hat{\theta}}^{AB}$ and show that

$$\begin{aligned}
\pi_J^{AB} &= |J|^{-1} \sum_{\hat{\theta}\in\Theta^*} |T_{\hat{\theta}}| \sum_{x\in\mathcal{X}} P_X(x) \sum_{\tilde{x}\in x+H_{\hat{\theta}}} P_{X|[X_r]}(\tilde{x}|x+H_{\hat{\theta}})|\tilde{x}\rangle\langle\tilde{x}|^A \otimes \rho_x^B \\
&= |J|^{-1} \sum_{\hat{\theta}} \sum_{\tilde{u}\in T_{\hat{\theta}}(u)} \sum_{x,\tilde{x}} \Pr\left(\phi(u)+V = x, \phi(\tilde{u})+V = \tilde{x}\right)|\tilde{x}\rangle\langle\tilde{x}|^A \otimes \rho_x^B,
\end{aligned} \tag{5.37}$$

where u is an arbitrary element of J.

Let Q^* be a positive operator that achieves $D_{\mathrm{H}}^{\epsilon'}(\pi^{AB}\|\pi_J^{AB})$. That is, $\mathrm{tr}[Q^*\pi^{AB}] \geqslant 1-\epsilon'$

and

$$D_{\mathrm{H}}^{\epsilon'}(\pi^{AB}\|\pi_J^{AB}) = -\log_2 \mathrm{tr}[Q^*\pi_J^{AB}].$$

We define the decoding POVM by its elements,

$$E_u = \left(\sum_{u'\in J} A_{x(u')}\right)^{-\frac{1}{2}} A_{x(u)} \left(\sum_{u'\in J} A_{x(u')}\right)^{-\frac{1}{2}},$$

where $x(u) \triangleq \phi(u) + V$ and $A_x \triangleq \mathrm{tr}_A\left[\left(|x\rangle\langle x|^A \otimes I^B\right) Q^*\right]$. For any state ρ^B on the Hilbert space $\mathbb{B}$, we have

$$\mathrm{tr}\left\{A_x\rho^B\right\} = \mathrm{tr}\left\{\left(|x\rangle\langle x|^A \otimes \rho^B\right) Q^*\right\}. \tag{5.38}$$

For a specific choice of ϕ and V and a message $u \in J$, the probability of error is given by

$$\Pr(\mathrm{error}|u,\phi,V) = \mathrm{tr}[(I - E_u)\rho_{x(u)}].$$

Applying Lemma 46 with $S = A_{x(u)}$ and $T = \sum_{u'\neq u} A_{x(u')}$, we may bound this term by

$$\Pr(\mathrm{error}|u,\phi,V) \leqslant (1+c)\left(1 - \mathrm{tr}[A_{x(u)}\rho_{x(u)}]\right) + (2+c+c^{-1})\sum_{u'\neq u}\mathrm{tr}[A_{x(u')}\rho_{x(u)}].$$

Taking expectation over all choices of ϕ, V, but keeping the transmitted message u fixed, we find

$$\Pr(\mathrm{error}|u) \leqslant (1+c)\left[1 - \mathbb{E}_{\phi,V}\ \mathrm{tr}[A_{X(u)}\rho_{X(u)}]\right] + (2+c+c^{-1})\sum_{u'\neq u}\mathbb{E}_{\phi,V}\mathrm{tr}[A_{X(u')}\rho_{X(u)}]. \tag{5.39}$$

Rewriting the first expectation via (5.38) and the linearity of the trace operator,

$$\begin{aligned}\mathbb{E}_{\phi,V}\mathrm{tr}[A_{X(u)}\rho_{X(u)}] &= \mathbb{E}_{\phi,V}\mathrm{tr}\left[\left(|X(u)\rangle\langle X(u)|^A \otimes \rho^B_{X(u)}\right) Q^*\right]\\ &= \mathrm{tr}\left[\mathbb{E}_{\phi,V}\left(|X(u)\rangle\langle X(u)|^A \otimes \rho^B_{X(u)}\right) Q^*\right]\\ &= \mathrm{tr}\left[Q^*\pi^{AB}\right] \geqslant 1-\epsilon'.\end{aligned}$$

Hence we may bound the first term by $(1+c)\left[1 - \mathbb{E}_{\phi,V}\ \mathrm{tr}[A_{X(u)}\rho_{X(u)}]\right] \leqslant (1+c)\epsilon'$.

For the sum in the second term of (5.39), we adopt an approach similar to that in

Section 5.4.1.

$$\begin{aligned}
\sum_{u'\neq u} \mathbb{E}_{\phi,V} \mathrm{tr}[A_{X(u')}\rho_{X(u)}] &= \sum_{\hat{\theta}\neq \mathbf{s}} \sum_{\tilde{u}\in T_{\hat{\theta}}(u)} \sum_{x,\tilde{x}} \Pr\left(\phi(u)+V=x, \phi(\tilde{u})+V=\tilde{x}\right) \mathrm{tr}[A_{\tilde{x}}\rho_x] \\
&= \sum_{\hat{\theta}\neq \mathbf{s}} \sum_{\tilde{u}\in T_{\hat{\theta}}(u)} \sum_{[x_r]} \sum_{x\in[x_r]} \sum_{\tilde{x}\in[x_r]} \Pr\left(\phi(u)+V=x, \phi(\tilde{u})+V=\tilde{x}\right) \mathrm{tr}[A_{\tilde{x}}\rho_x] \\
&= \sum_{\hat{\theta}\neq \mathbf{s}} P_{err}(u,\hat{\theta}), \qquad (5.40)
\end{aligned}$$

where $P_{err}(u,\hat{\theta}) \triangleq \sum_{\tilde{u}\in T_{\hat{\theta}}(u)} P_{err}(u,\tilde{u})$ and, for $\tilde{u}\in T_{\hat{\theta}}(u)$,

$$\begin{aligned}
P_{err}(u,\tilde{u}) &\triangleq \sum_{[x_r]} \sum_{x\in[x_r]} \sum_{\tilde{x}\in[x_r]} \Pr\left(\phi(u)+V=x, \phi(\tilde{u})+V=\tilde{x}\right) \mathrm{tr}\left[A_{\tilde{x}}\rho_x\right] \\
&= \sum_{[x_r]} P_{[X_r]}([x_r]) \sum_{x\in[x_r]} \sum_{\tilde{x}\in[x_r]} P_{X|[X_r]}(x|[x_r]) P_{X|[X_r]}(\tilde{x}|[x_r]) \mathrm{tr}\left[A_{\tilde{x}}\rho_x\right] \\
&= \sum_{[x_r]} P_{[X_r]}([x_r]) \sum_{\tilde{x}\in[x_r]} P_{X|[X_r]}(\tilde{x}|[x_r]) \mathrm{tr}\left[A_{\tilde{x}}\rho_{[x_r]}\right]. \qquad (5.41)
\end{aligned}$$

Leveraging equation (5.38), the above equation can be written as

$$P_{err}(u,\tilde{u}) = \sum_{[x_r]} P_{[X_r]}([x_r]) \sum_{\tilde{x}\in[x_r]} P_{X|[X_r]}(\tilde{x}|[x_r]) \mathrm{tr}\left[\left(|\tilde{x}\rangle\langle\tilde{x}|^A \otimes \rho_{[x_r]}\right) Q^*\right] = \mathrm{tr}[Q^*\pi_{\hat{\theta}}^{AB}].$$

Therefore the sum in (5.40) is given by

$$\begin{aligned}
\sum_{u'\neq u} \mathbb{E}_{\phi,V} \mathrm{tr}[A_{X(u')}\rho_{X(u)}] &= \sum_{\hat{\theta}\neq \mathbf{s}} |T_{\hat{\theta}}(u)| \, \mathrm{tr}[Q^*\pi_{\hat{\theta}}^{AB}] \leqslant \sum_{\hat{\theta}\in\Theta^*} |T_{\hat{\theta}}(u)| \, \mathrm{tr}[Q^*\pi_{\hat{\theta}}^{AB}] \\
&= \mathrm{tr}\left[Q^*\left(\sum_{\hat{\theta}\in\Theta^*} |T_{\hat{\theta}}(u)| \, \pi_{\hat{\theta}}^{AB}\right)\right] = \mathrm{tr}\left[Q^*\left(\sum_{\hat{\theta}\in\Theta^*} 2^{(1-\omega_{\hat{\theta}})} 2^R \pi_{\hat{\theta}}^{AB}\right)\right] \\
&= 2^R \, \mathrm{tr}[Q^*\pi_J^{AB}] = 2^{R - D_{\mathrm{H}}^{\epsilon'}(\pi^{AB}\|\pi_J^{AB})}.
\end{aligned}$$

The bound in (5.39) is thus

$$\Pr(\text{error}|u) \leqslant (1+c)\epsilon' + (2+c+c^{-1}) 2^{R - D_{\mathrm{H}}^{\epsilon'}(\pi^{AB}\|\pi_J^{AB})},$$

which is independent of u. Therefore we have demonstrated the existence of a $(2^R,\epsilon)$-code with

$$R \geqslant D_{\mathrm{H}}^{\epsilon'}(\pi^{AB}\|\pi_J^{AB}) - \log_2 \frac{2+c+c^{-1}}{\epsilon-(1+c)\epsilon'}\ . \tag{5.42}$$

Optimized over c, this bound implies the following theorem

Theorem 47. *Given a classical-quantum channel $\mathcal{N} = \{\rho_x\}_{x\in\mathcal{X}}$, where $\mathcal{X} = G$, an $(2^R, \epsilon)$ coding scheme exists for $\mathcal{N}$ with*

$$R \geqslant D_{\mathrm{H}}^{\epsilon'}(\pi^{AB}\|\pi_J^{AB}) - \log_2 \frac{4\epsilon}{(\epsilon-\epsilon')^2}$$

for any $\epsilon' \in (0,\epsilon)$.

5.5.2 Converse

Based on Lemma 40, for each group code $\mathbb{C} \leqslant G$, there exists a group J and a homomorphism such that $\mathbb{C}$ is the image of the homomorphism. Assume now that a group transmission system with parameters $(1, |J|, \epsilon)$ exists over a classical-quantum channel $\mathcal{N} = \left\{\rho_x^B\right\}_{x\in\mathcal{X}}$, where $\mathcal{X} = G$ is an Abelian group, and that the group J takes the form in equation (5.17), i.e., $J = \bigoplus_{p\in\mathcal{P}(G)} \bigoplus_{s=1}^{r_p} \mathbb{Z}_{p^s}^{k_{p,s}}$. Assume that the homomorphism ϕ for the group code $\mathbb{C}$ is a one-to-one mapping. We can express the code complactly as follows:

$$\mathbb{C} = \left\{ \bigoplus_{(p,r,m)\in\mathcal{G}(G)} \sum_{s=1}^{r_p} u_{p,s} g_{(p,s)\to(r,m)} + V : u_{p,s} \in \mathbb{Z}_{p^s}^{k_{p,s}}, \forall (p,s) \in \mathcal{S}(G) \right\}.$$

The rate of this code is given by $R = \log|J| = k\sum_{(p,s)\in\mathcal{S}(G)} s\omega_{p,s}\log p$.

