# Polar Codes and Polar Lattices for Efficient Communication and Source Quantization

Ling Liu

A Dissertation 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

2016

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

# Statement of Originality

Substantial parts of this thesis are believed to be original contributions to the field of lattice coding thoery. 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:

- **Explicit construction of capacity-achieving polar lattices for AWGN channels. Especially on the implementation of lattice Gaussian shaping for the AWGN-good polar lattices.**
- **Explicit construction of Polar lattices which achieve the strong secrecy capacity of Gaussian wiretap channels.**
- **Rate-distortion bound achieving polar lattices for Gaussian sources.**
- **Solving the Gaussian Wyner-Ziv problem and the Gaussian Gelfand-Pinsker problem through nested polar lattices.**
- **Extension of polar lattices for the general Poltyrev capacity of i.i.d. fading channels.**

# Abstract

In the past several decays, lattice codes played an important role in coding theory and information theory [1]. Lattice codes with good performance in communication and source compression have attracted considerable interest. A typical method of constructing good lattice codes is to use existing linear codes. For instance, the famous Barnes-Wall lattices are generated by Reed-Muller (RM) codes, and more recently, the emerging low density Construction-A (LDA) lattices are resulted from low density parity check (LDPC) codes.

In this thesis, we develop a new class of lattices, called polar lattices, based on polar codes. The invention of polar codes is considered to be one of the major breakthroughs in coding theory for the past ten years. We show that polar lattices provide explicit solutions for many interesting problems in information theory. For channel coding, we prove that polar lattices are capable of achieving the capacity of the additive white Gaussian noise (AWGN) channel. For the dual side, i.e., source compression, polar lattices can also achieve the rate-distortion bound for the independent and identically distributed (i.i.d.) Gaussian source. Moreover, a combining design of polar lattices for both channel coding and source coding gives us explicit solutions to the Gaussian version of the Wyner-Ziv and Gelfand-Pinsker problems. For physical layer security, we prove that polar lattices are able to approach the secrecy capacity of the Gaussian wiretap channel under the strong secrecy criterion. Two more applications of polar lattices are achieving the capacity of the i.i.d. fading channel and extracting the common information of two joint Gaussian sources.

The explicit construction of polar lattices provides us better insights on the study of lattice coding. Many interesting problems of lattice coding, such as AWGN-

goodness, secrecy-goodness, lattice shaping, and lattice Gaussian distribution will be addressed from the perspective of polar lattices.

*[redacted] my family and my little boy pi*

*for their endless love and support*

# Acknowledgments

I would like to thank my supervisor Dr. Cong Ling for his continuous support throughout the research. Cong has provided me with patient guidance, necessary resources and freedom to conduct my research during my Ph.D study. His passion for his work and optimistic attitudes towards life always inspires me and encourages me to meet the challenges encountered and move forward. I am so lucky to have a supervisor and mentor who cares about my work and my life and responds to my inquiries and questions so promptly. I would also like to extend my sincere thanks to Dr. Yanfei Yan and Dr. Zheng Wang, who spent a lot of time on helping me with my research.

I would like to thank all my colleagues in the Communications and Signal Processing Group. It has been a great pleasure to work with you for the past four years. I would also like to express my appreciation to the China Scholarship Council and Huawei Technologies Co. Ltd. Throughout my Ph.D, I was funded by CSC Imperial Scholarship and the HIRP Flagship Program by Huawei's Shield Lab.

I am deeply grateful to my parents for your tolerance, understanding, love, and faith. Thank you for encouraging me to pursue a Ph.D in Great Britain and giving me the motivation, strength, and determination to complete my Ph.D.

I want to thank my loving and caring wife Emily, who is also a Ph. D student at Imperial College London. You always inspire me when I feel discouraged. I have enjoyed every moment we spent together in the past four years, as part of this special journey in my life and I am grateful you are by my side. Thank you for your patience, trust, encouragement and tolerance. At last I want to thank my lovely son pi, you are the best thing ever happen to me. Thank you for all the joy you bring to me and

making my life meaningful and complete.

# Contents

**Statement of Originality** 3

**Abstract** 4

**Acknowledgments** 7

**Publications** 13

**Abbreviations** 20

**Notations** 22

**1 Introduction** 1

1.1 Overview . . . . . 1

1.2 Motivation and Aims . . . . . 2

1.3 Organization of the Thesis . . . . . 4

1.4 Statement of Originality . . . . . 4

**2 Preliminaries for Polar Codes** 6

2.1 Polar Codes for Channel Coding . . . . . 7

2.2 Polar Codes for Lossless Compression [2] . . . . . 13

2.3 Polar Codes for Lossy Compression . . . . . 17

2.4 Polar Codes for Wiretap Coding . . . . . 23

**3 From Polar Codes to Polar Lattices** 29

3.1 Latices and lattice codes . . . . . 29

3.1.1 Lattices . . . . . 30

3.1.2 Lattice Partition . . . . . 31

3.1.3 Lattice Codes . . . . . 34
3.2 AWGN-Goodness and Lattice Gaussian Distribution . . . . . 35
3.2.1 AWGN-goodness of Lattices . . . . . 35
3.2.2 Lattice Gaussian Distribution and Flatness Factor . . . . . 37
3.2.3 Existence of Capacity-Achieving Lattices . . . . . 39
3.3 Lattices constructed from error correction codes . . . . . 43
3.3.1 Single level constructions . . . . . 44
3.3.2 Multilevel constructions . . . . . 46
3.3.3 Other lattices . . . . . 52
3.4 Summary . . . . . 52

**4 Polar Lattices for AWGN Channels 54**
4.1 Introduction . . . . . 54
4.2 Construction of Polar Lattices . . . . . 56
4.2.1 Forney *et al.*'s Construction Revisited . . . . . 56
4.2.2 Polar Lattices . . . . . 60
4.2.3 AWGN Goodness . . . . . 62
4.3 Polarization-Based Lattice Gaussian Shaping . . . . . 64
4.3.1 Asymmetric Channels in Multilevel Lattice Coding . . . . . 64
4.3.2 Polar Codes for Asymmetric Channels . . . . . 67
4.3.3 Multilevel Polar Codes . . . . . 73
4.3.4 Achieving Channel Capacity . . . . . 78
4.4 Design Examples . . . . . 81
4.4.1 Design Examples without Power Constraint . . . . . 81
4.4.2 Design Examples with Power Constraint . . . . . 85
4.5 Summary . . . . . 88

**5 Polar Lattices for Gaussian Wiretap Channels 89**

5.1 Introduction . . . 89

5.2 Secrecy-good Polar Lattices for the Mod-$\Lambda_s$ GWC . . . 93

5.2.1 Secrecy-goodness . . . 94

5.2.2 Reliability over Mod-$\Lambda_s$ GWC . . . 104

5.3 Secrecy-good polar lattices with discrete Gaussian shaping . . . 106

5.3.1 Gaussian shaping over polar lattices . . . 107

5.3.2 Three-dimensional partition . . . 108

5.3.3 Strong secrecy . . . 111

5.3.4 Semantic security . . . 116

5.3.5 Reliability over GWC . . . 121

5.4 Summary . . . 127

**6 Polar Lattices for Quantization 129**

6.1 Introduction . . . 129

6.2 Polar Lattices for Gaussian Sources . . . 134

6.3 Simulation Results . . . 142

6.4 Gaussian Wyner-Ziv Coding . . . 143

6.4.1 System model . . . 143

6.4.2 A solution using continuous auxiliary variable . . . 144

6.4.3 A practical solution using lattice Gaussian distribution . . . 146

6.5 Gaussian Gelfand-Pinsker coding . . . 154

6.6 Polar Lattices for Wyner's Common Information . . . 161

6.6.1 Gray-Wyner Network and Common Information . . . 161

6.6.2 Common Information for Joint Gaussian Sources . . . 162

6.7 Summary . . . 169

**7 Polar Codes and Polar Lattices for Independent Fading Channels 170**

7.1 Introduction . . . . . 170

7.2 Polar Codes for Binary-Input Fading Channels . . . . . 173

7.3 Polar Lattices for i.i.d. Fading Channels . . . . . 179

7.3.1 Polar Lattices for Fading Channels without Power Constraint 180

7.3.2 Polar Lattices with Gaussian Shaping . . . . . 188

7.4 Summary . . . . . 193

**8 Conclusions and Future Work 195**

**A Proof of Lemma 4.2.2 200**

**B Proof of Lemma 4.3.1 202**

**C Proof of Theorem 4.3.7 205**

**D Proof of Theorem 4.3.9 208**

**E Proof of Lemma 5.2.2 211**

**F Proof of Theorem 6.2.3 212**

**G Proof of Theorem 6.2.4 215**

**H Proof of Theorem 6.2.5 217**

**I Proof of Theorem 6.4.4 221**

**J Proof of Theorem 6.5.1 226**

**Bibliography 230**

# Publications

The following publications support the material given in this thesis.

## Journal Papers

1. Y. Yan, **L. Liu**, C. Ling and X. Wu, “Construction of Capacity-Achieving Lattice Codes: Polar Lattices,” *submitted to IEEE Trans. Inform. Theory, 2013*.

2. **L. Liu**, Y. Yan, and C. Ling, “Achieving Secrecy Capacity of the Gaussian Wiretap Channel with Polar Lattices," *submitted to IEEE Trans. Inform. Theory, 2014*.

3. **L. Liu**, and C. Ling, “Polar Lattices for Lossy Compression," *submitted to IEEE Trans. Inform. Theory, 2015*.

4. **L. Liu**, and C. Ling, “Polar Codes and Polar Lattices for Independent Fading Channels," *accepted by IEEE Trans. Communications*.

5. J. Shi, **L. Liu** and C. Ling, “Extracting Wyner’s Common Information Using Polar Codes and Polar Lattices," *submitted to IEEE Trans. Inform. Theory, 2016*.

## Conference Proceedings

1. Y. Yan, **L. Liu**, and C. Ling, “Polar Lattices for Strong Secrecy over the Mod-$\Lambda$ Gaussian Wiretap Channel," *Proc. IEEE Int. Symp. Inform. Theory*, 2014.

2. **L. Liu**, Y. Yan, and C. Ling, “Secrecy-Good Polar Lattices with Optimal Shaping for the Gaussian Wiretap Channels," *IEEE Inform. Theory Workshop (Israel)*, 2015.

3. **L. Liu**, and C. Ling, “Polar Lattices are Good for Lossy Compression," *Proc. IEEE Inform. Theory Workshop (South Korea)*, 2015.

4. **L. Liu**, and C. Ling, “Polar Codes and Polar Lattices for Independent Fading Channels," *Proc. IEEE Int. Symp. Inform. Theory*, 2016.

5. **L. Liu**, J. Shi, and C. Ling, “Extracting Wyner’s Common Randomness Using Polar Codes," *Proc. IEEE Wireless Commun. Signal Process.*, 2016.

# List of Figures


2.1 The recursive construction of polar codes based of $G_2$, $G_4$, and $G_N$. 7

2.2 Plot of $I(W_N^{(i)})$ versus $i = 1, ..., N = 2^{12}$ for a binary erasure channel (BEC) with erasure probability $\epsilon = 0.5$. . . . . 9

2.3 Example of a factor graph for a polar code with block length of $2^3 = 8$. 14

2.4 Result of the numerical computation of $E[R]$ for source entropies of 0.9, 0.5 and 0.1. . . . . 17

2.5 Simulation results for the polar source encoder for $n = 10, 12, 14, 16, 18$. 18

2.6 The probabilistic model used for source coding. The source word $\bar{Y}$ is treated as the output of a channel. The code is defined through its frozen set $F = 0, 1, 2$, The codewords are indexed using $U_3^7$. As we will see later, the frozen bits for the lossy source coding problem match with the frozen bits of the channel coding problem, with slight difference in the unpolarized region, which is of vanishing proportion. 20

2.7 The rate-distortion performance for the SC encoding algorithm with randomized rounding for $n = 10, 12, 14, 16, 18$. . . . . 22

3.1 Lattice partitions induced by the $\mathbb{Z}^4/2\mathbb{Z}^4$ . . . . . 31

3.2 The encoding and decoding system of the $D_4$ lattice. The binary code is the Reed-Muller code with $N = 4$ and $k = 3$. . . . . 33

3.3 Geometric picture of lattices [1]. . . . . 35

3.4 The comparison of the $\Lambda$-aliased Gaussian distributions with different flatness factors. . . . . 38

4.1 Lattice Gaussian distribution $D_{\mathbb{Z},\sigma_s}$ and the associated labelling. . . 65

4.2 The relationship between the asymmetric channel $W$ and the symmetrized channel $\tilde{W}$. . . . . 69

4.3 Polarization for symmetric and asymmetric channels. . . . . 71

4.4 The first step of polarization in the construction for the second level. 74

4.5 The second step of polarization in the construction for the second level. 75

4.6 Channel capacity for partition chain $\mathbb{Z}/2\mathbb{Z}/\cdots/2^r\mathbb{Z}$. The discrete approximation used in Sect. 4.2.2 is also shown. . . . . 81

4.7 A polar lattice with two levels, where $\sigma_1 = \tilde{\sigma}$. . . . . 82

4.8 Block error probabilities of polar lattices and Barnes-Wall (BW) lattices of length $N = 1024$ with multistage decoding. . . . . 83

4.9 SER of lattices with dimension around 1000. . . . . 85

4.10 Channel capacity for each level as a function of SNR. . . . . 86

4.11 The proportions of the shaping set, information set, and frozen set on each level when $N = 2^{16}$ and $\mathsf{SNR} = 15$ dB. . . . . 87

4.12 Lower bounds on the rates achieved by polar lattices with block error probability $5 \times 10^{-5}$ for block lengths $2^{10}$, $2^{12}$, ..., $2^{20}$. . . . . 87

5.1 The Gaussian wiretap channel. . . . . 91

5.2 The mod-$\Lambda_s$ Gaussian wiretap channel. . . . . 95

5.3 The multilevel lattice coding system over the mod-$\Lambda_s$ Gaussian wiretap channel. . . . . 99

5.4 Partitions of the index set $[N]$ with shaping. . . . . 109

5.5 Block diagram of the shaping-induced channel $\mathcal{Q}_N(W, \mathcal{S})$. . . . . 111

5.6 Block diagram of the new induced channel $\mathcal{Q}_N(W, \mathcal{S}, \mathcal{R})$. . . . . 112

5.7 Markov block coding scheme without sharing the secret mapping. . 124

6.1 The relationship between the quantization of source $Y'$ and $Y$. . . . 136

6.2 Quantization performance of polar lattices for the Gaussian source with $\sigma_s = 3$. . . . . 142

6.3 Variational distance $\mathbb{V}(f_{Y'}, f_Y)$ and quantization noise for polar lattice quantizer. . . . . 143

6.4 Wyner-Ziv coding for the Gaussian case. The variances of $Z$, $Y$ and $X$ are given by $\sigma_z^2$, $\sigma_y^2$ and $\sigma_x^2 = \sigma_y^2 + \sigma_z^2$, respectively. . . . . 144

6.5 A solution of the Gaussian Wyner-Ziv problem using a continuous Gaussian random variable $X'$. . . . . 145

6.6 The MMSE rescaled channel blocks $(a)$ and $(b)$ for the Gaussian channels $X \to X'$ and $Y \to X'$, respectively. . . . . 146

6.7 A reverse solution of the Gaussian Wyner-Ziv problem, which is more compatible with lattice Gaussian distribution. . . . . 148

6.8 The partitions of $U_\ell^{1:N}$ for quantization lattice $L_1$ (left) and channel coding lattice $L_2$ (right). $\mathcal{F}_\ell^Q \subseteq \mathcal{F}_\ell^C$, $\mathcal{I}_\ell^C \subseteq \mathcal{I}_\ell^Q$, and $\mathcal{S}_\ell^Q \subseteq \mathcal{S}_\ell^C$. Without the side information, $U_\ell^{\mathcal{I}_\ell^Q}$ should be sent to achieve the target distortion. With the side information, however, $U_\ell^{\mathcal{I}_\ell^C}$ can be decoded and hence only $U_\ell^{d\mathcal{I}_\ell}$ need to be sent. . . . . 149

6.9 Gelfand-Pinsker coding for the Gaussian case. . . . . 154

6.10 A solution of the Gaussian Gelfand-Pinsker problem using continuous Gaussian random variable $S'$. . . . . 156

6.11 A reverse solution of the Gaussian Gelfand-Pinsker problem. . . . . 156

6.12 The partitions of $U_\ell^{1:N}$ for quantization lattice $L_2$ (left) and channel coding lattice $L_1$ (right). $\mathcal{F}_\ell^C \subseteq \mathcal{F}_\ell^Q$ and $\mathcal{S}_\ell^Q \subseteq \mathcal{S}_\ell^C$. Without the power constraint, $U_\ell^{\mathcal{I}_\ell^C}$ can be sent as message bits. With the power constraint, however, $U_\ell^{\mathcal{I}_\ell^Q}$ should be selected according to the interference $S^{1:N}$ and hence only $U_\ell^{d\mathcal{F}_\ell}$ can be fed with the message bits. 158

6.13 Gray-Wyner source coding network. . . . . 162

7.1 Binary-input ergodic fading channel with CSI available at the end of receiver. . . . . 173

7.2 Typical boundaries of $A_i$ for channel quantization. . . . . 176

7.3 Comparisons between $C(\tilde{W}_Q)$ and $C(\tilde{W})$, and between $C(\tilde{V}_Q)$ and $C(\tilde{V})$, when $Q = 128$. Here $\tilde{W}$ denotes the channel $X \to (Y, H)$, and $\tilde{V}$ denotes the channel $X \to Y$, i.e., the two channel models when the receiver knows the CSI and the CDI, respectively. $\tilde{W}_Q$ and $\tilde{V}_Q$ denote the quantized version of $\tilde{W}$ and $\tilde{V}$, respectively. . . . . 177

7.4 Performance of polar codes for the Rayleigh fading channel with CSI available to the receiver when $N = 2^{10}, 2^{11}, ..., 2^{14}$. . . . . 178

7.5 Performance of polar codes for the Rayleigh fading channel with CDI available to the receiver when $N = 2^{10}, 2^{11}, ..., 2^{14}$. . . . . 179

7.6 A block diagram of the fading mod-$\Lambda$ channel. . . . . 181

7.7 Performance of polar codes for the $\mathbb{Z}/2\mathbb{Z}$, $2\mathbb{Z}/4\mathbb{Z}$, $4\mathbb{Z}/8\mathbb{Z}$ and $8\mathbb{Z}/16\mathbb{Z}$ fading channels with $\sigma = 1$, $\sigma_h = 1.2575$ and $N = 2^{14}$. The capacities of these four channels are about 0.1172, 0.4929, 0.8200 and 0.9500, respectively. FER denotes the frame (block) error probability, and BER denotes the bit error probability. . . . . 185

7.8 The upper-bounds of the block error probability of polar lattices under the SC decoding when $\sigma_s = 3$, $\sigma = 1$, $\sigma_h = 1.2575$ and $N = 2^{10}, 2^{12}, ..., 2^{20}$. . . . . 193

A.1 Let $X \in \mathcal{R}(\Lambda_{\ell+1})$ denote the channel input. Let $N_1$ and $N_2$ denote two independent additive Gaussian noise with variances $\sigma_1^2$ and $\sigma_2^2 - \sigma_1^2$, respectively. Clearly, the two $\Lambda_\ell/\Lambda_{\ell+1}$ channels with noise variances $\sigma_1^2$ and $\sigma_2^2$ can be described by channel (a) and (b), respectively. By the property of modulo operation, channel (b) is equivalent to channel (c), which is a concatenated channel made by concatenating channel (a) with an intermediate mod-$\Lambda_{\ell+1}$ channel. . 201

# List of Tables

3.1 Lattice constructions from error correction codes [3]. . . . . 43

6.1 Performance comparison with TCQ for Gaussian source (SNR in dB). 143

# Abbreviations

| | |
|---|---|
| **AWGN** | Additive White Gaussian Noise |
| **BEC** | Binary Erasure Channel |
| **BMA** | Binary Memoryless Asymmetric |
| **BMS** | Binary Memoryless Symmetric |
| **BP** | Belief-Propogation |
| **BSC** | Binary Symmetric Channel |
| **BW** | Barnes-Wall |
| **CDI** | Channel Distribution Information |
| **CSI** | Channel State Information |
| **ECDQ** | Entropy-Coded Dithered Quantization |
| **GWC** | Gaussian Wiretap Channel |
| **IC** | Infinite Constellation |
| **I.I.D.** | Independent and Identically Distributed |
| **LDA** | Low-Density Construction-A lattice |
| **LDLC** | Low-Density Lattice Codes |
| **LDPC** | Low-Density Parity Check |
| **MIMO** | Multiple-Input and Multiple-Output |
| **ML** | Maximum Likelihood |
| **MMSE** | Minimum Mean-Square Error |
| **NLD** | Normalized Logarithmic Density |
| **PDF** | Probability Density Function |

**QKD** Quantum Key Distribution

**RV** Random Variable

**SC** Successive Cancellation

**SER** Symbol Error Rate

**SNR** Signal-to-Noise Ratio

**TCM** Trellis Codes Modulation

**TCQ** Trellis Codes Quantization

**VNR** Volume-to-Noise Ratio

**VQ** Vector Quantization

# 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 $X$ |
| $X_\ell$ | a RV $X$ at level $\ell$ in the multilevel coding scheme |
| $x_\ell^i$ | an $i$-th realization of $X_\ell$ |
| $x_\ell^{i:j}$ | 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 $\ell$ 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 $\mathcal{I}$ |
| $[N]$ or $1:N$ | 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 $\ell$-th level |
| $W_{\ell,N}^{(i)}$ | an $i$-th subchannel for the $\ell$-th level |
| $\mathsf{SNR}_b$ ($\mathsf{SNR}_e$) | SNR of the main (wiretap) channel |

CHAPTER 1

# Introduction

## 1.1 Overview

Polar codes [4] are regarded as the first kind of error correction codes which can be proved to achieve the capacity of binary-input memoryless symmetric (BMS) channels. The key idea behind polar codes in the context of channel coding is the channel polarization technique, which converts two i.i.d. copies of a general noisy symmetric channel into a noisier synthetic channel and a less noisy synthetic channel. By increasing the block length, the synthetic channels get more polarized and eventually become either noiseless or useless channels. Efficient construction methods of polar codes for classical BMS channels such as binary erasure channels (BECs), binary symmetric channels (BSCs), and binary-input additive white Gaussian noise (BAWGN) channels were proposed in [5, 6, 7]. Besides channel coding, polar codes were then extended to source coding and their asymptotic performance was proved to be optimal [8, 9]. As a combination of the application of polar codes for channel coding and lossless source coding, polar codes were further studied for binary-input memoryless asymmetric channels (BMACs) in [10, 11, 12]. The versatility of polar codes makes them attractive and promising for coding over many other channels, such as wiretap channels [13], broadcast channels [14], multiple access channels (MACs) [15], compound channels [16, 17] and even quantum channels [18].

As the counterpart of linear codes in the Euclidean space, lattice codes provide more freedom over signal constellation for communication systems. The existence

of lattice codes achieving the point-to-point additive white Gaussian noise (AWGN) channel capacity was established in [19, 20]. Besides point-to-point communications, lattice codes are also useful in a wide range of applications in multiterminal communications, such as information-theoretical security [21], compute-and-forward [22], distributed source coding [23], and $K$-user interference channel [24] (see [1] for an overview).

There has been a long history of using error correction codes to construct good lattice codes. We will give some examples in Chapter 3. The invention of polar codes provides us a new method to construct lattice codes with explicit structure. In simple words, a polar lattice is generated by a series of nested polar codes according to the "Construction D" [25]. We will show that polar lattices inherit the versatility of polar codes in the Euclidean space, namely they are able to achieve the AWGN channel capacity with certain power constraint and also approach the rate-distortion bound for a Gaussian source.

Besides data transmission and source compression, the issues of data confidentiality and security have taken on an increasingly important role in current communication systems. At present, the widely used cryptographic protocols are designed and implemented assuming the physical layer has already been established and provides an error-free link. The wireless network especially needs the physical layer security due to its broadcast nature of wireless medium. We will see that polar lattices have great potential to offer such security in the context of Gaussian wiretap channels.

## 1.2 Motivation and Aims

Although much work has been carried out to find good lattice codes with optimal asymptotic properties, most of them are based on random lattice coding arguments.

The recently proposed LDA lattices have been proved to exhibit such optimal asymptotic properties under belief prorogation (BP) decoding. Their construction is still not completely explicit. Similar to LDPC codes, the proof of LDA lattices are based on random ensembles. The aim of this thesis is to find explicit lattice coding solutions to some of the key problems in information theory.

Since polar codes were originally proposed for binary-input symmetric channels, as a direct extension, our first step is to design polar lattices for the AWGN channel with a certain power constraint. This work includes the construction of the AWGN-good polar lattices and the implementation of lattice Gaussian shaping. The designed polar lattices for AWGN channels can then be used in the Gaussian wiretap channel to guarantee efficient and reliable transmission for the legitimate receiver. Our second step is to design the secrecy-good polar lattices to prevent the eavesdropper from intercepting the information about the secure message. Since lattice codes can also be used to compress continuous sources, our third step is to design polar lattices which are good for source compression. The aim of this work is not just to find the rate-distortion bound achieving lattices through the polarization technique, but also to solve the Gaussian Wyner-Ziv coding problem and its dual problem, the Gaussian Gelfand-Pinsker channel coding problem. Another application of the quantization polar lattices is to extract the common information of two joint Gaussian sources. Our final step is to return to channel coding and design polar lattices for the i.i.d. fading channels. This work can be viewed as a generalization of the work for the AWGN channel, since the AWGN channel is a special i.i.d. fading channel with constant fading gain.

