Construction of lattices for communications and security

Yanfei Yan

A Thesis Submitted in Fulfilment of Requirements for the Degree of

Doctor of Philosophy of Imperial College London and

Diploma of Imperial College

Communications and Signal Processing Group

Department of Electrical and Electronic Engineering

Imperial College London

2014

Statement of Originality

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Substantial parts of this thesis are believed to be original contributions to the field of information theory. This is also supported by publications. As far as the author is aware, the following aspects of the thesis are believed to be important and original contributions:

- **The equivalence between Λ/Λ' channel and the channel generated from the chain rule of mutual information in terms of polar coding (Lemma 4.6)**
- **Explicit construction of AWGN-good polar lattices (Theorem 3.1)**
- **Lattice Gaussian shaping scheme for polar lattices with any given SNR, which can be proved to achieve the channel capacity of the AWGN channel $\frac{1}{2}\log(1+\mathsf{SNR})$ (Theorem 4.6)**
- **Polar lattice coding scheme achieving the strong secrecy capacity of the Gaussian wiretap channel (Theorem 5.1)**

Abstract

In this thesis, we propose a new class of lattices based on polar codes, namely polar lattices. Polar lattices enjoy explicit construction and provable goodness for the additive white Gaussian noise (AWGN) channel, *i.e.*, they are *AWGN-good* lattices, in the sense that the error probability (for infinite lattice coding) vanishes for any fixed volume-to-noise ratio (VNR) greater than $2\pi e$. Our construction is based on the multilevel approach of Forney *et al.*, where on each level we construct a capacity-achieving polar code. We show the component polar codes are naturally nested, thereby fulfilling the requirement of the multilevel lattice construction. We present a more precise analysis of the VNR of the resultant lattice, which is upper-bounded in terms of the flatness factor and the capacity losses of the component codes. The proposed polar lattices are efficiently decodable by using multi-stage decoding. Design examples are presented to demonstrate the superior performance of polar lattices.

However, there is no infinite lattice coding in the practical applications. We need to apply the power constraint on the polar lattices which generates the polar lattice codes. We prove polar lattice codes can achieve the capacity $\frac{1}{2}\log(1+\mathsf{SNR})$ of the power-constrained AWGN channel with a novel shaping scheme. The main idea is that by implementing the lattice Gaussian distribution over the AWGN-good polar lattices, the maximum error-free transmission rate of the resultant coding scheme can be arbitrarily close to the capacity $\frac{1}{2}\log(1+\mathsf{SNR})$. The shaping technique is based on discrete lattice Gaussian distribution, which leads to a binary asymmetric

channel at each level for the multilevel lattice codes. Then it is straightforward to employ multilevel asymmetric polar codes which is a combination of polar lossless source coding and polar channel coding. The construction of polar codes for an asymmetric channel can be converted to that for a related symmetric channel, and it turns out that this symmetric channel is equivalent to an minimum mean-square error (MMSE) scaled Λ/Λ' channel in lattice coding in terms of polarization, which eventually simplifies our coding design.

Finally, we investigate the application of polar lattices in physical layer security. Polar lattice codes are proved to be able to achieve the strong secrecy capacity of the Mod-Λ AWGN wiretap channel. The Mod-Λ assumption was due to the fact that a practical shaping scheme aiming to achieve the optimum shaping gain was missing. In this thesis, we use our shaping scheme and extend polar lattice coding to the Gaussian wiretap channel. By employing the polar coding technique for asymmetric channels, we manage to construct an AWGN-good lattice and a secrecy-good lattice with optimal shaping simultaneously. Then we prove the resultant wiretap coding scheme can achieve the strong secrecy capacity for the Gaussian wiretap channel.

Acknowledgments

First and foremost, I want to express my deepest appreciation to my Ph.D. supervisor Dr. Cong Ling for his enlightening guidance, invaluable suggestions, generous support, and fatherlike encouragement. He not only teaches me academically with his in-depth professional knowledge, but also shows me effective methodologies for solving complicated problems systematically. More importantly, Dr. Ling sets a true example to me on how a world-class researcher should behave, which I firmly believe is my lifetime standard to contribute the scientific research community.

I would like to thank my examiners, Prof. Michael Huth from Imperial College London and Prof. Alister Burr from York University, for their valuable time and insightful questions. Without their comments and suggestions, this thesis would have been less worthy.

The long days at the Imperial College would not have been so pleasant without all the friends that I have met there: my first thought goes to Ling Liu, who received me as a friend in our office and with whom I have spent two wonderful years. I will never forget those long nights we spent together discussing polar codes. I can not finish this thesis without his help. I am also indebted to the wonderful colleagues and friends in the Communications and Signal Processing group at Imperial College for making my years in London much more enjoyable than it would have been otherwise.

I gratefully acknowledge the funding sources that made my Ph.D. work possible. Throughout my Ph.D., I was funded by UK-China scholarship for excellence and

European Commission FP7 project on physical layer security.

Lastly, I gratefully appreciate the most solid and fundamental energy source of my life: my wife, my baby and my parents. I am grateful for their unconditional love, care and support. They have always encouraged me to deal with life's challenges bravely and to persevere to the end. I would not have achieved anywhere near as much without them. I can never thank them enough for what they have done for me, but to say simple words - thank you!

Yanfei Yan

Imperial College London

November, 2014

Contents

Publications

- **Journal Papers**

1. "Polar lattice codes can achieve the strong secrecy capacity of the Gaussian wiretap channel," *In preparation.*
2. **Y. Yan**, L. Liu, C. Ling and X. Wu, "Construction of capacity-achieving codes for the AWGN channel: Polar lattice codes ," *submitted to IEEE Trans. Inform. Theory*, 2013.

- **Conference Papers**

1. **Y. Yan**, L. Liu and C. Ling, "Polar lattices for strong secrecy over the mod-Λ Gaussian wiretap channel," *in Proc. 2014 IEEE Int. Symp. Inform. Theory (ISIT)*, 2014, Hawaii, USA.
2. **Y. Yan**, and C. Ling, "Polar lattices: Where Arıkan meets Forney," *in Proc. 2013 IEEE Int. Symp. Inform. Theory (ISIT)*, 2013, Istanbul, Turkey.
3. **Y. Yan**, and C. Ling, "A construction of lattices from polar codes," *in Proc. IEEE Inform. Theory Workshop (ITW)*, 2012, Lausanne, Switzerland.
4. **Y. Yan**, C. Ling and J.-C. Belfiore, "Secrecy gain of trellis codes: The other side of the union bound," *in Proc. IEEE Inform. Theory Workshop (ITW)*, 2011, Paraty, Brazil.

List of Tables

List of Figures

Abbreviations

AWGN	Additive White Gaussian Noise
BEC	Binary Erasure Channel
BMA	Binary Memoryless Asymmetric Channel
BMS	Binary Memoryless Symmetric Channel
BP	Belief-Propogation
BSC	Binary Symmetric Channel
BW	Barnes-Wall
IC	Infinite Constellation
LDA	Low-Density Construction A lattice
LDLC	Low-Density Lattice Codes
LDPC	Low-Density Parity Check
ML	Maximum Likelihood
MMSE	Minimum Mean-Square Error
PDF	Probability Density Function
RV	Random Variable
SC	Successive Cancellation
SER	Symbol Error Rate
SNR	Signal-to-Noise Ratio
VNR	Volume-to-Noise Ratio

Notations

X	a RV X (All the random variables will be denoted by capital letters)
$P(X)$	the probability distribution of a RV X taking values in a set $\mathcal{X}$
$H(X)$	the entropy of a RV X
X_ℓ	a RV X at level ℓ in the multilevel coding scheme
x_ℓ^i	an i-th realization of X_ℓ
$x_\ell^{i:j}$	a vector $(x_\ell^i, ..., x_\ell^j)$, which is a realization of RVs $X_\ell^{i:j} = (X_\ell^i, ..., X_\ell^j)$
$x_{\ell:\jmath}^i$	a vector of the i-th RVs at levels from the ℓ to $\jmath$, i.e., the RVs $X_{\ell:\jmath}^i = (X_\ell^i, ..., X_\jmath^i)$
$\mathcal{I}^c$	the compliment set of the set $\mathcal{I}$
$\lvert\mathcal{I}\rvert$	the cardinality of the set $\mathcal{I}$
$[N]$	a set of all integers from 1 to N
$\tilde{W}$	a binary memoryless symmetric (BMS) channel
W^N	N independent uses of channel W
W_N	a combined channel of polar codes
$W_N^{(i)}$	an i-th subchannel generated by the channel combining and splitting of polar codes
$W_{\ell,N}$	a combined channel for the ℓ-th level
$W_{\ell,N}^{(i)}$	an i-th subchannel for the ℓ-th level
SNR_b (SNR_b)	SNR of the main (wiretap) channel

CHAPTER 1

Introduction

THe fundamental theorem of channel coding is undoubtedly the most important result of information theory which started with Claude Shannon's 1948 landmark paper [1]. A fast-decodable, structured code that could achieve the capacity (the Shannon limit) of well-understood channels such as the AWGN channel is the "holy grail" of coding theory. After more than 60 years, by standing on the shoulders of giants we propose a lattice coding scheme based on polar codes to achieve this final destination.

We then apply the above theory to the physical layer security. The issues of data confidentiality and security have taken on an increasingly important role in current communication systems. Traditionally, security is viewed as an independent design addressed above the physical layer, and all widely used cryptographic protocols are designed and implemented assuming the physical layer has already been established and provides an error-free link. Given that data security is so critically important, it is reasonable to argue that security measures should be implemented at all layers. Furthermore, with the emergence of ad-hoc and decentralized networks, higher-layer techniques, such as encryption and key distribution, are complex and difficult to implement. The wireless network especially needs the physical layer security due to its broadcast nature of wireless medium. Therefore my research also focuses on this promising area.

1.1 Latices and lattice codes for the AWGN channel

Lattice codes are useful in many communication scenarios with continuous-output channels, i.e., Gaussian channel. The first thing is to understand the difference between lattices and lattice codes. In practice, only a finite set of points of a lattice Λ can be used as a signal constellation in a communication system. This set consists of those points of Λ that are contained in a bounded shaping region S, and is known as the lattice code $\mathbb{C}(\Lambda, S)$ based on S and Λ. The performance of a lattice code $\mathbb{C}(\Lambda, S)$ on the AWGN channel depends not only on the underlying lattice Λ (packing problem, coding gain) [2] but also on the shape of the support region S (covering problem, shaping gain) [3]. A lattice code is generated by applying the power constraint to an infinite lattice.

It has been recognized in recent years that, at least for codes of moderate complexity on high-SNR AWGN channels, the problems of coding (packing) and constellation shaping are largely separable [2]. That is, if $\mathbb{C} = \Lambda \cap S$ is a discrete constellation consisting of the points in an infinite lattice Λ that lie in a shaping region S, then when the number of constellations is large the properties of the constellation are largely determined by the properties of Λ. This is also supported by the fact that the largest shaping gain is only 1.53 dB. In this case, it is usually assumed that the decoder is unaware of the shaping, i.e., it always decodes to the nearest lattice point, whether or not this point lies in S. Such a decoder will be called a lattice decoder. Note that the attractive symmetry properties commonly associated with lattice codes, such as congruent decoding regions, uniform distance profile, and codeword-independent error probability, apply only to a lattice decoder. If using minimum-distance (maximum-likelihood) decoding rather than lattice (infinite-constellation) decoding, many of the benefits of lattice structure are lost. Therefore, assuming infinite constellation can simplify the performance analysis of lattices.

The purpose of this section is to introduce some basic concepts of lattices for the AWGN channel to the reader. The system model is shown in Figure 1.1. It is a real channel, meaning that the coordinates of the noise are N i.i.d. random variables which follow a Gaussian distribution with average 0 and fixed variance σ^2. Moreover, the noise is independent of the channel input X^N. Intuitively, for any fixed noise variance σ^2, if all the lattice points (codewords) are far away from each other the communication can be reliable. However, this will require a large transmission power which is not a good assumption for practical reasons. The transmission power can not be infinite. There should be a power constraint over all the codewords:

$$\frac{1}{N}E\Big[\parallel X^N \parallel^2\Big] \leq P.$$

Therefore the tradeoff between the maximum reliable transmission rate and the transmission power is an interesting and challenge problem. The best tradeoff is known as the channel capacity which is a well-known result of Shannon [1] for the AWGN channel:

$$C = \frac{1}{2}\log(1+\frac{P}{\sigma^2}) = \frac{1}{2}\log(1+\mathsf{SNR}) \text{ bits per transmission.}$$

For both theoretical and practical reasons, we take the following two steps to tackle this problem. First we consider the unbounded codewords as the constellation and try to find a good structure of lattice points (the shape of its fundamental region) to deal with the Gaussian noise in Chapter 3. Then a practical shaping scheme for lattices is proposed in Chapter 4.

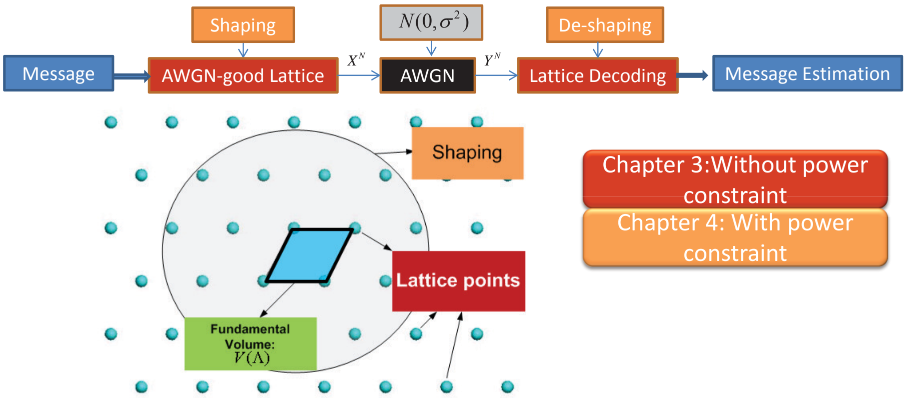


Figure 1.1: Lattices for the AWGN channel.

1.1.1 Lattices

Mathematically, a lattice is defined as a module over a certain ring and embedded in a vector space over a field. For our purposes, we will only consider real lattices, that is $\mathbb{Z}$-modules in the Euclidean space. And we will only deal with full rank lattices, that is n-dimensional lattices in an n-dimensional Euclidean space. They are a discrete subgroup of $\mathbb{R}^n$ which can be described by [4]

$$\Lambda = \{\boldsymbol{\lambda} = \mathbf{B}\mathbf{x} : \mathbf{x} \in \mathbb{Z}^n\},$$

where the columns of the generator matrix $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_n]$ are linearly independent.

For a vector $\mathbf{x} \in \mathbb{R}^n$, the nearest-neighbor quantizer associated with Λ is $Q_\Lambda(\mathbf{x}) = \arg\min_{\boldsymbol{\lambda}\in\Lambda} \parallel \boldsymbol{\lambda} - \mathbf{x} \parallel$. We define the modulo lattice operation by $\mathbf{x} \bmod \Lambda \triangleq \mathbf{x} - Q_\Lambda(\mathbf{x})$ [4]. The Voronoi region of Λ, defined by $\mathcal{V}(\Lambda) = \{\mathbf{x} : Q_\Lambda(\mathbf{x}) = 0\}$, specifies the nearest-neighbor decoding region. The Voronoi cell is one example of fundamental region of the lattice. A measurable set $\mathcal{R}(\Lambda) \subset \mathbb{R}^n$ is a fundamental region of the lattice Λ if $\cup_{\boldsymbol{\lambda}\in\Lambda}(\mathcal{R}(\Lambda) + \boldsymbol{\lambda}) = \mathbb{R}^n$ and if $(\mathcal{R}(\Lambda) + \boldsymbol{\lambda}) \cap (\mathcal{R}(\Lambda) + \boldsymbol{\lambda}')$ has measure 0 for any $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}'$ in Λ. The volume of a fundamental region is equal to that of the Voronoi region $\mathcal{V}(\Lambda)$, which is given by $V(\Lambda) =\mid \det(\mathbf{B}) \mid$. The minimum distance of a lattice Λ is $d_{\text{min}}(\Lambda) = \min_{x \in \Lambda} \mid x \mid$.

In this section, we only consider transmitting lattice points without power constraint over the AWGN channel. Since a lattice has infinite lattice points, it is known as infinite constellation (IC) or coding without power constraint which was proposed by Poltyrev [5]. Although the assumption is not realistic, it can provide some insights into the construction of good lattice codes for the power-constraint AWGN channel. If a lattice is "good“ under this framework, it is known as "AWGN-good“ (the formal definition will be given in Chapter 2). This scenario is simpler than the power constraint case in the sense that the decoding does not take account of the shaping region. Such a lattice decoder simply returns the closest lattice point to the decoder input. Due to the symmetry of the lattice, the performance of such a lattice decoder does not depend on the transmitting lattice points but only depends on the fundamental region of the lattice. Therefore both the transmitting (just sending all zeros) and decoding have been greatly simplified making it appealing for both theoretical analysis and practical implementation. Here we use a very simple lattice D_4 to illustrate the framework and its difference with conventional modulations. Again formal definitions of the lattice constructions from error correcting codes will be introduced in Chapter 2.

Example 1.1: D_4 is famous for the densest packing among 4 dimensional known lattices. It can be constructed from $(N = 4, k = 3)$ Reed-Muller Code by Construction A. The lattice partition is $\mathbb{Z}/2\mathbb{Z}$. The code formula is

$$D_4 = C(4, 3) + 2\mathbb{Z}^4.$$

Each lattice construction from error-correcting codes needs a lattice partition. This is different with the conventional modulations. The beauty of the lattice system

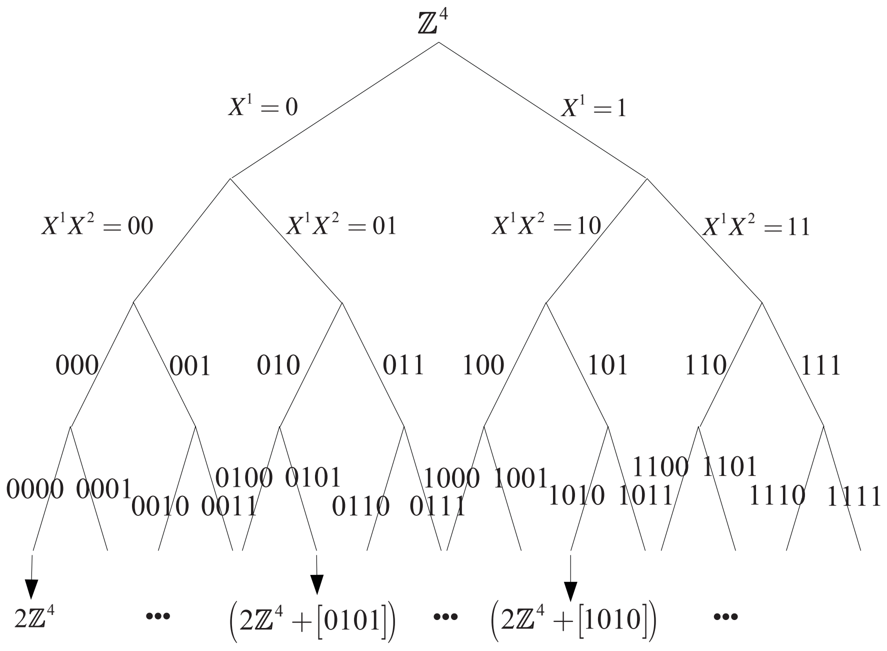


Figure 1.2: Lattice partitions induced by the $\mathbb{Z}^4/2\mathbb{Z}^4$

is that it is able to merge the channel coding and the modulation as one process. The lattice partition defines the available cosets and how to choose cosets to construct a lattice depends on the binary codes (a coset of a lattice can be simply regarded as a shift of this lattice). The process is depicted in Figure 1.2. We call $\mathbb{Z}$ and $2\mathbb{Z}$ as the top lattice and the bottom lattice in this partition tree. There are 2^3 codewords of this binary code. Therefore D_4 is the combination of the 2^3 cosets of $2\mathbb{Z}^4$ which are chosen from 2^4 cosets by the $(4,3)$ code (the number of the cosets is $\mid \mathbb{Z}^4/2\mathbb{Z}^4 \mid = 2^4$):

$$D_4 = \cup_{c_i \in C}(2\mathbb{Z}^4 + c_i).$$

Its generator matrix $\mathbf{B}$ is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

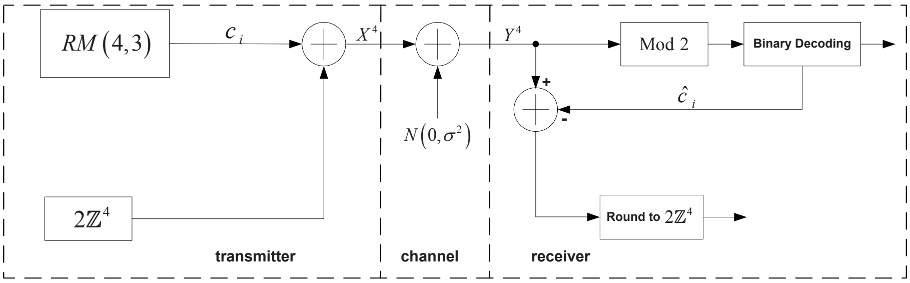


Figure 1.3: The encoding and decoding system of the D_4 lattice. The binary code is the Reed-Muller code with $N = 4$ and $k = 3$.

Then the transmitting symbols X^N for the AWGN channel are $[0, 0, 0, 0]$, $[1, 0, 0, 1]$, $[0, 0, 0, 2]$, $\cdots$. The encoding and decoding system is shown in Figure 1.3.

1.1.2 Lattice Codes

Since a lattice is infinite, shaping is needed to bound power. The common practice is to apply a finite shaping region.

Definition 1.1 (Lattice codes): Given a lattice $\Lambda \in \mathbb{R}^n$ and a bounded region $S \in \mathbb{R}^n$, a lattice code (or lattice constellation) $\mathbb{C}$ is the intersection of Λ and S:

$$\mathbb{C}(\Lambda, S) = \Lambda \cap S.$$

S is called the shaping region and, if M is the cardinality of the lattice code, its rate is defined as

$$R_{\mathbb{C}} = \frac{\log M}{n}.$$

The power of this lattice code is

$$P = \sum_{\lambda \in \Lambda \cap S} |\lambda|^2.$$

Here the shaping region can be generalized to the notion of a shaping technique.

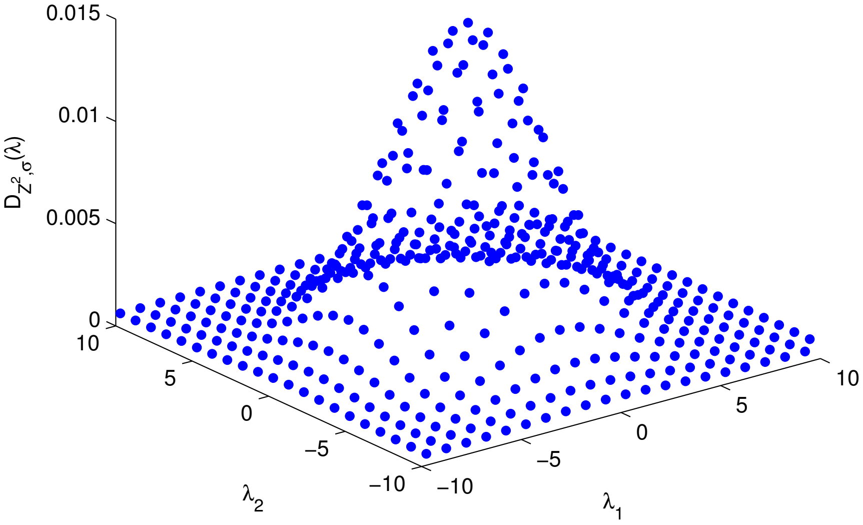


Figure 1.4: Discrete Gaussian distribution over $\mathbb{Z}^2$.

As long as we can control the transmitting power by selecting points from an infinite lattice, we obtain a lattice code. In this work, we use the probabilistic shaping over an infinite lattice proposed by [4]. The main idea is that the distribution of each coordinate of the input X^N $P_X(x)$ is a discrete Gaussian distribution. It is equivalent to a discrete Gaussian distribution over an N-dimensional lattice (IC). Although the constellation is infinite, both the rate and the transmission power are finite. Specifically, the rate of such lattice code is the entropy rate of the discrete Gaussian distribution $H(X)$. The transmission power is the variance of the discrete Gaussian distribution $E[|x|]^2$ (assume its mean is 0). A discrete Gaussian distribution is shown in Figure 1.4. The formal definition of the discrete Gaussian distribution will be given in Chapter 2.

Next we want to show the basic idea of the probabilistic shaping. The encoding process is shown in Figure 1.5. $X^{1:4} = U^{1:4}G$. If the distribution of U^i for $i = 1, 2, 3, 4$ is uniformly distributed, it is easy to verify that the input distribution $P_X(x = 0) = P_X(x = 1) = \frac{1}{2}$. If we assign the value of U^4 according to the

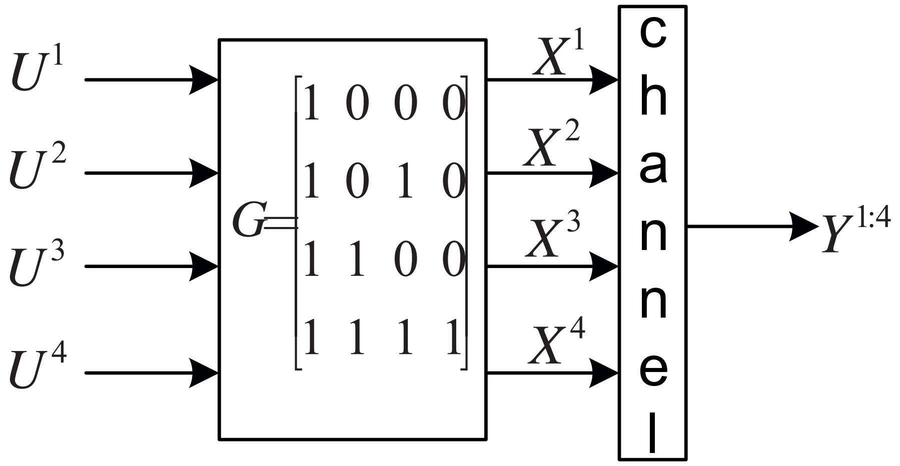

U^1
U^2
U^3
U^4
$G=\begin{bmatrix}1&0&0&0\\1&0&1&0\\1&1&0&0\\1&1&1&1\end{bmatrix}$
X^1
X^2
X^3
X^4
channel
$Y^{1:4}$

Figure 1.5: The encoding process of a polar code with block length $N = 4$.

following mappings:

$$U^4 = \begin{cases} 0 & U^3 = 0, \\ 1 & U^3 = 1, \end{cases}$$

the input distribution will certainly not be $P_X(x = 0) = P_X(x = 1) = \frac{1}{2}$. This example demonstrates that the key of the probability shaping is to find the connections between the input bits $U^{1:N}$. This is known as the inverse lossless source coding problem. Polar codes provide us a convenient tool to find these complicated mappings for any target distribution P_X and they have been proved to be optimal for this problem [6, 7]. Assume the target distribution of the input is P_X, as the block length $N \to \infty$, the output of the polar encoding process can generate a vector X^N and the distribution of each coordinate X^i for $1 \leq i \leq N$ is arbitrarily close to P_X. The simulation results are shown in Table 1.1. A shaping framework based on the polar source coding technique is presented in [8]. We will extend the framework to the multilevel codes to implement a discreet Gaussian distribution in Chapter 4.

Table 1.1: The simulation results of polar coding for the inverse lossless source coding problem. The distribution of each coordinate of X^N is getting closer to the target distribution $P_X(x = 0) = 0.89, P_X(x = 1) = 0.11$ as N increases. The settings of the simulations can be found in [6].

Block length N	P_X
256	$P_X(x = 0) = 0.8711, P_X(x = 1) = 0.1289$
1024	$P_X(x = 0) = 0.8838, P_X(x = 1) = 0.1162$
4196	$P_X(x = 0) = 0.8914, P_X(x = 1) = 0.1086$
8192	$P_X(x = 0) = 0.8899, P_X(x = 1) = 0.1101$

1.2 Road to the Capacity of the AWGN channel

Lattice codes are the counterpart of linear codes over a finite field (Hamming space) in the Euclidean space. Due to their large alphabets, lattice codes are useful in a wide range of applications in communications for the continuous channels with the Gaussian noise, such as theoretically achieving the AWGN channel capacity [4], information-theoretical security [9], compute-and-forward [10], and distributed source coding [11] (see [12, 13] for an overview). The story of lattice codes achieving the channel capacity of the AWGN channel can be traced back to de Buda's work at 1975 and followed by some corrections of his work and Poltyrev's proof [5]. All these results claimed that the achievable rate of lattice codes is $\frac{1}{2}\log(\mathsf{SNR})$ rather than the real channel capacity $\frac{1}{2}\log(1+\mathsf{SNR})$. This achievable rate can be explained by the Minkowski-Hlawka theorem. We recommend [14] to readers for a detailed introduction. After nearly 50 years, Erez and Zamir successfully proved that lattice codes can achieve this $\frac{1}{2}\log(1+\mathsf{SNR})$ by using MMSE scaling and Voronio shaping [15]. But they need AWGN-good lattices and quantization-good lattices, which are nonconstructive at that time. After another 10 years, Ling and Belfiore proved that lattice codes can achieve the capacity with only AWGN-good lattices and a discrete lattice Gaussian shaping [4]. This greatly simplifies the task but still the construction is not explicit.

On the other hand, there is a major breakthrough in the binary coding theory. Polar codes, proposed by Arıkan in [16], can provably achieve the capacity of binary memoryless symmetric (BMS) channels. Then there are considerable efforts to generalize polar codes to discrete memoryless channels and to nonbinary polar codes [17, 18, 19, 20]. An attempt from the theoretical side to construct polar codes for the AWGN channel was given in [21, 22], based on nonbinary polar codes and on the technique for the multi-access channel. The polar codes for asymmetric channels were introduced in [8] which provided an efficient shaping technique for symmetric polar codes. However, it is still an open problem to construct practical polar codes to achieve the capacity of the AWGN channel. In this thesis, we propose polar lattices to fulfil this goal, based on a combination of binary polar codes, lattice codes and discrete lattice Gaussian shaping. The main advantage of lattice Gaussian shaping is that it allows for multi-stage decoding. Our approach is different from the standard Voronoi shaping which involves a quantization-good lattice [15]. We note however that the explicit construction of quantization-good lattices is not available in literature. In contrast, the proposed Gaussian shaping does not require such a quantization-good lattice, therefore bypassing this difficulty.

1.3 Physical Layer Security

Information-theoretic physical layer security [23] is a technique that exploits the channel difference between the legitimate receiver and the eavesdropper (passive attacker) to provide security. Its security comes from information theoretic security. By adding randomness and modify the conventional channel codes, perfect secrecy can be ensured under a certain channel condition; in other words, the eavesdropper cannot obtain any useful information from the received signal. An intuitive example is shown here. Assume the confidential message is S and let ξ be a uniform binary random variable which is independent of S. Let the transmitted message be $X = (\xi, \xi \oplus S)$. The addition is modulo two addition. If the eavesdropper only can decode either coordinate of X due to the limitation of his channel capacity, he gains no information about S, hence the perfect secrecy has been achieved. The legitimate receiver, however, can obtain S by adding the two components of X because he has bigger channel capacity. This scheme shows the basic idea of physical layer security exploiting the difference of channels (channel capacity in this example) to provide secrecy.

The advantage of physical layer is that its security is based on the information theoretic security. It does not impose any restrictions on the computational power of the eavesdropper. Existing cryptosystems based on the hardness of certain computational problems are vulnerable to quantum attacks. Namely, if a large-scale quantum computer is successfully built (which looks increasingly likely in the future), then some problems such as integer factorization will be solved easily. The consequence is that existing crypto systems will be considerably easier to break. This puts existing crypto schemes at high risk.

Therefore, considerable attention has recently been paid on the research of physical layer security. The theoretical idea was originated from Shannon's notion of perfect secrecy [24]. Perfect security can be achieved if the encoding of information bits M into a transmitted codeword X is such that the mutual information $I(M; X) = 0$. Later Wyner [23] proved that both robustness to transmission errors and a prescribed degree of data confidentiality could be attained by channel coding technology without any key bits if the transmitted channels satisfy some conditions. Leung-Yan-Cheong extended this conclusion to Gaussian Wiretap Channel in [25]. Since Gaussian channels is the most fundamental channel model in communication theory, we restrict ourselves to the Gaussian wiretap channel in this work. The sys-

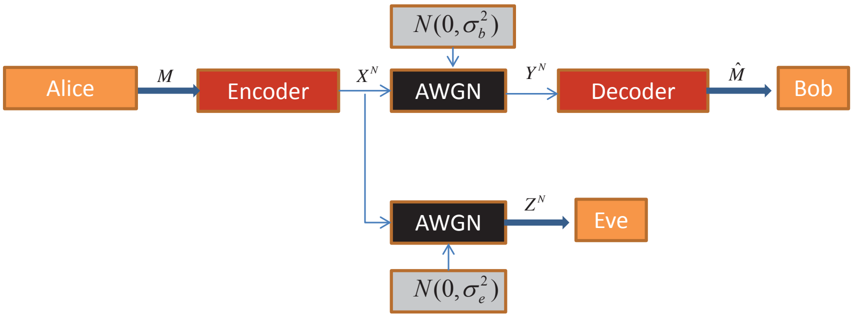


Figure 1.6: The Gaussian Wiretap channel.

tem model is depicted in Figure 1.6. A practical scenario for this model is that Bob connects to WIFT inside his room and Eve tries to eavesdrop on conversations outside the room. The Rayleigh fading wiretap channel is our future work. Alice wants to send information to Bob. Eve is an eavesdropper. The channel between Alice and Bob is called the main channel C_1. The channel between Alice and Eve is called the wiretapper channel C_2. Wyner showed that if the difference in terms of channel capacity between C_1 and C_2 is positive it is possible to achieve perfect secrecy. Then Csiszär and Körner [26] showed that the secrecy can be ensured for the cases when C_1 is less noisy than C_2 in wiretap channel.

The information theoretic analysis on physical layer security is all about secrecy capacity. Strong secrecy capacity is a theoretic limit of the transmission rate that can guarantee both reliability between Alice and Bob and strong secrecy $I(M;Z^N) \to 0$ between Alice and Eve. This theoretic limit can be a guideline for the practical coding design as what happened in the channel coding history. An exposition of progress in this area can be found in [27]. Much attention has been paid on the research on secrecy capacity of various kinds of channels, including fading wiretap channels [28], multi-input multi-output (MIMO) wiretap channels [29], multi-access channels [30], broadcast channels with confidential messages [31], and interference channels with confidential messages [32]. In this paper, we aim to propose a coding scheme that can achieve the strong secrecy capacity of the Gaussian wiretap channel.

1.3.1 Lattice codes for the Gaussian Wiretap Channel

The first wiretap code constructions were proposed in [33], where the nested codes were proved to achieve the secrecy capacity. The lattice codes have large alphabets. Furthermore, lattice codes have nested structure which are essential to the coding scheme for secrecy. Therefore we believe lattice codes are good candidates to provide secrecy. There has been some progress in wiretap lattice coding for the Gaussian wiretap channel. On the theoretical aspect, the achievable rate for lattice coding achieving weak secrecy over the Gaussian wiretap channel has been derived [34]. Furthermore, the existence of lattice codes approaching the strong secrecy capacity was demonstrated in [9]. On the practical aspect, wiretap lattice codes were proposed in [35] to maximize the eavesdropper's decoding error probability.

1.4 Research Challenges and Objectives

1.4.1 Challenges

To some extent, the major issues in the road to secrecy capacity are similar to those that exist in the path to channel capacity: the random coding arguments used to prove the achievability of capacity do not provide explicit code constructions. For the record, in 1993, the remarkable discovery of turbo codes [36] with their associated iterative decoding phenomenally brought the best performance of known codes so close to the Shannon limit that probably no one could have expected. It took nearly 60 years to design a practical code to achieve the channel capacity since Shannon's proof. It took even longer to propose the explicit capacity-achieving polar codes. However, the same thing will not happen in terms of codes for secrecy. This is because the achievements in the channel coding may help us to construct powerful codes for secrecy. Note that the codes used for secrecy of course should be different with the original channel codes. The channel codes call for the introduction of redundancy to resist the effect of channel noise; on the other hand, creating too much redundancy is likely to leak some information to the eavesdropper. Therefore the design of codes for secrecy must take into account both reliability and secrecy.

Compared with the explosive outcome on the information theoretic analysis in various wiretap channels, the problem of designing practical coding scheme for secure communication over wiretap channels has not received much attention. The design of codes for secrecy turns out to be surprisingly difficult. Although it has been proved that polar codes can achieve the secrecy capacity of the binary symmetric wiretap channels, no practical wiretap code constructions can achieve secrecy capacity of the Gaussian wiretap channel. This great challenge have fostered the development of alternative secrecy metrics in Gaussian wiretap channels. For example, the decoding error probability of the eavesdropper. If the decoding error probability of the eavesdropper goes to one, then his probability of correct decision tends to zero. The most exploited approaches to design practical Gaussian wiretap codes so far is to use LDPC codes [37] and unimodular lattice codes [35] using this secrecy metric. However, the connection between the decoding error probability and the leakage mutual information is still not clear.

1.4.2 Objectives

In this thesis, the objectives are to address the above issues by solving the challenging problems below:

(i) What is the role of the MMSE scaling to achieve the capacity of the AWGN channel? Is it necessary?

(ii) How to construct lattice codes to achieve the capacity of the AWGN channel $\frac{1}{2}\log(1+\mathsf{SNR})$ for any given SNR?

(iii) How to construct lattice codes to achieve the strong secrecy capacity for the

Gaussian wiretap channel?

1.5 Summary of Contributions

We study all above problems by establishing efficient constructions, encoding and decoding algorithms. The significance of this research is to provide insights and guidance for building reliable and secure coding scheme in real communication systems. Our contributions are four-fold:

- The equivalence between Λ/Λ' channel and the channel generated from the chain rule of mutual information in terms of polar coding (Lemma 4.6)
- Explicit construction of AWGN-good polar lattices (Theorem 3.1)
- Explicit discrete Gaussian shaping scheme for AWGN-good polar lattices, which can be proved to achieve the channel capacity of the AWGN channel $\frac{1}{2}\log(1+\mathsf{SNR})$ for any SNR (Theorem 4.6)
- Explicit discrete Gaussian shaping scheme for Secrecy-good polar lattices achieving the strong secrecy capacity of the Gaussian wiretap channel (Theorem 5.1)

1.6 Thesis Organization

The organization of the rest of the thesis is as follows.