Let $\hat{\theta}$ be a vector indexed by $(p,s) \in \mathcal{S}(G)$ with $0 \leqslant \hat{\theta}_{p,s} \leqslant s$. For an message $u \in J$, construct a one-to-one correspondence between $u_{p,s} \in \mathbb{Z}_{p^s}^{k_{p,s}}$ and the tuple $(\tilde{u}_{p,s}, \hat{u}_{p,s})$ where $\tilde{u}_{p,s} \in p^{\hat{\theta}_{p,s}} \mathbb{Z}_{p^s}^{k_{p,s}}$ and $\hat{u}_{p,s} \in \mathbb{Z}_{p^{\hat{\theta}_{p,s}}}^{k_{p,s}}$. Consider the subcode of $\mathbb{C}$:

$$\mathbb{C}_1(\hat{\theta}, \hat{u}) = \left\{ \bigoplus_{(p,r,m)\in\mathcal{G}(G)} \sum_{s=1}^{r_p} (\tilde{u}_{p,s} + \hat{u}_{p,s}) g_{(p,s)\to(r,m)} + V : \tilde{u}_{p,s} \in p^{\hat{\theta}_{p,s}} \mathbb{Z}_{p^s}^{k_{p,s}}, \forall (p,s) \in \mathcal{S}(G) \right\}.$$

Let $x = \phi(u) + V$ be the channel input and $H_{\hat{\theta}}$ be given as in (5.13). Then $\mathbb{C}_1(\hat{\theta}, \hat{u}) = [x]_{\hat{\theta}} = x + H_{\hat{\theta}}$. That is, there is an one-to-one correspondence between $\hat{U}$ and $[X]_{\hat{\theta}}$. Also, $|T_{\hat{\theta}}(u)| = \left|\mathbb{C}_1(\hat{\theta}, \hat{u})\right| = |H_{\hat{\theta}}|$ for all $u \in J$.

Define a one-to-one correspondence between x and the tuple $(\tilde{x}_{\hat{\theta}}, [x]_{\hat{\theta}})$ where $\tilde{x}_{\hat{\theta}} = \phi(\tilde{u})$.

Consider a genie-aided receiver which gets access to $\hat{U}$ and denote it by $\mathcal{D}^{ga}$,. Equivalently, this receiver has access to the coset information $[X]_{\hat{\theta}}$ of X and can be realized by a family of POVMs $\left\{E_x^{[x]}\right\}$. Clearly the average probability of error for this decoder must be not greater than ϵ. Let $X' \in \mathcal{X}$ be the output of $\mathcal{D}^{ga}$. For every $\hat{\theta}$ with $0 \leqslant \hat{\theta}_{p,s} \leqslant s$, $\hat{\theta} \neq \mathbf{s}$, the average probability of error for this decoder is

$$\sum_{\hat{u}} \sum_{x,x'} \Pr(\hat{u}) P_{XX'|\hat{U}}(x, x' \mid \hat{u}) \mathbf{1}_{\{x' \neq x\}} = \sum_{x,x'} P_{XX'}(x, x') \mathbf{1}_{\{x' \neq x\}} \leqslant \epsilon,$$

where $P_{XX'|\hat{U}}(x, x'|\hat{u}) \triangleq P_{X|[X_r]}(x|[x]) \mathrm{tr}\left[E_{x'}^{[x]} \rho_x\right]$ and $[x] = [x]_{\hat{\theta}} = x + \mathcal{C}_1(\hat{\theta}, \hat{u})$.

Note that the decoding POVM can be viewed as a CPM. This CPM maps π^{AB} to the (classical) state $P_{XX'}$ denoting the joint distribution of the transmitted codeword X and the decoder's guess X'. Similarly, it maps $\pi_{\hat{\theta}}^{AB}$ to $P_{\hat{U}}(P_{X|\hat{U}} \otimes P_{X'|\hat{U}})$. Hence, it follows from the DPI for $D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma)$ that

$$D_{\mathrm{H}}^{\epsilon}(P_{XX'} \| P_{\hat{U}}(P_{X|\hat{U}} \otimes P_{X'|\hat{U}})) \leqslant D_{\mathrm{H}}^{\epsilon}(\pi^{AB} \| \pi_{\hat{\theta}}^{AB}) \ .$$

Consider a strategy to distinguish $P_{XX'}$ and $P_{\hat{U}}(P_{X|\hat{U}} \otimes P_{X'|\hat{U}})$ as follows. The strategy guesses $P_{XX'}$ if it sees $X = X'$, and guesses $P_{\hat{U}}(P_{X|\hat{U}} \otimes P_{X'|\hat{U}})$ otherwise. When $P_{XX'}$ is the true underlying distribution, the type-I error probability is exactly the probability that $X \neq X'$ computed from $P_{XX'}$, namely, the average probability of a decoding error, and is thus not larger than ϵ. When $P_{\hat{U}}(P_{X|\hat{U}} \otimes P_{X'|\hat{U}})$ is the true underlying distribution, the probability of type-II error (misdetection) is

$$\begin{aligned}
&\sum_{\hat{u}} P_{\hat{U}}(\hat{u}) \sum_{x,x'} P_{X|\hat{U}}(x \mid \hat{u}) P_{X'|\hat{U}}(x' \mid \hat{u}) \mathbf{1}_{\{x'=x\}} \\
&= \sum_{[x_r]} P_{[X_r]}([x_r]) \sum_{x,x'} P_{X|[X_r]}(x \mid [x_r]) P_{X'|[X_r]}(x' \mid [x_r]) \mathbf{1}_{\{x'=x\}} \qquad (5.43) \\
&= \sum_{[x_r]} \frac{|H|}{|G|} \sum_{x} P_{X|[X_r]}(x \mid [x_r]) P_{X'|[X_r]}(x \mid [x_r]) \\
&= \sum_{[x_r]} \frac{|H|}{|G|} \sum_{x \in [x_r]} \frac{1}{|H|} P_{X'|[X_r]}(x \mid [x_r]) = \frac{1}{|H|},
\end{aligned}$$

where we wrote H for $H_{\hat{\theta}}$, and (5.43) follows from the one-to-one mapping between $\hat{U}$ and

$[X]_{\hat{\theta}}$. That is,

$$\begin{aligned} D_{\mathrm{H}}^{\epsilon}(\pi^{AB}\|\pi^{AB}_{\hat{\theta}}) &\geqslant D_{\mathrm{H}}^{\epsilon}(P_{XX'}\|P_{\hat{U}}(P_{X|\hat{U}}\otimes P_{X'|\hat{U}})) = \log_2|H| = \log_2|T_{\hat{\theta}}(u)| \\ &= (1-\omega_{\hat{\theta}})k\sum_{(p,s)\in\mathcal{S}(G)} s\omega_{p,s}\log p., \end{aligned} \tag{5.44}$$

where the last equality is shown in equation (5.15) in Section 5.3.3. Equivalently, we may rewrite (5.44) compactly as $D_{\mathrm{H}}^{\epsilon}(\pi^{AB}\|\pi^{AB}_{\hat{\theta}}) \geqslant (1-\omega_{\hat{\theta}})R$. Thus, we have the following theorem

Theorem 48. *Assume that a group transmission system with parameters* $(1,|J|,\epsilon)$ *exists over a classical-quatum channel* $\mathcal{N}=\left\{\rho_x^B\right\}_{x\in\mathcal{X}}$, *and that the group* J *takes the form as in equation* (5.17)*. Then the rate of the code,* $R=\log|J|$*, is bounded as:*

$$R\leqslant \min_{\hat{\theta}\neq\mathbf{s}}\frac{1}{1-\omega_{\hat{\theta}}}D_{\mathrm{H}}^{\epsilon}(\pi^{AB}\|\pi^{AB}_{\hat{\theta}}).$$

APPENDIX A

Supplementary material for Chapter 3

A.1 Proof for Section 3.3

Proof of Proposition 11: Since G_l is a polarization kernel, there is at least one column in G_l with weight at least 2. To see this, note that G_l being invertible implies that all rows and columns are nonzero vectors. If all the columns of G_l have weights equal to 1, then all the rows must also have weights equal to 1, i.e., G_l is a permutation matrix. Then $D_i = 1, \forall i$, and $E(G_l) = 0$, which implies that G_l can not be polarization kernel. The contradiction shows that at least one column in G_l have a weight at least 2.

Let $k \geqslant 1$ denote the number of columns in G_l with a weight at least 2. Let $\mathbf{v}$ be a randomly uniformly chosen column of $G_l^{\otimes n}$, and $w(\mathbf{v})$ be the Hamming weight of $\mathbf{v}$. For $\frac{1}{l} > r > 0$,

$$\begin{aligned}&\Pr\left(w(\mathbf{v}) = \mathcal{O}(N^{r\log_l 2})\right)\\ &\leqslant \Pr\left(2^{\sum_{i=1}^n F_i} = \mathcal{O}(N^{r\log_l 2}) = \mathcal{O}(2^{nr})\right),\end{aligned}$$

where F_i is the indicator variable that one of k non-unit-weight columns is used in the i-th Kronecker product of G_l to form $\mathbf{v}$. The variables $F_1, F_2, \ldots, F_n$ are i.i.d. as $\mathrm{Ber}(k/l)$. The law of large numbers implies that $\sum_{i=1}^n F_i \geqslant \frac{kn}{l} > nr$ with high probability. Thus, $\Pr\left(2^{\sum_{i=1}^n F_i} = \mathcal{O}(2^{nr})\right) \to 0$ as $n \to \infty$ for any $r > 0$. ■

A.2 Proofs for Section 3.4

A.2.1 Proof of Lemma 12

In order to understand the DRS algorithm's effect on $G = G_2^{\otimes n}$, we first study how the order of two special Kronecker product operations affects the number of output vectors. We present the following Lemmas 49 and 50 toward the proof of Lemma 12.