## 1.3 Organization of the Thesis

We focus on the construction of polar lattices for efficient communication and source quantization. The organization of the rest of the thesis is as follows. In chapter 2, we provide the background of polar codes for channel coding, wiretap coding and source coding. Chapter 3 presents the preliminaries for lattice codes. Chapter 4 is devoted to the construction of the AWGN-good polar lattices and the implementation of the lattice Gaussian shaping to achieve the AWGN capacity. The secrecy-good polar lattices are discussed in Chapter 5, where we prove that polar lattices achieve the secrecy capacity of the Gaussian wiretap channel under the strong secrecy criterion. We then move source quantization in Chapter 6, where we show how to achieve rate-distortion bound for Gaussian sources. A combining scheme of polar lattices for channel coding and source quantization gives solutions to the Gaussian Wyner-Ziv coding problem and the Gaussian Gelfand-Pinsker coding problem. We will also give a brief introduction on how to use polar lattices to extract the common information of joint Gaussian sources. The extension of polar lattices for i.i.d. fading channels is given in Chapter 7.

## 1.4 Statement of Originality

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

- **Explicit construction of capacity-achieving polar lattices for AWGN channels. Especially on the implementation of lattice Gaussian shaping for the AWGN-good polar lattices.**

- **Explicit construction of Polar lattices which achieve the strong secrecy capacity of Gaussian wiretap channels.**
- **Rate-distortion bound achieving polar lattices for Gaussian sources.**
- **Solving the Gaussian Wyner-Ziv problem and the Gaussian Gelfand-Pinsker problem through nested polar lattices.**
- **Extension of polar lattices for the general Poltyrev capacity of i.i.d. fading channels.**

CHAPTER 2

# Preliminaries for Polar Codes

Explicit construction method of provably capacity-achieving codes has been an essential challenge in information theory since Shannon only proved the existence of such codes, without presenting any specific example [26]. To solve this problem, many error correction codes with good practical performance were invented, such as Reed-Solomon (RS) codes, low density parity check (LDPC) codes, and convolutional codes. Although some of these codes have been shown to be capable of reaching the Shannon limit from their simulation results, a theoretical proof of such property is still lacking. In 2008, Arıkan proposed polar codes [4] and proved that they can achieve the capacity of any given binary-input discrete memoryless symmetric (BDMS) channel with low encoding and decoding complexity. We also note that the construction of polar codes was independently described by N. Stolte earlier in [27], without developing the theory of channel polarization. Being considered as the first kind of codes with such provable property, polar codes have spurred much recent attention. These codes have been investigated in many areas of classical information theory, including their efficient constructions [6, 28], and the bounds on error probabilities [29]. After that, polar codes have also been generalized to lossless and lossy source coding [30, 31], with a similar asymptotical property.

Moreover, in terms of communication security, polar codes are proved to achieve the strong secrecy for binary symmetric wiretap channels, which provides a wider range of applications for the polarization technique. It was then proved that the channel polarization also exists in quantum communication channels, which makes

it quite potential to construct quantum polar codes. Wilde, Renes and Guha showed that polar codes can be extended to different quantum channels such as classical-quantum channels, private classical channels, quantum erasure channels, and quantum depolarizing channels [32, 33, 34]. Our work will focus on the classical polar coding techniques and we will give a brief introduction of them in this chapter.

## 2.1 Polar Codes for Channel Coding

Since polar codes were originally proposed as a type of capacity achieving codes for BDMS channels, for convenience, we firstly introduce the concept of polar codes in the scenario of channel coding.

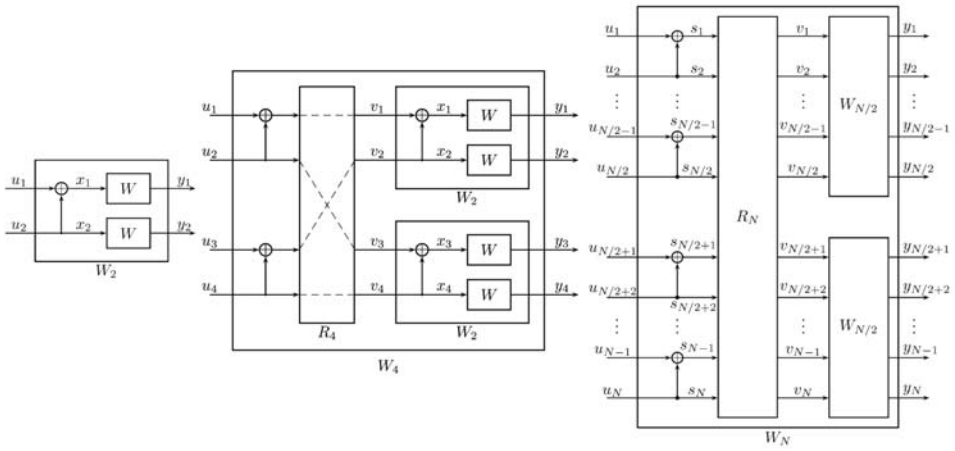



Figure 2.1: The recursive construction of polar codes based of $G_2$, $G_4$, and $G_N$.

Consider a polar code with block length $N$ (assume that $N = 2^n$ is an integer power of 2) transmitted over a BDMS channel $W$. According to [4], the main idea of channel polarization is based on an $N \times N$ generator matrix $G_N$, which is defined as

$$G_N = \mathbf{B} \cdot \mathbf{F}^{\otimes n}, \tag{2.1}$$

where $\otimes$ denotes Kronecker product, $\mathbf{F} = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$ denotes the $2 \times 2$ kernel of polar

codes and **B** denotes a bit-reverse matrix [1]. The recursive construction of polar codes based on $G_N$ is shown in Fig. 2.1. For a $W$ with input $X \in \{0,1\}$ and output $Y$, let $u_1^N = (u_1, u_2, ..., u_N)$ denote the data vector, $x_1^N$ denote the transmitted codeword and $y_1^N$ denote the received data vector, the combining channel $W_N$ can be represented as

$$W_N(y_1^N|u_1^N) = W^N(y_1^N|x_1^N) = W^N(y_1^N|u_1^N G_N), \tag{2.2}$$

where $W^N(y_1^N|x_1^N) = \prod_{i=1}^N W(y_i|x_i)$ denotes $N$ parallel uses of $W$. Let $u_i^j = (u_i, u_{i+1}, ..., u_j)$ be a subvector of $u_1^N$, the $i$-th subchannel $W_N^{(i)}$ after channel splitting is defined as

$$W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) = \frac{1}{2^{N-1}} \sum_{u_{i+1}^N} W_N(y_1^N|u_1^N), \tag{2.3}$$

treating $u_i$ as the only input and $(y_1^N, u_1^{i-1})$ as the outputs. Due to the recursive structure of polar codes, generator matrix $G_{2N}$ can be constructed by $G_N$ and the channel pair $(W_{2N}^{(2i-1)}, W_{2N}^{(2i)})$ can be calculated by $(W_N^{(i)}, W_N^{(i)})$ as

$$\begin{aligned} &W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-2}|u_{2i-1}) \\ &\quad = \sum_{u_{2i}} \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2}|u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2}|u_{2i}) \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} &W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1}|u_{2i}) \\ &\quad = \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2}|u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2}|u_{2i}), \end{aligned} \tag{2.5}$$

where $u_{1,o}^{2i-2}$ ($u_{1,e}^{2i-2}$) denotes the subsequence of $u_1^{2i-2}$ with odd (even) indices.

[1] Since **B** is a re-ordering of the indices $[1, ..., N]$, sometimes we ignore it and write $G_N = \mathbf{F}^{\otimes n}$ for simplicity

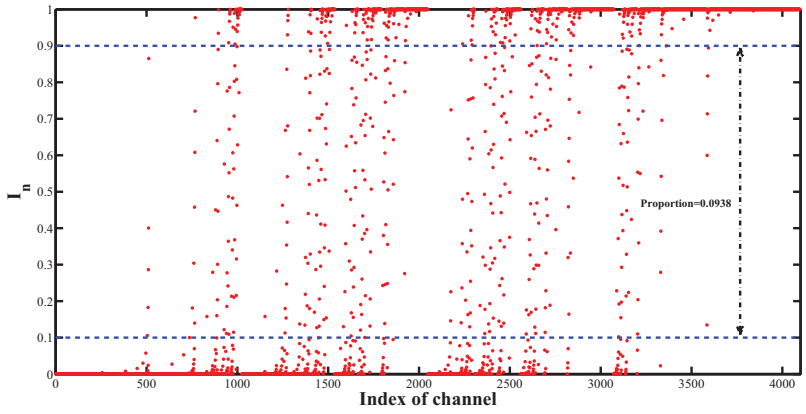



Figure 2.2: Plot of $I(W_N^{(i)})$ versus $i = 1, ..., N = 2^{12}$ for a binary erasure channel (BEC) with erasure probability $\epsilon = 0.5$.

An important property of polar codes is that the mutual information between inputs and outputs is preserved after channel combining and splitting process, namely

$$I(U_1^N; Y_1^N) = NI(W) = \sum_{i=1}^{N} I(W_N^{(i)}), \tag{2.6}$$

where $I(W)$ is the symmetric capacity of $W$. Moreover, after a single channel combining and splitting operation, we have

$$I(W_{2N}^{(2i)}) \leq I(W_N^{(i)}) \leq I(W_{2N}^{(2i+1)}), \tag{2.7}$$

which means that we obtain a better channel and a worse channel, respectively. As $N$ grows large, $I(W_N^{(i)})$ takes a value approximate to 0 or 1, and the fraction of subchannels $W_N^{(i)}$ with capacity equal to one is close to $I(W)$, which means that we can directly design a capacity-approaching code by selecting the elements of $U_1^N$ whose corresponding subchannels are with mutual information near 1 as information bits and simply freezing the rest elements (assumed to be all zero). An example of such polarization in mutual information is shown in Fig. 2.2. To sum up, we give the definition of polar codes as follows.

**Definition 2.1.1.** The polar code $C_N(F, u_F)$ of length $N = 2^n$, defined for any frozen set $F \subseteq \{1, ..., N\}$ and frozen bits $u_F \in \mathbb{F}_2^{|F|}$, is a linear code given by

$$C_N(F, u_F) = \{x_1^N = u_1^N G_N : u_{F^c} \in \mathbb{F}_2^{|F^c|}\}, \tag{2.8}$$

where $F^c$ denotes the complement of $F$ with respect to the set $\{1, ...N\}$ and $\mathbb{F}_2 = \{0, 1\}$.

Note that we can also choose the indices of subchannels according to their associated Bhattacharyya parameters, which is defined as follows.

**Definition 2.1.2** (Bhattacharyya parameter for BDMS channels)**.**

$$\tilde{Z}(W_N^{(i)}) = \sum_{y_1^N, u_1^{i-1}} \sqrt{W_N^{(i)}(y_1^N, u_1^{i-1}|0) W_N^{(i)}(y_1^N, u_1^{i-1}|1)}. \tag{2.9}$$

We note that this definition is for BDMS channels. In Chapter 4, we will introduce a more general definition of the Bhattacharyya parameter, which can also be used for asymmetric channels.

Here we conclude the construction process of polar codes as follows.

1) Given a fixed $N$, calculate the Bhattacharyya parameter of all split channels and obtain the sequence $\{\tilde{Z}(W_N^{(1)}), ..., \tilde{Z}(W_N^{(N)})\}$.
2) Given a fixed $K$, sort the sequence $\{\tilde{Z}(W_N^{(1)}), ..., \tilde{Z}(W_N^{(N)})\}$ and select a $K$-element subset of $\{1, ..., N\}$ such that $\tilde{Z}(W_N^{(i)}) \leq \tilde{Z}(W_N^{(j)})$ for all $i \in \Omega$, $j \in \Omega^c$. Note that $\Omega$ is the set consists of information index and $\Omega^c$ is the complementary set of $\Omega$.
3) Choose the information set $u_\Omega$ to be the source bits and freeze bit set $u_{\Omega^c}$ to be all-zero sequence, and obtain the codeword $x_1^N$ by $(u_\Omega, u_{\Omega^c}) \times G_N$.

Because of the assumption that the subsequence $u_1^{i-1}$ is already known for each subchannel $W_N^{(i)}$, Arıkan considered successive cancellation (SC) decoding which

processes bit by bit to achieve capacity within $O(N \log N)$ complexity. For the SC decoding of the $i$-th bit, given the decision of previous $i-1$ bits $\hat{u}_1^i$, if $i$ is in the set of frozen bits, we simply decode it as 0; if $i$ is in the set of information set, we decode it as

$$\hat{u}_i = \begin{cases} 0, & \text{if } \frac{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|0)}{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|1)} \geq 1 \\ 1, & \text{otherwise.} \end{cases} \tag{2.10}$$

The probability $W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|0)$ and $W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|1)$ can be calculated by (2.4) and (2.5) recursively. The next theorem shows that polar codes are capacity achieving under SC decoding.

**Theorem 2.1.1.** *[Polar Codes Achieve the Symmetric Capacity [35]]. Given a BMS channel $W$ and fixed $R < I(W)$, for any $0 < \beta < 1/2$ there exists a sequence of polar codes of rate $R_N > R$ such that its block error probability under the SC decoding satisfies*

$$P_B(R_N) = O(2^{-N^\beta}). \tag{2.11}$$

The following definition and remark will be important to the construction of polar lattices, which require a series of nested polar codes.

**Definition 2.1.3.** (Channel degradation): Consider two channels $W_1 : \mathcal{X} \to \mathcal{Y}_1$ and $W_2 : \mathcal{X} \to \mathcal{Y}_2$. $W_1$ is said to be (stochastically) degraded with respect to $W_2$ if there exists a distribution $Q : \mathcal{Y}_2 \to \mathcal{Y}_1$ such that

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

*Remark* 2.1.1 ([30, Lemma 4.7]). Let $\tilde{W}$ and $\tilde{V}$ be two BMS channels. If $\tilde{V}$ is degraded with respect to $\tilde{W}$, after channel polarization, we have that $\tilde{V}_N^{(i)}$ is degraded with respect to $\tilde{W}_N^{(i)}$, and $\tilde{Z}(\tilde{W}_N^{(i)}) \leq \tilde{Z}(\tilde{V}_N^{(i)})$. For any given constant

$0 < \delta < 1$, the index set $\{i \in [N] : \tilde{Z}(\tilde{V}_N^{(i)}) \leq \delta\}$ is a subset of the index set $\{i \in [N] : \tilde{Z}(\tilde{W}_N^{(i)}) \leq \delta\}$. Therefore, the polar code $C_V$ constructed according to the Bhattacharyya parameter rule for $\tilde{V}$ is a subcode of the polar code $C_W$ for $\tilde{W}$, i.e., $C_V \subseteq C_W$.

In spite of owning such good asymptotic property, polar codes still have two main drawbacks. One is their finite-length performance under SC decoding, which is far behind that of some existing error correction codes such as LDPC codes and Turbo codes. To improve the decoding performance of SC decoding, successive cancellation list (SCL) decoding is introduced [36], which makes the decoding performance of polar codes better than that of LDPC and Turbo codes with the assistance of cyclic redundancy check (CRC). Maximum likelihood (ML) decoding using Viterbi and BCJR algorithms on the codeword trellis of polar codes was proposed as well [37]. Unfortunately, this decoding method can only work on very short polar codes due to its high complexity. In [38], Arıkan also suggested that BP decoding can be utilized for polar codes based on the factor graph representation, and Hussami et. al. showed that BP decoding is significantly superior to SC decoding [31]. An intensive study of BP decoding on polar codes is presented in [39], where many interesting issues on BP decoding including stopping set, girth, and error floor have been discussed. In addition, there are some other kinds of decoding methods such as linear programming (LP) decoding [40].

Another drawback of polar codes is their construction in arbitrary binary-input memoryless symmetric (BMS) channels. The originally proposed construction method in [4] can only work for BDMS channels and its computational complexity is exponential in $N$ for other BDMS channels except binary erasure channels (BEC). Then Arikan developed a heuristic method for arbitrary BMS channel by treating any given BMS channel as a BEC with the same channel capacity [38]. However, this heuristic method can not guarantee capacity achieving polar codes. To solve this

problem, a construction method using density evolution was proposed [6], without finite-precision implementation details. Based on a similar idea, Tal and Vardy introduced smart quantization methods using channel merging functions for a practical implementation [28].

## 2.2 Polar Codes for Lossless Compression [2]

Consider a sequence of $N = 2^n$ i.i.d. source symbols $X_1^N$ generated from a Bernoulli source Ber($p$). As in channel coding, the two main ingredients of polar source coding are sequential encoding and polarization. Instead of encoding $X_1^N$ directly we consider encoding the sequence $U_1^N = X_1^N G_N$. Since the transform is invertible, we have

$$H(U_1^N) = H(X_1^N) = N h_2(p), \tag{2.12}$$

where $h_2(p)$ is the binary entropy function. Furthermore, we have the chain rule of entropy

$$N \cdot h_2(p) = H(U_1^N) = \sum_{i=1}^{N} H(U_i|U_1^{i-1}). \tag{2.13}$$

Polarization of entropy would imply that the terms on the right of (2.13) become either close to 1 or close to 0 when we increase $N$. Now if we encode $U_1^N$ in a sequential fashion from index 1 to $N$, we do not need to encode the $U_i$ for which $H(U_i|U_1^{i-1})$ is close to 0, since they can be estimated from the source model given $U_1^{i-1}$. Therefore, the $U_i$ for which $H(U_i|U_1^{i-1})$ is close to 1 are stored. The main difference with polar coding for a channel coding setting is that the encoder can perform the estimation and compare with the realization $u_i$. When the estimate turns out to be wrong, the index of the wrong estimate should be announced to the decoder

and this will involve rate loss. From (2.13) it follows that the fraction of bits whose conditional entropy approaches 1 is $h_2(p)$ which results in a compression rate of $h_2(p)$.

The source encoder is defined by a set of frozen indices $F$ and each index $i \in F$ corresponds to a entropy term $H(U_i|U_1^{i-1})$ close to 1. The frozen set $F$ and the source model are assumed to be known to the encoder as well as to the decoder. Let $\bar{x}$ ($x_1^N$) denote the realization of the source sequence. A source encoder performs the following steps.

1) Computes $\bar{u} = \bar{x}G_N$.
2) For each $i \in F^c$, calculate the probability $P(U_i|U_1^{i-1} = u_1^{i-1})$, and estimates the value of $u_i$ as

$$\hat{u_i} = \text{round}(P(U_i|U_1^{i-1} = u_1^{i-1})). \tag{2.14}$$

3) Record a set of indices $T$ which is defined as $T = \{i : i \in F^c, \hat{u_i} \neq u_i\}$. The set $T$ contains all indices for which the estimation of the corresponding $u_i$ would be incorrect. Finally, the output of the encoder is given by $(\hat{u}_F, T)$.

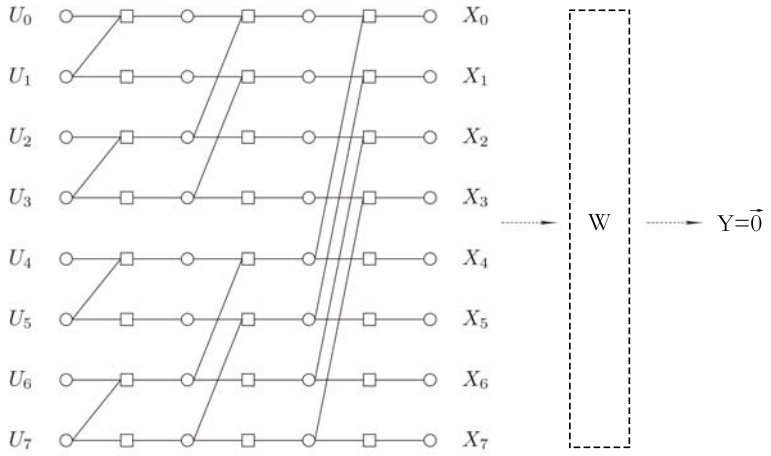



Figure 2.3: Example of a factor graph for a polar code with block length of $2^3 = 8$.

Now we discuss how to obtain $P(U_i|U_1^{i-1} = u_1^{i-1})$. The computation can be performed in an efficient way by considering the factor graph between $\bar{u}$ ($u_1^N$) and $\bar{x}$ ($x_1^N$). Fig. 2.3 shows an example of such a factor graph for $n = 3$. Due to the duality between the source compression and the channel coding problems, the task of recovering $\bar{x}$ from $u_F$ can be formulated as the following channel decoding problem for a BSC with crossover probability $p$ (We denote it by BSC($p$) for short). Let $\bar{x}$ be the input of a channel whose output is always $\bar{0}$, which means the noise during the transmission can be given by $\bar{z} = \bar{x}(\bar{y} = \bar{x} \oplus \bar{z} = \bar{0})$. Since $\bar{x}$ is an i.i.d. Ber($p$) source, it implies that the noise $\bar{z}$ is also i.i.d. Ber($p$) vector. By construction, $\bar{x} \in C_N(F, u_F)$. Therefore, recovering $\bar{x}$ from $u_F$ is equivalent to decoding at the output of $W$, which can be viewed as a BSC($p$) with output $\bar{y} = \bar{0}$. Consequently, the probability $P(U_i|U_1^{i-1})$ is equal to the posterior probability $P(U_i|U_1^{i-1}, Y_1^N = \bar{0})$ in the channel coding context. The SC algorithm indeed estimates the bits in $F^c$ based on these posteriors.

The input of the decoder is given by $(u_F, T)$. The decoder performs the following steps. First, the decoder computes for $i \in F^c$ an estimate of $u_i$. If $i \in T$ the value of the estimate is flipped. Once $u_{F^c}$ has been recovered, the encoder knows $\bar{u}$ and can recover $\bar{x}$ as $\bar{x} = \bar{u}G_N^{-1} = \bar{u}G_N$.

**Theorem 2.2.1** (Polar Code is Optimal for Lossless Compression [2])**.** *Let $X$ be a Ber(p) random variable. For any rate $R > h_2(p)$ there exists a sequence of polar codes of length $N$ and rate $R_N < R$ so that the source can be compressed losslessly with rate $R_N$.*

*Proof.* Let $\epsilon > 0$, $R = h_2(p)+\epsilon$, and $F_N$ be a frozen set such that $\frac{|F_N|}{N} \leq h_2(p)+\epsilon/2$. Because $T$ is the set of erroneous indices, the average rate of compression is given by

$$\frac{1}{N}(|F_N| + E[|T|\log N]) \leq \frac{1}{N}(|F_N| + Pr(|T| > 0)N\log N), \tag{2.15}$$

As we have discussed previously, the probability $Pr(|T| > 0)$ is equal to the block error probability of using a polar code $C_N(F_N, u_{F_N})$ for the BSC($p$) under SC decoding. By Theorem 2.1.1, the required rate is at most

$$R_N \leq \frac{|F_N| + P_B(F_N)N\log N}{N} \leq \frac{|F_N|}{N} + O(2^{-N^\beta}). \tag{2.16}$$

When $N$ is sufficiently large, we have $R_N < R$. □

Let $\mathcal{E}_i$ denote the failure probability of encoding bit $U_i$ given the past $U_1^{i-1}$. We can express the expected encoding length $E[L]$ as

$$\begin{aligned} E[L] &= E[|F| + \sum_{i \in F^c} \mathbb{1}(\hat{U}_i \neq U_i) \cdot \log N] \\ &= |F| + \log N \cdot E[\sum_{i \in F^c} \mathbb{1}(\hat{U}_i \neq U_i)] \\ &= |F| + \log N \cdot \sum_{i \in F^c} \mathcal{E}_i. \end{aligned} \tag{2.17}$$

Furthermore, the expected encoding rate is defined as $E[R] = \frac{E[L]}{N}$. In order to get $E[R]$, we need to know how to calculate $\mathcal{E}_i$. By the previous analysis, it is clear that $\mathcal{E}_i$ is equivalent to the error probability of the subchannel $W_N^{(i)}$ when the channel $W$ is the BSC($p$). We can use density evolution method or the smart channel quantization method to estimate this error probability. After obtaining the $\mathcal{E}_i$ for all $i \in \{1, ...N\}$, we have the following lemma.

**Lemma 2.2.2.** *The expected encoding length $E[L]$ is minimized when $F$ is chosen as*

$$F = \{i : \mathcal{E}_i \geq \frac{1}{\log N}\}. \tag{2.18}$$

For each $i$ we have the choice of including it in $F$ or $F_c$, contributing 1 bit or $\mathcal{E}_i \log N$ bits to $E[L]$ respectively, which implies the above construction principle is

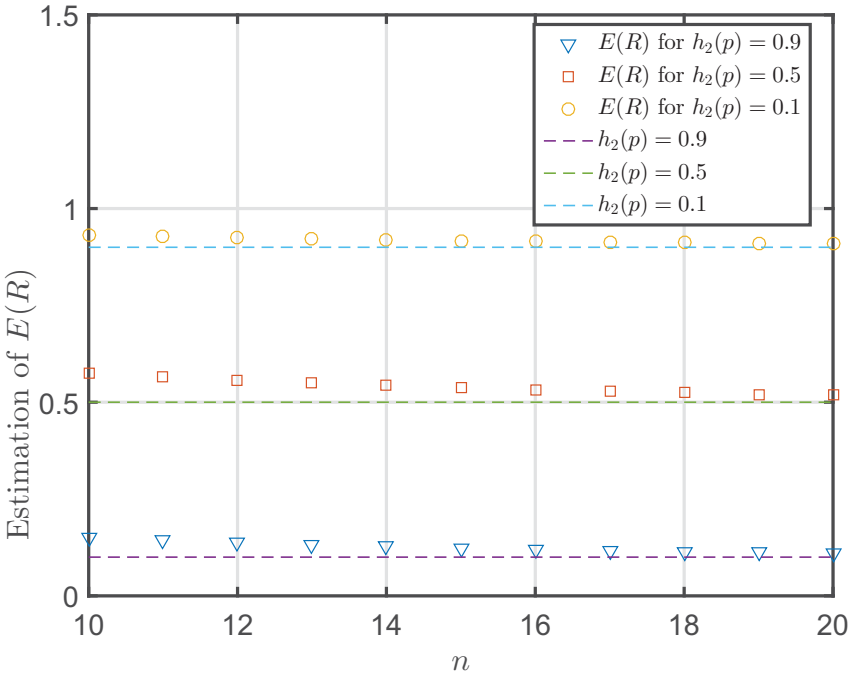



Figure 2.4: Result of the numerical computation of $E[R]$ for source entropies of 0.9, 0.5 and 0.1.

optimal.