In Chapter 2, we provide the literature review for lattice constructions from error-correcting codes. Chapter 3 presents a proof that polar lattices are AWGN-good. In regards to the power constraint, Chapter 4 addresses an efficient shaping technique for polar lattices such that the resultant polar lattice codes can achieve the channel capacity of the AWGN channel $\frac{1}{2}\log(1+\mathsf{SNR})$. In terms of security, Chapter 5 proves that polar lattice codes can achieve the strong secrecy capacity of the Gaussian wiretap channel. Finally, Chapter 6 concludes the thesis and identifies future work.

CHAPTER 2

Backgrounds on Lattices

In this chapter, we restrict ourselves to the construction of lattices which are good for AWGN channel coding without power constraint (its formal definition “AWGN-good" is given in the following sections). This is because such lattices are the most important lattices in communication systems. There are many other goodness of lattices, such as quantization-good lattices [38], secrecy-good lattices [9] and capacity-good lattices [4]. The following is the basics of lattices and the definition of AWGN-good lattices.

2.1 Basics

For $\sigma > 0$, we define the noise distribution of the AWGN channel with zero mean and variance σ^2 as

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

for all $\mathbf{x} \in \mathbb{R}^n$. Given σ, the volume-to-noise ratio (VNR) of an n-dimension lattice Λ is defined by

$$\gamma_\Lambda(\sigma) \triangleq \frac{V(\Lambda)^{\frac{2}{n}}}{\sigma^2}.$$

We also need the Λ-periodic function

$$f_{\sigma,\Lambda}(\mathbf{x}) = \sum_{\boldsymbol{\lambda} \in \Lambda} f_{\sigma,\boldsymbol{\lambda}}(\mathbf{x}) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \sum_{\boldsymbol{\lambda} \in \Lambda} e^{-\frac{\|\mathbf{x}-\boldsymbol{\lambda}\|^2}{2\sigma^2}},$$

for all $\mathbf{x} \in \mathbb{R}^n$.

We note that $f_{\sigma,\Lambda}(\mathbf{x})$ is a probability density function (PDF) if $\mathbf{x}$ is restricted to the the fundamental region $\mathcal{R}(\Lambda)$. This distribution for $\mathbf{x} \in \mathcal{R}(\Lambda)$ is actually the PDF of the Λ-aliased Gaussian noise, i.e., the Gaussian noise after the mod-Λ operation [39]. It gets flat as σ increases as shown in Figure 2.1. In order to describe such phenomenon, the flatness factor of a lattice Λ is defined as [9]

$$\epsilon_\Lambda(\sigma) \triangleq \max_{\mathbf{x} \in \mathcal{R}(\Lambda)} \mid V(\Lambda) f_{\sigma,\Lambda}(\mathbf{x}) - 1 \mid,$$

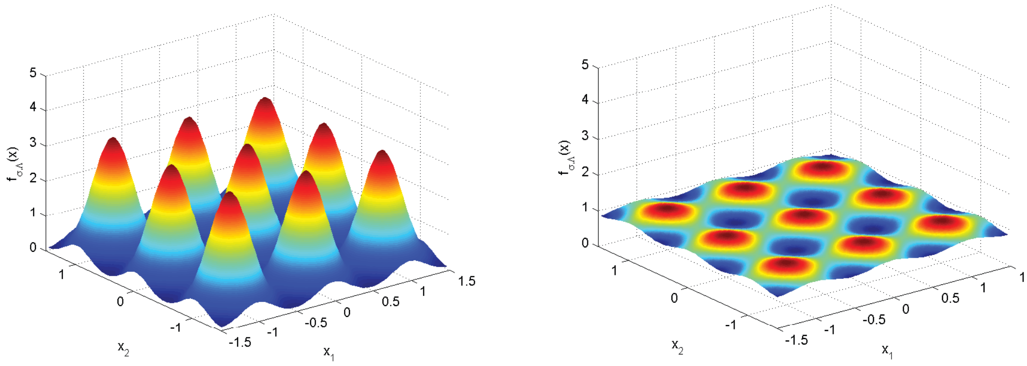
where $f_{\sigma,\Lambda}(\mathbf{x}) \to \frac{1}{V(\Lambda)}$ when it approaches uniform distribution.

We define the discrete Gaussian distribution over Λ centered at $\mathbf{c}$ as the discrete distribution taking values in $\lambda \in \Lambda$:

$$D_{\Lambda,\sigma,\mathbf{c}}(\lambda) = \frac{f_{\sigma,\mathbf{c}}(\lambda)}{f_{\sigma,\mathbf{c}}(\Lambda)}, \ \forall \lambda \in \Lambda,$$

where $f_{\sigma,\mathbf{c}}(\Lambda) = \sum_{\lambda \in \Lambda} f_{\sigma,\mathbf{c}}(\lambda)$. For convenience, we write $D_{\Lambda,\sigma} = D_{\Lambda,\sigma,\mathbf{0}}$. This distribution has been proved to achieve the optimum shaping gain when the flatness factor is small [9].

A sublattice $\Lambda' \subset \Lambda$ induces a partition (denoted by Λ/Λ') of Λ into equivalence



(a) When σ is small, the effect of aliasing becomes insignificant and the Λ-aliased Gaussian density $f_{\sigma,\Lambda}(\mathbf{x})$ approaches the Gaussian distribution. The flatness factor $\epsilon_\Lambda(\sigma)$ is large.

(b) When σ is large, $f_{\sigma,\Lambda}(\mathbf{x})$ approaches the uniform distribution. The flatness factor $\epsilon_\Lambda(\sigma)$ is small.

Figure 2.1: The comparison of the Λ-aliased Gaussian distributions with different flatness factors.

groups modulo Λ'. The order of the partition is denoted by $|\Lambda/\Lambda'|$, which is equal to the number of the cosets. If $|\Lambda/\Lambda'| = 2$, we call this a binary partition. Let $\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda_r$ for $r \geq 2$ be an n-dimensional lattice partition chain. If only one level is applied ($r = 2$), the construction is known as "Construction A". If multiple levels are used, the construction is known as "Construction D" [40, p.232]. For each partition $\Lambda_\ell/\Lambda_{\ell+1}$ $(1 \leq \ell \leq r-1)$ a code C_ℓ over $\Lambda_\ell/\Lambda_{\ell+1}$ selects a sequence of coset representatives a_ℓ in a set A_ℓ of representatives for the cosets of $\Lambda_{\ell+1}$. This construction requires a set of nested linear binary codes C_ℓ with block length N and dimension of information bits k_ℓ which are represented as $[N, k_\ell]$ for $1 \leq \ell \leq r-1$ and $C_1 \subseteq C_2 \cdots \subseteq C_{r-1}$. Let ψ be the natural embedding of $\mathbb{F}_2^N$ into $\mathbb{Z}^N$, where $\mathbb{F}_2$ is the binary field. Let $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_N$ be a basis of $\mathbb{F}_2^N$ such that $\mathbf{b}_1, \cdots \mathbf{b}_{k_\ell}$ span C_ℓ. When $n = 1$, the binary lattice L consists of all vectors of the form

$$\sum_{\ell=1}^{r-1} 2^{\ell-1} \sum_{j=1}^{k_\ell} \alpha_j^{(\ell)} \psi(\mathbf{b}_j) + 2^{r-1}\mathbf{l}, \tag{2.1}$$

where $\alpha_j^{(\ell)} \in \{0, 1\}$ and $\mathbf{l} \in \mathbb{Z}^N$.

A mod-Λ channel is a Gaussian channel with a modulo-Λ operator in the front end [39]. The capacity of the mod-Λ channel is [39]

$$C(\Lambda, \sigma^2) = \log V(\Lambda) - h(\Lambda, \sigma^2), \tag{2.2}$$

where $h(\Lambda, \sigma^2)$ is the differential entropy of the Λ-aliased noise over $\mathcal{V}(\Lambda)$:

$$h(\Lambda, \sigma^2) = -\int_{\mathcal{V}(\Lambda)} f_{\sigma,\Lambda}(\mathbf{x}) \log f_{\sigma,\Lambda}(\mathbf{x}) d\mathbf{x}.$$

The differential entropy is maximized to $\log V(\Lambda)$ by the uniform distribution over $\mathcal{V}(\Lambda)$. It is known that the Λ/Λ' channel (i.e., the mod-Λ' channel whose input is drawn from $\Lambda \cap \mathcal{V}(\Lambda')$) is regular, and the optimum input distribution is uniform [39]. Furthermore, the Λ/Λ' channel is a BMS if $|\Lambda/\Lambda'| = 2$ [41]. The capacity of the Λ/Λ' channel for Gaussian noise of variance σ^2 is given by [39]

$$\begin{aligned} C(\Lambda/\Lambda', \sigma^2) &= C(\Lambda', \sigma^2) - C(\Lambda, \sigma^2) \\ &= h(\Lambda, \sigma^2) - h(\Lambda', \sigma^2) + \log(V(\Lambda')/V(\Lambda)). \end{aligned}$$

2.2 AWGN-goodness of Lattices

In this section, we give an introduction about the AWGN-goodness of the lattices (infinite constellation (IC)). It is the best possible tradeoff between the volume of a lattice and the error probability $P_e(L, \sigma^2)$ when transmitting in the additive white Gaussian noise (AWGN) channel without power restriction. It is also known as achieving the Poltyrev capacity [5] or sphere-bound-achieving lattices [39]. In this thesis, we adapt these terms under different context accordingly.

Let $V(\Lambda)$ be the fundamental volume of Λ, which is the volume of the Voronoi region of Λ. Packing radius $r_{\Lambda}^{\text{pack}}$ shown in Figure 2.2 is the radius of the largest n-

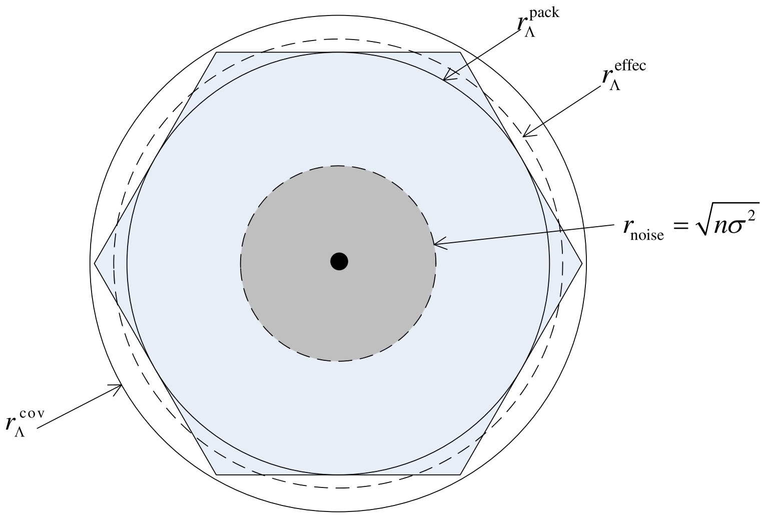

r_Λ^{pack}
r_Λ^{effec}
$r_{\text{noise}} = \sqrt{n\sigma^2}$
r_Λ^{cov}

Figure 2.2: Geometric picture of lattices [12].

dimensional ball contained in the Voronoi region of Λ. $r_\Lambda^{\text{pack}} = \frac{d_{\text{min}}(\Lambda)}{2}$, where $d_{\text{min}}(\Lambda)$ is the minimum distance between two lattice points of Λ.

The effective radius r_Λ^{effec} shown in Figure 2.2 is the radius of a sphere with the volume $V(\Lambda)$. It is known [40, p9] that the volume of a unit sphere V_n in $\mathbb{R}^n$ is

$$V_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} = \begin{cases} \dfrac{\pi^k}{k!}, & n = 2k \\ \dfrac{2^n \pi^k k!}{n!}, & n = 2k+1 \end{cases}$$

where $\Gamma(t) = \int_0^\infty u^{t-1}e^{-u}du$ is the gamma function.

Since $V_n(r_\Lambda^{\text{effec}})^n = V(\Lambda)$, the effective radius r_Λ^{effec} is

$$r_\Lambda^{\text{effec}} = \frac{V(\Lambda)^{1/n}}{V_n^{1/n}} = \frac{V(\Lambda)^{1/n}\Gamma(\frac{n}{2}+1)}{\sqrt{\pi}}.$$

Let σ^2 be the variance of the Gaussian noise. Best possible performance is achieved when the Voronoi regions of the lattice is approximately a sphere as $n \to \infty$. For example, the error probability is lower bounded by the probability that the noise leaves a sphere with the same volume as a Voronoi region. In other words, as n grows, the Voronio regions of the optimal lattice becomes closer to a sphere with squared radius that is equal to the mean squared radius of the noise, $n\sigma^2$. Therefore, a plausible way to describe this goodness with presence of the noise would be to measure the ratio between the squared effective radius of the lattices and the expected square noise amplitude [42], i.e.

$$\alpha^2(\Lambda,\sigma^2) = \frac{(r_\Lambda^{\text{effec}})^2}{n\sigma^2} \approx \frac{V(\Lambda)^{\frac{2}{n}}}{2\pi e\sigma^2},$$

where the approximation is obtained by using the Stirling approximation of $k! \approx (k/e)^k$ for even n. This is the volume-to-noise ratio (VNR).

We are concerned with the block error probability of lattices $P_e(\Lambda,\sigma^2)$. It is the probability $\mathbb{P}\{X^n \notin \mathcal{V}(\Lambda)\}$ that an n-dimensional independent and identically distributed (i.i.d.) Gaussian noise vector X^n with zero mean and variance σ^2 per dimension falls outside the Voronoi region of Λ.

Then we are ready to introduce the notion of lattices which are good for the AWGN channel without power constraint:

Definition 2.1 (AWGN-good [12]): A sequence of lattices $\Lambda^{(n)}$ of increasing dimension n is AWGN-good if, for any fixed $P_e(\Lambda^{(n)},\sigma^2) \in (0,1)$,

$$\lim_{n\to\infty} \gamma_{\Lambda^{(n)}}(\sigma) = 2\pi e$$

and if, for a fixed VNR greater than $2\pi e$, $P_e(\Lambda^{(n)},\sigma^2)$ goes to 0 as $n \to \infty$.

2.3 Constructions of AWGN-good lattices from error-correcting codes

There are many ways to construct lattices. For example, in mathematics, people construct lattices from sphere packing theory. In cryptography, people construct lattices from group theory (ring). However, in communications, most lattices are constructed from error-correcting codes by the coset codes construction [2]. In other words, using the codewords of the error-correcting codes to choose cosets which can be combined to a lattice. The following table is the summary of current lattice constructions from error-correcting codes. [1]

Table 2.1: Lattice constructions from error-correcting codes

Name	References	Descriptions	Lattices
Construction A	[43] [40, p137]	Single level, binary and non-binary codes	D_4, $(8,4,4)$ Hamming Code and E_8, Random mod-p lattices [14] [15], LDA lattices [44].
Construction B	[43] [40, p141]	Single level, the weight of the lattice vector must be divisible by 4	$(8,1,8)$ Repetition Code and E_8, $(16,5,8)$ Reed-Muller codes and Barnes-Wall lattice Λ_{16}, $(24,12,8)$ Golay code and Leech lattice Λ_{24},.
Construction D	[45] [40, p232]	Multilevel, nested binary linear codes, deals with generator matrix	Barnes-Wall lattices, Polar lattices [41], Turbo lattices [46].
Construction D'	[45] [40, p235]	Multilevel, nested binary linear codes, deals with parity-check matrix	LDPC lattices [47]
Construction E	[48] [40, p236]	Multilevel, nested binary linear codes, higher dimensional lattice partition	Polar lattices [41]

The following subsections will give a brief introduction of current AWGN-good practical lattices. All the lattices can be put in the frame known as finite dimensional infinite constellations [49].

2.3.1 Single level constructions

In this subsection, we focus on Construction A. This is because Construction B is limited to the class of codes with even Hamming weight of each codeword. More details of Construction B can be found in [40, p141,p191].

The general definition of Construction A can be found at [40, p211]. Let $C(N,k,d_{min})$ be an $\mathbb{F}_p$-linear code of length N, dimension k and rate $R = k/N$. Let $\Pi : \Lambda^N \rightarrow$

[1] We do not use Construction B to construct lattices because it requires stringent condition on the weight of codewords which makes it impractical. We omit Construction C because it generates nonlattice packing.

$(\Lambda/\Lambda')^N$ be the natural projection, the lattice L generated from Construction A is defined as:

$$L = \{\overline{x} \in \Lambda^N \mid \Pi(\overline{x}) \in C\}.$$

The most simple case for Construction A is $\Lambda/\Lambda' = \mathbb{Z}/p\mathbb{Z}$, namely mod-$p$ lattices [14]. The fundamental volume of such a lattice is

$$V(L) = p^{N-k}.$$

The generator matrix for L has the form

$$\begin{pmatrix} I_k & \Phi(B) \\ 0 & pI_{N-k} \end{pmatrix}$$

where $(I_k \;\; B)$ is a $k \times N$ generator matrix in systematic form for the code C and where $\Phi : \mathbb{F}_p \to \mathbb{Z}$ is a natural embedding of $\mathbb{F}_p$ into $\mathbb{Z}$. The minimum distance of L is

$$d_{\text{min}}(L) = \min\{\sqrt{d_{\text{min}}(C)}, d_{\text{min}}(\Lambda')\} = \min\{\sqrt{d_{\text{min}}(C)}, p\}.$$

Remark 2.1: This is why Construction A with $p = 2$ cannot be successful for large dimension (the coding gain $\theta(L) = \frac{d^2_{\text{min}}(L)}{V(L)^{2/N}}$ cannot be very large). The minimum distance of the mod-p lattices can be improved if using a more powerful code C and a larger finite field size p. Another direction is to use a set of nested binary codes for the multilevel construction.

2.3.1.1 Gosset lattice E_8

E_8 is famous for the densest packing among 8 dimensional known lattices. It can be constructed from $(8, 4, 4)$ Hamming Code by Construction A. The lattice partition is $\mathbb{Z}/2\mathbb{Z}$. The code formula is

$$L = C(8, 4, 4) + 2\mathbb{Z}^8.$$

There are 2^4 codewords of this code. From the viewpoint of coset codes [2], this construction can be interpreted as that E_8 is the combination of the 2^4 cosets of $2\mathbb{Z}^8$ which are chosen from 2^8 cosets by the $(8, 4, 4)$ Hamming Code (the number of the cosets is $\mid \mathbb{Z}^8/2\mathbb{Z}^8 \mid= 2^8$).

2.3.1.2 Leech lattice Λ_{24}

Let $C \subseteq \mathbb{F}_2^N$ be a binary code with the property that all codewords have even weight. The lattice L constructed by Construction B is defined as [40, p141]

$$L = \{\overline{x} \in \Lambda^N; \Pi(\overline{x}) \in C; \sum_{i=1}^{N} x_i \equiv 0 \text{ mod } 4\},$$

where $d_{\text{min}}(L) = \min\{\sqrt{d_{\text{min}}(C)}, \sqrt{8}\}$ if $\Lambda/\Lambda' = \mathbb{Z}/2\mathbb{Z}$.

Taking C as the extended binary Golay code with parameters $(24, 12, 8)$. This generates a lattice L_e which is half of the famous Leech lattice. Consider the translate [50, p366]

$$L_o = L_e + ((1/2)^{23}, -3/2).$$

It is easy to know that any pair of lattice points $x \in L_e$ and $y \in L_o$ differ by squared

Euclidean distance at least

$$23 \cdot (1/2)^2 + (3/2)^2 = 8.$$

This follows because each codeword in the Golay code has even weight, so there must always be at least one coordinate where x and y differ by $3/2$. But 8 is also the minimum squared distance in each one of L_e and L_o. It follows that the union

$$L = L_e \cup L_o$$

has the same minimum distance as L_e and L_o, and hence twice the density. This L is known as the Leech lattice.

2.3.1.3 Low-density integer lattices (2012)

Low-density integer lattices (LDA) are constructed from Construction A and non-binary LDPC codes [44]. The authors derived the factor graph for Construction A lattices which can be used by the iterative message-passing decoder. They gave an interesting example which uses a $(2, 5)$-regular LDPC code with $p = 11$ where the column degree is 2, the row degree is 5 and p is the alphabet size of the code. The construction is simple but the complexity of the decoder is $O(p^2 N \log N)$.

In 2013, the authors proved that two particular families of the low-density integer lattices can achieve the Poltyrev capacity under lattice decoding [51, 52]. The difference between them is the the number of non-zero coefficients h_i in a parity-check equation $\sum_{i=1}^{n} h_i x_i \equiv 0$ mod p. This number is called the degree of the parity-check equation (row degree). [51] shows how Poltyrev capacity can be achieved with LDA lattices the parity-check equations of which have degrees logarithmically growing in the dimension of the lattices. The proof is inspired by the proof given by Gallager which demonstrates that binary LDPC codes need logarithmically growing parity-check equation degrees to achieve the capacity of the binary symmetric channel. [52] gives a stronger statement, showing that Poltyrev capacity can be attained also by LDA lattices with constant parity-check equation degrees. The basic idea of the proofs is as follows: Define a decoding sphere centred at the channel output and containing (very probably) the channel input. If the sent lattice point is inside the sphere and is the only one, no error occurs. They managed to show that the probability that there is only one sent lattice point inside the sphere tends to 1 when n goes to infinity. The typical settings of the LDA lattices to guarantee achieving the Poltyrev capacity are $p = n^{\frac{1}{2}}$ and the column degree should at least be 7. The constant of the column degree can not get any closer to the constant 2 which appears to be an experimentally good choice for non-binary LDPC codes over binary-input channels. This was posed as an open question.

2.3.2 Multilevel constructions

From Remark 2.1 we know that the single level construction cannot generate high dimensional lattices with large coding gain if the component codes are binary codes. By using multilevel construction, we can take advantage of using capacity-achieving binary codes and generate high dimensional lattices with large coding gain. The following is the definition of Construction D.

Let ψ be the natural embedding of $\mathbb{F}_2^N$ into $\mathbb{Z}^N$, where $\mathbb{F}_2$ is the binary field. Let $C_0 \subseteq C_1 \subseteq \cdots \subseteq C_{a-1} \subseteq C_a = \mathbb{F}_2^N$ be a family of nested binary linear codes, where C_i has parameters (N, k_i, d_i) and C_a is the trivial $[N, N, 1]$ code. Let $k_i = \dim(C_i)$ and let $\bar{b}_1, \bar{b}_2, \cdots, \bar{b}_N$ be a basis of $\mathbb{F}_2^N$ such that $\bar{b}_1, \cdots \bar{b}_{k_i}$ span C_i. The lattice Λ_D consists of all vectors of the form [45]

$$\sum_{i=0}^{a-1} 2^i \sum_{j=1}^{k_i} \alpha_j^{(i)} \psi(\bar{b}_j) + 2^a \bar{l} \tag{2.2}$$

where $\alpha_j^{(i)} \in \{0, 1\}$ and $\bar{l} \in \mathbb{Z}^N$. The fundamental volume of a lattice of Construction D is given by

$$V(\Lambda_D) = 2^{-N \sum_{i=1}^{a-1} k_i} V(\Lambda_a)^N = (2^{a-1})^N \cdot 2^{N - \sum_{i=1}^{a-1} k_i},$$

where the lattice partition is $\Lambda_0/\Lambda_1/\cdots/\Lambda_a = \mathbb{Z}/2\mathbb{Z}/\cdots/2^a\mathbb{Z}$.

It is worth noting here that there is another version of Construction D, referred to Construction $\overline{D}$ by [53], which has been used in Barnes-Wall lattices [54]. Its code formula is defined as

$$\Gamma_{\overline{D}} = \psi(C_0) + 2\psi(C_1) + \cdots + 2^{a-1}\psi(C_{a-1}) + 2^a(\mathbb{Z})^N.$$

As claimed by [53], this Construction $\overline{D}$ may not necessarily generate a lattice. Only if the set of nested binary linear codes is closed under the Schur product, which means $\bar{c}_1, \bar{c}_2 \in C_i$ and $\bar{c}_1 * \bar{c}_2 \in C_{i+1}$ for $i = 0, \cdots, a-1$, $\Gamma_{\overline{D}}$ is a lattice and $\Gamma_{\overline{D}} = \Lambda_D$, where $*$ represents the coordinate-wise product between any two binary codewords.

Remark 2.2: The vectors from $\Gamma_{\overline{D}}$ can be written as

$$\sum_{i=0}^{a-1} 2^i \cdot \mathrm{mod}\left(\sum_{j=1}^{k_i} \alpha_j^{(i)} \psi(\bar{b}_j), 2\right) + 2^a \bar{l}, \tag{2.3}$$

where $\alpha_j^{(i)} \in \{0, 1\}$ and $\bar{l} \in \mathbb{Z}^N$. The difference between (2.2) and (2.3) is that this Construction $\overline{D}$ adds modulo operations which may break the linear property.

We use the example in [53] to demonstrate that $\Gamma_{\overline{D}}$ may not be a lattice.

Example 2.1: Consider nested binary linear codes $C_0 \subseteq C_1 \subseteq C_2 = \mathbb{F}_2^4$ and

$\bar{b}_1, \bar{b}_2, \bar{b}_3, \bar{b}_4$ is a basis for $\mathbb{Z}_2^4$:

$$\begin{aligned}\bar{b}_1 &= (1,1,0,0)\\ \bar{b}_2 &= (1,0,1,0)\\ \bar{b}_3 &= (1,0,0,1)\\ \bar{b}_4 &= (1,0,0,0)\end{aligned}$$

where $\bar{b}_1, \bar{b}_2$ span C_0, $\bar{b}_1, \bar{b}_2, \bar{b}_3$ span C_1.

Then,

$$\Lambda_D = (\alpha_1^{(0)}\bar{b}_1 + \alpha_2^{(0)}\bar{b}_2) + 2(\alpha_1^{(1)}\bar{b}_1 + \alpha_2^{(1)}\bar{b}_2 + \alpha_3^{(1)}\bar{b}_2) + 4\bar{l},$$

where $\alpha_j^{(i)} \in \{0,1\}$ and $\bar{l} \in \mathbb{Z}^N$.

Also,

$$\begin{aligned}\Gamma_{\overline{D}} &= \psi(C_0) + 2\psi(C_1) + 4\mathbb{Z}^4\\ &= \{\bar{c}_0 + 2\bar{c}_1 + 4\bar{l} \mid \bar{c}_0 \in \psi(C_0), \bar{c}_1 \in \psi(C_1), \bar{l} \in \mathbb{Z}^4\}.\end{aligned}$$

Since $(1,1,0,0),(1,0,1,0),(0,1,1,0) \in \Gamma_{\overline{D}}$, but $(2,0,0,0) = (1,1,0,0) + (1,0,1,0) - (0,1,1,0) \notin \Gamma_{\overline{D}}$. Therefore $\Gamma_{\overline{D}}$ is not a lattice. On the contrary, Λ_D is always a lattice.

Remark 2.3: It is a fact that the family of Reed-Muller codes is closed under the Schur product. Therefore Construction D and Construction $\overline{D}$ generate the same Barnes-Wall lattices.

For construction D with the lattice partition $\Lambda_1/\Lambda_2 \cdots = \mathbb{Z}/2\mathbb{Z}\cdots$. Let $X_{1:r} = X_1, X_2, \cdots, X_r$ and Y denote the input and output for AWGN channel where $X_i \in \mathcal{X} = \{0,1\}, Y \in \mathcal{Y}$. Therefore the channel of the ℓ-th level is a well-defined $2^{\ell-1}\mathbb{Z}/2^{\ell}\mathbb{Z}$ channel [39]. Given uniformly distributed $x_{1:\ell-1}$, let $\mathcal{A}_\ell(x_{1:\ell})$ denote the set of the chosen constellation, i.e., $\mathcal{A}_\ell(x_{1:\ell}) = x_1 + \cdots + 2^{\ell-1}x_\ell + 2^\ell\mathbb{Z}$, the conditional PDF of this channel with input $x_\ell \in \{0,1\}$ and output $\bar{y}_\ell = y \bmod 2^\ell\mathbb{Z}$ is [39]

$$\begin{aligned} P_{\bar{Y}_\ell|X_\ell,X_{1:\ell-1}}(\bar{y}_\ell|x_\ell,x_{1:\ell-1}) &= f_{\sigma,2^\ell\mathbb{Z}}(\bar{y}_\ell - x_1 - \cdots - 2^{\ell-1}x_\ell) \\ &= \sum_{\boldsymbol{\lambda}\in 2^\ell\mathbb{Z}} f_{\sigma,\boldsymbol{\lambda}}(\bar{y}_\ell - x_1 - \cdots - 2^{\ell-1}x_\ell) \\ &= \frac{1}{\sqrt{2\pi}\sigma}\sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} \exp\left(-\frac{1}{2\sigma^2}|\bar{y}_\ell - a|^2\right). \quad (2.4) \end{aligned}$$

This channel is the key to construct AWGN-good multilevel lattices. More details are introduced in Chapter 3 which follows the proof of [39].

2.3.2.1 Barnes-Wall lattices

Reed-Muller codes RM(N,k,d) are a class of linear block codes over GF(2), where n is the length of the codeword, k is the length of the information block and d is the minimum Hamming distance of this block code. Conventionally, Reed-Muller codes are denoted by RM(r,m) $(0 \le r \le m)$ with following relation between N, k and d:

$$N = 2^m, k = 1 + \begin{pmatrix} m \\ 1 \end{pmatrix} + \cdots + \begin{pmatrix} m \\ r \end{pmatrix}, d = 2^{m-r}.$$

Reed-Muller codes are famous for the recursively construction, which means larger Reed-Muller codes can be constructed from smaller ones.

Barnes-Wall lattices are an infinite family of lattices, which are the densest lattices known in 4, 8, and 16 dimensions[2]. Their constructions involve the family of Reed-Muller codes. We use the same notation as Reed-Muller codes. The m-th

[2] We give the example of Barnes-Wall lattices as a benchmark particularly because of the connection between Reed-Muller codes and polar codes [16]. The advantage of polar codes over Reed-Muller codes will translate into the advantage of polar lattices over Barnes-Wall lattices. This is shown in Figure 3.7

member of this family, denoted by $\Lambda(r = 0, m)$, is a $N = 2^m$ dimensional complex lattice or $2N$ dimensional real lattice. Following the code formulas given in [54], one may interpret the close relationship between Barnes-Wall lattices and Reed-Muller codes as follows:

For $m - r$ even:

$$\Lambda(r,m) = 2^{(m-r)/2}\mathbb{Z}^{2N} + \sum_{\substack{r+1 \leq r' \leq m \\ m-r' \text{ odd}}} \text{RM}(r', m+1)2^{(r'-r-1)/2}.$$

For $m - r$ odd:

$$\Lambda(r,m) = 2^{(m-r+1)/2}\mathbb{Z}^{2N} + \sum_{\substack{r+1 \leq r' \leq m \\ m-r' \text{ even}}} \text{RM}(r', m+1)2^{(r'-r-1)/2}.$$

Equivalently, one may use the complex code formula:

$$\Lambda(r,m) = \phi^{(m-r)/2}\mathbb{G}^N + \sum_{r \leq r' < n} \text{RM}(r', m)\phi^{r'-r},$$

where $\phi = 1 + i$ and $\mathbb{G}$ is the lattice of Gaussian integers.

For example, the code formula of the 1024-dimensional Barnes-Wall lattice is:

$$BW_{1024} = \text{RM}(1, 10) + 2\text{RM}(3, 10) + \cdots + 2^5\mathbb{Z}^{1024}. \tag{2.5}$$

The code formulas show a construction of Barnes-Wall lattices from Reed-Muller codes. The normalized fundamental volume of Barnes-Wall lattices is $2^{\frac{m}{2}}$. The minimum squared Euclidean distance of Barnes-Wall lattices is 2^m, leading to an asymptotic coding gain of $2^{\frac{m}{2}}$.

It is worth mentioning that Barnes-Wall lattices can be decoded with the bounded distance decoder efficiently [55].

2.3.2.2 Low-density parity-check lattices (2006)

Low-density parity-check lattices are constructed by applying Construction D' [40, p232] to a set of parity checks defining a family of nested LDPC codes [47]. In other words, this is also a multilevel construction with regular LDPC codes. This construction provides a Tanner graph representation of lattices, which in turn is used to efficiently decode the lattice by the generalized min-sum algorithm. One can also use multistage decoding to decode each level's LDPC codes. This reduces the complexity at the expense of some possible degradation in performance. The authors used the progressive-edge-growth algorithm (PEG) to find good component LDPC codes. Furthermore, the irregular LDPC lattices were proposed in [56].

The following is the definition of Construction D'. Examples and decoding algorithms can be found in [47]..

Let $C_0 \supseteq C_1 \supseteq \cdots \supseteq C_{a-1} \supseteq C_a$ be a family of nested binary linear codes, where C_i has parameters (N, k_i, d_i). Let $\{\overline{h}_1, \cdots, \overline{h}_N\}$ be linearly independent vectors in $\mathbb{F}_2^N$ such that each dual code C_i^* is generated by $\{\overline{h}_1, \cdots, \overline{h}_{r_i}\}$, where $r_i = N - k_i$. Let $\Lambda_{D'}$ be the corresponding lattice given by Construction D'. Its parity-check matrix is

$$H = \begin{pmatrix} \overline{h}_1 \\ \vdots \\ \overline{h}_{r_0} \\ \vdots \\ 2^a \overline{h}_{r_{a-1}+1} \\ \vdots \\ 2^a \overline{h}_{r_a} \end{pmatrix}.$$

Then $\overline{x} \in \mathbb{Z}^N$ is in $\Lambda_{D'}$ if and only if

$$H\overline{x}^T \equiv 0 \bmod 2^{a+1}.$$

2.3.2.3 Turbo lattices (2010)

Turbo lattices are constructed by applying Construction D to a set of nested turbo codes [46]. Each turbo code is generated by the tail-biting convolutional codes. Let $C_0 \subseteq C_1 \subseteq \cdots \subseteq C_{a-1} \subseteq C_a = \mathbb{F}_2^N$ be a family of nested binary linear codes, where C_i has parameters (N, k_i, d_i). The code formula for the turbo lattice is

$$\Lambda_{\text{TC}} = C_1 + 2C_2 + \cdots + 2^{a-1}C_{a-1} + 2^a\mathbb{Z}^N.$$

Consider two nested turbo codes $C_0 \subseteq C_1$ generated by a generator matrix G of a convolutional code with the size $K \times N$ and a random interleaver Π with size $k = LK$. Each interleaver can be represented by a matrix $P_{k\times k}$ which has only one 1 in each column and row. Therefore the generator matrix of C_2 is

$$G_{\text{TC}} = \begin{pmatrix} I_k & F & PF \end{pmatrix},$$

where F is a $LK \times L(N-K)$ submatrix including only parity columns of the tail-biting generator matrix G. Then the generator matrix of C_1 consists of the first k_1 rows of G_{TC}. Therefore the generator matrix of the turbo lattice is

$$G_{\text{TL}} = \begin{pmatrix} I_{k_1} & 0 & F_1 & P_1F_1 \\ 0 & 2I_{k_2} & 2F_2 & 2P_2F_2 \\ 0 & 0 & 4I_{k_3} & 0 \\ 0 & 0 & 0 & 4I_{k_3} \end{pmatrix},$$

For example, the generator matrix of the component encoders for C_2 is

$$\begin{pmatrix} 1 & 0 & \frac{1+x+x^2+x^3}{1+x^2+x^3} \\ 0 & 1 & \frac{1+x+x^2}{1+x^2+x^3} \end{pmatrix}.$$

The resulting turbo code has rate $R_2 = \frac{1}{2}$ and $d_2 = 13$ for block information bits of length 400. Then, only use the first row to be the generator matrix of the component encoders for C_1 which is

$$\begin{pmatrix} 1 & 0 & \frac{1+x+x^2+x^3}{1+x^2+x^3} \end{pmatrix}.$$

The block turbo code C_1 has rate $R_1 = \frac{1}{3}$ and $d_1 = 28$ for information block length of 576.

Suppose a block of information bits of size 1000 is used. Since C_2 is a rate $\frac{1}{2}$ block turbo code, the dimension of the turbo lattice is 2000. The generator matrix is

$$G_{\text{TL}} = \begin{pmatrix} I_{576} & 0 & F_1 & P_1F_1 \\ 0 & 2I_{324} & 2F_2 & 2P_2F_2 \\ 0 & 0 & 4I_{500} & 0 \\ 0 & 0 & 0 & 4I_{500} \end{pmatrix}.$$

The minimum distance of Λ_{TL} is $\min\{4, d_1, d_2\} = 4$.

2.3.2.4 Polar lattices (2013)

Polar lattices are generated by applying Construction D and construction E to a set of nested polar codes. Since polar lattices are as explicit as polar codes, their construction is equally efficient. Furthermore, compared with the above existing schemes [47, 44, 57, 46], polar lattices are distinguished by their provable AWGN-goodness and low complexity, namely, they asymptotically achieve the sphere bound with multistage decoding. More details about polar lattices are in Chapter 3.

2.3.3 Other lattices

Besides constructing from error-correcting codes, the following two lattices are designed directly from the Euclidean space. These constructions may be alternatives to the well known techniques (constructions A-D) that generate lattices from finite alphabet linear codes.

2.3.3.1 Signal Codes: Convolutional Lattice Codes (2008)

Signal codes (or convolutional lattice codes) [58] are lattice codes designed directly in the Euclidean space without any finite alphabet codes. As the name may suggest, a convolutional lattice codeword (or lattice point) is generated by convolving an integer sequence, representing the information sequence, with a fixed, continuous-valued, finite impulse response (FIR) filter pattern. The FIR length is small yet, as shown in the paper, by proper choice of short FIR filters it generates a lattice with surprisingly large minimal distance. It is due to the signal processing interpretation of the code construction. For practical usage the filter output has an increased power and this shaping gain loss will cancel the coding gain, therefore the code construction includes a shaping mechanism inspired by precoding techniques such as the Tomlinson-Harashima filter. Convolutional lattice codes can be decoded efficiently by sequential decoding or for better performance by bi-directional sequential decoding. Error analysis and simulation results indicate that with a very large stack length of 10^6, and for frame error rate of 10^{-3}, the stack decoder can work as close as 2.9 dB from channel capacity, where the bidirectional stack decoder can work as close as 2.3 dB from channel capacity.

In this section, we briefly review its construction. Convolutional lattice codes are defined as lattice codes which are based on an n-dimensional lattice whose $(n +$

$P) \times n$ generator matrix G has the following Toeplitz form

$$\begin{pmatrix} 1 & 0 & \vdots & \vdots & \vdots \\ g_1 & 1 & \ddots & \vdots & \vdots \\ \vdots & g_1 & \ddots & 0 & \vdots \\ g_P & \vdots & \ddots & 1 & 0 \\ 0 & g_P & \ddots & g_1 & 1 \\ \vdots & 0 & \ddots & \vdots & g_1 \\ \vdots & \vdots & \ddots & g_P & \vdots \\ \vdots & \vdots & \vdots & 0 & g_P \end{pmatrix}$$

where $1, g_1, g_2, \cdots, g_P$ are the impulse response coefficients of a monic and causal FIR filter.

