Achieving Channel Capacities with Nested Linear/Lattice Codes: A Unified Approach

by

Renming Qi

B.Eng., University of Science and Technology of China, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES

(Electrical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

August 2018

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Achieving Channel Capacities with Nested Linear/Lattice Codes: A Unified Approach

submitted by Renming Qi in partial fulfillment of the requirements for
the degree of Master of Applied Science in Electrical Engineering

Dr. Chen Feng, School of Engineering

Supervisor

Dr. Julian Cheng, School of Engineering

Supervisory Committee Member

Dr. Anas Chaaban, School of Engineering

Supervisory Committee Member

Dr. Heinz Bauschke, Irving K. Barber School of Arts and Sciences

University Examiner

Abstract

Random nested lattice codes have played an important role in network information theory. However, they are less accessible than conventional random codes because their achievability proofs are often involved, even for the case of the additive white Gaussian noise (AWGN) channel. In sharp contrast, their finite field counterparts, nested linear codes, enjoy much simpler achievability proofs. In this thesis, we make use of an intriguing connection between nested lattice codes and nested linear codes to handle their achievability proofs in a unified approach. As a by-product of this unified approach, we show it's capable of proving that the algebraic lattice codes constructed using number field could achieve the AWGN channel capacity.

Lay Summary

It's usually considered as an involved problem to use the lattice codes to achieve the capacities of noisy channel. This thesis provides a simpler and more transparent proof by using the underlying algebraic structure of the lattice codes.

Preface

The work outlined in this thesis was conducted by Renming Qi under the supervision of Dr. Chen Feng. A version of Chapter 4 of this thesis has been published in IEEE Information Theory Workshop, Kaohsiung, Taiwan, 2017, with a title of “A simpler proof for the existence of Capacity-Achieving nested lattice codes.” I am responsible for conducting the proofs.

Table of Contents

List of Figures

List of Symbols

Symbols	Definitions
$\mathbb{F}, \mathbb{Q}, \mathbb{R}, \mathbb{F}_{\mathsf{q}}$	a field, the rational numbers, the real numbers and a field of order q, respectively
$\mathcal{X}, \mathcal{Y}$	the alphabets
x, X	constant and random variable, respectively
$\boldsymbol{x}, \boldsymbol{X}$	constant and random row vector, respectively
$\mathsf{C}, \mathbf{\mathsf{C}}$	constant and random linear code, respectively
$\Lambda, \boldsymbol{\Lambda}$	constant and random lattices, respectively
$\mathcal{V}(\Lambda), V(\Lambda)$	the voronoi region of lattice Λ and the volume of $\mathcal{V}(\Lambda)$, respectively
$\mathsf{G}, \mathbf{G}$	constant and random matrice, respectively
$\mathcal{B}(\boldsymbol{s}, r)$	the ball centered at $\boldsymbol{s}$ with radius r
$\mathbb{I}(\cdot)$	the indicator function
$p_X(\cdot), \mathsf{E}(X), \mathrm{Var}(X)$	the pmf, expectation and variance of X
$\pi(x \mid \boldsymbol{x})$	the empirical pmf of $\boldsymbol{x}$
$H(\cdot)$	the entropy
$I(X;Y)$	the mutual information between X and Y
$\mathcal{T}_\epsilon^{(n)}(X)$	the typical set
$\mathcal{T}_\epsilon^{(n)}(X,Y)$	the joint typical set
$\mathcal{T}_\epsilon^{(n)}(X \mid \boldsymbol{y})$	the conditional typical set
$Q_\Lambda(\cdot)$	the nearest neighbor quantizer with respect to Λ
K	a field extension over the field $\mathbb{Q}$
$\mathcal{O}_K$, $\mathfrak{p}$	the ring of integers over K and a prime ideal of $\mathcal{O}_K$, respectively
$\langle a, b \rangle$	an ideal generated by the numbers a and b

Acknowledgements

The life in Kelowna taught me many things. Two of them are most important. The first one is the following

> Youth is to face the reality of the ability to imagine, not in accordance with other people's imagination to deceive themselves. (William Somerset Maugham)

The second is the principle of Maximizing Mutual Profits (MMP principle). To people who are willing to co-operate, this principle leads to steady progress.

Chapter 1

Introduction

1.1 Motivation

In 1948, Claude E. Shannon established the maximum rate at which information can be transmitted reliably over a noisy channel [1]. The mathematical setup is shown in Figure 1.1, where the channel is modeled as a probabilistic mapping from the input to the output, and the encoder and decoder are to be designed. Under this setup, Shannon proved a remarkable "phase transition" result: There is a fundamental rate limit—referred to as the channel capacity—under which one can design the encoder and decoder to achieve an arbitrarily small probability of error, but above which the probability of error is bounded away from zero (i.e., it cannot be made arbitrarily small no matter how we design the encoder and decoder) [1].

Shannon's channel coding theorem consists of two parts. The *achievability* part says that the probability of error can be made arbitrarily small for any rate below the channel capacity. The *converse* part states that the probability of error is bounded away from zero for any rate above the capacity. While the converse part applies to *any* decoder, the achievability part often involves several specific decoders, such as the maximum-likelihood (ML) decoder [2, p.37] and the joint typicality decoder [1][3, p.199]. These decoders, together with a random coding argument where the encoder generates independent and identically distributed (i.i.d.) codewords according to some codeword distribution, are used to prove the existence of good codes (without explicitly constructing them).

Practical communication systems are subject to complexity constraint. To control the computational complexity of encoding and decoding operations, codes with (algebraic) structures are used

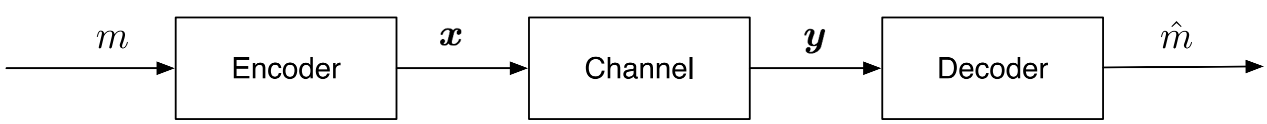

Figure 1.1: The model of a point-to-point communication system.

in practice. This motivates a study of structured codes, such as linear codes [4] and lattice codes [5–7]. In the sequel, we formally present the system setup and then discuss the use of structured codes in this setup.

1.2 System Setup

Here we describe Shannon's mathematical model of a point-to-point communication system depicted in Figure 1.1. Let $\mathcal{X}$ and $\mathcal{Y}$ denote the input and output alphabets, respectively. The channel maps an input sequence (of length n) $\boldsymbol{x} = (x_1, \ldots, x_n)$ to an output sequence (of length n) $\boldsymbol{y} = (y_1, \ldots, y_n)$ in a symbol-by-symbol manner. For example, when $\mathcal{X}$ and $\mathcal{Y}$ are finite, the conditional probability for the channel to output $\boldsymbol{y} \in \mathcal{Y}^n$ given $\boldsymbol{x} \in \mathcal{X}^n$ is

$$p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} p(y_i|x_i),$$

where $p(y|x)$ is a conditional probability mass function (pmf). This channel model is called a discrete memoryless channel (DMC). When $\mathcal{X}$ and $\mathcal{Y}$ are continuous alphabets, conditional probability density function (pdf) $f(y|x)$ should be used instead of $p(y|x)$. In particular, when

$$f(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-x)^2}{2\sigma^2}},$$

the corresponding channel model is called an additive white Gaussian noise (AWGN) channel.

The encoder maps a message $m \in \{1, \ldots, M\}$ to its corresponding codeword $\boldsymbol{x}(m)$ from a codebook $\mathsf{C} = \{\boldsymbol{x}(1), \ldots, \boldsymbol{x}(M)\}$. The decoder receives an output sequence $\boldsymbol{y}$ from the channel, and finds an "estimate" $\hat{m}$ of m according to certain decoding rule (such as ML decoding and joint typicality decoding).

We say an error occurs if $\hat{m} \neq m$ and denote this error probability as

$$\mathsf{P}_e(m; \mathsf{C}) \triangleq \mathsf{P}(\hat{m} \neq m),$$

where the randomness comes from the channel noise. We define the average error probability as

$$\mathsf{P}_e(\mathsf{C}) \triangleq \frac{1}{M} \sum_{m=1}^{M} \mathsf{P}_e(m; \mathsf{C}).$$

A rate R is said to be *achievable* if there exists a sequence of codebooks $\mathsf{C}^{(n)}$ of length n and size $M^{(n)}$ such that $M^{(n)} \geq 2^{nR}$ and $\mathsf{P}_e\left(\mathsf{C}^{(n)}\right) \to 0$ as $n \to \infty$. Achievable rates are often derived using a random coding argument. For a DMC with $p(y|x)$, we can fix a pmf $p(x)$ and construct a random i.i.d. ensemble in which each symbol of each codeword is generated independently according to $p(x)$. More specifically, we randomly and independently generate $M^{(n)} = \lceil 2^{nR} \rceil$ codewords $\boldsymbol{x}(m)$ for $m \in \{1, \ldots, M^{(n)}\}$, each according to $p(\boldsymbol{x}) = \prod_{i=1}^{n} p(x_i)$. Hence, the probability of generating a particular codebook $\mathsf{C}^{(n)}$ in the ensemble is

$$p\left(\mathsf{C}^{(n)}\right) = \prod_{m=1}^{M^{(n)}} p\left(\boldsymbol{x}(m)\right).$$

The key idea behind Shannon's random coding argument is the following. Although the error probability $\mathsf{P}_e\left(\mathsf{C}^{(n)}\right)$ for a particular codebook $\mathsf{C}^{(n)}$ is often hard to evaluate, the expected error probability averaged over all the codebooks in the ensemble is much simpler to analyze. In other words, random coding argument is an instance of the probabilistic method [8]. Using random coding argument, Shannon proved that random i.i.d. ensembles achieve both DMC capacity and AWGN channel capacity under joint typicality decoding in his 1948 paper [1].

1.3 Structured Codes

Instead of random i.i.d. ensembles, we can make use of random structured ensembles (such as random linear codes and random lattice codes) for the achievability proof. For example, Elias used random linear codes to establish the achievable rate for the binary symmetric channel (which is a special case of DMC) in 1955 [9]. Perhaps surprisingly, in their seminal work [10], Körner and Marton demonstrated that random linear codes yield better achievable rates than random i.i.d. ensembles for a multi-user source coding problem. Modern developments along this direction include coding problems from relay networks [11–20], interference channels [21–28], distributed source coding [29–33], and physical-layer secrecy [34–36], where random structured codes achieve better rates than random i.i.d. codes.

The use of random structured codes is also of practical value. For instance, random linear codes allow for computationally efficient encoding (since the encoding operation is essentially a matrix-vector multiplication), and random lattice codes allow for lattice decoding (which enjoys lower complexity than ML decoding and joint typicality decoding). Hence, the following two questions naturally arise

1. Can random linear codes achieve the DMC capacity?

2. Can random lattice codes achieve the AWGN channel capacity?

Unlike random i.i.d. codes, random structured codes are much less well understood. For example, it is only recently that Padakandla and Pradhan have demonstrated nested linear code ensembles achieve DMC capacity under joint typicality encoding and decoding [28, 37, 38]. In an independent work, Miyake and Muramatsu showed that nested linear code ensembles with special structures based on sparse matrices can also achieve DMC capacity under ML decoding [39–41]. In 2004, Erez and Zamir showed that nested lattice code ensembles achieve the AWGN channel capacity under lattice encoding and decoding [42]. See [42–50] for a history of this long-standing problem and Zamir's book [51] for a survey of recent results.

Despite these exciting developments, the achievability proofs associated with random structured codes are sometimes involved, making them much less accessible than their counterparts—random i.i.d. codes. Very recently, several attempts have been made towards simplifying the proofs related to random nested linear/lattice codes [52–54]. In this thesis, we will review these new developments and simplifications, with a particular focus on presenting a unified approach based on elementary probability, linear algebra, and number theory.

In the meanwhile, lattices used in the previous achievability proofs can rarely solve problems related to fading channels. Algebraic number theory turns out to be a very useful mathematical tool that enables the design of good lattice codes for fading channels. The lattice codes constructed using algebraic number theory (known as algebraic lattice codes) have good diversity and product distance [55]. In [56], algebraic lattice codes are used to achieve the ergodic fading channel capacity under Gaussian shaping. Very recently, the same authors also applied algebraic lattice codes to the Compute-and-Forward over compound fading channels [57]. However, whether the capacity is achievable under lattice encoding and lattice decoding remains an open problem. In this thesis, this problem is not tackled but we will take a minor step by showing the lattice codes of this kind could achieve the AWGN channel capacity by adopting the unified approach utilized in nested linear/lattice codes.

1.4 Organization of the Thesis

In Chapter 1, we introduce the model of the communication system and the motivation of using nested linear/lattice codes. In Chapter 2, we present definitions related to nested linear/lattice codes and introduce several elementary results from number theory that we use in our proofs. In Chapter 3, we prove that nested linear codes achieve the DMC channel capacity. In Chapter 4, we prove that nested lattice codes achieve the AWGN channel capacity. We make a particular effort in keeping these two proofs in parallel. In Chapter 5, we extend our techniques to lattice constructed using the algebraic number theory. We first briefly introduce a generalized version of construction A from [58] and then use it to construct AWGN-capacity-achieving lattice codes from the number field $\mathbb{Z}[i]$.

1.5 Notations

We closely follow the notations in [59]. We use the notation $\mathbb{F}, \mathbb{Q}, \mathbb{R}, \mathbb{F}_{\mathsf{q}}$ to denote a (general) field, the rational numbers, the real numbers, and the field of order q, respectively. We use $\mathcal{X}, \mathcal{Y}$ to denote the alphabets. We use lowercase letters $x, y, ...$ to denote constants. We use bold lowercase letters $\boldsymbol{x}, \boldsymbol{y}, ...$ to denote constant row vectors. The i-th component of $\boldsymbol{x}$ is denoted as x_i. An all-zero vector $(0, \dots, 0)$ with a specified dimension is denoted as $\mathbf{0}$. The i-th unit vector is denoted as $\boldsymbol{e}_i$. We use uppercase, sans-serif font letters to denote constant matrix and codebooks, e.g., a linear code C, and a matrix $\mathsf{G} \in \mathbb{F}_{\mathsf{q}}^{k \times n}$. We use uppercase letters $X, Y, \dots$ to denote random variables. We use bold uppercase letters $\boldsymbol{X}, \boldsymbol{Y}$ to denote random row vectors. The i-th component of $\boldsymbol{X}$ is denoted as X_i. We use bold, uppercase, sans-serif font letters to denote random matrix, e.g., a random linear code $\boldsymbol{\mathsf{C}}$ and a random matrix $\mathbf{G}$. As for the notations for the algebraic number theory, we use K to denote a algebraic number field and $\mathcal{O}_K$ to denote its ring of integers. The ideals of $\mathcal{O}_K$ are denoted by gothic font letters as $\mathfrak{p}, \mathfrak{a}$. A summary of our key notations is provided in the list of symbols at the beginning of this thesis.

Chapter 2

Preliminaries

2.1 Nested Linear Codes

An (n, k) *linear code* over $\mathbb{F}_\mathsf{q}$ is a k-dimensional subspace of the vector space $\mathbb{F}_\mathsf{q}^n$. Such a code can be expressed as

$$\mathsf{C} = \{\boldsymbol{a}\mathsf{G} : \boldsymbol{a} \in \mathbb{F}_\mathsf{q}^k\}$$

for some full-rank matrix $\mathsf{G} \in \mathbb{F}_\mathsf{q}^{k \times n}$ (called a *generator matrix* of C).

A *nested linear code* is a pair of linear codes $(\mathsf{C}_f, \mathsf{C}_c)$ such that $\mathsf{C}_c \subset \mathsf{C}_f$, i.e., each codeword of C_c is also a codeword of C_f. For convenience, C_f is called the *fine code* and C_c is called the *coarse code*. A *coset* of C_c in C_f is defined as

$$\boldsymbol{c}_f + \mathsf{C}_c = \{\boldsymbol{c}_f + \boldsymbol{c} : \boldsymbol{c} \in \mathsf{C}_c\},$$

where $\boldsymbol{c}_f$ is some codeword of C_f. Two cosets are either identical or disjoint [60]. The number of (distinct) cosets of C_c in C_f is called the *index* of C_c in C_f and is denoted by $[\mathsf{C}_f : \mathsf{C}_c]$. By Lagrange's theorem [60],

$$[\mathsf{C}_f : \mathsf{C}_c] = \frac{|\mathsf{C}_f|}{|\mathsf{C}_c|},$$

where $|\mathsf{C}_f|$ and $|\mathsf{C}_c|$ denote the cardinalities of C_f and C_c, respectively.