Fig. 2.4 shows the expected encoding rate for a Ber($p$) source with an entropy of 0.9, 0.5 and 0.1, respectively. $n$ is chosen in the range $n = 8, ..., 22$. For $h(p) = 0.5$, the estimation of $E[R]$ is $0.6050, 0.5837, 0.5687, 0.5549, 0.5440...$ respectively. We observe that for increasing block lengths the average encoding rate approaches the theoretical entropy limit. Even for the smaller block lengths the gap is not that large. Simulation results for an actual encoder are shown in Fig. 2.5.

## 2.3 Polar Codes for Lossy Compression

In this part, we discuss polar lossy source coding. We model the source as a sequence of i.i.d. realizations of a random variable $Y \in \mathcal{Y}$. Let $\mathcal{X}$ denote the reconstruction space. Let $\mathrm{d} : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}_+$ denote the distortion function with maximum value $\mathrm{d}_{max}$. The distortion function naturally extends to vectors as $\mathrm{d}(y_1^N, x_1^N) = \sum_{i=1}^{N} \mathrm{d}(y_i, x_i)$.

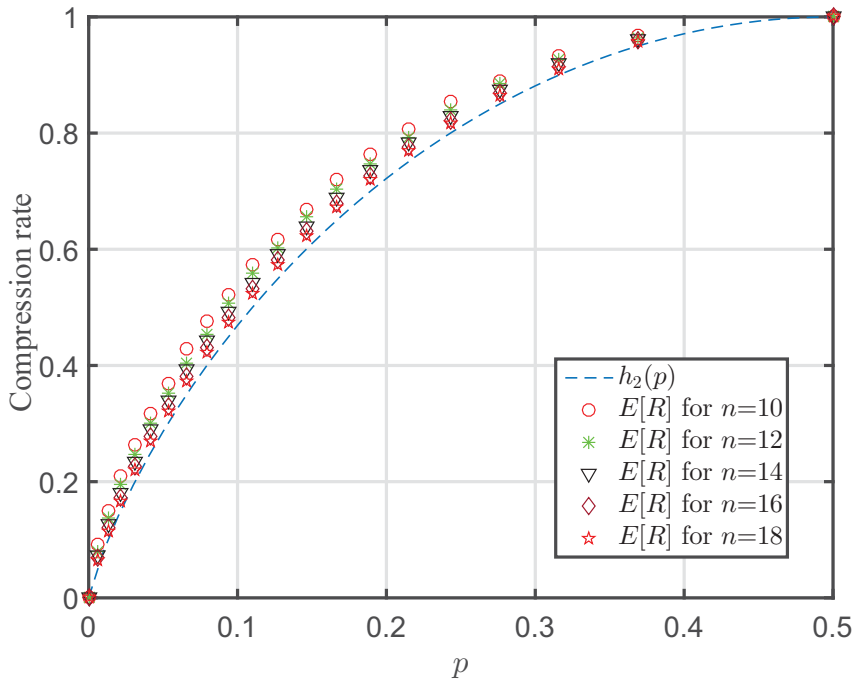



Figure 2.5: Simulation results for the polar source encoder for $n = 10, 12, 14, 16, 18$.

Let $C_N \subseteq \mathcal{X}^N$ be the code used for lossy compression. The encoder maps a source sequence $\bar{Y}(Y^N)$ to an index $f_N(\bar{Y})$. The index refers to a codeword $\bar{X}(X^N) \in C_N$. Therefore, the number of bits required for the encoder to convey the index is at most $\log |C_N|$, The decoder reconstructs the codeword $\bar{X}$ from the index. Let $g_N$ denote the reconstruction function. The rate of such a scheme is given by $\frac{\log |C_N|}{N}$ and the resulting average distortion is given by the single-letter distortion measure $D_N = \frac{1}{N} E[\mathrm{d}(\bar{Y}, g_N(f_N(\bar{Y})))]$.

Shannon's rate-distortion theorem characterizes the minimum rate required to achieve a given distortion.

**Theorem 2.3.1** (Shannon's Rate-Distortion Theorem)**.** *Consider an i.i.d. source $Y$ with probability distribution $P_Y(y)$. To achieve an average distortion $D$ the required rate is at least*

$$R(D) = \min_{p(y,x): E_p[d(y,x)] \leq D, p(y) = P_Y(y)} I(Y;X). \tag{2.19}$$

Since polar codes are utilized in the symmetrical scenarios, similarly to the def-

inition of the symmetric channel capacity, we define the symmetric rate-distortion function as follows.

**Definition 2.3.1** (Symmetric Rate-Distortion Function)**.** For a random variable $Y$ with probability distribution $P_Y(y)$, the symmetric rate distortion function $R_s(D)$ is defined as

$$R_s(D) = \min_{p(y,x):E_p[\mathrm{d(y,x)}]\leq D, p(y)=P_Y(y), p(x)=\frac{1}{|\mathcal{X}|}} I(Y;X). \tag{2.20}$$

In this definition we have the additional constraint that the induced probability distribution over $\mathcal{X}$ must be uniform. This is is similar in spirit to the symmetric capacity of a channel which is defined as the mutual information between the input and output of the channel when the input distribution is uniform. Clearly, $R(D) > R_s(D)$. Let $p^*(y,x)$ be a probability distribution that minimizes (2.20). Therefore, the marginal of $p^*(y,x)$ over $\mathcal{Y}$ is $P_Y$ and its marginal over $\mathcal{X}$ is uniform. The conditional distribution $p^*(y|x)$ is defined as test channel and it is denoted by $W(Y|X)$.

Let us consider using a polar code $C_N(F,u_F)$ for lossy source coding. The frozen set $F$ as well as the vector $u_F$ are known both to the encoder and the decoder. A source word $\bar{Y}$ is mapped to a codeword $\bar{X} \in C_N(F,u_F)$. This codeword can be described by the index $u_{F^c}$. Therefore, the required rate is $\frac{|F^c|}{N}$. Since the decoder knows $u_F$ in advance, it recovers the codeword $\bar{X}$ by performing the operation $u_1^N G_N$. Fig. 2.6 provides a pictorial description of the relationship between $\bar{U}$, $\bar{X}$ and $\bar{Y}$.

Let $\bar{y}$ denote $N$ i.i.d. realizations of the source $Y$ and $\hat{U}(\bar{y},u_F)$ denote the output of encoder. The encoding operation can be described as follows. Given $\bar{y}$, for each $i$ in the range 1 till $N$:

1) If $i \in F$, then set $\hat{u}_i$=$u_i$.

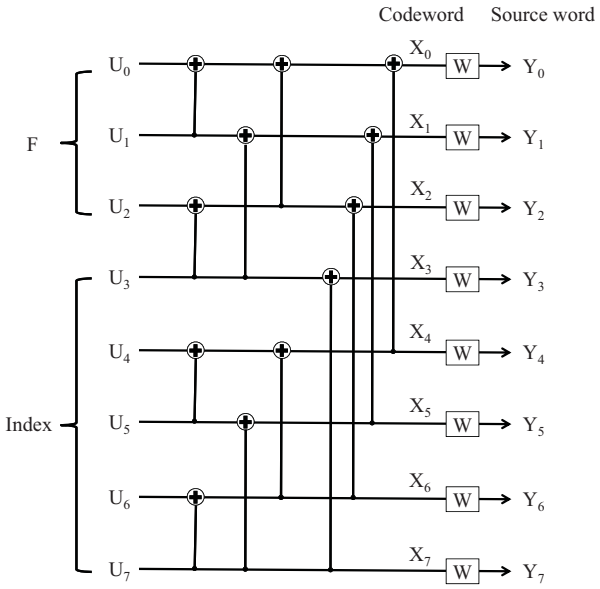



Figure 2.6: The probabilistic model used for source coding. The source word $\bar{Y}$ is treated as the output of a channel. The code is defined through its frozen set $F = 0, 1, 2$, The codewords are indexed using $U_3^7$. As we will see later, the frozen bits for the lossy source coding problem match with the frozen bits of the channel coding problem, with slight difference in the unpolarized region, which is of vanishing proportion.

2) If $i \in F^c$, then compute $L_N^{(i)}(\bar{y}, \hat{u}_1^{i-1})$, which is defined as $\frac{W_N^{(i)}(\bar{y},\hat{u}_1^{i-1}|0)}{W_N^{(i)}(\bar{y},\hat{u}_1^{i-1}|1)}$, and set

$$\hat{u}_i = \begin{cases} 0, & \text{w.p. } \frac{L_N^{(i)}}{1+L_N^{(i)}} \\ 1, & \text{w.p. } \frac{1}{1+L_N^{(i)}}. \end{cases} \tag{2.21}$$

There is a slight difference between (2.10) and (2.21). Note that unlike in channel coding, where we made a hard decision based on $L_N^{(i)}$, $\hat{u}_i$ is chosen randomly with a probability based on $L_N^{(i)}$ (MAP rule). We refer to this decision rule as randomized rounding. Actually, randomized rounding and the MAP rule perform similarly in practice. The purpose of using randomized rounding here is for the analysis convenience (theoretical proof).

**Theorem 2.3.2** (Polar Codes Achieve the Symmetric Rate-Distortion Bound)**.** *Con-*

*sider an i.i.d. source $Y$ and a distortion function. Fix a distortion $D$ and $0 < \beta < \frac{1}{2}$. For any rate $R > R_s(D)$, there exists a sequence of polar codes of length $N$ and rate $R_N \leq R$, such that under SC encoding using randomized rounding they achieve expected distortion $D_N$ satisfying*

$$D_N \leq D + O(2^{(-N^\beta)}). \tag{2.22}$$

A detailed proof of the above theorem can be found in [30]. The main idea behind the proof is that if we use all the variables $U_i$ to represent the source word i.e., if $F$ is empty, then the algorithm results in an average distortion $D$ because $G_N$ is invertible. But the rate of such a code would be 1. Fortunately, by polarization, there exists some "essentially random" bits among $\bar{U}$, which means their posteriori probabilities $P_{U_i|U_1^{i-1},\bar{Y}}(0|U_1^{i-1},\bar{Y}) \cong \frac{1}{2}$. If we choose $F$ to consist of those variables and make a random guess for them when decoding, there will be only a small distortion penalty. Note that the additional distortion is caused by the fact that some frozen bits are exactly uniformly random, especially when $N$ is not sufficiently large.

**Lemma 2.3.3** (How to Construct Polar Lossy Source Coding)**.** *For a small real number $\delta_N$, if the frozen set $F$ is chosen to be all the indices $i$ such that $\tilde{Z}(W_N^{(i)}) \geq 1 - 2\delta_N^2$, then we have*

$$E_P[|\frac{1}{2} - P_{U_i|U_1^{i-1},\bar{Y}}(0|U_1^{i-1},\bar{Y})|] \leq \delta_N, \tag{2.23}$$

*and the distortion achieved by this polar code satisfies*

$$D_N(F_N) \leq D + 2|F_N|d_{max}\delta_N \leq D + O(2^{-N^\beta}). \tag{2.24}$$

*when $\delta_N$ is chosen to be $\frac{1}{2Nd_{max}}2^{-N^\beta}$.*

Let us consider how polar codes behave in practice. We consider the binary sym-

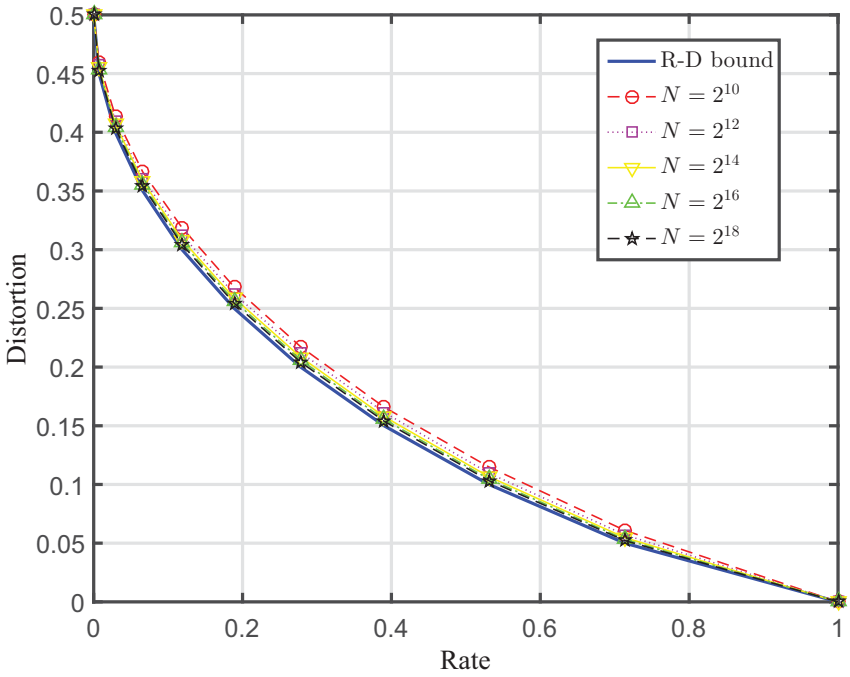



Figure 2.7: The rate-distortion performance for the SC encoding algorithm with randomized rounding for $n = 10, 12, 14, 16, 18$.

metric source (BSS) and the Hamming distortion measure. Hence the rate distortion function is given by $R(D) = 1 - h_2(D)$. The test channel is the BSC($D$). Fig. 2.7 shows the performance of polar codes based on the mentioned construction method. As the block length increases the points move closer to the rate-distortion bound.

Polar coding is a powerful tool which is able to handle the problems of channel coding, lossless source coding, and lossy source coding. Moreover, its asymptotic property in these problems can be theoretically proved. Note that the construction methods of polar codes in these problems are different. In the case of channel coding and lossy source coding, frozen set carries no information and can be treated as all zero set. Whereas in the case of lossless source coding, frozen set carries most information. The reason for this difference is that for lossless source coding, polarization is caused by the asymmetry of the source itself, while for channel coding and lossy source coding, polarization is caused by communication channel or test channel. The former one can be viewed as maintaining the system uncertainty as far

as possible, while the purpose of the latter ones is to avoid communication errors or compression distortions as far as possible.

The code construction for channel coding and lossy source coding requires knowledge of $\tilde{Z}(W_N^{(i)}) : \forall i \in \{1, ..., N\}$. The set $F$ for channel coding is given by $\{i : \tilde{Z}(W_N^{(i)}) \geq 2^{-N^\beta}\}$. Whereas the set $F$ for lossy compression is given by $\{i : \tilde{Z}(W_N^{(i)}) \geq 1 - 2^{-N^\beta}\}$. This also implies for channel coding we always have $R < I(W)$ and to source coding we always have $R > I(W)$. For lossless source coding, we need the knowledge of the error probability $P(W_N^{(i)})$ of each subchannel. By using density evolution [6] and smart channel states quantization [28], we can obtain a good estimation of $P(W_N^{(i)})$ and $\tilde{Z}(W_N^{(i)})$ for any BMS channels.

## 2.4 Polar Codes for Wiretap Coding

Wyner [41] introduced the wiretap channel and showed that both reliability to transmission errors and a prescribed degree of data confidentiality could be attained by channel coding without any key bits, if the channel between the sender and the eavesdropper (wiretapper's channel) is a degraded version of the channel between the sender and the legitimate receiver (main channel). The goal is to design a coding scheme that makes it possible to communicate both reliably and securely, as the block length of transmitted codeword $N$ tends to infinity. Reliability is measured by the decoding error probability of the legitimate user, namely $\lim_{N\to\infty} \Pr\{\hat{M} \neq M\} = 0$, where $M$ is the confidential message and $\hat{M}$ is its estimation. Secrecy is measured by the mutual information between $M$ and the signal received by the eavesdropper $Z_1^N$. Currently the widely accepted strong secrecy condition was proposed by Csiszár [42]: $\lim_{N\to\infty} I(M; Z_1^N) = 0$. In simple terms, the secrecy capacity is the maximum achievable rate of any coding scheme that can satisfy both the reliability and strong secrecy conditions.

Polar codes have shown their great potential of solving this wiretap coding problem. The polar coding scheme proposed in [43] is proved to achieve the strong secrecy capacity with explicit construction when the main channel and the wiretapper's channel are both binary-input memoryless channels, although it is not able to guarantee the reliability condition. A subsequently modified scheme [44] fixes this issue and finally satisfies the two conditions. However, for continuous channels such as the Gaussian wiretap channel, the problem of achieving strong secrecy with a practical code is still open. To solve this problem, we firstly need to modify the polar wiretap coding scheme proposed in [43]. We will introduce the modified scheme in this part and the modification is for the purpose of the shaping scheme, which is going to be discussed in the next section.

Now 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

$$\begin{aligned} \mathcal{G}(\tilde{V}) &= \{i : \tilde{Z}(\tilde{V}_N^{(i)}) \leq 2^{-N^\beta}\}, \\ \mathcal{N}(\tilde{W}) &= \{i : \tilde{Z}(\tilde{W}_N^{(i)}) \geq 1 - 2^{-N^\beta}\}. \end{aligned} \tag{2.25}$$

In [43], the information-bad set $\mathcal{N}(\tilde{W})$ is defined as $\{i : I(\tilde{W}_N^{(i)}) \leq 2^{-N^\beta}\}$, which was directly from the mutual information of the subchannels. However, our new criterion is based on the Bhattacharyya parameter. The following lemma shows that the new criterion is similar to 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 2.4.1.** *Let $\tilde{W}_N^{(i)}$ be the $i$-th subchannel after the polarization transform on independent $N$ 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*

$$I(\tilde{W}_N^{(i)}) \leq 2^{-N^{\beta'}}, 0 < \beta' < \beta < 0.5. \tag{2.26}$$

*Proof.* Since $\tilde{W}$ is symmetric, $\tilde{W}_N^{(i)}$ is symmetric as well. By the [4, Proposition 1], we have

$$\begin{aligned} I(\tilde{W}_N^{(i)}) &\leq \sqrt{1 - \tilde{Z}(\tilde{W}_N^{(i)})^2} \\ &\leq \sqrt{2 \cdot 2^{-N^\beta}} \leq 2^{-N^{\beta'}}, \end{aligned} \tag{2.27}$$

where the last inequality holds for sufficiently large $N$. □

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 the strong secrecy can be achieved using the technique proposed in [43]. Similarly, we divide the index set $[N]$ into the following four sets:

$$\begin{aligned} \mathcal{A} &= \mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W}) \\ \mathcal{B} &= \mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W})^c \\ \mathcal{C} &= \mathcal{G}(\tilde{V})^c \cap \mathcal{N}(\tilde{W}) \\ \mathcal{D} &= \mathcal{G}(\tilde{V})^c \cap \mathcal{N}(\tilde{W})^c. \end{aligned} \tag{2.28}$$

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 2.4.2.** *According to the partitions of the index set shown in* (2.28)*, if we assign the four sets as follows*

$$\begin{aligned} \mathcal{A} &\leftarrow M \\ \mathcal{B} &\leftarrow R \\ \mathcal{C} &\leftarrow F \\ \mathcal{D} &\leftarrow R, \end{aligned} \tag{2.29}$$

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

$$I(M; Z^{1:N}) \leq N \cdot 2^{-N^{\beta'}}. \tag{2.30}$$

*Proof.* As has been shown in [43], the induced channel $MF \to Z^{1:N}$ 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 $\mathcal{A} \cup \mathcal{C}$,

$$\begin{aligned} I(MF; Z^{1:N}) &\leq I(\tilde{U}_{\mathcal{A}}\tilde{U}_{\mathcal{C}}; \tilde{Z}^{1:N}) \\ &= \sum_{j=1}^{|\mathcal{A}\cup\mathcal{C}|} I(\tilde{U}^{i_j}; \tilde{Z}^{1:N} | \tilde{U}^{i_1}, ..., \tilde{U}^{i_{j-1}}) \\ &= \sum_{j=1}^{|\mathcal{A}\cup\mathcal{C}|} I(\tilde{U}^{i_j}; \tilde{Z}^{1:N}, \tilde{U}^{i_1}, ..., \tilde{U}^{i_{j-1}}) \\ &\leq \sum_{j=1}^{|\mathcal{A}\cup\mathcal{C}|} I(\tilde{U}^{i_j}; \tilde{Z}^{1:N}, \tilde{U}^{1:i_j-1}) \\ &= \sum_{j=1}^{|\mathcal{A}\cup\mathcal{C}|} I(\tilde{W}_N^{(i_j)}) \leq N \cdot 2^{-N^{\beta'}}. \end{aligned} \tag{2.31}$$

□

Note that there is no specific assumption on the distribution on $M$ and $F$ and a similar proof can be found in [44]. With regard to the secrecy rate, we show that the modified polar coding scheme can also achieve the secrecy capacity.

**Lemma 2.4.3.** *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.,*

$$\lim_{N\to\infty} |\mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W})|/N = C(\tilde{V}) - C(\tilde{W}). \tag{2.32}$$

*Proof.* According to the definitions of $\mathcal{G}(\tilde{V})$ and $\mathcal{N}(\tilde{W})$ presented in (2.25),

$$\lim_{N\to\infty} \frac{|\mathcal{G}(\tilde{V})|}{N} = \lim_{N\to\infty} \frac{1}{N} |\{i : \tilde{Z}(\tilde{V}_N^{(i)}) \leq 2^{-N^\beta}\}| = C(\tilde{V}), \tag{2.33}$$

$$\lim_{N\to\infty} \frac{|\mathcal{N}(\tilde{W})|}{N} = \lim_{N\to\infty} \frac{1}{N} |\{i : \tilde{Z}(\tilde{W}_N^{(i)}) \geq 1 - 2^{-N^\beta}\}| = 1 - C(\tilde{W}). \tag{2.34}$$

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

$$\begin{aligned} \bar{\mathcal{G}}(\tilde{V}) &= \{i : \tilde{Z}(\tilde{V}_N^{(i)}) \geq 1 - 2^{-N^\beta}\}, \\ \bar{\mathcal{N}}(\tilde{W}) &= \{i : \tilde{Z}(\tilde{W}_N^{(i)}) \leq 2^{-N^\beta}\}. \end{aligned} \tag{2.35}$$

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

$$\lim_{N\to\infty} \frac{|\bar{\mathcal{G}}(\tilde{V}) \cup \bar{\mathcal{N}}(\tilde{W})|}{N} = 1 - C(\tilde{V}) + C(\tilde{W}). \tag{2.36}$$

By the property of polarization, the proportion of the unpolarized part is vanishing

as $N$ goes to infinity, i.e.,

$$\lim_{N\to\infty} \frac{|\mathcal{G}(\tilde{V}) \cup \bar{\mathcal{G}}(\tilde{V})|}{N} = 1, \tag{2.37}$$

$$\lim_{N\to\infty} \frac{|\mathcal{N}(\tilde{W}) \cup \bar{\mathcal{N}}(\tilde{W})|}{N} = 1, \tag{2.38}$$

Finally, we have

$$\lim_{N\to\infty} \frac{|\mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W})|}{N} = 1 - \lim_{N\to\infty} \frac{|\bar{\mathcal{G}}(\tilde{V}) \cup \bar{\mathcal{N}}(\tilde{W})|}{N} = C(\tilde{V}) - C(\tilde{W}). \tag{2.39}$$

□

It is not difficult to observe that the proportion of the problematic set $\mathcal{D}$ is arbitrarily small. This is 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 [43], the reliability condition cannot be proved due to the existence of set $\mathcal{D}$. Fortunately, since the proportion of set $\mathcal{D}$ can still be made arbitrarily small, we can use the blocking technique proposed in [44, 45] to achieve reliability and strong secrecy condition simultaneously.

CHAPTER 3

# From Polar Codes to Polar Lattices

In this chapter, we give a brief introduction of lattices and their construction methods. Some classical lattices will be revisited as examples. We also discuss some newly emerging lattices including the polar lattices for the following chapters.

As mentioned in Chapter 1, our first application of polar lattices will focus on achieving the capacity of AWGN channels. To this end, we will need two essential ingredients, i.e., the construction of lattices which are good for AWGN channel coding without power constraint (the AWGN-good lattices) and the lattice Gaussian shaping. We will present the definitions of those concepts in this chapter. Some definitions of other good lattices, such as the secrecy-good lattices [21] and quantization-good lattices [46], will be discussed in Chapter 5 and Chapter 6, respectively. For more details of lattices, we recommend [25] and [1].

## 3.1 Latices and lattice codes

Lattice codes are useful in many communication scenarios with continuous-output channels, such as the AWGN 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 $\Lambda$ can be used as a signal constellation in a communication system. This set consists of those points of $\Lambda$ that are contained in a bounded shaping region $S$, and is known as the lattice code $\mathbb{C}(\Lambda, S)$ based on $S$ and $\Lambda$. The performance of a lattice code $\mathbb{C}(\Lambda, S)$ on the AWGN channel depends not only on the underlying lattice $\Lambda$

(packing problem, coding gain) [47] but also on the shape of the support region $S$ (covering problem, shaping gain) [48]. A lattice code is generated by applying the power constraint to an infinite lattice.