From this generator matrix, it is easy to prove if $n \gg P$, the fundamental volume of this lattice $V(\Lambda)^{2/n} = [\text{det}(G'G)]^{1/n} \rightarrow 1$. Therefore, for large n, the volume will almost not increase. It has been proved that the minimum distance of the proposed lattices is equal or greater than 1. Let n_0 be the smallest index for the non-zero component of $\overline{b}_{\text{min}}$, where $\overline{x}_{\text{min}} = G\overline{b}_{\text{min}}$ and $\overline{x}_{\text{min}}$ is the shortest non-zero lattice point. Since the filter is monic and causal, $\overline{x}_{\text{min}}(n_0) = \overline{b}_{\text{min}}(n_0)$, and thus

$$d^2_{\text{min}}(\Lambda) \geq \mid \overline{x}_{\text{min}}(n_0) \mid^2 = \mid \overline{b}_{\text{min}}(n_0) \mid^2 \geq 1.$$

In order to obtain high $d^2_{\text{min}}(\Lambda)$, a numerical algorithm is proposed to choose the parameters of the FIR filter. Methods that were developed for finding the minimum distance between output sequences of intersymbol interference (ISI) channels can be applied here to find the minimum distance given a FIR filter. The resulting search algorithm is to compare all the lattice points within a given hypersphere, by developing a tree of all possible integer sequences, and truncating tree branches as soon as it can identify that all the corresponding lattice points will lie outside the hypersphere. In fact, this search algorithm is equivalent to a sphere decoder. It can be seen that the squared minimum distance improves as the spectral null of the filter becomes deeper by increasing the number of zeros P. However, the range of the shaped integers becomes larger, which increases the decoding and shaping implementation complexity.

The coordinates of a lattice point $\overline{x} = G\overline{b}$, where $\overline{b}$ is an n-dimensional vector of integers, are the convolution of the sequence of $\overline{b}$ with the filter

$$x_k = b_k + \sum_{p=1}^{P} g_p b_{k=p}$$

for $k = 1, \cdots, n + P$.

Shaping is essential for convolutional lattice codes, otherwise the power increase due to filtering operation is higher than the increase in minimal distance. The shaping operation has a close relation to the precoding operation for ISI channels. The purpose of precoding is pre-equalizing the distortion of a linear channel, which is known at the transmitter, in order to avoid the need for equalization at the receiver. The most simple way to do this is to filter the data with the inverse channel filter, but the inverse filtering operation can significantly increase the signal's power. The solution is to map the sequence of information symbols to another sequence such that the constraints at the channel input can be meet after this procoding operation. Therefore three shaping methods (Tomlinson-Harashima shaping, systematic shaping and nested lattice shaping) are introduced in the paper, where the first two are indeed based on well-known procoding schemes for ISI channels. The encoding diagram is shown in Figure 2.3.

The decoder consists of two blocks. First to do an inverse shaping operation. Then do the detection for b'. Unfortunately, Viterbi decoding [59] cannot be used.

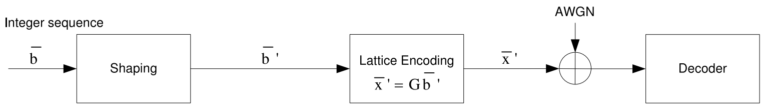


Figure 2.3: Convolutional lattice codes [58]. Instead of mapping the information $\bar{b}$ to the lattice point $\bar{x} = G\bar{b}$ directly, it should be mapped to some other lattice point $\bar{x}' = G\bar{b}'$ that belongs to the shaping region.

This is because the shaping operation increases substantially the range of possible integer values for any filter tap, and hence the number of states in the Viterbi decoder. The authors proposed to use the sequential decoder. The computational complexity of sequential decoding of any tree codes increases abruptly below some cutoff SNR. The simulations show that the sequential decoder works well even close to the cutoff rate. Furthermore, they also use bidirectional sequential decoder with large complexity and large computational resources to demonstrate that the low error rate can be achieved beyond the cutoff rate. Only an approximated upper bound of the union bound is given. Further check is needed to verify the actual code performance.

2.3.3.2 Low-density lattice codes (2008)

A low-density lattice code [57] is a dimension n lattice with a non-singular generator matrix G, for which $H = G^{-1}$ is sparse with constant row and column weight d. For a given $V = \mid \det G \mid$ and a a sorted sequence of these d values $w_1 \geq w_2 \geq \cdots \geq w_d > 0$,, the inverse generator H is designed as follows. Let

$$\begin{pmatrix} w_1 & w_2 & \cdots & w_d & 0 & \cdots & 0 \end{pmatrix}$$

be a row vector with d positive values, followed by $n - d$ zeros. The matrix H can be written as permutations π_i of h, followed by a random sign change S_i, followed by scaling by $k > 0$:

$$H = k \begin{pmatrix} S_1 \cdot \pi_1(h) \\ S_2 \cdot \pi_2(h) \\ \vdots \\ S_n \cdot \pi_n(h) \end{pmatrix}$$

such that the permutations result in H having exactly exactly d non-zero values in each column. The sign-change matrix S_i is a square, diagonal matrix, where the diagonal entries are $+1$ or -1 with probability $\frac{1}{2}$. Then k is selected to normalize the determinant to the given V. For example:

$$H = \begin{pmatrix} 0 & -0.8 & 0 & -0.5 & 1 & 0 \\ 0.8 & 0 & 0 & 1 & 0 & -0.5 \\ 0 & 0.5 & 1 & 0 & 0.8 & 0 \\ 0 & 0 & -0.5 & -0.8 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0.5 & 0.8 \\ 0.5 & -1 & -0.8 & 0 & 0 & 0 \end{pmatrix}$$

is the parity-check matrix for $(n = 6, d = 3)$ LDLC lattices. The factor graph which can be used for belief-propagation decoding is shown in Figure 2.4. This is a special case of the standard LDLC constructions, which are characterized by a parameter $\alpha \geq 0$. Belief-propagation decoding of LDLC lattices will converge exponentially fast if and only if $\alpha \leq 1$ [57, Theorem 1], where $\alpha = \frac{\sum_{i=2}^{d} w_i^2}{h_1^2}$. Therefore for our example, $\alpha = 0.8^2 + 0.5^2 = 0.89$.

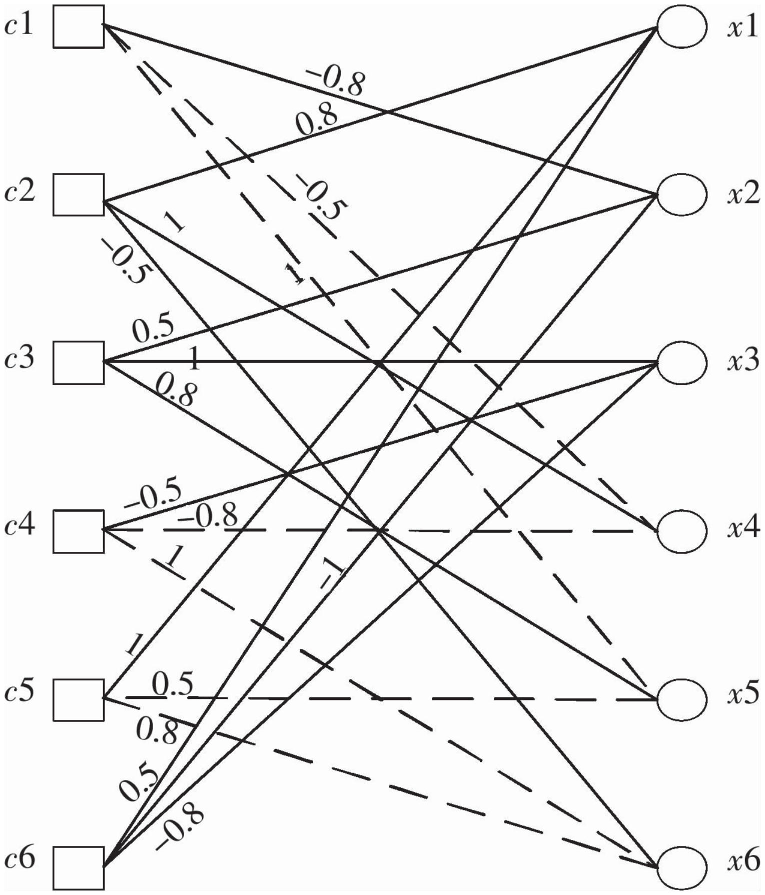


Figure 2.4: Factor graph of an LDLC [57].

In conclusion, a codeword x of LDLC is also generated directly at the Euclidean space as a linear transformation of a corresponding integer message vector b, i.e., $x = Gb$, where $H = G^{-1}$ is restricted to be sparse. The non-zero elements of H are real numbers. An iterative sum-product algorithm is used to decode lattice codewords. The variable node messages are Gaussian mixtures. The convergence analysis imply some necessary conditions on H.

2.3.4 Comparison

The performance comparison of recently introduced lattices approaching the Poltyrev capacity with dimension around 1000 is shown in Figure 2.5 in terms of SER[3].

[3] The reason to shift from block error probability to SER is that most of the lattices are using SER. Since one lattice point consists n symbols, it is a fair comparison as long as the dimensions n are the same. The curve for the LDPC lattice was plotted with the normalized block error probability (NEP).

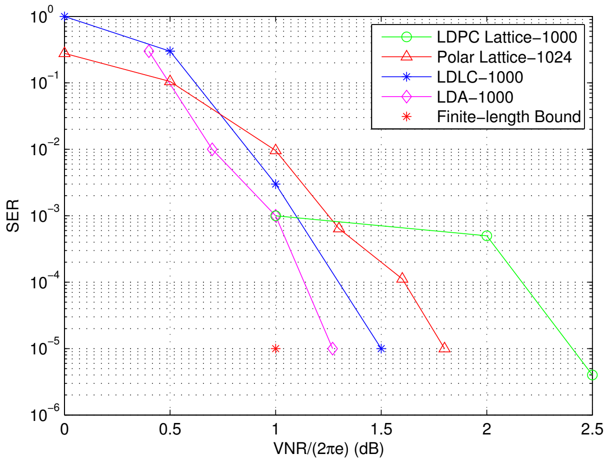


Figure 2.5: Transmitting lattice points over the AWGN channel without power constraint. The decoding method is the lattice decoding. We investigate the SER of lattices with dimension around 1000. All the results are obtained from their own papers.

The polar lattice used here is constructed from the two-dimensional lattice partition ($N = 512, n = 2$). The simulation curves of other lattices are obtained from their corresponding papers. We note that the theoretic minimum gap to the Poltyrev capacity is about 1 dB for dimension 1000 [49]. Among the four types of lattices compared, the LDPC lattice [47] has the weakest performance, while all other three have similar performance at this dimension (the difference is within 0.5 dB). In contrast to the polar lattice and LDA lattice [44, 51], analytic results of the LDLC [57] are not available; therefore, they are less understood in theory. The performance of the LDA lattice is analyzed on the basis of a random ensemble of nonbinary LDPC codes with a somewhat nonstandard assumption [51]; consequently, it is unclear whether standard LDPC codes satisfy the assumption. The LDA lattice has slightly better performance than the polar lattice at the expense of higher decoding complexity ($O(p^2 N \log N)$). Polar lattices have an excellent compromise between decoding complexity ($O(rN \log N)$) and error performance. It is worth pointing out there is potential to improve the multi-stage decoding of polar lattices. For example, a soft-output multi-stage decoding algorithm will outperform the current hard-output multi-stage decoding. We add this as our future work.

2.4 Poltyrev capacity

Poltyrev [5] has given an existence proof of capacity-achieving mod-p lattices, with exponential error bounds that are tight near capacity. Loeliger [14] reproved the result using standard averaging arguments for linear codes over the p-ary field GF(p) lifted to mod-p lattices. All the results above are based on the Minkowski-Hlawka theorem of lattice theory.

2.4.1 Poltyrev's result

There are several ways to define coding rate and capacity per unit volume of a general infinite constellation, not necessarily a lattice. One simple way is to count the number of codewords per unit volume within a “large" cube and translate it into bits.

Definition 2.2 (Normalized logarithmic density): The normalized logarithmic density (NLD) is defined as

$$\delta = \frac{1}{n} \limsup_{a \to \infty} \log \left(\frac{|\mathbb{C}_a|}{a^n} \right),$$

where

$$\mathbb{C}_a = \Lambda \cap \text{CUBE}(a)$$

is the intersection of the IC with the n-dimensional cube $\text{CUBE}(a) = [-a/2, a/2)^n$.

Definition 2.3 (Poltyrev capacity): The largest normalized logarithmic density that allows reliable communication (i.e., a vanishing error probability) over a large block of channel uses of AWGN channel is

$$\delta^* = \frac{1}{2}\log\frac{1}{2\pi e\sigma^2}.$$

Suppose now that the IC is a lattice Λ. Since the diameter d of the Voronoi cell is finite, the number of codewords inside the cube is bounded between $(a-d)^n/V(\Lambda)$ and $(a+d)^n/V(\Lambda)$. Therefore, the total number of codewords approximates to $a^n/V(\Lambda)$ for large a. The normalized logarithmic density of a lattice is

$$\delta = \frac{1}{n}\log\frac{1}{V(\Lambda)}.$$

The error exponent $E(\delta)$ for the unconstrained AWGN is defined as $P_e = e^{-n(E(\delta)+o(1))}$. The following is the restatement of the main results in [5]. The lower bound on the error exponent is the random coding exponent $E_r(\delta)$, given by

$$E_r(\delta) = \begin{cases} \frac{1}{2}\log\frac{1}{8\pi\sigma^2} - \delta, & \delta \leq \delta_{cr}; \\ \frac{e^{-2\delta}}{4\pi e\sigma^2} + \delta + \frac{1}{2}\log 2\pi\sigma^2, & \delta_{cr} \leq \delta \leq \delta^*; \\ 0, & \delta \geq \delta^*, \end{cases}$$

where $\delta_{cr} = \frac{1}{2}\log\frac{1}{4\pi e\sigma^2}$. Poltyrev also provided an expurgation-type argument to improve the error exponent at low NLD values (below $\delta_{ex} = \frac{1}{2}\log\frac{1}{8\pi e\sigma^2}$). We refer the reader to [5] for details.

An upper bound on the error exponent is the sphere packing exponent. It is given by

$$E_{sp}(\delta) = \frac{1}{4\pi e^{2\delta+1}} + \delta + \frac{1}{2}\log 2\pi\sigma^2,$$

which is derived from the sphere bound. The decoding region for this sphere bound is the equivalent sphere with effective radius $r_{\Lambda}^{\text{effec}}$, which leads to the lower bound on the error probability.

Therefore the upper and lower bounds on the error exponent give the value of $P_e(n, \delta)$:

$$e^{-n(E_{sp}(\delta)+o(1))} \leq P_e(n, \delta) \leq e^{-n(E_r(\delta)+o(1))},$$

where $P_e(n, \delta) \leq e^{n\delta} n V_n \int_0^{2r} w^{n-1} \text{Pr}\{X^n \in D(r, w)\} dw + \text{Pr}\{\|X^n\| > r\}$. $D(r, w)$ denotes the section of the sphere with radius r that is cut off by a hyperplane at a distance $\frac{w}{2}$ from the origin. The first part is the weak noise and the second part is the strong noise. In order to derive the error exponent, Poltyrev used the asymptotic value of $r = \sqrt{n}\delta e^{\delta^* - \delta}$. However this is not the optimal value which minimizes the upper bound.

It is also shown in [5] that $P_e(n, \delta) \geq 0.5$ for $n = 1, 2, \cdots$ with $\delta > \delta^*$.

2.4.2 Loeliger's result

Loeliger applied Minkowski-Hlawka theorem to scaled mod-p lattices L_C where C is the component code which is used to construct the mod-p lattice by Construction A. The proof of Minkowski-Hlawka theorem for the lattice version is shown in Appendix A. The mod-p lattices L_C can be replaced by any other set of lattices for which the Minkowski-Hlawka theorem can be proved.

Theorem 2.1 (Loeliger's coding theorem [14]): Let E be a Jordan measurable bounded subset of $\mathbb{R}^n$, for any $\alpha > 0$, for all sufficiently small ϵ, and all sufficiently large primes p, the arithmetic average of the ambiguity probability over all lattices ϵL_C is upper bounded by [14]

$$\overline{P_{\text{amb}|E}} \lesssim (1 + \alpha) V(E)/V,$$

where $V = \epsilon^n p^{n-k}$ is the fundamental volume of the scaled mod-p lattices ϵL_C, $C \in \mathcal{C}$. $\mathcal{C}$ is any balanced set of linear (n, k) codes over $\mathbb{Z}_p$.

Proof. Let $f_{\mathbf{e}|E}$ be the probability density function of **e** conditioning on the event $\mathbf{e} \in E$. For any $e \in E$, the event $\mathbf{e} = e$ is an ambiguity if and only if $(L \setminus \{0\}) \cap (e - E) \neq 0$. By applying the Minkowski-Hlawka Theorem, we have

$$\begin{aligned}
\overline{P_{\text{amb}|E}} &= \frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} P_{\text{amb}|E} \\
&\leq \int_E f_{\mathbf{e}|E}(v) \frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} \mid (\epsilon L_C \setminus \{0\}) \cap (v - E) \mid dv \\
&\approx \int_E f_{\mathbf{e}|E}(v) V(E)/V dv \\
&= V(E)/V,
\end{aligned}$$

and the approximation becomes exact in the limit $\epsilon \to 0$, $p \to \infty$. □

The above bound can be rewritten as

$$\overline{P_{\text{amb}|E}} \lesssim (1 + \alpha) 2^{n(\delta - h(E))},$$

where $\delta = \frac{1}{n} \log \frac{1}{V(\epsilon L_C)}$ is NLD and $h(E) = \frac{1}{2} \log V(E)$ is the geometric entropy rate of E. For $n \to \infty$,$h(E)$ converges to the information-theoretic differential entropy rate $h(e) = \frac{1}{2} \log 2\pi e \sigma^2$ for Gaussian noise. It is obvious that, for reliable transmission, the fundamental volume $V(\epsilon L)$ cannot be smaller than $V(E)$. In other words, reliable transmission is not possible if $\delta > \delta^*$.

If we consider the shaping region $S \subset \mathbb{R}^n$, we would expect to obtain a code with about $M = V(S)/V$ codewords. The bound of the ambiguity probability is

$$\overline{P_{\text{amb}|E}} \lesssim (1 + \alpha) 2^{-n[h(S) - h(E) - R]},$$

where $R \triangleq 1/n\log_2 M$ is the information rate of the code in bits per dimension, and where $h(S) \triangleq 1/n\log_2 V(S)$. Since $h(E) = h(e)$ as $n \to \infty$ and $\lim_{n\to\infty} h(S) = 1/2\log_2(2\pi eP)$ where P is the signal power per dimension (S becomes an n-dimensional sphere). Therefore, it is safe to say that arbitrarily small (but positive) error probability is achievable with lattice codes and lattice decoding at any rate below $1/2\log_2(P/N)$.

2.5 Summary

The concept of lattice codes for the AWGN channel is introduced in Chapter 1. Some basics about lattices, including discrete Gaussian distribution, AWGN-goodness of lattices, lattice constructions from error-correcting codes and theoretical analysis based on radome lattices are introduced in this Chapter. No mathematical novelty but all the background which is needed to understand the sequel is presented in this chapter. We will show how to construct AWGN-good lattices from polar codes explicitly and prove their AWGN-goodness in the next chapter Chapter 3.

CHAPTER 3

Polar lattices are AWGN-good

In this chapter, we introduce the detailed constructions of polar lattice and give the proof that they are AWGN-good. We also provide some new insights about Forney's construction.

3.1 Forney *et al.*'s Construction Revisited

Forney *et al.* gave constructions of AWGN-good lattices for noise variance σ^2 in [39]. We now revisit their constructions by applying the properties of the flatness factor and provide new guidelines to the constructions. These new guidelines have a close connection to the later analysis of achieving the capacity in Chapter 4.

3.1.1 Component Lattices

Recall that a mod-Λ channel is a Gaussian channel with a modulo-Λ operator in the receiver front end [60, 39]. The capacity of the mod-Λ channel is [39]

$$C(\Lambda, \sigma^2) = \log V(\Lambda) - h(\Lambda, \sigma^2), \tag{3.1}$$

where $h(\Lambda, \sigma^2)$ is the differential entropy of the Λ-aliased noise over $\mathcal{V}(\Lambda)$:

$$h(\Lambda, \sigma^2) \quad = \quad -\int_{\mathcal{V}(\Lambda)} f_{\sigma,\Lambda}(\mathbf{x}) \log f_{\sigma,\Lambda}(\mathbf{x}) d\mathbf{x}.$$

The uniform distribution over $\mathcal{V}(\Lambda)$ maximizes the differential entropy which is $\log V(\Lambda)$. Therefore $C(\Lambda, \sigma^2) \geq 0$ where the equality holds for the uniform distribution.

Consider the lattice partition $\Lambda_1/\cdots/\Lambda_r$ where Λ_1 is the top lattice and Λ_r is the bottom lattice, both of dimension n and let r denote the number of levels if each level is a binary partition. It is worth pointing out that Λ_1 and Λ_r can be scaled versions of simple low-dimensional lattices such as $\mathbb{Z}$ and $\mathbb{Z}^2$. It is known that the Λ_1/Λ_r channel (i.e., the mod-Λ_r channel whose input is restricted to $|\Lambda_1/\Lambda_r|$ discrete lattice points in Λ_1) is regular, and the optimum input distribution is uniform [39].

Recall that the capacity of the Λ_1/Λ_r channel for Gaussian noise of variance σ^2 is given by [39]

$$\begin{aligned} C(\Lambda_1/\Lambda_r, \sigma^2) &= C(\Lambda_r, \sigma^2) - C(\Lambda_1, \sigma^2) \\ &= h(\Lambda_1, \sigma^2) - h(\Lambda_r, \sigma^2) + \log V(\Lambda_r)/V(\Lambda_1). \end{aligned} \tag{3.2}$$

Forney *et al.* chose Λ_1 and Λ_r as follows [39]:

- $V(\Lambda_1)$ is small enough that the mod-Λ_1 noise is almost uniform and

$$C(\Lambda_1, \sigma^2) \approx 0; \tag{3.3}$$

- $V(\Lambda_r)$ is large enough that the error probability $P_e(\Lambda_r, \sigma^2) \approx 0$. This means that the mod-Λ_r noise is almost unaliased and

$$C(\Lambda_r, \sigma^2) \approx \frac{n}{2} \log \left(\frac{V(\Lambda_r)^{2/n}}{2\pi e \sigma^2} \right). \tag{3.4}$$

We now give a new upper bound on the first condition and make the second condition more precise.

First, we bound the condition (3.3) by the flatness factor. The proof is in Appendix B.

Lemma 3.1: The capacity of the mod-Λ_1 channel is bounded by

$$C(\Lambda_1, \sigma^2) \leq \log\left(1 + \epsilon_{\Lambda_1}(\sigma)\right) \leq \log(e) \cdot \epsilon_{\Lambda_1}(\sigma). \tag{3.5}$$

The second condition (3.4) means that r is large. We now link r with N in a quantitative manner.

Lemma 3.2: Let the dimension n of Λ_r be fixed. It is required that $r = nO(\log N)$ so that $P_e(\Lambda_r, \sigma^2)$ vanishes sub-exponentially with N.

Proof. For this purpose, we assume $\Lambda_1 = a\mathbb{Z}^n$ where a is determined by the flatness factor condition and $\Lambda_r = b\mathbb{Z}^n$ where b is to be estimated. a and b are both real numbers. We note that for all partition chains in [39], this is always possible: if the bottom lattice does not take the form of $b\mathbb{Z}^n$, one may simply further extend the partition chain (which will lead to an upper bound on r).

By the union bound, the error probability of Λ_r is upper-bounded by

$$P_e(\Lambda_r, \sigma^2) \leq nQ\left(\frac{b}{2\sigma}\right) \leq ne^{-\frac{b^2}{8\sigma^2}}$$

where we apply the Chernoff bound on the Q-function. We want

$$P_e(\Lambda_r, \sigma^2) = e^{-O(N)},$$

which leads to $b = O(\sqrt{N})$. Without loss of generality, consider the binary lattice partition. In this case we have $(b/a)^n = 2^{r-1}$. Thus,

$$\begin{aligned} r &= n \log\left(\frac{b}{a}\right) + 1 \\ &= nO(\log N) \end{aligned}$$

if we fix n. Obviously, this estimate still holds for non-binary lattice partitions. □

Note that the error probability of polar codes (hence that of polar lattices) decreases as $e^{-O(\sqrt{N})}$ rather than $e^{-O(N)}$. Lemma 3.2 shows that the number of levels r needs to grow with n and $\log N$. In practical designs, it is best to estimate the hidden constant numerically.

3.1.2 Gap to Poltyrev Capacity

Let $\mathcal{C}$ denote the “composite" code corresponding to the weighted sum of $r-1$ binary codes in (2.1). The total decoding error probability for L is bounded by

$$P_e(L, \sigma^2) \leq P_e(\mathcal{C}, \sigma^2) + P_e(\Lambda_r^N, \sigma^2). \tag{3.6}$$

To make $P_e(L, \sigma^2) \to 0$, we need to choose the lattice Λ_r such that $P_e(\Lambda_r^N, \sigma^2) \to 0$ and the code $\mathcal{C}$ for the Λ_1/Λ_r channel whose error probability also tends to zero.

Since $V(L) = 2^{-NR_\mathcal{C}} V(\Lambda_r)^N$, the logarithmic VNR of L is

$$\begin{aligned} \log\left(\frac{\gamma_L(\sigma)}{2\pi e}\right) &= \log \frac{V(L)^{\frac{2}{nN}}}{2\pi e \sigma^2} \\ &= \log \frac{2^{-\frac{2}{n}R_\mathcal{C}} V(\Lambda_r)^{\frac{2}{n}}}{2\pi e \sigma^2} \\ &= -\frac{2}{n} R_\mathcal{C} + \frac{2}{n} \log V(\Lambda_r) - \log 2\pi e \sigma^2. \end{aligned} \tag{3.7}$$

Define

$$\begin{cases} \epsilon_1 = C(\Lambda_1, \sigma^2) \\ \epsilon_2 = h(\sigma^2) - h(\Lambda_r, \sigma^2) \\ \epsilon_3 = C(\Lambda_1/\Lambda_r, \sigma^2) - R_{\mathcal{C}}, \end{cases} \tag{3.8}$$

where $h(\sigma^2) = \frac{n}{2}\log 2\pi e\sigma^2$ is the differential entropy of the Gaussian noise. We note that, $\epsilon_1 \geq 0$ represents the capacity of the mod-Λ_1 channel, $\epsilon_2 \geq 0$ (due to the data processing theorem) is the difference between the entropy of the Gaussian noise and that of the mod-Λ_r Gaussian noise, and $\epsilon_3 \geq 0$ is the capacity loss of the composite code $\mathcal{C}$.

Then we have

$$\begin{aligned} \log\left(\frac{\gamma_L(\sigma)}{2\pi e}\right) &= \frac{2}{n}\left(\log V(\Lambda_r) - C(\Lambda_1/\Lambda_r, \sigma^2) + \epsilon_3 - \frac{n}{2}\log 2\pi e\sigma^2\right) \\ &= \frac{2}{n}\left(\log V(\Lambda_r) - \log V(\Lambda_r) + h(\Lambda_r, \sigma^2) + \epsilon_1 + \epsilon_3 - \frac{n}{2}\log 2\pi e\sigma^2\right) \\ &= \frac{2}{n}(\epsilon_1 - \epsilon_2 + \epsilon_3). \end{aligned}$$

Since $\epsilon_2 \geq 0$, we obtain the upper bound[1]

$$\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\right) \leq \frac{2}{n}(\epsilon_1 + \epsilon_3). \tag{3.9}$$

Since $\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\right) = 0$ represents the Poltyrev capacity, the right hand side of (3.9) gives an upper bound on the gap to the Poltyrev capacity. The bound is equal to $\frac{6.02}{n}(\epsilon_1 + \epsilon_3)$ decibels (dB), by conversion of the binary logarithm into the base-10 logarithm.

[1] It was shown in [39] that $\epsilon_2 \approx \pi P_e(\Lambda_r, \sigma^2)$, which is negligible compared to the other two terms.

Table 3.1: Comparison of the choices of component lattices Λ_1 and Λ_r

	One-dimensional partition chain	Multi-dimensional partition chain
Top lattice Λ_1	$\Lambda_1 = a\mathbb{Z}$, a small	$\epsilon_{\Lambda_1}(\sigma) \approx 0$
Bottom lattice Λ_r	$\Lambda_r = ap\mathbb{Z}$, ap large	dense

3.1.2.1 Design Guidelines

To approach the Poltyrev capacity, we would like to have $\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\right) \to 0$ while $P_e(L, \sigma^2) \to 0$. Thus, from (3.9), we need both $\frac{\epsilon_1}{n} \to 0$ and $\frac{\epsilon_3}{n} \to 0$. Now we have the design criteria:

- The top lattice Λ_1 has a small normalized flatness factor $\frac{1}{n}\epsilon_{\Lambda_1}(\sigma)$, namely, it is almost impossible to decode.

- The bottom lattice Λ_r has a small error probability $P_e(\Lambda_r, \sigma^2)$, i.e., it can be decoded almost perfectly, which in the low-dimensional case essentially means Λ_2 is a dense lattice.

- $\mathcal{C}$ is a capacity-approaching code for the Λ_1/Λ_r channel.

In the following, we provide some guidelines to select the component lattices Λ_1 and Λ_r. In Table 3.1, we compare constructions of L in [39] with one and multi-dimensional lattice partition chains, using the new insights obtained in this paper.

The first method is based on the one-dimensional partition chain, corresponding to the choice $\Lambda_1 = a\mathbb{Z}$ and $\Lambda_r = ap\mathbb{Z}$ for some scaling factor a (p is a prime). Obviously, a has to be small while ap has to be large[2] so that $\epsilon_{\Lambda_1}(\sigma)$ is small and $P_e(\Lambda_r, \sigma^2)$ is small.

The second method is to use a multi-dimensional partition chain ($n > 1$). This method has the following advantages: (a) a smaller value of $\frac{1}{n}\epsilon_{\Lambda_1}(\sigma)$, thereby a smaller gap to the Poltyrev capacity; (b) a lower error probability $P_e(\Lambda_r, \sigma^2)$ if a dense lattice Λ_r is used. Figure 3.1 compares the normalized flatness factor $\frac{1}{n}\epsilon_{\Lambda_1}(\sigma)$

[2] p does not have to be very large for the target error probability in practice, e.g., 10^{-5} [39].

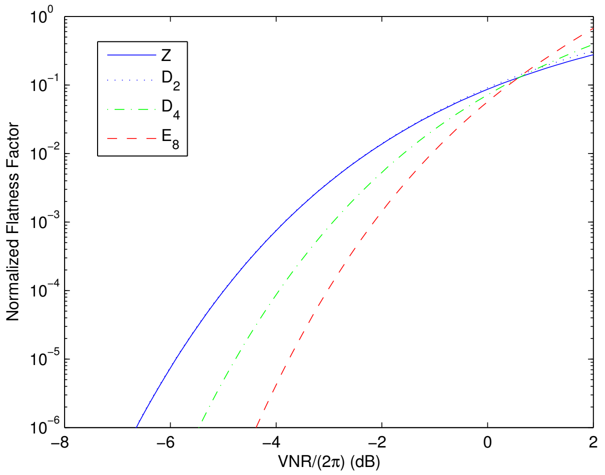


Figure 3.1: Normalized flatness factor $\frac{1}{n}\epsilon_{\Lambda_1}(\sigma)$ as a function of VNR for $\mathbb{Z}$, D_2, D_4 and E_8.

for $\mathbb{Z}$, D_2, D_4 and E_8. Clearly, the normalized flatness factors of D_4 and E_8 are much smaller than that of $\mathbb{Z}$ when the VNR is negative.

When it comes to the design of the code $\mathcal{C}$, one may apply single or multilevel constructions. The single-level construction corresponds to the well-known mod-p lattices, which are widely used in the theory of lattice coding, since $\mathcal{C}$ is a linear code over $\mathrm{GF}(p)$ [14, 15]. However, for large p, such non-binary codes are generally hard to decode. Lately, a practical design of such lattices, namely integer low-density lattice codes (LDA), is reported in [44, 51]. Its decoding complexity scales as p^2 under belief-propagation (BP) decoding.

Analogously, one might use non-binary polar codes to construct single-level polar lattices. The technical challenges associated with this approach are the design of non-binary polar codes for the Λ_1/Λ_r channel and the decoding complexity [22]. In the rest of this paper, we address such challenges by using multilevel constructions (Construction D). Since both encoding and decoding are done independently on different levels, the multilevel approach benefits from multi-stage decoding with lower complexity. More precisely, if r levels are used, it corresponds to $p = 2^{r-1}$ in mod-p lattices. The decoding complexity scales linearly in r rather than $p^2 = 2^{2(r-1)}$. On

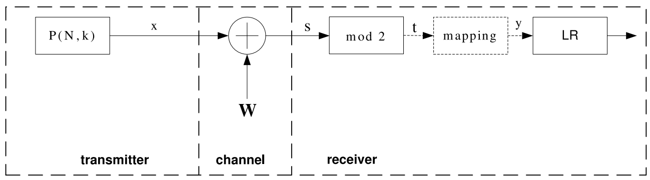


Figure 3.2: Signal flow of the mod-2 BAWGN channel. $P(N,k)$ represents the polar code with block length N and k information bits.

the other hand, this approach suffers from error propagation from level to level, and consequently the multi-stage decoder has to be carefully designed.

3.2 Polar Codes for Mod-2 BAWGN Channel

In this section, we will show how to construct a polar code for each level in Construction D. For clarity, we exemplify the construction for the one-dimensional partition chain, while the extension to the multi-dimensional partition chains is straightforward. In this case, the channel in each level is in fact a mod-2 binary-input AWGN (BAWGN) channel which is equivalent to a $\mathbb{Z}/2\mathbb{Z}$ channel. The only difference between the individual channels is the noise variance: the upper channels are noisier, while the lower channels are less noisy. The signal flow of the mod-2 BAWGN channel is shown in Figure 3.2, where the mod-2 operation is applied within $[-1,1]$, not $[0,2]$.

3.2.1 Capacity of Mod-2 BAWGN Channel

After the mod-$2\mathbb{Z}$ (mod-2) operation we have the PDF

$$f_{\sigma,2\mathbb{Z}}(w) = \sum_{\lambda\in 2\mathbb{Z}} f_\sigma(w+\lambda), \quad w\in[-1,1]. \tag{3.10}$$

With this density we can compute the capacity of the $\mathbb{Z}/2\mathbb{Z}$ channel $C(\mathbb{Z}/2\mathbb{Z})$ by applying (3.1) and (3.2), which is shown in Figure 3.3. Also shown are the capacities of the $2\mathbb{Z}/4\mathbb{Z}$ and $4\mathbb{Z}/8\mathbb{Z}$ channels, which can also be regarded as mod-2 BAWGN channels, but are upgraded versions (with smaller σ^2) compared with $\mathbb{Z}/2\mathbb{Z}$. For example, coding over $2\mathbb{Z}/4\mathbb{Z}$ is not very different with coding over $\mathbb{Z}/2\mathbb{Z}$. If we scale the $2\mathbb{Z}/4\mathbb{Z}$ channel by a factor $\frac{1}{2}$, then it becomes a $\mathbb{Z}/2\mathbb{Z}$ channel with the noise standard deviation $\frac{\sigma}{2}$. This observation simplifies our task to find good polar codes for the component channels. We just deal with the $\mathbb{Z}/2\mathbb{Z}$ channel with different noise variances.

We generalize this finding to the following lemma:

Lemma 3.3: Consider a lattice L constructed by a binary lattice partition chain $\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda_r$, where the channel of the ℓ-th level is the $\Lambda_\ell/\Lambda_{\ell+1}$ channel for $1 \leq \ell \leq r-1$. Then, the $\Lambda_{\ell-1}/\Lambda_\ell$ channel is degraded with respect to the $\Lambda_\ell/\Lambda_{\ell+1}$ channel.

Proof. By scaling the $\Lambda_\ell/\Lambda_{\ell+1}$ channel by $\frac{1}{2}$, the $\Lambda_\ell/\Lambda_{\ell+1}$ channel with σ is equivalent to the $\Lambda_{\ell-1}/\Lambda_\ell$ channel with $\sigma/2$. To see this, we use $\mathbb{Z}/2\mathbb{Z}$ channel and $2\mathbb{Z}/4\mathbb{Z}$ channel as an example. Let W_1 and W_2 represent $\mathbb{Z}/2\mathbb{Z}$ channel and $2\mathbb{Z}/4\mathbb{Z}$ channel respectively. The input and output of $\mathbb{Z}/2\mathbb{Z}$ channel $X_1 \in \{0,1\}$, $Y_1 \in [0,2)$. The input and output of $2\mathbb{Z}/4\mathbb{Z}$ channel $X_2 \in \{0,2\}$, $Y_2 \in [0,4)$. The transition PDF of $2\mathbb{Z}/4\mathbb{Z}$ channel is given by [39]

$$\begin{aligned}
W_2(y_2|x_2,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}}\sum_{a\in 4\mathbb{Z}} \exp\left[-\frac{(y_2-a-x_2)^2}{2\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\sum_{a\in 4\mathbb{Z}} \exp\left[-\frac{(\frac{1}{2}y_2-\frac{1}{2}a-\frac{1}{2}x_2)^2}{2(\frac{\sigma}{2})^2}\right] \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\sum_{a'\in 2\mathbb{Z}} \exp\left[-\frac{(\frac{1}{2}y_2-a'-\frac{1}{2}x_2)^2}{2(\frac{\sigma}{2})^2}\right] \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\sum_{a'\in 2\mathbb{Z}} \exp\left[-\frac{(y_1'-a'-x_1')^2}{2(\frac{\sigma}{2})^2}\right] \\
&= W_1(y_1'|x_1',(\frac{\sigma}{2})^2)
\end{aligned}$$

where $x_1' \in \{0,1\}$ and $y_1' \in [0,2)$.