Suppose that a nested linear code consists of an (n, k_f) fine code C_f and an (n, k_c) coarse code C_c. Then the index $[\mathsf{C}_f : \mathsf{C}_c]$ is $\mathsf{q}^{k_f - k_c}$, since $|\mathsf{C}_f| = \mathsf{q}^{k_f}$ and $|\mathsf{C}_c| = \mathsf{q}^{k_c}$. Moreover, there exist two generator matrices $\mathsf{G}_f \in \mathbb{F}_\mathsf{q}^{k_f \times n}$ and $\mathsf{G}_c \in \mathbb{F}_\mathsf{q}^{k_c \times n}$ for C_f and C_c, respectively, such that

$$\mathsf{G}_f = \begin{bmatrix} \mathsf{G}_c \\ \mathsf{G}' \end{bmatrix},$$

where G' is a matrix of size $(k_f - k_c) \times n$.

2.2 Nested Lattice Codes

A *lattice* is a discrete subgroup (under vector addition) of $\mathbb{R}^n$. Any (full-rank) lattice Λ in $\mathbb{R}^n$ can be expressed in terms of some (full-rank) $n \times n$ generator matrix $\mathsf{G}_\Lambda \in \mathbb{R}^{n\times n}$ as

$$\Lambda = \{\boldsymbol{a}\mathsf{G}_\Lambda : \boldsymbol{a} \in \mathbb{Z}^n\}.$$

That is, Λ is the set of all integer combinations of the rows of G_Λ.

A *nearest neighbour quantizer* $Q_\Lambda : \mathbb{R}^n \to \Lambda$ associated with the lattice Λ maps a vector in $\mathbb{R}^n$ to the closest lattice point

$$Q_\Lambda(\boldsymbol{x}) = \arg\min_{\boldsymbol{\lambda}\in\Lambda} \|\boldsymbol{x} - \boldsymbol{\lambda}\|, \tag{2.1}$$

where ties in (2.1) are broken systematically. The *Voronoi region* of Λ, denoted by $\mathcal{V}(\Lambda)$, is the set of all vectors in $\mathbb{R}^n$ which are quantized to $\mathbf{0}$, i.e., $\mathcal{V}(\Lambda) = \{\boldsymbol{x} \in \mathbb{R}^n : Q_\Lambda(\boldsymbol{x}) = \mathbf{0}\}$. The volume of the Voronoi region is denoted by $V(\Lambda)$.

The modulo-Λ operation is ‘defined as

$$\boldsymbol{x} \bmod \Lambda = \boldsymbol{x} - Q_\Lambda(\boldsymbol{x})$$

A *nested lattice* is a pair of lattices (Λ_c, Λ_f) such that $\Lambda_c \subset \Lambda_f$. Similar to nested linear codes, Λ_f is called the *fine lattice* and Λ_c is called the *coarse lattice*. A coset of Λ_c in Λ_f is defined as

$$\boldsymbol{\lambda}_f + \Lambda_c = \{\boldsymbol{\lambda}_f + \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \Lambda_c\}.$$

A *nested lattice code* $\mathcal{L}(\Lambda_c, \Lambda_f)$ consists of the lattice points of Λ_f in the Voronoi region $\mathcal{V}(\Lambda_c)$, i.e.,

$$\mathcal{L}(\Lambda_c, \Lambda_f) = \Lambda_f \cap \mathcal{V}(\Lambda_c).$$

For this reason, $\mathcal{L}(\Lambda_c, \Lambda_f)$ is also known as a Voronoi codebook. The number of codewords in $\mathcal{L}(\Lambda_c, \Lambda_f)$ is

$$|\mathcal{L}(\Lambda_c, \Lambda_f)| = \frac{V(\Lambda_c)}{V(\Lambda_f)}.$$

Intuitively, each lattice point of Λ_f “occupies” a Voronoi region of volume $V(\Lambda_f)$, and so the number of lattice points inside $\mathcal{V}(\Lambda_c)$ is $V(\Lambda_c)/V(\Lambda_f)$.

There is an alternative characterization of nested lattice codes: $\mathcal{L}(\Lambda_c, \Lambda_f)$ *consists of the short-*

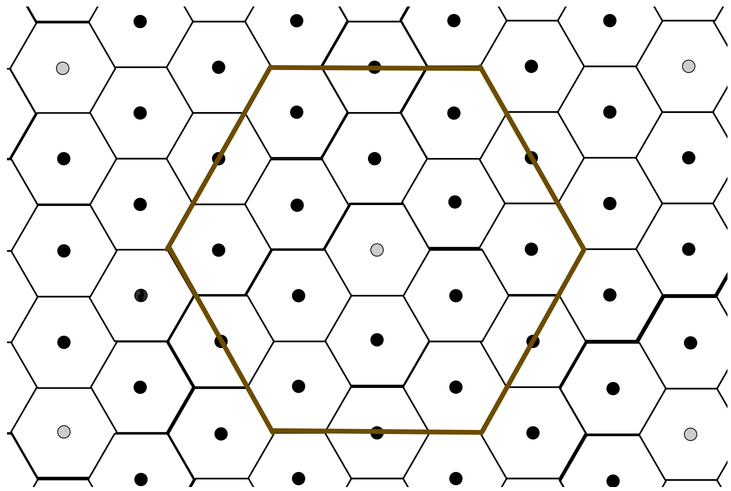

Figure 2.1: An example of nested lattices

est vectors of distinct cosets. To see this, for each coset $\boldsymbol{\lambda}_f + \Lambda_c$, let us take a particular coset representative $\boldsymbol{\lambda}_f - Q_{\Lambda_c}(\boldsymbol{\lambda}_f)$. First, $\boldsymbol{\lambda}_f - Q_{\Lambda_c}(\boldsymbol{\lambda}_f)$ is the shortest vector in the coset $\boldsymbol{\lambda}_f + \Lambda_c$ by the definition of $Q_{\Lambda_c}(\cdot)$. Second, $\boldsymbol{\lambda}_f - Q_{\Lambda_c}(\boldsymbol{\lambda}_f)$ is in the Voronoi region $\mathcal{V}(\Lambda_c)$ of Λ_c.

In Fig. 2.2, we present an example of nested lattices. Black (grey) points belong to the fine (coarse) lattice. The small (large) hexagon area is the Voronoi region of the fine (coarse) lattice. The lattice points inside the large hexagon form the Voronoi codebook (the ties on the boundaries are broken systematically). There are 16 lattice points in the codebook due to the tie breaking. Also note that the volume of the large hexagon is 16 times of the volume of the small one.

2.3 Nested Construction A

A nested lattice code can be constructed from a nested linear code. Consider two linear codes C_1 and C_2 over the field $\mathbb{Z}_\mathsf{p} = \{0, 1, \ldots, \mathsf{p}-1\}$, where each code C_i is determined by a (full-rank) $k_i \times n$ generator matrix G_i for $i = 1, 2$. Suppose that the generator matrices are related as

$$\mathsf{G}_1 = \begin{bmatrix} \mathsf{G}_2 \\ \mathsf{G}' \end{bmatrix}, \tag{2.2}$$

where G' is a matrix of size $(k_1 - k_2) \times n$. Clearly, we have $\mathsf{C}_2 \subset \mathsf{C}_1 \subset \mathbb{Z}_\mathsf{p}^n$. By "lifting" these linear codes to $\mathbb{Z}^n$ via Construction A, we obtain two lattices

$$\Lambda_1 = \{\boldsymbol{x} \in \mathbb{Z}^n : \boldsymbol{x} \bmod \mathsf{p} \in \mathsf{C}_1\}$$

and

$$\Lambda_2 = \{\boldsymbol{x} \in \mathbb{Z}^n : \boldsymbol{x} \bmod \mathsf{p} \in \mathsf{C}_2\}$$

with $\Lambda_2 \subset \Lambda_1 \subset \mathbb{Z}^n$.

Finally, we apply some positive scaling factor γ to obtain a fine lattice

$$\Lambda_f = \gamma\Lambda_1 \triangleq \{\gamma\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \Lambda_1\}$$

and a coarse lattice

$$\Lambda_c = \gamma\Lambda_2 \triangleq \{\gamma\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \Lambda_2\}$$

with $\Lambda_c \subset \Lambda_f \subset \gamma\mathbb{Z}^n$. The volumes of the Voronoi regions of Λ_f and Λ_c are $V(\Lambda_f) = \gamma^n \mathsf{p}^{n-k_1}$ and $V(\Lambda_c) = \gamma^n \mathsf{p}^{n-k_2}$, respectively.

To facilitate encoding and decoding operations, we "label" each (discrete) point of $\gamma\mathbb{Z}^n$ as follows. Let $\varphi : \gamma\mathbb{Z}^n \to \mathbb{Z}_\mathsf{p}^n$ be a map from points in $\gamma\mathbb{Z}^n$ to vectors in $\mathbb{Z}_\mathsf{p}^n$ given by

$$\varphi(\boldsymbol{x}) = \frac{1}{\gamma}\boldsymbol{x} \bmod \mathsf{p}.$$

Clearly, a point $\boldsymbol{x}$ is in Λ_f (or Λ_c, respectively) if and only if its label $\varphi(\boldsymbol{x})$ is a codeword in C_1 (or C_2, respectively). Moreover, the map φ is homomorphic, i.e.,

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \gamma\mathbb{Z}^n,\ \varphi(\boldsymbol{x} + \boldsymbol{y}) = \varphi(\boldsymbol{x}) + \varphi(\boldsymbol{y}).$$

A visualization of $\varphi(\cdot)$ when $\mathsf{p} = 5$ is provided in Fig. 2.2. The labels of the points in $\gamma\mathbb{Z}^n$ can be obtained by periodically shifting the labels in the rectangle.

It is also convenient to define an inverse operation that maps a vector in $\mathbb{Z}_\mathsf{p}^n$ to a point in $\gamma\mathbb{Z}^n$. This can be done through an embedding map $\tilde{\varphi} : \mathbb{Z}_\mathsf{p}^n \to \gamma\mathbb{Z}^n$: for any $\boldsymbol{c}$ in $\mathbb{Z}_\mathsf{p}^n$, we choose a point $\boldsymbol{x}$ in $\gamma\mathbb{Z}^n$ of the shortest Euclidean norm such that $\varphi(\boldsymbol{x}) = \boldsymbol{c}$. Clearly, such a point $\boldsymbol{x} = \tilde{\varphi}(\boldsymbol{c})$ must live in the grid $\gamma\mathbb{Z}^n \cap [-\frac{\gamma\mathsf{p}}{2}, \frac{\gamma\mathsf{p}}{2}]^n$. Reader can view this from a more algebraic perspective. The kernel $\ker\varphi$ is a subgroup of $\mathbb{Z}_\mathsf{p}^n$ and thus is also a lattice. All the points that will be mapped to $\boldsymbol{c}$ by φ belong to $\tilde{\varphi}(\boldsymbol{c}) + \ker\varphi$. The point of the shortest Euclidean norm among $\tilde{\varphi}(\boldsymbol{c}) + \ker\varphi$ must belong to the Voronoi region of $\ker\varphi$, which is exactly $[-\frac{\gamma\mathsf{p}}{2}, \frac{\gamma\mathsf{p}}{2}]^n$. For convenience, we denote $\ker\varphi$ as Λ_p.

In fact, the embedding map $\tilde{\varphi}$ can be viewed as a *Euclidean embedding* for the vector space $\mathbb{Z}_\mathsf{p}^n$,

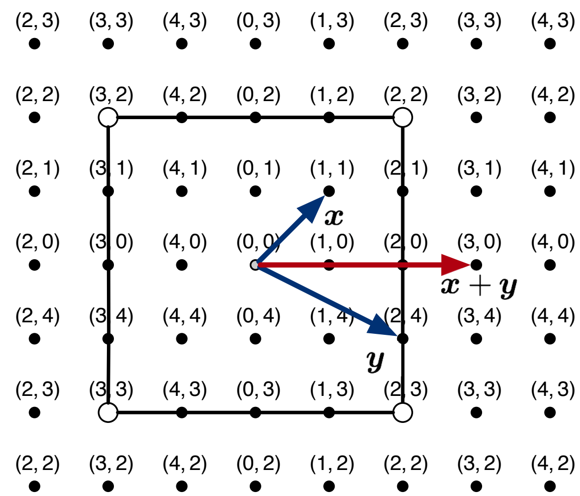

Figure 2.2: A visualization of $\varphi(\cdot)$ when $\mathsf{p} = 5$.

which connects the nested lattice codes with the underlying nested linear codes.

2.4 Useful Lemmas

In this section, we present some useful lemmas related to nested linear/lattice codes. Readers who are not familiar with entropy and typical sequence are also encouraged to read Appendix A and Appendix B.

Let $\mathbf{G}$ be a random matrix uniform over $\mathbb{Z}_{\mathsf{p}}^{k\times n}$, i.e., each entry of $\mathbf{G}$ is drawn uniformly and independently from $\mathbb{Z}_{\mathsf{p}}$.

Lemma 1(Uniformity): For any fixed non-zero vector $\boldsymbol{a}$, $\boldsymbol{a}\mathbf{G}$ is uniform over $\mathbb{Z}_{\mathsf{p}}^{n}$.

Proof. We leave it as an exercise to our readers. □

Lemma 2(Linear independence ⇒ statistical independence): For any linearly independent vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, the random vectors $\boldsymbol{a}\mathbf{G}$ and $\boldsymbol{b}\mathbf{G}$ are statistically independent.

Proof. Since $\boldsymbol{a}$ and $\boldsymbol{b}$ are linearly independent, there exists a full rank matrix $\mathsf{A} \in \mathbb{Z}_{\mathsf{p}}^{k\times n}$ whose first row vector is $\boldsymbol{a}$, and the second row vector is $\boldsymbol{b}$, i.e., $\boldsymbol{e}_1\mathsf{A} = \boldsymbol{a}, \boldsymbol{e}_2\mathsf{A} = \boldsymbol{b}$. For any fixed vectors $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathbb{Z}_{\mathsf{p}}^{k}$, $\boldsymbol{e}_1\mathsf{A}\mathbf{G} = \boldsymbol{a}\mathsf{G} = \boldsymbol{c}_1$ and $\boldsymbol{e}_2\mathsf{A}\mathbf{G} = \boldsymbol{b}\mathbf{G} = \boldsymbol{c}_2$, if and only if the first and second row vector of AG are $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$. Let $S_{\boldsymbol{c}_1,\boldsymbol{c}_2} = \{\mathsf{B} \in \mathbb{Z}_{\mathsf{p}}^{k\times n} \mid \boldsymbol{e}_1\mathsf{B} = \boldsymbol{c}_1, \boldsymbol{e}_2\mathsf{B} = \boldsymbol{c}_2\}$, then $|S_{\boldsymbol{c}_1,\boldsymbol{c}_2}| = \mathsf{p}^{(k-2)n}$. Hence

$$\mathsf{P}(\boldsymbol{a}\mathsf{G} = \boldsymbol{c}_1, \boldsymbol{b}\mathsf{G} = \boldsymbol{c}_2) = \sum_{\mathsf{B}\in S_{\boldsymbol{c}_1,\boldsymbol{c}_2}} \mathsf{P}(\mathbf{G} = \mathsf{A}^{-1}\mathsf{B}) = \frac{1}{p^{2n}}$$

Hence, $\mathsf{P}(\boldsymbol{a}\mathbf{G} = \boldsymbol{c}_1, \boldsymbol{b}\mathbf{G} = \boldsymbol{c}_2) = \mathsf{P}(\boldsymbol{a}\mathbf{G} = \boldsymbol{c}_1)\,\mathsf{P}(\boldsymbol{b}\mathbf{G} = \boldsymbol{c}_2)$, which means $\boldsymbol{a}\mathbf{G}$ and $\boldsymbol{b}\mathbf{G}$ are statistically independent. □

Lemma 3(Crypto lemma): Let Λ be a lattice. Let $\boldsymbol{D}$ be a random variable uniformly distributed over $\mathcal{V}(\Lambda)$. Let $\boldsymbol{T}$ be a random variable over $\mathcal{V}(\Lambda)$, and is independent from $\boldsymbol{D}$, then $\boldsymbol{X} = \boldsymbol{D} + \boldsymbol{T} \bmod \Lambda$ is uniformly distributed over $\mathcal{V}(\Lambda)$, and is independent from $\boldsymbol{T}$.

Remark 1: This lemma is a discrete parallel of [42, Lemma 1].