Lemma 49. *Let a column vector v and a column weight threshold $w_{u.b.}$ be given. Then the outputs of the DRS algorithm for $(v \otimes [1,1]^T) \otimes [0,1]^T$ and $(v \otimes [0,1]^T) \otimes [1,1]^T$ contain the same number of vectors.*

Proof: The input vectors can be denoted by:

$$(v \otimes [1,1]^T) \otimes [0,1]^T \equiv v_{LR}$$
$$(v \otimes [0,1]^T) \otimes [1,1]^T \equiv v_{RL}.$$

We note that $2w_H(v) = w_H(v_{LR}) = w_H(v_{RL})$, and prove the lemma in two cases:

1. $2w_H(v) \leqslant w_{u.b.}$: In this case, the algorithm will not split either v_{LR} or v_{RL}. Both outputs contain exactly one vector.

2. $2w_H(v) > w_{u.b.}$: Let $n_{DRS}(v)$ denote the size, or more precisely, the number of column vectors the DRS algorithm returns when it is applied to v.

   For v_{LR}, the DRS algorithm observes $\mathbf{x}_h = \mathbf{0}$, hence the number of output vectors is the same as the size of DRS-SPLIT$(w_{u.b.}, (v^T, v^T)^T)$ (see Section 3.4.1). With $2w_H(v) > w_{u.b.}$, the size of DRS-SPLIT$((v^T, v^T)^T)$ is the sum of the sizes of $Y_h =$ DRS-SPLIT$(w_{u.b.}, \mathbf{x}_h = v)$ and $Y_t =$ DRS-SPLIT$(w_{u.b.}, \mathbf{x}_t = v)$. By assumption, $|Y_h| = |Y_t| = n_{DRS}(v)$, giving $n_{DRS}(v_{LR}) = 2\, n_{DRS}(v)$.

   For v_{RL}, the number of vectors in the DRS algorithm output is the sum of the sizes of two sets $Y_h =$ DRS-SPLIT$(w_{u.b.}, \mathbf{x}_h = (\mathbf{0}^T, v^T)^T)$ and $Y_t =$ DRS-SPLIT$(w_{u.b.}, \mathbf{x}_t = (\mathbf{0}^T, v^T)^T)$. It is easy to see that $|Y_h| = |Y_t| = |$DRS-SPLIT$(w_{u.b.}, v)|$, which equals $n_{DRS}(v)$. Thus, $n_{DRS}(v_{RL}) = 2\, n_{DRS}(v)$.

■

The next lemma shows the effect of the DRS algorithm from a different perspective. If there are two vectors with the same column weights, and numbers of vectors of the DRS algorithm outputs are identical when they are the inputs, the properties will be preserved when they undergo some basic Kronecker product operations.

Lemma 50. *Let u_1 and u_2 be two vectors with equal Hamming weights. Assume, for a given $w_{u.b.}$, the DRS algorithm splits u_1 and u_2 into the same number of vectors. Then the DRS algorithm also returns the same number of vectors for $u_1 \otimes [1,1]^T$ and $u_2 \otimes [1,1]^T$, as well as for $u_1 \otimes [0,1]^T$ and $u_2 \otimes [0,1]^T$.*

Proof: We first discuss the case when $u_1 \otimes [1,1]^T$ and $u_2 \otimes [1,1]^T$ are processed by the DRS algorithm. If $2w_H(u_1) = 2w_H(u_2) \leqslant w_{u.b.}$, no splitting is done. If $2w_H(u_1) =$

$2w_H(u_2) > w_{u.b.}$, the size of the DRS algorithm output for the input $u_1 \otimes [1,1]^T$ is the sum of the sizes of $Y_h = \text{DRS-S}\textsc{plit}(w_{u.b.}, \mathbf{x}_h = u_1)$ and $Y_t = \text{DRS-S}\textsc{plit}(w_{u.b.}, \mathbf{x}_t = u_1)$, both of which are $n_{DRS}(u_1)$. The size of the output for the input $u_2 \otimes [1,1]^T$ can be found in a similar way to be $2n_{DRS}(u_2)$. Note that $n_{DRS}(u_1) = n_{DRS}(u_2)$ by assumption. Therefore, the sizes of the outputs of the DRS algorithm, when $u_1 \otimes [1,1]^T$ and $u_2 \otimes [1,1]^T$ are the inputs, are equal. Similarly, one can easily show that when $u_1 \otimes [0,1]^T$ and $u_2 \otimes [0,1]^T$ are processed by the DRS algorithm, the number of output columns are equal. ■

Proof of Lemma 12: Suppose that there is an index i such that $(s_i, s_{i+1}) = (+,-)$. Let $v^{(i+1)}$ and $(v^{(i+1)})'$ be defined by (3.1) with sequences $(s_1, \ldots, s_{i-1}, s_i = +, s_{i+1} = -)$ and $(s_1, \ldots, s_{i-1}, s_i' = -, s_{i+1}' = +)$, respectively. We note that

$$\begin{aligned} v^{(i+1)} &= \left(v^{(i-1)} \otimes [0,1]^T\right) \otimes [1,1]^T \\ (v^{(i+1)})' &= \left(v^{(i-1)} \otimes [1,1]^T\right) \otimes [0,1]^T. \end{aligned}$$

Lemma 49 shows that the DRS algorithm splits $v^{(i+1)}$ and $(v^{(i+1)})'$ into the same number of columns. Furthermore, Lemma 50 shows that the number of output vectors of the DRS algorithm for $v^{(n)} = [v^{(i+1)}]^{(s_{i+2},\ldots,s_n)}$ and $(v^{(n)})' = [(v^{(i+1)})']^{(s_{i+2},\ldots,s_n)}$ are equal.

Therefore, an occurrence of $(s_i, s_{i+1}) = (+,-)$ in a sequence can be replaced by $(s_i, s_{i+1}) = (-,+)$ without changing the number of output vectors of the DRS algorithm. Since any sequence $(s_1, s_2, \ldots, s_n)$ with n_1 minus signs and n_2 plus signs can be permuted into $(s_1', s_2', \ldots s_n')$, where $s_i' = -$ for $i \leqslant n_1$ and $s_i' = +$ for $i > n_1$, by repeatedly replacing any occurrence of $(+,-)$ by $(-,+)$, the above arguments show $n_{DRS}(v^{(n)}) = n_{DRS}(v^{(s_1', s_2', \ldots s_n')})$ always holds. Hence, the size of DRS algorithm output for $v^{(n)}$ depends only on the values n_1 and n_2. ■

A.2.2 Proof for Proposition 13

First note that there is a bijection between $\{-,+\}^n$ and the columns of $G_2^{\otimes n}$ as follows. For each $\mathbf{s} = (s_1, \ldots, s_n) \in \{-,+\}^n$, there is exactly one column of $G_2^{\otimes n}$ in the form $1^{(\mathbf{s})}$ (see equation (3.1) and the paragraph following it, where we use $v = v^{(0)} = [1] \in \mathbb{F}_2$ to be the length-1 vector). The term γ can be characterized as follows:

$$\gamma = \left[\frac{1}{N} \sum_{\mathbf{s} \in \{-,+\}^n} n_{DRS}(1^{(\mathbf{s})})\right] - 1.$$

By Lemma 12, the terms in the summation can be grouped according to the number of minus and plus signs in the sequence. Hence,

$$\gamma = \left[\frac{1}{N} \sum_{i=0}^{n} \binom{n}{i} n_{DRS}(1^{(s_1=-,\ldots,s_i=-,s_{i+1}=+,\ldots,s_n=+)}) \right] - 1.$$

Let u_i denote the vector $1^{(s_1,\ldots,s_n)}$ with $s_l = -$ for $l \leqslant i$ and $s_l = +$ for $l > i$. Without loss of generality, let $n\lambda \in \mathbb{N}$. For $i \leqslant n\lambda$, the Hamming weight of u_i is $2^i \leqslant 2^{n\lambda} = w_{u.b.}$. Hence, $n_{DRS}(u_i) = 1$. For $i > n\lambda$, u_i is split into $2^{i-n\lambda}$ vectors, each of which having weight equal to $2^{n\lambda}$. Therefore, $n_{DRS}(u_i) = 2^{i-n\lambda}$.

The term γ can be written as follows:

$$\gamma = \sum_{i=0}^{n\lambda} \frac{1}{N} \binom{n}{i} + \sum_{i=n\lambda+1}^{n} \frac{1}{N} \binom{n}{i} 2^{i-n\lambda} - 1 = \sum_{i=n\lambda+1}^{n} a_i, \tag{A.1}$$

where $a_i \triangleq \frac{1}{N}\binom{n}{i}(2^{i-n\lambda} - 1)$. Now, let $\alpha = i/n$. Since $i > n\lambda$ for each summand a_i, we consider $\alpha > \lambda > \frac{1}{2}$ in the following calculations. The term a_i can be written as

$$\begin{aligned} a_i = a_{n\alpha} &= 2^{-n} \binom{n}{n\alpha} 2^{n\alpha - n\lambda + o(1)} \\ &= 2^{-n} \cdot 2^{nh_b(\alpha)+o(1)} \cdot 2^{n\alpha - n\lambda + o(1)} \\ &= 2^{n \cdot f(\alpha,\lambda) + o(1)}, \end{aligned} \tag{A.2}$$

where the third equality is due to an asymptotic approximation of the binomial coefficient, and $f(\alpha, \lambda) \triangleq h_b(\alpha) + \alpha - \lambda - 1$.

Consider $f(\alpha, \lambda)$ as a function of α over the interval $[0, 1]$. We find its first and second derivatives with respect to α as follows:

$$\frac{\partial f(\alpha,\lambda)}{\partial \alpha} = 1 - \log \frac{\alpha}{1-\alpha}, \tag{A.3}$$

$$\frac{\partial^2 f(\alpha,\lambda)}{\partial^2 \alpha} = -\frac{1}{\ln 2}\left(\frac{1}{\alpha} + \frac{1}{1-\alpha}\right) < 0, \text{ for } 0 < \alpha < 1. \tag{A.4}$$

Thus, for any fixed λ, $f(\alpha, \lambda)$ is a concave function of α and has local maximum when ${}^{\partial f(\alpha,\lambda)}/{}_{\partial \alpha} = 0$. From (A.3), the equality holds if and only if $\alpha = \frac{2}{3}$, and the maximum is

$$\sup_{\alpha \in (0,1)} f(\alpha, \lambda) = h_b\left(\frac{2}{3}\right) + \frac{2}{3} - \lambda - 1 = \lambda^* - \lambda. \tag{A.5}$$

Also, when $\alpha = 0$, $f(0,\lambda) = -\lambda - 1$, and when $\alpha = 1$, $f(1,\lambda) = -\lambda$. Hence, $\sup_{\alpha\in[0,1]} f(\alpha,\lambda) = \lambda^* - \lambda$. When $\lambda > \lambda^*$, we know $f(\frac{i}{n},\lambda) \leqslant \sup_\alpha f(\alpha,\lambda) < 0$ for all integers $0 \leqslant i \leqslant n$, and (A.2) implies that $a_i \to 0$ exponentially fast for each i. Equation (A.1) then shows that γ vanishes exponentially fast in n. ■

A.3 Proof for Section 3.5

Proof for Lemma 14: We show the claim by proving the following: when we encode the source bits according to G', the bit-channels observed by the source bits are BECs and that the erasure probabilities are less than or equal to those when G is used. We use proof by induction on n. For ease of notation, we use M' to denote DRS(M) for a given matrix M in this proof.