Since we construct polar codes for the $\Lambda_{\ell-1}/\Lambda_\ell$ channel and the $\Lambda_\ell/\Lambda_{\ell+1}$ channel independently, we only need to prove $W_1(y_1|x_1,\sigma^2)$ is degraded with respect to $W_1'(y_1'|x_1,\sigma'^2)$ where $\sigma \geq \sigma'$. Note that they both are mod-2 BAWGN channels with $X \in \{0,1\}$ and $Y_1, Y_1' \in [0,2)$. Then

$$\begin{cases} W_1(y_1|x_1,\sigma^2) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \displaystyle\sum_{i\in 2\mathbb{Z}} \exp\left[-\dfrac{(y_1 - x_1 - i)^2}{2\sigma^2}\right], \\ W_1'(y_1'|x_1,\sigma'^2) = \dfrac{1}{\sqrt{2\pi\sigma'^2}} \displaystyle\sum_{i\in 2\mathbb{Z}} \exp\left[-\dfrac{(y_1' - x_1 - i)^2}{2\sigma'^2}\right]. \end{cases}$$

Assume there is a channel from Y_1' to Y_1 such that

$$W(y_1|y_1') = \frac{1}{\sqrt{2\pi(\sigma^2 - \sigma'^2}} \sum_{i\in 2\mathbb{Z}} \exp\left[-\frac{(y_1 - y_1' - i)^2}{2(\sigma^2 - \sigma'^2)}\right].$$

Then we have

$$\begin{aligned} &\int W_1'(y_1'|x_1,\sigma'^2) W(y_1|y_1') dy_1' \\ &= \frac{1}{2\pi\sqrt{\sigma'^2(\sigma^2-\sigma'^2)}} \sum_{i\in 2\mathbb{Z}} \sum_{i'\in 2\mathbb{Z}} \int_0^2 \exp\left[-\frac{(y_1' - x_1 - i)^2}{2\sigma'^2}\right] \exp\left[-\frac{(y_1 - y_1' - i')^2}{2(\sigma^2 - \sigma'^2)}\right] dy_1' \\ &= \frac{1}{2\pi\sqrt{\sigma'^2(\sigma^2-\sigma'^2)}} \sum_{i\in 2\mathbb{Z}} \sum_{i'\in 2\mathbb{Z}} \exp\left[-\frac{(y_1 - i' - x_1 - i)^2}{2\sigma^2}\right] \int_0^2 \exp\left[-\frac{(y_1' - i' - c)^2}{2\sigma'^2(\sigma^2-\sigma'^2)/\sigma^2}\right] dy_1' \\ &= \frac{1}{2\pi\sqrt{\sigma'^2(\sigma^2-\sigma'^2)}} \sum_{i\in 2\mathbb{Z}} \sum_{i'\in 2\mathbb{Z}} \exp\left[-\frac{(y_1 - i' - x_1 - i)^2}{2\sigma^2}\right] \int_{\mathbb{R}} \exp\left[-\frac{(y_1' - c)^2}{2\sigma'^2(\sigma^2-\sigma'^2)/\sigma^2}\right] dy_1' \\ &= \frac{1}{2\pi\sqrt{\sigma'^2(\sigma^2-\sigma'^2)}} \frac{\sqrt{2\pi\sigma'^2(\sigma^2-\sigma'^2)}}{\sigma^2} \sum_{i\in 2\mathbb{Z}} \sum_{i'\in 2\mathbb{Z}} \exp\left[-\frac{(y_1 - i' - x_1 - i)^2}{2\sigma^2}\right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{i\in 2\mathbb{Z}} \sum_{i'\in 2\mathbb{Z}} \exp\left[-\frac{(y_1 - i' - x_1 - i)^2}{2\sigma^2}\right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{i\in 2\mathbb{Z}} \exp\left[-\frac{(y_1 - x_1 - i)^2}{2\sigma^2}\right] \\ &= W_1(y_1|x_1,\sigma^2), \end{aligned}$$

where c is a constant.

The definition of the degradation is given as [61, Definition 1.7]: Let $W_1 : \mathcal{X} \to \mathcal{Y}_1$ and $W_2 : \mathcal{X} \to \mathcal{Y}_2$ be two channels. W_1 is degraded with respect to W_2 if there

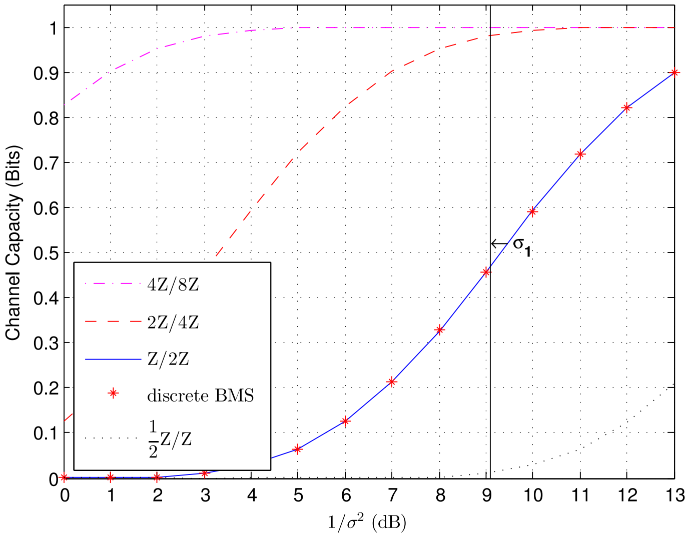


Figure 3.3: Channel capacity of the mod-2 BAWGN channel. The capacity of the discrete BMS channel is calculated in Section 3.2.3 to show the negligible difference between the continuous mod-2 BAWGN channel and the quantized discrete channel.

exists a channel $W : \mathcal{Y}_2 \to \mathcal{Y}_1$ such that

$$W_1(y_1|x) = \sum_{y_2 \in \mathcal{Y}_2} W_2(y_2|x) W(y_1|y_2).$$

Therefore, according to the definition, the $\Lambda_{\ell-1}/\Lambda_\ell$ channel is degraded with respect to the $\Lambda_\ell/\Lambda_{\ell+1}$ channel. □

3.2.2 Symmetry of Mod-2 BAWGN Channel

Polar codes are proved to be able to achieve the capacity of output-symmetric channels. Therefore we need to show this mod-2 BAWGN channel is indeed an output-symmetric channel. The output of the mapping operation in Figure 3.2 is

$$y = 2|t| - 1,$$

where the conditional PDFs of t can be derived from (3.10).

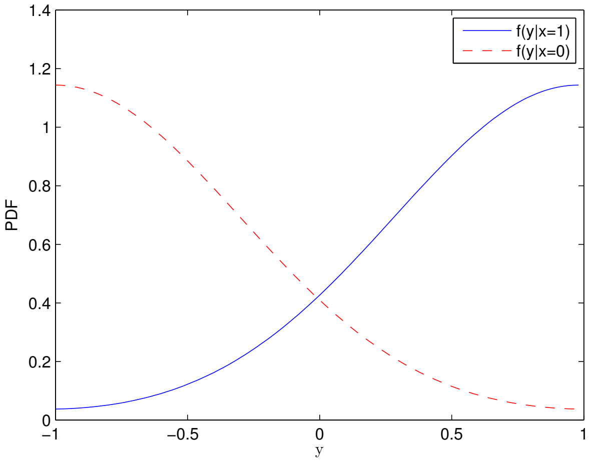


Figure 3.4: Conditional PDFs of y with $\sigma = 0.3380$.

Then the conditional PDFs of y are

$$\begin{cases} f(y|x=1) = \dfrac{1}{2\sqrt{2\pi\sigma^2}} \displaystyle\sum_{j=-\infty}^{+\infty} \exp\left[-\dfrac{(y-1+4j)^2}{8\sigma^2}\right], \\ f(y|x=0) = \dfrac{1}{2\sqrt{2\pi\sigma^2}} \displaystyle\sum_{j=-\infty}^{+\infty} \exp\left[-\dfrac{(y+1+4j)^2}{8\sigma^2}\right]. \end{cases}$$

From Figure 3.4, it is clearly a binary-input output-symmetric channel. The PDFs of the output satisfy

$$f(y|x=0) = f(-y|x=1), \text{ for all } y \in [-1,1].$$

3.2.3 Construction of Polar Codes for Mod-2 BAWGN channel

Let $W(y|x)$ be a BMS channel with input alphabet $\mathcal{X} = \{0,1\}$ and output alphabet $\mathcal{Y} \subseteq \mathbb{R}$. Polar codes are block codes of length $N = 2^m$ with input bits $\{u^i\}^{i=1:N}$. Let $I(W)$ be the capacity of W. Given the rate $R < I(W)$, the information bits are indexed by a set of RN rows of the generator matrix $G = \left[\begin{smallmatrix}1 & 0\\ 1 & 1\end{smallmatrix}\right]^{\otimes m}$. This gives an N-dimensional channel $W_N(y^{1:N}|u^{1:N})$. The channel seen by each bit [16] can be defined by

$$W_N^{(i)}(y^{1:N}, u^{1:i-1}|u^i) = \sum_{u^{i+1:N} \in \mathcal{X}^{N-i}} \frac{1}{2^{N-1}} W_N(y^{1:N}|u^{1:N}).$$

Arıkan proved that as N grows $W_N^{(i)}$ approaches either an error-free channel or a completely noisy channel. The set of completely noisy (resp. error-free) subchannels is called the frozen set $\mathcal{A}^{\mathcal{C}}$ (resp. free set $\mathcal{A}$). One sets $u^i = 0$ for $i \in \mathcal{A}^{\mathcal{C}}$ and only sends information bits within $\mathcal{A}$.

The decoding rule using the successive cancellation (SC) decoding is defined as

$$\hat{u}_i = \begin{cases} 0 & i \in \mathcal{A}^{\mathcal{C}} \quad \text{or} \quad \dfrac{W_N^{(i)}(y^{1:N}, \hat{u}^{1:i-1}|u^i = 0)}{W_N^{(i)}(y^{1:N}, \hat{u}^{1:i-1}|u^i = 1)} \geq 1 \quad \text{when } i \in \mathcal{A}, \\ 1 & \text{otherwise.} \end{cases}$$

Let P_B denote the block error probability of a binary polar code and by $P_e(W_N^{(i)})$ the error probability for the i-th subchannel. P_B can be upper-bounded by the sum of the subchannels' Bhattacharyya parameters $P_B \leq \Sigma_{i \in \mathcal{A}} Z(W_N^{(i)})$. It was shown in [62] that for any $\beta < \frac{1}{2}$,

$$\liminf_{m \to \infty} \frac{1}{N} \left| \{i : Z(W_N^{(i)}) < 2^{-N^\beta}\} \right| = I(W). \tag{3.11}$$

This means if the RN rows are chosen properly, as $m \to \infty$, the fraction of channels which have capacity close to 1 is about $I(W)$. Therefore, constructing polar codes is equivalent to choosing all the good indexes. However, the complexity of the exact calculation appears to be exponential in the block length. The authors in [63] proposed a practical quantization method which can transform a BMS channel with a continuous alphabet to a BMS channel with a specified finite output alphabet size. The authors in [64] proposed a practical approximation method to construct polar codes efficiently over any discrete-output BMS channel. We combined these two methods together in order to construct polar codes for the mod-2 BAWGN channel.

First of all, we need a collection of binary symmetric channels (BSCs) to approximate this mod-2 BAWGN channel (see also [65]). The output can be divided into several intervals A_i (positive) and $-A_i$ (negative), $1 \leq i \leq K$, which are symmetric with respect to $y = 0$. The i-th BSC is chosen with probability p_i and let the cross-over probability be x_i, which means $\mathbb{P}(A_i|0) = \mathbb{P}(-A_i|1) = x_i$. According to the conditional PDFs and the definition of cross-over probability, p_i and x_i of the i-th BSC channel can be calculated as

$$\begin{cases} p_i = \displaystyle\int_{A_i} f(y|x=1) + f(y|x=0)dy, \\ x_i = \dfrac{\int_{A_i} f(y|x=0)dy}{\int_{A_i} f(y|x=1) + f(y|x=0)dy}. \end{cases}$$

Therefore a set of BSCs are obtained from the mod-2 BAWGN channel.

In [63], the partition on the continuous alphabet is done by a function of the likelihood ratio (LR). For $y \geq 0$, the LR of y is given by

$$\zeta_y = \frac{f(y|x=1)}{f(y|x=0)}.$$

The symmetric capacity of W is

$$I(W) = \int_0^1 (f(y|x=1) + f(y|x=0))C[\zeta_y]dy, \tag{3.12}$$

where $C[\zeta]$ for $\zeta \geq 1$ is defined as

$$C[\zeta] = 1 - \frac{\zeta}{\zeta+1}\log\left(1 + \frac{1}{\zeta}\right) - \frac{1}{\zeta+1}\log(\zeta+1).$$

ζ_y and $C[\zeta_y]$ are both strictly increasing in y for $y \geq 0$. In our case, we let the maximum value of $C[\zeta]$ be $C_{\text{max}} = C[\zeta_1]$. For $1 \leq i \leq K$, each interval is defined

Table 3.2: The Channel Capacity Using Different Quantization Methods for the Mod-2 BAWGN Channel with $\sigma = 0.3380$

K	**Uniform** $I(Q')$	**Non-uniform** $I(Q)$	ς	**Upper bound**
8	0.5129	0.5124	0.0021	0.0998
16	0.5140	0.5138	0.0007	0.0499
32	0.5143	0.5143	0.0002	0.025
64	0.5144	0.5144	0.0001	0.0125

as

$$A_i = \left\{ y \geq 0 : \frac{i-1}{K} C_{\text{max}} \leq C[\zeta] \leq \frac{i}{K} C_{\text{max}} \right\}.$$

Thus, the number of discrete output symbols is $2K$. According to Lemma 13 from [63], the difference in symmetric capacities of the discrete-output BMS and the original continuous-output BMS can be bounded by $\frac{1}{K} C_{\text{max}}$. Although this bound may not be very tight, it is enough for our theoretic proof.

Remark 3.1: The afore-mentioned non-uniform partition gives us a theoretic guarantee. Yet, in numerical experiments, there is no essential difference between this method and “uniform quantization". Therefore, one can also use equal interval partitions in the practical design. For example, the capacity of mod-2 BAWGN channel with $\sigma = 0.3380$ (we choose this value for future reference) is $I(W) = 0.5145$ when calculated from the continuous density. Let Q be the non-uniformly quantized channel of W and let $\varsigma = I(W) - I(Q)$. Also, let $I(Q')$ be the uniformly quantized channel Q'. The results are compared in Table 3.2. We can see that if the quantization is sufficiently fine, the mod-2 BAWGN channel can be approximated by a discrete BMS channel.

Then we calculate $I(Q)$ with different values of σ^2. We set $K = 32$, which is sufficient for the approximation error to be negligible. We note that $1 - I(W)$ is the expectation of the function $-x \log(x) - (1-x)\log(1-x)$ over the distribution P_χ, where the random variable $\chi \in [0, 1/2]$ and its PDF is $P_\chi(x) = \sum_{i=1}^K p_i \delta(x - x_i)$. Actually this calculation is just the discrete version of (3.12). The results are shown in Figure 3.3. It can be seen that the capacity of this BMS channel is almost the same as the theoretic capacity of the mod-2 BAWGN channel.

After a discrete BMS is obtained from a mod-2 BAWGN channel, we can use the merging algorithm in [64] to construct polar codes[3]. The main idea behind their method is to perform the calculation approximately by restricting the number of output symbols in each level, whose complexity is $O(K^2N)$.

The details are given in Construction 3.1, where the function $g(x) = 2\sqrt{x(1-x)}$ for the Bhattacharyya parameter $Z(W)$ and $g(x) = -x\log(x) - (1-x)\log(1-x)$ for mutual information $I(W)$. This algorithm starts with distribution P_χ. It generates a tree from a BMS channel W as the root node according to (1) and (2) in [64]. Then it performs quantization on each level of the tree to reduce the size of the output alphabet for the next level. Finally, the transmitting channels with the least Bhattacharyya parameters are chosen.

Construction 3.1 Construction of Polar Codes

```
1: Start from the root node with parameters obtained by quantization: $(p_1, x_1), \cdots, (p_K, x_K)$ $(x_1 < \cdots < x_K)$.
2: Calculate the parameters of two new channels according to the rule of polarization.
3: Merge the parameter which has the minimum value of $p_j d_j$ $(d_j = g(x_{j+1}) - g(x_j))$ with its right neighbour and repeat this merge until the number of output symbols is $2K$.
4. Calculate the parameters of the next level.
5. Merge the parameters of new channels.
⋮
Stop until reaching the block length.
Then choose the transmitting channels with the least Bhattacharyya parameters.
```

Until here, we have shown how to construct polar codes over the mod-2 BAWGN channel. Although the above quantization algorithm results in an approximation

[3] In fact, two merging algorithms were presented in [64] where the second algorithm gives a slightly better performance guarantee. We just use the first algorithm which is sufficient for the purpose of this paper.

error, it can be bounded in a precise manner. All the merging algorithms in [63] and [64] share the same approximation error bound. It was shown in [64] that the “average" approximation error of the N channels is $O\left(\frac{m}{K}\right)$ and $K = m^2$ is enough that the approximation error decays to zero.

Now we introduce the *capacity loss* ϵ_{loss} under the quantization-merging algorithm and finite length. More precisely, it means that we can construct a polar code of length N over a channel with the symmetric capacity C such that this polar code is assured to have a block error probability $P_B \leq N2^{-N^\beta}$ $(\beta < 1/2)$ at the rate $C - \epsilon_{\text{loss}}$. This capacity loss is caused by the approximation error and the finite block length. We now give the following Lemma on the capacity loss, which is essentially a restatement of [63, Theorem 1].

Lemma 3.4: Let $\overline{P}_e(\tilde{W}_N^{(i)}, K)$ denote the upper bound on the probability of error under SC decoding for the degraded subchannel $\tilde{W}_N^{(i)}$ resulting from the quantization-merging algorithm. Given any constant $0 < \beta < 1/2$, let the capacity loss ϵ_{loss} be defined by

$$\frac{1}{N}\left|\{i : \overline{P}_e(\tilde{W}_N^{(i)}, K) < 2^{-N^\beta}\}\right| = I(W) - \epsilon_{\text{loss}}.$$

Then $\liminf\limits_{m\to\infty} \epsilon_{\text{loss}}$ is not a function of N, and

$$\lim_{K\to\infty} \liminf_{m\to\infty} \epsilon_{\text{loss}}(W, K, \beta) = 0.$$

Remark 3.2: From (57) in [63], it is shown that for any positive value ϵ_{loss} there exists K such that

$$\liminf_{m\to\infty} \frac{1}{N}\left|\{i : \overline{P}_e(\tilde{W}_N^{(i)}, K) < 2^{-N^\beta}\}\right| = I(W) - \epsilon_{\text{loss}}.$$

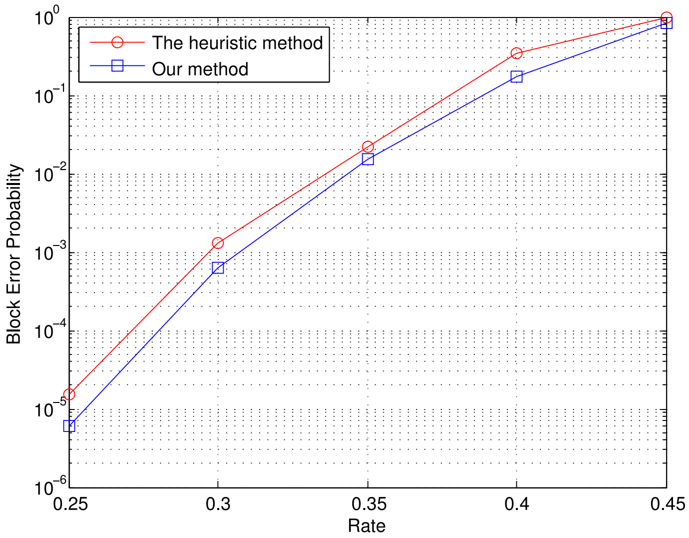


Figure 3.5: Block error probabilities of polar codes over the mod-2 BAWGN channel with $\sigma = 0.3380$ and $N = 1024$, respectively constructed by the heuristic BEC approximation [16] and by our method.

We may take K large enough so that the channel $\tilde{W}_N^{(i)}$ is arbitrarily close to the channel $W_N^{(i)}$ and $\overline{P}_e(\tilde{W}_N^{(i)}) \to P_e(W_N^{(i)})$. Then according to the polarization theory (3.11), this capacity loss ϵ_{loss} vanishes if $m \to \infty$. This lemma shows that we can get arbitrarily close to the optimal construction of a polar code as K increases.

3.2.4 Simulation Results

Arıkan proposed a heuristic method in [16] in which any BMS W is regarded as a binary erasure channel (BEC) with erasure probability $1 - I(W)$. Then one can construct a polar code over this BEC channel instead of the BMS channel. We compare the performance of different methods for the mod-2 BAWGN channel with $\sigma = 0.338$ and codeword $N = 1024$. Multiple rates are tested. From the simulation results in Figure 3.5, polar codes constructed by our method have better performance. This is because the other method assumes a mismatched channel model. Accordingly, our quantization method will be used to construct the component polar codes for the polar lattice. We also note that in [66], the transmission channel of each level is approximated by a Gaussian channel with the same capacity, which results in a suboptimal component polar code.

3.3 Construction of Polar Lattices

Polar lattices are constructed with the lattice partition chain $\Lambda_1/\Lambda_2 \cdots /\Lambda_r$ and the associated $(r-1)$ nested polar codes with block length N. More precisely, using the method given in the preceding section, we build a component polar code $P(N, k_\ell)$ at the ℓ-th level to achieve the capacity of the $\Lambda_\ell/\Lambda_{\ell+1}$ channel $(1 \leq \ell \leq r-1)$. In this section, we prove that polar lattices can asymptotically achieve the Poltyrev capacity as $N \to \infty$. We also present a performance analysis for finite N.

3.3.1 Nested Polar Codes

We start by showing that the component polar codes constructed at all levels are nested. This requirement is to guarantee that the multilevel construction creates a lattice. We consider two rules to determine the component codes, for theoretic and practical purposes, respectively. One is the *capacity rule* [39, 67], where we select the channel indexes according to a threshold on the mutual information. The other is the *equal-error-probability rule* [67], namely, the same error probability for each level, where we select the channel indexes according to a threshold on the Bhattacharyya parameter. The advantage of the equal-error-probability rule based on the Bhattacharyya parameter is that it leads to an upper bound on the error probability. For this reason, we use the equal-error-probability rule in the practical design. It is well-known that these two rules will converge as the block length goes to infinity [16].

Lemma 3.5: For either the capacity rule or the equal-error-probability rule, the component polar codes built in the multilevel construction are nested, i.e., $P(N, k_1) \subseteq P(N, k_2) \subseteq \cdots \subseteq P(N, k_{r-1})$.

Proof. Firstly, consider the equal-error-probability rule. By Lemma 4.7 in [61], if a BMS channel $\tilde{W}$ is a degraded version of W, then the subchannel $\tilde{W}_N^{(i)}$ is also degraded with respect to $W_N^{(i)}$ and $Z(\tilde{W}_N^{(i)}) \geq Z(W_N^{(i)})$. Let the threshold be, say 2^{-N^β} for some $\beta < 1/2$. Then, the block error probability of the polar code with SC decoding is upper-bounded by $N2^{-N^\beta}$. The codewords are generated by $x_1^N = u_\mathcal{A}G_\mathcal{A}$, where $G_\mathcal{A}$ is the submatrix of G formed by rows with indexes in the free set $\mathcal{A}$. The free sets for these two channels are respectively given by

$$\begin{cases} \mathcal{A} & =\{i : Z(W_N^{(i)}) < 2^{-N^\beta}\}, \\ \tilde{\mathcal{A}} & =\{i : Z(\tilde{W}_N^{(i)}) < 2^{-N^\beta}\}. \end{cases}$$

Due to the fact that $Z(\tilde{W}_N^{(i)}) \geq Z(W_N^{(i)})$, we have $\tilde{\mathcal{A}} \subseteq \mathcal{A}$. If we construct polar codes $P(N, \mathcal{A})$ over W and $P(N, \tilde{\mathcal{A}})$ over $\tilde{W}$, $G_{\tilde{\mathcal{A}}}$ is a submatrix of $G_\mathcal{A}$. Therefore $P(N, \tilde{\mathcal{A}}) \subseteq P(N, \mathcal{A})$.

From Lemma 3.3, the channel of the ℓ-th level is always degraded with respect to the channel of the $(\ell + 1)$-th level, and consequently, $P(N, k_\ell) \subseteq P(N, k_{\ell+1})$.

Then, consider the capacity rule. The nesting relation still holds if we select the channel indexes according to a threshold on the mutual information. This is because, by Lemma 4.7 in [61], $I(\tilde{W}_N^{(i)}) \leq I(W_N^{(i)})$ if a BMS channel $\tilde{W}$ is a degraded version of W. □

Concerning the effect of approximation errors of Construction 3.1, we find the nesting property of Lemma 3.5 still holds, as long as the number of quantization symbols K is sufficiently large.

3.3.2 Polar lattices are AWGN-good lattices

As we will use a multi-stage decoding algorithm, the overall (block) error probability $P_e(L, \sigma^2)$ of a polar lattice is upper-bounded by the sum of the block error probabilities at individual levels, by the union bound. Let $P_e(\mathcal{C}_\ell, \sigma^2)$ denote the block error probability of the polar code for the ℓ-th level $(1 \leq \ell \leq r-1)$, and by $P_e(\Lambda_r^N, \sigma^2)$ the error probability for the r-th level (i.e., product of the bottom lattice). Then

$$P_e(L, \sigma^2) \quad \leq \quad \sum_{\ell=1}^{r-1} P_e(\mathcal{C}_\ell, \sigma^2) + P_e(\Lambda_r^N, \sigma^2),$$

where the error probability of Λ_r^N is given by

$$P_e(\Lambda_r^N, \sigma^2) = 1 - \int_{\mathcal{V}(\Lambda_r^N)} f_{\sigma^2}(x)dx. \tag{3.13}$$

Let $\epsilon_{\text{loss}}(\ell)$ be the capacity loss of level ℓ. As shown in the previous subsection, we can construct nested polar codes (input length k_ℓ and block length N) with rates $R_\ell = \frac{k_\ell}{N} = C(\Lambda_\ell/\Lambda_{\ell+1}, \sigma^2) - \epsilon_{\text{loss}}(\ell)$ such that the block error probability in each level $P_B(\mathcal{C}_\ell, \sigma^2)$ is upper-bounded by $N2^{-N^{\beta_\ell}}$, for any $\beta_\ell < \frac{1}{2}$. Note that for finite N, the capacity loss is a function of N, the channel $\Lambda_\ell/\Lambda_{\ell+1}$, K, and β_ℓ. The capacity loss ϵ_3 is the sum of the capacity losses of the component codes:

$$\epsilon_3 = \sum_{\ell=1}^{r-1} \epsilon_{\text{loss}}(\ell). \tag{3.14}$$

By Lemma 3.4, we have

$$\lim_{N\to\infty} \lim_{K\to\infty} \epsilon_3 = \lim_{K\to\infty} \sum_{\ell=1}^{r-1} \epsilon(\Lambda_\ell/\Lambda_{\ell+1}, K, \beta_\ell) = 0,$$

As mentioned earlier, K does not need to be very large in practice. The sum error probability of polar codes is upper-bounded by

$$\sum_{\ell=1}^{r-1} P_e(\mathcal{C}_\ell, \sigma^2) \leq \sum_{\ell=1}^{r-1} N2^{-N^{\beta_\ell}}$$

which can be made arbitrarily small by increasing the block length N.

In conclusion, we have the following theorem:

Theorem 3.1: Let polar lattice L be constructed from the n-dimensional binary lattice partition chain $\Lambda_1/\Lambda_2 \cdots /\Lambda_r$ and $r-1$ nested polar codes with block length N, where $r = nO(\log N)$. Then, the error probability of L under multi-stage decoding is bounded by

$$P_e(L, \sigma^2) \leq \sum_{\ell=1}^{r-1} N2^{-N^{\beta_\ell}} + N\left(1 - \int_{\mathcal{V}(\Lambda_r)} f_{\sigma^2}(x)dx\right), \tag{3.15}$$

with the logarithmic VNR bounded by (3.9). In the limit as $\epsilon_{\Lambda_1}(\sigma) \to 0$, $N \to \infty$ and $K \to \infty$, L can achieve the Poltyrev capacity, i.e., $\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\right) \to 0$ and $P_e(L, \sigma^2) \to 0$.

Remark 3.3: It is worth pointing out that Theorem 3.1 only requires mild conditions. In practice, r and K need not be very large. The condition $\epsilon_{\Lambda_1}(\sigma) \to 0$ is also easily satisfied by properly scaling the top lattice Λ_1. Therefore, the essential condition in practice is $N \to \infty$.

3.3.3 Finite-Length Performance Analysis

In this subsection, we investigate the finite-length performance of polar lattices.

The finite-length analysis of polar codes was given in [68, 69, 70]. The analysis concerns the relationship between the block length and the rate for a fixed error probability. In other words, given a code and a desired (and fixed) error probability P_e, what is the block length N required, in terms of the rate R, so that the code has error probability less than P_e? It was proved that polar codes need a polynomial block length with respect to the gap to capacity $\epsilon_{\text{loss}} = I(W) - R = O(N^{-\frac{1}{\mu}})$ [68, 69], where μ is known as the scaling exponent. The lower bound of the gap is $\epsilon_{\text{loss}} \geq \underline{\beta} N^{-\frac{1}{\underline{\mu}}}$, where $\underline{\beta}$ is a constant that depends only on $I(W)$ and $\underline{\mu} = 3.55$ [68]. The upper bound of the gap is $\epsilon_{\text{loss}} \leq \bar{\beta} N^{-\frac{1}{\bar{\mu}}}$, where $\bar{\beta}$ is a constant that depends only on the block error probability P_B and $\bar{\mu} = 7$ was given in [68]. Later this scaling factor $\bar{\mu}$ has been improved to 5.77 [70].

From (3.9) and (3.14), the gap to the Poltyrev capacity of finite-dimensional polar lattices is

$$\begin{aligned}\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\right) &\leq \frac{2}{n}\left(\epsilon_1 + \sum_{\ell=1}^{r-1}\epsilon_{\text{loss}}(\ell)\right)\\ &\leq \frac{2}{n}\left(\epsilon_1 + (r-1)\bar{\beta}N^{-\frac{1}{\bar{\mu}}}\right)\end{aligned}$$

with the corresponding block error probability

$$P_e(L,\sigma^2) \leq (r-1)P_B + P_e(\Lambda_r^N, \sigma^2),$$

where the constant $\bar{\beta}$ depends only on P_B (assuming equal error probabilities for the component polar codes). Since $n \ll N$ is fixed, the gap to the Poltyrev capacity of polar lattices also scales polynomially in the dimension $n_L = nN$.

In comparison, the optimal bound for finite-dimensional lattices is given by [49]

$$\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\right)_{\text{opt}} = \sqrt{\frac{2}{n_L}}Q^{-1}(P_e(L,\sigma^2)) - \frac{1}{n_L}\log n_L + O\left(\frac{1}{n_L}\right). \tag{3.16}$$

At finite dimensions, this is more precise than the exponential error bound for lattices constructed from random linear codes given in [39]. Thus, given $P_e(L,\sigma^2)$, the scaling exponent of optimum random lattices is 2 which is smaller than that of polar lattices $\bar{\mu}$ (smaller scaling exponent means smaller block length needed to guarantee the same error probability). The result is consistent with the fact that polar codes require larger block length than random codes to achieve the same rate and error probability.

3.4 Decoding Algorithm

In the previous section we proposed a lattice construction from polar codes. In this section, we present a multi-stage decoding algorithm based on SC decoding of the component polar codes, exemplified by the one-dimensional lattice partition chain.

3.4.1 SC Decoding for Each Level

The noise analysis is crucial to the construction of polar codes on each level. Besides, it is also important to the performance of our SC decoder due to the calculation of the channel LR. This subsection will discuss the effective noise in view of the SC decoder on the first level. The derivation of other levels is the same but with different noise variances. Note that we use $\{0, 1\}$ notation to construct lattices. Referring to Figure 3.2, each element of the received vector $\mathbf{s}$ has the conditional PDF

$$\begin{cases} f(s|x=1) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\text{exp}\left[-\dfrac{(s-1)^2}{2\sigma^2}\right], \\ f(s|x=0) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\text{exp}\left[-\dfrac{(s)^2}{2\sigma^2}\right]. \end{cases}$$

The next step is to apply the mod-2 operation in $[-1, 1]$. From (3.10), the conditional PDFs of each element of the resultant $\mathbf{t}$ are

$$\begin{cases} f(t|x=1) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\displaystyle\sum_{j=-\infty}^{+\infty}\text{exp}\left[-\dfrac{(t-1+2j)^2}{2\sigma^2}\right], \\ f(t|x=0) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\displaystyle\sum_{j=-\infty}^{+\infty}\text{exp}\left[-\dfrac{(t+2j)^2}{2\sigma^2}\right]. \end{cases}$$

The channel LR for the SC decoder is derived from the conditional PDFs:

$$\zeta_t = \frac{f(t|x=1)}{f(t|x=0)}. \tag{3.17}$$

To calculate the LR exactly one needs to add infinite terms. Some truncation is necessary which depends on the level. We can truncate more terms for the levels closer to the bottom lattice.

3.4.2 Multi-Stage Decoder for Multilevel Construction

A multi-stage decoding algorithm is introduced to feed soft values of the received vector s into the binary decoder in each level. Algorithm 3.1 works on the real domain for the one-dimensional lattice partition. It describes the implementation of this multi-stage soft decoding algorithm, where the index ℓ represents the ℓ-th level $(1 \leq \ell \leq r-1)$, s is the received vector and σ^2 is the variance of the mod-2 BAWGN channel used to calculate the channel LR for the SC decoder. The calculation of the LR needs the information about the conditional PDFs after the mod-2 operation. We apply the mod-2 operation, use the SC decoder to estimate the codeword, and subtract it out from the received vector s. After that the received vector is divided by 2 and we run the same decoding for the second level. This is because the channel of the second level is equivalent to the channel of the first level with half of its noise standard deviation. We keep running this decoding until the $(r-1)$-th level. Finally we return the nearest lattice point $\mathbb{Z}^N$ to the received vector of the r-th level.

Algorithm 3.1 Multi-Stage Soft Decoding Algorithm for Multilevel Construction

```
function MULTI-STAGEDECODER($\ell, \mathrm{s}, \sigma$)
    if $\ell = r$ then
        return $\mathbf{z} =$ LatticeDecoder($\mathbb{Z}^N, \mathrm{s}$)
    else
        $\mathbf{t} = \mathrm{mod}(\mathrm{s}, 2)$                ▷ mod-2 operation in [-1,1]
        $\zeta_{\mathbf{t}} = \frac{f(\mathbf{t}|\mathbf{0})}{f(\mathbf{t}|\mathbf{1})}$
        $\mathbf{c} \leftarrow$ SCdecoder($\ell, \zeta_{\mathbf{t}}, \sigma$)
        $\mathbf{v} \leftarrow$ Multi-StageDecoder($\ell + 1, (\mathbf{s} - \mathbf{c})/2, \sigma/2$)
        return $\mathbf{z} = \mathbf{c} + 2 \cdot \mathbf{v}$
    end if
end function
```

3.4.3 Decoding Complexity

The complexity of the mod-2 operation and calculating the LR is negligible compared to the complexity of SC decoding which is $O(N \log N)$. Therefore the overall complexity of decoding such a multilevel lattice is $O(rN \log N)$, the number of levels times the complexity of the SC decoder.

3.5 Design Examples for the infinite constellation

In this section, we give design examples of polar lattices based on one and two-dimensional partition chains. The design follows the equal-error-probability rule. If the total target error probability is P_e, then the target error probability for each level should be $\frac{P_e}{r}$ where r is the number of levels. It is not difficult to extend the design procedure to higher-dimensional partition chains.

3.5.1 One-Dimensional Lattice Partition

In this subsection, we use the one-dimensional lattice partition $\mathbb{Z}/2\mathbb{Z}/\cdots/2^r\mathbb{Z}$. To construct a multilevel lattice, one needs to determine the number of levels of lattice partitions and the actual rates according to the the target error probability for a given noise variance. In addition to the guidelines given in Section 3.1, we have the following rule of thumb:

If a component code cannot achieve the target error probability at the rate $1/N$ for a reference σ, the corresponding level is not needed. This is because not even one bit will be polarized. On the other hand, if the rate of a level is almost 1 and still can achieve the target error probability for a reference σ, it is also not needed. This is because the rate from this level will be canceled by the increment of the logarithmic volume $\log V(\Lambda_r)$ in (3.7), leaving the VNR unchanged. The effective levels are

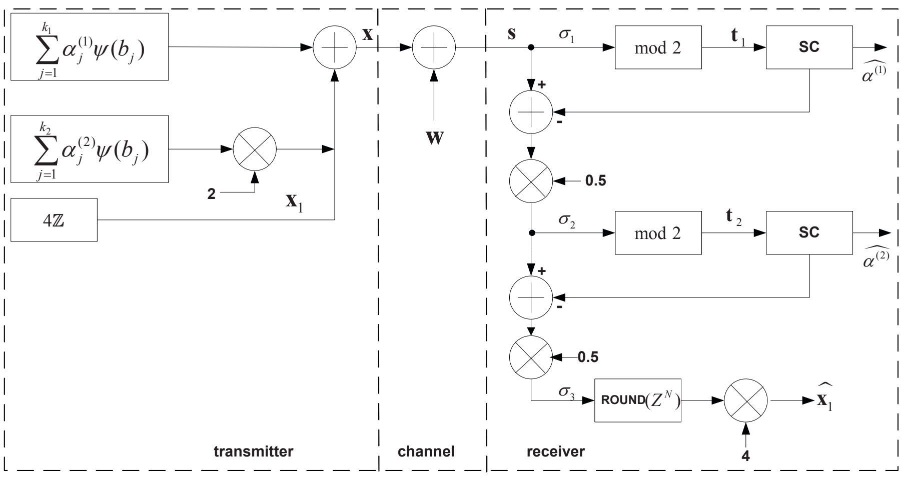


Figure 3.6: A polar lattice with two levels, where $\sigma = \sigma_1$.

those which can achieve the target error probability with an actual rate not too close to either 0 or 1. Therefore, one can determine the number of effective levels with the help of capacity curves in Figure 3.3. In other words, adding levels whose capacities are close to 1 or 0 do not noticeably improve the performance.

For example, at the given noise variance indicated by the straight line in Figure 3.3, one may choose three levels. However, the first level has an almost zero rate for the target error probability. Therefore, we choose two levels of component codes, which was indeed suggested in [39].

The multilevel construction and the multi-stage decoding are shown in Figure 3.6. For the ℓ-th level, $\alpha^{(\ell)}$ are information bits, $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_{k_\ell}$ are a set of code generators which are chosen from the matrix $G_N = \left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]^{\otimes m}$ according to the polarization rule for the ℓ-th level's channel, and σ_ℓ is the standard deviation of the noise.