Proof. Note that $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{T} = \boldsymbol{t}) = \mathsf{P}(\boldsymbol{D} = [\boldsymbol{x} - \boldsymbol{t}] \bmod \Lambda \mid \boldsymbol{T} = \boldsymbol{t})$. By the fact that $\boldsymbol{D}$ and $\boldsymbol{T}$ are independent, we obtain $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{T} = \boldsymbol{t}) = \mathsf{P}(\boldsymbol{D} = [\boldsymbol{x} - \boldsymbol{t}] \bmod \Lambda)$. Since $\boldsymbol{D}$ is uniform over $\mathcal{V}(\Lambda)$, $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{T} = \boldsymbol{t})$ is constant for all possible combinations of $\boldsymbol{x}$ and $\boldsymbol{t}$. Hence, $\boldsymbol{X}$ is uniformly distributed over $\mathcal{V}(\Lambda)$, and is independent from $\boldsymbol{T}$. □

Let $\mathcal{B}(\boldsymbol{s}, r)$ denote a ball of radius $r > 0$ centered at the point $\boldsymbol{s} \in \mathbb{R}^n$, i.e., $\mathcal{B}(\boldsymbol{s}, r)$ is the set $\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{s}\| \leq r\}$. For convenience, we denote $\mathcal{B}(\boldsymbol{0}, r)$ as $\mathcal{B}(r)$. The volume of $\mathcal{B}(r)$ is given by $r^n V_n$, where V_n is the volume of the unit-radius ball.

Lemma 4(Lattice points inside a ball [58, Lemma 4]): Let $\Lambda \in \mathbb{R}^n$. Let $l = \sup_{\boldsymbol{x} \in \mathcal{V}(\Lambda)} \|\boldsymbol{x}\|$, for any $r > l$, we have

$$(r - l)^n \frac{V_n}{V(\Lambda)} \leq |\mathcal{B}(r) \cap \Lambda| \leq (r + l)^n \frac{V_n}{V(\Lambda)}$$

Specifically, we can choose Λ as $\mathbb{Z}^n$. For $\mathbb{Z}^n$, we have $l = \frac{\sqrt{n}}{2}$ and $\mathcal{V}(\mathbb{Z}^n) = 1$. We then obtain the following lemma

Lemma 5(Integer points inside a ball [53, Lemma 1]): For any $\boldsymbol{s} \in \mathbb{R}^n$, the number of points of $\mathbb{Z}^n$ inside $\boldsymbol{s} + \mathcal{B}(r)$ can be bounded as

$$V_n \left(\max\left\{ r - \frac{\sqrt{n}}{2}, 0 \right\} \right)^n \leq |\mathbb{Z}^n \cap \mathcal{B}(\boldsymbol{s}, r)| \leq V_n \left(r + \frac{\sqrt{n}}{2} \right)^n .$$

Lemma 6(Bertrand's postulate [61]): For any integer n that's larger than 3, there exists a prime p such that $n < \mathsf{p} < 2n - 2$ and $\mathsf{p} \bmod 4 = 1$.

Chapter 3

Achievable Rate of Nested Linear Codes

We begin with the case of a pre-determined nested linear code, to get readers familiar with the encoding and decoding methods. We then introduce a random ensemble of nested linear codes. We will show the average error probability of this ensemble will vanish as the length of codewords goes to infinity and the achievable rate is close to the desired channel capacity. By the above facts, it's then clear that we find some pre-determined codebooks that achieve the channel capacity.

3.1 The Case of a Pre-Determined Nested Linear Code

Codebook generation. Given a pair of linear codes $(\mathsf{C}_f, \mathsf{C}_c)$ and a dither vector $\boldsymbol{d} \in \mathbb{F}_{\mathsf{q}}^n$, we construct a codebook whose codewords are shifted cosets of the form $\{\boldsymbol{c}_f + \boldsymbol{d} + \mathsf{C}_c : \boldsymbol{c}_f \in \mathsf{C}_f\}$. The number of (distinct) codewords is $[\mathsf{C}_f : \mathsf{C}_c]$, which doesn't depend on the dither vector $\boldsymbol{d}$. These codewords can be expressed using generator matrices as follows.

Let $\mathsf{G}_f \in \mathbb{F}_{\mathsf{q}}^{k_f \times n}$ and $\mathsf{G}_c \in \mathbb{F}_{\mathsf{q}}^{k_c \times n}$ be two generator matrices for C_f and C_c, respectively, such that

$$\mathsf{G}_f = \begin{bmatrix} \mathsf{G}_c \\ \mathsf{G}' \end{bmatrix}.$$

Then all the codewords (i.e., the shifted cosets) can be expressed as

$$\left\{ \boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c : \boldsymbol{m} \in \mathbb{F}_{\mathsf{q}}^{k_f - k_c} \right\}.$$

Note that there is a one-to-one correspondence between the vectors in $\mathbb{F}_{\mathsf{q}}^{k_f - k_c}$ and the shifted cosets of C_c. Hence, $\boldsymbol{m}$ can be viewed as the "index" of the shifted coset $\boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c$, and the codebook contains $\mathsf{q}^{k_f - k_c}$ (distinct) codewords.

Encoding. To send a message vector $\boldsymbol{m} \in \mathbb{F}_{\mathsf{q}}^{k_f - k_c}$, the encoder first finds an "information-carrying" shifted coset $\boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c$. We also define the following typical set

$$\mathcal{T}_{\epsilon'}^{(n)}(X) = \{\boldsymbol{x} : |\pi(x \mid \boldsymbol{x}) - p_X(x)| \leq \epsilon p_X(x) \quad \text{for all } x \in \mathcal{X}\},$$

where the distribution $p_X(\cdot)$ can be arbitrary distribution. However, in order to achieve the channel capacity, we will choose it as the distribution that maximize the mutual information between the channel input and channel output. The encoder then checks the intersection

$$\boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon'}^{(n)}(X).$$

If the intersection is nonempty, the encoder transmits a vector $x \in \mathbb{F}_{\mathsf{q}}^n$ chosen uniformly at random from the intersection. Otherwise, the encoder declares a failure and then transmits a vector $\boldsymbol{x} \in \mathbb{F}_{\mathsf{q}}^n$ chosen uniformly at random from the shifted coset $\boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c$ (which is not in $\mathcal{T}_{\epsilon'}^{(n)}(X)$).

Decoding. Upon receiving $\boldsymbol{y} \in \mathbb{F}_{\mathsf{q}}^n$, the decoder searches for a unique index $\hat{\boldsymbol{m}} \in \mathbb{F}_{\mathsf{q}}^{k_f - k_c}$ such that the corresponding shifted coset has a non-empty intersection with $\mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y})$. The set $\mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y})$ consists of all the good codewords that are close to $\boldsymbol{y}$ and is defined as

$$\mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y}) = \{\boldsymbol{x} : (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{T}_{\epsilon}^{(n)}(X, Y)\},$$

where $\mathcal{T}_{\epsilon}^{(n)}(X, Y)$ is the typical set defined with respect to the joint distribution $p_{X,Y}(\cdot,\cdot)$ which is induced by the input distribution $p_X(\cdot)$ and the channel $p_{Y|X}(\cdot)$. In other words, we will find $\hat{\boldsymbol{m}}$ such that

$$\hat{\boldsymbol{m}}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y}) \neq \emptyset.$$

If there is none or more than one such vector, the decoder declares a failure.

Analysis. For any given message vector $\boldsymbol{m}$, we say the decoding is successful if the unique index $\hat{\boldsymbol{m}} = \boldsymbol{m}$. This occurs if all of the following events happen

- $\boldsymbol{m}\mathsf{G}' + \boldsymbol{D} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon'}^{(n)}(X) \neq \emptyset$;
- $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{T}_{\epsilon}^{(n)}(X, Y)$ (which implies that $\boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y}) \neq \emptyset$);
- $\forall \boldsymbol{m}' \neq \boldsymbol{m} : \boldsymbol{m}'\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y}) = \emptyset$.

3.2 The Case of a Random Nested Linear Code

We then proceed to the case of a random nested linear code, which allows us to apply the probabilistic method.

Random codebook generation. Randomly generate a matrix $\mathbf{G}_f \in \mathbb{F}_{\mathsf{q}}^{k_f \times n}$ and a vector $\boldsymbol{D} \in \mathbb{F}_{\mathsf{q}}^n$ where each entry of $\mathbf{G}_f$ and $\boldsymbol{D}$ is drawn independently and uniformly from $\mathbb{F}_{\mathsf{q}}$. As before, let

$$\mathbf{G}_f = \begin{bmatrix} \mathbf{G}_c \\ \mathbf{G}' \end{bmatrix}.$$

If $\mathbf{G}_f$ is full rank, then $\mathbf{G}_c$ is also full rank and, in particular, they are valid generator matrices. In this case, the codebook consists of $\mathsf{q}^{k_f - k_c}$ shifted cosets of the form

$$\left\{ \boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \mathsf{C}_c : \boldsymbol{m} \in \mathbb{F}_{\mathsf{q}}^{k_f - k_c} \right\}.$$

If $\mathbf{G}_f$ is not full rank, we declare a codebook failure.

Encoding. The same as before.

Decoding. The same as before.

Analysis of the probability of error. For any given message vector $\boldsymbol{m}$, a successful decoding occurs upon receiving $\boldsymbol{Y}$ if all of the following events happen

- $\mathbf{G}_f$ is full rank;
- $\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon'}^{(n)}(X) \neq \emptyset$;
- $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathcal{T}_{\epsilon}^{(n)}(\boldsymbol{X}, \boldsymbol{Y})$;
- $\forall \boldsymbol{m}' \neq \boldsymbol{m}, \boldsymbol{l} : (\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c, \boldsymbol{Y}) \notin \mathcal{T}_{\epsilon}^{(n)}(X, Y)$.

To conduct the error analysis, we define the following events

- $\mathcal{E}_1 = \{\mathbf{G}_f \text{ is not full rank}\}$;
- $\mathcal{E}_2(\boldsymbol{m}) = \{\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon'}^{(n)}(X) = \emptyset\}$;
- $\mathcal{E}_3(\boldsymbol{m}) = \{(\boldsymbol{X}, \boldsymbol{Y}) \notin \mathcal{T}_{\epsilon}^{(n)}(X, Y)\}$;
- $\mathcal{E}_4(\boldsymbol{m}) = \{\exists \boldsymbol{m}' \neq \boldsymbol{m}, \boldsymbol{l} :\ (\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c, \boldsymbol{Y}) \in \mathcal{T}_{\epsilon}^{(n)}(X, Y)\}$.

Let $\mathsf{P}_e(\boldsymbol{m})$ be the error probability for message $\boldsymbol{m}$. Then, by the union bound, we have

$$\mathsf{P}_e(\boldsymbol{m}) \leq \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2(\boldsymbol{m})) + \mathsf{P}(\mathcal{E}_3(\boldsymbol{m})) + \mathsf{P}(\mathcal{E}_4(\boldsymbol{m})).$$

3.2.1 Analysis of the Codebook Failure

It is a well known result that

$$\mathsf{P}(\mathcal{E}_1) = 1 - \prod_{i=0}^{k_f-1} \left(1 - \frac{\mathsf{q}^i}{\mathsf{q}^n}\right).$$

Moreover, it is easy to show that

$$\mathsf{P}(\mathcal{E}_1) \leq \frac{1}{\mathsf{q}-1} \frac{1}{\mathsf{q}^{n-k_f}}.$$

Hence, $\mathsf{P}(\mathcal{E}_1) \to 0$ as $\mathsf{q} \to \infty$ or $(n - k_f) \to \infty$.

3.2.2 Analysis of the Encoding Failure

Bounding $\mathsf{P}(\mathcal{E}_2(\boldsymbol{m}))$

Note that $\mathcal{E}_2(\boldsymbol{m})$ is equivalent to

$$\sum_{\boldsymbol{l} \in \mathbb{F}_{\mathsf{q}}^{k_c}} \mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right) = 0.$$

Since $\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c$ is uniformly distributed over $\mathbb{F}_{\mathsf{q}}^n$, we have

$$\mathsf{E}\left(\mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right) = \frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}$$

and

$$\mathrm{Var}\left(\mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right) = \frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}\left(1 - \frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}\right).$$

Note that for any $\boldsymbol{l}' \neq \boldsymbol{l}$, $\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}'\mathbf{G}_c$ and $\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c$ are independent. Hence,

$$\mathsf{E}\left(\sum_{\boldsymbol{l} \in \mathbb{F}_{\mathsf{q}}^{k_c}} \mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right) = \mathsf{q}^{k_c} \frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}$$

and

$$\operatorname{Var}\left(\sum_{\boldsymbol{l}\in\mathbb{F}_{\mathsf{q}}^{k_c}}\mathbb{I}\left(\boldsymbol{m}\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c\in\mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right)=\frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}\mathsf{q}^{k_c}\left(1-\frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}\right).$$

Finally, by Chebyshev's inequality, we have

$$\begin{aligned}\mathsf{P}(\mathcal{E}_2(\boldsymbol{m}))&=\mathsf{P}\left(\sum_{\boldsymbol{l}\in\mathbb{F}_{\mathsf{q}}^{k_c}}\mathbb{I}\left(\boldsymbol{m}\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c\in\mathcal{T}_{\epsilon'}^{(n)}(X)\right)=0\right)\\&\le\frac{\operatorname{Var}\left(\sum_{\boldsymbol{l}\in\mathbb{F}_{\mathsf{q}}^{k_c}}\mathbb{I}\left(\boldsymbol{m}\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c\in\mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right)}{\mathsf{E}\left(\sum_{\boldsymbol{l}\in\mathbb{F}_{\mathsf{q}}^{k_c}}\mathbb{I}\left(\boldsymbol{m}\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c\in\mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right)^2}\\&\le\frac{\mathsf{q}^{n-k_c}}{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}.\end{aligned}$$

Bounding $\mathsf{P}(\mathcal{E}_3(\boldsymbol{m}))$

By the law of total probability, we have

$$\begin{aligned}&\mathsf{P}\left((\boldsymbol{X},\boldsymbol{Y})\notin\mathcal{T}_{\epsilon}^{(n)}(X,Y)\right)\\&=\mathsf{P}(\boldsymbol{X}\in\mathcal{T}_{\epsilon'}^{(n)}(X))\,\mathsf{P}((\boldsymbol{X},\boldsymbol{Y})\notin\mathcal{T}_{\epsilon}^{(n)}(X,Y)|\boldsymbol{X}\in\mathcal{T}_{\epsilon'}^{(n)}(X))\\&\quad+\mathsf{P}(\boldsymbol{X}\notin\mathcal{T}_{\epsilon'}^{(n)}(X))\,\mathsf{P}((\boldsymbol{X},\boldsymbol{Y})\notin\mathcal{T}_{\epsilon}^{(n)}(X,Y)|\boldsymbol{X}\notin\mathcal{T}_{\epsilon'}^{(n)}(X))\\&\le\mathsf{P}((\boldsymbol{X},\boldsymbol{Y})\notin\mathcal{T}_{\epsilon}^{(n)}(X,Y)|\boldsymbol{X}\in\mathcal{T}_{\epsilon'}^{(n)}(X))+\mathsf{P}(\boldsymbol{X}\notin\mathcal{T}_{\epsilon'}^{(n)}(X)).\end{aligned}$$

By the conditional typicality lemma [59, p. 27], $\mathsf{P}((\boldsymbol{X},\boldsymbol{Y})\notin\mathcal{T}_{\epsilon}^{(n)}(X,Y)|\boldsymbol{X}\in\mathcal{T}_{\epsilon'}^{(n)}(X))\to 0$, as $n\to\infty$. Finally, note that $\boldsymbol{X}\notin\mathcal{T}_{\epsilon'}^{(n)}(X)$ is equivalent to the event $\mathcal{E}_2(\boldsymbol{m})$. Hence, we obtain $\mathsf{P}\left((\boldsymbol{X},\boldsymbol{Y})\notin\mathcal{T}_{\epsilon}^{(n)}(X,Y)\right)\to 0$, as long as $\mathsf{P}(\mathcal{E}_2(\boldsymbol{m}))\to 0$.