For $n = 1$, if $G_2 = G_2'$, we naturally have $Z_{G_2'}^{(s_1)} = Z_{G_2}^{(s_1)}$ for $s_1 \in \{-,+\}$. If $G_2 \neq G_2'$, the latter must be $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$, corresponding to the encoding block diagram in Figure A.1, where the solid black circles indicate a *split* of the XOR operation, i.e., the two operands of the original XOR operation are transmitted through two copies of channel W. The bit-channels observed by U_1 and U_2, denoted by $W^{\boxminus}$ and $W^{\boxplus}$, are BECs with erasure probability ϵ and ϵ^2, respectively. Note that the Bhattacharyya parameters satisfy the following:

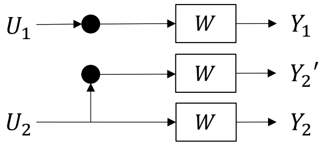


Figure A.1: Encoding Block for G_2'

$$Z_{G_2'}^{(-)} = Z(W^{\boxminus}) = \epsilon \leqslant Z_{G_2}^{(-)} = Z(W^-) = 2\epsilon - \epsilon^2,$$
$$Z_{G_2'}^{(+)} = Z(W^{\boxplus}) = \epsilon^2 = Z_{G_2}^{(+)} = Z(W^+).$$

Suppose that, for a fixed $w_{u.b.}$, the claim holds for all $n \leqslant m$ for some integer $m \geqslant 1$. Let B_m denote the encoding block corresponding to the generator matrix $(G_2^{\otimes m})'$, with inputs $U_1, \ldots, U_{2^m}$ and encoded bits $X_1, \ldots, X_{f(m)}$, where $f(m)$ is the number of columns in $(G_2^{\otimes m})'$. Using the matrix notation, the relation between the input bits and encoded bits is:

$$(U_1, \ldots, U_{2^m})(G_2^{\otimes m})' = (X_1, \ldots, X_{f(m)}).$$

For $n = m+1$, the matrix $(G_2^{\otimes m+1})'$ is associated with $(G_2^{\otimes m})'$ as follows:

$$(G_2^{\otimes m+1})' = \mathrm{DRS}\left(\begin{bmatrix} (G_2^{\otimes m})' & \mathbf{0} \\ (G_2^{\otimes m})' & (G_2^{\otimes m})' \end{bmatrix}\right). \tag{A.6}$$

Since $(G_2^{\otimes m})'$ consists of the outputs of the DRS algorithm, the columns in the right half of the input matrix in equation (A.6) remain unaltered in the output. For the columns in the left half, they are of the form $[v^T, v^T]^T$ for some column v of $(G_2^{\otimes m})'$. If $2w_H(v) > w_{u.b.}$, the outputs of the DRS algorithm are $[\mathbf{0}^T, v^T]^T$ and $[v^T, \mathbf{0}^T]^T$ because the vector v must have weight no larger than the threshold. If $2w_H(v) \leqslant w_{u.b.}$, the algorithm leaves the vector unchanged. We may represent the encoding block B_{m+1} as in Figure A.2, where it is assumed that the j-th column of the input matrix in (A.6) is halved by the DRS algorithm.

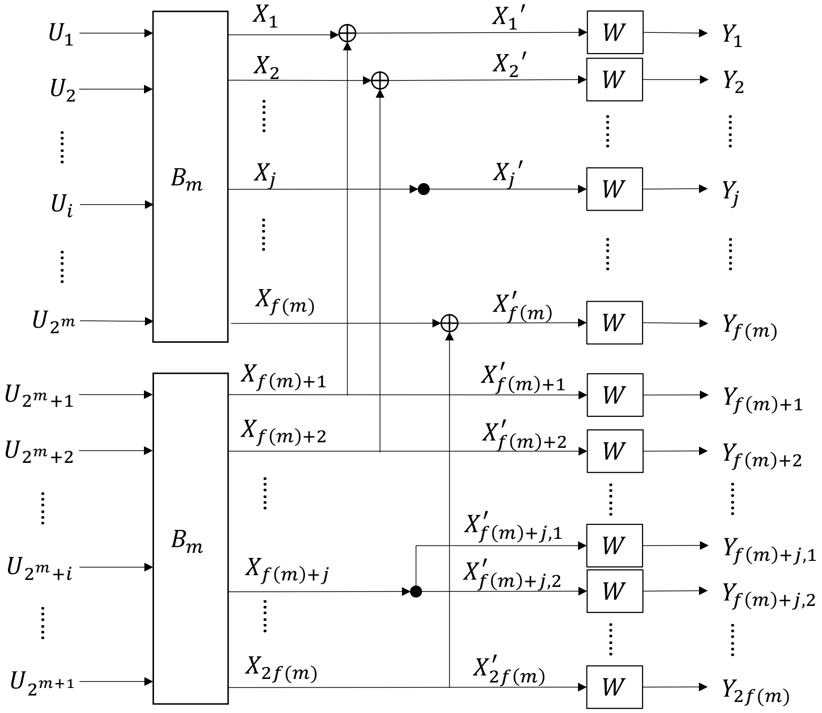


Figure A.2: Encoding Block B_{m+1}

The erasure probabilities for the bit-channels observed by X_i, denoted here as $W(X_i)$, are less than or equal to $2\epsilon - \epsilon^2$ for $1 \leqslant i \leqslant f(m)$, and are equal to ϵ^2 for $f(m)+1 \leqslant i \leqslant 2f(m)$, respectively. Hence we may replace the XOR operations to the right of the X_i's as well as the transmission over W's by BECs $W(X_1), \ldots, W(X_{f(m)})$, $W(X_{f(m)+1}), \ldots, W(X_{2f(m)})$, as in Figure A.3.

One may observe that the erasure probability for the bit-channel observed by U_i is a non-decreasing function of those of $W(X_1), \ldots, W(X_{f(m)})$ for $i \leqslant 2^m$, and of $W(X_{f(m)+1}), \ldots,$

$W(X_{2f(m)})$ for $i > 2^m$. Thus, for $i \leqslant 2^m$, we have

$$\begin{aligned}
&Z_{(G_2^{\otimes m+1})'}\left(U_i \mid Z(W)=\epsilon\right)\\
&\leqslant Z\left[W(U_i) \mid Z(W(X_j)) = 2\epsilon-\epsilon^2,\ \text{for } 1 \leqslant j \leqslant f(m)\right]\\
&= Z_{(G_2^{\otimes m})'}\left(U_i \mid Z(W) = 2\epsilon-\epsilon^2\right)\\
&\leqslant Z_{G_2^{\otimes m}}\left(U_i \mid Z(W) = 2\epsilon-\epsilon^2\right)\\
&= Z_{G_2^{\otimes m+1}}\left(U_i \mid Z(W)=\epsilon\right),
\end{aligned}$$

where the first inequality is due to $Z(W(X_j)) \leqslant 2\epsilon - \epsilon^2$ for $1 \leqslant j \leqslant f(m)$ and the second inequality follows from the hypothesis of the induction.

Similarly, for $i > 2^m$, we have

$$\begin{aligned}
&Z_{(G_2^{\otimes m+1})'}\left(U_i \mid Z(W)=\epsilon\right)\\
&= Z\left[W(U_i) \mid Z(W(X_j)) = \epsilon^2,\ \text{for } f(m)+1 \leqslant j \leqslant 2f(m)\right]\\
&= Z_{(G_2^{\otimes m})'}\left(U_{i-f(m)} \mid Z(W) = \epsilon^2\right)\\
&\leqslant Z_{G_2^{\otimes m}}\left(U_{i-f(m)} \mid Z(W) = \epsilon^2\right)\\
&= Z_{G_2^{\otimes m+1}}\left(U_i \mid Z(W)=\epsilon\right).
\end{aligned}$$

Hence the inequality holds when $n = m+1$ as well. ■

A.4 Proofs for Section 3.6

A.4.1 Proof of Proposition 16

First note that each XOR operation at the j-th recursion can be associated with exactly one vector $\mathbf{s} = (s_1, s_2, \ldots, s_n) \in \{-,+\}^n$ and $s_j = -$. For example, assume that the number of minus signs in $\mathbf{s}$, denoted as $m(\mathbf{s})$, is larger than $n_{lub} = \log w_{u.b.} = n\lambda$, and let $\tau = \tau(\mathbf{s})$ be the index such that $m(s_\tau, s_{\tau+1}, \ldots, s_n) = n_{lub}$ and $s_\tau = -$. Then for each index i in the set $\{k : 1 \leqslant k < \tau, s_k = -\}$, there is a bijection between the pair $(\mathbf{s}, i)$ and an XOR operation at the i-recursion which is split and modified in the ADRS scheme. Hence, the extra complexity of the SC decoder for the ADRS scheme, compared to that of the SC decoder for the code