Now, we give an example for length $N = 1024$ and target error probability $P_e(L, \sigma^2) = 10^{-5}$. Since the bottom level is a $\mathbb{Z}^N$ lattice decoder, $\sigma_3 \approx 0.0845$ for target error probability $\frac{1}{3}\cdot 10^{-5}$. For the middle level, $\sigma_2 = 2\cdot\sigma_3 = 0.1690$. From Figure 3.3, the channel capacity of the middle level is $C(\mathbb{Z}/2\mathbb{Z}, \sigma_2^2) = C(2\mathbb{Z}/4\mathbb{Z}, \sigma_1^2) = 0.9874$. For the top level, $\sigma = \sigma_1 = 0.3380$ and the capacity is 0.5145. Our goal is

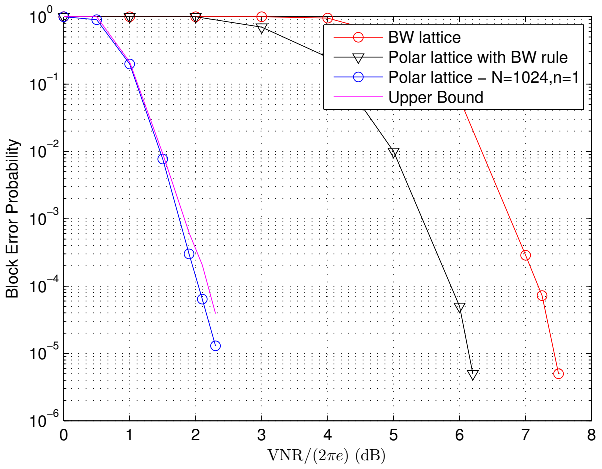


Figure 3.7: Block error probabilities of polar lattices and Barnes-Wall (BW) lattices of length $N = 1024$ with multi-stage decoding. BW rule means following the structure of the Barnes-Wall lattice, but changing the Reed-Muller code to a polar code on each level to construct lattices.

to find two polar codes approaching the respective capacities and block error probabilities $\leq \frac{1}{3} \cdot 10^{-5}$ over these mod-2 BAWGN channels.

For $N = 1024$, we found the first polar code with $\frac{k_1}{N} = 0.23$ for $P_e(\mathcal{C}_1, \sigma_1^2) \approx \frac{1}{3} \cdot 10^{-5}$, and the second polar code with $\frac{k_2}{N} = 0.9$ for $P_e(\mathcal{C}_2, \sigma_2^2) \approx \frac{1}{3} \cdot 10^{-5}$. Thus, the sum rate of component polar codes $R_{\mathcal{C}} = 0.23 + 0.9$, implying a capacity loss $\epsilon_3 = 0.3719$. Meanwhile, the factor $\epsilon_1 = C(\mathbb{Z}, 0.3380^2) = 0.0160$. From (3.9), the logarithmic VNR is given by

$$\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\right) \leq 2\left(\epsilon_1 + \epsilon_3\right) = 0.7758, \tag{3.18}$$

which is 2.34 dB. Figure 3.7 shows the simulation results for this example. It is seen that the estimate 2.34 dB is very close to the actual gap at $P_e(L, \sigma_1^2) \approx 10^{-5}$. This simulation indicates that the performance of the component codes is very important to the multilevel lattice. The gap to the Poltyrev capacity is largely due to the capacity losses of component codes.

Thanks to density evolution [71], the upper bound $\sum_{i\in\mathcal{A}}(Z(W_N^{(i)}))$ on the block error probability of a polar code with finite-length can be calculated numerically.

According to (3.15), we plot the upper bound on the block error probability $P_e(L, \sigma^2)$ of the polar lattice in Figure 3.7, which is quite tight.

Now we revisit the Barnes-Wall lattice with its performance shown in Figure 3.7. We know that there are only 2 effective levels, but the Barnes-Wall lattice (2.5) has 5 levels for $N = 1024$. The reason for its relatively poor performance is that it violates the capacity rule: at some levels, the rate of the code exceeds the capacity of the equivalent channel. For example, the rate of the first level is 0.01, which exceeds the capacity of the first level[4]. Another reason is the relatively weak error-correction ability of Reed-Muller codes. Therefore, the error probability of the first level will be high in the low VNR region. Also shown in Figure 3.7 is our prior design [72], where we followed the structure of the Barnes-Wall lattice, but changed the Reed-Muller code to a polar code on each level. It is seen that replacing the Barnes-Wall rule with our new design yields significantly improved performance.

We use the same multi-stage decoder for both polar lattices and Barnes-Wall lattices. Thus, the encoding and decoding complexity of polar lattices is almost the same as that of Barnes-Wall lattices.

3.5.2 Two-Dimensional Lattice Partition

In this subsection, we use the two-dimensional lattice partition

$$\mathbb{Z}^2/R\mathbb{Z}^2/2\mathbb{Z}^2/2R\mathbb{Z}^2/4\mathbb{Z}^2$$

as an example. With some abuse of notation, here R denotes the rotation operator represented by matrix $\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ [2]. The capacities of the component channels are shown in Figure 3.8. According to the rule of level selection, the number of effective levels for component codes is 4.

[4] Similar performance degradation was observed in multilevel coding, where an excessive rate at the lowest level results in a tremendous increase of nearest neighbors [67].

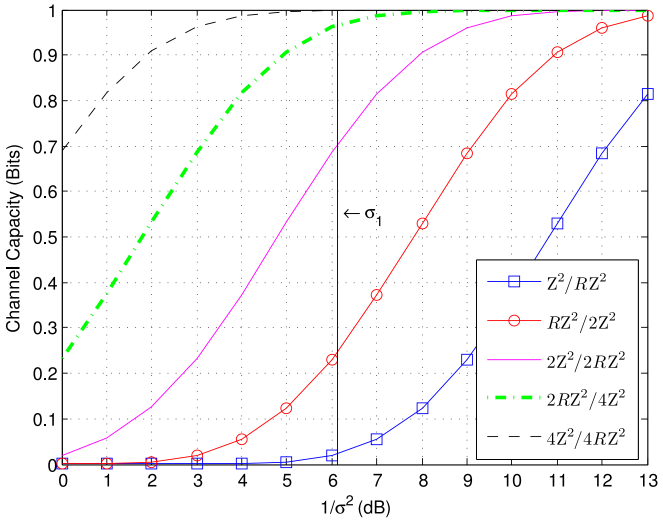


Figure 3.8: Channel capacity of the two-dimensional lattice partition.

This polar lattice with two-dimensional lattice partition is depicted in Figure 3.9.

It consists of all vectors of the form

$$\sum_{j=1}^{k_1} \alpha_j^{(1)} \psi(b_j) \otimes \mathbf{g}_1 + \cdots + \sum_{j=1}^{k_4} \alpha_j^{(4)} \psi(b_j) \otimes \mathbf{g}_4 + \mathbf{l}, \tag{3.19}$$

where $\alpha_j^{(\ell)} \in \{0,1\}$ for $1 \leq \ell \leq 4$, $\mathbf{l} \in (4\mathbb{Z}^2)^N$ and $\mathbf{g}_\ell$ $(1 \leq \ell \leq 4)$ is the generator of the coset representative $[\Lambda_\ell/\Lambda_{\ell+1}]$ for the partition $\Lambda_\ell/\Lambda_{\ell+1}$ [2]. $\mathbf{g}_\ell$ is an element of Λ_ℓ but not of $\Lambda_{\ell+1}$. To be more specific, $\mathbf{g}_1 = (1,0)$, $\mathbf{g}_2 = (1,1)$, $\mathbf{g}_3 = (2,0)$, and $\mathbf{g}_4 = (2,2)$. Let $G_R = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ be the generator matrix of $R\mathbb{Z}^2$, then $\mathbf{g}_{\ell+1} = \mathbf{g}_\ell \cdot G_R$. Each level is a “mod-$R\mathbb{Z}^2$ BAWGN" channel. We note here $\sigma_\ell = \sqrt{2}\sigma_{\ell+1}$ for $1 \leq \ell \leq 4$. Then (3.19) can be written as

$$\sum_{\ell=1}^{4} \sum_{j=1}^{k_\ell} \alpha_j^{(\ell)} \psi(b_j) \otimes \mathbf{g}_1 \cdot G_R^{\ell-1} + \mathbf{l}.$$

The calculation the channel likelihood ratio is a coset decoding problem. After

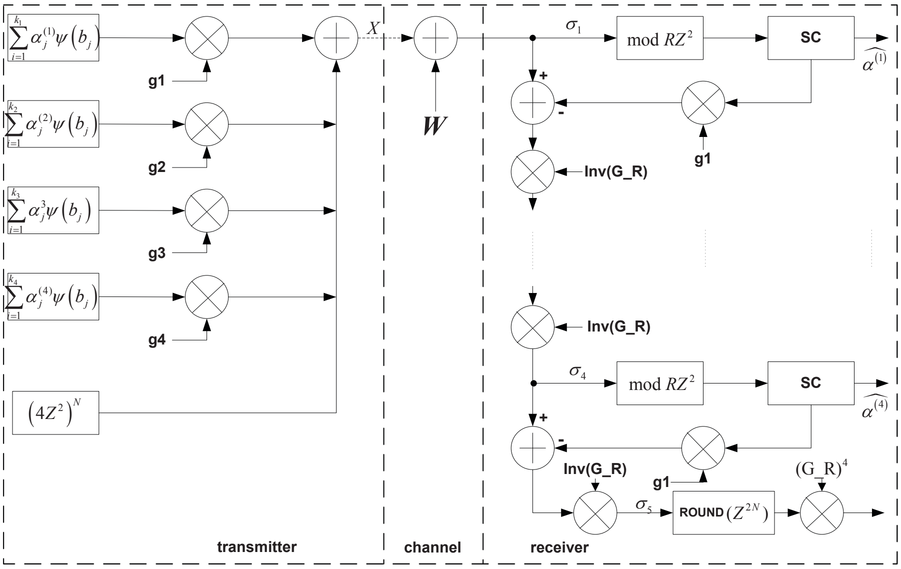


Figure 3.9: A polar lattice with five levels, where $\sigma = \sigma_1$.

the mod-$R\mathbb{Z}^2$ operation, the PDF of the Gaussian noise is given by

$$f_{\sigma,R\mathbb{Z}^2}(\mathbf{x}) = \sum_{\lambda \in R\mathbb{Z}^2} f_\sigma(\mathbf{x}+\lambda), \quad \mathbf{x} \in \mathcal{R}(R\mathbb{Z}^2).$$

Therefore the likelihood ratio is

$$\mathrm{LR} = \frac{f_{\sigma,R\mathbb{Z}^2}(\mathbf{x})}{f_{\sigma,R\mathbb{Z}^2}(\mathbf{x}-(1,0))}.$$

For a fair comparison, let us design such a polar lattice for $n = 2$ and $N = 512$, thus the same dimension $n_L = 1024$, for target error probability $P_e(L, \sigma^2) = 10^{-5}$. Since $r = 5$, $\sigma_5 \approx 0.083$ to make sure the error probability of the bottom level is $\frac{1}{5} \cdot 10^{-5}$. We also have $\sigma_1 = \sigma_5 \cdot (\sqrt{2})^4 = 0.332$. The actual rates of the component codes to achieve $\frac{1}{5} \cdot 10^{-5}$ are found to be 0.07, 0.40, 0.825 and 0.981, respectively. The channel capacities for each level are 0.2488, 0.7064, 0.9666 and 0.9996. Meanwhile, $\epsilon_1 = C(\mathbb{Z}^2, 0.332^2) = 0.0374$ and $\epsilon_3 = 0.6453$. From (3.9), the

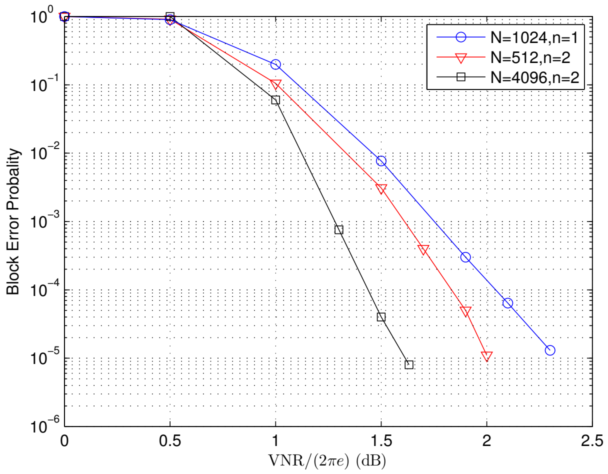


Figure 3.10: Block error probabilities of polar lattices with multi-stage decoding.

gap to the Poltyrev capacity is

$$\log\left(\frac{\gamma_L(\tilde{\sigma})}{2\pi e}\right) = \epsilon_1 + \epsilon_3 = 0.6827,$$

which is 2.05 dB. Again, the gap to the Poltyrev capacity is largely due to the capacity losses of component codes.

The simulation result of a polar lattice with $N = 4096$ and $n = 2$ (so the dimension is $n_L = 8192$) is also shown in Figure 3.10. For this lattice, the gap to the Poltyrev capacity is only 1.5 dB at block error probability 10^{-5}.

3.6 Summary

We show how to construct AWGN-good polar lattices in this chapter. In particular, polar codes constructed for each level is capacity-achieving. Furthermore, due to the degradation between each level, the component polar codes are nested. This is the requirement of Construction D. Polar lattices has been proved to be good without power constraint. We will propose a shaping scheme over polar lattices in order to communicate with power constraint in the next chapter Chapter 4.

CHAPTER 4

Polar lattice codes can achieve the channel capacity of the AWGN channel $\frac{1}{2}\log(1+\mathsf{SNR})$

In this Chapter, we show that an AWGN-good polar lattice with a good constellation can achieve the channel capacity of the AWGN channel. This is equivalent to implementing the shaping over the AWGN-good polar lattice. Recall that the basic idea of shaping is to generate the distribution of the input by finding the connections between input bits. The recently introduced asymmetric polar codes [8] are powerful tool to find such connections and implement the shaping. [1]

4.1 Good constellations for multilevel lattice codes

In order to achieve the AWGN channel capacity, a good constellation for the AWGN-good lattice is necessary. As shown in [4], the mutual information between the discrete Gaussian lattice distribution $D_{\eta\Lambda,\sigma_s}$ (Figure 1.4) and the output of the AWGN channel approaches $\frac{1}{2}\log(1+\mathsf{SNR})$ as the flatness factor $\epsilon_{\eta\Lambda}\left(\frac{\sigma_s\sigma}{\sqrt{\sigma_s^2+\sigma^2}}\right) \rightarrow 0$ where η is a scaling factor. Note that Λ is the top lattice in our lattice partition chain. There-

[1] Thanks Ling Liu for his contribution on the proof of Theorem 4.4.

fore throughout this work, we use the lattice Gaussian distribution $P(X) \sim D_{\eta\Lambda}, \sigma_s$ as the constellation. This gives us $\lim_{r\to\infty} P(X_{1:r}) = P(X) \sim D_{\eta\Lambda,\sigma_s}$ which is shown in Figure 4.1.

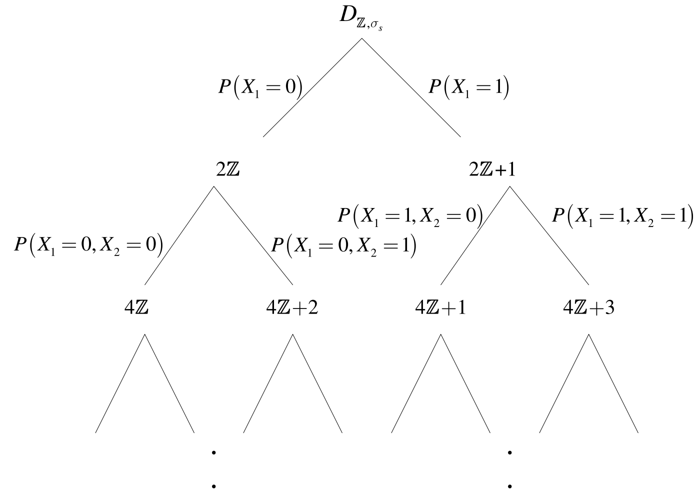


Figure 4.1: The lattice Gaussian distribution for $D_{\mathbb{Z},\sigma_s}$.

From the chain rule of mutual information,

$$I(Y;X_{1:r}) = \sum_{\ell=1}^{r} I(Y;X_\ell|X_{1:\ell-1}), \tag{4.1}$$

we have r binary-input channels. Given $x_{1:\ell-1}$, let $\mathcal{A}_\ell(x_{1:\ell})$ denote the set of the chosen constellation. According to [67], the channel transition PDF of the ℓ-th channel is given by

$$\begin{aligned}
&P_{Y|X_\ell,X_{1:\ell-1}}(y|x_\ell,x_{1:\ell-1}) = \frac{1}{P\{\mathcal{A}_\ell(x_{1:\ell})\}} \sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} P(a)P_{Y|A}(y|a) \\
&= \frac{1}{f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))} \sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} \exp\left(-\frac{|y-a|^2}{2\sigma^2} - \frac{a^2}{2\sigma_s^2}\right) \\
&= \exp\left(-\frac{y^2}{2(\sigma_s^2+\sigma^2)}\right) \frac{1}{f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))} \frac{1}{2\pi\sigma\sigma_s} \sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} \exp\left(-\frac{1}{2}\left(\frac{\sigma_s^2+\sigma^2}{\sigma_s^2\sigma^2}\left|\frac{\sigma_s^2}{\sigma_s^2+\sigma^2}y-a\right|^2\right)\right) \\
&= \exp\left(-\frac{y^2}{2(\sigma_s^2+\sigma^2)}\right) \frac{1}{f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))} \frac{1}{2\pi\sigma\sigma_s} \sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} \exp\left(-\frac{1}{2\tilde{\sigma}^2}\left(|\alpha y-a|^2\right)\right), \qquad (4.2)
\end{aligned}$$

where $\alpha = \frac{\sigma_s^2}{\sigma_s^2+\sigma^2}$ is asymptotically equal to the MMSE coefficient $\frac{P_s}{P_s+\sigma^2}$, and $\tilde{\sigma} = \frac{\sigma_s\sigma}{\sqrt{\sigma_s^2+\sigma^2}}$, with P_s and σ^2 denoting signal power and noise power, respectively.

Therefore if we use $D_{\eta\Lambda,\sigma_s}$ as the constellation, the ℓ-th channel is generally asymmetric with the input distribution $P(X_\ell|X_{1:\ell-1})$ ($\ell \leq r$), unless

$$f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))/f_{\sigma_s}(\mathcal{A}_{\ell-1}(x_{1:\ell-1})) = \frac{1}{2}$$

which means $\epsilon_{\mathcal{A}_\ell(x_{1:\ell})}(\sigma_s)$ is negligible.

From the above we know that the number of levels r needs to be infinity such that the input distribution is $D_{\eta\mathbb{Z},\sigma_s}$. We now describe this in a quantitative manner showing that how large the number of levels should be in order to achieve the channel capacity. In other words, the number of levels should be large enough to guarantee a vanishing mutual information of the bottom level.

Lemma 4.1: The mutual information of the bottom level $I(Y;X_r|X_{1:r-1})$ goes to 0 if the number of levels $r = O(\log N)$ and N goes infinity. Moreover, using the first $r-1$ levels would involve a capacity loss $\sum_{\ell \geq r} I(Y;X_\ell|X_{1:\ell-1}) \leq O(\frac{1}{N})$.

Proof. For level r, note that $\mathcal{A}_r$ is defined as $x_1 + \cdots 2^{r-1}x_r + 2^r\mathbb{Z}$. Clearly, $\mathcal{A}_r$ is a subset of $\mathcal{A}_{r-1}$. Let λ_1 and λ_2 denote the two lattice points with smallest norm in set $\mathcal{A}_{r-1}$. Without loss of generality, we assume $\lambda_1 \leq 0 \leq \lambda_2$ and $|\lambda_1| \leq |\lambda_2|$. Observe that $\lambda_2 - \lambda_1 = 2^{r-1}$. For a Gaussian distribution with variance σ_s^2, we can find a positive integer T, making the probability

$$\int_{-T\sigma_s}^{T\sigma_s} \frac{1}{\sqrt{2\pi\sigma_s^2}}\text{exp}(-\frac{x^2}{2\sigma_s^2})dx \to 1.$$

Actually, this T does not need to be very large. For instance, when $T = 6$, the above probability is larger than $1 - 2e^{-9}$. Now we assume $2^{r-1} = 3T\sigma_s$, and $T = \delta N$ for some constant δ, then λ_1 and λ_2 cannot be in the interval $[-T\sigma_s, T\sigma_s]$ simultaneously. If the two points are both outside of $[-T\sigma_s, T\sigma_s]$, then we have

$$P(\mathcal{A}_{r-1}) < 2\int_{-\infty}^{-T\sigma_s} \frac{1}{\sqrt{2\pi\sigma_s^2}}\text{exp}(-\frac{x^2}{2\sigma_s^2})dx \to 0,$$

which means the probability of choosing $\mathcal{A}_{r-1}$ goes to zero. This is in contradiction to the assumption. Therefore, we have that the point λ_1 is in the interval $[-T\sigma_s, T\sigma_s]$ and λ_2 is outside the interval. The two cosets according to $x_r = 0$ and $x_r = 1$ are $\lambda_1 + 2^r\mathbb{Z}$ and $\lambda_2 + 2^r\mathbb{Z}$ respectively. We have

$$\begin{aligned}
\frac{P(x_r = 0|x_{1:r-1})}{P(x_r = 1|x_{1:r-1})} &= \frac{\sum_{2^r\mathbb{Z}} \text{exp}(-\frac{(x+\lambda_1)^2}{2\sigma_s^2})}{\sum_{2^r\mathbb{Z}} \text{exp}(-\frac{(x+\lambda_2)^2}{2\sigma_s^2})} \\
&\geq \frac{\text{exp}(-\frac{\lambda_1^2}{2\sigma_s^2})}{2\sum_{2^r\mathbb{Z}+} \text{exp}(-\frac{\lambda_2^2}{2\sigma_s^2})} \\
&\geq \frac{\text{exp}(-\frac{\lambda_1^2}{2\sigma_s^2})}{2\cdot\text{exp}(-\frac{\lambda_2^2}{2\sigma_s^2})}(1-\text{exp}(-\frac{(2^r)^2}{2\sigma_s^2}))
\end{aligned}$$

Since $\lambda_2 - \lambda_1 = 2^{r-1} = 3T\sigma_s$ and $\lambda_2 + \lambda_1 \geq T\sigma_s$, we have

$$\begin{aligned}
\frac{P(x_r = 0|x_{1:r-1})}{P(x_r = 1|x_{1:r-1})} &\geq \frac{1}{2}\text{exp}(\frac{3}{2}T^2)(1-\text{exp}(-18T^2)) \\
&\geq \frac{3}{4}T^2 = \frac{3}{4}\delta^2N^2.
\end{aligned}$$

Assume that $\frac{3}{4}\delta^2N^2 = M$, we can get $P(x_r = 0|x_{1:r-1}) \geq \frac{M}{M+1}$ and $P(x_r = 1|x_{1:r-1}) \leq \frac{1}{M+1}$. Then we have,

$$I(Y;X_r|X_{1:r-1}) \leq H(X_r|X_{1:r-1}) \leq h_2(\frac{1}{M+1}),$$

where $h_2(p) = p\text{log}(\frac{1}{p}) + (1-p)\text{log}(\frac{1}{(1-p)})$ denotes the binary entropy function. By the relationship $\ln(x) \leq \frac{x-1}{\sqrt{x}}$ when $x \geq 1$, we finally have

$$I(Y; X_r|X_{1:r-1}) \leq \log(e)(\frac{1}{\sqrt{M}} + \frac{1}{M}) = \epsilon_1 \frac{1}{2^r} + \epsilon_2 \frac{1}{2^{2r}},$$

where ϵ_1 and ϵ_2 are two positive constants. Therefore, when $r = O(\log N)$, we have $I(Y; X_r|X_{1:r-1}) \to 0$, and $\sum_{\ell \geq r} I(Y; X_\ell|X_{1:\ell-1}) \leq O(\frac{1}{N})$. □

4.2 Asymmetric Polar Codes

The polar codes for the BMAs are introduced in [8]. It provides a feasible way to do the shaping over polar codes.

Definition 1 (Bhattacharyya Parameter for BMA Channel [7, 8]): Let W be a BMA channel with input $X \in \mathcal{X} = \{0, 1\}$ and output $Y \in \mathcal{Y}$. The input distribution and channel transition probability is denoted by P_X and $P_{Y|X}$ respectively. The Bhattacharyya parameter Z for W is then defined as

$$\begin{aligned} Z(X|Y) &= 2\sum_y P_Y(y)\sqrt{P_{X|Y}(0|y)P_{X|Y}(1|y)} \\ &= 2\sum_y \sqrt{P_{X,Y}(0,y)P_{X,Y}(1,y)}. \end{aligned}$$

The following lemma shows that by adding an observable at the output of W, Z will not decrease.

Lemma 4.2 (Conditioning reduces Bhattacharyya parameter Z): Let $(X, Y, Y') \sim P_{X,Y,Y'}, X \in \mathcal{X} = \{0, 1\}, Y \in \mathcal{Y}, Y' \in \mathcal{Y}'$, we have

$$Z(X|Y, Y') \leq Z(X|Y).$$

Proof.

$$\begin{aligned}
Z(X|Y,Y') &= 2\sum_{y,y'}\sqrt{P_{X,Y,Y'}(0,y,y')P_{X,Y,Y'}(1,y,y')}\\
&= 2\sum_{y}\sum_{y'}\sqrt{P_{X,Y,Y'}(0,y,y')}\sqrt{P_{X,Y,Y'}(1,y,y')}\\
&\overset{(a)}{\leq} 2\sum_{y}\sqrt{\sum_{y'}P_{X,Y,Y'}(0,y,y')}\sqrt{\sum_{y'}P_{X,Y,Y'}(1,y,y')}\\
&= 2\sum_{y}\sqrt{P_{X,Y}(0,y)P_{X,Y}(0,y)}
\end{aligned}$$

where (a) follows from Cauchy-Schwartz inequality. □

Let $X^{1:N}$ and $Y^{1:N}$ be the input and output vector after N independent uses of W. For each $i \in [N]$, $(X^i, Y^i) \sim P_{XY} = P_X P_{Y|X}$. Let $N = 2^n$ for integer $n \geq 1$. Consider polarized random variables $U^{1:N} = X^{1:N}G_N$ generated by the matrix $G_N = \left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]^{\otimes n}$, where $\otimes$ denotes the Kronecker product.

Theorem 4.1 (Polarization of Random Variables [8]): For any $\beta \in (0, 0.5)$,

$$\begin{cases}
\lim_{N\to\infty}\frac{1}{N}\left|\left\{i: Z(U^i|U^{1:i-1}) \geq 1-2^{-N^\beta}\right\}\right| = H(X),\\
\lim_{N\to\infty}\frac{1}{N}\left|\left\{i: Z(U^i|U^{1:i-1}) \leq 2^{-N^\beta}\right\}\right| = 1-H(X),\\
\lim_{N\to\infty}\frac{1}{N}\left|\left\{i: Z(U^i|U^{1:i-1},Y^{1:N}) \geq 1-2^{-N^\beta}\right\}\right| = H(Y|X),\\
\lim_{N\to\infty}\frac{1}{N}\left|\left\{i: Z(U^i|U^{1:i-1},Y^{1:N}) \leq 2^{-N^\beta}\right\}\right| = 1-H(Y|X),
\end{cases} \tag{4.3}$$

and

$$\begin{cases}
\lim_{N\to\infty}\frac{1}{N}\left|\left\{i: Z(U^i|U^{1:i-1},Y^{1:N}) \leq 2^{-N^\beta} \text{ and } Z(U^i|U^{1:i-1}) \geq 1-2^{-N^\beta}\right\}\right| = I(X;Y),\\
\lim_{N\to\infty}\frac{1}{N}\left|\left\{i: Z(U^i|U^{1:i-1},Y^{1:N}) \geq 2^{-N^\beta} \text{ or } Z(U^i|U^{1:i-1}) \leq 1-2^{-N^\beta}\right\}\right| = 1-I(X;Y).
\end{cases}$$

The Bhattacharyya parameter for BMA channels was firstly defined as the Bhattacharyya parameter of a source X given Y as its side information. The definition is for the distributed source coding problem in [7]. By the duality between channel coding and source coding, it can be also used to construct capacity achieving polar codes for BMA channels [8]. Actually, $Z(U^i|U^{1:i-1})$ is the Bhattacharyya parameter for a single source X (without side information). Consider the case that the output Y of W is a random variable which is independent of X, then $Z(U^i|U^{1:i-1,Y^{1:N}}) = Z(U^i|U^{1:i-1})$ and $H(X|Y) = H(X)$. Moreover, the calculation of Z can be converted to the calculation of the Bhattacharyya parameter $\tilde{Z}$ for a related binary-input memoryless symmetric (BMS) channel. Now we construct a BMS channel $\tilde{W}$ based on the BMS channel W. The following lemma is hidden in [8], we make it explicit.

Lemma 4.3 (From Asymmetric to Symmetric): Let $\tilde{W}$ be a binary input channel corresponding to the asymmetric channel W with input $\tilde{X} \in \mathcal{X} = \{0,1\}$ and output $\tilde{Y} \in \{\mathcal{Y}, \mathcal{X}\}$. The input of $\tilde{W}$ is uniformly distributed, i.e., $P_{\tilde{X}}(\tilde{x} = 0) = P_{\tilde{X}}(\tilde{x} = 1) = \frac{1}{2}$. The relationship between $\tilde{W}$ and W is shown in Figure4.2. Then $\tilde{W}$ is a binary symmetric channel in the sense that $P_{\tilde{Y}|\tilde{X}}(y, x \oplus \tilde{x}|\tilde{x}) = P_{Y,X}(y, x)$.

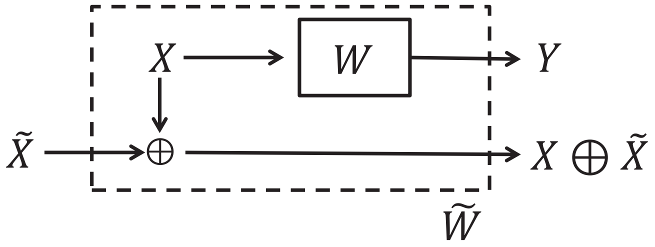


Figure 4.2: The relationship between $\tilde{W}$ and W.

Proof.

$$\begin{aligned}
P_{\tilde{Y}|\tilde{X}}(y, x \oplus \tilde{x}|\tilde{x}) &= \frac{P_{\tilde{Y},\tilde{X}}(y, x \oplus \tilde{x}, \tilde{x})}{P_{\tilde{X}}(\tilde{x})} = \frac{\sum_{x' \in X} P_{\tilde{Y},X,\tilde{X}}(y, x \oplus \tilde{x}, x', \tilde{x})}{P_{\tilde{X}}(\tilde{x})} \\
&\overset{(a)}{=} \frac{\sum_{x' \in X} P_{Y|X}(y|x') P_{X \oplus \tilde{X},X,\tilde{X}}(x \oplus \tilde{x}, x', \tilde{x})}{P_{\tilde{X}}(\tilde{x})} \\
&\overset{(b)}{=} \frac{\sum_{x' \in X} P_{Y|X}(y|x') P_{X \oplus \tilde{X}|X,\tilde{X}}(x \oplus \tilde{x}|x', \tilde{x}) P_X(x') P_{\tilde{X}}(\tilde{x})}{P_{\tilde{X}}(\tilde{x})} \\
&\overset{(c)}{=} P_{Y,X}(y, x).
\end{aligned}$$

The equalities (a)-(c) follow from (a) Y is only dependent on X, (b) X and $\tilde{X}$ are independent to each other and (c) $P_{X \oplus \tilde{X}|X,\tilde{X}}(x \oplus \tilde{x}|x', \tilde{x}) = \ (x' = x)$. □

The following definition of Bhattacharyya parameter for a BMS channel is from the seminal paper of polar codes [16]. This kind of Bhattacharyya parameter can be calculated recursively.

Definition 2 (Bhattacharyya Parameter for Symmetric Channel[16]): Let $\tilde{W}$ be a binary-input memoryless symmetric channel with transition probability $P_{Y|X}$, the Bhattacharyya parameter $\tilde{Z} \in [0, 1]$ is defined as

$$\tilde{Z}(\tilde{W}) \triangleq \sum_{y} \sqrt{P_{Y|X}(y|0) P_{Y|X}(y|1)}.$$

Note that Definition 1 and Definition 2 are the same when P_X is uniform.

The following theorem describes how to construct polar codes for a BMA channel W from a BMS channel $\tilde{W}$. Let $X^{1:N}$ and $Y^{1:N}$ be the input and output vectors of W. Let $\tilde{X}^{1:N}$ and $\tilde{Y}^{1:N} = \left(X^{1:N} \oplus \tilde{X}^{1:N}, Y^{1:N}\right)$ be the input and output vectors of $\tilde{W}$. Consider polarized random variables $U^{1:N}$=$X^{1:N}G_N$ and $\tilde{U}^{1:N}$=$\tilde{X}^{1:N}G_N$. Let W_N and $\tilde{W}_N$ denote the combining channel of N uses of W and $\tilde{W}$.

Theorem 4.2 (Construction of Polar Codes for BMA Channels[8]): The Bhattacharyya parameter for each subchannel of W_N is equal to that of each subchannel of $\tilde{W}_N$, i.e.,

$$Z(U^i|U^{1:i-1}, Y^{1:N}) = \tilde{Z}(\tilde{U}^i|\tilde{U}^{1:i-1}, X^{1:N} \oplus \tilde{X}^{1:N}, Y^{1:N}).$$

We formally define the frozen set $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{I}}$ of the symmetric polar codes as follows:

$$\begin{cases} \text{its frozen set: } \tilde{\mathcal{F}} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{1:N}) > 2^{-N^\beta}\} \\ \text{its information set: } \tilde{\mathcal{I}} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{1:N}) \leq 2^{-N^\beta}\}. \end{cases}$$

By Theorem 4.2, the Bhattacharyya parameters of the symmetric channel $\tilde{W}$ and the asymmetric channel W are the same. However, the channel capacity of $\tilde{W}$ is $I(\tilde{X}; X \oplus \tilde{X}) + I(\tilde{X}; Y|X \oplus \tilde{X}) = 1 - H(X) + I(X;Y)$, which is $1 - H(X)$ more than the capacity of W. This is because the choice of the input $\tilde{X}^{1:N}$ is more than that of $X^{1:N}$ ($\tilde{X}$ is uniform while X is selected according to P_X). If we fix the input of $\tilde{W}$ to be $X^{1:N}$, then we receive $(Y^{1:N}, 0^{1:N})$ as the output of $\tilde{W}_N$, where $0^{1:N}$ denotes the all zero vector. In this case, the mutual information becomes $I(X; Y, 0) = I(X; Y)$. Therefore, to obtain the real capacity $I(X;Y)$ of W, the input distribution of W needs to be adjusted to P_X. By the polar lossless source coding, the indices with very small $Z(U^i|U^{1:i-1})$ should be stripped off from the information set $\tilde{\mathcal{I}}$ of the symmetric channel, and the proportion of this part is $1 - H(X)$ as N goes to infinity. We name this set as the information set $\mathcal{I}$ for the symmetric channel W. And the remaining part $\mathcal{I}^c$ is the frozen set. According to [8], the bits within this frozen set can be determined by a certain mapping and the bits within the information set $\mathcal{I}$. We further find out that there are some bits which can be made independent to the information bits and uniformly distributed. The purpose of extracting such an independent frozen set is for the interest of our lattice construction which will be shown in Section 4.4. We name this part as the independent frozen set $\mathcal{F}$. In order to generate the input distribution P_X, the remaining frozen bits are determined by the bits in $\mathcal{F} \cup \mathcal{I}$. We name the set of all those deterministic bits as the shaping frozen set $\mathcal{S}$. The process is depicted in Figure 4.3. We formally define the above three sets

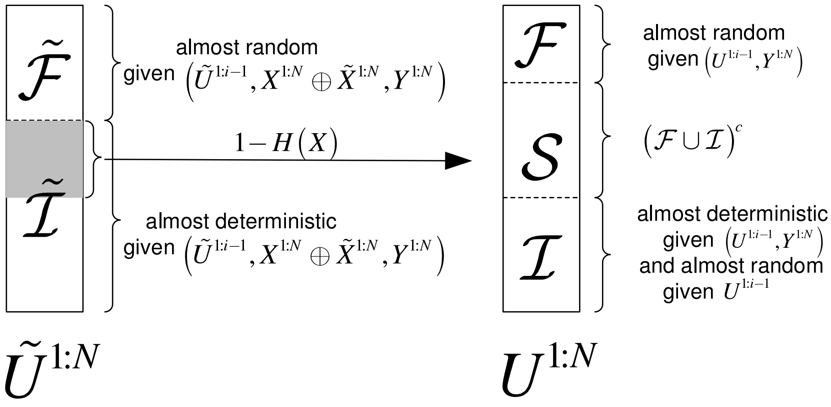

$\tilde{\mathcal{F}}$
$\tilde{\mathcal{I}}$
almost random given $(\tilde{U}^{1:i-1}, X^{1:N} \oplus \tilde{X}^{1:N}, Y^{1:N})$
$1 - H(X)$
almost deterministic given $(\tilde{U}^{1:i-1}, X^{1:N} \oplus \tilde{X}^{1:N}, Y^{1:N})$
$\tilde{U}^{1:N}$
$\mathcal{F}$
$\mathcal{S}$
$\mathcal{I}$
almost random given $(U^{1:i-1}, Y^{1:N})$
$(\mathcal{F} \cup \mathcal{I})^c$
almost deterministic given $(U^{1:i-1}, Y^{1:N})$ and almost random given $U^{1:i-1}$
$U^{1:N}$

Figure 4.3: Polarization for symmetric and asymmetric channels.

as follows:

$$\begin{cases} \text{its independent frozen set: } \mathcal{F} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \text{its information set: } \mathcal{I} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{1:N}) \leq 2^{-N^\beta} \text{ and } Z(U^i|U^{1:i-1}) \geq 1 - 2^{-N^\beta}\} \\ \text{its shaping frozen set: } \mathcal{S} = (\mathcal{F} \cup \mathcal{I})^c . \end{cases} \tag{4.4}$$

To find these three sets, one can use Theorem 4.2 to calculate $Z(U^i|U^{1:i-1}, Y^{1:N})$ using the known constructing techniques for symmetric polar codes [71, 63]. We note that $Z(U^i|U^{1:i-1})$ can be computed using a similar way. We construct a symmetric channel between $\tilde{X}$ and $X \oplus \tilde{X}$, which is actually a binary symmetric channel with cross probability $P_X(x = 1)$. This method has been used in lossless source coding [6]. We realize that the above operation is equivalent to implementing the shaping over the polar codes for the symmetric channel $\tilde{W}$. This is consistent with the concept that shaping can be dealt with as a source coding problem.

Besides the construction, the decoding process for the asymmetric polar codes can also be converted to the decoding for the symmetric polar codes. When $X^{1:N} \oplus \tilde{X}^{1:N} = 0$, we have $U^{1:N} = \tilde{U}^{1:N}$, which means the decoding results of $U^{1:N}$ equals to that of $\tilde{U}^{1:N}$. This explains why the decoding of polar codes on W can be treated as the decoding of polar codes on $\tilde{W}$ given $X \oplus \tilde{X} = 0$. We conclude this as the following lemma.