3.2.3 Analysis of the Decoding Failure

By the union of events bound, we have

$$\mathsf{P}(\mathcal{E}_4(\boldsymbol{m}))\le\sum_{\boldsymbol{m}'\neq\boldsymbol{m}}\sum_{\boldsymbol{l}}\mathsf{P}((\boldsymbol{m}'\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c,\boldsymbol{Y})\in\mathcal{T}_{\epsilon}^{(n)}(X,Y)).$$

For each term, by the law of total probability, we have

$$\mathsf{P}((\boldsymbol{m}'\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c,\boldsymbol{Y})\in\mathcal{T}_\epsilon^{(n)}(X,Y))=\sum_{\boldsymbol{y}}\mathsf{P}(\boldsymbol{Y}=\boldsymbol{y})\,\mathsf{P}\left(\boldsymbol{m}'\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c\in\mathcal{T}_\epsilon^{(n)}(X\mid\boldsymbol{y})\middle|\boldsymbol{Y}=\boldsymbol{y}\right).$$

Note that, for any $\boldsymbol{m}'\neq\boldsymbol{m}$ and any $\boldsymbol{l}$, the random vector $\boldsymbol{m}'\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c$ is independent of the random shifted coset $\boldsymbol{m}\mathbf{G}'+\boldsymbol{D}+\mathsf{C}_c$. This implies that $\boldsymbol{m}'\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c$ is independent of $\boldsymbol{Y}$. Hence,

$$\mathsf{P}(\boldsymbol{m}'\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c\in\mathcal{T}_\epsilon^{(n)}(X\mid\boldsymbol{y})|\boldsymbol{Y}=\boldsymbol{y})=\mathsf{P}(\boldsymbol{m}'\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c\in\mathcal{T}_\epsilon^{(n)}(X\mid\boldsymbol{y})).$$

Since $\mathsf{P}(\boldsymbol{m}'\mathbf{G}'+\boldsymbol{D}+\boldsymbol{l}\mathbf{G}_c\in\mathcal{T}_\epsilon^{(n)}(X\mid\boldsymbol{y}))=\frac{|\mathcal{T}_\epsilon^{(n)}(X|\boldsymbol{y})|}{\mathsf{q}^n}$, we have

$$\begin{aligned}\mathsf{P}(\mathcal{E}_4(\boldsymbol{m})\mid\boldsymbol{Y}=\boldsymbol{y})&\leq\left(\mathsf{q}^{k_f-k_c}-1\right)\mathsf{q}^{k_c}\frac{|\mathcal{T}_\epsilon^{(n)}(X\mid\boldsymbol{y})|}{\mathsf{q}^n}\\&<\mathsf{q}^{k_f}\frac{|\mathcal{T}_\epsilon^{(n)}(X\mid\boldsymbol{y})|}{\mathsf{q}^n}.\end{aligned}$$

Hence, we have

$$\mathsf{P}(\mathcal{E}_4(\boldsymbol{m}))\leq\sum_{\boldsymbol{y}}\mathsf{P}(\boldsymbol{Y}=\boldsymbol{y})\frac{|\mathcal{T}_\epsilon^{(n)}(X\mid\boldsymbol{y})|}{\mathsf{q}^{n-k_f}}.$$

3.3 Analysis of the Error Probability

Our goal is to select k_c and k_f (as functions of n) such that

$$n-k_f\to\infty \tag{3.1}$$

$$\frac{\mathsf{q}^{n-k_c}}{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}\to 0 \tag{3.2}$$

$$\forall\boldsymbol{y}:\ \frac{|\mathcal{T}_\epsilon^{(n)}(X\mid\boldsymbol{y})|}{\mathsf{q}^{n-k_f}}\to 0. \tag{3.3}$$

We show in Appendix B that

$$\begin{aligned}|\mathcal{T}_{\epsilon'}^{(n)}(X)|&\geq(1-\epsilon')2^{n(1-\epsilon')H(X)},\\ \forall\boldsymbol{y}\in\mathcal{Y}^n:|\mathcal{T}_\epsilon^{(n)}(X\mid\boldsymbol{y})|&\leq 2^{n(1+\epsilon)H(X|Y)}.\end{aligned}$$

Let $\delta > 0$ be some constant. We choose $\mathsf{q}^{n-k_c} = 2^{n(1-\epsilon'-\delta)H(X)}$ and $\mathsf{q}^{n-k_f} = 2^{n(1+\epsilon+\delta)H(X|Y)}$. More precisely, we choose

$$k_c = \left\lceil n - \frac{(1-\epsilon'-\delta)H(X)}{\log_2 \mathsf{q}} n \right\rceil$$

and

$$k_f = \left\lfloor n - \frac{(1+\epsilon+\delta)H(X|Y)}{\log_2 \mathsf{q}} n \right\rfloor .$$

We can easily verify that conditions (3.1), (3.2) are satisfied. This implies the average error probability of the random ensemble we use is vanishing. In other words, there exists a non-zero portion of pre-determined codebooks in our ensemble that have vanishing error probability.

Finally, we calculate the achievable rate

$$\frac{1}{n} \log_2 \mathsf{q}^{k_f - k_c} \geq I(X;Y) - (\epsilon' + \delta)H(X) - (\epsilon + \delta)H(X|Y) - 2\frac{\log_2 \mathsf{q}}{n}.$$

Since ϵ, ϵ' and δ can be arbitrarily small, any rate below $I(X;Y)$ is achievable as $n \to \infty$. Since we can choose the distribution $p_X(\cdot)$ to be the one that maximizes $I(X;Y)$, the achievable rate then can be arbitrarily close to the channel capacity. Hence, we claim there exist pre-determined codebooks in our ensemble that achieve the channel capacity.

Chapter 4

Achievable Rate of Nested Lattice Codes

Similar to the case of nested linear codes, we begin with the case of a pre-determined nested linear code, to get readers familiar with the encoding and decoding methods. We then introduce a random ensemble of nested lattice codes. We will show the average error probability of this ensemble will vanish as the length of codewords goes to infinity and the achievable rate is close to the desired channel capacity. By the above facts, it's then clear that we find some pre-determined codebooks that achieve the channel capacity.

4.1 The Case of a Pre-Determined Nested Lattice Code

Codebook generation. Given a pair of lattice codes (Λ_f, Λ_c) and a dither vector $\boldsymbol{u} \in \mathbb{R}^n$, we construct a codebook whose codewords are shifted cosets of the form $\{\boldsymbol{\lambda}_f + \boldsymbol{u} + \Lambda_c : \boldsymbol{\lambda}_f \in \Lambda_c\}$. The number of codewords is $V(\Lambda_c)/V(\Lambda_f)$, which doesn't depend on the dither vector $\boldsymbol{u}$.

Suppose that the pair (Λ_f, Λ_c) is constructed via Nested Construction A using generating matrices $(\mathsf{G}_f, \mathsf{G}_c)$ and a scaling factor γ. Then all the codewords (i.e., the shifted cosets) can be expressed as

$$\left\{\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c : \boldsymbol{m} \in \mathbb{F}_{\mathsf{p}}^{k_f - k_c}\right\}.$$

Note that there is a one-to-one correspondence between the vectors in $\mathbb{F}_{\mathsf{p}}^{k_f - k_c}$ and the shifted cosets of Λ_c. Hence, $\boldsymbol{m}$ can be viewed as the "index" of the shifted coset $\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c$, and the codebook contains $\mathsf{p}^{k_f - k_c}$ (distinct) codewords.

Encoding. To send a message vector $\boldsymbol{m} \in \mathbb{F}_{\mathsf{p}}^{k_f - k_c}$, the encoder first finds an "information-carrying" shifted coset $\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c$. The encoder then transmits a shortest vector $\boldsymbol{x} \in \mathbb{R}^n$ in the shifted coset, i.e.,

$$\boldsymbol{x} = \tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} \mod \Lambda_c.$$

Decoding. The channel considered here is the AWGN channel, so instead of using typicality decoding as we did in last chapter, we will follow [62] to use lattice decoding. In other words, upon receiving $\boldsymbol{y} \in \mathbb{R}^n$, the decoder searches for a unique index $\hat{\boldsymbol{m}} \in \mathbb{F}_{\mathsf{p}}^{k_f - k_c}$ such that the distance between its corresponding shifted coset $\tilde{\varphi}(\hat{\boldsymbol{m}}\mathsf{G}') + \boldsymbol{u} + \Lambda_c$ and $\alpha\boldsymbol{y}$ is the shortest among all the shifted cosets, where $\alpha = \frac{P}{P+N}$ is some scaling factor (whose role will be explained later). P and N are the average power of the codeword and the noise per dimension, respectively. That is,

$$\hat{\boldsymbol{m}} = \arg\min_{\boldsymbol{m}} \ \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c, \alpha\boldsymbol{y}\right).$$

In fact, one can easily show that the unique shifted coset with the shortest distance is given by $Q_{\Lambda_f}(\alpha\boldsymbol{y} - \boldsymbol{u}) + \boldsymbol{u} + \Lambda_c$.

Analysis. For any given message vector $\boldsymbol{m}$, the average power constraint is satisfied if

- $\|\boldsymbol{x}\|^2 \leq nP$.

The decoding is successful if

- $\forall \boldsymbol{m}' \neq \boldsymbol{m} : \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathsf{G}') + \boldsymbol{u} + \Lambda_c, \alpha\boldsymbol{y}\right) > \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c, \alpha\boldsymbol{y}\right)$.

In Fig. 4.1, we provide a counter example in which the signal is decoded wrongly. The transmitted vector is $\boldsymbol{x}$, which is then "shifted" by the Gaussian noise $\boldsymbol{z}$ to $\boldsymbol{y}$. The received signal $\boldsymbol{y}$ is scaled by α to $\alpha\boldsymbol{y}$. The decoder will find the nearest coset to $\alpha\boldsymbol{y}$. In this example, the nearest coset to $\alpha\boldsymbol{y}$ is the coset containing $\hat{\boldsymbol{x}}$ (the star points) instead of the one containing $\boldsymbol{x}$ (the rectangle points). Hence, a decoding failure happens.

4.2 The Case of a Random Nested Lattice Code

We then proceed to the case of a random nested lattice code, which also allows us to apply probabilistic methods.

Random codebook generation. Randomly generate a matrix $\mathbf{G}_f \in \mathbb{Z}_{\mathsf{p}}^{k_f \times n}$ and a vector $\boldsymbol{U} \in \mathbb{Z}_{\mathsf{p}}^n$ where each entry of $\mathbf{G}_f$ and $\boldsymbol{U}$ is drawn independently and uniformly over $\mathbb{Z}_{\mathsf{p}}$. As before, let

$$\mathbf{G}_f = \begin{bmatrix} \mathbf{G}_c \\ \mathbf{G}' \end{bmatrix},$$

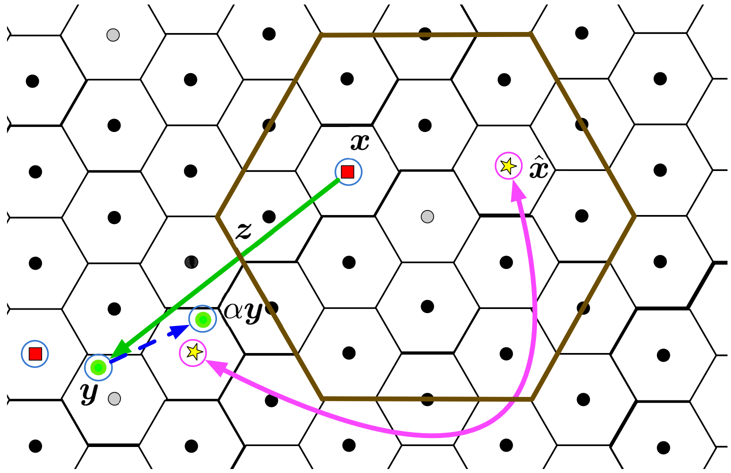

Figure 4.1: An example of decoding failure

and if $\mathbf{G}_f$ is full rank, so is $\mathbf{G}_c$. In this case, the codebook consists of $\mathsf{p}^{k_f-k_c}$ shifted cosets of the form

$$\left\{\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c : \boldsymbol{m} \in \mathbb{F}_{\mathsf{p}}^{k_f-k_c}\right\}.$$

If $\mathbf{G}_f$ is not full rank, we declare a codebook failure.

Encoding. The same as before.

Decoding. The same as before.

4.2.1 Analysis of the Codebook Failure.

Let $\mathcal{E}_1 = \{\mathbf{G}_f \text{ is not full rank}\}$. As before

$$\mathsf{P}(\mathcal{E}_1) \leq \frac{1}{\mathsf{p}-1}\frac{1}{\mathsf{p}^{n-k_f}}.$$

Hence, $\mathsf{P}(\mathcal{E}_1) \to 0$, as $\mathsf{p} \to \infty$ or $(n-k_f) \to \infty$.

4.2.2 Analysis of the Encoding Failure.

Recall that $\|\boldsymbol{X}\|^2 \leq nP$ if and only if $\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c \cap \mathcal{B}\left(\sqrt{nP}\right) \neq \emptyset$, where $\mathcal{B}\left(\sqrt{nP}\right)$ is the ball centred at the origin with radius $\sqrt{nP}$. Let

$$\mathcal{E}_2(\boldsymbol{m}) = \{\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c \cap \mathcal{B}\left(\sqrt{nP}\right) = \emptyset\}.$$

We will show that $\mathsf{P}(\mathcal{E}_2(\boldsymbol{m})) \to 0$ under certain condition.

Note that when $\mathcal{B}\left(\sqrt{nP}\right) \subset \mathcal{V}(\Lambda_{\mathsf{p}})$, where $\Lambda_{\mathsf{p}} = \ker \varphi = (\mathsf{p}\mathbb{Z})^n$ and $\mathcal{V}(\Lambda_{\mathsf{p}}) = [-\frac{\gamma\mathsf{p}}{2}, \frac{\gamma\mathsf{p}}{2}]^n$, $\mathcal{E}_2(\boldsymbol{m})$ is equivalent to

$$\sum_{\boldsymbol{l}\in\mathbb{Z}_{\mathsf{p}}^{k_c}} \mathbb{I}\left(\tilde{\varphi}(\boldsymbol{m}\mathsf{G}' + \boldsymbol{U} + \boldsymbol{l}\mathbf{G}_c) \in \mathcal{B}\left(\sqrt{nP}\right)\right) = 0,$$

because the set $\{\tilde{\varphi}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{U} + \boldsymbol{l}\mathbf{G}_c\right) : \boldsymbol{l} \in \mathbb{Z}_{\mathsf{p}}^{k_c}\}$ generates all the points of $\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c$ inside $\mathcal{V}(\Lambda_{\mathsf{p}})$.

Since $\tilde{\varphi}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{U} + \boldsymbol{l}\mathbf{G}_c\right)$ is uniformly distributed over the grid $\gamma\mathbb{Z}^n \cap \mathcal{V}(\Lambda_{\mathsf{p}})$ and there are exactly p^n different points inside $\mathcal{V}(\Lambda_{\mathsf{p}})$, we have

$$\mathsf{E}\left(\mathbb{I}\left(\tilde{\varphi}(\boldsymbol{m}\mathbf{G}' + \boldsymbol{U} + \boldsymbol{l}\mathbf{G}_c) \in \mathcal{B}\left(\sqrt{nP}\right)\right)\right) = \frac{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}{|\gamma\mathbb{Z}^n \cap \mathcal{V}(\Lambda_{\mathsf{p}})|} = \frac{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}{\mathsf{p}^n}$$

and

$$\mathrm{Var}\left(\mathbb{I}\left(\tilde{\varphi}(\boldsymbol{m}\mathbf{G}' + \boldsymbol{U} + \boldsymbol{l}\mathbf{G}_c) \in \mathcal{B}\left(\sqrt{nP}\right)\right)\right) = \frac{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}{\mathsf{p}^n}\left(1 - \frac{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}{\mathsf{p}^n}\right).$$

Similar to the case of nested linear codes, we have

$$\mathsf{P}(\mathcal{E}_2(\boldsymbol{m})) \leq \frac{\mathsf{p}^{n-k_c}}{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}. \tag{4.1}$$

4.2.3 Analysis of the Decoding Failure.

Recall that a successful decoding occurs upon receiving $\boldsymbol{Y}$ if

$$\forall \boldsymbol{m}' \neq \boldsymbol{m} : \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha\boldsymbol{Y}\right) > \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha\boldsymbol{Y}\right).$$

Let

$$\mathcal{E}_3(\boldsymbol{m}) = \{\exists \boldsymbol{m}' \neq \boldsymbol{m} : \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha\boldsymbol{Y}\right) \leq \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha\boldsymbol{Y}\right)\}.$$

Recall that $\boldsymbol{X} = \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) \mod \boldsymbol{\Lambda}_c$, and, in particular, $\boldsymbol{X} \in \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c$. Hence,

$$\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha\boldsymbol{Y}\right) \leq \|\boldsymbol{X} - \alpha\boldsymbol{Y}\| = \|(\alpha - 1)\boldsymbol{X} + \alpha\boldsymbol{Z}\|.$$

Let $\boldsymbol{W} = (\alpha - 1)\boldsymbol{X} + \alpha \boldsymbol{Z}$ be the "effective noise." By the Total Probability Theorem, we have