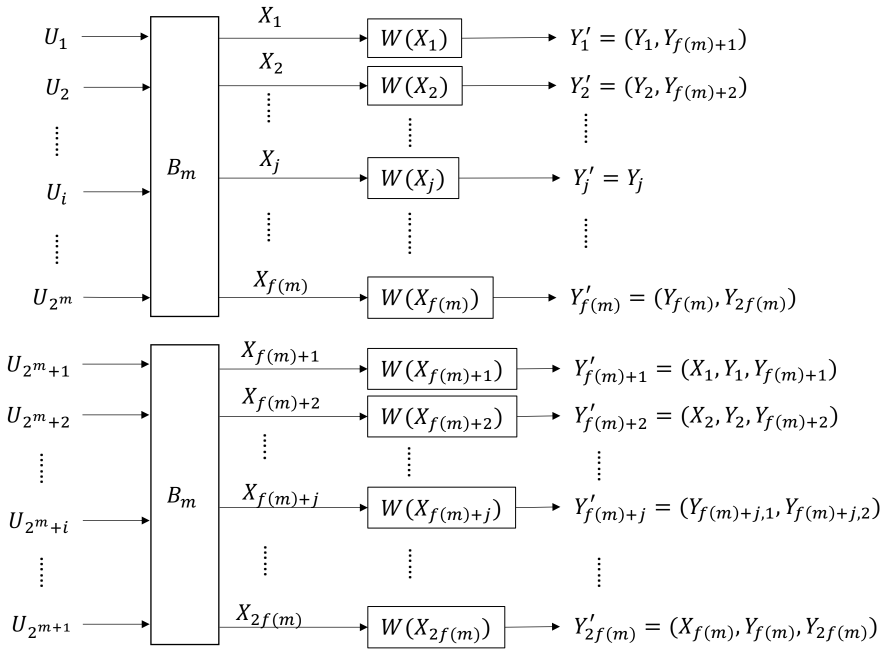


Figure A.3: Equivalent Encoding Block B_{m+1}

based on $G_2^{\otimes n}$, is given by

$$\begin{aligned}
&\sum_{l=1}^{n-n_{lub}+1} \left|\{\mathbf{s} \in \{-,+\}^n : \tau(\mathbf{s}) > l, s_l = -\}\right| 2(2^{l+1}-2)c \\
=& \sum_{l=1}^{n-n_{lub}+1} \sum_{k=l+1}^{n-n_{lub}+1} \left|\{\mathbf{s} \in \{-,+\}^n : \tau(\mathbf{s}) = k, s_l = -\}\right| \cdot \\
&\qquad 2(2^{l+1}-2)c \\
=& \sum_{k=1}^{n-n_{lub}+1} \left[\binom{n-k+1}{n_{lub}} 2^{k-2} \sum_{l=1}^{k-1} 2(2^{l+1}-2)c \right] \\
\leqslant& \, 4c \sum_{k=1}^{n-n_{lub}+1} \binom{n-k+1}{n_{lub}} 2^{k-2} \sum_{l=1}^{k-1} 2^l \\
=& \, 4c \sum_{k=1}^{n-n_{lub}+1} \binom{n-k+1}{n_{lub}} 2^{k-2}(2^k-2) \\
\leqslant& \, 4c \sum_{k=0}^{n-n_{lub}} \binom{n-k}{n_{lub}} 2^{2k}.
\end{aligned} \tag{A.7}$$

Now, let $\alpha = \frac{k}{n} \in [0, 1-\lambda]$. Using Stirling's approximation we have

$$\binom{n-k}{n_{lub}} 2^{2k} = \binom{n(1-\alpha)}{n\lambda} 2^{2\alpha n} \approx 2^{n(1-\alpha)h_b(\frac{\lambda}{1-\alpha})+2\alpha n},$$

for all sufficiently large n. It suffices to assume that $\lambda < \frac{3}{4}$. For $\lambda \geqslant \frac{3}{4}$, note that fewer XOR operations are split and modified, and that the resulting additional decoding complexity is not larger than when $\lambda < \frac{3}{4}$ is used. Let $f(\alpha, \lambda) = (1-\alpha)h_b(\frac{\lambda}{1-\alpha}) + 2\alpha$. We find its maximum, for a given λ, by solving

$$\begin{aligned} 0 = \frac{\partial}{\partial \alpha} f(\alpha, \lambda) = \frac{\partial}{\partial \alpha} \Bigg[&- (1-\alpha) \left(\frac{\lambda}{1-\alpha} \log \frac{\lambda}{1-\alpha} \right) \\ &- (1-\alpha) \left(1 - \frac{\lambda}{1-\alpha} \right) \log \left(1 - \frac{\lambda}{1-\alpha} \right) + 2\alpha \Bigg] \\ = \frac{1}{\ln 2} &\left[-\ln(1-\alpha) + \ln(1-\alpha-\lambda) \right] + 2, \end{aligned}$$

which is true if and only if $\alpha = 1 - \frac{4}{3}\lambda$. And note that $\frac{\partial^2 f(\alpha,\lambda)}{\partial \alpha^2} = \frac{1}{\ln 2}\left[\frac{1}{1-\alpha} - \frac{1}{1-\alpha-\lambda}\right] < 0$ for all $\alpha \in [0, 1-\lambda]$. The maximum of the function is then given by $f(1-\frac{4}{3}\lambda, \lambda) = 2 - \lambda \log 3$. Using the union bound, the sum in (A.7) can be bounded by $4cn2^{n(2-\lambda\log 3)}$, and the ratio of the sum to $N = 2^n$, denoted by γ_C, is bounded from above as $\gamma_C \leqslant 4cn2^{n(1-\lambda\log 3)}$. Since the exponent $n(1-\lambda\log 3)$ goes to negative infinity as n grows when $\lambda > \lambda^{\dagger} = 1/\log 3 \approx 0.631$, the ratio γ_C vanishes exponentially in n when $\lambda > \lambda^{\dagger}$. The proposition follows by noting that the SC decoding complexity for the code based on $G_2^{\otimes n}$ is $N \log N$. ■

A.4.2 Proof of Proposition 17

Similar to the proof of Proposition 16, the number of additional channels due to the ADRS scheme modification is given by

$$\begin{aligned}
&\sum_{l=1}^{n-n_{lub}+1} \left|\{\mathbf{s} \in \{-,+\}^n : \tau(\mathbf{s}) > l, s_l = -\}\right| 2^l \\
&= \sum_{l=1}^{n-n_{lub}+1} \sum_{k=l+1}^{n-n_{lub}+1} \left|\{\mathbf{s} \in \{-,+\}^n : \tau(\mathbf{s}) = k, s_l = -\}\right| 2^l \\
&\leqslant \sum_{k=1}^{n-n_{lub}+1} \binom{n-k+1}{n_{lub}} 2^{k-2} \sum_{l=1}^{k-1} 2^l \\
&\leqslant \sum_{k=0}^{n-n_{lub}} \binom{n-k}{n_{lub}} 2^{2k}. \qquad \text{(A.8)}
\end{aligned}$$

By the argument in the proof of Proposition 16, the sum in (A.8) can be upper bounded by $n2^{n(2-\lambda\log 3)}$, and the ratio of the sum to $N = 2^n$, denoted by γ, is bounded from above as $\gamma \leqslant n2^{n(1-\lambda\log 3)}$. Since the exponent $n(1-\lambda\log 3)$ goes to negative infinity as n grows when $\lambda > \lambda^\dagger = 1/\log_2 3 \approx 0.631$, the ratio γ vanishes exponentially in n when $\lambda > \lambda^\dagger$. ■

APPENDIX B

Supplementary material for Chapter 4

B.1 Harmonic Analysis

We compile in this section harmonic analysis preliminaries as in [106, 72]. See [72] for a more detailed treatment. Here we list several necessary definitions and simple facts.

Consider the abelian group structure $\mathbb{F}_2^n = (\mathbb{Z}/2\mathbb{Z})^n$ on the hypercube $\{0,1\}^n$. The characters of the abelian group $\mathbb{F}_2^n$ are $\{\chi_z\}_{z\in\mathbb{F}_2^n}$, where $\chi_z : \{0,1\}^n \to \{-1,1\}$ is given by $\chi_z(x) = (-1)^{\langle x,z\rangle}$ and $\langle x,z\rangle = \sum_{i=1}^n x_i z_i$.

Consider the $\mathbb{R}$-vector space $\mathcal{L}(\mathbb{F}_2^n) = \{f : \mathbb{F}_2^n \to \mathbb{R}\}$ endowed with the inner product $\langle\cdot,\cdot\rangle$, associated with the uniform distribution on $\{0,1\}^n$:

$$\langle f,g\rangle = \mathbb{E}_{U_n} fg = \frac{1}{2^n}\sum_{x\in\mathbb{F}_2^n} f(x)g(x). \tag{B.1}$$

The set of 2^n characters $\{\chi_z\}_{z\in\mathbb{F}_2^n}$ form an orthonormal basis in the space $\mathcal{L}(\mathbb{F}_2^n)$, equipped with uniform probability distribution. That is, for each $z, z' \in \{0,1\}^n$, $\langle\chi_z,\chi_{z'}\rangle = \delta_{z,z'}$, where δ is the Kronecker delta function. The *Fourier transform* of a function $f \in \mathcal{L}(\mathbb{F}_2^n)$ is the function $\mathcal{F}(f) = \widehat{f} \in \mathcal{L}(\mathbb{F}_2^n)$ given by the coefficients of the unique expansion of f in terms of the characters:

$$f(x) = \sum_z \widehat{f}(z)\chi_z(x) \text{ or equivalently, } \widehat{f}(z) = \langle f,\chi_z\rangle. \tag{B.2}$$

One may show that $\mathcal{F}(\mathcal{F}(f)) = 2^n f$, and $\mathbb{E}\, f = \widehat{f}(0)$. For $f,g \in \mathcal{L}(\mathbb{F}_2^n)$, the Parseval's identity holds: $\langle f,g\rangle = \sum_z \widehat{f}(z)\widehat{g}(z) = 2^n\left\langle \widehat{f},\widehat{g}\right\rangle$. A special case of the above equality is the following equality: $\mathbb{E}\, f^2 = \sum_z \widehat{f}(z)^2$.

The convolution of f and g is defined by $(f * g)(x) = \mathbb{E}_y\, f(y)g(x+y) = \frac{1}{2^n}\sum_{y\in\mathbb{F}_2^n} f(y)g(x+y)$. The convolution transforms to dot product: $\widehat{f * g} = \widehat{f}\cdot\widehat{g}$. The convolution operator is commutative and associative. For arbitrary functions $f, g, h \in \mathcal{L}(\mathbb{F}_2^n)$, the following equality holds:

$$\langle f * g, h\rangle = \langle f, g * h\rangle. \tag{B.3}$$

Also, it can be shown that $\mathbb{E}(f * g) = \mathbb{E}\, f \cdot \mathbb{E}\, g$ for all functions $f, g \in \mathcal{L}(\mathbb{F}_2^n)$.

In this section and in Appendix-B.2, $L \in \mathcal{L}(\mathbb{F}_2^n)$ is a function defined by $L(x) = 2^n$ for $x \in \{0,1\}^n$ with $w_H(x) = 1$, and $L(x) = 0$ otherwise. Let A denote the $2^n \times 2^n$ adjacency matrix of $\mathbb{F}_2^n$, such that $Af(x) = (Af)(x) = \sum_{y\in\mathbb{F}_2^n : d_H(x,y)=1} f(y)$. For any $f \in \mathcal{L}(\mathbb{F}_2^n)$ holds $Af = f * L$ because for $x \in \mathbb{F}_2^n$, $Af(x) = \sum_{y:d_H(x,y)=1} f(y) = \sum_{y:w_H(y)=1} f(x+y)$ $= \mathbb{E}_y\, L(y) f(x+y) = (L * f)(x) = (f * L)(x)$. The Fourier transform of L is the function $\mathcal{F}(L) = \widehat{L}$ given by $\widehat{L}(z) = \langle L, \chi_z\rangle = \sum_{x:w_H(x)=1} (-1)^{\langle x,z\rangle} = n - 2\cdot w_H(z)$.