Lemma 4.4 (Decoding for Asymmetric Channel [8]): Let $y^{1:N}$ be a realization of $Y^{1:N}$ and $\hat{u}^{1:i-1}$ be the previous $i-1$ estimation of $u^{1:N}$. The ratio of the posterior probability of the u^i can be calculated as

$$\frac{P_{U^i|U^{1:i-1},Y^{1:N}}(0|\hat{u}^{1:i-1},y^{1:N})}{P_{U^i|U^{1:i-1},Y^{1:N}}(1|\hat{u}^{1:i-1},y^{1:N})}=\frac{\tilde{W}_N^{(i)}((y^{1:N},0^{1:N}),\hat{u}^{1:i-1}|0)}{\tilde{W}_N^{(i)}((y^{1:N},0^{1:N}),\hat{u}^{1:i-1}|1)}, \tag{4.5}$$

where $\tilde{W}_N^{(i)}$ is the transition probability of the i-th subchannel of $\tilde{W}_N$ and can be computed by the successive cancellation (SC) decoding algorithm with complexity $O(N\log N)$.

In [8], the bits in $\mathcal{F}\cup\mathcal{S}$ are all chosen according to $P(U^i|U^{1:i-1})$. However, in order to construct polar lattices, we change the scheme slightly by making the bits in $\mathcal{F}$ uniformly distributed from $\{0,1\}$ and the bits in $\mathcal{S}$ are still chosen according to $P(U^i|U^{1:i-1})$. The expectation of the decoding error probability still vanishes with N. This is an extension of the results from [8, Theorem 3]. We give the proof in Appendix C for completeness. Consider a polar code with the following encoding and decoding strategies for a BMA.

- Encoding: Before sending the codeword $x^{1:N}=u^{1:N}G_N$, the index set $[N]$ should be divided into three parts: the independent frozen set $\mathcal{F}$, the information set $\mathcal{I}$ and the shaping frozen set $\mathcal{S}$ which are defined in (4.4). The encoder first places the uniformly distributed information bits in $\mathcal{I}$. We fill $\mathcal{F}$ with a uniformly distributed sequence from $\{0,1\}$ which are shared between the encoder and the decoder. The bits in $\mathcal{S}$ are generated according to the family of randomized mapping $\Phi_{\mathcal{S}}$ as follows:

$$u^i=\begin{cases}0 & \text{with probability } P_{U^i|U^{1:i-1}}(0|u^{1:i-1}),\\ 1 & \text{with probability } P_{U^i|U^{1:i-1}}(1|u^{1:i-1}).\end{cases}$$

- Decoding: The decoder receives $y^{1:N}$ and estimates $\hat{u}^{1:N}$ of $u^{1:N}$ according to the rule

$$\hat{u}^i = \begin{cases} u^i, & \text{if } i \in \mathcal{F} \\ \phi_i(\hat{u}^{1:i-1}), & \text{if } i \in \mathcal{S} \\ \underset{u}{\operatorname{argmax}}\ P_{U^i|U^{1:i-1},Y^{1:N}}(u|\hat{u}^{1:i-1}, y^{1:N}), & \text{if } i \in \mathcal{I} \end{cases}.$$

where $\phi_{\mathcal{S}} \triangleq \{\phi_i\}_{i\in\mathcal{S}}$ and $\phi_{\mathcal{S}} \in \Phi_{\mathcal{S}}$.

Theorem 4.3: With the above encoding and decoding, the message rate can be arbitrarily close to $I(Y;X)$ and the expectation of the decoding error probability over the randomized mappings satisfies $E_{\Phi_{\mathcal{S}}}[P_e(\Phi_{\mathcal{S}})] = O(2^{-N^{\beta'}})$ for any $\beta' < \beta < 0.5$. Consequently, there exists a deterministic mapping $\phi_{\mathcal{S}}$ such that $P_e(\phi_{\mathcal{S}}) = O(2^{-N^{\beta'}})$.

Practically, to share the mapping $\phi_{\mathcal{S}}$ between the encoder and the decoder, we can let them have access to the same source of randomness, which can be achieved by forcing the pseudorandom number generators at both sides to be in the same state.

4.3 Multilevel asymmetric polar codes

As we have mentioned in Section 4.1, if we use $D_{\eta\Lambda,\sigma_s}$ as the constellation, the ℓ-th channel of the multilevel system is generally asymmetric and its channel transition PDF is shown by (4.2). Our task is to construct asymmetric polar codes for each level in order to achieve its mutual information $I(Y;X_\ell|X_{1:\ell-1})$. The construction of the polar code for the first level is already given in Section 4.2. We take the channel of the second level W_2 as an example to demonstrate our construction. This channel is also a BMA with input $X_2 \in \mathcal{X} = \{0,1\}$, output $Y \in \mathcal{Y}$ and side information X_1 at the transmitter. To construct explicit asymmetric polar codes we propose the following two-step algorithm.

- At the first step, construct a polar code for a BMS with the input vector $\tilde{X}_2^{1:N} = [\tilde{X}_2^1, \tilde{X}_2^2, \cdots, \tilde{X}_2^N]$ and the output vector $\tilde{Y}^{1:N} = \left(X_2^{1:N} \oplus \tilde{X}_2^{1:N}, Y^{1:N}, X_1^{1:N}\right)$

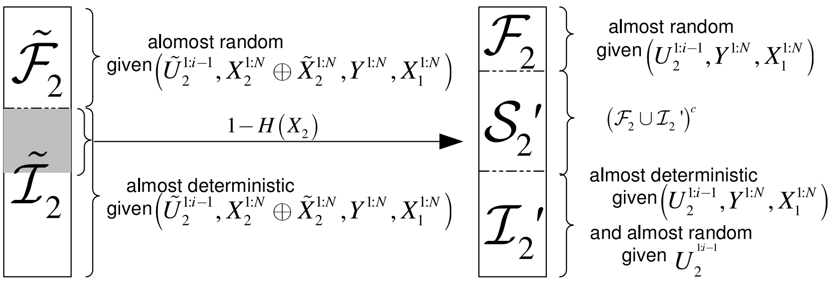

$\tilde{\mathcal{F}}_2$
$\tilde{\mathcal{I}}_2$
alomost random
given$(\tilde{U}_2^{1:i-1}, X_2^{1:N} \oplus \tilde{X}_2^{1:N}, Y^{1:N}, X_1^{1:N})$
$1-H(X_2)$
almost deterministic
given$(\tilde{U}_2^{1:i-1}, X_2^{1:N} \oplus \tilde{X}_2^{1:N}, Y^{1:N}, X_1^{1:N})$
$\mathcal{F}_2$
$\mathcal{S}_2{}'$
$\mathcal{I}_2{}'$
almost random
given$(U_2^{1:i-1}, Y^{1:N}, X_1^{1:N})$
$(\mathcal{F}_2 \cup \mathcal{I}_2{}')^c$
almost deterministic
given$(U_2^{1:i-1}, Y^{1:N}, X_1^{1:N})$
and almost random
given $U_2^{1:i-1}$

Figure 4.4: The first step of polarization.

where $\tilde{X}_2^i \in \mathcal{X} = \{0,1\}$ is uniformly distributed. At this step X_1 is regarded as the output. Then the distribution of the input X_2 becomes the marginal distribution $\sum_{x_1,x_{3:r}} P_{X_{1:r}}(x_{1:r})$. Consider polarized random variables $U_2^{1:N} = X_2^{1:N} G_N$ and $\tilde{U}_2^{1:N} = \tilde{X}_2^{1:N} G_N$. Then according to Theorem 4.1, the polarization gives us the three sets $\mathcal{F}_2$, $\mathcal{I}_2'$ and $\mathcal{S}_2'$ as shown in Figure 4.4. Similarly, we can prove that $\frac{|\mathcal{I}_2'|}{N} \to I(Y, X_1; X_2)$ and $\frac{|\mathcal{F}_2 \cup \mathcal{S}_2'|}{N} \to 1 - I(Y, X_1; X_2)$ as N goes to infinity. The definitions of these three sets are as follows:

$$\begin{cases} \text{its independent frozen set: } \mathcal{F}_2 = \{i \in [N] : Z(U_2^i | U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \text{its information set: } \mathcal{I}_2' = \{i \in [N] : Z(U_2^i | U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}) \leq 2^{-N^\beta} \text{ and } Z(U_2^i | U_2^{1:i-1}) \geq 1 - 2^{-N^\beta}\} \\ \text{its shaping frozen set: } \mathcal{S}_2' = \left(\mathcal{F} \cup \mathcal{I}_2'\right)^c. \end{cases} \tag{4.6}$$

More explicitly,

$$\begin{cases} \lim_{N\to\infty} \dfrac{|\mathcal{I}_2'|}{N} = I(Y, X_1; X_2), \\ \lim_{N\to\infty} \dfrac{|\mathcal{F}_2 \cup \mathcal{S}_2'|}{N} = 1 - I(Y, X_1; X_2). \end{cases}$$

- At the second step, we consider $X_1^{1:N}$ as the side information for the encoder. Given $X_1^{1:N}$, the choices of $X_2^{1:N}$ should be further restriced since X_1 and X_2 are generally correlated. For example, $P_{X_1,X_2}(x_1, x_2) = f_{\sigma_s}(\mathcal{A}(x1, x2))/f_{\sigma_s}(\eta\Lambda)$ from Figure 4.1. X_1 and X_2 will be independent only if $\epsilon_{\mathcal{A}(x1,x2)}(\sigma_s) = 0$. By stripping off the bits which are almost deterministic given $U_2^{1:i-1}$ and $X_1^{1:N}$ from $\mathcal{I}_2'$, we obtain the information set $\mathcal{I}_2$ for W_2. Then the distribution of the input X_2 becomes the conditional distribution $P_{X_2|X_1}(x_2|x_1)$. The process is shown in Figure 4.5. More explicitly, consider the proportions of the indices divided as

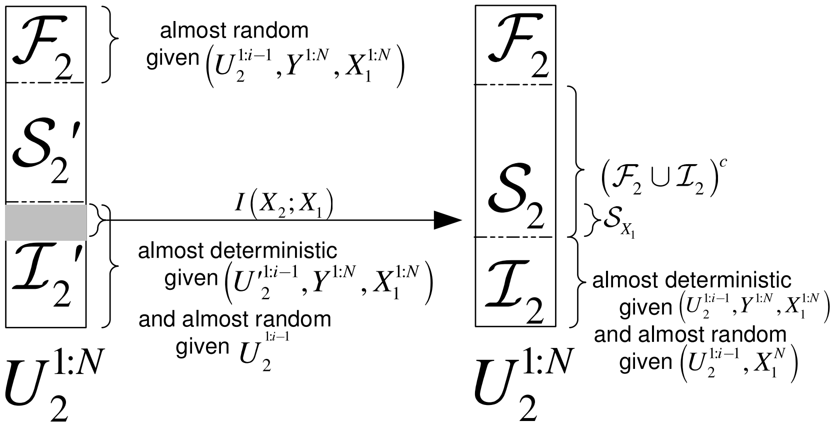

$\mathcal{F}_2$
almost random given $\left(U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}\right)$
$\mathcal{S}_2{}'$
$I(X_2;X_1)$
$\mathcal{I}_2{}'$
almost deterministic given $\left(U'^{1:i-1}_2, Y^{1:N}, X_1^{1:N}\right)$
and almost random given $U_2^{1:i-1}$
$U_2^{1:N}$
$\mathcal{F}_2$
$\mathcal{S}_2$
$\left(\mathcal{F}_2 \cup \mathcal{I}_2\right)^c$
$\mathcal{S}_{X_1}$
$\mathcal{I}_2$
almost deterministic given $\left(U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}\right)$
and almost random given $\left(U_2^{1:i-1}, X_1^{N}\right)$
$U_2^{1:N}$

Figure 4.5: The second step of polarization.

followings:

$$\begin{aligned}
1 &= \underbrace{1 - I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y)}_{\mathcal{F}_2} + I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y) \\
&\overset{\text{Step1}}{=} \underbrace{1 - I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y)}_{\mathcal{F}_2} + \underbrace{I(\tilde{X}_2; \tilde{X}_2 \oplus X_2)}_{\mathcal{S}'_2} + \underbrace{I(\tilde{X}_2; X_1, Y|\tilde{X}_2 \oplus X_2)}_{\mathcal{I}'_2} \\
&\overset{\text{Step2}}{=} \underbrace{1 - I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y)}_{\mathcal{F}_2} + \underbrace{I(\tilde{X}_2; \tilde{X}_2 \oplus X_2)}_{\mathcal{S}'_2} + \underbrace{I(\tilde{X}_2; X_1|\tilde{X}_2 \oplus X_2)}_{\mathcal{S}_{X_1}} + \underbrace{I(\tilde{X}_2; Y|X_1, \tilde{X}_2 \oplus X_2)}_{\mathcal{I}_2} \\
&= \underbrace{1 - I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y)}_{\mathcal{F}_2} + \underbrace{1 - H(X_2)}_{\mathcal{S}'_2} + \underbrace{I(X_2; X_1)}_{\mathcal{S}_{X_1}} + \underbrace{I(X_2; Y|X_1)}_{\mathcal{I}_2} \\
&= \underbrace{1 - I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y)}_{\mathcal{F}_2} + \underbrace{1 - H(X_2|X_1)}_{\mathcal{S}_2} + \underbrace{I(X_2; Y|X_1)}_{\mathcal{I}_2}
\end{aligned}$$

Remark 4.1: This also gives a method of lossless source coding for discreet source with arbitrary alphabet size.

We give the formal statement in the following lemma.

Lemma 4.5: After the first step of polarization, we obtain the three sets $\mathcal{F}_2$, $\mathcal{I}'_2$ and $\mathcal{S}'_2$ according to (4.6). Let the set $\mathcal{S}_{X_1}$ denote the indices whose Bhattacharyya parameters satisfy $Z(U_2^i|U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}) \leq 2^{-N^\beta}$ and $Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}) \leq 1 - 2^{-N^\beta}$ and $Z(U_2^i|U_2^{1:i-1}) \geq 1 - 2^{-N^\beta}$. The proportion of $\mathcal{S}_{X_1}$ satisfies that $\lim_{N\to\infty} \frac{|\mathcal{S}_{X_1}|}{N} = I(X_2; X_1)$. Then by stripping $\mathcal{S}_{X_1}$ from $\mathcal{I}'_2$, we obtain the true information set $\mathcal{I}_2$ for W_2. Therefore the three sets are obtained as follows:

$$\begin{cases}
\text{its independent frozen set: } \mathcal{F}_2 = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}) \geq 1 - 2^{-N^\beta}\} \\
\text{its information set: } \mathcal{I}_2 = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}) \leq 2^{-N^\beta} \text{ and } Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}) \geq 1 - 2^{-N^\beta}\} \\
\text{its shaping frozen set: } \mathcal{S}_2 = (\mathcal{F} \cup \mathcal{I}_2)^c\,.
\end{cases}$$

Proof. Firstly we show the proportion of set $\mathcal{S}_{X_1}$ goes to $I(X_1; X_2)$ when the block length N is sufficiently large. Here we define a set which is slightly different from $\mathcal{S}_{X_1}$ as $\mathcal{S}'_{X_1} = \{i \in [N] : Z(U_2^i | U_2^{1:i-1}, X_1^{1:N}) \leq 2^{-N^\beta}$ and $Z(U_2^i | U_2^{1:i-1}) \geq 1 - 2^{-N^\beta}\}$. Consider we are constructing asymmetric polar codes over the channel from X_1 to X_2, it is not difficult to find that $\lim_{N\to\infty} \frac{|\mathcal{S}'_{X_1}|}{N} = I(X_2; X_1)$ by Theorem 4.3. Furthermore, by Lemma 4.2, if $Z(U_2^i | U_2^{1:i-1}, X_1^{1:N}) \leq 2^{-N^\beta}$, we can immediately have $Z(U_2^i | U_2^{1:i-1}, X_1^{1:N}, Y^{1:N}) \leq 2^{-N^\beta}$. Therefore, the difference between the definitions of $\mathcal{S}_{X_1}$ and $\mathcal{S}'_{X_1}$ is the part of $Z(U_2^i | U_2^{1:i-1}, X_1^{1:N})$. Let $\bar{\mathcal{P}}_{X_1}$ denote the unpolarized set with $2^{-N^\beta} \leq Z(U_2^i | U_2^{1:i-1}, X_1^{1:N}) \leq 1 - 2^{-N^\beta}$, we have

$$\lim_{N\to\infty} \frac{|\mathcal{S}_{X_1}|}{N} - \frac{|\mathcal{S}'_{X_1}|}{N} \leq \lim_{N\to\infty} \frac{|\bar{\mathcal{P}}_{X_1}|}{N} = 0.$$

As a result, $\lim_{N\to\infty} \frac{|\mathcal{S}_{X_1}|}{N} = \lim_{N\to\infty} \frac{|\mathcal{S}'_{X_1}|}{N} = I(X_2; X_1)$.

Now we prove that $\mathcal{S}_{X_1} \cup \mathcal{I}_2 = \mathcal{I}'_2$. Again, by Lemma 4.2, if $Z(U_2^i | U_2^{1:i-1}, X_1^{1:N}) \geq 1 - 2^{-N^\beta}$, we get $Z(U_2^i | U_2^{1:i-1}) \geq 1 - 2^{-N^\beta}$ and the difference between the definitions of $\mathcal{S}_{X_1}$ and $\mathcal{I}'_2$ only lays on the part of $Z(U_2^i | U_2^{1:i-1}, X_1^{1:N})$. Observe that the union set $\mathcal{S}_{X_1} \cup \mathcal{I}_2$ would remove the condition on $Z(U_2^i | U_2^{1:i-1}, X_1^{1:N})$, and therefore $\mathcal{S}_{X_1} \cup \mathcal{I}_2 = \mathcal{I}'_2$. It can be also found that the proportion of $\mathcal{I}_2$ goes to $I(X_2; Y | X_1)$ as N goes to infinity. □

We summarize our main results in the following theorem. The proof is in Appendix D. Consider a polar code with the following encoding and decoding strategies for the channel of the second level W_2 with the channel transition PDF $P_{Y|X_2,X_1}(y|x_2, x_1)$ shown in (4.2).

- Encoding: Before sending the codeword $x_2^{1:N} = u_2^{1:N} G_N$, the index set $[N]$ should be divided into three parts: the independent frozen set $\mathcal{F}_2$, the information set $\mathcal{I}_2$, and the shaping frozen set $\mathcal{S}_2$. The encoder first places the uniformly distributed information bits in $\mathcal{I}_2$. Then the independent frozen set $\mathcal{F}_2$ is filled with a uniformly distributed sequence which are shared between the encoder and the decoder. The bits in $\mathcal{S}_2$ are generated according to the family of randomized mapping $\Phi_{\mathcal{S}_2}$ as follows:

$$u_2^i = \begin{cases} 0 & \text{with probability } P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(0|u_2^{1:i-1}, x_1^{1:N}), \\ 1 & \text{with probability } P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(1|u_2^{1:i-1}, x_1^{1:N}). \end{cases} \tag{4.7}$$

- Decoding: The decoder receives $y^{1:N}$ and estimates $\hat{u}_2^{1:N}$ based on the previously recovered $x_1^{1:N}$ according to the rule

$$\hat{u}_2^i = \begin{cases} u_2^i, & \text{if } i \in \mathcal{F}_2 \\ \phi_i(\hat{u}_2^{1:i-1}), & \text{if } i \in \mathcal{S}_2 \\ \operatorname*{argmax}_u P_{U_2^i|U_2^{1:i-1},X_1^{1:N},Y^{1:N}}(u|\hat{u}_2^{1:i-1}, x_1^{1:N}, y^{1:N}), & \text{if } i \in \mathcal{I}_2 \end{cases} .$$

where $\phi_{\mathcal{S}_2} \triangleq \{\phi_i\}_{i \in \mathcal{S}_2}$ and $\phi_{\mathcal{S}_2} \in \Phi_{\mathcal{S}_2}$.

Theorem 4.4 (Coding Theorem for Multilevel Asymmetric Polar Codes): With the above encoding and decoding, the message rate can be arbitrarily close to $I(Y; X_2|X_1)$ and the expectation of the decoding error probability over the randomized mappings satisfies $E_{\Phi_{\mathcal{S}_2}}[P_e(\Phi_{\mathcal{S}_2})] = O(2^{-N^{\beta'}})$ for any $\beta' < \beta < 0.5$. Consequently, there exists a deterministic mapping $\phi_{\mathcal{S}_2}$ such that $P_e(\phi_{\mathcal{S}_2}) = O(2^{-N^{\beta'}})$.

We note that Theorem 4.4 can be generalized to the construction of asymmetric polar codes for the channel of the ℓ-th level W_ℓ. The only difference is that the side information changes from $X_1^{1:N}$ to $X_{1:\ell-1}^{1:N}$. As a result, we can construct an asymmetric polar code which achieves a rate arbitrarily close to $I(Y; X_\ell|X_{1:\ell})$ with vanishing error probability. We omit the proof for the sake of brevity.

4.4 Polar lattices with lattice Gaussian shaping can achieve the capacity

In this section, we first show that polar lattice codes with the lattice Gaussian distribution can achieve the AWGN channel capacity $\frac{1}{2}\log(1+\mathsf{SNR})$. Then we further demonstrate that constructing asymmetric multilevel polar codes is equivalent to implementing the shaping over an AWGN-good polar lattice. This is consistent with the theory proved in [4].

As shown in [4], the mutual information between the discrete Gaussian lattice distribution and the output of the AWGN channel approaches $\frac{1}{2}\log(1+\mathsf{SNR})$ as the flatness factor goes to 0. Assume the distribution of X is $D_{\eta\mathbb{Z},\sigma_s}$ where η is the scale factor. Therefore by applying polar codes over the asymmetric channels for each level, if $\epsilon_{\eta\mathbb{Z}}\left(\tilde{\sigma}\right)\to 0$ where $\tilde{\sigma}=\frac{\sigma_s\sigma}{\sqrt{\sigma_s^2+\sigma^2}}$, $r=O(\log N)$ and N goes infinity, the total message rate of such polar lattice code with shaping can be arbitrarily close to the channel capacity $\frac{1}{2}\log(1+\mathsf{SNR})$.

$$\begin{aligned}\lim_{\epsilon_{\eta\mathbb{Z}}(\tilde{\sigma})\to 0} I(Y;X) &= \lim_{\epsilon_{\eta\mathbb{Z}}(\tilde{\sigma})\to 0, r\to\infty} I(Y;X_{1:r}) \\ &= \lim_{\epsilon_{\eta\mathbb{Z}}(\tilde{\sigma})\to 0, r\to\infty} I(Y;X_1)+I(Y;X_2|X_1)+\cdots+I(Y;X_r|X_{1:r-1}) \\ &= \frac{1}{2}\log(1+\mathsf{SNR}).\end{aligned}$$

We use multistage SC decoding and the LR of each level can be computed according to Lemma 4.4. The block error probability of each level can be guaranteed to be exponentially vanished by Theorem 4.4. Let $P_e(\mathcal{C}_\ell,\sigma^2)$ denote the block error probability of the polar code for the ℓ-th level $(1\leq\ell\leq r-1)$. Then the block error probability of the asymmetric multilevel polar code can be bounded as

$$\begin{aligned} P_e(L,\sigma^2) &\leq \sum_{\ell=1}^{r} P_e(\mathcal{C}_\ell,\sigma^2) \\ &\leq \sum_{\ell=1}^{r} 2^{-N^{\beta_\ell}}, \end{aligned}$$

where $0<\beta_\ell<\beta<0.5$ for $1\leq\ell\leq r$.

In conclusion, we have the following theorem:

Theorem 4.5: Consider the above multilevel asymmetric polar code, where $r=O(\log N)$. In the limit as $\epsilon_{\eta\mathbb{Z}}(\tilde{\sigma})\to 0$, $N\to\infty$, with the transmitting rate up to $\frac{1}{2}\log(1+\mathsf{SNR})$, the error probability of this multilevel code under multi-stage SC decoding is bounded by

$$P_e(L,\sigma^2)<\sum_{\ell=1}^{r} 2^{-N^{\beta_\ell}},$$

where $0<\beta_\ell<\beta<0.5$. In other words, the above multilevel asymmetric polar code can achieve the full AWGN channel capacity.

Next, we explain that this asymmetric multilevel polar coding scheme is equivalent to implementing Gaussian shaping over a coset of an AWGN-good polar lattice $L+c$. First we need to find the AWGN-good polar lattice. As discussed in the previous section, the shaping over polar codes for symmetric channels is implemented by generating the bits in the source coding set $\mathcal{S}$ randomly according to the probability $P_{U_\ell^i|U_\ell^{1:i-1},X_{1:\ell-1}^{1:N}}$. Therefore the AWGN-good lattice L is constructed by polar codes for all the corresponding symmetric channels. We note here that the frozen bits of the polar codes for symmetric channels must be set to all-zeros in order to be obtain a polar lattice.

The following lemma shows the connection between multilevel codes and lattices which can simplify our polar codes construction for the symmetric channels. Recall the ℓ-th channel is a BMA with the input distribution $P(X_\ell|X_{1:\ell-1})$ ($\ell \leq r$). It is clear that $P_{X_\ell}(x_{1:\ell}) = f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))/f_{\sigma_s}(\eta\Lambda)$. According to Lemma 4.3 and (4.2), the channel transition probability of the ℓ-th corresponding symmetric channel $\tilde{W}_\ell$ is

$$\begin{aligned}
&P_{\tilde{W}_\ell}(y, x_{1:\ell-1}, x_\ell \oplus \tilde{x}_\ell|\tilde{x}_\ell) = P_{Y,X_{1:\ell}}(y, x_{1:\ell})\\
&\quad = P_{X_{1:\ell}}(x_{1:\ell})P_{Y|X_\ell,X_{1:\ell-1}}(y|x_\ell, x_{1:\ell-1})\\
&\quad = \exp\left(-\frac{y^2}{2(\sigma_s^2+\sigma^2)}\right)\frac{1}{2\pi\sigma\sigma_s}\frac{1}{f_{\sigma_s}(\eta\Lambda)}\sum_{a\in\mathcal{A}_\ell(x_{1:\ell})}\exp\left(-\frac{1}{2\tilde{\sigma}^2}\left(|\alpha y-a|^2\right)\right).
\end{aligned}$$

Therefore, for example we use $D_{\mathbb{Z},\sigma_s}$ as the constellation, we can conclude that the channel likelihood ratio (LR) of ℓ-th symmetric channel $\tilde{W}_\ell$ is in the same form as that of the $2^{\ell-1}\mathbb{Z}/2^\ell\mathbb{Z}$ channel shown in (2.4). The only difference is an MMSE scaling factor α on y and σ^2. We note here that the multistage SC decoding at the receiving end is actually performed on the MMSE scaled received signal αy (See Lemma 4.4). We summarize the foregoing analysis in the following lemma:

Lemma 4.6: Consider a multilevel lattice code with the constellation $D_{\mathbb{Z},\sigma_s}$. Constructing a polar code for the ℓ-th symmetric channel $\tilde{W}_\ell$ transformed from the asymmetric channel W_ℓ is equivalent to constructing a polar code for the MMSE scaled $2^{\ell-1}\mathbb{Z}/2^\ell\mathbb{Z}$ channel defined in the lattice literature [39].

Proof. The proof is straightforward by applying the definitions of Bhattacharyya parameters or mutual information of $\tilde{W}_\ell$ and MMSE-scaled $2^{\ell-1}\mathbb{Z}/2^\ell\mathbb{Z}$ channel. As a result, the Bhattacharyya parameters and mutual information of the polarized subchannels are equal for these two channels.

For the sake of simplicity, we only give the proof of the case with uniform inputs. To see this, it suffices to show that the mutual information and Bhattacharyya parameters of the resultant bit-channels which are polarized from $W(\Lambda_1/\Lambda_2 = \mathbb{Z}/2\mathbb{Z}, \sigma^2)$ and $W'(Y; X_1)$ with input X_1 and output Y created by (4.1) are the same. Let $Q(y|x)$ be a BMS channel with binary input alphabet $\mathcal{X} \in \{0, 1\}$ and output alphabet $\mathcal{Y} \in \mathbb{R}$. Consider a random vector U^2 that is uniformly distributed over $\mathcal{X}^2$. Let $X^2 = U^2 \cdot \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$ be the input to two independent copies of the channel Q and let Y^2 be the corresponding outputs. After the channel combining and splitting, the resultant bit-channels [16] are defined as

$$\begin{aligned} Q_2^{(1)}(y^2|u_1) &= \frac{1}{2}\sum_{u_2} Q(y_1|u_1 \oplus u_2)Q(y_2|u_2), \\ Q_2^{(2)}(y^2, u_2|u_1) &= \frac{1}{2}Q(z_1|u_1 \oplus u_2)Q(y_2|u_2). \end{aligned}$$

Then we apply this polarization transformation to $W(\Lambda_1/\Lambda_2, \sigma^2)$ and $W'(Y; X_1)$, respectively. After some manipulations, we get

$$\begin{aligned} W_2^{(1)}(y^2|0) &= \frac{1}{2}(f_{\sigma,\Lambda_2}(y_1)f_{\sigma,\Lambda_2}(y_2) + f_{\sigma,\Lambda_2}(y_1-1)f_{\sigma,\Lambda_2}(y_2-1)), \\ W_2^{(1)}(y^2|1) &= \frac{1}{2}(f_{\sigma,\Lambda_2}(y_1-1)f_{\sigma,\Lambda_2}(y_2) + f_{\sigma,\Lambda_2}(y_1)f_{\sigma,\Lambda_2}(y_2-1)), \end{aligned}$$

and

$$\begin{aligned} W_2'^{(1)}(y^2|0) &= \frac{1}{2|\Lambda_2/\Lambda_r|^2}\Big(\sum_{x\in\Lambda_2/\Lambda_r} f_{\sigma,\Lambda_2}(y_1-x)f_{\sigma,\Lambda_2}(y_2-x) + \sum_{x\in\Lambda_2/\Lambda_r} f_{\sigma,\Lambda_2}(y_1-x-1)f_{\sigma,\Lambda_2}(y_2-x-1)\Big), \\ W_2'^{(1)}(y^2|1) &= \frac{1}{2|\Lambda_2/\Lambda_r|^2}\Big(\sum_{x\in\Lambda_2/\Lambda_r} f_{\sigma,\Lambda_2}(y_1-x-1)f_{\sigma,\Lambda_2}(y_2-x) + \sum_{x\in\Lambda_2/\Lambda_r} f_{\sigma,\Lambda_2}(y_1-x)f_{\sigma,\Lambda_2}(y_2-x-1)\Big). \end{aligned}$$

By the definitions of the mutual information and the Bhattacharyya parameter of a BMS channel [16]

$$\begin{cases} I(Q) \triangleq \int \sum_x \frac{1}{2}Q(y|x)\log\frac{Q(y|x)}{\frac{1}{2}Q(y|0)+\frac{1}{2}Q(y|1)}dy \\ Z(Q) \triangleq \int \sqrt{Q(y|0)Q(y|1)}dy \end{cases},$$

we have

$$\begin{cases} I(W_2^{(1)}(y^2|x_1)) = I(W_2'^{(1)}((y^2|x_1))) \\ Z(W_2^{(1)}(y^2|x_1)) = Z(W_2'^{(1)}(y^2|x_1))) \end{cases}.$$

And it is not difficult to verify that

$$\begin{cases} I(W_2^{(1)}(y^2, x_2|x_1)) = I(W_2'^{(1)}((y^2, x_2|x_1))) \\ Z(W_2^{(1)}(y^2, x_2|x_1)) = Z(W_2'^{(1)}(y^2, x_2|x_1))) \end{cases}.$$

Since the construction of polar codes are based on either the mutual information or the Bhattacharyya parameter of the bit-channels, polar codes constructed for $W(\Lambda_1/\Lambda_2, \sigma^2)$ and $W'(Y; X_1)$ are the same. The validation of the equivalence between the ℓ-th channel $W(\Lambda_\ell/\Lambda_{\ell+1}, \sigma^2)$ and $W'(Y; X_\ell|X_1, \cdots, X_{\ell-1})$ is similar. □

Remark 4.2: This lemma unifies the multilevel coding theory and lattice coding theory which were hidden in [39, 67]. One can expect the equivalence in a more general sense than the construction of polar codes. The proof may be based on the equivalence between coset decoding [12] and maximum likelihood (ML) decoding of the fine lattice Λ_ℓ in the presence of the $\Lambda_{\ell+1}$-aliased Gaussian noise.

The following lemma shows that these polar codes for all the corresponding symmetric channels are nested which is an important requirement to construct lattices [39].

Lemma 4.7: Let $\tilde{W}_\ell$ and $\tilde{W}_{\ell+1}$ denote the corresponding symmetric channels which are transformed from the ℓ-th and the $(\ell+1)$-th asymmetric channel for $1 \leq \ell \leq r$. $\tilde{W}_\ell$ is degraded with respect to $\tilde{W}_{\ell+1}$ and polar codes constructed for $\tilde{W}_\ell$ and $\tilde{W}_{\ell+1}$ are nested.

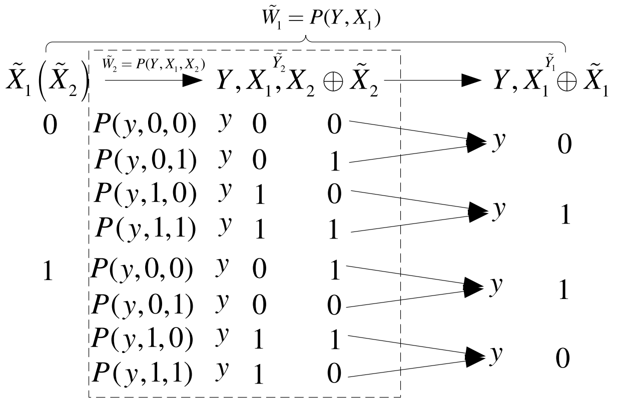


Figure 4.6: The relation between $\tilde{W}_1$ and $\tilde{W}_2$.

Proof. From Theorem 4.2, the channel transition probabilities of $\tilde{W}_\ell$ and $\tilde{W}_{\ell+1}$ are

$$\begin{cases} P_{\tilde{W}_\ell} = P_{\tilde{Y}_\ell|\tilde{X}_\ell}(y, x_{1:\ell-1}, x_\ell \oplus \tilde{x}_\ell|\tilde{x}_\ell) = P_{Y,X_{1:\ell}}(y, x_{1:\ell}) \\ P_{\tilde{W}_{\ell+1}} = P_{\tilde{Y}_{\ell+1}|\tilde{X}_{\ell+1}}(y, x_{1:\ell}, x_{\ell+1} \oplus \tilde{x}_{\ell+1}|\tilde{x}_{\ell+1}) = P_{Y,X_{1:\ell+1}}(y, x_{1:\ell+1}) \end{cases}. \quad (4.8)$$

By the definition of the degradation [61, Definition 1.7], we need to show that there always exists a channel that can transform $\tilde{W}_{\ell+1}$ to $\tilde{W}_\ell$. For simplicity, we use $\tilde{W}_1$ and $\tilde{W}_2$ as an illustrative example. The relation between $\tilde{W}_1$ and $\tilde{W}_2$ is depicted in Figure 4.6. It is clear that the mapping between $\tilde{Y}_2$ and $\tilde{Y}_1$ is independent of the input. According to the definition of the degraded channel, $\tilde{W}_1$ is degraded with respect to $\tilde{W}_2$. This degradation can also be proved by using the equivalence lemma Lemma 4.6 since it is proved that the $2^{\ell-1}\mathbb{Z}/2^\ell\mathbb{Z}$ channel is degraded with respect to the $2^\ell\mathbb{Z}/2^{\ell+1}\mathbb{Z}$ channel in Lemma 3.3.

Since $\tilde{W}_\ell$ is degraded with respect to $\tilde{W}_{\ell+1}$ we have $\tilde{Z}(\tilde{W}_{\ell,N}^{(i)}) \geq \tilde{Z}(W_{\ell+1,N}^{(i)})$, where $\tilde{W}_{\ell,N}^{(i)}$ and $\tilde{W}_{\ell+1,N}^{(i)}$ denote the i-th subchannel at ℓ-th and $(\ell+1)$-th level. Then $\mathcal{F}_\ell \supseteq \mathcal{F}_{\ell+1}$. By Theorem 4.3 and 4.4, these sets can be filled with uniformly random bits. Then we generate a uniformly distributed binary sequence with size $|\mathcal{F}_1|$. We fill $\mathcal{F}_\ell$ with the first $|\mathcal{F}_\ell|$ bits of the sequence and fill $\mathcal{F}_{\ell+1}$ with the first $|\mathcal{F}_{\ell+1}|$ bits of the sequence. Therefore the uniformly distribution requirement for the sets $\mathcal{F}_\ell$ and $\mathcal{F}_{\ell+1}$ can be guaranteed and $\mathcal{F}_\ell \supseteq \mathcal{F}_{\ell+1}$. Recall that the original definition of frozen set for symmetric polar codes is the bits which satisfy $Z(U_\ell^i|U_\ell^{1:i-1}, Y^{1:N}, X_{1:\ell-1}^{1:N}) \geq 2^{-N^\beta}$ [16]. If the above original frozen set is a all-zero vector, then this multilevel polar code is a polar lattice denoted by L. Otherwise, this multilevel polar code is a coset of the polar lattice $L+\chi$ where χ is a bitwise addition of the bits in the original frozen sets of all levels. Therefore it is clear that $2^{r-1}\mathbb{Z}^N \subseteq L+\chi \subseteq \mathbb{Z}^N$. □

From Lemma 4.6 we know that the polar lattice L is equivalent to the multilevel construction of lattices with one dimensional lattice partition $\mathbb{Z}/2\mathbb{Z} \cdot \cdot\cdot$ and the MMSE scaling factor. Such polar lattices are AWGN-good lattices corresponding to the Gaussian noise variance $\tilde{\sigma}^2$. As pointed out in the previous section, constructing the polar code for an asymmetric channel is equivalent to implementing the shaping over the codewords of the polar code for the corresponding symmetric channel. If we do not share the independent frozen set $\mathcal{F}_\ell$ in each level before each communication, then the above shaping scheme implements a lattice Gaussian distribution over $\mathbb{Z}^N$ for the constellation $D_{\mathbb{Z},\sigma_s}$. Since we need to share the frozen set during a communication process (we have already proved in Theorem 4.3 and 4.4 that there exists at least one frozen set which are good for communication.), the above shaping scheme implements a lattice Gaussian distribution over a coset of the AWGN-good lattice $L+\chi$ which is because the sublattice of a lattice Gaussian is still a lattice Gaussian. We can conclude now that by constructing multilevel asymmetric polar codes with the constellation $D_{\mathbb{Z},\sigma_s}$ is equivalent to implementing a lattice Gaussian distribution $D_{L+\chi,\sigma_s}$ where L is an AWGN-good lattice constructed from Construction D corresponding to the Gaussian noise variance $\tilde{\sigma}^2$.