### 3.1.1 Lattices

Mathematically, a lattice is defined as a modulo over a certain ring and embedded in a vector space over a field. For our purposes, we will only consider real lattices, that is $\mathbb{Z}$-modulo in the Euclidean space. And we will only deal with full rank lattices, namely $n$-dimensional lattices in an $n$-dimensional Euclidean space.

**Definition 3.1.1.** A lattice is a discrete subgroup of $\mathbb{R}^n$ which can be described by [20]

$$\Lambda = \{\boldsymbol{\lambda} = \mathbf{B}\mathbf{x} : \mathbf{x} \in \mathbb{Z}^n\}, \tag{3.1}$$

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 $\Lambda$ is $Q_\Lambda(\mathbf{x}) = \arg\min_{\boldsymbol{\lambda}\in\Lambda} \|\boldsymbol{\lambda} - \mathbf{x}\|$. We define the modulo lattice operation by $\mathbf{x} \bmod \Lambda \triangleq \mathbf{x} - Q_\Lambda(\mathbf{x})$ [20]. The Voronoi region of $\Lambda$, 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 $\Lambda$ 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 $\Lambda$. 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 $\Lambda$ is $d_{\min}(\Lambda) = \min_{\mathbf{x}\in\Lambda} \|\mathbf{x}\|$.

Since a lattice has infinite lattice points, it is known as infinite constellation (IC) or coding without power constraint [49]. This scenario is simpler than the power

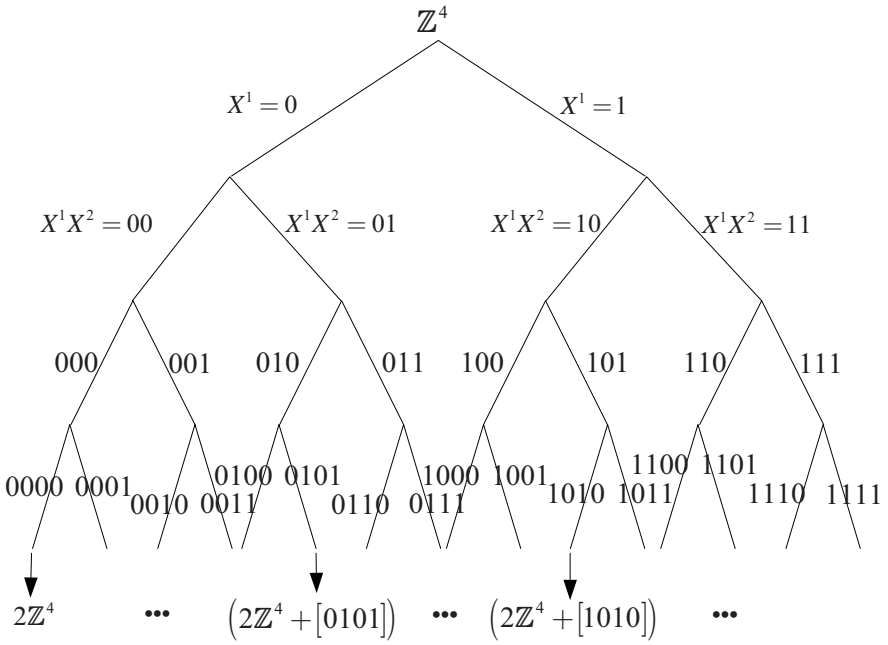



Figure 3.1: Lattice partitions induced by the $\mathbb{Z}^4/2\mathbb{Z}^4$

constrained case in the sense that the decoding does not take the shaping region into account. 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.

### 3.1.2 Lattice Partition

A sublattice $\Lambda' \subset \Lambda$ induces a partition (denoted by $\Lambda/\Lambda'$) of $\Lambda$ into equivalence groups modulo $\Lambda'$. 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" [25, p.232].

For each partition $\Lambda_\ell/\Lambda_{\ell+1}$ ($1 \leq \ell \leq r-1$) a code $C_\ell$ over $\Lambda_\ell/\Lambda_{\ell+1}$ selects a sequence of coset representatives $a_\ell$ in a set $A_\ell$ of representatives for the cosets of

$\Lambda_{\ell+1}$. This construction requires a set of nested linear binary codes $C_\ell$ with block length $N$ and dimension of information bits $k_\ell$ for $1 \leq \ell \leq r-1$ and $C_1 \subseteq C_2 \cdots \subseteq C_{r-1}$. Let $\psi$ 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 $C_\ell$. When $n = 1$, the binary lattice $L$ consists of all vectors of the form

$$\sum_{\ell=1}^{r-1} 2^{\ell-1} \sum_{j=1}^{k_\ell} \alpha_j^{(\ell)} \psi(\mathbf{b}_j) + 2^{r-1}\mathbf{l}, \tag{3.2}$$

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

A mod-$\Lambda$ channel is a Gaussian channel with a modulo-$\Lambda$ operator in the front end [50]. The capacity of the mod-$\Lambda$ channel is [50]

$$C(\Lambda, \sigma^2) = \log V(\Lambda) - h(\Lambda, \sigma^2), \tag{3.3}$$

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

$$h(\Lambda, \sigma^2) = -\int_{\mathcal{V}(\Lambda)} f_{\sigma,\Lambda}(\mathbf{x}) \log f_{\sigma,\Lambda}(\mathbf{x}) d\mathbf{x}. \tag{3.4}$$

The differential entropy is maximized to $\log V(\Lambda)$ by the uniform distribution over $\mathcal{V}(\Lambda)$. It is known that the $\Lambda/\Lambda'$ channel (i.e., the mod-$\Lambda'$ channel whose input is drawn from $\Lambda \cap \mathcal{V}(\Lambda')$) is regular, and the optimum input distribution is uniform [50]. Furthermore, the $\Lambda/\Lambda'$ channel is a symmetric binary-input channel if $|\Lambda/\Lambda'| = 2$ [51]. The capacity of the $\Lambda/\Lambda'$ channel for Gaussian noise of variance $\sigma^2$ is given by [50]

$$\begin{aligned} C(\Lambda/\Lambda', \sigma^2) &= C(\Lambda', \sigma^2) - C(\Lambda, \sigma^2) \\ &= h(\Lambda, \sigma^2) - h(\Lambda', \sigma^2) + \log(V(\Lambda')/V(\Lambda)). \end{aligned} \tag{3.5}$$

Each lattice construction from error correction codes needs a lattice partition.

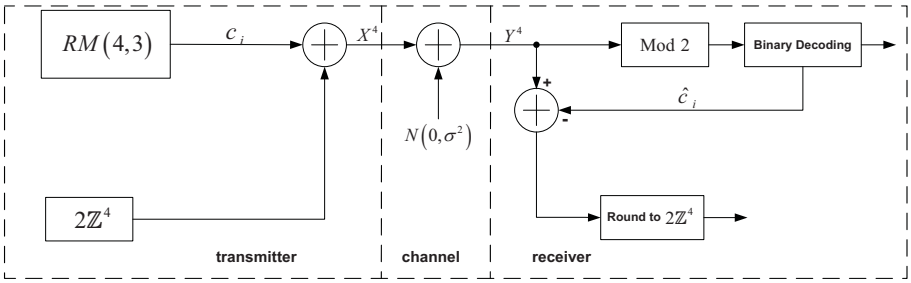



Figure 3.2: The encoding and decoding system of the $D_4$ lattice. The binary code is the Reed-Muller code with $N = 4$ and $k = 3$.

This is different with the conventional modulations. The beauty of the lattice system 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 Fig. 3.1. 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 $|\mathbb{Z}^4/2\mathbb{Z}^4| = 2^4$):

**Example 3.1.1.**

$$D_4 = \cup_{c_i \in C}(2\mathbb{Z}^4 + c_i). \tag{3.6}$$

Its generator matrix $\mathbf{B}$ is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}. \tag{3.7}$$

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 Fig. 3.2.

### 3.1.3 Lattice Codes

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

**Definition 3.1.2** (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 $\Lambda$ (possibly a shifted version of $\Lambda$) and $S$:

$$\mathbb{C}(\Lambda, S) = (\Lambda + c) \cap S, \tag{3.8}$$

where $S$ is called the shaping region, and $c$ is a shift of $\Lambda$. Letting $M$ be the cardinality of the lattice code, its rate is defined as

$$R_{\mathbb{C}} = \frac{\log M}{n}. \tag{3.9}$$

The average power of this lattice code per dimension is

$$P = \frac{1}{nM} \sum_{\lambda \in \Lambda \cap S} \|\lambda\|^2. \tag{3.10}$$

Here the shaping region can be generalized to the notion of a shaping technique. As long as we can control the transmit power by selecting points from an infinite lattice, we obtain a lattice code. There are some general shaping methods such as the Cubic shaping and the Voronoi shaping [1]. In this work, we use the probabilistic shaping over an infinite lattice proposed by [20]. The main idea is that each coordinate of the input follows a discrete Gaussian distribution.

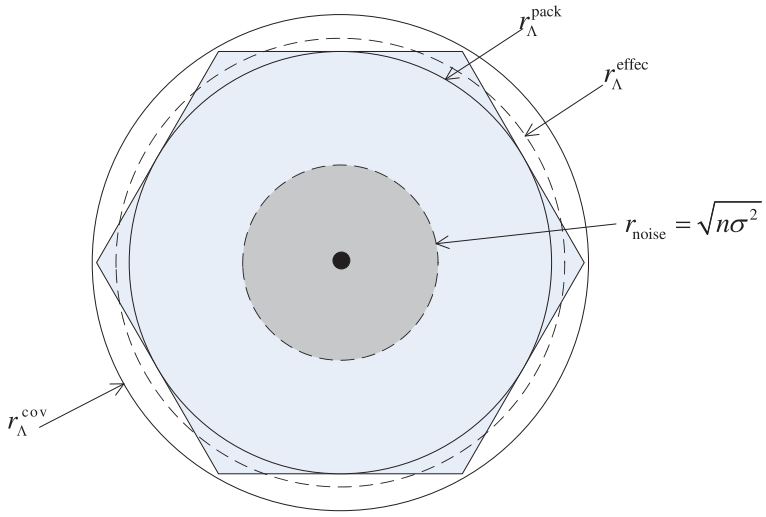



Figure 3.3: Geometric picture of lattices [1].

## 3.2 AWGN-Goodness and Lattice Gaussian Distribution

### 3.2.1 AWGN-goodness of Lattices

In this section, we give an introduction about the AWGN-goodness of the lattices. 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 [49] or sphere-bound-achieving lattices [50]. In this thesis, we adapt these terms under different context accordingly.

Let $V(\Lambda)$ be the fundamental volume of $\Lambda$, which is the volume of the Voronoi region of $\Lambda$. Packing radius $r_\Lambda^{\text{pack}}$ shown in Fig. 3.3 is the radius of the largest $n$-dimensional ball contained in the Voronoi region of $\Lambda$. $r_\Lambda^{\text{pack}} = \frac{d_{\min}(\Lambda)}{2}$, where $d_{\min}(\Lambda)$ is the minimum distance between two lattice points of $\Lambda$.

The effective radius $r_\Lambda^{\text{effec}}$ shown in Fig. 3.3 is the radius of a sphere with the

volume $V(\Lambda)$. It is known [25, p9] that the volume of a unit sphere $V_n$ in $\mathbb{R}^n$ is

$$V_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} = \begin{cases} \frac{\pi^k}{k!}, & n = 2k \\ \frac{2^n \pi^k k!}{n!}, & n = 2k+1 \end{cases} \tag{3.11}$$

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 $r_\Lambda^{\text{effec}}$ is

$$r_\Lambda^{\text{effec}} = \frac{V(\Lambda)^{1/n}}{V_n^{1/n}} = \frac{V(\Lambda)^{1/n}\Gamma(\frac{n}{2}+1)}{\sqrt{\pi}}. \tag{3.12}$$

Let $\sigma^2$ 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 [52], i.e.

$$\alpha^2(\Lambda, \sigma^2) = \frac{(r_\Lambda^{\text{effec}})^2}{n\sigma^2} \approx \frac{V(\Lambda)^{\frac{2}{n}}}{2\pi e \sigma^2}, \tag{3.13}$$

where the approximation is obtained by using the Stirling approximation of $k! \approx (k/e)^k$ for even $n$. Here $\frac{V(\Lambda)^{\frac{2}{n}}}{\sigma^2}$ is called the volume-to-noise ratio (VNR).

**Definition 3.2.1.** Given $\sigma$, the VNR of an $n$-dimension lattice $\Lambda$ is defined by

$$\gamma_\Lambda(\sigma) \triangleq \frac{V(\Lambda)^{\frac{2}{n}}}{\sigma^2}. \tag{3.14}$$

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 i.i.d. Gaussian noise vector $X^n$ with zero mean and variance $\sigma^2$ per dimension falls outside the Voronoi region of $\Lambda$.

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

**Definition 3.2.2** (AWGN-goodness[20])**.** A sequence of lattices $\Lambda^{(n)}$ of increasing dimension $n$ is AWGN-good if, for any fixed $P_e(\Lambda^{(n)}, \sigma^2) \in (0, 1)$,

$$\lim_{n\to\infty} \gamma_{\Lambda^{(n)}}(\sigma) = 2\pi e \tag{3.15}$$

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

### 3.2.2 Lattice Gaussian Distribution and Flatness Factor

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

$$f_\sigma(\mathbf{x}) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{\|\mathbf{x}\|^2}{2\sigma^2}}, \tag{3.16}$$

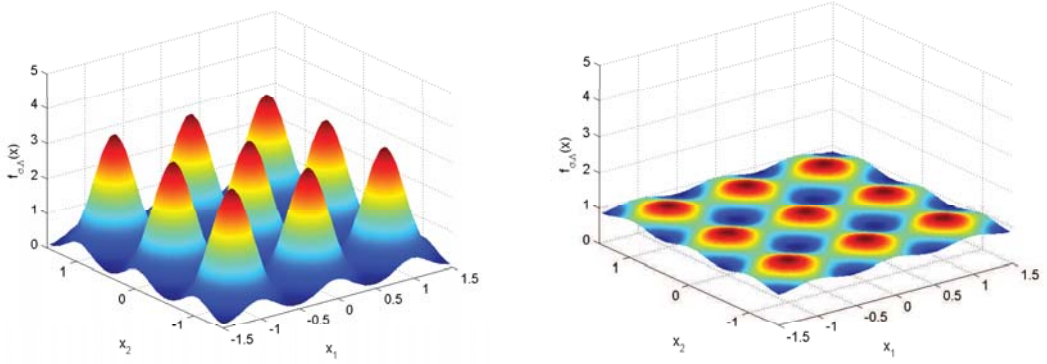
for all $\mathbf{x} \in \mathbb{R}^n$.

We also need the $\Lambda$-periodic function

$$f_{\sigma,\Lambda}(\mathbf{x}) = \sum_{\boldsymbol{\lambda}\in\Lambda} f_{\sigma,\boldsymbol{\lambda}}(\mathbf{x}) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \sum_{\boldsymbol{\lambda}\in\Lambda} e^{-\frac{\|\mathbf{x}-\boldsymbol{\lambda}\|^2}{2\sigma^2}}, \tag{3.17}$$

for all $\mathbf{x} \in \mathbb{R}^n$, where $f_{\sigma,\boldsymbol{\lambda}}(\mathbf{x}) = f_\sigma(\mathbf{x} - \boldsymbol{\lambda})$ is a $\boldsymbol{\lambda}$-shifted version of $f_\sigma(\mathbf{x})$.

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 $\Lambda$-aliased Gaussian noise, i.e., the Gaussian noise after the mod-$\Lambda$ operation [50]. It gets flat as $\sigma$ increases as shown in Fig. 3.4. In order to describe



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

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

Figure 3.4: The comparison of the $\Lambda$-aliased Gaussian distributions with different flatness factors.

such phenomenon, the flatness factor of a lattice $\Lambda$ is defined as [21]

$$\epsilon_\Lambda(\sigma) \triangleq \max_{\mathbf{x}\in\mathcal{R}(\Lambda)} \mid V(\Lambda) f_{\sigma,\Lambda}(\mathbf{x}) - 1 \mid, \tag{3.18}$$

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

*Remark* 3.2.1. $\epsilon_\Lambda(\sigma_1) < \epsilon_\Lambda(\sigma_2)$, if $\sigma_1 > \sigma_2$ [21].

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

$$D_{\Lambda,\sigma,\mathbf{c}}(\lambda) = \frac{f_{\sigma,\mathbf{c}}(\lambda)}{f_{\sigma,\mathbf{c}}(\Lambda)},\ \forall \lambda \in \Lambda, \tag{3.19}$$

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 [21].

**Theorem 3.2.1** (Mutual information of lattice Gaussian distribution [20])**.** *Consider an AWGN channel $Y = X + Z$ where the input constellation $X$ has a discrete Gaussian distribution $D_{\Lambda-c,\sigma_s}$ for arbitrary $c \in \mathbb{R}^n$, and where the variance of the noise*

*$Z$ is $\sigma^2$. Let the average signal power be $P$, and let $\tilde{\sigma} \triangleq \frac{\sigma_s \sigma}{\sqrt{\sigma_s^2+\sigma^2}}$ be the minimum mean square error (MMSE) re-scaled noise deviation. Then, if $\varepsilon = \epsilon_\Lambda(\tilde{\sigma}) < \frac{1}{2}$ and $\frac{\pi\varepsilon_t}{1-\epsilon_t} \leq \varepsilon$, where*

$$\varepsilon_t \triangleq \begin{cases} \epsilon_\Lambda\left(\sigma_s/\sqrt{\frac{\pi}{\pi-t}}\right), & t \geq 1/e, \\ (t^{-4}+1)\epsilon_\Lambda\left(\sigma_s/\sqrt{\frac{\pi}{\pi-t}}\right), & 0 < t < 1/e, \end{cases} \tag{3.20}$$

*the discrete Gaussian constellation results in mutual information*

$$I_D \geq \frac{1}{2}\log\left(1+\frac{P}{\sigma^2}\right) - \frac{5\varepsilon}{n} \tag{3.21}$$

*per channel use.*

Motivated by Theorem 3.2.1, one may choose a low-dimensional $\Lambda$ such as $\mathbb{Z}$ and $\mathbb{Z}^2$ whose mutual information has a negligible gap to the AWGN channel capacity, and then construct lattices to achieve the capacity. This idea will be used for the shaping of polar lattices in Chapter 4.

### 3.2.3 Existence of Capacity-Achieving Lattices

In [49], Poltyrev has presented a proof of the existence of capacity-achieving by using mod-$p$ lattices, with a tight error bounds which decays exponentially near capacity. Loeliger [53] then reproved the result using standard averaging arguments for the mod-$p$ lattices which are lifted from linear codes over the $p$-ary field GF($p$). Both of the two proofs are based on the Minkowski-Hlawka theorem.

For a general infinite constellation (not necessarily a lattice), there are several ways to define its coding rate or capacity per unit volume. One simple way used in [1] is to count the number of codewords per unit volume within a "large" cube and translate it into bits.

**Definition 3.2.3** (Rate per unit volume [1])**.** The coding rate per unit volume of an infinite constellation is defined as

$$\delta = \frac{1}{n} \limsup_{a\to\infty} \log\left(\frac{|\mathbb{C}_a|}{a^n}\right), \tag{3.22}$$

where

$$\mathbb{C}_a = \Lambda \cap \text{CUBE}(a) \tag{3.23}$$

is the intersection of the infinite constellation with the $n$-dimensional cube $\text{CUBE}(a) = [-a/2, a/2)^n$. We note that the rate per unit volume is also called normalized logarithmic density (NLD) [49].

**Definition 3.2.4** (Poltyrev capacity)**.** The highest rate per unit volume that allows reliable communication (i.e., with a vanishing error probability) over a large block of independent channel uses of AWGN channel is given by

$$\delta^* = \frac{1}{2}\log\frac{1}{2\pi e\sigma^2}. \tag{3.24}$$

Let the infinite constellation be a lattice $\Lambda$. The number of codewords contained in the cube is bounded between $(a-d)^n/V(\Lambda)$ and $(a+d)^n/V(\Lambda)$, where $d$ is the finite diameter of the Voronoi cell of $\Lambda$. Therefore, when $a$ is large and $\frac{d}{a}$ is negligible, the total number of codewords inside the cube is approximately $\frac{a^n}{V(\Lambda)}$.

In this case, the rate per unit volume of a lattice is given by

$$\delta = \frac{1}{n}\log\frac{1}{V(\Lambda)}. \tag{3.25}$$

The error performance of a lattice for unconstrained AWGN channels can be described by the error exponent $E(\delta)$. The decoding error probability is defined as

$P_e = e^{-n(E(\delta)+o(1))}$. In [49], the author derived a lower bound and a upper bound for $E(\delta)$. The lower bound is given by the random coding exponent $E_r(\delta)$, which is

$$E_r(\delta) = \begin{cases} \frac{1}{2}\log\frac{1}{8\pi\sigma^2} - \delta, & \delta \leq \delta_{cr}; \\ \frac{e^{-2\delta}}{4\pi e\sigma^2} + \delta + \frac{1}{2}\log 2\pi\sigma^2, & \delta_{cr} \leq \delta \leq \delta^*; \\ 0, & \delta \geq \delta^*, \end{cases} \tag{3.26}$$

where $\delta_{cr} = \frac{1}{2}\log\frac{1}{4\pi e\sigma^2}$. This low bound can be further improved by an expurgation-type argument at low rate region. More details can be found in [49] and [54].

The upper bound is given by the sphere packing exponent

$$E_{sp}(\delta) = \frac{1}{4\pi e^{2\delta+1}} + \delta + \frac{1}{2}\log 2\pi\sigma^2. \tag{3.27}$$

Please note that for a lattice with high dimension, the optimal decoding region is the equivalent sphere with effective radius $r_\Lambda^{\text{effec}}$, which leads to the upper bound on the error exponent (lower bound on the error probability).

By combining both the lower bound and upper bound on $E(\delta)$, we have

$$e^{-n(E_{sp}(\delta)+o(1))} \leq P_e(n,\delta) \leq e^{-n(E_r(\delta)+o(1))}, \tag{3.28}$$

where $P_e(n,\delta)$ is the decoding error probability.

We also note that $P_e(n,\delta)$ can be upper bounded by the union bound as

$$P_e(n,\delta) \leq e^{n\delta} n V_n \int_0^{2r} w^{n-1}\text{Pr}\{X^n \in D(r,w)\}dw + \text{Pr}\{\|X^n\| > r\}, \tag{3.29}$$

where the first term denotes the ambiguity error and $D(r,w)$ is defined as 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 second term is independent of the lattice.

Loeliger [53] reproved the existence of capacity-achieving lattices by using the mod-$p$ lattices generated by Construction A. The proof is again based on the Minkowski-Hlawka theorem.

**Theorem 3.2.2** (Loeliger's coding theorem [53])**.** *Let $E$ be a Jordan measurable bounded subset of $\mathbb{R}^n$, for any $\alpha > 0$, for all sufficiently small $\epsilon$, and all sufficiently large primes $p$, the arithmetic average of the ambiguity probability over all lattices $\epsilon L_C$ is upper bounded by [53]*

$$\overline{P_{amb|E}} \lesssim (1+\alpha)V(E)/V, \tag{3.30}$$

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

Recall that $\delta = \frac{1}{n}\log\frac{1}{V(\epsilon L_C)}$ is the rate per unit volume and $h(E) = \frac{1}{2}\log V(E)$ is the geometric entropy rate of $E$, and thus (3.30) can be rewritten as

$$\overline{P_{\text{amb}|E}} \lesssim (1+\alpha)2^{n(\delta+h(E))}. \tag{3.31}$$

For $n \to \infty$, $h(E)$ converges to the differential entropy $h(e) = \frac{1}{2}\log 2\pi e\sigma^2$ of Gaussian noise. For reliable transmission, we can see that the fundamental volume $V(\epsilon L)$ should be larger than $V(E)$, i.e., $\delta \leq \delta^*$.

The lattice shaping problem was also considered in [53]. For a properly chosen shaping region $S \subset \mathbb{R}^n$, we would expect to obtain a code with about $M = V(S)/V$ codewords. The bound in (3.31) is rewritten as

$$\overline{P_{\text{amb}|E}} \lesssim (1+\alpha)2^{-n[h(S)-h(E)-R]}, \tag{3.32}$$

where $R \triangleq 1/n\log_2 M$ is the information rate of the lattice code in bits per dimension, and where $h(S) \triangleq 1/n\log_2 V(S)$. When $n \to \infty$, both $E$ and $S$ become

$n$-dimensional spheres, then we have $\lim_{n\to\infty} h(E) = h(e)$ and $\lim_{n\to\infty} h(S) = 1/2\log_2(2\pi eP)$, where $P$ is the average signal power per dimension. Therefore, it is concluded that the AWGN channel capacity $\frac{1}{2}\log_2(P/N)$ is achievable by using lattice codes and lattice decoding in the sense that the error probability can be made arbitrarily small (but positive).

## 3.3 Lattices constructed from error correction codes

Lattices can be constructed by different ways. In mathematics, lattices are usually constructed from the sphere packing theory [25]. In cryptography, people prefer the group theory for lattice construction. However, in communications, most lattices are lifted from the existing error correction codes [47]. The following table presents the summary of current lattice constructions from error correction codes.