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_3(\boldsymbol{m})|\mathbf{G}_c = \mathsf{G}_c) \\
&\le \mathsf{P}(\boldsymbol{W} \notin \mathcal{B}\left(r_e\right)|\mathbf{G}_c = \mathsf{G}_c) + \mathsf{P}(\boldsymbol{W} \in \mathcal{B}\left(r_e\right)|\mathbf{G}_c = \mathsf{G}_c)\,\mathsf{P}(\mathcal{E}_3(\boldsymbol{m})|\boldsymbol{W} \in \mathcal{B}\left(r_e\right), \mathbf{G}_c = \mathsf{G}_c),
\end{aligned}$$

where $\mathcal{B}\left(r_e\right)$ is the "typical ball" for the effective noise $\boldsymbol{W}$ with radius r_e. It will be specified in Sec 4.2.3. It follows that

$$\mathsf{P}(\mathcal{E}_3(\boldsymbol{m})) \le \mathsf{P}(\boldsymbol{W} \notin \mathcal{B}\left(r_e\right)) + \sum_{\mathsf{G}_c} \mathsf{P}(\boldsymbol{W} \in \mathcal{B}\left(r_e\right), \mathbf{G}_c = \mathsf{G}_c)\,\mathsf{P}(\mathcal{E}_3(\boldsymbol{m})|\boldsymbol{W} \in \mathcal{B}\left(r_e\right), \mathbf{G}_c = \mathsf{G}_c).$$

Bounding $\mathsf{P}(\boldsymbol{W} \notin \mathcal{B}\left(r_e\right))$

Let ϵ be a small positive constant. We set $\alpha = \frac{P}{P+N}$ and set the radius

$$\begin{aligned}
r_e &= \sqrt{(1+\epsilon)n((\alpha-1)^2 P + \alpha^2 N)} \\
&= \sqrt{(1+\epsilon)\frac{nPN}{P+N}}.
\end{aligned}$$

Let

$$\begin{aligned}
\mathcal{E}_{\boldsymbol{X}} &= \{\|\boldsymbol{X}\| > \sqrt{nP}\}, \\
\mathcal{E}_{\boldsymbol{Z}} &= \{\|\boldsymbol{Z}\| > \sqrt{(1+\epsilon/2)nN}], \\
\mathcal{E}_{P} &= \{\|\boldsymbol{X}\boldsymbol{Z}^T\| > n^{\frac{1}{4}}\sqrt{nPN}\}.
\end{aligned}$$

It's clear that when n is large, $\mathcal{E}_{\boldsymbol{X}}^c \cap \mathcal{E}_{\boldsymbol{Z}}^c \cap \mathcal{E}_P^c$ implies $\|\boldsymbol{W}\| \le r_e$. Hence,

$$\mathsf{P}(\boldsymbol{W} \notin \mathcal{B}\left(r_e\right)) \le \mathsf{P}(\mathcal{E}_{\boldsymbol{X}}) + \mathsf{P}(\mathcal{E}_{\boldsymbol{Z}}) + \mathsf{P}(\mathcal{E}_P).$$

Note that $\mathcal{E}_{\boldsymbol{X}}$ is the same event as $\mathcal{E}_2$, which is bounded via (4.1). Since $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, NI_n)$, we obtain $\mathsf{P}(\mathcal{E}_{\boldsymbol{Z}}) \le 8\epsilon^2 n^{-1}$ by Chebyshev's inequality. The probability of $\mathcal{E}_P$ can be bouned as

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_P) &\le \mathsf{P}(\mathcal{E}_P \mid \|\boldsymbol{X}\| \le \sqrt{nP}) + \mathsf{P}(\|\boldsymbol{X}\| > \sqrt{nP}) \\
&= \mathsf{P}(\|\boldsymbol{X}\boldsymbol{Z}^T\|^2 > n^{\frac{3}{2}} PN \mid \|\boldsymbol{X}\| \le \sqrt{nP}) + \mathsf{P}(\mathcal{E}_2)
\end{aligned}$$

$$\leq \frac{\mathsf{E}(\|\boldsymbol{X}\boldsymbol{Z}^T\|^2 \mid \|\boldsymbol{X}\| \leq \sqrt{nP})}{n^{\frac{3}{2}}PN} + \mathsf{P}(\mathcal{E}_2)$$

where the last inequality follows from the Markov's inequality. Note that for any given $\boldsymbol{X} = \boldsymbol{x}$ with $\|\boldsymbol{x}\| \leq \sqrt{nP}$, $\boldsymbol{x}\boldsymbol{Z}^T \sim \mathcal{N}(0, \|\boldsymbol{x}\|^2 N)$, we then obtain $\mathsf{E}(\|\boldsymbol{X}\boldsymbol{Z}^T\|^2 \mid \|\boldsymbol{X}\| \leq \sqrt{nP}) \leq nPN$. Hence,

$$\mathsf{P}(\mathcal{E}_P) \leq n^{-\frac{1}{2}} + \mathsf{P}(\mathcal{E}_2).$$

Therefore,

$$\mathsf{P}(\boldsymbol{W} \notin \mathcal{B}(r_e)) \leq 8\epsilon^2 n^{-1} + n^{-\frac{1}{2}} + 2 \times \frac{\mathsf{p}^{n-k_c}}{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}.$$

Bounding $\mathsf{P}(\mathcal{E}_3(\boldsymbol{m}) \mid \boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c)$

Note that

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_3(\boldsymbol{m}) \mid \boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c) \\
&\leq \mathsf{P}\left(\exists \boldsymbol{m}' \neq \boldsymbol{m} : \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \Lambda_c, \alpha\boldsymbol{Y}\right) \leq \|\boldsymbol{W}\| \mid \boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c\right) \\
&\leq \sum_{\boldsymbol{m}' \neq \boldsymbol{m}} \mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \Lambda_c, \alpha\boldsymbol{Y}\right) \leq \|\boldsymbol{W}\| \mid \boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c\right).
\end{aligned}$$

Note also that

$$\begin{aligned}
&\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha\boldsymbol{Y}\right) \\
&= \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \boldsymbol{X} + (\alpha - 1)\boldsymbol{X} + \alpha\boldsymbol{Z}\right) \\
&= \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \boldsymbol{X} + \boldsymbol{W}\right) \\
&= \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \boldsymbol{\Lambda}_c, \boldsymbol{W}\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_3(\boldsymbol{m}) | \boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c) \\
&\leq \sum_{\boldsymbol{m}' \neq \boldsymbol{m}} \mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \Lambda_c, \boldsymbol{W}\right) \leq \|\boldsymbol{W}\| \mid \boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c\right) \\
&\leq \sum_{\boldsymbol{m}' \neq \boldsymbol{m}} \mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \Lambda_c, \boldsymbol{W}\right) \leq r_e \mid \boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c\right).
\end{aligned}$$

Next, we observe that $\mathbf{G}'$ and $\boldsymbol{W} = (\alpha - 1)\boldsymbol{X} + \alpha\boldsymbol{Z}$ are conditionally independent when given $\mathbf{G}_c = \mathsf{G}_c$. To see this, note that conditioned on $\mathbf{G}_c = \mathsf{G}_c$, $\boldsymbol{X}$ is uniformly distributed over $\gamma\mathbb{Z}^n \cap \mathcal{V}(\Lambda_c)$ and is independent of $\mathbf{G}'$ by Lemma 3. By the total probability theorem, we have

$$\begin{aligned} &\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \Lambda_c, \boldsymbol{W}\right) \leq r_e | \boldsymbol{W} \in \mathcal{B}\left(r_e\right), \mathbf{G}_c = \mathsf{G}_c\right) \\ = &\int_{\boldsymbol{w} \in \mathcal{B}(r_e)} \tilde{f}_{\boldsymbol{W}|\mathsf{G}_c}(\boldsymbol{w} \mid \mathsf{G}_c) \, \mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \Lambda_c, \boldsymbol{w}\right) \leq r_e | \mathbf{G}_c = \mathsf{G}_c\right) d\boldsymbol{w} \end{aligned}$$

where

$$\tilde{f}_{\boldsymbol{W}|\mathsf{G}_c}(\boldsymbol{w} \mid \mathsf{G}_c) = \frac{f_{\boldsymbol{W}|\mathsf{G}_c}(\boldsymbol{w} \mid \mathsf{G}_c)}{\mathsf{P}(\boldsymbol{W} \in \mathcal{B}\left(r_e\right) \mid \mathbf{G}_c = \mathsf{G}_c)}.$$

It turns out that the term $\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \Lambda_c, w\right) \leq r_e | \mathbf{G}_c = \mathsf{G}_c\right)$ can be bounded following Loeliger's approach [48].

Since $\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \Lambda_c, \boldsymbol{w}\right) \leq r_e$ implies

$$[\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}')] \bmod \Lambda_c \in [\boldsymbol{w} + \mathcal{B}\left(r_e\right)] \bmod \Lambda_c,$$

we have

$$\begin{aligned} &\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \Lambda_c, \boldsymbol{w}\right) \leq r_e | \mathbf{G}_c = \mathsf{G}_c\right) \\ \leq\ &\mathsf{P}\left([\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}')] \bmod \Lambda_c \in [\boldsymbol{w} + \mathcal{B}\left(r_e\right)] \bmod \Lambda_c \mid \mathbf{G}_c = \mathsf{G}_c\right). \end{aligned}$$

On the other hand, $([\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}')] \bmod \Lambda_c)$ is uniformly distributed over $\gamma\mathbb{Z}^n \cap \mathcal{V}(\Lambda_c)$, and so

$$\begin{aligned} &\mathsf{P}\left([\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}')] \bmod \Lambda_c \in ([\boldsymbol{w} + \mathcal{B}\left(r_e\right)] \bmod \Lambda_c) | \mathbf{G}_c = \mathsf{G}_c\right) \\ = &\frac{|\gamma\mathbb{Z}^n \cap \mathcal{V}(\Lambda_c) \cap (\boldsymbol{w} + \mathcal{B}\left(r_e\right))|}{\mathsf{p}^{n-k_c}} \\ \leq &\frac{|\gamma\mathbb{Z}^n \cap (\boldsymbol{w} + \mathcal{B}\left(r_e\right))|}{\mathsf{p}^{n-k_c}}. \end{aligned}$$

Therefore,

$$\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \Lambda_c, \alpha\boldsymbol{Y}\right) \leq \|\boldsymbol{W}\| \mid \boldsymbol{W} \in \mathcal{B}\left(r_e\right), \mathbf{G}_c = \mathsf{G}_c\right) \leq \max_{\boldsymbol{w} \in \mathcal{B}(r_e)} \frac{|\gamma\mathbb{Z}^n \cap (\boldsymbol{w} + \mathcal{B}\left(r_e\right))|}{\mathsf{p}^{n-k_c}}$$

and

$$\begin{aligned}\mathsf{P}(\mathcal{E}_3(\boldsymbol{m})|\boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c) &\leq \mathsf{p}^{k_f-k_c} \max_{\boldsymbol{w}\in\mathcal{B}(r_e)} \frac{|\gamma\mathbb{Z}^n \cap (\boldsymbol{w} + \mathcal{B}(r_e))|}{\mathsf{p}^{n-k_c}} \\ &\leq \max_{\boldsymbol{w}\in\mathcal{B}(r_e)} \frac{|\gamma\mathbb{Z}^n \cap (\boldsymbol{w} + \mathcal{B}(r_e))|}{\mathsf{p}^{n-k_f}}.\end{aligned}$$

4.3 Analysis of the Error Probability

By the union bound, the error probability P_e of the coding scheme is bounded by

$$\mathsf{P} \leq \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2) + \mathsf{P}_e(\mathcal{E}_3), \tag{4.2}$$

because the decoding is successful if $\mathbf{G}_c$ is full rank, $\|\boldsymbol{X}\|^2 \leq nP$, and the shifted coset containing $\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U})$ is the closest coset to $\alpha\boldsymbol{Y}$. In Chapter 4.3.2, we will show that, for any $\epsilon > 0$, we can select parameters $k_f, k_c, \mathsf{p}, \gamma$ as functions of n such that a rate of

$$R = \frac{1}{2}\log_2\left(\frac{1 + P/N}{1+\epsilon}\right)$$

is achievable with error probability $\mathsf{P}_e \to 0$ as $n \to \infty$.

However, the above result *doesn't* imply our random ensemble achieves the AWGN capacity, because the power constraint is not always satisfied. In fact, the power constraint is violated with probability $\mathsf{P}(\mathcal{E}_2)$. To address this issue, we introduce a spherical shaping strategy, which is in parallel with the minor change introduced in [59, p.47] for proving channel coding theorem with input cost constraint.

4.3.1 Spherical Shaping

We apply a "truncated" spherical shaping to $\boldsymbol{X}$ as follows

$$\boldsymbol{X}_S = \begin{cases} \boldsymbol{X}, & \text{if } \|\boldsymbol{X}\| \leq nP, \\ \mathbf{0}, & \text{otherwise.} \end{cases}$$

Clearly, the power constraint is always satisfied for the new coding scheme. Note that the error probability for the new coding scheme is still bounded by $\mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2) + \mathsf{P}(\mathcal{E}_3)$, because the spherical shaping converts an encoding failure to a decoding failure.

4.3.2 The Selection of Parameters.

To complete the proof that our random ensemble achieves the AWGN capacity with lattice encoding and decoding, we carefully select the values of $k_f, k_c, \mathsf{p}, \gamma$ so that P_e goes to zero and the rate of our coding scheme goes to the AWGN capacity as n goes to infinity.

We have already bounded the error probability as

$$\begin{aligned}\mathsf{P}_e &\leq \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2) + \mathsf{P}(\mathcal{E}_3)\\ &\leq \frac{1}{\mathsf{p}-1}\frac{1}{\mathsf{p}^{n-k_f}} + 8\epsilon^2 n^{-1} + n^{-\frac{1}{2}} + 3\times\frac{\mathsf{p}^{n-k_c}}{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|} + \max_{\boldsymbol{w}\in\mathcal{B}(r_e)}\frac{|\gamma\mathbb{Z}^n\cap(\boldsymbol{w}+\mathcal{B}(r_e))|}{\mathsf{p}^{n-k_f}}.\end{aligned}$$

Using Lemma 5, we obtain

$$\mathsf{P}_e \leq \frac{1}{\mathsf{p}-1}\frac{1}{\mathsf{p}^{n-k_f}} + 8\epsilon^2 n^{-1} + n^{-\frac{1}{2}} + 3\times\frac{\mathsf{p}^{n-k_c}}{\left(\max\left\{\frac{\sqrt{nP}}{\gamma}-\frac{\sqrt{n}}{2},0\right\}\right)^n V_n} + \frac{\left(\frac{r_e}{\gamma}+\frac{\sqrt{n}}{2}\right)^n V_n}{\mathsf{p}^{n-k_f}}.$$

Now our goal is to select p, γ, k_c and k_f (as functions of n) such that

$$\frac{1}{\mathsf{p}-1}\frac{1}{\mathsf{p}^{n-k_f}} \to 0, \tag{4.3}$$

$$\frac{\mathsf{p}^{n-k_c}}{\left(\max\left\{\frac{\sqrt{nP}}{\gamma}-\frac{\sqrt{n}}{2},0\right\}\right)^n V_n} \to 0, \tag{4.4}$$

$$\frac{\left(\frac{r_e}{\gamma}+\frac{\sqrt{n}}{2}\right)^n V_n}{\mathsf{p}^{n-k_f}} \to 0, \tag{4.5}$$

under the constraint $\mathcal{B}\left(\sqrt{nP}\right) \subset \mathcal{V}(\Lambda_{\mathsf{p}})$. Recall that $\mathcal{V}(\Lambda_{\mathsf{p}}) = [-\frac{\gamma\mathsf{p}}{2}, \frac{\gamma\mathsf{p}}{2}]^n$ which is equivalent to

$$\gamma\mathsf{p} \geq 2\sqrt{nP}. \tag{4.6}$$

Let $\eta > 0$ and $\delta \in (0,1)$ be two constants. Then let $\gamma = n^{-\frac{1}{2}\eta}$. Let p be the smallest prime larger than $n^{1+\eta}$ which statisfies $\mathsf{p} \bmod 4 = 1$. By Lemma 6, we can write $\mathsf{p} = \mu n^{1+\eta}$ where μ is a bounded constant. We then assign

$$k_c = \left\lceil n\left(1 - \frac{\log_2(\sqrt{P}n^{\frac{1}{2}\eta} - \frac{1}{2}) + \frac{1}{2}\log_2((1-\delta)nV_n^{\frac{2}{n}})}{\log_2 \mathsf{p}}\right)\right\rceil,$$

and

$$k_f = \left\lfloor n \left(1 - \frac{\log_2(\sqrt{\frac{1}{n} r_e^2} n^{\frac{1}{2}\eta} + \frac{1}{2}) + \frac{1}{2}\log_2(\frac{1}{1-\delta} n V_n^{\frac{2}{n}})}{\log_2 \mathsf{p}} \right) \right\rfloor .$$

Since $\gamma p \geq n^{\frac{1}{2}+\frac{1}{2}\eta}$, it grows faster than $n^{\frac{1}{2}}$ and then the constraint (4.6) is met when n is large. By the facts that $\lim_{n\to\infty} n V_n^{\frac{2}{n}} = 2\pi e$ from [53, (2)] and that $\frac{1}{n} r_e^2 < P$ for small ϵ, one can verify that $1 \leq k_c < k_f < n$ when n is large. We now substitute p, k_1 and k_2 into (4.3),(4.4) and (4.5). It is clear (4.3),(4.4) and (4.5) vanish as $n \to \infty$. In other words, in our random ensemble, there exist a non-zero portion of pre-determined codebooks whose error probabilities go to zero.