For $C \subset \mathbb{F}_2^n$, let $1_C \in \mathcal{L}(\mathbb{F}_2^n)$ be the indicator function of C. It can be shown that a code C has minimum distance d if and only if $(1_C * 1_C)(x) = 0$ for all $0 < w_H(x) < d$.

B.2 Proof of Proposition 34

Let f_B be an eigenfunction supported on B corresponding to its maximal eigenvalue λ_B. That is $\lambda_B = \langle Af_B, f_B\rangle / \langle f_B, f_B\rangle$. It is known that the maximum can be attained with an non-negative function f_B, and further we have $Af_B \geqslant \lambda_B f_B$ (see [46, p.13-15 and appendix C]) for details). We write $f = f_B$ and $\lambda = \lambda_B$ interchangeably, and denote the Hamming weight of $x \in \mathbb{F}_2^n$ by $|x| = w_H(x)$, in this proof. As f is supported on B, Cauchy-Schwarz inequality yields the following:

$$\mathbb{E}^2 f = \langle f, 1_B\rangle^2 \leqslant \mathbb{E}\, f^2 \cdot \mathbb{E}(1_B)^2 = \mathbb{E}\, f^2 \cdot |B|/2^n. \tag{B.4}$$

Let $\phi \in \mathcal{L}(\mathbb{F}_2^n)$ be a function such that $(\widehat{\phi})^2 = \widehat{\phi * \phi} = 1_C * 1_C$. Equivalently, $\phi * \phi = 2^n \widehat{1_C * 1_C} = 2^n \widehat{1_C}^2$. Therefore we have

$$\phi * \phi \geqslant 0 \text{ and } \frac{\mathbb{E}(\phi^2)}{\mathbb{E}^2(\phi)} = \frac{(\phi * \phi)(0)}{\widehat{\phi}^2(0)} = |C|\,. \tag{B.5}$$

Now let $F = \phi * f$. We estimate the product $\langle AF, F\rangle$ in two ways. First,

$$\begin{aligned} \langle AF, F\rangle &= \langle (\phi * f) * L, \phi * f\rangle = \langle \phi * \phi * f, f * L\rangle \\ &= \langle \phi * \phi * f, Af\rangle \geqslant \langle \phi * \phi * f, \lambda f\rangle \\ &= \lambda\langle \phi * f, \phi * f\rangle = \lambda\langle F, F\rangle = \lambda\, \mathbb{E}\, F^2. \end{aligned}$$

Second, by Parseval's identity,

$$\begin{aligned}\langle AF, F\rangle &= 2^n \left\langle \widehat{AF}, \widehat{F} \right\rangle = 2^n \left\langle \widehat{L} \cdot \widehat{F}, \widehat{F} \right\rangle \\ &= \sum_z (n - 2\,|z|)\, \widehat{F}^2(z).\end{aligned}$$

Since $\widehat{F} = \widehat{\phi} \cdot \widehat{f}$ and $(\widehat{\phi})^2(z) = (1_C * 1_C)(z)$, $\widehat{F}(z) = 0$ for all $0 < |z| < d$. We can estimate $\langle AF, F\rangle$ by

$$\begin{aligned}\sum_z (n - 2\,|z|)\, \widehat{F}^2(z) &= n\widehat{F}^2(0) + \sum_{z:|z|\geqslant d} (n - 2\,|z|)\, \widehat{F}^2(z) \\ &\leqslant n\widehat{F}^2(0) + (n - 2d) \sum_z \widehat{F}^2(z) = n\,\mathbb{E}^2 F + (n-2d)\,\mathbb{E}\, F^2.\end{aligned}$$

Combining the two estimates, we have the following inequality: $n\,\mathbb{E}^2 F \geqslant (\lambda - (n-2d))$ $\mathbb{E}\, F^2$. Since

$$\mathbb{E}^2 F = \mathbb{E}^2(\phi * f) = [\widehat{\phi * f}(0)]^2 = [\widehat{\phi}(0)\widehat{f}(0)]^2 = \mathbb{E}^2 \phi\, \mathbb{E}^2 f,$$

$$\begin{aligned}\mathbb{E}\, F^2 &= \langle F, F\rangle = \langle \phi * f, \phi * f\rangle = \langle \phi * \phi, f * f\rangle \\ &\geqslant 1/2^n (\phi * \phi)(0)(f * f)(0) = 1/2^n\, \mathbb{E}\,\phi^2\, \mathbb{E}\, f^2,\end{aligned}$$

as $\phi * \phi = 2^n \cdot \widehat{1_C}^2 \geqslant 0$, we now have

$$n\,\mathbb{E}^2 \phi\, \mathbb{E}^2 f \geqslant (\lambda - (n-2d))\, \frac{1}{2^n}\, \mathbb{E}\,\phi^2\, \mathbb{E}\, f^2. \tag{B.6}$$

Leveraging equations (B.4), (B.5), and (B.6), the size of any code C with minimum distance d is

$$|C| = \frac{\mathbb{E}\,\phi^2}{\mathbb{E}^2 \phi} \leqslant \frac{n}{\lambda - (n-2d)} \cdot 2^n \frac{\mathbb{E}^2 f}{\mathbb{E}\, f^2} \leqslant \frac{n}{\lambda - (n-2d)}\, |B|\,. \qquad \blacksquare$$

BIBLIOGRAPHY

[1] Fariba Abbasi, Hessam Mahdavifar, and Emanuele Viterbo. Hybrid non-binary repeated polar codes for low-SNR regime. In *IEEE International Symposium on Information Theory*, pages 1742–1747, 2021.

[2] Fariba Abbasi, Hessam Mahdavifar, and Emanuele Viterbo. Polar coded repetition for low-capacity channels. In *IEEE Information Theory Workshop*, pages 1–5, 2021.

[3] Fariba Abbasi, Hessam Mahdavifar, and Emanuele Viterbo. Hybrid non-binary repeated polar codes. *IEEE Transactions on Wireless Communications*, 21(9):7582–7594, 2022.

[4] Fariba Abbasi, Hessam Mahdavifar, and Emanuele Viterbo. Polar coded repetition. *IEEE Transactions on Communications*, 70(10):6399–6409, 2022.

[5] Emmanuel Abbe. Polarization and randomness extraction. *Proceedings of IEEE International Symposium on Information Theory*, pages 184–188, 2011.

[6] Erik Agrell. Bounds for unrestricted binary codes. Available at `https://codes.se/bounds/unr.html` (223/07/12).

[7] Rudolf Ahlswede. Group codes do not achieve shannon's channel capacity for general discrete channels. *The Annals of Mathematical Statistics*, pages 224–240, 1971.

[8] Rudolf Ahlswede and J Gemma. Bounds on algebraic code capacities for noisy channels I. *Information and Control*, 19(2):124–145, 1971.

[9] Rudolf Ahlswede and J Gemma. Bounds on algebraic code capacities for noisy channels II. *Information and Control*, 19(2):146–158, 1971.

[10] Kwangjun Ahn, Kangwook Lee, and Changho Suh. Community recovery in hypergraphs. *arXiv preprint arXiv:1709.03670*, 2017.

[11] Mattias Andersson, Vishwambhar Rathi, Ragnar Thobaben, Jorg Kliewer, and Mikael Skoglund. Nested polar codes for wiretap and relay channels. *IEEE Communications Letters*, 14(8):752–754, 2010.

[12] E. Arıkan and E. Telatar. On the rate of channel polarization. In *IEEE International Symposium on Information Theory*, pages 1493–1495, June 2009.

[13] Erdal Arıkan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, 2009.

[14] Erdal Arıkan. Source polarization. In *IEEE International Symposium on Information Theory*, pages 899–903, 2010.

[15] Erdal Arıkan. On the origin of polar coding. *IEEE Journal on Selected Areas in Communications*, 34(2):209–223, 2015.

[16] Hassan Ashtiani, Shrinu Kushagra, and Shai Ben-David. Clustering with same-cluster queries. In *NIPS*, pages 3216–3224, 2016.

[17] Alexander Barg, Sugi Guritman, and Juriaan Simonis. Strengthening the Gilbert–Varshamov bound. *Linear Algebra and its Applications*, 307(1-3):119–129, 2000.

[18] Alexander Barg and David B Jaffe. Numerical results on the asymptotic rate of binary codes. *Codes and Association Schemes*, 56:25–32, 1999.

[19] Alexander Barg and Dmitry Nogin. A functional view of upper bounds on codes. In *Coding and Cryptology*, pages 15–24. World Scientific, 2008.

[20] Alexander M Barg and D Yu Nogin. Spectral approach to linear programming bounds on codes. *Problems of Information Transmission*, 42(2):77–89, 2006.

[21] Raj Chandra Bose and Dwijendra K Ray-Chaudhuri. On a class of error correcting binary group codes. *Information and control*, 3(1):68–79, 1960.

[22] Andries Brouwer. Small binary codes: Table of general binary codes. Available at https://www.win.tue.nl/%7Eaeb/codes/binary-1.html (223/07/12).

[23] D. Burshtein, M. Krivelevich, S. Litsyn, and G. Miller. Upper bounds on the rate of LDPC codes. *IEEE Trans. on Info. Theory*, 48(9):2437–2449, Sept 2002.

[24] D. Burshtein, M. Krivelevich, S. Litsyn, and G. Miller. Upper bounds on the rate of ldpc codes. *IEEE Trans. on Info. Theory*, 48(9):2437–2449, Sept 2002.

[25] Francesco Buscemi and Nilanjana Datta. The quantum capacity of channels with arbitrarily correlated noise. *IEEE Transactions on Information theory*, 56(3):1447–1460, 2010.

[26] Giuseppe Caire, Shlomo Shamai, and Sergio Verdú. Noiseless data compression with low-density parity-check codes. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 66, 08 2004.

[27] Venkat Venkat Bala Chandar. *Sparse graph codes for compression, sensing, and secrecy.* PhD thesis, Massachusetts Institute of Technology, 2010.

[28] Giacomo Como and Fabio Fagnani. The capacity of finite abelian group codes over symmetric memoryless channels. *IEEE Transactions on Information Theory*, 55(5):2037–2054, 2009.

[29] Leonardo Nagami Coregliano, Fernando Granha Jeronimo, and Chris Jones. A complete linear programming hierarchy for linear codes. *arXiv preprint arXiv:2112.09221*, 2021.

[30] Eren Şaşoğlu, Emre Telatar, and Edmund Yeh. Polar codes for the two-user binary-input multiple-access channel. *IEEE Transactions on Information Theory*, 59(10):6583–6592, 2013.

[31] Nilanjana Datta and Tony C Dorlas. The coding theorem for a class of quantum channels with long-term memory. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8147, 2007.

[32] Philippe Delsarte. An algebraic approach to the association schemes of coding theory. *Philips Res. Rep. Suppl.*, 10:vi+–97, 1973.