Table 3.1: Lattice constructions from error correction codes [3].

| Name | References | Descriptions | Lattices |
|---|---|---|---|
| Construction $A$ | [55] [25, p137] | Single level, binary and non-binary codes | $D_4$, $(8,4,4)$ Hamming Code and $E_8$, Random mod-$p$ lattices [53] [19], LDA lattices [56]. |
| Construction $B$ | [55] [25, 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 $\Lambda_{16}$, $(24,12,8)$ Golay code and Leech lattice $\Lambda_{24}$,. |
| Construction $D$ | [57] [25, p232] | Multilevel, nested binary linear codes, deals with generator matrix | Barnes-Wall lattices, Polar lattices [51], Turbo lattices [58]. |
| Construction $D'$ | [57] [25, p235] | Multilevel, nested binary linear codes, deals with parity-check matrix | LDPC lattices [59] |
| Construction $E$ | [60] [25, p236] | Multilevel, nested binary linear codes, higher dimensional lattice partition | Polar lattices [51] |

### 3.3.1 Single level constructions

In this subsection, we briefly introduce Construction $A$. The general definition of Construction $A$ can be found at [25, p211]. Let $\Pi : \Lambda^N \to (\Lambda/\Lambda')^N$ be the natural projection. The lattice $L$ generated from Construction $A$ is defined as:

$$L = \{\overline{x} \in \Lambda^N \mid \Pi(\overline{x}) \in C\}, \tag{3.33}$$

where $C(N, k, d_{min})$ is an $\mathbb{F}_p$-linear code with length $N$, dimension $k$ and minimum distance $d_{min}$.

A simple case for Construction $A$ is the mod-$p$ lattice used in [53], where $\Lambda/\Lambda' = \mathbb{Z}/p\mathbb{Z}$. The fundamental volume of such a lattice is $V(L) = p^{N-k}$.

Let $I_k$ denote the $k$-by-$k$ identity matrix. The systematic form of the code $C$ is given by $(I_k,\ B)$. The generator matrix for $L$ has the form

$$\begin{pmatrix} I_k & \Phi(B) \\ 0 & pI_{N-k} \end{pmatrix} \tag{3.34}$$

where $\Phi : \mathbb{F}_p \to \mathbb{Z}$ is a natural embedding from $\mathbb{F}_p$ to $\mathbb{Z}$. Moreover, the minimum distance of $L$ is given by

$$d_{\min}(L) = \min\{\sqrt{d_{\min}(C)}, d_{\min}(\Lambda')\} = \min\{\sqrt{d_{\min}(C)}, p\}. \tag{3.35}$$

Consequently, Construction $A$ cannot be successful for large dimension when $p$ is small. To improve the minimum distance of the mod-$p$ lattices, we may use a more powerful code $C$ and a finite field with larger size $p$.

#### 3.3.1.1 Gosset lattice $E_8$

In this subsection, we give a brief review on the Gosset lattice $E_8$. The Gosset lattice $E_8$ is famous for having the densest packing efficiency among all known 8-dimensional lattices. One code formula of $E_8$ is given by

$$L = C(8, 4, 4) + 2\mathbb{Z}^8, \tag{3.36}$$

where $C(8, 4, 4)$ denotes the $(8, 4, 4)$ Hamming code. Clearly, it is constructed by the binary partition $\mathbb{Z}/2\mathbb{Z}$ and Construction A.

There are $2^4$ codewords of the $(8, 4, 4)$ Hamming code. From the viewpoint of coset codes [47], $E_8$ is the combination of the $2^4$ cosets of $2\mathbb{Z}^8$ which are chosen by the $(8, 4, 4)$ Hamming Code.

#### 3.3.1.2 Low-density integer lattices [56]

Low-density integer lattices (LDA) are constructed from non-binary LDPC codes [56] and Construction A. Similar to LDPC codes, LDA lattices can be decoded efficiently by the iterative message-passing algorithm based on their factor graphs. In [56], the authors gave an example using a $(2, 5)$-regular LDPC code and a non-binary partition $\mathbb{Z}/p\mathbb{Z}$ with $p = 11$. The decoding complexity is $O(p^2 N \log N)$.

In 2013, it was further proved that two particular families of the LDA lattices can achieve the Poltyrev capacity under lattice decoding [61, 62]. The difference lies in the number of non-zero coefficients $h_i$ in a parity-check equation $\sum_{i=1}^{n} h_i x_i \equiv 0 \bmod p$, which is called the row degree of LDA lattices. [61] shows that Poltyrev capacity can be achieved with LDA lattices when the row degree is growing logarithmically. This proof is inspired by the result proved by Gallager that binary LDPC codes need logarithmically growing row degrees to achieve the capacity of the binary symmetric channel. A stronger statement is then given in [62], which

shows that Poltyrev capacity can be attained also by LDA lattices with constant row degrees. The typical settings of the LDA lattices to guarantee achieving the Poltyrev capacity are $p = n^{\frac{1}{2}}$ and the degree should at least be 7. It remains an open problem that whether the degree can be reduced to 2.

### 3.3.2 Multilevel constructions

From the previous section, we know that the single level construction cannot generate high dimensional lattices with large coding gain if the component codes are binary codes. However, by using multilevel construction, we can stack a series of binary nested capacity-achieving codes and generate high dimensional lattices with large coding gain. We define Construction $D$ as follows.

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 $\overline{b}_1, \overline{b}_2, \cdots, \overline{b}_N$ be a basis of $\mathbb{F}_2^N$ such that $\overline{b}_1, \cdots \overline{b}_{k_i}$ span $C_i$. The lattice $\Lambda_D$ consists of all vectors of the form [57]

$$\sum_{i=0}^{a-1} 2^i \sum_{j=1}^{k_i} \alpha_j^{(i)} \psi(\overline{b}_j) + 2^a \mathbb{Z}^N, \tag{3.37}$$

where $\alpha_j^{(i)} \in \{0, 1\}$, and $\psi$ be the natural embedding of $\mathbb{F}_2^N$ into $\mathbb{Z}^N$.

Particularly, when the lattice partition chain is $\mathbb{Z}/2\mathbb{Z}/\cdots/2^a\mathbb{Z}$, the fundamental volume of $\Lambda_D$ is given by

$$V(\Lambda_D) = 2^{-N\sum_{i=1}^{a-1} k_i} V(\Lambda_a)^N = (2^{a-1})^N \cdot 2^{N-\sum_{i=1}^{a-1} k_i}. \tag{3.38}$$

For construction D with the 1-dimensional binary 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 $X_i \in \mathcal{X} = \{0, 1\}, Y \in \mathcal{Y}$. The channel of the $\ell$-th level is a well-

defined $2^{\ell-1}\mathbb{Z}/2^{\ell}\mathbb{Z}$ channel [50]. Given uniformly distributed $x_{1:\ell-1}$, letting $\mathcal{A}_\ell(x_{1:\ell})$ 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 [50]

$$\begin{aligned} P_{\bar{Y}_\ell|X_\ell,X_{1:\ell-1}}(\bar{y}_\ell|x_\ell, x_{1:\ell-1}) &= f_{\sigma,2^{\ell}\mathbb{Z}}(\bar{y}_\ell - x_1 - \cdots - 2^{\ell-1}x_\ell) \\ &= \sum_{\boldsymbol{\lambda}\in 2^{\ell}\mathbb{Z}} f_{\sigma,\boldsymbol{\lambda}}(\bar{y}_\ell - x_1 - \cdots - 2^{\ell-1}x_\ell) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} \exp\left(-\frac{1}{2\sigma^2}|\bar{y}_\ell - a|^2\right). \end{aligned} \tag{3.39}$$

This channel is the key to construct AWGN-good multilevel lattices. More details are introduced in Chapter 4.

#### 3.3.2.1 Barnes-Wall lattices

Barnes-Wall lattices are constructed from Reed-Muller (RM) codes. Conventionally, RM codes are denoted by RM$(r, m)$ $(0 \le r \le m)$, with parameters $N$, $k$ and $d$ given by

$$N = 2^m, k = 1 + \begin{pmatrix} m \\ 1 \end{pmatrix} + \cdots + \begin{pmatrix} m \\ r \end{pmatrix}, d = 2^{m-r}, \tag{3.40}$$

respectively.

Barnes-Wall lattices are the densest lattices known in $4$, $8$, and $16$ dimensions[1]. Following the code formulas given in [64], one may interpret the close relationship between Barnes-Wall lattices and RM codes as follows:

[1] We give the example of Barnes-Wall lattices as a benchmark particularly because of the close connection between RM codes and polar codes [63]. As we show in Figure 4.8, the advantage of polar codes over RM codes will translate into the advantage of polar lattices over Barnes-Wall lattices.

For even $m-r$:

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

For odd $m-r$:

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

Alternatively, one may use the complex code formula:

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

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

For example, the code formula of the Barnes-Wall lattice with dimension 1024 can be written as:

$$BW_{1024} = \text{RM}(1,10) + 2\text{RM}(3,10) + \cdots + 2^5\mathbb{Z}^{1024}. \tag{3.41}$$

We note that the normalized fundamental volume of Barnes-Wall lattices is $2^{\frac{m}{2}}$, and the minimum squared Euclidean distance is $2^m$, which also means 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 (BDD) efficiently [65].

#### 3.3.2.2 Low-density parity-check lattices [59]

Low-density parity-check lattices are constructed by a set of nested LDPC codes [59] according to Construction D' [25, p232]. This kind of lattices can be represented by a Tanner graph, 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 sequentially, using the standard belief propogation algorithm. This reduces the complexity at the expense of some loss in error performance. It was suggested to use the progressive-edge-growth algorithm (PEG) to find good component LDPC codes with lager girth. Furthermore, lattices based on irregular LDPC codes were proposed in [66]. Examples and decoding algorithms can be found in [59].

Instead of defining a lattice by the generator matrix, Construction $D'$ gives a definition based on the parity check matrix of the linear code at each level.

Suppose $C_0 \supseteq C_1 \supseteq \cdots \supseteq C_{a-1} \supseteq C_a$ be a series of nested binary linear codes, where $C_i$ has parameters $(N, k_i, d_i)$. Let $C_i^*$ denote the dual code of $C_i$. 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 constructed by Construction $D'$. The parity-check matrix of $\Lambda_{D'}$ is given by

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

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

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

#### 3.3.2.3 Turbo lattices (2010)

When turbo codes are used in Construction D, we get turbo lattices [58]. The turbo code for each level 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 turbo codes. The code formula for the turbo lattice is

$$\Lambda_{\text{TC}} = C_1 + 2C_2 + \cdots + 2^{a-1}C_{a-1} + 2^a\mathbb{Z}^N, \tag{3.42}$$

which is similar to that of Barnes-Wall lattices.

Consider two nested turbo codes $C_0 \subseteq C_1$ generated by a generator matrix $G$ of a convolutional code with size $K \times N$ and a random interleaver $\Pi$ with size $k = LK$. Each interleaver can be represented by a permutation matrix $P_{k\times k}$. Therefore the generator matrix of $C_2$ has form

$$G_{\text{TC}} = \begin{pmatrix} I_k & F & PF \end{pmatrix}, \tag{3.43}$$

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

$$G_{\text{TL}} = \begin{pmatrix} I_{k_1} & 0 & F_1 & P_1F_1 \\ 0 & 2I_{k_2} & 2F_2 & 2P_2F_2 \\ 0 & 0 & 4I_{k_3} & 0 \\ 0 & 0 & 0 & 4I_{k_3} \end{pmatrix}, \tag{3.44}$$

Taking the same example in [58], consider a component turbo code $C_2$ with

generator matrix

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

The resulting turbo code has rate $R_2 = \frac{1}{2}$, and $d_2 = 13$ when $k_2 = 400$. Consider,the component encoder for $C_1$ with the following generator matrix

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

which is the first row of the generator matrix of $C_2$. It can be checked that $C_1$ has rate $R_1 = \frac{1}{3}$ and $d_1 = 28$ when $k_1 = 576$.

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 of the turbo lattice $\Lambda_{\text{TL}}$ is

$$G_{\text{TL}} = \begin{pmatrix} I_{576} & 0 & F_1 & P_1F_1 \\ 0 & 2I_{324} & 2F_2 & 2P_2F_2 \\ 0 & 0 & 4I_{500} & 0 \\ 0 & 0 & 0 & 4I_{500} \end{pmatrix}, \tag{3.45}$$

and its minimum distance is $\min\{4, d_1, d_2\} = 4$.

#### 3.3.2.4 Polar lattices (2013)

Polar lattices are generated by a set of nested polar codes according to Construction D or Construction E [2]. Since polar lattices are as explicit as polar codes, their construction is equally efficient. Furthermore, compared with the above existing schemes [59, 56, 67, 58], polar lattices are distinguished by their provable AWGN-

[2] We usually say that we use Construction D to build polar lattices, because we choose low-dimensional partition chain for simplicity.

goodness, namely, they asymptotically achieve the sphere bound with multistage decoding, as we will show in Chapter 4.

In simple words, the construction of polar lattices requires a set of nested binary polar codes $\mathcal{P}_\ell$ with block length $N$ and dimension of information bits $k_\ell$ for $1 \leq \ell \leq r$. Let $\mathcal{P}_1 \subseteq \mathcal{P}_2 \cdot \cdot \subseteq \mathcal{P}_r$ for the nesting requirement. Denote by $\psi$ be the natural embedding of $\mathbb{F}_2^N$ into $\mathbb{Z}^N$. Consider $\mathrm{g}_1, \mathrm{g}_2, \cdots, \mathrm{g}_N$ be a basis of $\mathbb{F}_2^N$ such that $\mathrm{g}_1, \cdots \mathrm{g}_{k_\ell}$ span $\mathcal{P}_\ell$. When $n = 1$, the polar lattice $L$ consists of all vectors of the form

$$\sum_{\ell=1}^{r} 2^{\ell-1} \sum_{j=1}^{k_\ell} \alpha_j^{(\ell)} \psi(\mathrm{g}_j) + 2^r z, \tag{3.46}$$

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

### 3.3.3 Other lattices

Besides constructing from error-correcting codes, there are other lattices, such as the convolutional lattice codes [68] (also called as signal codes) and low-density lattice codes [67], which 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. For brevity, we skip the details of these lattices, and the rest of the thesis will mainly focus on polar lattices.

## 3.4 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 novel-

ty 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 4.

CHAPTER 4

# Polar Lattices for AWGN Channels

## 4.1 Introduction

A fast-decodable, structured code achieving the capacity of the power-constrained additive white Gaussian-noise (AWGN) channel is the dream goal of coding theory. Polar codes, proposed by Arıkan in [63], can provably achieve the capacity of binary memoryless symmetric (BMS) channels. There are considerable efforts to extend polar codes to general discrete memoryless channels, to nonbinary polar codes, and to asymmetric channels [69, 70, 71, 72, 10, 73]. A largely theoretic attempt to construct polar codes for the AWGN channel was given in [74, 75], based on nonbinary polar codes or on the technique for the multi-access channel. However, these methods turn out to be inefficient in the sense of coding complexity. In this chapter, we propose polar lattices to fulfil this goal, based on a combination of binary polar codes and multilevel lattice codes.

Lattice codes are the counterpart of linear codes in the Euclidean space. It is well known that the design of a lattice code consists of two essentially separate problems: AWGN coding and shaping. AWGN coding is addressed by the notion of AWGN-good lattices [49, 19]. Informally, AWGN-goodness means that if the fundamental volume of the lattice is slightly greater than that of the “noise sphere", the error probability of infinite lattice decoding could be made arbitrarily small. Recently, several new lattice constructions with good performance have been introduced [59, 56, 76, 67]. On the other hand, shaping takes care of the finite power constraint

of the Gaussian channel. Capacity-achieving shaping techniques include Voronoi shaping [19] and lattice Gaussian shaping [20]. Despite these significant progresses, an explicit construction of lattice codes achieving the capacity of the Gaussian channel has been missing up until now.

We settle this open problem by employing the powerful tool of polarization in lattice construction. The technical ingredients of this work are the following:

- The construction of polar lattices and the proof of their AWGN-goodness. We follow the multilevel construction of Forney, Trott and Chung [50], where for each level we build a polar code to achieve its capacity. A salient feature of the proposed method is that it naturally leads to a set of nested polar codes, as required by the multilevel construction. This compares favorably with existing multilevel constructions [59], where extra efforts are needed to nest the component codes.

- The Gaussian shaping technique for polar lattices in the power-constrained AWGN channel. This is based on source polarization. We are able to achieve the capacity $\frac{1}{2}\log(1+\mathsf{SNR})$ with low-complexity multistage successive cancellation (SC) decoding for any given signal-to-noise ratio (SNR). It is worth mentioning that our proposed shaping scheme is not only a practical implementation of lattice Gaussian shaping, but also an improvement in the sense that we successfully remove the restriction $\mathsf{SNR} > e$ in [20, Theorem 3].

Both source and channel polarization are employed in the construction, resulting in an integrated approach in the sense that error correction and shaping are performed by one single polar code on each level. Further, it is worth pointing out that each aspect may also be of independent interest. AWGN-good lattices have many applications in network information theory (e.g., the aforementioned compute-and-forward and Wyner-Ziv coding), while lattice Gaussian shaping, i.e., generating a

Gaussian distribution over a lattice, is useful in lattice-based cryptography as well [77].

Both theoretic and practical aspects of polar lattices are addressed in this chapter. We not only prove the theoretic goodness of polar lattices, but also give practical rules for designing these lattices.

## 4.2 Construction of Polar Lattices

As reviewed in the preceding chaptor, achieving the channel capacity involves an AWGN-good lattice. Forney *et al.* gave single and multilevel constructions of AWGN-good lattices in [50]. We now follow their multilevel approach to construct polar lattices. In order to achieve the capacity of the AWGN channel with the noise variance $\sigma^2$, the concerned noise variance for the AWGN-good lattice is $\tilde{\sigma}^2$ (recall $\tilde{\sigma} \triangleq \frac{\sigma_s \sigma}{\sqrt{\sigma_s^2+\sigma^2}}$), which is the variance of the equivalent noise after MMSE rescaling [20]. This methodology can also be justified by the equivalence lemma in the next section (see Lemma 4.3.11).

### 4.2.1 Forney *et al.*'s Construction Revisited

Recall that a mod-$\Lambda$ Gaussian channel is a Gaussian channel with an input in any fundamental region $\mathcal{R}(\Lambda)$ and with a mod-$\mathcal{R}(\Lambda)$ operator at the receiver front end [50]. The capacity of the mod-$\Lambda$ channel for noise variance $\tilde{\sigma}^2$ is

$$C(\Lambda, \tilde{\sigma}^2) = \log V(\Lambda) - h(\Lambda, \tilde{\sigma}^2), \tag{4.1}$$

where $h(\Lambda, \tilde{\sigma}^2) = -\int_{\mathcal{V}(\Lambda)} f_{\tilde{\sigma},\Lambda}(x) \log f_{\tilde{\sigma},\Lambda}(x) dx$ is the differential entropy of the $\Lambda$-aliased noise over $\mathcal{V}(\Lambda)$.

*Remark* 4.2.1. A mod-$\Lambda$ Gaussian channel with noise variance $\sigma_1^2$ is degraded with

respect to one with noise variance $\sigma_2^2$ if $\sigma_1^2 > \sigma_2^2$. Let $\tilde{W}_1$ and $\tilde{W}_2$ denote the two channels respectively. Consider an intermediate channel $\tilde{W}'$ which is also a mod-$\Lambda$ channel, with noise variance $\sigma_1^2 - \sigma_2^2$. By the property $[X \mod \Lambda + Y] \mod \Lambda = [X + Y] \mod \Lambda$, it is easy to see that $\tilde{W}_1$ is stochastically equivalent to a channel constructed by concatenating $\tilde{W}_2$ with $\tilde{W}'$. Therefore, $C(\Lambda, \sigma_1^2) < C(\Lambda, \sigma_2^2)$, and $\mathfrak{h}(\Lambda, \sigma_1^2) > \mathfrak{h}(\Lambda, \sigma_2^2)$.

Given lattice partition $\Lambda/\Lambda'$, the $\Lambda/\Lambda'$ channel is a mod-$\Lambda'$ channel whose input is restricted to discrete lattice points in $(\Lambda + a) \cap \mathcal{R}(\Lambda')$ for some translate $a$. The capacity of the $\Lambda/\Lambda'$ channel is given by

$$\begin{aligned} C(\Lambda/\Lambda', \tilde{\sigma}^2) &= C(\Lambda', \tilde{\sigma}^2) - C(\Lambda, \tilde{\sigma}^2) \\ &= h(\Lambda, \tilde{\sigma}^2) - h(\Lambda', \tilde{\sigma}^2) + \log V(\Lambda')/V(\Lambda). \end{aligned} \tag{4.2}$$

Further, if $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'$ is a lattice partition chain, then

$$C(\Lambda/\Lambda', \tilde{\sigma}^2) = C(\Lambda/\Lambda_1, \tilde{\sigma}^2) + \cdots + C(\Lambda_{r-1}/\Lambda', \tilde{\sigma}^2). \tag{4.3}$$

The key idea of [50] is to use a good component code $\mathcal{C}_\ell$ to achieve the capacity $C(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}^2)$ for each level $\ell = 1, 2, \ldots, r$ in Construction D. For such a construction, the total decoding error probability with multistage decoding is bounded by

$$P_e(L, \tilde{\sigma}^2) \le \sum_{\ell=1}^{r} P_e(\mathcal{C}_\ell, \tilde{\sigma}^2) + P_e((\Lambda')^N, \tilde{\sigma}^2). \tag{4.4}$$

To make $P_e(L, \tilde{\sigma}^2) \to 0$, we need to choose the lattice $\Lambda'$ such that $P_e((\Lambda')^N, \tilde{\sigma}^2) \to 0$ and the codes $\mathcal{C}_\ell$ for the $\Lambda_{\ell-1}/\Lambda_\ell$ channels whose error probabilities also tend to zero.

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

$$\begin{aligned} \log\left(\frac{\gamma_L(\tilde{\sigma})}{2\pi e}\right) &= \log \frac{V(L)^{\frac{2}{nN}}}{2\pi e\tilde{\sigma}^2} \\ &= \log \frac{2^{-\frac{2}{n}R_{\mathcal{C}}} V(\Lambda')^{\frac{2}{n}}}{2\pi e\tilde{\sigma}^2} \\ &= -\frac{2}{n}R_{\mathcal{C}} + \frac{2}{n}\log V(\Lambda') - \log 2\pi e\tilde{\sigma}^2. \end{aligned} \tag{4.5}$$

Define

$$\begin{cases} \epsilon_1 = C(\Lambda_1, \tilde{\sigma}^2) \\ \epsilon_2 = h(\tilde{\sigma}^2) - h(\Lambda', \tilde{\sigma}^2) \\ \epsilon_3 = C(\Lambda/\Lambda', \tilde{\sigma}^2) - R_{\mathcal{C}} = \sum_{\ell=1}^{r} C(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}^2) - R_\ell, \end{cases} \tag{4.6}$$

where $h(\tilde{\sigma}^2) = \frac{n}{2}\log 2\pi e\tilde{\sigma}^2$ is the differential entropy of the Gaussian noise. We note that, $\epsilon_1 \geq 0$ represents the capacity of the mod-$\Lambda_1$ 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-$\Lambda_r$ Gaussian noise, and $\epsilon_3 \geq 0$ is the total capacity loss of component codes.

Then we have

$$\log\left(\frac{\gamma_L(\tilde{\sigma})}{2\pi e}\right) = \frac{2}{n}(\epsilon_1 - \epsilon_2 + \epsilon_3). \tag{4.7}$$

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

$$\log\left(\frac{\gamma_L(\tilde{\sigma})}{2\pi e}\right) \leq \frac{2}{n}(\epsilon_1 + \epsilon_3). \tag{4.8}$$

Since $\log\left(\frac{\gamma_L(\tilde{\sigma})}{2\pi e}\right) = 0$ represents the Poltyrev capacity, the right hand side of (4.8) gives an upper bound on the gap to the Poltyrev capacity. The bound is equal to

[1] It was shown in [50] that $\epsilon_2 \approx \pi P_e(\Lambda', \tilde{\sigma}^2)$, which is negligible compared to the other two terms.

$\frac{6.02}{n}(\epsilon_1 + \epsilon_3)$ decibels (dB), by conversion of the binary logarithm into the base-10 logarithm.

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 (4.8), we need that both $\epsilon_1$ and $\epsilon_3$ are negligible. In [78, Appendix A], we prove the following lemma.

**Lemma 4.2.1.** *The capacity of the mod-$\Lambda$ channel is bounded by*

$$C(\Lambda, \sigma^2) \le \log\left(1 + \epsilon_\Lambda(\sigma)\right) \le \log(e) \cdot \epsilon_\Lambda(\sigma). \tag{4.9}$$

Thus, we have the following design criteria:

- The top lattice $\Lambda$ has a negligible flatness factor $\epsilon_\Lambda(\sigma)$.
- The bottom lattice $\Lambda'$ has a small error probability $P_e(\Lambda', \sigma^2)$.
- Each component code $\mathcal{C}_\ell$ is a capacity-approaching code for the $\Lambda_{\ell-1}/\Lambda_\ell$ channel.

These conditions are essentially the same as those of Forney et al. [50], except that we impose a slightly stronger condition on the top lattice. In [50], the top lattice satisfies $C(\Lambda, \sigma^2) \approx 0$. The reason why we require negligible $\epsilon_\Lambda(\sigma)$ is to achieve the capacity of the power-constrained Gaussian channel. This will become clear in the next section.

Asymptotically, the error probability of polar codes decrease as $e^{-O(\sqrt{N})}$ and we may desire the same for the error probability of polar lattices. To make this happen, we may want $P_e(\Lambda', \sigma^2)$ to decrease exponentially since $P_e((\Lambda')^N, \sigma^2) \le N P_e(\Lambda', \sigma^2)$. It is easy to see that the number of levels $r = O(\log N)$ is sufficient such that $P_e(\Lambda', \sigma^2) = e^{-O(N)}$ (see [78, Appendix B] for a proof). In practical designs, if the target error probability (e.g., $P_e(L, \sigma^2) = 10^{-5}$) is fixed, a small number of levels will suffice.