Finally, we calculate the achievable rate

$$\lim_{n\to\infty} \frac{1}{n} \log_2 p^{k_f - k_c} = \lim_{n\to\infty} \frac{1}{2} \log_2 \left(\frac{nP}{r_e^2} \right) = \frac{1}{2} \log_2 \left(\frac{1 + P/N}{1+\epsilon} \right),$$

where ϵ can be arbitrarily small. Hence, we claim there exist pre-determined codebooks in our ensemble that achieve the channel capacity.

Chapter 5

Achievable Rates of Nested Algebraic Lattice Codes

In Chapter 2.3, we constructed a pair of nested lattice codes by using the map $\varphi : \gamma\mathbb{Z}^n \to \mathbb{F}_{\mathsf{p}}^n$ and its associated map $\tilde{\varphi}$. In this chapter, we will first consider constructing lattice codes from a more general map ϕ_{p} which maps a lattice point in $\Lambda \subset \mathbb{R}^m$ to a point in $\mathbb{F}_{\mathsf{p}}^n$, where $m = tn$ and t is a constant integer. This construction is proposed in [58] and we briefly recapture it in Chapter 5.1. We then analyze the error probability of the codes we just constructed using almost the same methods in Chapter 4.2. This analysis only relies on the abstract properties of ϕ_{p} as we will show in Chapter 5.3. To build concrete examples of such ϕ_{p}, we need to make use of the algebraic number theory and we will briefly introduce it in Chapter 5.4 and Chapter 5.5. For convenience, we also call the lattice codes constructed by such ϕ_{p} *algebraic lattice codes*. At the end of this chapter, we will show that some algebraic lattice codes could achieve the AWGN channel capacity.

5.1 A Generalized Reduction

Let Λ be a lattice in $\mathbb{R}^m$. Let $\phi_{\mathsf{p}} : \Lambda \to \mathbb{F}_{\mathsf{p}}^n$ be a surjective homomorphism. Given a linear code C in $\mathbb{F}_{\mathsf{p}}^n$, its associated lattice via ϕ_{p} is defined as $\Lambda_{\mathsf{p}}(\mathsf{C}) \triangleq \phi_{\mathsf{p}}^{-1}(\mathsf{C})$. The kernel of ϕ is denoted as $\ker(\phi_{\mathsf{p}}) = \Lambda_{\mathsf{p}}(\{\mathbf{0}\}) \triangleq \Lambda_{\mathsf{p}}$. It's clear that $\Lambda_{\mathsf{p}} \subset \Lambda_{\mathsf{p}}(\mathsf{C}) \subset \Lambda$ by noting that $\Lambda = \phi_{\mathsf{p}}^{-1}(\mathbb{F}_{\mathsf{p}}^n)$ and that $\{\mathbf{0}\} \subset \mathsf{C} \subset \mathbb{F}_{\mathsf{p}}^n$. Moreover, the quotient $\Lambda_{\mathsf{p}}(\mathsf{C})/\Lambda_{\mathsf{p}} \simeq \mathsf{C}$ and therefore $V(\Lambda_{\mathsf{p}}(\mathsf{C})) = |\mathsf{C}|^{-1}\mathsf{p}^n V(\Lambda)$.

Similar to $\tilde{\varphi}$, we can also define $\widetilde{\phi}_{\mathsf{p}}$, as the associated map of ϕ_{p}, which embeds $\mathbb{F}_{\mathsf{p}}^n$ into $\mathbb{R}^m$. For a point $\boldsymbol{c}$ in $\mathbb{F}_{\mathsf{p}}^n$, we define $\widetilde{\phi}_{\mathsf{p}}(\boldsymbol{c})$ as the point of the shortest Euclidean norm in $\phi_{\mathsf{p}}^{-1}(\boldsymbol{c})$. Similar to $\tilde{\varphi}$, $\widetilde{\phi}_{\mathsf{p}}(\boldsymbol{c})$ must lie in $\mathcal{V}(\Lambda_{\mathsf{p}})$. Unlike $\tilde{\varphi}$ which embeds $\mathbb{F}_{\mathsf{p}}^n$ into $\mathbb{R}^n$, the generalized map $\widetilde{\phi}_{\mathsf{p}}$ embeds $\mathbb{F}_{\mathsf{p}}^n$ into $\mathbb{R}^m$ where n and m do not need to be equal.

Equipped with $\widetilde{\phi}_{\mathsf{p}}$, we can naturally construct a pair of nested lattice codes (Λ_f, Λ_c) in $\mathbb{R}^m$ from a given pair of nested linear codes $(\mathsf{G}_f, \mathsf{G}_c)$ in $\mathbb{F}_{\mathsf{p}}$ as we did in Chapter 4.1.

5.2 Generalized Codebook Generalization

We first consider the case of a pre-determined nested lattice code. Given a pair of lattice codes (Λ_f, Λ_c) and a dither vector $\boldsymbol{u} \in \mathbb{R}^m$, we construct a codebook whose codewords are shifted cosets of the form $\{\boldsymbol{\lambda}_f + \boldsymbol{u} + \Lambda_c : \boldsymbol{\lambda}_f \in \Lambda_c\}$ using the same procedure introduced in Chapter 4.1. That is, all the codewords (i.e., the shifted cosets) can be expressed as

$$\left\{ \widetilde{\phi}_{\mathsf{p}}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c : \boldsymbol{m} \in \mathbb{F}_{\mathsf{p}}^{k_f - k_c} \right\}.$$

Note that $\widetilde{\phi}_{\mathsf{p}}$ is a one-to-one map, so that here is still a one-to-one correspondence between the vectors in $\mathbb{F}_{\mathsf{p}}^{k_f-k_c}$ and the shifted cosets of Λ_c. Hence, $\boldsymbol{m}$ can still be viewed as the "index" of the shifted coset $\widetilde{\phi}_{\mathsf{p}}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c$, and the codebook contains $\mathsf{p}^{k_f-k_c}$ (distinct) codewords.

Encoding. To send a message vector $\boldsymbol{m} \in \mathbb{F}_{\mathsf{p}}^{k_f-k_c}$, the encoder transmits

$$\boldsymbol{x} = \widetilde{\phi}_{\mathsf{p}}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} \mod \Lambda_c.$$

Decoding. Upon receiving $\boldsymbol{y} \in \mathbb{R}^m$, we estimate $\boldsymbol{m}$ as

$$\hat{\boldsymbol{m}} = \arg\min_{\boldsymbol{m}} \ \mathrm{d}\left(\widetilde{\phi}_{\mathsf{p}}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c, \alpha \boldsymbol{y} \right).$$

In fact, this is almost the same as the decoding procedure in Chapter 4.1. One can easily show that the unique shifted coset with the shortest distance is given by $Q_{\Lambda_f}(\alpha\boldsymbol{y} - \boldsymbol{u}) + \boldsymbol{u} + \Lambda_c$.

As for the random case, we first randomly generate a matrix $\mathbf{G}_f \in \mathbb{Z}_{\mathsf{p}}^{k_f \times n}$ and a vector $\boldsymbol{U} \in \mathbb{Z}_{\mathsf{p}}^n$ where each entry of $\mathbf{G}_f$ and $\boldsymbol{U}$ is drawn independently and uniformly over $\mathbb{Z}_{\mathsf{p}}$ as we did in Chapter 4.1. As before, let

$$\mathbf{G}_f = \begin{bmatrix} \mathbf{G}_c \\ \mathbf{G}' \end{bmatrix},$$

and if $\mathbf{G}_f$ is full rank, so is $\mathbf{G}_c$. We then generate all the codewords in the random nested lattice codes as

$$\left\{ \widetilde{\phi}_{\mathsf{p}}(\boldsymbol{m}\mathbf{G}') + \widetilde{\phi}_{\mathsf{p}}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c : \boldsymbol{m} \in \mathbb{F}_{\mathsf{p}}^{k_f-k_c} \right\}.$$

5.3 Analysis of the Error Probability

The error probability of the generalized scheme can be analyzed using the same methods in Chapter 4.2. The sperical shaping in Chapter 4.3.1 is still needed. There are two differences between the current analysis and the one in Chapter 4.2. The first is that the energy constraint becomes $\boldsymbol{X} \in \mathcal{B}\left(\sqrt{mP}\right)$ instead of $\boldsymbol{X} \in \mathcal{B}\left(\sqrt{nP}\right)$ since the lattices in this chapter lie in $\mathbb{R}^m$ instead of $\mathbb{R}^n$. The second is that the domain of $\widetilde{\phi}_{\mathsf{p}}$ is a more general lattice Λ instead of the lattice $\gamma\mathbb{Z}^n$ used by $\tilde{\varphi}$.

To make the error probability of the generalized scheme goes to zero, we need three conditions that are similar to the ones (4.3), (4.4), and (4.5) in Chapter 4.3,

$$\frac{\mathsf{p}^{n-k_c}}{|\Lambda \cap \mathcal{B}\left(\sqrt{mP}\right)|} \to 0, \tag{5.1}$$

$$\max_{\boldsymbol{w} \in \mathcal{B}(r_e)} \frac{|\Lambda \cap (\boldsymbol{w} + \mathcal{B}(r_e))|}{\mathsf{p}^{n-k_f}} \to 0, \tag{5.2}$$

$$\mathcal{B}\left(\sqrt{mP}\right) \subset \mathcal{V}(\Lambda_{\mathsf{p}}), \tag{5.3}$$

where $r_e = \sqrt{(1+\epsilon)\frac{mPN}{P+N}}$. By Lemma 4, the above becomes

$$\frac{V(\Lambda)\mathsf{p}^{n-k_c}}{(\sqrt{mP}-l)^m V_m} \to 0, \tag{5.4}$$

$$\frac{(r_e+l)^m V_m}{\mathsf{p}^{n-k_f} V(\Lambda)} \to 0, \tag{5.5}$$

$$\mathcal{B}\left(\sqrt{mP}\right) \subset \mathcal{V}(\Lambda_{\mathsf{p}}), \tag{5.6}$$

where $l = \sup_{\boldsymbol{x} \in \mathcal{V}(\Lambda)} \|\boldsymbol{x}\|$.

Clearly, the above requirements rely on geometric properties of the base lattice Λ and the kernel lattice Λ_{p}. However, till now, we only rely on the abstract properties of ϕ_{p}, so that we lack detailed geometric measures. We thus will offer some concrete examples on ϕ_{p} to demonstrate those measures. The naive example is to choose ϕ_{p} as φ. By doing so, we get the same result as the one in Chapter 4. In the rest of this chapter, we introduce how to construct ϕ_{p} and nested lattice codes using knowledge of algebraic number field.

5.4 Algebraic Number Field

In this chapter, we will introduce basics of algebraic number theory briefly. The readers need some common knowledge of abstract algebra, including the definitions of ring, field, and group.

To find the roots of a polynomial $f(X)$ over a field K, it's often necessary to pass to a larger field L containing K. In these cases, the field L is usually called a *field extension* of K. For example $f(X) = X^2 - 2$ has no roots in $\mathbb{Q}$. However when considering $f(X)$ as a polynomial in a field that contains $\sqrt{2}$, we can naturally find roots $\pm\sqrt{2}$. We also denote the field extension relationship between L and K as L/K. As a field extension over K, the field L naturally owns a structure as a vector space over K. The dimension of this vector space is called as the *degree* of L over K and is denoted as $[L : K]$. If $[L : K]$ is finite, we call L a *finite extension* of K.

A field K is called a *number field* if it is a finite extension of $\mathbb{Q}$. Assume the degree of this extension is n. We know that for any $\alpha \in K$, there must exist a $\mathbb{Q}$-linear dependency $\{1, \alpha, \alpha^2, \cdots, \alpha^n\}$. In other words, there exists a polynomial f whose coefficients lie in $\mathbb{Q}$ such that $f(\alpha) = 0$. We call α an *algebraic number*. Among all such polynomials which have a root α, we can find a polynomial with the smallest degree and call it the *minimal polynomial* of α.

For example, we can build a number field by "adding" $\sqrt{2}$ to $\mathbb{Q}$. To make this new set a field, we need to add all multiples and all powers of $\sqrt{2}$ to $\mathbb{Q}$. It turns out the new set is $\{a + b\sqrt{2} | a, b \in \mathbb{Q}\}$. Readers can check this new set is actually a field. Since $\sqrt{2}$ is the root of $X^2 - 2 = 0$, $\sqrt{2}$ is an algebraic number. Also, $X^2 - 2 = 0$ is the minimal polynomial of $\sqrt{2}$. Similarly, $X - 1 = 0$ is the minimal polynomial of 1. By adding more numbers, we can get larger number field. For instance, by adding $\sqrt[3]{5}$ to $\mathbb{Q}(\sqrt{2})$, we get $\mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$, which is the smallest field extension over $\mathbb{Q}$ that contains $\sqrt{2}$ and $\sqrt[3]{5}$.

Since a number field K is a finite extension of $\mathbb{Q}$, we can write K as $\mathbb{Q}(\alpha_1, \alpha_2, \ldots, \alpha_s)$ for finite many algebraic numbers $(\alpha_1, \alpha_2, \ldots, \alpha_s)$. We have a stronger result.

Lemma 7([63, Theorem 2.2]): If K is a number field then $K = \mathbb{Q}(\theta)$ for some algebraic number θ, which is also called the *primitive element*.

The key observation of this lemma is that for a number field $\mathbb{Q}(\alpha, \beta)$, we can always find a suitable $c \in \mathbb{Q}$ such that $\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\alpha + c\beta)$. For example, $\mathbb{Q}(\sqrt{2}, \sqrt[3]{5}) = \mathbb{Q}(\sqrt{2} + \sqrt[3]{5})$. Of course, the representation of K as $\mathbb{Q}(\theta)$ is not unique since $\mathbb{Q}(\theta) = \mathbb{Q}(-\theta) = \mathbb{Q}(\theta + 1) = \ldots$ etc.

Also, as a consequence of this lemma, we find a $\mathbb{Q}$-basis for K as

$$\{1, \theta, \theta^2, \ldots, \theta^{n-1}\}.$$

A number field $K = \mathbb{Q}(\theta)$ can be embedded into the complex field $\mathbb{C}$ by several distinct homomorphism $\sigma_i : K \to \mathbb{C}$. For example, if $K = \mathbb{Q}(i)$ where $i = \sqrt{-1}$, we have two possibilities

$$\sigma_1(x + yi) = x + yi,$$
$$\sigma_2(x + yi) = x - yi.$$

This observation can be described by the following lemma.

Lemma 8([63, Theorem 2.4]): Let $K = \mathbb{Q}(\theta)$ be a number field of degree n over $\mathbb{Q}$. Then there are exactly n distinct homomorphisms $\sigma_i : K \to \mathbb{C}, i = 1, 2, \ldots, n$. The element $\sigma_i(\theta) = \theta_i$ is the i-th root in $\mathbb{C}$ of the minimal polynomial of θ over $\mathbb{Q}$.

If $\sigma_i(K) \in \mathbb{R}$, which happens if and only if $\sigma_i(\theta) \in \mathbb{R}$, we say that σ_i is *real*; otherwise, σ_i is said *complex*. As usual, denote the complex conjugate by bars and define

$$\bar{\sigma_i}(\alpha) = \overline{\sigma_i(\alpha)}.$$

Suppose there are r_1 real homomorphisms and $2r_2$ complex homomorphisms, then the degree of the field extension n is equal to $r_1 + 2r_2$. The couple (r_1, r_2) is known as the *signature* of K. For example, the signature of $\mathbb{Q}(i)$ is $(0, 1)$ since there are 2 complex homomorphisms. The signature of $\mathbb{Q}(\sqrt[3]{(2)})$ is $(1, 1)$. It's because there are exactly 3 σ_i's, $\sigma_1(\sqrt[3]{2}) = \sqrt[3]{2}, \sigma_2(\sqrt[3]{2}) = \omega\sqrt[3]{2}, \sigma_3(\sqrt[3]{2}) = \omega^2\sqrt[3]{2}$, where ω is the cubic root of unity in $\mathbb{C}$.