According to [4, Lemma 1], the average power of $D_{L+\chi,\sigma_s}$ P_s will never be greater than σ_s^2 regardless of the shift vector χ. Then we don't need the flatness factor condition on L that $\epsilon_L(\sigma_s) \to 0$ in [4]. Therefore our coding scheme can achieve the AWGN channel capacity for any SNR. We note here that our coding scheme is not only a practical implementation of [4], but also an improvement in the sense that we successfully remove the restriction that $\mathsf{SNR} > e$ in [4, Theorem 3]. We summarize the results in the following theorem:

Theorem 4.6: Consider a coset of polar lattice $L+\chi$ constructed from the lattice partition $\eta\Lambda/\Lambda'$ with the noise variance $\tilde{\sigma} = \frac{\sigma_s \sigma}{\sqrt{\sigma_s^2+\sigma^2}}$. By further manipulating the shaping frozen sets of the symmetric polar codes in $L + \chi$ according to the constellation $D_{\eta\Lambda,\sigma_s}$, we get a coset of lattice code whose codewords are distributed as the discrete Gaussian distribution $D_{L+\chi,\sigma_s}$. In the limit as $\epsilon_{\eta\Lambda}(\tilde{\sigma}) \to 0$, $N \to \infty$ and $r = O(\log N)$, with any transmitting rate up to the channel capacity $\frac{1}{2}\log(1 + \mathsf{SNR})$ where $\mathsf{SNR} = \frac{P_s}{\sigma^2}$, the error probability of multi-stage SC decoding vanishes exponentially which is bounded by

$$P_e(L, \sigma^2) < \sum_{\ell=1}^{r} 2^{-N^{\beta_\ell}},$$

where $0 < \beta_\ell < \beta < 0.5$.

4.5 Summary

Asymmetric polar codes provide us a powerful tool to find the connections between input bits to implement the probabilistic shaping. The proposed multilevel asymmetric polar codes can be proved to achieve the capacity of the AWGN channel for any SNR. It is equivalent to implementing the shaping over an AWGN-good polar lattice by our equivalence lemma Lemma 4.6. This coding scheme is the first explicit construction of lattice codes achieving the capacity of the AWGN channel. We will expand this technique to the Gaussian wiretap channel in the next chapter Chapter 5.

CHAPTER 5

Polar lattices can achieve the strong secrecy capacity of the Gaussian wiretap channel

Now we are ready to introduce our polar lattice coding scheme on the Gaussian wiretap channel and the system model is shown in Figure 1.6. The target in this chapter is to prove the proposed coding scheme can achieve the strong secrecy capacity.[1] In [73], a nested polar lattice structure was proposed to achieve the strong secrecy on the Mod-Λ Gaussian wiretap channel. Although power constrain was taken into consideration, the shaping lattice was non-constructive, which makes the problem of constructing practical strong secrecy achieving polar lattice on the Gaussian wiretap channel still an open question. As we have shown in the above sections, the discrete lattice Gaussian distribution provides us an alternative way to obtain the shaping gain. In this chapter, we implement this shaping scheme on the same nested polar lattice structure proposed in [73]. In order to make the shaping scheme compatible, we have to modify the construction method of the multilevel wiretap polar codes. We notice that the modified wiretap polar coding scheme is still based on

[1] Thanks Ling Liu for his contribution on the proof of Lemma 5.4 and the concept of the shaping induced channel.

the original scheme introduced in [74]. However, we change the selection criteria of the information bad set for Eve, which is defined in terms of the Bhattacharyya parameter instead of the information of the subchannels. It turns out that our modified wiretap polar coding scheme can also be proved to achieve the strong secrecy capacity and it is more suitable for the further shaping implementation.

5.1 Modified Binary Symmetric Wiretap Polar Coding

In this part we consider the construction of polar codes on the binary symmetric wiretap channel. With some abuse of notation, we use $\tilde{V}$ and $\tilde{W}$ to denote the main channel between Alice and Bob and the wiretap channel between Alice and Eve respectively. Both $\tilde{V}$ and $\tilde{W}$ are with binary input X and $\tilde{W}$ is degraded with respect to $\tilde{V}$. Let Y and Z denote the output of $\tilde{V}$ and $\tilde{W}$. After the channel combination and splitting of N independent uses of the $\tilde{V}$ and $\tilde{W}$ by the polarization transform $U^{1:N} = X^{1:N}G_N$, we define the sets of reliability-good indices for Bob and information-bad indices for Eve as

$$\begin{aligned} \mathcal{G}(\tilde{V}) &= \{i : \tilde{Z}(\tilde{V}_N^{(i)}) \leq 2^{-N^\beta}\}, \\ \mathcal{N}(\tilde{W}) &= \{i : \tilde{Z}(\tilde{W}_N^{(i)}) \geq 1 - 2^{-N^\beta}\}. \end{aligned} \tag{5.1}$$

In [74], the information-bad set $\mathcal{N}(\tilde{W})$ was defined according to the mutual information of the subchannels $\{i : I(\tilde{W}_N^{(i)}) \leq 2^{-N^\beta}\}$. However, our new criterion is based on the Bhattacharyya parameter. The following lemma shows that the new criterion is stricter than the original one in the sense that the mutual information of the subchannels with indices in the new set $\mathcal{N}(\tilde{W})$ can also be bounded in the same form.

Lemma 5.1: Let $\tilde{W}_N^{(i)}$ be the i-th subchannel after the polarization transform on N independent uses of a BMS channel $\tilde{W}$. For any $0 < \beta < 0.5$, if $\tilde{Z}(\tilde{W}_N^{(i)}) \geq 1 - 2^{-N^\beta}$, the mutual information of the i-th subchannel can be upper-bounded as

$$I(\tilde{W}_N^{(i)}) \leq 2^{-N^{\beta'}}, 0 < \beta' < \beta < 0.5.$$

Proof. Since $\tilde{W}$ is symmetric, $\tilde{W}_N^{(i)}$ is symmetric as well. By the [Proposition 1, [16]], we have

$$\begin{aligned} I(\tilde{W}_N^{(i)}) &\leq \sqrt{1 - \tilde{Z}(\tilde{W}_N^{(i)})^2} \\ &\leq \sqrt{2 \cdot 2^{-N^\beta}} \leq 2^{-N^{\beta'}}. \end{aligned}$$

□

Since the mutual information of subchannels in $\mathcal{N}(\tilde{W})$ can be upper-bounded in the same form, it is not difficult to understand that strong secrecy can be achieved using the technique proposed in [74]. Similarly, we divide the index set $[N]$ into the following four sets shown in Figure 5.1:

$$\begin{aligned} &\mathcal{A} = \mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W}) \\ &\mathcal{B} = \mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W})^c \\ &\mathcal{C} = \mathcal{G}(\tilde{V})^c \cap \mathcal{N}(\tilde{W}) \\ &\mathcal{D} = \mathcal{G}(\tilde{V})^c \cap \mathcal{N}(\tilde{W})^c. \end{aligned} \tag{5.2}$$

Clearly, $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{D} = [N]$. Then we assign set $\mathcal{A}$ with message bits M, set $\mathcal{B}$ with random bits R, set $\mathcal{C}$ with frozen bits F which are known to both Bob and Eve prior to transmission and set $\mathcal{D}$ with random bits R.

The next lemma shows that this assignment achieves strong secrecy.

Lemma 5.2: According to the partitions of the index set shown in (5.2), if we

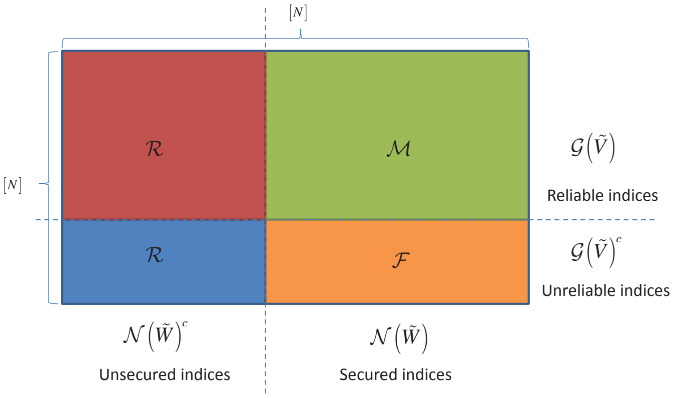


Figure 5.1: The partition of the index $[N]$ for the binary wiretap channel [74]. Intuitively, if the message bits are assigned in the reliable and secured set, both the reliability and secrecy can be guaranteed.

assign the four sets as follows

$$\begin{aligned} \mathcal{A} &\leftarrow M \\ \mathcal{B} &\leftarrow R \\ \mathcal{C} &\leftarrow F \\ \mathcal{D} &\leftarrow R, \end{aligned} \tag{5.3}$$

then the information leakage $I(M; Z^{1:N})$ can be upper-bounded as

$$I(M; Z^{1:N}) \leq N \cdot 2^{-N^{\beta'}}. \tag{5.4}$$

Proof. As has been shown in [74], the induced channel $MF \rightarrow Z^{1:N}$ is symmetric when $\mathcal{B}$ and $\mathcal{D}$ are fed with random bits R. For a symmetric channel, the maximum mutual information is achieved by uniform input distribution. Let $\tilde{U}_{\mathcal{A}}$ and $\tilde{U}_{\mathcal{C}}$ denote independent and uniform versions of M and F and $\tilde{Z}^{1:N}$ be the corresponding channel output. Letting $i_1 < i_2 < ... < i_{|\mathcal{A}\cup\mathcal{C}|}$ be the indices in $\mathcal{A} \cup \mathcal{C}$.

$$I(MF; Z^{1:N}) \quad \leq \quad I(\tilde{U}_{\mathcal{A}}\tilde{U}_{\mathcal{C}}; \tilde{Z}^{1:N})$$

$$\begin{aligned}
&= \sum_{j=1}^{|\mathcal{A}\cup\mathcal{C}|} I(\tilde{U}^{i_j}; \tilde{Z}^{1:N} | \tilde{U}^{i_1}, ..., \tilde{U}^{i_{j-1}}) \\
&= \sum_{j=1}^{|\mathcal{A}\cup\mathcal{C}|} I(\tilde{U}^{i_j}; \tilde{Z}^{1:N}, \tilde{U}^{i_1}, ..., \tilde{U}^{i_{j-1}}) \\
&\leq \sum_{j=1}^{|\mathcal{A}\cup\mathcal{C}|} I(\tilde{U}^{i_j}; \tilde{Z}^{1:N}, \tilde{U}^{1:i_j-1}) \\
&= \sum_{j=1}^{|\mathcal{A}\cup\mathcal{C}|} I(\tilde{W}_N^{(i_j)}) \leq N \cdot 2^{-N^{\beta'}}.
\end{aligned}$$

□

Due to the symmetry of the induced channel [74], there is no specific assumption on the distribution on M and F and a similar proof can be found in [75].

With regard to the secrecy rate, we show that the modified polar coding scheme can also achieve the secrecy capacity.

Lemma 5.3: Let $C(\tilde{V})$ and $C(\tilde{W})$ denote the channel capacity of the main channel $\tilde{V}$ and wiretap channel $\tilde{W}$ respectively. Since $\tilde{W}$ is degraded with respect to $\tilde{V}$, the secrecy capacity, which is given by $C(\tilde{V}) - C(\tilde{W})$, is achievable using the modified wiretap coding scheme, i.e.,

$$\lim_{N\to\infty} |\mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W})|/N = C(\tilde{V}) - C(\tilde{W}).$$

Proof. According to the definitions of $\mathcal{G}(\tilde{V})$ and $\mathcal{N}(\tilde{W})$ presented in (5.1),

$$\begin{aligned}
\lim_{N\to\infty} \frac{|\mathcal{G}(\tilde{V})|}{N} &= \lim_{N\to\infty} \frac{1}{N} |\{i : \tilde{Z}(\tilde{V}_N^{(i)}) \leq 2^{-N^{\beta}}\}| = C(\tilde{V}), \\
\lim_{N\to\infty} \frac{|\mathcal{N}(\tilde{W})|}{N} &= \lim_{N\to\infty} \frac{1}{N} |\{i : \tilde{Z}(\tilde{W}_N^{(i)}) \geq 1 - 2^{-N^{\beta}}\}| = 1 - C(\tilde{W}).
\end{aligned}$$

Here we define another two sets $\bar{\mathcal{G}}(\tilde{V})$ and $\bar{\mathcal{N}}(\tilde{W})$ as

$$\begin{aligned}\bar{\mathcal{G}}(\tilde{V}) \quad &= \{i : \tilde{Z}(\tilde{V}_N^{(i)}) \geq 1 - 2^{-N^\beta}\},\\ \bar{\mathcal{N}}(\tilde{W}) \quad &= \{i : \tilde{Z}(\tilde{W}_N^{(i)}) \leq 2^{-N^\beta}\}.\end{aligned}$$

Similarly, we have $\lim_{N\to\infty} \frac{|\bar{\mathcal{G}}(\tilde{V})|}{N} = 1 - C(\tilde{V})$ and $\lim_{N\to\infty} \frac{|\bar{\mathcal{N}}(\tilde{W})|}{N} = C(\tilde{W})$. Since $\tilde{W}$ is degraded with respect to $\tilde{V}$, $\bar{\mathcal{G}}(\tilde{V})$ and $\bar{\mathcal{N}}(\tilde{W})$ are disjoint with each other, then we have

$$\lim_{N\to\infty} \frac{|\bar{\mathcal{G}}(\tilde{V}) \cup \bar{\mathcal{N}}(\tilde{W})|}{N} = 1 - C(\tilde{V}) + C(\tilde{W}).$$

By the property of polarization, the proportion of the unpolarized part is vanishing as N goes to infinity, i.e.,

$$\begin{aligned}\lim_{N\to\infty} \frac{|\mathcal{G}(\tilde{V}) \cup \bar{\mathcal{G}}(\tilde{V})|}{N} &= 1,\\ \lim_{N\to\infty} \frac{|\mathcal{N}(\tilde{W}) \cup \bar{\mathcal{N}}(\tilde{W})|}{N} &= 1,\end{aligned}$$

Finally, we have

$$\lim_{N\to\infty} \frac{|\mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W})|}{N} = 1 - \lim_{N\to\infty} \frac{|\bar{\mathcal{G}}(\tilde{V}) \cup \bar{\mathcal{N}}(\tilde{W})|}{N} = C(\tilde{V}) - C(\tilde{W}).$$

□

It is not difficult to observe that the proportion of the problematic set $\mathcal{D}$ is arbitrarily small. This because set $\mathcal{D}$ is a subset of the unpolarized set $\{i : 2^{-N^\beta} < \tilde{Z}(\tilde{V}_N^{(i)}) < 1 - 2^{-N^\beta}\}$. As has been shown in [74], the reliability condition cannot be proved due to the existence of the set $\mathcal{D}$. Fortunately, a blocking technique can solve the issue of the set $\mathcal{D}$. Since we do not need this technique in our lattice coding design, we refer the reader to [75] for more details.

5.2 Secrecy-Good Lattices

In [73], we have already discussed how to obtain the AWGN good lattice Λ_b and secrecy good lattice Λ_e for the mod-Λ Gaussian wiretap channel. In fact, the result also holds for the polar lattices when the input distribution of each level is uniform for the genuine Gaussian wiretap channel. The setting without power constraint is similar to the Poltyrev setting in the Gaussian point-to-point channel.

Definition 5.1 (Secrecy-good): Alice sends the confidential message M which is mapped to the coset leaders of the lattice partition Λ_b/Λ_e in a Gaussian wiretap channel. If the above coding scheme results in fast-vanishing information leakage $I(M; Z^{1:N})$ where $Z^{1:N}$ is the signal received by Eve. Then the lattice Λ_e is regarded as a secrecy-good lattice.

Note that this definition is more general than the definition proposed in [9] which is based on the flatness factor.

As we have analyzed, Λ_b and Λ_e can be viewed as the lattices constructed according to the related symmetric channels at each level. Briefly, our construction method of Λ_b and Λ_e is based on the previously mentioned polar coding scheme for binary symmetric wiretap channels. A polar lattice L is constructed by a set of nested polar codes $C_1(N, k_1) \subseteq C_2(N, k_2) \subseteq \cdots \subseteq C_{r-1}(N, k_{r-1})$ and a binary lattice partition chain $\Lambda_1/\Lambda_2/\cdots/\Lambda_r$. The block length of polar codes is N. Alice splits the message M into $M_1, \cdots, M_{r-1}$. We follow (5.3) to assign bits in the component polar codes to achieve strong secrecy. Define $V_\ell = W(\Lambda_\ell/\Lambda_{\ell+1}, \sigma_b^2)$ and $W_\ell = W(\Lambda_\ell/\Lambda_{\ell+1}, \sigma_e^2)$ and W_ℓ is degraded with respect to V_ℓ for $1 \leq \ell \leq r-1$. Then we can get $\mathcal{A}_\ell$, $\mathcal{B}_\ell$, $\mathcal{C}_\ell$

and $\mathcal{D}_\ell$ for $1 \leq \ell \leq r-1$ which are defined as

$$\begin{aligned}
\mathcal{A}_\ell &= \mathcal{G}(V_\ell) \cap \mathcal{N}(W_\ell) \\
\mathcal{B}_\ell &= \mathcal{G}(V_\ell) \cap \mathcal{N}(W_\ell)^c \\
\mathcal{C}_\ell &= \mathcal{G}(V_\ell)^c \cap \mathcal{N}(W_\ell) \\
\mathcal{D}_\ell &= \mathcal{G}(V_\ell)^c \cap \mathcal{N}(W_\ell)^c.
\end{aligned}$$

Similarly we assign the bits as follows

$$\begin{aligned}
\mathcal{A}_\ell &\leftarrow M_\ell \\
\mathcal{B}_\ell &\leftarrow R \\
\mathcal{C}_\ell &\leftarrow F \\
\mathcal{D}_\ell &\leftarrow R
\end{aligned} \tag{5.5}$$

for $1 \leq \ell \leq r-1$. Since W_ℓ(and V_ℓ) is degraded with respect to $W_{\ell+1}$(and $V_{\ell+1}$), it is easy to obtain that $\mathcal{C}_\ell \supseteq \mathcal{C}_{\ell+1}$ which means $\mathcal{A}_\ell \cup \mathcal{B}_\ell \cup \mathcal{D}_\ell \subseteq \mathcal{A}_{\ell+1} \cup \mathcal{B}_{\ell+1} \cup \mathcal{D}_{\ell+1}$. This construction is clearly a lattice construction as polar codes constructed on each level are nested.

Interestingly, this polar lattice construction generates an AWGN-good lattice Λ_b and a secrecy-good lattice Λ_e simultaneously. Λ_b is constructed from a set of nested polar codes $C_1(N, |\mathcal{A}_1| + |\mathcal{B}_1| + |\mathcal{D}_1|) \subseteq \cdots \subseteq C_{r-1}(N, |\mathcal{A}_{r-1}| + |\mathcal{B}_{r-1}| + |\mathcal{D}_{r-1}|)$ and the lattice partition chain $\Lambda_1/\cdots/\Lambda_r$, while Λ_e is constructed from a set of nested polar codes $C_1(N, |\mathcal{B}_1| + |\mathcal{D}_1|) \subseteq \cdots \subseteq C_{r-1}(N, |\mathcal{B}_{r-1}| + |\mathcal{D}_{r-1}|)$ and the same lattice partition chain $\Lambda_1/\cdots/\Lambda_r$.

By using the above assignments and Lemma 5.2, such polar codes can guarantee an upper bound on the mutual information between the input message M_ℓ and the output of the Eve's ℓ-th level channel Z_ℓ^N as shown in the following inequality:

$$I(M_\ell; Z_\ell^N) \leq N2^{-N^{\beta'}},$$

where $Z_\ell^N = Z^{1:N} \bmod \Lambda_{\ell+1}$. Recall $Z^{1:N}$ is the signal received by Eve.

From the equivalence lemma Lemma 4.6, this polar code can also guarantee the same upper bound on the mutual information between the input message and the output of the channel derived by the chain rule of the mutual information (4.1) as shown in the following inequality:

$$I(M_\ell; Z^{1:N}, X_{1:\ell-1}^{1:N}) \leq N2^{-N^{\beta'}}.$$

From the chain rule of mutual information,

$$\begin{aligned} I(Z^{1:N}; M) &= \sum_{i=1}^{r} I(Z^{1:N}; M_\ell | M_{1:\ell-1}) \qquad (5.6) \\ &= \sum_{\ell=1}^{r} h(M_\ell | M_{1:\ell-1}) - h(M_\ell | Z^{1:N}, M_{1:\ell-1}) \\ &= \sum_{\ell=1}^{r} h(M_\ell) - h(M_\ell | Z^{1:N}, M_{1:\ell-1}) \\ &= \sum_{\ell=1}^{r} I(M_\ell; Z^{1:N}, M_{1:\ell-1}) \\ &\leq \sum_{\ell=1}^{r} I(M_\ell; Z^{1:N}, X_{1:\ell-1}^{1:N}) \leq rN2^{-N^{\beta'}}, \end{aligned}$$

where the first inequality is because by adding more random variables cannot decrease the mutual information. Therefore strong secrecy is achieved as $\lim_{N\to\infty} I(M; Z^{1:N}) = 0$.

As we have analyzed in Lemma 4.6, the secrecy-good polar lattice constructed above is based on the symmetric channel at each level. Without considering the shaping, the lattice can be viewed as the polar lattice constructed on an MMSE scaled

Gaussian wiretap channel, i.e., with main channel noise variance $\tilde{\sigma}_b^2$ and wiretap channel noise variance $\tilde{\sigma}_e^2$. According to the main theory of [73], we have

$$\begin{aligned}\lim_{N\to\infty} R &= \sum_{\ell=1}^{r} \lim_{N\to\infty} \frac{|\mathcal{A}_\ell|}{N} \\ &= \sum_{\ell=1}^{r} I(\tilde{X}_\ell; Y, \tilde{X}_\ell \oplus \mathsf{X}_\ell, X_{1:\ell-1}) - I(\tilde{X}_\ell; Z, \tilde{X}_\ell \oplus \mathsf{X}_\ell, X_{1:\ell-1}) \qquad (5.7) \\ &= \frac{1}{2}\log(\frac{\tilde{\sigma}_e^2}{\tilde{\sigma}_b^2}) = \frac{1}{2}\log\left(\frac{1+\mathsf{SNR}_b}{1+\mathsf{SNR}_e}\right).\end{aligned}$$

Remark 5.1: In the Poltyrev setting, the capacities of the main channel and the wiretap channel are both infinity. But the secrecy capacity is finite. Interestingly it equals the secrecy capacity under the power constraint. Therefore we can predict here that the shaping operation on both Λ_b and Λ_e should not change the secrecy capacity, which is shown in the following section.

5.3 Shaping over Λ_b and Λ_e

Now we consider shaping for both AWGN-good and secrecy-good lattices. Since the shaping scheme is implemented by Alice, who is the sender of both main channel and wiretap channel, the shaping is implemented on Λ_b and Λ_e simultaneously. According to Chapter 4, we have to strip off those indices with small $Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:l-1}^{1:N})$ from the information set of the symmetric channels. Therefore, Alice cannot send messages on those subchannels with $Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta}$. Note that this part is the same for $\tilde{V}_\ell$ and $\tilde{W}_\ell$, because it only depends on the shaping distribution. At the ℓ-th level, to make the input distribution satisfying $P_{X_\ell|X_{1:\ell-1}}$, the index set which is used for shaping is given as

$$\mathcal{S}_\ell \triangleq \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta}\}.$$

The index set which is shaping free is denoted by $\mathcal{S}_\ell^c$. Recall that for the index set $[N]$, we already have two partition criteria, i.e, reliability-good and information-bad (see (5.1)). We rewrite the reliability-good index set $\mathcal{G}_\ell$ and information-bad index set $\mathcal{N}_\ell$ at level ℓ as

$$\begin{aligned} \mathcal{G}_\ell \triangleq & \quad \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}, Y^{1:N}) \leq 2^{-N^\beta}\}, \\ \mathcal{N}_\ell \triangleq & \quad \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}, Z^{1:N}) \geq 1 - 2^{-N^\beta}\}. \end{aligned} \tag{5.8}$$

Note that $\mathcal{G}_\ell$ and $\mathcal{N}_\ell$ are defined by the asymmetric Bhattacharyya parameters. However, by Theorem 4.3 and Lemma 4.6, we have $\mathcal{G}_\ell = \mathcal{G}(\tilde{V}_\ell)$ and $\mathcal{N}_\ell = \mathcal{N}(\tilde{W}_\ell)$, where $\tilde{V}_\ell$ and $\tilde{W}_\ell$ are the corresponding symmetric channels for Bob and Eve at level ℓ. The four sets $\mathcal{A}_\ell$, $\mathcal{B}_\ell$, $\mathcal{C}_\ell$ and $\mathcal{D}_\ell$ are defined in the same fashion as (5.5) with $\mathcal{G}_\ell$ and $\mathcal{N}_\ell$ replacing $\mathcal{G}(\tilde{V}_\ell)$ and $\mathcal{N}(\tilde{W}_\ell)$, respectively.

As a result, we have three criteria: shaping-dependent, reliability-good and information-bad. The whole index set $[N]$ is divided in a cube according to these three directions which give us eight sets:

$$\begin{aligned} &\mathcal{A}_\ell^{\mathcal{S}} = \mathcal{A}_\ell \cap \mathcal{S}_\ell, \ \mathcal{A}_\ell^{\mathcal{S}^c} = \mathcal{A}_\ell \cap \mathcal{S}_\ell^c \\ &\mathcal{B}_\ell^{\mathcal{S}} = \mathcal{B}_\ell \cap \mathcal{S}_\ell, \ \mathcal{B}_\ell^{\mathcal{S}^c} = \mathcal{B}_\ell \cap \mathcal{S}_\ell^c \\ &\mathcal{C}_\ell^{\mathcal{S}} = \mathcal{C}_\ell \cap \mathcal{S}_\ell, \ \mathcal{C}_\ell^{\mathcal{S}^c} = \mathcal{C}_\ell \cap \mathcal{S}_\ell^c \\ &\mathcal{D}_\ell^{\mathcal{S}} = \mathcal{D}_\ell \cap \mathcal{S}_\ell, \ \mathcal{D}_\ell^{\mathcal{S}^c} = \mathcal{D}_\ell \cap \mathcal{S}_\ell^c \end{aligned}$$

By a careful analysis which will be given in the journal paper, we can obtain the following lemma.

Lemma 5.4: Consider the reliability-good indices set $\mathcal{G}_\ell$ and information-bad indices set $\mathcal{N}_\ell$ defined as in (5.8). By striping off the source coding set $\mathcal{S}_\ell$, we get the new message set $\mathcal{A}_\ell^{\mathcal{S}^c} = \mathcal{G}_\ell \cap \mathcal{N}_\ell \cap \mathcal{S}_\ell^c$, the proportion of $|\mathcal{A}_\ell^{\mathcal{S}^c}|$ equals to that of $|\mathcal{A}_\ell|$, and the message rate after shaping can still be arbitrarily close to $\frac{1}{2} \log \left(\frac{1+\mathsf{SNR}_b}{1+\mathsf{SNR}_e} \right)$.

Proof. $\mathcal{A}_\ell^{\mathcal{S}c} = \mathcal{A}_\ell$ and (5.7). □

5.4 Strong secrecy

In [74], an induced channel is defined to prove the strong secrecy. Here we call it randomness induced channel because it is caused by feeding the subchannels in the set $\mathcal{B}_\ell$ and $\mathcal{D}_\ell$ with uniformly random bits. Following the same fashion, we will define an induced channel for the wiretap coding scheme with shaping. However, this new induced channel is different from the randomness induced channel because we are no longer feeding uniformly random bits to the subchannels in the set $\mathcal{B}_\ell$ and $\mathcal{D}_\ell$. In fact, some subchannels (covered by the mapping) should be fed with the bits according to the distribution $P(U_\ell^i|U_\ell^{1:i-1}, X_{1:l-1}^{1:N})$. We define the channel induced by the shaping bits as the shaping induced channel.

Definition 5.2 (Shaping induced channel): The shaping induced channel $\mathcal{Q}_N(W, \mathcal{S})$ is defined in terms of N uses of an asymmetric channel W, and a shaping subset $\mathcal{S}$ of $[N]$ of size $|\mathcal{S}| = s$. The input alphabet of $\mathcal{Q}_N(W, \mathcal{S})$ is $\{0, 1\}^{N-s}$ and the bits in $\mathcal{S}$ are determined by the input bits according to a specific mapping.

According to the analysis in Sec. 5.3, we can set $\mathcal{S}$ as the set which consists of all the bits decided by the mapping (including set $\mathcal{B}$). Based on the shaping induced channel, we define the new induced channel, which is caused by feeding a part of the input bits of the shaping induced channel with uniformly random bits. It is a combination of the shaping induced channel and randomness induced channel. The input alphabet of $\mathcal{Q}_N(W, \mathcal{S}, \mathcal{R})$ is $\{0, 1\}^{N-s-r}$ and the bits in $\mathcal{R}$ are uniformly and independently random. This is different from the definition given in [74] because the bits in $\mathcal{S}$ are neither independent to the message bits nor uniformly distributed. As long as the input bits of the new induced channel are uniform and the shaping bits are chosen according to all possible mappings (randomly pick one of the family of mappings each time), the new induced channel can still generate 2^N possible realizations of $X_\ell^{1:N}$ as N goes to infinity, and those $x_\ell^{1:N}$ can be viewed as the output of N i.i.d binary sources with input distribution $P_{X_\ell|X_{1:\ell-1}}$. These two results are exactly the conditions required by Theorem 4.2. Specifically, we have $Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}, Z^{1:N}) = \tilde{Z}(\tilde{U}_\ell^i|\tilde{U}_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}, X_\ell^{1:N} \oplus \tilde{X}_\ell^{1:N}, Z^{1:N})$. In simple words, this equation holds when $x_\ell^{1:N}$ and $x_\ell^{1:N} \oplus \tilde{x}_\ell^{1:N}$ are all selected from $\{0,1\}^N$ according to their distributions. Then we can exploit the relation between the asymmetric channel and the corresponding symmetric channel to help us to bound the mutual information of the asymmetric channel . Therefore, we have to stick to the input distribution (uniform) of our new induced channel and also the distribution of the mappings. This is similar to the setting of the randomness induced channel in [74], where the input distribution and the randomness distribution are both set to be uniform. However, the randomness induced channel is further proved to be symmetric, then any other input distribution can still achiev the strong secrecy and the symmetry finally results in the semantic security. In this work, unfortunately, we do not have the symmetry of the new induced channel, and the input distribution which includes message bits and the independent frozen bits should be restricted to be uniform. In other words, we can not fix the independent frozen bits as [74] did.

Lemma 5.5: Let M_ℓ be the message and F_ℓ be the independent frozen bits at the input of the channel at the ℓ-th level after shaping, we have

$$I(M_\ell F_\ell; Z^{1:N}, X_{1:\ell-1}^{1:N}) \leq 2N2^{-N^{\beta'}}.$$

Proof. For the shaping induced channel $\mathcal{Q}_N(W_\ell, \mathcal{S}_\ell, \mathcal{R})$, we write the indices of the input bits $(\mathcal{S}_\ell \cup \mathcal{R})^c = [N] \setminus (\mathcal{S}_\ell \cup \mathcal{R})$ as $(\mathcal{S}_\ell \cup \mathcal{R})^c = \{i_1, i_2, ..., i_{N-s_\ell-r}\}$, where

$|\mathcal{R}| = r$ and $|\mathcal{S}_\ell| = s_\ell$, and assume that $i_1 < i_2 < \cdots < i_{N-s_\ell-r}$. We have

$$\begin{aligned}
I(M_\ell F_\ell; Z^{1:N}, X^{1:N}_{1:\ell-1}) &= I(U_\ell^{(\mathcal{S}_\ell \cup \mathcal{R})^c}; Z^{1:N}, X^{1:N}_{1:\ell-1}) \\
&= I(U_\ell^{i_1}, U_\ell^{i_2}, ..., U_\ell^{i_{N-r-s_\ell}}; Z^{1:N}, X^{1:N}_{1:\ell-1}) \\
&= \sum_{j=1}^{N-r-s_\ell} I(U_\ell^{i_j}; Z^{1:N}, X^{1:N}_{1:\ell-1} | U_\ell^{i_1}, U_\ell^{i_2}, ..., U_\ell^{i_{j-1}}) \\
&= \sum_{j=1}^{N-r-s_\ell} I(U_\ell^{i_j}; Z^{1:N}, X^{1:N}_{1:\ell-1}, U_\ell^{i_1}, U_\ell^{i_2}, ..., U_\ell^{i_{j-1}}) \\
&\overset{(a)}{\leq} \sum_{j=1}^{N-r-s_\ell} I(U_\ell^{i_j}; Z^{1:N}, X^{1:N}_{1:\ell-1}, U_\ell^{1}, U_\ell^{2}, ..., U_\ell^{i_j-1})
\end{aligned}$$

where (a) holds because adding more variables will not decrease the mutual information.

Then the above mutual information can be bounded by the mutual information of the symmetric channel plus an infinitesimal term as follows:

$$\begin{aligned}
&\sum_{j=1}^{N-r-s_\ell} I(U_\ell^{i_j}; Z^{1:N}, X^{1:N}_{1:\ell-1}, U_\ell^{1:i_j-1}) \\
&\quad \overset{(a)}{\leq} \sum_{j=1}^{N-r-s_\ell} I(\tilde{U}_\ell^{i_j}; Z^{1:N}, X^{1:N}_{1:\ell-1}, \tilde{X}_\ell^{1:N} \oplus X_\ell^{1:N}, \tilde{U}_\ell^{1:i_j-1}) + H(\tilde{U}_\ell^{i_j} | Z^{1:N}, X^{1:N}_{1:\ell-1}, \tilde{X}_\ell^{1:N} \oplus X_\ell^{1:N}, \tilde{U}_\ell^{1:i_j-1}) \\
&\quad - \sum_{j=1}^{N-r-s_\ell} H(U_\ell^{i_j} | Z^{1:N}, X^{1:N}_{1:\ell-1}, U_\ell^{1:i_j-1}) \\
&\quad \overset{(b)}{\leq} \sum_{j=1}^{N-r-s_\ell} I(\tilde{U}_\ell^{i_j}; Z^{1:N}, X^{1:N}_{1:\ell-1}, \tilde{X}_\ell^{1:N} \oplus X_\ell^{1:N}, \tilde{U}_\ell^{1:i_j-1}) \\
&\quad + \sum_{j=1}^{N-r-s_\ell} Z(U_\ell^{i_j} | Z^{1:N}, X^{1:N}_{1:\ell-1}, U_\ell^{1:i_j-1}) - (Z(U_\ell^{i_j} | Z^{1:N}, X^{1:N}_{1:\ell-1}, U_\ell^{1:i_j-1}))^2 \\
&\quad \overset{(c)}{\leq} \sum_{j=1}^{N-r-s_\ell} I(\tilde{U}_\ell^{i_j}; Z^{1:N}, X^{1:N}_{1:\ell-1}, \tilde{X}_\ell^{1:N} \oplus X_\ell^{1:N}, \tilde{U}_\ell^{1:i_j-1}) + N2^{-N^\beta} \\
&\quad \overset{(d)}{\leq} N2^{-N^{\beta'}} + N2^{-N^\beta} \\
&\quad \leq 2N2^{-N^{\beta'}}
\end{aligned}$$

for $0 < \beta' < \beta < 0.5$ and inequalities (a)-(d) follows from

(a) $\tilde{U}_\ell^{i_j}$ is uniformly distributed,

(b) [7, Proposition 2] gives $H(X|Y)-H(X|Y,Z) \leq Z(X|Y)-(Z(X|Y,Z)^2)$ and Theorem 4.2,

(c) Our coding scheme can guarantee that $Z(U_\ell^{i_j}|Z^{1:N}, X_{1:\ell-1}^{1:N}, U_\ell^{1:i_j-1})$ is either smaller than 2^{-N^β} or greater than $1-2^{-N^\beta}$,

(d) Lemma 5.1.

□

Finally, the strong secrecy can be proved in the same fashion as shown in (5.6).

5.5 Reliability

In the original setting of polar coding scheme for binary wiretap channel [74], how to assign $\mathcal{D}$ is a problem. Assigning freezing bits to $\mathcal{D}$ guarantees the reliability but achieves the weak secrecy, whereas assigning random bits to $\mathcal{D}$ guarantees the strong secrecy but may violate the reliability requirement because $\mathcal{D}$ may be nonempty. In order to ensure strong security, $\mathcal{D}$ is assigned with random bits ($\mathcal{D} \in \mathcal{R}$), which results in the fact that this scheme failed to accomplish the theoretical reliability. More explicitly, for any ℓ-th level channel $W(\Lambda_i/\Lambda_{i+1}, \sigma_b^2)$ at Bob's end, the probability of error is upper bounded by the sum of the Bhattacharyya parameters $Z(W_N^{(j)}(\Lambda_i/\Lambda_{i+1}, \sigma_b^2))$ of those bit-channels that are not frozen to zero. For each bit-channel index j and $\beta < 0.5$, we have

$$j \in \mathcal{A} \cup \mathcal{R} = \mathcal{G}(W(\Lambda_i/\Lambda_{i+1}, \sigma_b^2), \beta) \cup \mathcal{D}.$$

By the definition (5.1), we can see that the sum of $Z(W_N^{(j)}(\Lambda_i/\Lambda_{i+1}, \sigma_b^2))$ over the set $\mathcal{G}(W(\Lambda_i/\Lambda_{i+1}, \sigma_b^2)$ is bounded by 2^{-N^β}, and therefore, the error probability of the ℓ-th level channel under the SC decoding, denoted by $P_e^{SC}(\Lambda_i/\Lambda_{i+1}, \sigma_b^2)$, can be upper bounded by

$$P_e^{SC}(\Lambda_i/\Lambda_{i+1},\sigma_b^2) \leq 2^{-N^\beta} + \sum_{j\in\mathcal{D}} Z(W_N^{(j)}(\Lambda_i/\Lambda_{i+1},\sigma_b^2)).$$

Since multistage decoding is utilized, by the union bound, the final decoding error probability for Bob is bounded as

$$\text{Pr}\{\hat{M} \neq M\} \leq \sum_{i=1}^{r-1} P_e^{SC}(\Lambda_i/\Lambda_{i+1},\sigma_b^2).$$

Unfortunately, a proof that this scheme satisfies the reliability condition cannot be arrived here because the bound of the sum $\sum_{j\in\mathcal{D}} Z(W_N^{(j)}(\Lambda_i/\Lambda_{i+1},\sigma_b^2))$ is not known. Note that significantly low probabilities of error can still be achieved in practice since the size of $\mathcal{D}$ is very small.

It is also worth mentioning that this reliability problem was recently solved in [75], where a new scheme dividing the information message of each Λ_i/Λ_{i+1} channel into several blocks is proposed. For a specific block, $\mathcal{D}$ is still assigned with random bits and transmitted in advance in the set $\mathcal{A}$ of the previous block. This scheme involves negligible rate loss and finally realizes reliability and strong security simultaneously. In this case, if the reliability of each partition channel can be achieved, i.e., for any ℓ-th level partition Λ_i/Λ_{i+1}, $P_e^{SC}(\Lambda_i/\Lambda_{i+1},\sigma_b^2)$ vanishes as N goes to infinity. Then the total decoding error probability for Bob can be made arbitrarily small. Actually, based on the new scheme of assigning the problematic bits in $\mathcal{D}$ [75], the error probability on level i can be upper bounded by

$$P_e^{SC}(\Lambda_i/\Lambda_{i+1},\sigma_b^2) \leq \epsilon_{N'}^i + k_i \cdot o(2^{-N'^\beta}),$$

where k_i is the number of information blocks on the ℓ-th level, N' is the length of each block which satisfies $N' \times k_i = N$ and ϵ_N^i is caused by the first separate block on the ℓ-th level consisting of the initial bits in $\mathcal{D}_i$. Since $|\mathcal{D}_i|$ is extremely small comparing to the block length N, the decoding failure probability for the first block can be made arbitrarily small when N is sufficiently large. Therefore, Λ_b is an AWGN-good lattice.