### 4.2.2 Polar Lattices

It is shown in [50] that the $\Lambda_{\ell-1}/\Lambda_\ell$ channel is symmetric, and that the optimum input distribution is uniform. Since we use a binary partition $\Lambda_{\ell-1}/\Lambda_\ell$, the input $X_\ell$ is binary for $\ell \in 1, 2, \ldots, r$. Associate $X_\ell$ with representative $a_\ell$ of the coset in the quotient group $\Lambda_{\ell-1}/\Lambda_\ell$. The fact that the $\Lambda_{\ell-1}/\Lambda_\ell$ channel is a BMS channel allows a polar code to achieve its capacity.

Let $Y$ denote the output of the AWGN channel. Given $x_{1:\ell-1}$, let $\mathcal{A}_\ell(x_{1:\ell})$ denote the coset chosen by $x_\ell$, i.e., $\mathcal{A}_\ell(x_{1:\ell}) = a_1 + \cdots + a_\ell + \Lambda_\ell$. Assuming a uniform input distribution for all $X_\ell$, the conditional PDF of this $\Lambda_{\ell-1}/\Lambda_\ell$ channel with input $x_\ell$ and output $\bar{y}_\ell = y \bmod \Lambda_\ell$ is given by[79, (5)]

$$\begin{aligned} P_{\bar{Y}_\ell|X_\ell,X_{1:\ell-1}}(\bar{y}_\ell|x_\ell, x_{1:\ell-1}) &= f_{\sigma,\Lambda_\ell}(\bar{y}_\ell - a_1 - \cdots - a_\ell) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \sum_{a \in \mathcal{A}_\ell(x_{1:\ell})} \exp\left(-\frac{\|\bar{y}_\ell - a\|^2}{2\sigma^2}\right). \quad (4.10) \end{aligned}$$

In [50], this conditional PDF is written in a somewhat different from. Namely, the conditional PDF is $f_{\sigma,\Lambda_\ell}(\bar{y}_\ell - a - a_\ell)$ with an offset $a$. Nevertheless, the two forms are equivalent because we can let the offset $a = a_1 + \cdots + a_{\ell-1}$. The regularity (symmetry) and capacity separability [50, Th. 4 and Th. 5] of the $\Lambda/\Lambda'$ channel hold for any offset on its input. In fact, the offset due to previous input bits $x_{1:\ell-1}$ would be removed by the multistage decoder at level $\ell$, which means that the code for level $\ell$ can be designed according to (5.5) with $x_{1:\ell-1} = 0$. For this reason, we will fix $x_{1:\ell-1} = 0$ to prove channel degradation in the following lemma. The reason why we use the form (5.5) is for consistency with the case of non-uniform input $x_{1:\ell-1}$ in Sect. 4.3, where one cannot always let $x_{1:\ell-1} = 0$.

The proof of the following lemma is given in Appendix A.

**Lemma 4.2.2.** *Consider a self-similar binary lattice partition chain* $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'$, *in which we have* $\Lambda_\ell = T^\ell\Lambda$ *for all* $\ell$, *with* $T = \alpha V$ *for*

*some scale factor* $\alpha > 1$ *and orthogonal matrix* $V$. *Then, the* $\Lambda_{\ell-1}/\Lambda_\ell$ *channel is degraded with respect to the* $\Lambda_\ell/\Lambda_{\ell+1}$ *channel for* $1 \le \ell \le r-1$.

With component polar codes $\mathcal{P}(N, k_\ell)$ for all the $\Lambda_{\ell-1}/\Lambda_\ell$ channels $(1 \le \ell \le r)$, we stack them as in Construction D to build the polar lattice. The following lemma shows that these component codes are nested, which is to guarantee that the multilevel construction creates a lattice [50]. We consider two rules to determine the component codes, for theoretical and practical purposes, respectively. One is the *capacity rule* [50, 79], where we select the channel indices according to a threshold on the mutual information. The other is the *equal-error-probability rule* [79], namely, the same error probability for each level, where we select the channel indices 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 [63]. This nesting relation is a consequence of [30, Lemma 4.7].

**Lemma 4.2.3.** *For either the capacity rule or the equal-error-probability rule, the component polar codes built in the multilevel construction are nested, i.e.,* $\mathcal{P}(N, k_1) \subseteq \mathcal{P}(N, k_2) \subseteq \cdots \subseteq \mathcal{P}(N, k_r)$.

*Proof.* Firstly, consider the equal-error-probability rule. By [30, Lemma 4.7], if a BMS channel $\tilde{V}$ is a degraded version of $\tilde{W}$, then the subchannel $\tilde{V}_N^{(i)}$ is also degraded with respect to $\tilde{W}_N^{(i)}$ and $\tilde{Z}(\tilde{V}_N^{(i)}) \ge \tilde{Z}(\tilde{W}_N^{(i)})$. Let the threshold be $2^{-N^\beta}$ for some $\beta < 1/2$. The codewords are generated by $x^{1:N} = u^{\mathcal{I}} G_{\mathcal{I}}$, where $G_{\mathcal{I}}$ is the submatrix of $G$ whose rows are indexed by information set $\mathcal{I}$. The information sets

for these two channels are respectively given by

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

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

From Lemma 4.2.2, the channel of the $\ell$-th level is always degraded with respect to the channel of the $(\ell+1)$-th level, and consequently, $\mathcal{P}(N, k_\ell) \subseteq \mathcal{P}(N, k_{\ell+1})$.

Then, consider the capacity rule. The nesting relation still holds if we select the channel indices according to a threshold on the mutual information. This is because, by [30, Lemma 4.7], $I(\tilde{V}_N^{(i)}) \leq I(\tilde{W}_N^{(i)})$ if a BMS channel $\tilde{V}$ is a degraded version of $\tilde{W}$. □

### 4.2.3 AWGN Goodness

For a threshold $2^{-N^\beta}$ of the Bhattacharyya parameter, the block error probability of the polar code with SC decoding is upper-bounded by $N2^{-N^\beta}$. It can be made arbitrarily small by increasing the block length $N$. Also, the capacity loss $\epsilon_3$ diminishes as $N \to \infty$. Therefore, we have the following theorem:

**Theorem 4.2.4.** *Suppose $\epsilon_\Lambda(\tilde{\sigma})$ is negligible. Construct polar lattice $L$ from the $n$-dimensional binary lattice partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'$ and $r$ nested polar codes with block length $N$, where $r = O(\log N)$. Then, the error probability of $L$ under multistage decoding is bounded by*

$$P_e(L, \tilde{\sigma}^2) \leq rN2^{-N^\beta} + N\left(1 - \int_{\mathcal{V}(\Lambda')} f_{\tilde{\sigma}^2}(x)dx\right), \tag{4.11}$$

*with the logarithmic VNR bounded by (4.8). As $N \to \infty$, $L$ can achieve the Poltyrev*

*capacity, i.e.,* $\log\left(\frac{\gamma_L(\tilde{\sigma})}{2\pi e}\right) \to 0$ *and* $P_e(L, \tilde{\sigma}^2) \to 0$.

*Remark* 4.2.2. It is worth pointing out that Theorem 4.2.4 only requires mild conditions. The condition $\epsilon_\Lambda(\tilde{\sigma}) \to 0$ is easily satisfied by properly scaling the top lattice $\Lambda$. In practice, if the target error probability is fixed (e.g., $10^{-5}$), $r$ can be a small constant, namely, $r$ does not have to scale as $\log N$. Thus, the essential condition is $N \to \infty$.

For finite $N$, however, the capacity loss $\epsilon_3$ is not negligible. We investigate the finite-length performance of polar lattices in the following.

The finite-length analysis of polar codes was given in [80, 81, 82]. It was proved that polar codes need a polynomial block length with respect to the gap to capacity $\epsilon_{\text{loss}} = I(\tilde{W}) - R = O(N^{-\frac{1}{\mu}})$ [80, 81], where $\mu$ 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 $I(\tilde{W})$ and $\underline{\mu} = 3.55$ [80]. 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 [80]. Later this scaling factor $\bar{\mu}$ has been improved to $5.77$ [82].

Thus, the gap to the Poltyrev capacity of finite-dimensional polar lattices is

$$\log\left(\frac{\gamma_L(\tilde{\sigma})}{2\pi e}\right) \leq \frac{2}{n}\left(\epsilon_1 + r\bar{\beta} N^{-\frac{1}{\bar{\mu}}}\right) \tag{4.12}$$

with the corresponding block error probability

$$P_e(L, \tilde{\sigma}^2) \leq r P_B + P_e(\Lambda'^N, \tilde{\sigma}^2), \tag{4.13}$$

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 [83]

$$\log\left(\frac{\gamma_L(\tilde{\sigma})}{2\pi e}\right)_{\text{opt}} = \sqrt{\frac{2}{n_L}}Q^{-1}(P_e(L,\tilde{\sigma}^2)) - \frac{1}{n_L}\log n_L + O\left(\frac{1}{n_L}\right). \tag{4.14}$$

At finite dimensions, this is more precise than the exponential error bound for lattices constructed from random linear codes given in [50]. Thus, given $P_e(L,\sigma^2)$, the scaling exponent of optimum random lattices is 2 which is smaller than that of polar lattices $\bar{\mu}$. 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.

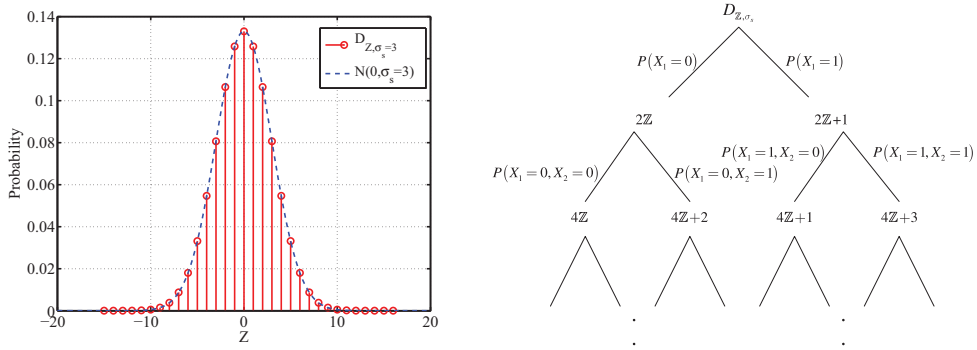
## 4.3 Polarization-Based Lattice Gaussian Shaping

To achieve the capacity of the power-constrained Gaussian channel, we can apply Gaussian shaping over the polar lattice $L$. However, it appears difficult to do so directly. In this section, we will apply Gaussian shaping to the top lattice $\Lambda$ instead, which is more friendly for implementation. This is motivated by Theorem 3.2.1, which implies that one may construct a capacity-achieving lattice code from a good constellation. More precisely, one may choose a low-dimensional top lattice such as $\mathbb{Z}$ and $\mathbb{Z}^2$ whose mutual information has a negligible gap to the channel capacity as bounded in Theorem 3.2.1, and then construct a multilevel code to achieve the capacity. We will show that this strategy is equivalent to implementing Gaussian shaping over the AWGN-good polar lattice. For this purpose, we will employ the recently introduced polar codes for asymmetric channels [10, 73].

### 4.3.1 Asymmetric Channels in Multilevel Lattice Coding

By Theorem 3.2.1, we choose a good constellation $D_{\Lambda,\sigma_s}$ such that the flatness factor $\epsilon_\Lambda(\tilde{\sigma})$ is negligible. Let the binary partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'/\cdots$ be la-

belled by bits $X_1, \cdots, X_r, \cdots$. Then, $D_{\Lambda,\sigma_s}$ induces a distribution $P_{X_{1:r}}$ whose limit corresponds to $D_{\Lambda,\sigma_s}$ as $r \to \infty$. An example for $D_{\mathbb{Z},\sigma_s}$ for $\sigma_s = 3$ is shown in Fig. 4.1. In this case, a shaping constellation with $M = 32$ points are actually sufficient, since the total probability of these points is rather close to 1.




(a) $D_{\mathbb{Z},\sigma_s}$ for $\sigma_s = 3$ (b) Bit labelling

Figure 4.1: Lattice Gaussian distribution $D_{\mathbb{Z},\sigma_s}$ and the associated labelling.

By the chain rule of mutual information

$$I(Y; X_{1:r}) = \sum_{\ell=1}^{r} I(Y; X_\ell | X_{1:\ell-1}), \tag{4.15}$$

we obtain $r$ binary-input channels $W_\ell$ for $1 \leq \ell \leq r$. Given $x_{1:\ell-1}$, denote again by $\mathcal{A}_\ell(x_{1:\ell})$ the coset of $\Lambda_\ell$ indexed by $x_{1:\ell-1}$ and $x_\ell$. According to [79], the channel

transition PDF of the $\ell$-th channel $W_\ell$ is given by

$$\begin{aligned}
&P_{Y|X_\ell,X_{1:\ell-1}}(y|x_\ell,x_{1:\ell-1}) \\
&\quad= \frac{1}{P\{\mathcal{A}_\ell(x_{1:\ell})\}} \sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} P(a)P_{Y|A}(y|a) \\
&\quad= \frac{1}{f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))} \sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} \frac{1}{2\pi\sigma\sigma_s}\exp\left(-\frac{\|y-a\|^2}{2\sigma^2}-\frac{\|a\|^2}{2\sigma_s^2}\right) \\
&\quad= \exp\left(-\frac{\|y\|^2}{2(\sigma_s^2+\sigma^2)}\right)\frac{1}{f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))}\frac{1}{2\pi\sigma\sigma_s} \\
&\qquad\qquad \cdot \sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} \exp\left(-\frac{\sigma_s^2+\sigma^2}{2\sigma_s^2\sigma^2}\left\|\frac{\sigma_s^2}{\sigma_s^2+\sigma^2}y-a\right\|^2\right) \\
&\quad= \exp\left(-\frac{\|y\|^2}{2(\sigma_s^2+\sigma^2)}\right)\frac{1}{f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))}\frac{1}{2\pi\sigma\sigma_s}\sum_{a\in\mathcal{A}_\ell(x_{1:\ell})}\exp\left(-\frac{\|\alpha y-a\|^2}{2\tilde{\sigma}^2}\right).
\end{aligned} \tag{4.16}$$

where we recall the MMSE coefficient $\alpha = \frac{\sigma_s^2}{\sigma_s^2+\sigma^2}$, and $\tilde{\sigma} = \frac{\sigma_s\sigma}{\sqrt{\sigma_s^2+\sigma^2}}$. In general, $W_\ell$ is asymmetric with the input distribution $P_{X_\ell|X_{1:\ell-1}}$, unless $\frac{f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))}{f_{\sigma_s}(\mathcal{A}_{\ell-1}(x_{1:\ell-1}))} \approx \frac{1}{2}$ which means that $\epsilon_{\Lambda_\ell}(\sigma_s)$ is negligible.

For a finite power, the number of levels does not need to be large. The following lemma shows in a quantitative manner how large $r$ should be in order to achieve the channel capacity. The proof can be found in Appendix B.

**Lemma 4.3.1.** *If $r = O(\log\log N)$, the mutual information of the bottom level $I(Y;X_r|X_{1:r-1}) \to 0$ as $N\to\infty$. Moreover, using the first $r$ levels only incurs a capacity loss $\sum_{\ell>r} I(Y;X_\ell|X_{1:\ell-1}) \le O(\frac{1}{N})$.*

*Remark* 4.3.1. The condition $r = O(\log\log N)$ is again theoretic. In practice, $r$ can be a small constant so that the difference between $I(Y;X_{1:r})$ and the capacity is negligible, as we will see from the example in the next section.

### 4.3.2 Polar Codes for Asymmetric Channels

Since the component channels are asymmetric, we need polar codes for asymmetric channels to achieve their capacity. Fortunately, polar codes for the binary memoryless asymmetric (BMA) channels have been introduced in [10, 73] recently.

**Definition 4.3.1** (Bhattacharyya Parameter for BMA Channel [8, 10])**.** Let $W$ be a BMA channel with input $X \in \mathcal{X} = \{0, 1\}$ and output $Y \in \mathcal{Y}$, and let $P_X$ and $P_{Y|X}$ denote The input distribution and channel transition probability, respectively. The Bhattacharyya parameter $Z$ for channel $W$ is the defined as

$$\begin{aligned} Z(X|Y) &= 2\sum_{y} P_Y(y)\sqrt{P_{X|Y}(0|y)P_{X|Y}(1|y)} \\ &= 2\sum_{y} \sqrt{P_{X,Y}(0,y)P_{X,Y}(1,y)}. \end{aligned} \tag{4.17}$$

Note that Definition 4.3.1 is the same as Definition 2.1.2 when $P_X$ is uniform.

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

**Lemma 4.3.2** (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*

$$Z(X|Y, Y') \leq Z(X|Y). \tag{4.18}$$

*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 simplicity, denote the distribution of $(X^i, Y^i)$ by $P_{XY} = P_X P_{Y|X}$ for $i \in [N]$. The following property of the polarized random variables $U^{1:N} = X^{1:N} G_N$ is well known.

**Theorem 4.3.3** (Polarization of Random Variables [10]). *For any $\beta \in (0, 0.5)$,*

$$\begin{cases} \lim\limits_{N\to\infty} \frac{1}{N} \left| \left\{ i : Z(U^i|U^{1:i-1}) \geq 1 - 2^{-N^\beta} \right\} \right| = H(X), \\ \lim\limits_{N\to\infty} \frac{1}{N} \left| \left\{ i : Z(U^i|U^{1:i-1}) \leq 2^{-N^\beta} \right\} \right| = 1 - H(X), \\ \lim\limits_{N\to\infty} \frac{1}{N} \left| \left\{ i : Z(U^i|U^{1:i-1}, Y^{1:N}) \geq 1 - 2^{-N^\beta} \right\} \right| = H(Y|X), \\ \lim\limits_{N\to\infty} \frac{1}{N} \left| \left\{ i : Z(U^i|U^{1:i-1}, Y^{1:N}) \leq 2^{-N^\beta} \right\} \right| = 1 - H(Y|X), \end{cases} \tag{4.19}$$

*and*

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

The Bhattacharyya parameter for asymmetric models was originally defined for distributed source coding in [8]. By the duality between channel coding and source coding, it can be also used to construct capacity-achieving polar codes for BMA channels [10]. Actually, $Z(U^i|U^{1:i-1})$ is the Bhattacharyya parameter for a single source $X$ (without side information).

The Bhattacharyya parameter of a BMA channel can be related to that of a symmetric channel. To this aim, we use a symmetrization technique which creates a binary-input symmetric channel $\tilde{W}$ from the BMA channel $W$. The following lemma was implicit in [10]; here we make it explicit.

**Lemma 4.3.4** (Symmetrization)**.** *Let $\tilde{W}$ be a binary-input channel with input $\tilde{X} \in \mathcal{X} = \{0, 1\}$ and output $\tilde{Y} \in \{\mathcal{Y}, \mathcal{X}\}$, built from the asymmetric channel $W$ as shown in Fig. 4.2. Suppose the input of $\tilde{W}$ is uniformly distributed, i.e., $P_{\tilde{X}}(\tilde{x} = 0) = P_{\tilde{X}}(\tilde{x} = 1) = \frac{1}{2}$. Then $\tilde{W}$ is a binary-input symmetric channel in the sense that $P_{\tilde{Y}|\tilde{X}}(y, x \oplus \tilde{x}|\tilde{x}) = P_{Y,X}(y, x)$.* [2]

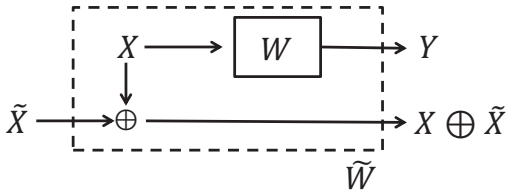



Figure 4.2: The relationship between the asymmetric channel $W$ and the symmetrized channel $\tilde{W}$.

*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 $(c)$ $P_{X \oplus \tilde{X}|X,\tilde{X}}(x \oplus \tilde{x}|x', \tilde{x}) = \mathbb{1}(x' = x)$. □

The following theorem connects the Bhattacharyya Parameter of a BMA channel $W$ and that of the symmetrized channel $\tilde{W}$. Denote by $W_N$ and $\tilde{W}_N$ the combining channels of $N$ uses of $W$ and $\tilde{W}$, respectively.

[2] We note that the definition of symmetric channel is slightly different from that given in [63] since we have a condition on the input distribution here. However, for the construction of polar codes and polar lattices, i.e., for the calculation of the Bhattacharyya parameter (see Theorem 4.3.5), we always assume uniform input distribution for the symmetrized channel. In this special case, the channel can be treated as a symmetric channel.

**Theorem 4.3.5** (Connection Between Bhattacharyya Parameters [10]). *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* $\tilde{W}$*, respectively, and let* $U^{1:N}{=}X^{1:N}G_N$ *and* $\tilde{U}^{1:N}{=}\tilde{X}^{1:N}G_N$*. The Bhattacharyya parameter of each subchannel of* $W_N$ *is equal to that of each subchannel of* $\tilde{W}_N$*, i.e.,*

$$Z(U^i|U^{1:i-1}, Y^{1:N}) = \tilde{Z}(\tilde{U}^i|\tilde{U}^{1:i-1}, X^{1:N} \oplus \tilde{X}^{1:N}, Y^{1:N}). \tag{4.20}$$

Now, we are in a position to construct polar codes for the BMA channel. Define the frozen set $\tilde{\mathcal{F}}$ and information set $\tilde{\mathcal{I}}$ of the symmetric polar codes as follows:

$$\begin{cases} \text{frozen set: } \tilde{\mathcal{F}} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{1:N}) > 2^{-N^\beta}\} \\ \text{information set: } \tilde{\mathcal{I}} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{1:N}) \leq 2^{-N^\beta}\}. \end{cases} \tag{4.21}$$

By Theorem 4.3.5, 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 $W$. To obtain the real capacity $I(X;Y)$ of $W$, the input distribution of $W$ needs to be adjusted to $P_X$. By polar lossless source coding, the indices with very small $Z(U^i|U^{1:i-1})$ should be removed from the information set $\tilde{\mathcal{I}}$ of the symmetrized channel, and the proportion of this part is $1 - H(X)$ as $N \to \infty$. We name this set as the information set $\mathcal{I}$ of the asymmetric channel $W$. Further, there are some bits which are uniformly distributed and can be made independent from the information bits; we name this set as the frozen set $\mathcal{F}$. In order to generate the desired input distribution $P_X$, the remaining bits are determined from the bits in $\mathcal{F} \cup \mathcal{I}$; we call it the shaping set $\mathcal{S}$. This process is depicted in Fig. 4.3. We formally

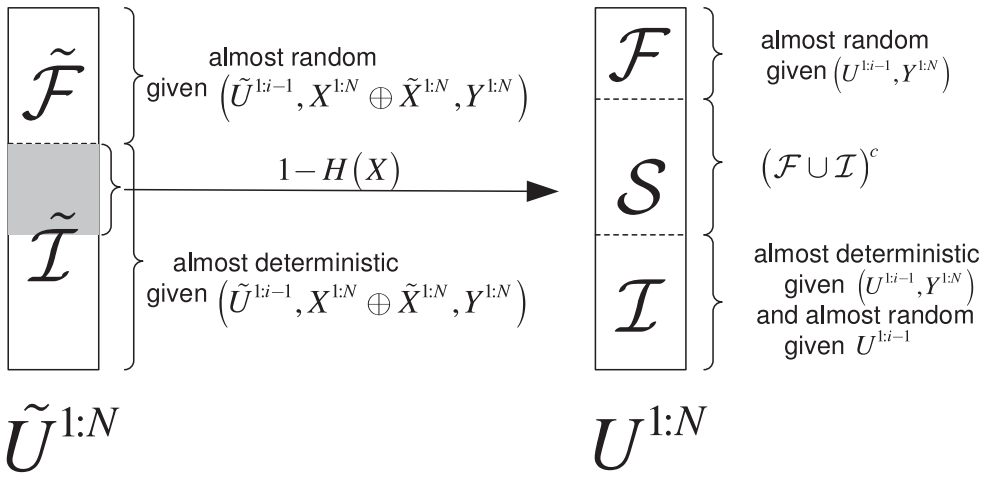



Figure 4.3: Polarization for symmetric and asymmetric channels.

define the three sets as follows:

$$\begin{cases} \text{frozen set: } \mathcal{F} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \text{information set: } \mathcal{I} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{1:N}) \leq 2^{-N^\beta} \text{ and} \\ \qquad\qquad Z(U^i|U^{1:i-1}) \geq 1 - 2^{-N^\beta}\} \\ \text{shaping set: } \mathcal{S} = (\mathcal{F} \cup \mathcal{I})^c \,. \end{cases} \tag{4.22}$$

To find these sets, one can use Theorem 4.3.5 to calculate $Z(U^i|U^{1:i-1}, Y^{1:N})$ with the known technique for symmetric polar codes [84, 5]. We note that $Z(U^i|U^{1:i-1})$ can be computed in a similar way: one constructs a symmetric channel between $\tilde{X}$ and $X \oplus \tilde{X}$, which is actually a binary symmetric channel with cross probability $P_X(x = 1)$. The above construction is equivalent to implementing shaping over the polar code for the symmetrized channel $\tilde{W}$.

Besides the construction, the decoding can also be converted to that of the symmetric polar code. If $X^{1:N} \oplus \tilde{X}^{1:N} = 0$, we have $U^{1:N} = \tilde{U}^{1:N}$, which means the decoding result of $U^{1:N}$ equals to that of $\tilde{U}^{1:N}$. Thus, decoding of the polar code for $W$ can be treated as decoding of the polar code for $\tilde{W}$ given that $X \oplus \tilde{X} = 0$. Clearly, the SC decoding complexity for asymmetric channel is also $O(N \log N)$.