Equipped with σ_i's, we can build the *canonical embedding* σ which sends a point in K to a point $\mathbb{R}^{r_1+2r_2}$ as

$$\sigma : K \mapsto \mathbb{R}^n$$
$$\sigma(x) = (\sigma_1(x), \ldots, \sigma_{r_1}(x), \mathrm{Re}(\sigma_{r_1+1}(x)), \mathrm{Im}(\sigma_{r_1+1}(x)), \ldots, \mathrm{Re}(\sigma_{r_1+r_2}(x)), \mathrm{Im}(\sigma_{r_1+r_2}(x))).$$

Readers can check that a $\mathbb{Q}$-basis $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ of K can generate vectors $\{\sigma(\alpha_1), \sigma(\alpha_2), \ldots, \sigma(\alpha_n)\}$ which are linearly independent over $\mathbb{Q}$. However, we can obtain more as stated in the following lemma.

Lemma 9([63, Theorem 8.1]): If $\alpha_1, \alpha_2, \ldots, \alpha_n$ is a basis for K over $\mathbb{Q}$, then $\sigma(\alpha_1), \sigma(\alpha_2), \ldots, \sigma(\alpha_n)$ are linearly independent over $\mathbb{R}$.

The following corollary clarifies a way on using the number field K to build lattices.

Corollary 1: If G is a finitely generated subgroup of $(K, +)$ with $\mathbb{Z}$-basis $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ then the image of G is a lattice in $\mathbb{R}^n$ with generators $\{\sigma(\alpha_1), \sigma(\alpha_2), \ldots, \sigma(\alpha_m)\}$.

In the following chapter, we will introduce the ring of integers over K, which is a finitely generated subgroup of K, as well as some useful properties of it.

5.5 Algebraic Integers

A number θ is an *algebraic integer* if there is a *monic* polynomial $p(X)$ with integer coefficients such that $p(\theta) = 0$. In other words,

$$\theta^n + a_{n-1}\theta^{n-1} + \cdots + a_0 = 0,$$

where $a_i \in \mathbb{Z}$ for all i. For example, $\frac{1+\sqrt{5}}{2}$ is an algebraic integer since its the root of $X^2 - X - 1 = 0$, but $\frac{1}{3}$ is not an algebraic integer.

For convenience, we denote the set of all algebraic integers $\mathcal{B}$. An insightful observation is given by the following lemma.

Lemma 10([63, Theorem 2.9]): The algebraic integers form a subring of the field of algebraic numbers.

For two algebraic integers α and β, it's not easy to show that $\alpha\beta$ and $\alpha + \beta$ lie in $\mathcal{B}$. We need the following lemma.

Lemma 11([63, Lemma 2.8]): A complex number θ is an algebraic integer if and only if the additive group generated by all powers $1, \theta, \theta^2, \cdots$ is finitely generated.

Since all powers of α and β are finitely generated, we know that powers of $\alpha\beta$ and $\alpha + \beta$ are also finitely generated. Hence, $\alpha\beta$ and $\alpha + \beta$ lie in $\mathcal{B}$.

For any number field K, we denote

$$\mathcal{O}_K = K \cap \mathcal{B},$$

and call $\mathcal{O}_K$ the *ring of integers* of K. Obviously, $\mathcal{O}_K$ is a subring of K and $\mathbb{Z} \subset \mathcal{O}_K$. One of the reasons that we are interested in $\mathcal{O}_K$ is stated as the following.

Lemma 12([63, Theorem 2.16]): Let K be a number field and $\mathcal{O}_K$ be the ring of integers of K. The additive group of $\mathcal{O}_K$ is a free abelian group of rank n equal to the degree of K.

In other words, there exist a $\mathbb{Z}$-basis $\alpha_1, \cdots, \alpha_n$ for $\mathcal{O}_K$ where $\alpha_i \in \mathcal{O}_K$ for all i. As a natural result, $\mathcal{O}_K$ is a finitely generated subgroup of K and by Corollary 1, we obtain that the image $\sigma(\mathcal{O}_K)$ generated by the cannoical embedding is a lattice in $\mathbb{R}^n$.

We already bridged the lattice in real field $\mathbb{R}^n$ and $\mathcal{O}_K$. We then introduce the connection between $\mathcal{O}_K$ and a certain finite field so that we can build lattices from linear codes in that finite field. It turns out the key ingredient of this connection is the unique factorization of ideals. Similar to the factorization of rational integers, we might factorize algebraic integers into product of irreducibles. However, we cannot always factorize an algebraic integer uniquely. For example, if we work in $\mathbb{Z}(\sqrt{-6})$, there are two factorizations, $6 = 2 \cdot 3$ and $6 = \sqrt{-6} \cdot \sqrt{-6}$. Though the numbers $2, 3$ and $\sqrt{6}$ are already irreducible, $2, 3$ and $\sqrt{6}$ are not prime since $2 \nmid \sqrt{-6}$, $3 \nmid \sqrt{-6}$, $\sqrt{-6} \nmid \sqrt{2}$ and $\sqrt{-6} \nmid \sqrt{3}$. Nevertheless, as we already stated, the facotrization into ideals can be unique.

We need to introduce two concepts first. Given two ideals $\mathfrak{a}, \mathfrak{b}$, the *product of ideals* $\mathfrak{ab}$ is the set of finite sums $\sum a_i b_i$ where $a_i \in \mathfrak{a}, b_i \in \mathfrak{b}$. An ideal $\mathfrak{p}$ is a *prime ideal* if given $\mathfrak{ab} \subset \mathfrak{p}$, then either $\mathfrak{a} \subset \mathfrak{p}$ or $\mathfrak{a} \subset \mathfrak{p}$. Now, we are prepared to introduce the following lemma.

Lemma 13([63, Theorem 5.6]): Every non-zero ideal of $\mathcal{O}_K$ can be written as a product of prime ideals, uniquely up to the order of the factors.

For example, in $\mathbb{Z}\sqrt{-17}$, we have the unique factorization of 3 as $\langle 3 \rangle = \langle 3, 1 + \sqrt{-17} \rangle \langle 3, 1 - \sqrt{-17} \rangle$, where both $\langle 3, 1 + \sqrt{-17} \rangle$ and $\langle 3, 1 - \sqrt{-17} \rangle$ are prime.

We provide two useful lemmas about prime ideals. Similar to the fact that the prime ideal of $\mathbb{Z}$ is a maximal ideal, we have

Lemma 14: Every non-zero prime ideal $\mathfrak{p}$ of $\mathcal{O}_K$ is a maximal ideal of $\mathcal{O}_K$. Moreover, the residue field ${}^{\mathcal{O}_K}\!/_{\mathfrak{p}}$ is a finite field.

By analogy with factorization of rational integers, for ideals $\mathfrak{a}, \mathfrak{b}$, we shall say that $\mathfrak{a}$ divides $\mathfrak{b}$ (written $\mathfrak{a} \mid \mathfrak{b}$) if there is an ideal $\mathfrak{c}$ such that $\mathfrak{b} = \mathfrak{ac}$. We have the following lemma.

Lemma 15([63, Propostion 5.7]): For ideals $\mathfrak{a}, \mathfrak{b}$ of $\mathcal{O}_K$,

$$\mathfrak{a} \mid \mathfrak{b} \text{ if and only if } \mathfrak{b} \subset \mathfrak{a}.$$

By Lemma 14, the ring of integers $\mathcal{O}_K$ is connected to a finite field by the fact that the residue field ${\mathcal{O}_K}/{\mathfrak{p}}$ is a finite field if $\mathfrak{p}$ is prime. Here we start to dig out more concrete descriptions about this residue field. When $\mathfrak{p}$ is a prime ideal of $\mathcal{O}_K$, it's easy to verify $\mathfrak{p} \cap \mathbb{Z}$ is a prime ideal of $\mathbb{Z}$. Therefore, there must exist a rational prime p such that $\mathfrak{p} \cap \mathbb{Z} = \mathsf{p}\mathbb{Z}$. In this case, we say that $\mathfrak{p}$ is above p. We claim that ${\mathcal{O}_K}/{\mathfrak{p}}$ is a finite extension of $\mathbb{F}_{\mathsf{p}}$. To see this, first construct a *projection* map π as

$$\pi : \quad \mathcal{O}_K \to {\mathcal{O}_K}/{\mathfrak{p}}$$
$$\pi(a) = a + \mathfrak{p}.$$

The kernel of π is $\mathfrak{p}$. Then we construct a map $\tau : \mathbb{Z} \to \mathcal{O}_K \to {\mathcal{O}_K}/{\mathfrak{p}}$. The first arrow is the canonical embedding of $\mathbb{Z}$ into $\mathcal{O}_K$ and the second arrow is the projection map π. The kernel of τ is exactly $\mathfrak{p} \cap \mathbb{Z} = \mathsf{p}\mathbb{Z}$. Therefore, there is an injection from $\mathbb{F}_{\mathsf{p}} = {\mathbb{Z}}/{\mathsf{p}\mathbb{Z}}$ to ${\mathcal{O}_K}/{\mathfrak{p}}$. Hence, ${\mathcal{O}_K}/{\mathfrak{p}}$ is a finite extension of $\mathbb{F}_{\mathsf{p}}$. The degree of this extension is called the *inertial degree* and is denoted as $f_{\mathfrak{p}} = [{\mathcal{O}_K}/{\mathfrak{p}} : \mathbb{F}_{\mathsf{p}}]$.

On the other hand, since $\mathsf{p} \in \mathfrak{p}$, we know that $\mathsf{p}\mathcal{O}_K \subset \mathfrak{p}$. Hence, $\mathfrak{p}$ must be a prime factor of $\mathsf{p}\mathcal{O}_K$. We write the factorization of $\mathsf{p}\mathcal{O}_K$ in $\mathcal{O}_K$ as

$$\mathsf{p}\mathcal{O}_K = \prod_{i=1}^{g} \mathfrak{p}_i^{e_{\mathfrak{p}_i}}, \tag{5.7}$$

where $e_{\mathfrak{p}_i} \in \mathbb{Z}$ for all $\mathfrak{p}_i$ and is called the *ramification index* of $\mathfrak{p}_i$. The inertial degree and ramification index are connected via the following lemma.

Lemma 16: Given the factorization $\mathsf{p}\mathcal{O}_K = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i}$ for a prime $\mathsf{p} \in \mathbb{Z}$ and the ring of integers over the number field K, we have

$$[K : \mathbb{Q}] = \sum_{i=1}^{g} f_{\mathfrak{p}_i} e_{\mathfrak{p}_i}.$$

When $g = [K : \mathbb{Q}], f_{\mathfrak{p}_i} = 1$ for all i, we say the prime p *splits*. We then have the following corollary.

Corollary 2: Given the factorization $\mathsf{p}\mathcal{O}_K = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i}$, when the prime p splits, we have

$${\mathcal{O}_K}/{\mathfrak{p}_i} \simeq \mathbb{F}_{\mathsf{p}}. \tag{5.8}$$

This result is used to construct the map ϕ_{p} in Chapter 5.6.1.

5.6 Construction of Nested Algebraic Lattice Codes

5.6.1 The Construction of ϕ_{p}

Let K be a number field whose signature is (r_1, r_2), i.e., it has r_1 real embeddings and $2r_2$ complex embeddings. Let $\sigma_1, \sigma_2, \ldots, \sigma_{r_1}$ be the real embeddings, and $\sigma_{r_1+1}, \sigma_{r_2+1}, \ldots, \sigma_{r_n}$ be the complex embeddings where $\sigma_{r_1+r_2+i} = \overline{\sigma_{r_1+i}}$.

Let $\mathcal{O}_K$ be the ring of integers of K. The cannonical emebedding σ from $\mathcal{O}_K$ to the real vector space of dimension $r_1 + 2r_2$ is denoted as

$$\sigma : \mathcal{O}_K \mapsto \mathbb{R}^{r_1+2r_2}$$
$$\sigma(x) = (\sigma_1(x), \ldots, \sigma_{r_1}(x), \mathrm{Re}(\sigma_{r_1+1}), \mathrm{Im}(\sigma_{r_1+1}), \ldots, \mathrm{Re}(\sigma_{r_1+r_2}), \mathrm{Im}(\sigma_{r_1+r_2})).$$

For example, the canonical embedding from $\mathbb{Z}[i]$ to $\mathbb{R}^2$ is $\sigma(a+bi) = (a, b)$.

Equipped with this canonical embedding, we can build the map ϕ_{p} from $\Lambda \subset \mathbb{R}^m$ to $\mathbb{F}_{\mathsf{p}}^n$ as follows. Let p be a prime that splits and $\mathfrak{p}$ be a prime ideal above p. By Corollary 2, we can find a projection map $\pi : \mathcal{O}_K \to \mathcal{O}_K/\mathfrak{p} \simeq \mathbb{F}_{\mathsf{p}}$. Let $\sigma : \mathcal{O}_K \to \mathbb{R}^t$ be the canonical embedding, where $t = r_1 + 2r_2$ and $m = tn$. Let Λ be the lattice $\sigma(\mathcal{O}_K)$ which is in $\mathbb{R}^{nt}$. By applying the projection map π element-wisely, we obtain a concrete map ϕ_{p} as

$$\phi_{\mathsf{p}} : \Lambda \to \mathbb{F}_{\mathsf{p}}^n,$$

$$\phi_{\mathsf{p}}\left(\gamma\sigma(x_1, \ldots, x_n)\right) = \left(\pi(x_1), \ldots, \pi(x_n)\right),$$

where γ is a scaling factor. The kernel $\Lambda_{\mathsf{p}} = \ker(\phi_{\mathsf{p}}) = \gamma\sigma(\mathfrak{p})^n$ and the point in Λ_{p} has a Euclidean norm at least $\gamma\sqrt{\frac{t}{2}}p^{\frac{1}{t}}$. Therefore, $\mathcal{B}\left(\gamma\sqrt{\frac{t}{8}}p^{\frac{1}{t}}\right) \subset \mathcal{V}(\Lambda_{\mathsf{p}})$.

For example, we can let $K = \mathbb{Q}(i), \mathcal{O}_K = \mathbb{Z}[i]$, $t = 2$, $\mathsf{p} = 5$ and $\mathfrak{p} = \langle 2+i \rangle$. Let $\Lambda \subset \mathbb{R}^4$ be the base lattice, $\gamma = 1$ and $\boldsymbol{\lambda} = (2, 2, 1, 1) \in \Lambda$ be a lattice point. Clearly, $\boldsymbol{\lambda} = \sigma(2+2i, 1+1i)$ and $\phi_{\mathsf{p}}(\boldsymbol{\lambda}) = (1, 3) \in \mathbb{F}_{\mathsf{p}}^2$. Also, $\mathcal{V}(\Lambda_{\mathsf{p}})$ can cover the ball $\mathcal{B}\left(\frac{\sqrt{5}}{2}\right)$. The process is visualized in Fig. 5.1. The yellow rectangle is $\mathcal{V}(\Lambda_{\mathsf{p}})$.

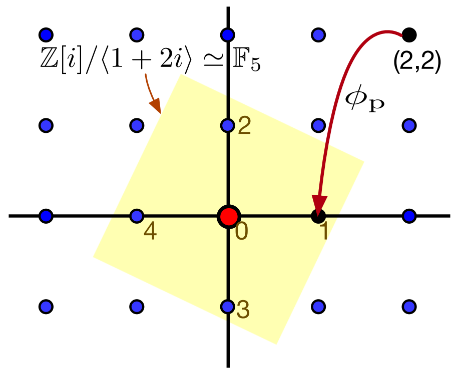

Figure 5.1: The visualization of ϕ_{p} when $\mathsf{p} = 5$.