[33] Philippe Delsarte, Jean-Marie Goethals, and F Jessie Mac Williams. On generalized reedmuller codes and their relatives. *Information and control*, 16(5):403–442, 1970.

[34] Alexandros G Dimakis, P Brighten Godfrey, Yunnan Wu, Martin J Wainwright, and Kannan Ramchandran. Network coding for distributed storage systems. *IEEE Transactions on Information Theory*, 56(9):4539–4551, 2010.

[35] RL Dobrushin. Asymptotic optimality of group and systematic codes for some channels. *Theory of Probability & Its Applications*, 8(1):47–60, 1963.

[36] RL Dobrusin. Asymptotic bounds of the probability of error for the transmission of messages over a discrete memoryless channel with a symmetric transition probability matrix. *Teor. Veroyatnost. i Primenen*, 7:283–311, 1962.

[37] Tony Dorlas and Ciara Morgan. Invalidity of a strong capacity for a quantum channel with memory. *Physical Review A*, 84(4):042318, 2011.

[38] Ilya Dumer and Navid Gharavi. Codes for high-noise memoryless channels. In *IEEE International Symposium on Information Theory and Its Applications*, pages 101–105, 2020.

[39] Ilya Dumer and Navid Gharavi. Codes approaching the shannon limit with polynomial complexity per information bit. In *IEEE International Symposium on Information Theory*, pages 238–243, 2021.

[40] Ilya Dumer and Navid Gharavi. Combined polar-LDPC design for channels with high noise. In *IEEE Information Theory Workshop*, pages 1–6, 2021.

[41] Frédéric Dupuis, Lea Kraemer, Philippe Faist, Joseph M Renes, and Renato Renner. Generalized entropies. In *XVIIth international congress on mathematical physics*, pages 134–153. World Scientific, 2014.

[42] Michele Elia. Some results on the existence of binary linear codes (corresp.). *IEEE Transactions on Information Theory*, 29(6):933–934, 1983.

[43] Cen Feng. Capacity-approaching joint source-channel coding for asymmetric channels with low-density parity-check codes. *Sensors & Transducers*, 168(4):203–209, 2014.

[44] Mohammad Fereydounian, Mohammad Vahid Jamali, Hamed Hassani, and Hessam Mahdavifar. Channel coding at low capacity. In *IEEE Information Theory Workshop*, pages 1–5, 2019.

[45] G David Forney and Mitchell D Trott. The dynamics of group codes: state spaces, trellis diagrams, and canonical encoders. *IEEE Transactions on Information Theory*, 39(5):1491–1513, 1993.

[46] Joel Friedman and Jean-Pierre Tillich. Generalized Alon–Boppana theorems and error-correcting codes. *SIAM Journal on Discrete Mathematics*, 19(3):700–718, 2005.

[47] Patricio Fuentes, Josu Etxezarreta Martinez, Pedro M Crespo, and Javier Garcia-Frias. Approach for the construction of non-Calderbank-Steane-Shor low-density-generator-matrix–based quantum codes. *Physical Review A*, 102(1):012423, 2020.

[48] Philippe Gaborit and Gilles Zemor. Asymptotic improvement of the Gilbert–Varshamov bound for linear codes. *IEEE Transactions on Information Theory*, 54(9):3865–3872, 2008.

[49] R. Gallager. Low-density parity-check codes. *IRE Trans. on Info. Theory*, 8(1):21–28, January 1962.

[50] Robert Gallager. Low-density parity-check codes. *IRE Transactions on Information Theory*, 8(1):21–28, 1962.

[51] Javier Garcia-Frias and Wei Zhong. Approaching Shannon performance by iterative decoding of linear codes with low-density generator matrix. *IEEE Communications Letters*, 7(6):266–268, 2003.

[52] Roberto Garello and Sergio Benedetto. Multilevel construction of block and trellis group codes. *IEEE Transactions on Information Theory*, 41(5):1257–1264, 1995.

[53] Edgar N Gilbert. A comparison of signalling alphabets. *The Bell system technical journal*, 31(3):504–522, 1952.

[54] Thomas John Goblick. *Coding for a discrete information source with a distortion measure*. PhD thesis, Massachusetts Institute of Technology, 1963.

[55] Naveen Goela, Emmanuel Abbe, and Michael Gastpar. Polar codes for broadcast channels. *IEEE Transactions on Information Theory*, 61(2):758–782, 2015.

[56] Jean-Marie Goethals. Nonlinear codes defined by quadratic forms over GF(2). *Information and control*, 31(1):43–74, 1976.

[57] JM Goethals. Two dual families of nonlinear binary codes. *Electronics Letters*, 10(23):471–472, 1974.

[58] Ahmad Golmohammadi, David GM Mitchell, Jörg Kliewer, and Daniel J Costello. Encoding of spatially coupled LDGM codes for lossy source compression. *IEEE Transactions on Communications*, 66(11):5691–5703, 2018.

[59] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. Available at https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/, 2022.

[60] Te Sun Han and Sergio Verdú. Approximation theory of output statistics. *IEEE Transactions on Information Theory*, 39(3):752–772, 1993.

[61] Gary Harris and Clyde Martin. Shorter notes: The roots of a polynomial vary continuously as a function of the coefficients. *Proceedings of the American Mathematical Society*, 100(2):390–392, 1987.

[62] Masahito Hayashi and Hiroshi Nagaoka. General formulas for capacity of classical-quantum channels. *IEEE Transactions on Information Theory*, 49(7):1753–1768, 2003.

[63] Ferdinand B Hergert. On the Delsarte-Goethals codes and their formal duals. *Discrete mathematics*, 83(2-3):249–263, 1990.

[64] Kenichi Hirose. Continuity of the roots of a polynomial. *The American Mathematical Monthly*, 127(4):359–363, 2020.

[65] Alexis Hocquenghem. Codes correcteurs d'erreurs. *Chiffers (in French)*, 2:147–156, 1959.

[66] Alexander S Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, 1998.

[67] S. Horii, T. Yoshida, M. Kobayashi, and T. Matsushima. Distributed stochastic gradient descent using LDGM codes. In *2019 IEEE International Symposium on Information Theory*, pages 1417–1421, 2019.

[68] Shunsuke Horii, Takahiro Yoshida, Manabu Kobayashi, and Toshiyasu Matsushima. Distributed stochastic gradient descent using LDGM codes. In *IEEE International Symposium on Information Theory*, pages 1417–1421. IEEE, 2019.

[69] Panagiotis G Ipeirotis. Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students*, 17(2):16–21, 2010.

[70] Mohammad Vahid Jamali and Hessam Mahdavifar. Massive coded-NOMA for low-capacity channels: A low-complexity recursive approach. *IEEE Transactions on Communications*, 69(6):3664–3681, 2021.

[71] Tao Jiang and Alexander Vardy. Asymptotic improvement of the Gilbert-Varshamov bound on the size of binary codes. *IEEE Transactions on Information Theory*, 50(8):1655–1664, 2004.

[72] Jeff Kahn, Gil Kalai, and Nathan Linial. *The influence of variables on Boolean functions.* Citeseer, 1989.

[73] A. Makhdoumi Kakhaki, H. Karkeh Abadi, P. Pad, H. Saeedi, F. Marvasti, and K. Alishahi. Capacity achieving linear codes with random binary sparse generating matrices over the binary symmetric channel. In *IEEE International Symposium on Information Theory Proceedings*, pages 621–625, 2012.

[74] David R Karger, Sewoong Oh, and Devavrat Shah. Iterative learning for reliable crowdsourcing systems. In *Advances in Neural Information Processing Systems*, pages 1953–1961, 2011.

[75] David R Karger, Sewoong Oh, and Devavrat Shah. Iterative learning for reliable crowdsourcing systems. In *NIPS*, pages 1953–1961, 2011.

[76] David R Karger, Sewoong Oh, and Devavrat Shah. Budget-optimal task allocation for reliable crowdsourcing systems. *Operations Research*, 62(1):1–24, 2014.

[77] Tadao Kasami, Shu Lin, and W Peterson. New generalizations of the Reed-Muller codes–I: Primitive codes. *IEEE Transactions on Information Theory*, 14(2):189–199, 1968.

[78] Satish Babu Korada, Eren Sasoglu, and Rüdiger Urbanke. Polar codes: Characterization of exponent, bounds, and constructions. *IEEE Transactions on Information Theory*, 56(12):6253–6264, 2010.

[79] Ilia Krasikov and Simon Litsyn. On upper bounds for the distance of codes of small size. In *Proceedings of IEEE International Symposium on Information Theory*, page 84, 1997.

[80] Dennis Kretschmann and Reinhard F Werner. Quantum channels with memory. *Physical Review A*, 72(6):062323, 2005.

[81] Dinesh Krithivasan and S Sandeep Pradhan. Distributed source coding using abelian group codes: A new achievable rate-distortion region. *IEEE Transactions on Information Theory*, 57(3):1495–1519, 2011.

[82] S. Kudekar, T. Richardson, and R. L. Urbanke. Spatially coupled ensembles universally achieve capacity under belief propagation. *IEEE Trans. on Info. Theory*, 59(12):7761–7813, Dec 2013.

[83] Farshad Lahouti and Babak Hassibi. Fundamental limits of budget-fidelity trade-off in label crowdsourcing. In *NIPS*, pages 5058–5066, 2016.

[84] Kangwook Lee, Maximilian Lam, Ramtin Pedarsani, Dimitris Papailiopoulos, and Kannan Ramchandran. Speeding up distributed machine learning using codes. *IEEE Transactions on Information Theory*, 64(3):1514–1529, 2018.

[85] Maxwell W Libbrecht and William Stafford Noble. Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6):321–332, 2015.

[86] Wenchao Lin, Suihua Cai, Baodian Wei, and Xiao Ma. Coding theorem for systematic LDGM codes under list decoding. In *2018 IEEE Information Theory Workshop (ITW)*, pages 1–5. IEEE, 2018.

[87] Simon Litsyn. An update table of the best binary codes known. *Handbook of Coding Theory*, 1998.

[88] H-A Loeliger. Signal sets matched to groups. *IEEE Transactions on Information Theory*, 37(6):1675–1682, 1991.

[89] H-A Loeliger and Thomas Mittelholzer. Convolutional codes over groups. *IEEE Transactions on Information Theory*, 42(6):1660–1686, 1996.

[90] Elyassaf Loyfer and Nati Linial. Linear programming hierarchies in coding theory: Dual solutions. *arXiv preprint arXiv:2211.12977*, 2022.

[91] Elyassaf Loyfer and Nati Linial. New LP-based upper bounds in the rate-vs.-distance problem for binary linear codes. *IEEE Transactions on Information Theory*, 2023.