We summarize this observation as the following lemma.

**Lemma 4.3.6** (Decoding for Asymmetric Channel [10])**.** *Let $y^{1:N}$ be a realization of $Y^{1:N}$ and $\hat{u}^{1:i-1}$ be the previous $i-1$ estimates of $u^{1:N}$. The likelihood ratio of $u^i$ is given by*

$$\frac{P_{U^i|U^{1:i-1},Y^{1:N}}(0|\hat{u}^{1:i-1},y^{1:N})}{P_{U^i|U^{1:i-1},Y^{1:N}}(1|\hat{u}^{1:i-1},y^{1:N})}=\frac{\tilde{W}_N^{(i)}((y^{1:N},0^{1:N}),\hat{u}^{1:i-1}|0)}{\tilde{W}_N^{(i)}((y^{1:N},0^{1:N}),\hat{u}^{1:i-1}|1)}, \tag{4.23}$$

*where $\tilde{W}_N^{(i)}$ denotes the transition probability of the $i$-th subchannel of $\tilde{W}_N$.*

In [10], the bits in $\mathcal{F}\cup\mathcal{S}$ are all chosen according to $P_{U^i|U^{1:i-1}}(u^i|u^{1:i-1})$, which can also be calculated using (4.23) (treating $Y$ as an independent variable and remove it). However, in order to be compatible with polar lattices, we modify the scheme such that the bits in $\mathcal{F}$ are uniformly distributed over $\{0,1\}$ while the bits in $\mathcal{S}$ are still chosen according to $P_{U^i|U^{1:i-1}}(u^i|u^{1:i-1})$. The expectation of the decoding error probability still vanishes with $N$. The following theorem is an extension of the result in [10, Theorem 3]. We give the proof in Appendix C for completeness.

**Theorem 4.3.7.** *Consider a polar code with the following encoding and decoding strategies for a BMA channel.*

- *Encoding: Before sending the codeword $x^{1:N}=u^{1:N}G_N$, the index set $[N]$ are divided into three parts: the frozen set $\mathcal{F}$, the information set $\mathcal{I}$ and the shaping set $\mathcal{S}$ which are defined in (5.17). The encoder places uniformly distributed information bits in $\mathcal{I}$, and fills $\mathcal{F}$ with a uniform random $\{0,1\}$ sequence which is shared between the encoder and the decoder. The bits in $\mathcal{S}$ are generated by a mapping $\phi_{\mathcal{S}}\triangleq\{\phi_i\}_{i\in\mathcal{S}}$ in the family of randomized mappings $\Phi_{\mathcal{S}}$, which*

*yields the following distribution:*

$$u^i = \begin{cases} 0 & \text{with probability } P_{U^i|U^{1:i-1}}(0|u^{1:i-1}), \\ 1 & \text{with probability } P_{U^i|U^{1:i-1}}(1|u^{1:i-1}). \end{cases} \tag{4.24}$$

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

$$\hat{u}^i = \begin{cases} u^i, & \text{if } i \in \mathcal{F} \\ \phi_i(\hat{u}^{1:i-1}), & \text{if } i \in \mathcal{S} \\ \underset{u}{\operatorname{argmax}}\ P_{U^i|U^{1:i-1},Y^{1:N}}(u|\hat{u}^{1:i-1}, y^{1:N}), & \text{if } i \in \mathcal{I}. \end{cases} . \tag{4.25}$$

*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 $E_{\Phi_\mathcal{S}}[P_e(\phi_\mathcal{S})] \leq N2^{-N^{\beta'}}$ for $\beta' < \beta < 0.5$. Consequently, there exists a deterministic mapping $\phi_\mathcal{S}$ such that $P_e(\phi_\mathcal{S}) \leq N2^{-N^{\beta'}}$.*

In practice, 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, e.g., using the same seed for the pseudorandom number generators.

### 4.3.3 Multilevel Polar Codes

Next, our task is to construct polar codes to achieve the mutual information $I(Y;X_\ell|X_{1:\ell-1})$ for all levels. The construction of the preceding subsection is readily applicable to the construction for the first level $W_1$. To demonstrate the construction for other levels, we take the channel of the second level $W_2$ as an example. This is also a BMA channel with input $X_2 \in \mathcal{X} = \{0,1\}$, output $Y \in \mathcal{Y}$ and side information $X_1$. Its channel transition probability is shown in (4.16). To construct a polar

code for the second level, we propose the following two-step procedure.

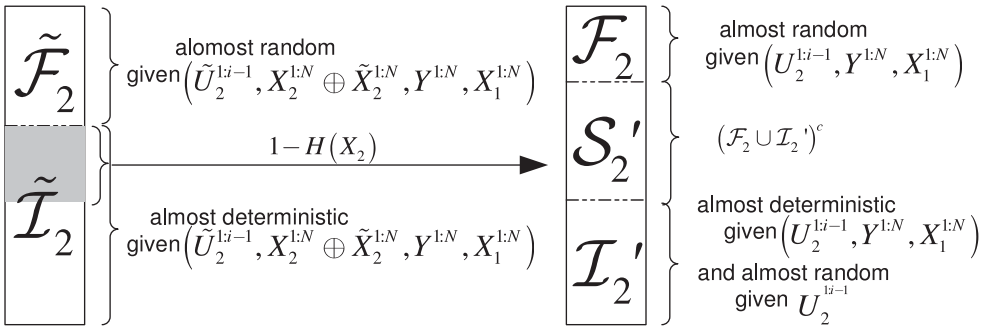



Figure 4.4: The first step of polarization in the construction for the second level.

Step 1: Construct a polar code for the BMS channel with input vector $\tilde{X}_2^{1:N} = [\tilde{X}_2^1, \tilde{X}_2^2, \cdots, \tilde{X}_2^N]$ and output vector $\tilde{Y}^{1:N} = \left(X_2^{1:N} \oplus \tilde{X}_2^{1:N}, Y^{1:N}, X_1^{1:N}\right)$ 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 $X_2$ 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 $\tilde{U}_2^{1:N} = \tilde{X}_2^{1:N} G_N$. According to Theorem 4.3.3, the polarization gives us the three sets $\mathcal{F}_2$, $\mathcal{I}_2'$ and $\mathcal{S}_2'$ as shown in Fig. 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 $N \to \infty$. These three sets are defined as follows:

$$\begin{cases} \text{frozen set: } \mathcal{F}_2 = \{i \in [N] : Z(U_2^i | U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \text{information set: } \mathcal{I}_2' = \{i \in [N] : Z(U_2^i | U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}) \leq 2^{-N^\beta} \\ \qquad\qquad \text{and } Z(U_2^i | U_2^{1:i-1}) \geq 1 - 2^{-N^\beta}\} \\ \text{shaping set: } \mathcal{S}_2' = (\mathcal{F}_2 \cup \mathcal{I}_2')^c. \end{cases} \tag{4.26}$$

Step 2: Treat $X_1^{1:N}$ as the side information for the encoder. Given $X_1^{1:N}$, the choices of $X_2^{1:N}$ are further restricted since $X_1$ and $X_2$ are generally correlated, i.e., $P_{X_1,X_2}(x_1, x_2) = f_{\sigma_s}(\mathcal{A}(x_1, x_2))/f_{\sigma_s}(\Lambda)$ (cf. Fig. 4.1). By removing from $\mathcal{I}_2'$ the bits which are almost deterministic given $U_2^{1:i-1}$ and $X_1^{1:N}$, we obtain

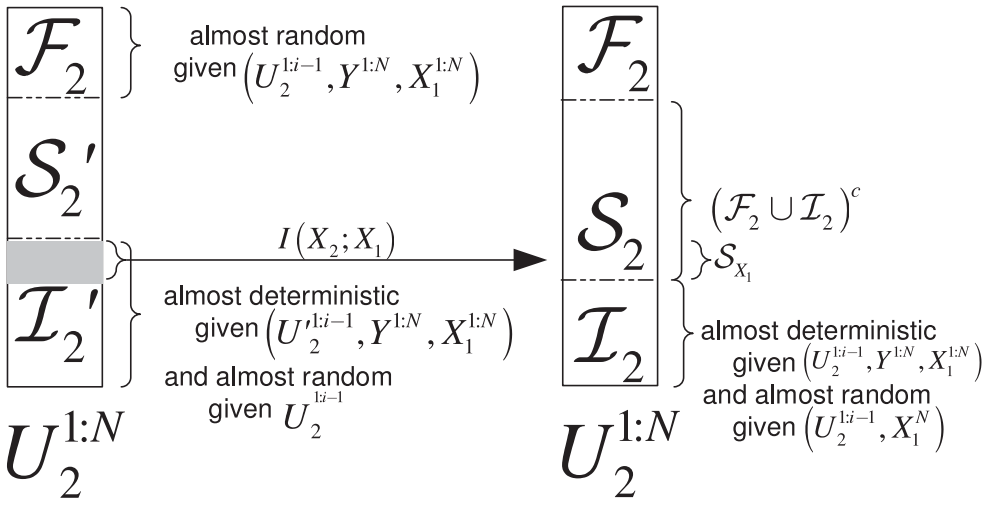



Figure 4.5: The second step of polarization in the construction for the second level.

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 Fig. 4.5. More precisely, the indices are divided into three portions as follows:

$$
\begin{aligned}
1 &= \underbrace{1 - I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y)}_{\mathcal{F}_2} + I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y) \\
&\overset{\text{Step1}}{=} \underbrace{1 - I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y)}_{\mathcal{F}_2} + \underbrace{I(\tilde{X}_2; \tilde{X}_2 \oplus X_2)}_{\mathcal{S}_2'} + \underbrace{I(\tilde{X}_2; X_1, Y | \tilde{X}_2 \oplus X_2)}_{\mathcal{I}_2'} \\
&\overset{\text{Step2}}{=} \underbrace{1 - I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y)}_{\mathcal{F}_2} + \underbrace{I(\tilde{X}_2; \tilde{X}_2 \oplus X_2)}_{\mathcal{S}_2'} + \underbrace{I(\tilde{X}_2; X_1 | \tilde{X}_2 \oplus X_2)}_{\mathcal{S}_{X_1}} \\
&\quad + \underbrace{I(\tilde{X}_2; Y | X_1, \tilde{X}_2 \oplus X_2)}_{\mathcal{I}_2} \\
&= \underbrace{1 - I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y)}_{\mathcal{F}_2} + \underbrace{1 - H(X_2)}_{\mathcal{S}_2'} + \underbrace{I(X_2; X_1)}_{\mathcal{S}_{X_1}} + \underbrace{I(X_2; Y | X_1)}_{\mathcal{I}_2} \\
&= \underbrace{1 - I(\tilde{X}_2; \tilde{X}_2 \oplus X_2, X_1, Y)}_{\mathcal{F}_2} + \underbrace{1 - H(X_2 | X_1)}_{\mathcal{S}_2} + \underbrace{I(X_2; Y | X_1)}_{\mathcal{I}_2}
\end{aligned}
\tag{4.27}
$$

We give the formal statement of this procedure in the following lemma.

**Lemma 4.3.8.** *After the first step of polarization, we obtain the three sets $\mathcal{F}_2$, $\mathcal{I}_2'$ and $\mathcal{S}_2'$ in (4.26). Let $\mathcal{S}_{X_1}$ denote the set of indices whose Bhattacharyya param-*

*eters satisfy* $Z(U_2^i|U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}) \leq 2^{-N^\beta}$, $Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}) \leq 1 - 2^{-N^\beta}$ *and* $Z(U_2^i|U_2^{1:i-1}) \geq 1 - 2^{-N^\beta}$. *The proportion of* $\mathcal{S}_{X_1}$ *is asymptotically given by* $\lim_{N\to\infty} \frac{|\mathcal{S}_{X_1}|}{N} = I(X_2; X_1)$. *Then by removing* $\mathcal{S}_{X_1}$ *from* $\mathcal{I}_2'$, *we obtain the true information set* $\mathcal{I}_2$ *for* $W_2$. *Formally, the three sets are obtained as follows:*

$$\begin{cases} \textit{frozen set: } \mathcal{F}_2 = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \textit{information set: } \mathcal{I}_2 = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, Y^{1:N}, X_1^{1:N}) \leq 2^{-N^\beta} \\ \qquad\qquad \textit{and } Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \textit{shaping set: } \mathcal{S}_2 = (\mathcal{F}_2 \cup \mathcal{I}_2)^c. \end{cases} \tag{4.28}$$

*Proof.* Firstly, we show the proportion of set $\mathcal{S}_{X_1}$ goes to $I(X_1; X_2)$ as $N \to \infty$. Here we define a slightly different set $\mathcal{S}'_{X_1} = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}) \leq 2^{-N^\beta}$ and $Z(U_2^i|U_2^{1:i-1}) \geq 1 - 2^{-N^\beta}\}$. Suppose we are constructing an asymmetric polar code for 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 4.3.7. Furthermore, by Lemma 4.3.2, 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 $\mathcal{S}'_{X_1}$ only lies on $Z(U_2^i|U_2^{1:i-1}, X_1^{1:N})$. Denoting by $\bar{\mathcal{P}}_{X_1}$ 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

$$\lim_{N\to\infty} \frac{|\mathcal{S}_{X_1}|}{N} - \frac{|\mathcal{S}'_{X_1}|}{N} \leq \lim_{N\to\infty} \frac{|\bar{\mathcal{P}}_{X_1}|}{N} = 0. \tag{4.29}$$

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

Secondly, we show that $\mathcal{S}_{X_1} \cup \mathcal{I}_2 = \mathcal{I}_2'$. Again, by Lemma 4.3.2, 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 lies on $Z(U_2^i|U_2^{1:i-1}, X_1^{1:N})$. Observe that the union $\mathcal{S}_{X_1} \cup \mathcal{I}_2$ would remove the condition on $Z(U_2^i|U_2^{1:i-1}, X_1^{1:N})$, and accordingly we have $\mathcal{S}_{X_1} \cup \mathcal{I}_2 = \mathcal{I}_2'$. It can be also found that the proportion of

$\mathcal{I}_2$ goes to $I(X_2;Y|X_1)$ as $N \to \infty$. □

We summarize our main results in the following theorem:

**Theorem 4.3.9** (Coding Theorem for Multilevel Polar Codes)**.** *Consider a polar code with the following encoding and decoding strategies for the channel of the second level $W_2$ with the channel transition probability $P_{Y|X_2,X_1}(y|x_2,x_1)$ shown in* (4.16)*.*

- *Encoding: Before sending the codeword $x_2^{1:N} = u_2^{1:N}G_N$, the index set $[N]$ are divided into three parts: the frozen set $\mathcal{F}_2$, information set $\mathcal{I}_2$, and shaping set $\mathcal{S}_2$. The encoder first places uniformly distributed information bits in $\mathcal{I}_2$. Then the frozen set $\mathcal{F}_2$ is filled with a uniform random sequence which are shared between the encoder and the decoder. The bits in $\mathcal{S}_2$ are generated by a mapping $\phi_{\mathcal{S}_2} \triangleq \{\phi_i\}_{i\in\mathcal{S}_2}$ form a family of randomized mappings $\Phi_{\mathcal{S}_2}$, which yields the following distribution:*

$$u_2^i = \begin{cases} 0 & \text{with probability } P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(0|u_2^{1:i-1},x_1^{1:N}), \\ 1 & \text{with probability } P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(1|u_2^{1:i-1},x_1^{1:N}). \end{cases} \tag{4.30}$$

- *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*

$$\hat{u}_2^i = \begin{cases} u_2^i, & \text{if} \quad i \in \mathcal{F}_2 \\ \phi_i(\hat{u}_2^{1:i-1}), & \text{if} \quad i \in \mathcal{S}_2 \\ \underset{u}{\operatorname{argmax}}\, P_{U_2^i|U_2^{1:i-1},X_1^{1:N},Y^{1:N}}(u|\hat{u}_2^{1:i-1},x_1^{1:N},y^{1:N}), & \text{if} \quad i \in \mathcal{I}_2 \end{cases}. \tag{4.31}$$

*Note that probability $P_{U_2^i|U_2^{1:i-1},X_1^{1:N},Y^{1:N}}(u|\hat{u}_2^{1:i-1},x_1^{1:N},y^{1:N})$ can be calculated by* (4.23) *efficiently, treating $Y$ and $X_1$ (already decoded by the SC decoder at level 1)*

*as the outputs of the asymmetric channel. 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* $E_{\Phi_{\mathcal{S}_2}}[P_e(\phi_{\mathcal{S}_2})] \leq N2^{-N^{\beta'}}$ *for any* $\beta' < \beta < 0.5$*. Consequently, there exists a deterministic mapping* $\phi_{\mathcal{S}_2}$ *such that* $P_e(\phi_{\mathcal{S}_2}) \leq N2^{-N^{\beta'}}$.

The proof of this theorem is given in Appendix D.

Obviously, Theorem 4.3.9 can be generalized to the construction of a polar code for the channel of the $\ell$-th level $W_\ell$. 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 a 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.

### 4.3.4 Achieving Channel Capacity

So far, we have constructed polar codes to achieve the capacity of the induced asymmetric channels for all levels. Since the sum capacity of the component channels nearly equals the mutual information $I(Y;X)$, and since we choose a good constellation such that $I(Y;X) \approx \frac{1}{2}\log(1+\mathsf{SNR})$, we have constructed a lattice code to achieve the capacity of the Gaussian channel. We summarize the construction in the following theorem:

**Theorem 4.3.10.** *Choose a good constellation with negligible flatness factor* $\epsilon_\Lambda(\tilde{\sigma})$ *and negligible* $\epsilon_t$ *as in Theorem 3.2.1, and construct a multilevel polar code with* $r = O(\log\log N)$ *as above. Then, for any SNR, the message rate approaches* $\frac{1}{2}\log(1+\mathsf{SNR})$*, while the error probability under multistage decoding is bounded by*

$$P_e \leq rN2^{-N^{\beta'}}, \quad 0 < \beta' < 0.5 \tag{4.32}$$

*as* $N \to \infty$.

*Remark* 4.3.2. It is simple to generate a transmitted codeword of the proposed scheme. For $n = 1$, let

$$\chi = \sum_{\ell=1}^{r} 2^{\ell-1} \left[ \sum_{i \in \mathcal{I}_\ell} u_\ell^i \mathbf{g}_i + \sum_{i \in \mathcal{S}_\ell} u_\ell^i \mathbf{g}_i + \sum_{i \in \mathcal{F}_\ell} u_\ell^i \mathbf{g}_i \right]. \tag{4.33}$$

The transmitted codeword $x$ is drawn from $D_{2^r\mathbb{Z}^N+\chi, \sigma_s}$. From the proof of Lemma 4.3.1, we know that the probability of choosing a point outside of the interval $[-2^{r-1}, 2^{r-1})$ is negligible if $r$ is sufficiently large, which implies there exists only one point in this interval with probability close to 1. Therefore, one may simply transmit $x = \chi \mod 2^r$, where the modulo operation is applied component-wise with range $[-2^{r-1}, 2^{r-1})$.

Next, we show that such a multilevel polar coding scheme is equivalent to Gaussian shaping over a coset $L + c'$ of a polar lattice $L$ for some translate $c'$. In fact, the polar lattice $L$ is exactly constructed from the corresponding symmetrized channels $\tilde{W}_\ell$. Recall that the $\ell$-th channel $W_\ell$ is a BMA channel with the input distribution $P(X_\ell | X_{1:\ell-1})$ $(1 \le \ell \le r)$. It is clear that $P_{X_{1:\ell}}(x_{1:\ell}) = f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))/f_{\sigma_s}(\Lambda)$. By Lemma 4.3.4 and (4.16), the transition probability of the symmetrized channel $\tilde{W}_\ell$ is

$$\begin{aligned}
&P_{\tilde{W}_\ell}((y, x_{1:\ell-1}, x_\ell \oplus \tilde{x}_\ell)|\tilde{x}_\ell) \\
&\quad = P_{Y,X_{1:\ell}}(y, x_{1:\ell}) \\
&\quad = P_{X_{1:\ell}}(x_{1:\ell}) P_{Y|X_\ell, X_{1:\ell-1}}(y|x_\ell, x_{1:\ell-1}) \\
&\quad = \exp\left(-\frac{\|y\|^2}{2(\sigma_s^2+\sigma^2)}\right) \frac{1}{f_{\sigma_s}(\Lambda)} \frac{1}{2\pi\sigma\sigma_s} \sum_{a \in \mathcal{A}_\ell(x_{1:\ell})} \exp\left(-\frac{\|\alpha y - a\|^2}{2\tilde{\sigma}^2}\right).
\end{aligned} \tag{4.34}$$

Note that the difference between the asymmetric channel (4.16) and symmetrized channel (4.34) is the *a priori* probability $P_{X_{1:\ell}}(x_{1:\ell}) = f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))/f_{\sigma_s}(\Lambda)$. Comparing with the $\Lambda_{\ell-1}/\Lambda_\ell$ channel (5.5), we see that the symmetrized channel (4.34) is equivalent to a $\Lambda_{\ell-1}/\Lambda_\ell$ channel, since the common terms in front of the sum will be

completely cancelled out in the calculation of the likelihood ratio[3]. We summarize the foregoing analysis in the following lemma:

**Lemma 4.3.11** (Equivalence lemma)**.** *Consider a multilevel lattice code constructed from constellation $D_{\Lambda,\sigma_s}$ for a Gaussian channel with noise variance $\sigma^2$. The $\ell$-th symmetric channel $\tilde{W}_\ell$ ($1 \le \ell \le r$) which is derived from the asymmetric channel $W_\ell$ is equivalent to the MMSE-scaled $\Lambda_{\ell-1}/\Lambda_\ell$ channel with noise variance $\tilde{\sigma}^2$.*

Thus, the resultant polar codes for the symmetrized channels are nested, and the polar lattice is AWGN-good for noise variance $\tilde{\sigma}^2$; also, the multistage decoding is performed on the MMSE-scaled signal $\alpha y$ (cf. Lemma 4.3.6). Since the frozen sets of the polar codes are filled with random bits (rather than all zeros), we actually obtain a coset $L + c'$ of the polar lattice, where the shift $c'$ accounts for the effects of all random frozen bits. Finally, since we start from $D_{\Lambda,\sigma_s}$, we would obtain $D_{\Lambda^N,\sigma_s}$ without coding; since $L + c' \subset \Lambda^N$ by construction, we obtain a discrete Gaussian distribution $D_{L+c',\sigma_s}$ over $L + c'$.

*Remark* 4.3.3. This analysis shows that our proposed scheme is an explicit construction of lattice Gaussian coding introduced in [20], which applies Gaussian shaping to an AWGN-good lattice (or its coset). Note that the condition of negligible $\epsilon_\Lambda(\tilde{\sigma})$ in Theorem 4.3.10 is the same as the condition on $\Lambda$ imposed in the construction of polar lattice in Section 4.2 (cf. Theorem 4.2.4). Again, it is always possible to scale down the top lattice $\Lambda$ such that both $\epsilon_\Lambda(\tilde{\sigma})$ and $\epsilon_t$ become negligible in Theorem 3.2.1. Thus, Theorem 4.3.10 holds for any SNR, meaning that we have removed the condition $\mathsf{SNR} > e$ required by [20, Theorem 3][4]. Moreover, if a good constellation of the form $D_{\Lambda-c,\sigma_s}$ for some shift $c$ is used in practice (e.g., a constellation taking values in $\{\pm 1, \pm 3, \ldots\}$), the proposed construction holds verbatim.

[3] Even if $y \in \mathbb{R}^n$ in (4.34), the sum over $\mathcal{A}_\ell(x_{1:\ell})$ is $\Lambda_\ell$-periodic. Hence, the likelihood ratio will be the same if one takes $\bar{y} = y \mod \Lambda_\ell$ and uses (5.5).

[4] The reason of the condition $\mathsf{SNR} > e$ in [20] is that a more stringent condition is imposed on the flatness factor of $L$, namely, $\epsilon_L\left(\frac{\sigma_s^2}{\sqrt{\sigma_s^2+\sigma^2}}\right)$ is negligible.

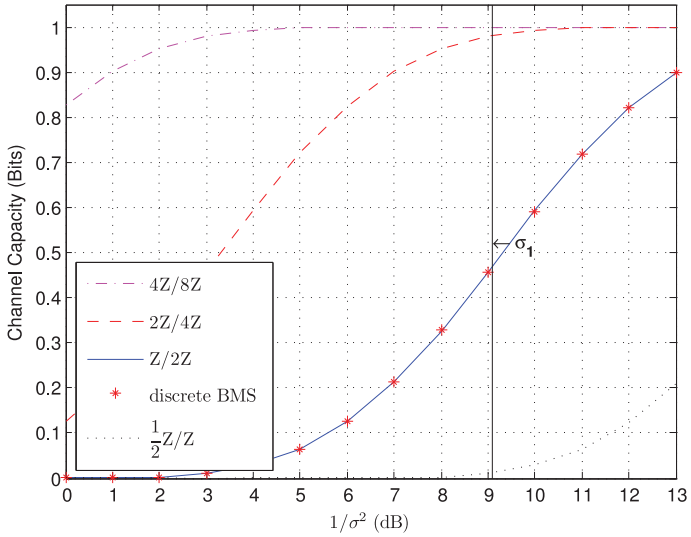



Figure 4.6: Channel capacity for partition chain $\mathbb{Z}/2\mathbb{Z}/\cdots/2^r\mathbb{Z}$. The discrete approximation used in Sect. 4.2.2 is also shown.