Note that the rate loss incurred by repeatedly transmitting bits in $\mathcal{D}_i$ is negligible because of its small size and the fact that only one block is wasted on each level. Explicitly, the actually achieved secrecy rate in the ℓ-th level is given by $\frac{k_i}{k_i+1}[C(\Lambda_i/\Lambda_{i+1}, \sigma_b^2) - C(\Lambda_i/\Lambda_{i+1}, \sigma_e^2)]$. Clearly, this rate can be made close to the maximum secrecy rate by choosing sufficiently large k_i as well.

The above analysis is for the coding design without shaping. When shaping is involved, the problematic set $\mathcal{D}_\ell$ at each level is included in the shaping $\mathcal{S}_\ell$. The bits in $\mathcal{D}_\ell$ can be recovered by Bob simply by the sharing mapping and do not need to use the blocking technique. By Theorem 4.3 and Theorem 4.4, the reliability at each level can be guaranteed by uniformly distributed independent frozen bits and random mapping with distribution $P(U_\ell^i|U_\ell^{1:i-1}, X_{1:l-1}^{1:N})$. Consequently, by the multilevel decoding and union bound, the expectation of the block error probability of our wiretap coding scheme is vanishing as N goes to infinity.

Now we present the main theorem of this Chapter.

Theorem 5.1: Consider a multilevel coset code constructed from polar codes based on asymmetric channels and lattice Gaussian shaping $D_{\mathbb{Z},\sigma_s}$. Given $\sigma_e^2 > \sigma_b^2$, as the number of levels $r = O(\log N)$, $N \to \infty$ and $\epsilon_{\mathbb{Z}}\left(\frac{\sigma_s\sigma_e}{\sqrt{\sigma_s^2+\sigma_e^2}}\right) \to 0$, all strong secrecy rates R satisfying $R < \frac{1}{2}\log\left(\frac{1+\mathsf{SNR}_b}{1+\mathsf{SNR}_e}\right)$ are achievable for the Gaussian wiretap channel.

Proof.

$$\begin{aligned}
\lim_{N\to\infty} R &= \sum_{i=1}^{r} \lim_{N\to\infty} \frac{|\mathcal{A}_\ell^{\mathcal{S}^c}|}{N} \\
&= \sum_{i=1}^{r} I(Y; X_i|X_1, \cdots, X_{i-1}) - I(Z; X_i|X_1, \cdots, X_{i-1}) \\
&= \frac{1}{2}\log\left(\frac{1+\mathsf{SNR}_b}{1+\mathsf{SNR}_e}\right).
\end{aligned}$$

□

5.6 Discussions

We would like to explain our coding scheme for the Gaussian wiretap channel further in terms of the lattice structures. As we discussed in the previous section, we constructed the AWGN-good lattice Λ_b and the secrecy-good lattice Λ_e without considering the power constraint. We note that these two lattices are generated only if the independent frozen bits in each level are all zeros. By using the lattice Gaussian shaping $D_{\sigma_s,\mathbb{Z}}$ as our constellation, we actually implemented the lattice Gaussian shaping over both $\Lambda_b + \chi$ and $\Lambda_e + \chi$, where χ is a uniformly distributed shift. This is because we can not fix the independent frozen bits F_ℓ in our scheme (due to the lack of the proof that the new induced channel is symmetric). However, the coset leaders of the partition $\Lambda_b + \chi/\Lambda_e + \chi$ are the same as the lattice partition Λ_b/Λ_e. To sum up our coding scheme, Alice first associates each message $m \in M$ to a coset leader of Λ_b/Λ_e, then randomly picks a point in the coset $\Lambda_e + \chi + \lambda_m$ according to the distribution $D_{\Lambda_e+\chi+\lambda_m,\sigma_s}$ to send. The above scheme is consistent with the theoretical model proposed in [9].

Another practical issue is that how to share the uniformly distributed bits in the independent frozen bits and the specific mapping. The solution is that we assume Alice shares a seed with both Bob and Eve. Then they can generate the independent frozen bits locally. Bob can recover the shaping frozen set according to the seed and the distribution $P(U_\ell^i|U_\ell^{1:i-1}, X_{1:l-1}^{1:N})$ which is available for Bob. We must admit that it is possible for Eve to obtain some bits in the shaping frozen set S_ℓ given F_ℓ and mapping even before the communication. An unavoidable question is that whether such shaping bits in $\mathcal{S}_\ell$ make the message $\mathcal{M}_\ell$ insecure when Eve knows F and the selected mapping in the current round of communication. Fortunately those bits turn out to be irrelevant to the message $\mathcal{M}_\ell$, and they can be viewed as another kind of frozen bits in $\mathcal{S}_\ell$. Therefore we can conclude that the whole shaping scheme is secure in the sense that the mutual information leakage between M and $Z^{1:N}$ vanishes subexponentially with the block length of polar codes N.

5.7 Summary

Polar lattices with discrete Gaussian shaping have been proved to be good in the AWGN channel. We apply this technique to the Gaussian wiretap channel. The design turns out to be a shaping over an AWGN-good polar lattice and a secrecy-good polar lattice simultaneously. Finally it can be proved to achieve the strong secrecy capacity of the Gaussian wiretap channel.

CHAPTER 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, we have refined Forney *et al.*'s multilevel approach to the construction of AWGN-good lattices. The channel capacity of each level is calculated and a polar code is constructed to achieve its capacity accordingly. This leads to the construction of polar lattices and the proof of their AWGN-goodness. Polar lattices are of both theoretic and practical interests. Since polar lattices are as explicit as polar codes, their construction is equally efficient. Both the analysis and simulation results show that the performance of polar lattices can be improved by increasing the dimension n of the lattice partition chain. Compared with existing schemes [47, 44, 57, 46], polar lattices are distinguished by their provable AWGN-goodness and low complexity, namely, they asymptotically achieve the Poltyrev capacity with multi-stage decoding. With discrete Gaussian shaping, polar lattices also achieve the capacity of the power-constrained AWGN channel.

Following our previous work in [73], an explicit shaping scheme is proposed to construct polar lattices which achieve the strong secrecy rate of Gaussian wiretap channels. Since our shaping scheme uses the discrete lattice Gaussian distribution, the equivalent channel at each level is no longer symmetric, which requires capacity achieving polar codes for asymmetric channel. Fortunately, this problem can be solved by combining the design of channel coding and source coding together over a symmetric channel. Merely considering the channel coding part would give us secrecy-good polar lattices without shaping, as has been shown in [73]. To obtain the optimum shaping gain, the input bits according to the source coding part should be carefully selected. This also follows the concept that the shaping problem can be actually viewed as a source coding problem. It is worth noting that the channel equivalence between the related symmetric channel and the Λ/Λ' channel provides us much convenience for the coding design.

6.2 Future Work

It is well known that shaping has a close relation with quantization. Out next work is to construct quantization-good polar lattices.

It also would be very interesting to apply polar lattices to network applications, for example, multiple access channel, interference alignment and compute-and-forward problem. Here is a list of potential starting point:

1. Construct quantization-good polar lattice by employing the shaping technique presented in this thesis. Compared the performance with trellis coded quantization.
2. If polar lattices can be proved to be quantization good. Then it is straightforward to prove them to achieve the Wyner-Ziv bound.
3. In order to improve the performance of polar lattices, a soft multi-stage decoding is a promising direction. Maybe this can be implemented by a very long SC decoder.

4. The compute-and-forward problem also consists Gaussian noise and power constraint. Some theoretical work has already been done with random lattices and discrete Gaussian shaping. It would be very interesting to apply polar lattices in this scenario.
5. Another direction to apply polar lattices is the broadcast channel and interference channel. The construction of component polar codes need to be modified according to different requirements.
6. Regarding the secrecy, the first problem is to prove polar lattice can achieve the semantic security of the Gaussian wiretap channel. This requires the strong secrecy for any distribution of the messages. The problem of the broadcasting channel with confidential messages is still open. Polar lattices have the potential to solve it.

Bibliography

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, Jul.-Oct. 1948. 20, 22

[2] G. D. Forney Jr., “Coset codes-Part I: Introduction and geometrical classification,” *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1123–1151, Sep. 1988. 21, 43, 45, 95, 96

[3] J. Forney, G.D. and L.-F. Wei, “Multidimensional constellations-Part I: Introduction, figures of merit, and generalized cross constellations,” *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 877 –892, Aug. 1989. 21

[4] C. Ling and J.-C. Belfiore, “Achieving the AWGN channel capacity with lattice Gaussian coding,” *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014. 23, 27, 29, 37, 99, 115, 121, 122

[5] G. Poltyrev, “On coding without restictions for the AWGN channel,” *IEEE Trans. Inf. Theory*, vol. 40, pp. 409–417, Mar. 1994. 24, 29, 40, 62, 63, 64

[6] H. S. Cronie and S. B. Korada, “Lossless source coding with polar codes,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2010, pp. 904–908. 14, 28, 29, 108

[7] E. Arıkan, “Source polarizations,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2010, pp. 899–903. 28, 103, 105, 137

[8] J. Honda and H. Yamamoto, “Polar coding without alphabet extension for asymmetric models,” *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013. 28, 30, 99, 103, 104, 105, 106, 107, 109, 159, 162

[9] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé, “Semantically secure lattice codes for the Gaussian wiretap channel,” *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014. 29, 33, 37, 38, 129, 140

[10] B. Nazer and M. Gastpar, “Compute-and-forward: Harnessing interference through structured codes,” *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011. 29

[11] R. Zamir, S. Shamai, and U. Erez, “Nested linear/lattice codes for structured multiterminal binning,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002. 29

[12] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge, UK: Cambridge Univ. Press, Aug. 2014. 15, 29, 41, 42, 119

[13] ——, “Lattice coding for signals and networks: Application and design,” MIT, USA, Tutorial at ISIT 2012. 29

[14] H. A. Loeliger, “Averaging bounds for lattices and linear codes,” *IEEE Trans. Inf. Theory*, vol. 43, pp. 1767–1773, Nov. 1997. 29, 43, 44, 62, 64, 73, 154, 155

[15] U. Erez and R. Zamir, “Achieving 1/2 log (1+SNR) on the AWGN channel with lattice encoding and decoding,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004. 29, 30, 43, 73

[16] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009. 16, 30, 50, 78, 84, 85, 106, 118, 121, 125

[17] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Oct. 2009, pp. 144–148. 30

[18] A. Sahebi and S. Pradhan, "Multilevel channel polarization for arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory,*, vol. 59, no. 12, pp. 7839–7857, Dec. 2013. 30

[19] W. Park and A. Barg, "Polar codes for q-ary channels, $q = 2^r$," *IEEE Trans. Inf. Theory,*, vol. 59, no. 2, pp. 955–969, Feb. 2013. 30

[20] R. Mori and T. Tanaka, "Non-binary polar codes using Reed-Solomon codes and algebraic geometry codes," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Aug. 2010, pp. 1–5. 30

[21] E. Abbe and E. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5437–5448, Aug. 2012. 30

[22] E. Abbe and A. Barron, "Polar coding schemes for the AWGN channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2011, pp. 194–198. 30, 73

[23] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975. 30, 31

[24] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949. 31

[25] S. K. Leung-Yan-Cheong and M. E. Hellman, “The Gaussian wire-tap channel,” *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978. 31

[26] I.Csiszár and J.Körner, “Broadcast channels with confidential messages,” *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978. 32

[27] Y. Liang, H. Poor, and S. Shamai (Shitz), “Information theoretic security,” *Found. and Trends in Commun. and Inform. Theory*, vol. 5, no. 4-5, pp. 355–580, Apr. 2008. 32

[28] Y. Liang, H. Poor, and S. Shamai, “Secure communication over fading channels,” *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2470–2492, Jun. 2008. 32

[29] S. Shafiee, N. Liu, and S. Ulukus, “Towards the secrecy capacity of the Gaussian mimo wire-tap channel: The 2-2-1 channel,” *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4033–4039, Sep. 2009. 32

[30] Y. Liang and H. Poor, “Multiple-access channels with confidential messages,” *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 976–1002, Mar. 2008. 32

[31] H. Y.Liang and S.Shamai, “Physical layer security in Broadcast networks,” *Security and Commun. Networks*, vol. 2, no. 3, pp. 227–238, May 2009. 32

[32] Y. Liang, A. Somekh-Baruch, H. V. Poor, S. Shamai, and S. Verdu, “Capacity of cognitive interference channel with and without secrecy,” *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 604–619, Feb. 2009. 32

[33] L.H.Ozarow and A.D.Wyner, “Wire-tap channel II,” *Bell Syst. Tech. J.*, vol. 63, no. 10, pp. 2135–2157, Dec. 1984. 33

[34] L.-C. Choo, C. Ling, and K.-K. Wong, “Achievable rates for lattice coding over the Gaussian wiretap channel,” in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Jun. 2011, pp. 1–5. 33

[35] F. E. Oggier, P. Solé, and J.-C. Belfiore, “Lattice codes for the wiretap Gaussian channel: Construction and analysis,” to be published. [Online]. Available: http://arxiv.org/abs/1103.4086 33, 34

[36] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near shannon limit error-correcting coding and decoding: Turbo-codes. I,” in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, May 1993, pp. 1064–1070. 33

[37] D. Klinc, J. Ha, S. McLaughlin, J. Barros, and B.-J. Kwak, “LDPC codes for the Gaussian wiretap channel,” in *Proc. IEEE Inf. Theory Workshop (ITW)*, Oct. 2009, pp. 95–99. 34

[38] U. Erez, S. Litsyn, and R. Zamir, “Lattices which are good for (almost) everything,” *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005. 37

[39] G. D. Forney Jr., M. Trott, and S.-Y. Chung, “Sphere-bound-achieving coset codes and multilevel coset codes,” *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 820–850, May 2000. 38, 40, 49, 50, 67, 68, 69, 71, 72, 75, 85, 89, 93, 117, 119

[40] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*, 3rd ed. New York: Springer-Verlag, 1998. 39, 41, 43, 45, 52

[41] Y. Yan, C. Ling, and X. Wu, “Polar lattices: Where Arıkan meets Forney,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 1292–1296. 40, 43

[42] V. Tarokh, A. Vardy, and K. Zeger, “Universal bound on the performance of lattice codes,” *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 670 –681, Mar. 1999. 42

[43] J. Leech and N. J. A. Sloane, “Sphere packings and error-correcting codes,” *Can. J. Math.*, vol. 23, pp. 718–745, 1971. 43

[44] N. Di Pietro, J. J. Boutros, G. Zémor, and L. Brunel, “Integer low-density lattices based on construction A,” in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2012, pp. 422 –426. 43, 46, 54, 61, 73, 142

[45] E. S. Barnes and N. J. A. Sloane, “New lattice packings of spheres,” *Can. J. Math.*, vol. 35, pp. 117–130, 1983. 43, 47

[46] A. Sakzad, M.-R. Sadeghi, and D. Panario, “Construction of turbo lattices,” in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2010, pp. 14–21. 43, 53, 54, 142

[47] M.-R. Sadeghi, A. Banihashemi, and D. Panario, “Low-density parity-check lattices: Construction and decoding analysis,” *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4481–4495, Oct. 2006. 43, 52, 54, 61, 142

[48] A. Bos, J. H. Conway, and N. J. A. Sloane, “Further lattice packings in high dimensions,” *Math.*, vol. 29, pp. 171–180, 1982. 43

[49] A. Ingber, R. Zamir, and M. Feder, “Finite dimensional infinite constellations,” *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1630–1656, Mar. 2013. 43, 61, 89

[50] T. Ericson and V. Zinoviev, *Codes on Euclidean Spheres*, 1st ed. San Diego, CA: Elsevier, 2001. 45

[51] N. Di Pietro, J. J. Boutros, G. Zémor, and L. Brunel, “New results on low-density integer lattices,” in *Proc. Inform. Theory and Applicat. Workshop (ITA)*, Feb. 2013, pp. 1–6. 46, 61, 73

[52] N. Di Pietro, G. Zemor, and J. Boutros, “New results on Construction A lattices based on very sparse parity-check matrices,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 1675–1679. 46, 47

[53] W. Kositwattanarerk and F. Oggier, "On Construction D and related constructions of lattices from linear codes," in *Proc. Int. Workshop Coding and Cryptography (WCC)*, Apr. 2013. 48

[54] G. D. Forney Jr., "Coset codes-Part II: Binary lattices and related codes," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1152–1187, Sep. 1988. 48, 51

[55] D. Micciancio and A. Nicolosi, "Efficient bounded distance decoders for Barnes-Wall lattices," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2008, pp. 2484–2488. 52

[56] I.-J. Baik and S.-Y. Chung, "Irregular low-density parity-check lattices," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2008, pp. 2479–2483. 52

[57] N. Sommer, M. Feder, and O. Shalvi, "Low-density lattice codes," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1561–1585, Apr. 2008. 15, 54, 58, 59, 60, 61, 142

[58] O. Shalvi, N. Sommer, and M. Feder, "Signal codes: Convolutional lattice codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5203–5226, Aug. 2011. 15, 55, 58

[59] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967. 57

[60] R. Fischer, "The modulo-lattice channel: The key feature in precoding schemes," *Int. J. Electron. and Commun. (AEÜ)*, vol. 59, no. 4, pp. 244–253, Jun. 2005. 67

[61] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, EPFL, Lausanne, Switzerland, May 2009. 76, 86, 120

[62] E. Arıkan and E. Telatar, “On the rate of channel polarization,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2009, pp. 1493–1495. 79

[63] I. Tal and A. Vardy, “How to construct polar codes,” *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013. 79, 80, 81, 83, 108

[64] R. Pedarsani, S. Hassani, I. Tal, and E. Telatar, “On the construction of polar codes,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2011, pp. 11–15. 79, 82, 83

[65] I. Land and J. Huber, “Information combining,” *Found. and Trends in Commun. and Inform. Theory*, vol. 3, no. 3, pp. 227–330, Nov. 2006. 80

[66] M. Seidl, A. Schenk, C. Stierstorfer, and J. B. Huber, “Multilevel polar-coded modulation,” *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4108–4119, Oct. 2013. 84

[67] U. Wachsmann, R. Fischer, and J. Huber, “Multilevel codes: Theoretical concepts and practical design rules,” *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999. 85, 95, 100, 119

[68] S. H. Hassani, K. Alishahi, and R. Urbanke, “Finite-length scaling for polar codes,” *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5875–5898, Oct. 2014. 88, 89

[69] V. Guruswami and P. Xia, “Polar codes: Speed of polarization and polynomial gap to capacity,” in *Proc. IEEE 54th Annu. Symp. Found. of Comput. Sci. (FOCS)*, Oct. 2013, pp. 310–319. 88

[70] D. Goldin and D. Burshtein, “Improved bounds on the finite length scaling of polar codes,” *IEEE Trans. Inf. Theory*, pp. 6966–6978, Sep. 2014. 88, 89

[71] R. Mori and T. Tanaka, “Performance of polar codes with the construction using density evolution,” *IEEE Commun. Lett.*, vol. 13, no. 7, pp. 519–521, Jul. 2009. 94, 108

[72] Y. Yan and C. Ling, “A construction of lattices from polar codes,” in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2012, pp. 124–128. 95

[73] Y. Yan, L. Liu, and C. Ling, “Polar lattices for strong secrecy over the mod-Λ Gaussian wiretap channel,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2014, pp. 961–965. 123, 129, 132, 142, 143

[74] H. Mahdavifar and A. Vardy, “Achieving the secrecy capacity of wiretap channels using polar codes,” *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011. 17, 124, 125, 126, 127, 128, 134, 135, 137

[75] E. Şaşoğlu and A. Vardy, “A new polar coding scheme for strong security on wiretap channels,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 1117–1121. 127, 128, 138

APPENDIX A

Minkowski-Hlawka Theorem

In this proof, we only use mod-p lattices for demonstration purpose, i.e., of the form $L_C \triangleq \{v \in \mathbb{Z}^n : v \equiv c \bmod(p\mathbb{Z}^n), c \in C\}$, where p is a prime and $C(n,k)$ is a linear code over $\mathbb{Z}_p$ (Construction A). The lattice partition is $\mathbb{Z}/p\mathbb{Z}$. The fundamental volume of a scaled mod-p lattice is

$$V(\gamma L_C) = \gamma^n p^{n-k},$$

for some $\gamma \in \mathbb{R}$.

For any Riemann integrable function f: $\mathbb{R}^n \to R$ of bounded support and any positive ϵ, there exists a lattice Λ in $\mathbb{R}^n$ with fundamental volume 1 such that

$$\sum_{x \in \Lambda \backslash \{0\}} f(x) < \int_{\mathbb{R}^n} f(x) dx + \epsilon.$$

Theorem A.1 (MH Theorem for mod-p lattices, [14]): Let f be a Riemann integrable function $\mathbb{R}^n \to \mathbb{R}$ of bounded support. Then, for any integer k, $0 < k < n$, and any fixed V, the approximation

$$\frac{1}{\mathcal{C}}\sum_{C\in\mathcal{C}}\sum_{v\in\gamma L_C\setminus\{0\}} f(v) \approx V^{-1}\int_{\mathbb{R}^n} f(v)dv$$

where $\mathcal{C}$ is any balanced set of linear (n,k) codes over $\mathbb{Z}_p$, becomes exact in the limit $p\to\infty$, $\gamma\to 0$, $\gamma^n p^{n-k} = V$ fixed.

Proof.

$$\begin{aligned}
&\frac{1}{\mathcal{C}}\sum_{C\in\mathcal{C}}\sum_{v\in\gamma L_C\setminus\{0\}} f(v)\\
&\quad=\frac{1}{\mathcal{C}}\sum_{C\in\mathcal{C}}\left[\sum_{v\in\mathbb{Z}^n\setminus\{0\}:v \bmod p=0} f(\gamma v) + \sum_{v\in\mathbb{Z}^n:v \bmod p\in C\setminus\{0\}} f(\gamma v)\right]\\
&\quad\overset{(a)}{=}\sum_{v\in\mathbb{Z}^n\setminus\{0\}:v \bmod p=0} f(\gamma v) + \frac{1}{\mathcal{C}}\sum_{C\in\mathcal{C}}\sum_{c\in C\setminus\{0\}}\left[\sum_{v\in\mathbb{Z}^n:v \bmod p=c} f(\gamma v)\right]\\
&\quad\overset{(b)}{=}\sum_{v\in\mathbb{Z}^n\setminus\{0\}:v \bmod p=0} f(\gamma v) + \frac{p^k-1}{p^n-1}\sum_{c\in\mathbb{Z}_p^n}\left[\sum_{v\in\mathbb{Z}^n:v \bmod p=c} f(\gamma v)\right]\\
&\quad\overset{(c)}{=}\sum_{v\in\mathbb{Z}^n\setminus\{0\}:v \bmod p=0} f(\gamma v) + \frac{p^k-1}{p^n-1}\sum_{v\in\mathbb{Z}^n:v \bmod p\neq 0} f(\gamma v),
\end{aligned}$$

where the step from (a) to (b) follows from the Basic Averaging Lemma in [14]. Since f has bounded support, the left term of (c) vanishes for sufficiently large γp (i.e. f is the error probability function). The right term of (c) becomes

$$\frac{p^k-1}{p^n-1}\sum_{v\in\mathbb{Z}^n:v \bmod p\neq 0} f(\gamma v) \approx p^{k-n}\gamma^{-n}\int_{\mathbb{R}^n} f(v)dv$$

which becomes exact in the limit $p\to\infty$, $\gamma\to 0$. □

All known proofs of the Minkowski-Hlawka theorem are obtained from averaging over a large, usually infinite, class of lattices; in this sense, the Minkowski-Hlawka theorem can be regarded as the random coding arguments. One can derive various existence results for packing lattices. For example, for a sequence of lattices Λ_n, the best known asymptotic lower bound for the packing efficiency $\frac{r_\Lambda^{\text{pack}}}{r_\Lambda^{\text{effec}}}$ is equal to $\frac{1}{2}$, a result known as the Minkowski-Hlawka theorem.

APPENDIX B

The proof of Lemma 3.1

By the definition of the flatness factor, we have

$$f_{\sigma,\Lambda_1}(\mathbf{x}) \leq \frac{1+\epsilon_{\Lambda_1}(\sigma)}{V(\Lambda_1)}.$$

Thus, the differential entropy of the mod-Λ_1 Gaussian noise is bounded by

$$\begin{aligned}
h(\Lambda_1,\sigma^2) &= -\int_{\mathcal{V}(\Lambda_1)} f_{\sigma,\Lambda_1}(\mathbf{x}) \log f_{\sigma,\Lambda_1}(\mathbf{x}) d\mathbf{x} \\
&\geq -\int_{\mathcal{V}(\Lambda_1)} f_{\sigma,\Lambda_1}(\mathbf{x}) \log \frac{1+\epsilon_{\Lambda_1}(\sigma)}{V(\Lambda_1)} d\mathbf{x} \\
&= -\log \frac{1+\epsilon_{\Lambda_1}(\sigma)}{V(\Lambda_1)} \\
&= \log V(\Lambda_1) - \log\left(1+\epsilon_{\Lambda_1}(\sigma)\right).
\end{aligned}$$

Therefore, from (3.1), $C(\Lambda_1,\sigma^2)$ is bounded by $\log\left(1+\epsilon_{\Lambda_1}(\sigma)\right)$. The second inequality in (3.5) follows from the fact $\log(1+x) = \log_2(e)\cdot\log_e(1+x) \leq \log(e)\cdot x$ for $x > 0$.

APPENDIX C

Proof of Theorem 4.3

Proof. Let $\mathcal{E}_i$ denote the set of pairs of $u^{1:N}$ and $y^{1:N}$ such that decoding error occurs at the ith bit, then the block decoding error event is given by $\mathcal{E} \equiv \bigcup_{i\in\mathcal{I}} \mathcal{E}_i$. According to our encoding scheme, each codeword $u^{1:N}$ appears with probability

$$2^{-(|\mathcal{I}|+|\mathcal{F}|)} \prod_{i\in\mathcal{S}} P_{U^i|U^{1:i-1}}(u^i|u^{1:i-1}).$$

Then the expectation of decoding error probability over all random mapping is expressed as

$$\begin{aligned} E[P_e] \quad = \quad & \sum_{u^{1:N},y^{1:N}} 2^{-(|\mathcal{I}|+|\mathcal{F}|)}(\prod_{i\in\mathcal{S}} P_{U^i|U^{1:i-1}}(u^i|u^{1:i-1})) \\ & \cdot P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N}) \;\; [(u^{1:N},y^{1:N}) \in \mathcal{E}]. \end{aligned}$$

Now we define the probability distribution $Q_{U^{1:N},Y^{1:N}}$ as

$$Q_{U^{1:N},Y^{1:N}}(u^{1:N},y^{1:N}) = 2^{-(|\mathcal{I}|+|\mathcal{F}|)}(\prod_{i\in\mathcal{S}} P_{U^i|U^{1:i-1}}(u^i|u^{1:i-1}))P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N}).$$

Then the variational distance between $Q_{U^{1:N},Y^{1:N}}$ and $P_{U^{1:N},Y^{1:N}}$ can be bounded as

$$\begin{aligned}
&2||Q_{U^{1:N},Y^{1:N}} - P_{U^{1:N},Y^{1:N}}|| = \sum_{u^{1:N},y^{1:N}} |Q(u^{1:N},y^{1:N}) - P(u^{1:N},y^{1:N})| \\
\overset{(a)}{=} &\sum_{u^{1:N},y^{1:N}} |\sum_{i}(Q(u^i|u^{1:i-1}) - P(u^i|u^{1:i-1}))(\prod_{j=1}^{i-1} P(u^i|u^{1:i-1}))(\prod_{j=i+1}^{N} Q(u^i|u^{1:i-1}))Q(y^{1:N}|u^{1:N})| \\
\leq &\sum_{i\in\mathcal{I}\cup\mathcal{F}} \sum_{u^{1:N},y^{1:N}} |Q(u^i|u^{1:i-1}) - P(u^i|u^{1:i-1})|(\prod_{j=1}^{i-1} P(u^i|u^{1:i-1}))(\prod_{j=i+1}^{N} Q(u^i|u^{1:i-1}))Q(y^{1:N}|u^{1:N}) \\
= &\sum_{i\in\mathcal{I}\cup\mathcal{F}} \sum_{u^{1:i-1}} 2P(u^{1:i-1})||Q_{U^i|U^{1:i-1}=u^{1:i-1}} - P_{U^i|U^{1:i-1}=u^{1:i-1}}|| \\
\overset{(b)}{\leq} &\sum_{i\in\mathcal{I}\cup\mathcal{F}} \sum_{u^{1:i-1}} P(u^{1:i-1})\sqrt{2\ln2 D(P_{U^i|U^{1:i-1}=u^{1:i-1}}||Q_{U^i|U^{1:i-1}=u^{1:i-1}})} \\
\leq &\sum_{i\in\mathcal{I}\cup\mathcal{F}} \sqrt{2\ln2 \sum_{u^{1;i-1}} P(u^{1:i-1})D(P_{U^i|U^{1:i-1}=u^{1:i-1}}||Q_{U^i|U^{1:i-1}=u^{1:i-1}})} \\
\leq &\sum_{i\in\mathcal{I}\cup\mathcal{F}} \sqrt{2\ln2 D(P_{U^i|U^{1:i-1}}||Q_{U^i|U^{1:i-1}})} \\
\leq &\sum_{i\in\mathcal{I}} \sqrt{2\ln2(1 - H(U^i|U^{1:i-1}))} + \sum_{i\in\mathcal{F}} \sqrt{2\ln2(1 - H(U^i|U^{1:i-1}))} \\
\leq &\sum_{i\in\mathcal{I}} \sqrt{2\ln2(1 - Z(U^i|U^{1:i-1})^2)} + \sum_{i\in\mathcal{F}} \sqrt{2\ln2(1 - Z(U^i|U^{1:i-1},Y^{1:N})^2)} \\
\leq &\ 2N\sqrt{4\ln2 \cdot 2^{-N^\beta}} = O(2^{-N^{\beta'}}),
\end{aligned} \tag{C.1}$$

where equality (a) follows from [8, Equation (56)] and $Q(y^{1:N}|u^{1:N}) = P(y^{1:N}|u^{1:N})$. $D(\cdot||\cdot)$ in the inequality (b) is the relative entropy, and this inequality holds because of the Pinsker's inequality. Then we have

$$\begin{aligned}
E[P_e] &= Q_{U^{1:N},Y^{1N}}(\mathcal{E}) \\
&\leq ||Q_{U^{1:N},Y^{1:N}} - P_{U^{1:N},Y^{1:N}}|| + P_{U^{1:N},Y^{1:N}}(\mathcal{E}) \\
&\leq ||Q_{U^{1:N},Y^{1:N}} - P_{U^{1:N},Y^{1:N}}|| + \sum_{i\in\mathcal{I}} P_{U^{1:N},Y^{1:N}}(\mathcal{E}_i),
\end{aligned} \tag{C.2}$$

where

$$\begin{aligned} P_{U^{1:N},Y^{1:N}}(\mathcal{E}_i) &\leq \sum_{u^{1:N},y^{1:N}} P(u^{1;i-1},y^{1:N})P(u^i|u^{1:i-1},y^{1:N})\cdot \quad [P(u^i|u^{1:i-1},y^{1:N}) \\ &\leq P(u^i\oplus 1|u^{1:i-1},y^{1:N})] \\ &\leq \sum_{u^{1:N},y^{1:N}} P(u^{1;i-1},y^{1:N})P(u^i|u^{1:i-1},y^{1:N})\sqrt{\frac{P(u^i\oplus 1|u^{1:i-1},y^{1:N})}{P(u^i|u^{1:i-1},y^{1:N})}} \\ &= Z(U^i|U^{1:i-1},Y^{1:N}) \leq 2^{-N^{\beta}}. \qquad \text{(C.3)} \end{aligned}$$

From (C.1), (C.2) and (C.3), we have $E[P_e] = O(2^{-N^{\beta'}})$ for any $\beta' < \beta < 0.5$. □

APPENDIX D

Proof of Theorem 4.4

Proof. Let $\mathcal{E}_i$ denote the set of triples of $u_2^{1:N}$, $x_1^{1:N}$ and $y^{1:N}$ such that decoding error occurs at the i-th bit, then the block decoding error event is given by $\mathcal{E} \equiv \bigcup_{i\in\mathcal{I}} \mathcal{E}_i$. According to our encoding scheme, each codeword $u_2^{1:N}$ appears with probability

$$2^{-(|\mathcal{I}_2|+|\mathcal{F}_2|)} \prod_{i\in\mathcal{S}_2} P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(u_2^i|u_2^{1:i-1},x_1^{1:N}).$$

Then the expectation of decoding error probability over all random mapping is expressed as

$$\begin{aligned} E[P_e] = & \sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} 2^{-(|\mathcal{I}_2|+|\mathcal{F}_2|)}(\prod_{i\in\mathcal{S}_2} P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(u_2^i|u_2^{1:i-1},x_1^{1:N})) \\ & \cdot P_{Y^{1:N},X_1^{1:N}|U_2^{1:N}}(y^{1:N},x_1^{1:N}|u_2^{1:N}) \ \ [(u_2^{1:N},x_1^{1:N},y^{1:N}) \in \mathcal{E}]. \end{aligned}$$

Now we define the probability distribution $Q_{U_2^{1:N},X_1^{1:N},Y^{1:N}}$ as

$$\begin{aligned} Q_{U_2^{1:N},X_1^{1:N},Y^{1:N}}(u_2^{1:N},x_1^{1:N},y^{1:N}) =& 2^{-(|\mathcal{I}_2|+|\mathcal{F}_2|)} \cdot Q_{X_1^{1:N}}(x_1^{1:N}) \\ & (\prod_{i\in\mathcal{S}_2} P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(u_2^i|u_2^{1:i-1},x_1^{1:N})) \cdot P_{Y^{1:N}|X_1^{1:N},U_2^{1:N}}(y^{1:N}|u_2^{1:N},x_1^{1:N}). \end{aligned}$$

Then the variational distance between $Q_{U_2^{1:N},X_1^{1:N},Y^{1:N}}$ and $P_{U_2^{1:N},X_1^{1:N},Y^{1:N}}$ can be bounded as

$$
\begin{aligned}
&2\|Q_{U_2^{1:N},X_1^{1:N},Y^{1:N}} - P_{U_2^{1:N},X_1^{1:N},Y^{1:N}}\| = \sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(u_2^{1:N},x_1^{1:N},y^{1:N}) - P(u_2^{1:N},x_1^{1:N},y^{1:N})| \\
&= \sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(u_2^{1:N}|x_1^{1:N})Q(x_1^{1:N})Q(y^{1:N}|u_2^{1:N},x_1^{1:N}) - P(u_2^{1:N}|x_1^{1:N})P(x_1^{1:N})P(y^{1:N}|u_2^{1:N},x_1^{1:N})| \\
&\overset{(a)}{\leq} \sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(u_2^{1:N}|x_1^{1:N}) - P(u_2^{1:N}|x_1^{1:N})|P(x_1^{1:N})P(y^{1:N}|u_2^{1:N},x_1^{1:N}) \\
&+ \sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(x_1^{1:N}) - P(x_1^{1:N})|Q(u_2^{1:N}|x_1^{1:N})P(y^{1:N}|u_2^{1:N},x_1^{1:N})
\end{aligned}
$$

where inequation (a) follows from [8, Equation (56)], $Q(y^{1:N}|u_2^{1:N},x_1^{1:N}) = P(y^{1:N}|u_2^{1:N},x_1^{1:N})$. For the first summation, following the same fashion as the proof of Theorem 4.3, we can prove

$$\sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(u_2^{1:N}|x_1^{1:N}) - P(u_2^{1:N}|x_1^{1:N})|P(x_1^{1:N})P(y^{1:N}|u_2^{1:N},x_1^{1:N}) \leq 2N\sqrt{4\ln2 \cdot 2^{-N^\beta}}.$$

According to the result of the coding scheme for level 1, we already have

$$2\|Q_{U_1^{1:N},Y^{1:N}} - P_{U_1^{1:N},Y^{1:N}}\| \leq 2N\sqrt{4\ln2 \cdot 2^{-N^\beta}}.$$

Since we have $P_{Y^{1:N}|U_1^{1:N}} = Q_{Y^{1:N}|U_1^{1:N}}$, we can write

$$2\|Q_{U_1^{1:N}} - P_{U_1^{1:N}}\| \leq 2N\sqrt{4\ln2 \cdot 2^{-N^\beta}}.$$

Clearly, there is a one to one mapping between $U_1^{1:N}$ and $X_1^{1:N}$, then we immediately have $2\|Q_{X_1^{1:N}} - P_{X_1^{1:N}}\| \leq 2N\sqrt{4\ln2 \cdot 2^{-N^\beta}}$. Therefore, for the second summation,

$$
\begin{aligned}
&\sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(x_1^{1:N}) - P(x_1^{1:N})|Q(u_2^{1:N}|x_1^{1:N})P(y^{1:N}|u_2^{1:N},x_1^{1:N}) \\
&\qquad = \sum_{x_1^{1:N}} |Q(x_1^{1:N}) - P(x_1^{1:N})| \leq 2N\sqrt{4\ln2 \cdot 2^{-N^\beta}}.
\end{aligned}
$$

Then we have $||Q_{U_2^{1:N},X_1^{1:N},Y^{1^N}} - P_{U_2^{1:N},X_1^{1:N},Y^{1^N}}|| \leq 4N\sqrt{4\text{ln}2 \cdot 2^{-N^{\beta}}}$, and

$$\begin{aligned} E[P_e] &= Q_{U_2^{1:N},X_1^{1:N},Y^{1^N}}(\mathcal{E}) \\ &\leq \|Q_{U_2^{1:N},X_1^{1:N},Y^{1^N}} - P_{U_2^{1:N},X_1^{1:N},Y^{1^N}}\| + P_{U_2^{1:N},X_1^{1:N},Y^{1^N}}(\mathcal{E}) \\ &\leq \|Q_{U_2^{1:N},X_1^{1:N},Y^{1^N}} - P_{U_2^{1:N},X_1^{1:N},Y^{1^N}}\| + \sum_{i \in \mathcal{I}} P_{U_2^{1:N},X_1^{1:N},Y^{1^N}}(\mathcal{E}_i), \end{aligned}$$

The rest part of the proof follows the same fashion of the proof of Theorem 4.3. Finally we have $E[P_e] \leq N2^{-N^{\beta'}}$ for any $\beta' < \beta < 0.5$. □