[92] David JC MacKay. Good error-correcting codes based on very sparse matrices. *IEEE Transactions on Information Theory*, 45(2):399–431, 1999.

[93] David JC MacKay and Radford M Neal. Good codes based on very sparse matrices. In *IMA International Conference on Cryptography and Coding*, pages 100–111. Springer, 1995.

[94] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error correcting codes*, volume 16. Elsevier, 1977.

[95] Hessam Mahdavifar. Scaling exponent of sparse random linear codes over binary erasure channels. In *IEEE International Symposium on Information Theory*, pages 689–693. IEEE, 2017.

[96] Hessam Mahdavifar, Mostafa El-Khamy, Jungwon Lee, and Inyup Kang. Polar coding for bit-interleaved coded modulation. *IEEE Transactions on Vehicular Technology*, 65(5):3115–3127, 2015.

[97] Hessam Mahdavifar, Mostafa El-Khamy, Jungwon Lee, and Inyup Kang. Achieving the uniform rate region of general multiple access channels by polar coding. *IEEE Transactions on Communications*, 64(2):467–478, 2016.

[98] Hessam Mahdavifar and Alexander Vardy. Achieving the secrecy capacity of wiretap channels using polar codes. *IEEE Transactions on Information Theory*, 57(10):6428–6443, 2011.

[99] Y. Matsunaga and H. Yamamoto. A coding theorem for lossy data compression by ldpc codes. *IEEE Trans. on Info. Theory*, 49(9):2225–2229, Sept 2003.

[100] Arya Mazumdar and Soumyabrata Pal. Semisupervised clustering, AND-queries and locally encodable source coding. In *Advances in Neural Information Processing Systems*, pages 6489–6499, 2017.

[101] Arya Mazumdar and Soumyabrata Pal. Semisupervised clustering, and-queries and locally encodable source coding. In *NIPS*, pages 6489–6499, 2017.

[102] Arya Mazumdar and Barna Saha. Clustering via crowdsourcing. *arXiv preprint arXiv:1604.01839*, 2016.

[103] Robert McEliece, Eugene Rodemich, Howard Rumsey, and Lloyd Welch. New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE Transactions on Information Theory*, 23(2):157–166, 1977.

[104] Marco Mondelli, Seyed Hamed Hassani, Igal Sason, and Rüdiger L Urbanke. Achieving Marton's region for broadcast channels using polar codes. *IEEE Transactions on Information Theory*, 61(2):783–800, 2015.

[105] Milán Mosonyi and Nilanjana Datta. Generalized relative entropies and the capacity of classical-quantum channels. *Journal of Mathematical physics*, 50(7), 2009.

[106] Michael Navon and Alex Samorodnitsky. Linear programming bounds for codes via a covering argument. *Discrete & Computational Geometry*, 41(2):199, 2009.

[107] Katie M O'Brien and Patrick Fitzpatrick. Bounds on codes derived by counting components in Varshamov graphs. *Designs, Codes and Cryptography*, 39(3):387–396, 2006.

[108] James Chin-Jen Pang, Hessam Mahdavifar, and S Sandeep Pradhan. Coding for crowdsourced classification with XOR queries. In *IEEE Information Theory Workshop*, pages 1–5, 2019.

[109] James Chin-Jen Pang, Hessam Mahdavifar, and S Sandeep Pradhan. Capacity-achieving polar-based ldgm codes with crowdsourcing applications. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 449–454. IEEE, 2020.

[110] James Chin-Jen Pang, Hessam Mahdavifar, and S. Sandeep Pradhan. New bounds on the size of binary codes with large minimum distance. In *IEEE International Symposium on Information Theory*, pages 1963–1968, 2022.

[111] James Chin-Jen Pang, Hessam Mahdavifar, and S Sandeep Pradhan. Capacity-achieving polar-based codes with sparsity constraints on the generator matrices. *IEEE Transactions on Communications*, 2023.

[112] James Chin-Jen Pang, Hessam Mahdavifar, and S. Sandeep Pradhan. New bounds on the size of binary codes with large minimum distance. *IEEE Journal on Selected Areas in Information Theory*, 4:219–231, 2023.

[113] James Chin-Jen Pang, S Sandeep Pradhan, and Hessam Mahdavifar. Abelian group codes for classical and classical-quantum channel coding: one-shot rate bounds. Unpublished Manuscript, 2023.

[114] Woomyoung Park and Alexander Barg. Polar codes for q-ary channels, $q = 2^r$. *IEEE Transactions on Information Theory*, 59(2):955–969, 2012.

[115] Morris Plotkin. Binary codes with specified minimum distance. *IRE Transactions on Information Theory*, 6(4):445–450, 1960.

[116] Rapeepat Ratasuk, Nitin Mangalvedhe, Yanji Zhang, Michel Robert, and Jussi-Pekka Koskinen. Overview of narrowband IoT in LTE Rel-13. In *2016 IEEE Conference on Standards for Communications and Networking*, pages 1–7. IEEE, 2016.

[117] Irving S Reed and Xuemin Chen. *Error-control coding for data networks*, volume 508. Springer Science & Business Media, 2012.

[118] Renato Renner, Stefan Wolf, and Jurg Wullschleger. The single-serving channel capacity. In *2006 IEEE International Symposium on Information Theory*, pages 1424–1427. IEEE, 2006.

[119] T. J. Richardson and R. L. Urbanke. The capacity of low-density parity-check codes under message-passing decoding. *IEEE Trans. on Info. Theory*, 47(2):599–618, Feb 2001.

[120] Aria G Sahebi and S Sandeep Pradhan. Multilevel polarization of polar codes over arbitrary discrete memoryless channels. In *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1718–1725. IEEE, 2011.

[121] Aria Ghasemian Sahebi and S Sandeep Pradhan. Abelian group codes for channel coding and source coding. *IEEE Transactions on Information Theory*, 61(5):2399–2414, 2015.

[122] Alex Samorodnitsky. One more proof of the first linear programming bound for binary codes and two conjectures. *arXiv preprint arXiv:2104.14587*, 2021.

[123] Eren Şaşoğlu, Emre Telatar, and Erdal Arikan. Polarization for arbitrary discrete memoryless channels. In *2009 IEEE Information Theory Workshop*, pages 144–148. IEEE, 2009.

[124] I. Sason and R. Urbanke. Parity-check density versus performance of binary linear block codes over memoryless symmetric channels. *IEEE Trans. on Info. Theory*, 49(7):1611–1635, July 2003.

[125] Benjamin Schumacher and Michael D Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131, 1997.

[126] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[127] Mahyar Shirvanimoghaddam and Sarah Johnson. Raptor codes in the low SNR regime. *IEEE Transactions on Communications*, 64(11):4449–4460, 2016.

[128] Vladimir Michilovich Sidel'nikov. On mutual correlation of sequences. In *Doklady Akademii Nauk*, volume 196, pages 531–534. Russian Academy of Sciences, 1971.

[129] David Slepian. Group codes for the gaussian channel. *Bell System Technical Journal*, 47(4):575–602, 1968.

[130] Dejan Spasov and Marjan Gushev. Some notes on the binary Gilbert-Varshamov bound. In *Sixth International Workshop on Optimal Codes and Related Topics, Varna, Bulgaria*, 2009.

[131] Aimo Tietäväinen. Bounds for binary codes just outside the Plotkin range. *Information and Control*, 47(2):85–93, 1980.

[132] Michael Tsfasman and Serge G Vladut. *Algebraic-geometric codes*, volume 58. Springer Science & Business Media, 2013.

[133] Mojtaba Vaezi, Amin Azari, Saeed R Khosravirad, Mahyar Shirvanimoghaddam, M Mahdi Azari, Danai Chasaki, and Petar Popovski. Cellular, wide-area, and non-terrestrial IoT: A survey on 5G advances and the road toward 6G. *IEEE Communications Surveys & Tutorials*, 24(2):1117–1174, 2022.

[134] Jacobus Hendricus Van Lint. *Introduction to coding theory*, volume 86. Springer Science & Business Media, 1998.

[135] Rom Rubenovich Varshamov. Estimate of the number of signals in error correcting codes. *Docklady Akad. Nauk, SSSR*, 117:739–741, 1957.

[136] Lav R. Varshney. Participation in crowd systems. *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 996–1001, 2012.

[137] F. J. Vazquez-Araujo, M. Gonzalez-Lopez, L. Castedo, and J. Garcia-Frias. Capacity approaching low-rate LDGM codes. *IEEE Trans. on Info. Theory*, 59(2):352–356, February 2011.

[138] A. Vempaty, L. R. Varshney, and P. K. Varshney. Reliable crowdsourcing for multi-class labeling using coding theory. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):667–679, Aug 2014.

[139] Aditya Vempaty, Lav R Varshney, and Pramod K Varshney. Reliable crowdsourcing for multi-class labeling using coding theory. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):667–679, 2014.

[140] Norases Vesdapunt, Kedar Bellare, and Nilesh Dalvi. Crowdsourcing algorithms for entity resolution. *Proceedings of the VLDB Endowment*, 7(12):1071–1082, 2014.

[141] Ramya Korlakai Vinayak and Babak Hassibi. Crowdsourced clustering: Querying edges vs triangles. In *NIPS*, pages 1316–1324, 2016.

[142] Ligong Wang and Renato Renner. One-shot classical-quantum capacity and hypothesis testing. *Physical Review Letters*, 108(20):200501, 2012.

[143] Zicheng Ye, Huazi Zhang, Rong Li, Jun Wang, Guiying Yan, and Zhiming Ma. Improving the Gilbert-Varshamov bound by graph spectral method. *arXiv preprint arXiv:2104.01403*, 2021.

[144] Jing Zhang, Xindong Wu, and Victor S Sheng. Learning from crowdsourced labeled data: a survey. *Artificial Intelligence Review*, 46:543–576, 2016.

[145] Ying Zhang, Sorina Dumitrescu, Jun Chen, and Zhibin Sun. LDGM-based multiple description coding for finite alphabet sources. *IEEE Transactions on Communications*, 60(12):3671–3682, 2012.

[146] W. Zhong, H. Lou, and J. Garcia-Frias. LDGM codes for joint source-channel coding of correlated sources. In *Proceedings 2003 International Conference on Image Processing*, volume 1, pages I–593, Sept 2003.

[147] Wei Zhong, Huiqiong Chai, and Javier Garcia-Frias. Approaching the Shannon limit through parallel concatenation of regular LDGM codes. In *IEEE International Symposium on Information Theory Proceedings, 2005.*, pages 1753–1757.

[148] Dengyong Zhou, Sumit Basu, Yi Mao, and John C Platt. Learning from the wisdom of crowds by minimax entropy. In *NIPS*, pages 2195–2203, 2012.