5.6.2 An Example from $\mathbb{Z}[i]$

We select $K = \mathbb{Q}(i)$, $\mathcal{O}_K = \mathbb{Z}[i]$ and accordingly $t = 2$. We also select $\Lambda = \gamma\mathbb{Z}^{2n}$ and build the map $\phi_{\mathsf{p}} : \gamma\mathbb{Z}^{2n} \mapsto \mathbb{F}_{\mathsf{p}}^n$ as described in last section. Accordingly, $l = \gamma\sqrt{\frac{n}{2}}$ and $\mathcal{V}(\Lambda_{\mathsf{p}})$ can cover the ball $\mathcal{B}\left(\frac{\gamma}{2}\mathsf{p}^{\frac{1}{2}}\right)$. Then a sufficient condition for the requirements (5.4)-(5.6) is

$$\begin{aligned}
&\frac{\gamma^{2n}\mathsf{p}^{n-k_c}}{(\sqrt{2nP} - \gamma\frac{\sqrt{n}}{2})^{2n}V_{2n}} \le (1-\delta)^{2n}, \\
&\frac{(r_e + \gamma\frac{\sqrt{n}}{2})^{2n}V_{2n}}{\mathsf{p}^{n-k_f}V(\Lambda)} \le (1-\delta)^{2n}, \\
&\mathcal{B}\left(\sqrt{2nP}\right) \le \frac{\gamma}{2}\mathsf{p}^{\frac{1}{2}}.
\end{aligned}$$

Let p be the smallest prime larger than $n^{1+\eta}$ which statisfies $\mathsf{p} \bmod 4 = 1$. By Lemma 6, we can write $\mathsf{p} = \mu n^{1+\eta}$ where μ is a bounded constant. Let $\gamma = n^{-\frac{1}{3}\eta}$. We then assign

$$\begin{aligned}
k_c &= \left\lceil n\left(1 - \frac{2\log_2(\sqrt{2P}n^{\frac{1}{3}\eta} - \frac{1}{2}) + \log_2((1-\delta)nV_{2n}^{\frac{1}{n}})}{\log_2 \mathsf{p}}\right)\right\rceil, \\
k_f &= \left\lfloor n\left(1 - \frac{2\log_2(\sqrt{\frac{1}{n}r_e^2}n^{\frac{1}{3}\eta} + \frac{1}{2}) + \log_2(\frac{1}{1-\delta}nV_{2n}^{\frac{1}{n}})}{\log_2 \mathsf{p}}\right)\right\rfloor.
\end{aligned}$$

It's easy to verify that the rate of the scheme is

$$\lim_{n\to\infty}\frac{k_f - k_c}{2n}\log_2 \mathsf{p} = \frac{1}{2}\log_2\left(\frac{2P}{r_e^2}\right) = \frac{1}{2}\log_2\left(\frac{1+P/N}{1+\epsilon}\right),$$

where ϵ can be made arbitrarily small.

Chapter 6

Conclusions

In this thesis, we first adopt the unified approach to handle the proofs related to nested linear/lattice code. As a result, the achievability proof of nested lattice code is more accessible. We then extend the unified approach to the case of nested algebraic lattice codes constructed using the algebraic number theory and show they can achieve the AWGN channel capacity. This extension is the first step towards achieving the fading channel capacity under lattice encoding and decoding.

Potential future work includes achieving the ergodic fading channel capacity using algebraic lattice codes, optimizing the exponent of the growth rate of the prime p as a function of n, removing the spherical shaping technique.

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 27(4), 623–656, 1948.

[2] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[3] T. M. Cover and J. A. Thomas, "Elements of information theory 2nd edition," 2006.

[4] E. R. Berlekamp, *Algebraic coding theory.* World Scientific Publishing Co, 2015.

[5] G. D. Forney, "Coset codes. I. introduction and geometrical classification," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1123–1151, 1988.

[6] ——, "Coset codes. II. binary lattices and related codes," *IEEE Trans. Inf. Theory*, vol. 34, pp. 1152–1187, 1988.

[7] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.

[8] N. Alon and J. H. Spencer, *The probabilistic method.* John Wiley & Sons, 2004.

[9] P. Elias, "Coding for noisy channels," in *IRE Convetion Record*, 1955, vol. 3, part 4, pp. 37–46.

[10] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 219–221, 1979.

[11] M. P. Wilson, K. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5641–5654, Nov. 2010.

[12] W. Nam, S.-Y. Chung, and Y. H. Lee, "Capacity of the Gaussian two-way relay channel to within $\frac{1}{2}$ bit," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5488–5494, Nov. 2010.

[13] B. Nazer and M. Gastpar, “Compute-and-forward: Harnessing interference through structured codes,” *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6463–6486, 2011.

[14] U. Niesen and P. Whiting, “The degrees-of-freedom of compute-and-forward,” *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5214–5232, Aug. 2012.

[15] Y. Song and N. Devroye, “Lattice codes for the Gaussian relay channel: Decode-and-forward and compress-and-forward,” *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4927–4948, Sep. 2013.

[16] S. N. Hong and G. Caire, “Compute-and-forward strategies for cooperative distributed antenna systems,” *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227–5243, Sep. 2013.

[17] Z. Ren, J. Goseling, J. H. Weber, and M. Gastpar, “Maximum throughput gain of compute-and-forward for multiple unicast,” *IEEE Communication Letters*, vol. 18, no. 7, pp. 1111–1113, Jul. 2014.

[18] A. Chaaban, H. Maier, A. Sezgin, and R. Mathar, “Three-way channels with multiple unicast sessions: Capacity approximation via network transformation,” *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7086–7102, Dec 2016.

[19] A. Chaaban and A. Sezgin, “The approximate capacity region of the Gaussian Y-channel via the deterministic approach,” *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 939–962, Feb 2015.

[20] A. Chaaban, A. Sezgin, and A. S. Avestimehr, “Approximate sum-capacity of the Y-channel,” *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5723–5740, Sept 2013.

[21] G. Bresler, A. Parekh, and D. N. C. Tse, “The approximate capacity of the many-to-one and one-to-many Gaussian interference channel,” *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4566–4592, Sep. 2010.

[22] A. S. Motahari, S. O. Gharan, M. A. Maddah-Ali, and A. K. Khandani, “Real interference alignment: Exploring the potential of single antenna systems,” *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4799–4810, Aug 2014.

[23] U. Niesen and M. A. Maddah-Ali, “Interference alignment: From degrees-of-freedom to constant-gap capacity approximations,” *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4855–4888, Aug. 2013.

[24] A. Chaaban and A. Sezgin, "The approximate capacity region of the symmetric K-user Gaussian interference channel with strong interference," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2592–2621, May 2016.

[25] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3450–3482, Jun. 2014.

[26] I. Shomorony and S. Avestimehr, "Degrees of freedom of two-hop wireless networks: Everyone gets the entire cake," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2417–2431, May 2014.

[27] V. Ntranos, V. R. Cadambe, B. Nazer, and G. Caire, "Integer-forcing interference alignment," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013.

[28] A. Padakandla, A. G. Sahebi, and S. S. Pradhan, "An achievable rate region for the three-user interference channel based on coset codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1250–1279, Mar. 2016.

[29] D. Krithivasan and S. S. Pradhan, "Lattices for distributed source coding: Jointly Gaussian sources and reconstruction of a linear function," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5628–5651, Dec. 2009.

[30] ——, "Distributed source coding using Abelian group codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1495–1519, Mar. 2011.

[31] A. B. Wagner, "On distributed compression of linear functions," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 79–94, Jan. 2011.

[32] D. N. C. Tse and M. A. Maddah-Ali, "Interference neutralization in distributed lossy source coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, June 2010.

[33] Y. Yang and Z. Xiong, "Distributed compression of linear functions: Partial sum-rate tightness and gap to optimal sum-rate," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2835–2855, May 2014.

[34] X. He and A. Yener, "Providing secrecy with structured codes: Tools and applications to two-user Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2121–2138, Apr. 2014.

[35] S. Vatedka, N. Kashyap, and A. Thangaraj, "Secure compute-and-forward in a bidirectional relay," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2531–2556, May 2015.

[36] J. Xie and S. Ulukus, "Secure degrees of freedom of one-hop wireless networks," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3359–3378, Jun. 2014.

[37] A. Padakandla and S. S. Pradhan, "Achievable rate region based on coset codes for multiple access channel with states," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6393–6415, Oct 2017.

[38] ——, "Achievable rate region for three user discrete broadcast channel based on coset codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2267–2297, April 2018.

[39] S. Miyake and J. Muramatsu, "A construction of channel code, joint source-channel code, and universal code for arbitrary stationary memoryless channels using sparse matrices," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 92, no. 9, pp. 2333–2344, 2009.

[40] J. Muramatsu and S. Miyake, "Hash property and coding theorems for sparse matrices and maximum-likelihood coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2143–2167, May 2010.

[41] S. Miyake, "Coding theorems for point-to-point communication systems using sparse matrix codes." Ph.D. Thesis, University of Tokyo, Tokyo, Japan, 2010.

[42] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[43] R. de Buda, "The upper error bound of a new near-optimal code," *IEEE Transactions on Information Theory*, vol. 21, no. 4, pp. 441–445, 1975.

[44] ——, "Some optimal codes have structure," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 6, pp. 893–899, 1989.

[45] T. Linder, C. Schlegal, and K. Zeger, "Corrected proof of de buda's theorem (lattice channel codes)," *IEEE transactions on information theory*, vol. 39, no. 5, pp. 1735–1737, 1993.

[46] G. Poltyrev, "On coding without restrictions for the awgn channel," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 409–417, 1994.

[47] H.-A. Loeliger, "On the basic averaging arguments for linear codes," *Kluwer International Series in Engineering and Computer Science*, pp. 251–251, 1994.

[48] ——, "Averaging bounds for lattices and linear codes," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1767–1773, 1997.

[49] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the awgn channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 273–278, 1998.

[50] "LDA lattices without dithering achieve capacity on the Gaussian channel."

[51] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation and Multiuser Information Theory.* Cambridge University Press, 2014.

[52] S. H. Lim, C. Feng, A. Pastore, B. Nazer, and M. Gastpar, "A joint typicality approach to algebraic network information theory," *arXiv preprint arXiv:1606.09548*, 2016.

[53] O. Ordentlich and U. Erez, "A simple proof for the existence of "good" pairs of nested lattices," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4439–4453, Aug. 2016.

[54] R. Qi, C. Feng, and Y.-C. Huang, "A simpler proof for the existence of Capacity-Achieving nested lattice codes," in *2017 IEEE Information Theory Workshop (ITW) (IEEE ITW 2017)*, Kaohsiung, Taiwan, Nov. 2017.

[55] F. Oggier, E. Viterbo *et al.*, "Algebraic number theory and code design for rayleigh fading channels," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 3, pp. 333–415, 2004.

[56] A. Campello, C. Ling, and J. C. Belfiore, "Algebraic lattices achieving the capacity of the ergodic fading channel," in *2016 IEEE Information Theory Workshop (ITW)*, Cambridge, UK, Sept 2016, pp. 459–463.

[57] S. Lyu, A. Campello, and C. Ling, "Ring compute-and-forward over block-fading channels," *arXiv preprint arXiv:1805.02073*, 2018.

[58] A. Campello, "Random ensembles of lattices from generalized reductions," *IEEE Trans. Inf. Theory*, pp. 1–9, 2018.

[59] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

[60] D. S. Dummit and R. M. Foote, *Abstract algebra*. Wiley Hoboken, 2004, vol. 3.

[61] R. Breusch, "Zur verallgemeinerung des bertrandschen postulates, daß zwischen x und 2x stets primzahlen liegen," *Mathematische Zeitschrift*, vol. 34, no. 1, pp. 505–526, Dec 1932. [Online]. Available: https://doi.org/10.1007/BF01180606

[62] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, 2004.

[63] I. Stewart and D. Tall, *Algebraic number theory and Fermat's last theorem*. AK Peters/CRC Press, 2001.

Appendix

Appendix A

Entropy

We briefly introduce various definitions related to entropy.

Entropy. Let X be a discrete random variable with probability mass function (pmf) $p(x)$. The "uncertainty" about the outcome of X is measured by its entropy

$$H(X) = -\,\mathsf{E}_X(\log p(X)).$$

Conditional entropy. Let X, Y be two discrete random variables. Since $p(y|x)$ is a pmf, we can define $H(Y|X=x)$ for every x. The conditional entropy is the average of $H(Y|X=x)$ over every X, i.e.,

$$H(Y|X) = \sum_x H(Y|x)p(x) = -\,\mathsf{E}_{X,Y}(\log(p(Y|X))).$$

Joint entropy. Let (X, Y) be a pair of discrete random variables with pmf $p(x, y)$. The joint entropy is

$$H(X,Y) = -\,\mathsf{E}(\log p(X,Y)).$$

Mutual information. The mutual information between X and Y is

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}.$$

It can be shown

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y).$$

Appendix B

Typical Sequences

Here we present basics about typical sequences.

Let $\mathcal{X}$ be a discrete alphabet. For a vector $\boldsymbol{x} = (x_1, x_2, \dots, x_n) \in \mathcal{X}^n$, we define its *empirical pmf* as

$$\pi(x \mid \boldsymbol{x}) = \frac{|\{i : x_i = x\}|}{n} \quad \text{for } x \in \mathcal{X}.$$

For $X \in \mathcal{X} \sim p_X(x_i)$ and $\epsilon \in (0,1)$, define the set of ϵ-typical n-sequences $\boldsymbol{x} \in \mathcal{X}^n$ (or the typical set in short) as

$$\mathcal{T}_\epsilon^{(n)}(X) = \{\boldsymbol{x} : |\pi(x \mid \boldsymbol{x}) - p_X(x)| \leq \epsilon p_X(x) \quad \text{for all } x \in \mathcal{X}\}.$$

Let $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ be a random vector in $\mathcal{X}^n$ whose elements are i.i.d. random variables with each element $x_i \sim p_X(x_i)$, $i \in [1, n]$. Then by weak law of large numbers, for each $x \in \mathcal{X}$,

$$\pi(x \mid \boldsymbol{X}) \to p_X(x) \quad \text{in probability.}$$

Hence,

$$\lim_{n \to \infty} \mathsf{P}(\boldsymbol{X} \in \mathcal{T}_\epsilon^{(n)}(X)) = 1.$$

Intuitively, for any $\boldsymbol{x} \in \mathcal{T}_\epsilon^{(n)}(X)$, the empirical average $\frac{1}{n}\sum_{i=1}^n x_i$ should be close to the expectation $\mathsf{E}(X)$. In fact, we have a more general result as follows.

Lemma 17(Typical average lemma): Let $\boldsymbol{x} \in \mathcal{T}_\epsilon^{(n)}(X)$. Then for any non-negative function $g(\cdot)$ on $\mathcal{X}$,

$$(1-\epsilon)\,\mathsf{E}(g(X)) \leq \frac{1}{n}\sum_{i=1}^n g(x_i) \leq (1+\epsilon)\,\mathsf{E}(g(X)).$$

The proof is direct by noting $\frac{1}{n}\sum_{i=1}^n g(x_i) = \sum_{x \in \mathcal{X}} \pi(x \mid \boldsymbol{x}) g(x)$. Let $g(x) = -\log p_X(x)$ and note that $\mathsf{E}(-\log p_X(x)) = H(X)$, we obtain

$$2^{-n(1+\epsilon)H(X)} \le p_{\boldsymbol{X}}(\boldsymbol{x}) \le 2^{-n(1-\epsilon)H(X)}.$$

Equipped with this, we can bound the size of $\mathcal{T}_\epsilon^{(n)}(X)$. Note that the $\sum_{\boldsymbol{x}\in\mathcal{T}_\epsilon^{(n)}(X)} p_{\boldsymbol{X}}(\boldsymbol{x}) \le 1$, we obtain

$$|\mathcal{T}_\epsilon^{(n)}(X)| \le 2^{n(1+\epsilon)H(X)}.$$

Also note that by the law of large numbers,

$$\lim_{n\to\infty} \mathsf{P}(X \in \mathcal{T}_\epsilon^{(n)}(X)) = 1.$$

That is to say when n is sufficiently large, $\mathsf{P}(X \in \mathcal{T}_\epsilon^{(n)}(X)) \ge 1-\epsilon$. Hence,

$$|\mathcal{T}_\epsilon^{(n)}(X)| \ge (1-\epsilon)2^{n(1-\epsilon)H(X)}.$$

The notion of typical set can be extended to multiple random variables. For $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$, define their *joint empirical pmf* as

$$\pi(x, y \mid \boldsymbol{x}, \boldsymbol{y}) = \frac{|\{i : (x_i, y_i) = (x, y)\}|}{n} \quad \text{for } (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

Let $(X, Y) \sim p_{X,Y}(x, y)$. The set of jointly ϵ-typical n-sequences is defined as

$$\mathcal{T}_\epsilon^{(n)}(X, Y) = \{(\boldsymbol{x}, \boldsymbol{y}) : |\pi(x, y \mid \boldsymbol{x}, \boldsymbol{y}) - p_{X,Y}(x, y)| \le \epsilon p_{X,Y}(x, y) \quad \text{for all } (x, y) \in \mathcal{X} \times \mathcal{Y}\}.$$

Also define the set of conditionally ϵ-typical n-sequences as

$$\mathcal{T}_\epsilon^{(n)}(X \mid \boldsymbol{y}) = \{\boldsymbol{x} : (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{T}_\epsilon^{(n)}(X, Y)\}.$$

It can be shown that for sufficiently large n,

$$\forall \boldsymbol{y} \in \mathcal{Y}^n : |\mathcal{T}_\epsilon^{(n)}(X \mid \boldsymbol{y})| \le 2^{n(1+\epsilon)H(X|Y)}. \tag{B.1}$$