*Remark* 4.3.4. By [20, Lemma 1], the power $P$ of a discrete Gaussian distribution $D_{L+c',\sigma_s}$ is never greater than $\sigma_s^2$.

*Remark* 4.3.5. A shaping method was proposed in [20, Section IV], where Gaussian shaping is only performed on the bottom lattice. However, it requires negligible $\epsilon_{\Lambda'}(\sigma_s)$, which does not hold in general.

## 4.4 Design Examples

In this section, we give design examples of polar lattices based on one and two-dimensional partition chains, with and without the power constraint. The design follows the equal-error-probability rule. Multistage SC decoding is applied. Since the complexity of SC decoding is $O(N \log N)$, the overall decoding complexity is $O(rN \log N)$.

### 4.4.1 Design Examples without Power Constraint

Consider the one-dimensional lattice partition $\mathbb{Z}/2\mathbb{Z}/\cdots/2^r\mathbb{Z}$. To construct a multi-level lattice, one needs to determine the number of levels of lattice partitions and the

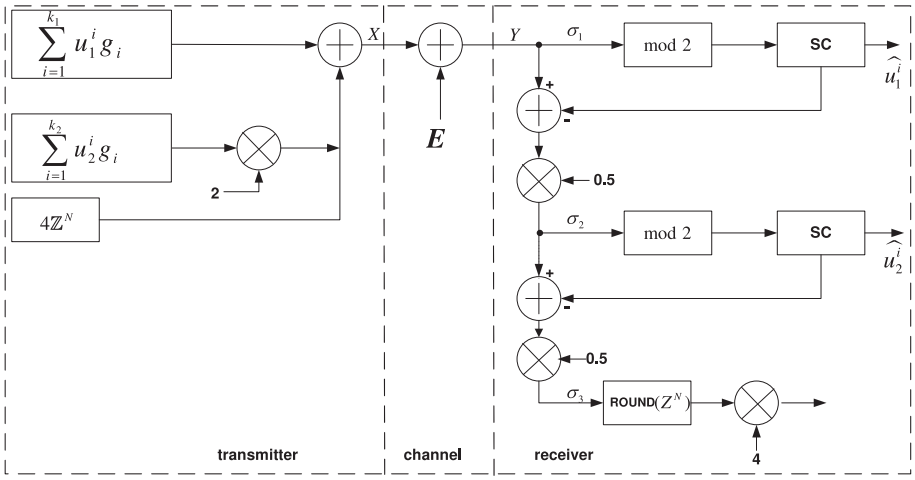



Figure 4.7: A polar lattice with two levels, where $\sigma_1 = \tilde{\sigma}$.

actual rates according to the target error probability for a given noise variance. By the guidelines of "Construction D" in Chapter 3, the effective levels are 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 Fig. 4.6. For example, at the given noise variance indicated by the straight line in Fig. 4.6, one may choose two levels of component codes, which was indeed suggested in [50].

The multilevel construction and the multistage decoding are shown in Fig. 4.7. For the $\ell$-th level, $\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_{k_\ell}$ are a set of code generators chosen from the matrix $G_N$, and $\sigma_\ell$ is the standard deviation of the noise.

Now, we give an example for length $N = 1024$ and target error probability $P_e(L, \tilde{\sigma}^2) = 10^{-5}$. 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 Fig. 4.6, 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, $\tilde{\sigma} = \sigma_1 = 0.3380$ and the capacity is $0.5145$. Our goal is to find two polar codes approaching the respective capacities at block error probabilities $\leq \frac{1}{3} \cdot 10^{-5}$ over these binary-input mod-2 channels.

For $N = 1024$, we found the first polar code with $\frac{k_1}{N} = 0.23$ for $P_e(\mathcal{C}_1, \sigma_1^2) \approx$

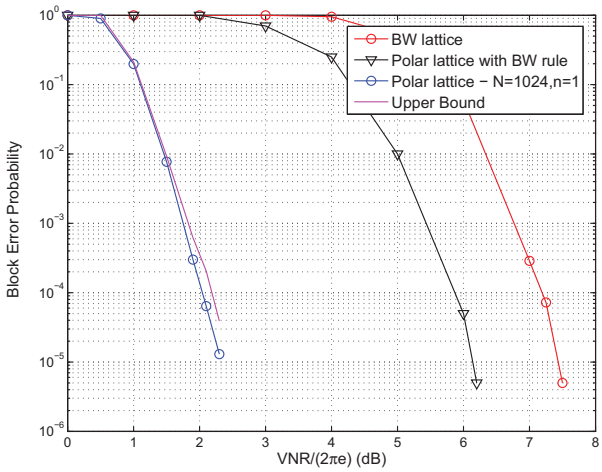



Figure 4.8: Block error probabilities of polar lattices and Barnes-Wall (BW) lattices of length $N = 1024$ with multistage decoding.

$\frac{1}{3} \cdot 10^{-5}$, 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 (7.29), the logarithmic VNR is given by

$$\log\left(\frac{\gamma_L(\tilde{\sigma})}{2\pi e}\right) \leq 2\left(\epsilon_1 + \epsilon_3\right) = 0.7758, \tag{4.35}$$

which is 2.34 dB. Fig. 4.8 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 [84], the upper bound $\sum_{i \in \mathcal{A}}(\tilde{Z}(\tilde{W}_N^{(i)}))$ on the block error probability of a polar code with finite-length can be calculated numerically. According to (7.23), we plot the upper bound on the block error probability $P_e(L, \tilde{\sigma}^2)$ of the polar lattice in Fig. 4.8, which is quite tight.

Now we revisit the Barnes-Wall lattice with its performance shown in Fig. 4.8. We know that there are only 2 effective levels, but the Barnes-Wall lattice (3.41)

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. 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 Fig. 4.8 is our prior design [85], 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 multistage 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.

The performance comparison of other competing lattices approaching the Poltyrev capacity with dimension around 1000 is shown in Fig. 4.9 in terms of the symbol error rate (SER)[5]. The polar lattice used here is constructed from the two-dimensional lattice partition $\mathbb{Z}^2/R\mathbb{Z}^2/2\mathbb{Z}^2/2R\mathbb{Z}^2/4\mathbb{Z}^2$, where $R$ denotes the rotation matrix $\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. The simulation curves of other lattices are obtained from their corresponding papers. We note that the theoretical minimum gap to the Poltyrev capacity is about $1$ dB for dimension $1000$ [83]. Among the four types of lattices compared, the LDPC lattice [59] 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 [56, 76], analytic results of the LDLC [67] are not available; therefore, they are less understood in theory. The LDA lattice has slightly better performance than the polar lattice at the expense of higher decoding complexity ($O(p^2 N \log N)$) if $p$-ary LDPC codes are employed. Assuming $p \approx 2^r$,

[5] SER is defined as the average error probability of the coordinates of the lattice codeword $\lambda$, which is commonly used in literature. The curve for the LDPC lattice was plotted with the normalized block error probability [59].

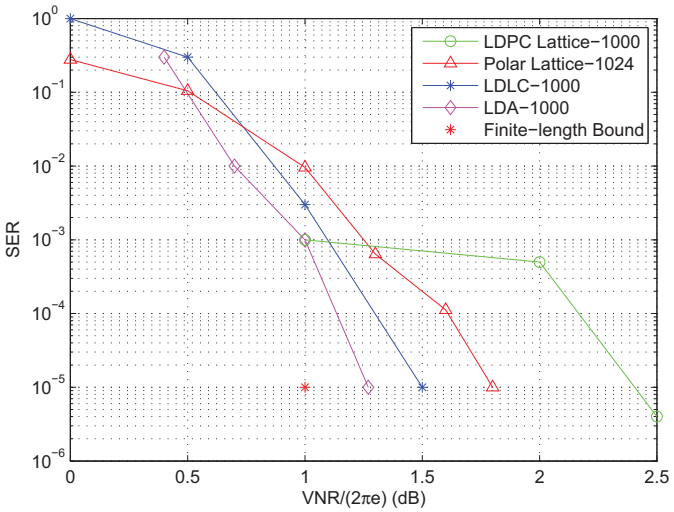



Figure 4.9: SER of lattices with dimension around 1000.

it would rquire complexity $O(2^{2r} N \log N)$.

### 4.4.2 Design Examples with Power Constraint

To satisfy the power constraint, we use discrete lattice distribution $D_{\mathbb{Z},\sigma_s}$ for shaping. The mutual information $I(Y; X_\ell | X_{1:\ell-1})$ at each level for different SNRs is shown in Fig. 4.10. We can see that for partition $\mathbb{Z}/2\mathbb{Z}/...$, five levels are enough to achieve the AWGN channel capacity for SNR ranging from $-5$ dB to $20$ dB. Note that this is more than the number of levels required in the design of the AWGN-good lattice itself.

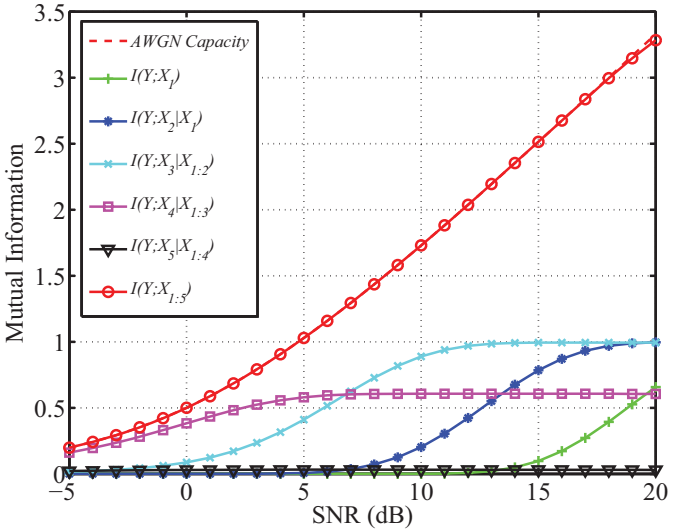



Figure 4.10: Channel capacity for each level as a function of SNR.

For each level, we estimate a lower bound on the code rate for block error probability $P_{B_\ell} = 1 \times 10^{-5}$. The is done by calculating an upper bound on the block error probability of the polar code, using the Bhattacharyya parameter. With this target error probability, the assignments of bits to the information, shaping and frozen sets on different levels are shown in Fig. 4.11 for $\mathsf{SNR} = 15$ dB and $N = 2^{16}$. In fact, $X_1$ and $X_2$ are nearly uniform such that there is no need for shaping on the first two levels (these levels actually correspond to the AWGN-good lattice). For the third level, most bits are information bits. In contrast, the fifth level is mostly for shaping; since its message rate is already small, adding another level clearly would not contribute to the overall rate of the lattice code. Finally, lower bounds on the rates achieved by polar lattices with various block lengths are shown in Fig. 4.12. We note that the gap to the channel capacity diminishes as $N$ increases, and it is only about $0.1$ bits/dimension when $N = 2^{20}$.

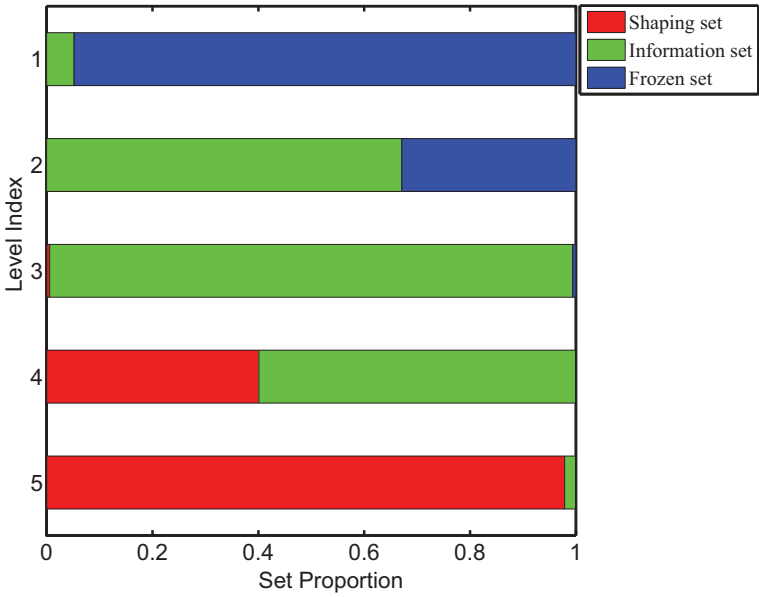



Figure 4.11: The proportions of the shaping set, information set, and frozen set on each level when $N = 2^{16}$ and SNR $= 15$ dB.

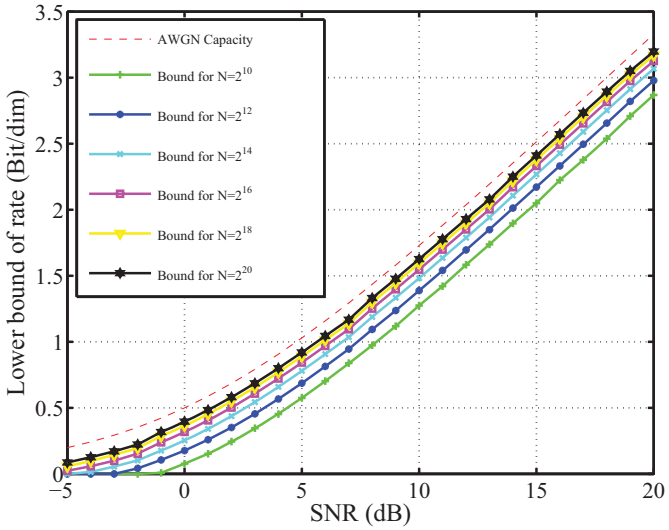



Figure 4.12: Lower bounds on the rates achieved by polar lattices with block error probability $5 \times 10^{-5}$ for block lengths $2^{10}$, $2^{12}$, ..., $2^{20}$.

## 4.5 Summary

In this chapter, we have constructed polar lattices to approach the capacity of the power-constrained Gaussian channel. The construction is based on a combination of channel polarization and source polarization. Without shaping, the constructed polar lattices are AWGN-good. The Gaussian shaping on a polar lattice deals with the power constraint but is technically more involved. The overall scheme is explicit and efficient, featuring quasi-linear complexity.

CHAPTER 5

# Polar Lattices for Gaussian Wiretap Channels

## 5.1 Introduction

In this chapter, an explicit wiretap coding scheme based on polar lattices is proposed to achieve the secrecy capacity of the additive white Gaussian noise (AWGN) wiretap channel. Firstly, polar lattices are used to construct secrecy-good lattices for the mod-$\Lambda_s$ Gaussian wiretap channel. Then we propose an explicit shaping scheme to remove this mod-$\Lambda_s$ front end and extend polar lattices to the genuine Gaussian wiretap channel. The shaping technique is based on the lattice Gaussian distribution, which leads to a binary asymmetric channel at each level for the multilevel lattice codes. By employing the asymmetric polar coding technique, we construct an AWGN-good lattice and a secrecy-good lattice with optimal shaping simultaneously. As a result, the encoding complexity for the sender and the decoding complexity for the legitimate receiver are both $O(N \log N \log(\log N))$. The proposed scheme is proven to be semantically secure.

The wiretap channel model was introduced by Wyner [41], who showed that both reliability and confidentiality could be attained by coding without any key bits if the channel between the sender and the eavesdropper (wiretapper's channel $W$) is degraded with respect to the channel between the sender and the legitimate receiver (main channel $V$). The goal of wiretap coding is to design a coding scheme that

makes it possible to communicate both reliably and securely between the sender and the legitimate receiver. Reliability is measured by the decoding error probability for the legitimate user, namely $\lim_{N\to\infty} \Pr\{\hat{M} \neq M\} = 0$, where $N$ is the length of transmitted codeword, $M$ is the confidential message and $\hat{M}$ is its estimation. Secrecy is measured by the mutual information between $M$ and the signal received by the eavesdropper $Z^{[N]}$. In this work, we will follow the strong secrecy condition proposed by Csiszár [42], i.e., $\lim_{N\to\infty} I(M; Z^{[N]}) = 0$, which is more widely accepted than the weak secrecy criterion $\lim_{N\to\infty} \frac{1}{N} I(M; Z^{[N]}) = 0$. In simple terms, the secrecy capacity is defined as the maximum achievable rate under both the reliability and strong secrecy conditions. When $W$ and $V$ are both symmetric, and $W$ is degraded with respect to $V$, the secrecy capacity is given by $C(V) - C(W)$ [86], where $C(\cdot)$ denotes the channel capacity.

In the study of strong secrecy, plaintext messages are often assumed to be random and uniformly distributed. From a cryptographic point of view, it is crucial that the security does not rely on the distribution of the message. This issue can be resolved by using the standard notion of *semantic security* [87] which means that, asymptotically, it is impossible to estimate any function of the message better than to guess it without accessing $Z^{[N]}$ at all. The relation between strong secrecy and semantic security was recently revealed in [88, 21], namely, semantic security is equivalent to achieving strong secrecy for all distributions $p_M$ of the plaintext messages:

$$\lim_{N\to\infty} \max_{p_M} I(M; Z^{[N]}) = 0. \tag{5.1}$$

In this chapter, we construct lattice codes for the Gaussian wiretap channel (G-WC) which is shown in Fig. 5.1. The confidential message $M$ drawn from the message set $\mathcal{M}$ is encoded by the sender (Alice) into an $N$-dimensional codeword $X^{[N]}$. The outputs $Y^{[N]}$ and $Z^{[N]}$ received by the legitimate receiver (Bob) and the

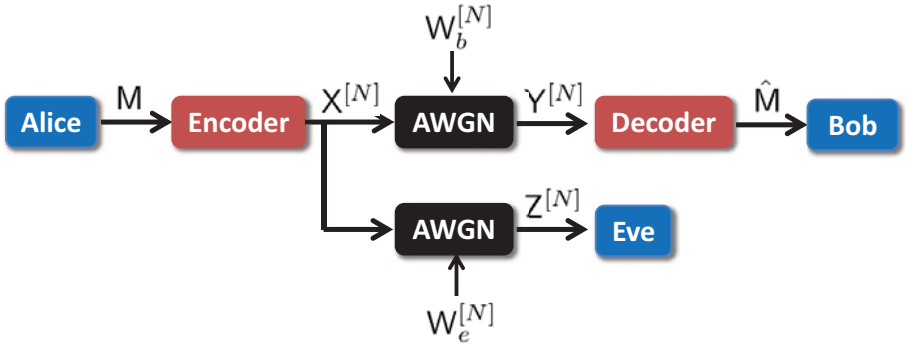



Figure 5.1: The Gaussian wiretap channel.

eavesdropper (Eve) are respectively given by

$$\begin{cases} Y^{[N]} = X^{[N]} + W_b^{[N]} \\ Z^{[N]} = X^{[N]} + W_e^{[N]}, \end{cases}$$

where $W_b^{[N]}$ and $W_e^{[N]}$ are $N$-dimensional Gaussian noise vectors with zero mean and variance $\sigma_b^2$, $\sigma_e^2$ respectively. The channel input $X^{[N]}$ satisfies the power constraint $P_s$, i.e.,

$$\frac{1}{N} E\left[\|X^{[N]}\|^2\right] \leq P_s.$$

Polar codes [4] have shown their great potential in solving the wiretap coding problem. The polar coding scheme proposed in [13], combined with the block Markov coding technique [44], was proved to achieve the strong secrecy capacity when $W$ and $V$ are both binary-input symmetric channels, and $W$ is degraded with respect to $V$. More recently, polar wiretap coding has been extended to general wiretap channels (not necessarily degraded or symmetric) in [89] and [90]. For continuous channels such as the GWC, there also has been notable progress in wiretap lattice coding. On the theoretical aspect, the existence of lattice codes achieving the secrecy capacity to within $\frac{1}{2}$ nat under the strong secrecy as well as semantic security criterion was demonstrated in [21]. On the practical aspect, wiretap lattice

codes were proposed in [91] and [92] to maximize the eavesdropper's decoding error probability.

Polar lattices, the counterpart of polar codes in the Euclidean space, have already been proved to be additive white Gaussian noise (AWGN)-good [51] and further to achieve the AWGN channel capacity with lattice Gaussian shaping in Chapter 4 [1]. Motivated by [13], we will propose polar lattices to achieve both strong secrecy and reliability over the mod-$\Lambda_s$ GWC. Conceptually, this polar lattice structure can be regarded as a secrecy-good lattice $\Lambda_e$ nested within an AWGN-good lattice $\Lambda_b$ ($\Lambda_e \subset \Lambda_b$). Further, we will propose a Gaussian shaping scheme over $\Lambda_b$ and $\Lambda_e$, using the multilevel asymmetric polar coding technique. As a result, we will accomplish the design of an explicit lattice coding scheme which achieves the secrecy capacity of the GWC. The novel technical ingredients of this work are the following:

- The construction of secrecy-good polar lattices for the mod-$\Lambda_s$ GWC and the proof of their secrecy capacity-achieving. This is an extension of the binary symmetric wiretap coding [13] to the multilevel coding scenario, and can also be considered as the construction of secrecy-good polar lattices for the GWC without the power constraint. The construction for the mod-$\Lambda_s$ GWC provides considerable insight into wiretap coding for the genuine GWC, without deviating to the technicality of Gaussian shaping. This work is also of independent interest to other problems of information theoretic security, e.g., secret key generation from Gaussian sources [95].

- The Gaussian shaping applied to the secrecy-good polar lattice, which follows the footpath of Section. 4.3. The resultant coding scheme is proved to achieve the secrecy capacity of the GWC. This coding scheme is further proved to be semantically secure. The idea follows the conception of [21], where lattice Gaussian sampling was employed to obtain semantic security. It is worth

[1] Please refer to [1, 93, 94] for other methods of achieving the AWGN channel capacity.

mentioning that our proposed coding scheme is not only a practical implementation of the secure random lattice coding in [21], but also an improvement in the sense that we successfully remove the constant $\frac{1}{2}$-nat gap to the secrecy capacity. [2]

Invertible randomness extractors were introduced into wiretap coding in [96, 97, 88]. The key idea is that an extractor is used to convert a capacity-achieving code with rate close to $C(V)$ for the main channel into a wiretap code with the rate close to $C(V) - C(W)$. Later, this coding scheme was extended to the GWC in [98]. Besides, channel resolvability [99] was proposed as a tool for wiretap codes. An interesting connection between the resolvability and the extractor was revealed in [100].

The proposed approach and the one based on invertible extractors have their respective advantages. The extractor-based approach is modular, i.e., the error-correction code and extractor are realized separately; it is possible to harness the results of invertible extractors in literature. The advantage of our lattice-based scheme is that the wiretap code designed for Eve is nested within the capacity-achieving code designed for Bob, which represents an integrated approach. More importantly, lattice codes are attractive for emerging applications in network information theory thanks to their useful structures [1], [101]; thus the proposed scheme may fit better with this landscape when security is a concern [102].

## 5.2 Secrecy-good Polar Lattices for the Mod-$\Lambda_s$ GWC

Recall that a sublattice $\Lambda' \subset \Lambda$ induces a partition (denoted by $\Lambda/\Lambda'$) of $\Lambda$ into equivalence classes modulo $\Lambda'$. The order of the partition is denoted by $|\Lambda/\Lambda'|$,

[2] The $\frac{1}{2}$-nat gap in [21] was due to a requirement on the volume-to-noise ratio of the secrecy-good lattice. When polar codes are used to construct the secrecy-good lattice, however, this requirement can be removed because the upper-bound of information leakage can be directly calculated.

which is equal to the number of cosets. If $|\Lambda/\Lambda'| = 2$, we call this a binary partition. Let $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'$ for $r \geq 1$ be an $n$-dimensional lattice partition chain. For each partition $\Lambda_{\ell-1}/\Lambda_\ell$ ($1 \leq \ell \leq r$ with convention $\Lambda_0 = \Lambda$ and $\Lambda_r = \Lambda'$) a code $C_\ell$ over $\Lambda_{\ell-1}/\Lambda_\ell$ selects a sequence of representatives $a_\ell$ for the cosets of $\Lambda_\ell$. Consequently, if each partition is binary, the code $C_\ell$ is a binary code.

From the previous chapter, we know that polar lattices are constructed by "Construction D" using a set of nested polar codes $C_1 \subseteq C_2 \cdot\cdot \subseteq C_r$. Suppose $C_\ell$ has block length $N$ and the number of information bits $k_\ell$ for $1 \leq \ell \leq r$. Choose a basis $\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N$ from the polar generator matrix $G_N$ such that $\mathbf{g}_1, \cdots \mathbf{g}_{k_\ell}$ span $C_\ell$. Recall that when the dimension $n = 1$, the lattice $L$ admits the form [50]

$$L = \left\{ \sum_{\ell=1}^{r} 2^{\ell-1} \sum_{i=1}^{k_\ell} u_\ell^i \mathbf{g}_i + 2^r \mathbb{Z}^N \mid u_\ell^i \in \{0,1\} \right\}, \tag{5.2}$$

where the addition is carried out in $\mathbb{R}^N$. The fundamental volume of a lattice obtained from this construction is given by

$$V(L) = 2^{-NR_C} \cdot V(\Lambda_r)^N,$$

where $R_C = \sum_{\ell=1}^{r} R_\ell = \frac{1}{N}\sum_{\ell=1}^{r} k_\ell$ denotes the sum rate of component codes. In this paper, we limit ourselves to the binary lattice partition chain and binary polar codes for simplicity.

### 5.2.1 Secrecy-goodness

Now we consider the construction of secrecy-good polar lattices over the mod-$\Lambda_s$ GWC shown in Fig. 5.2. The difference between the mod-$\Lambda_s$ GWC and the genuine GWC is the mod-$\Lambda_s$ operation on the received signal of Bob and Eve. With some abuse of notation, the outputs $Y^{[N]}$ and $Z^{[N]}$ at Bob and Eve's ends respectively

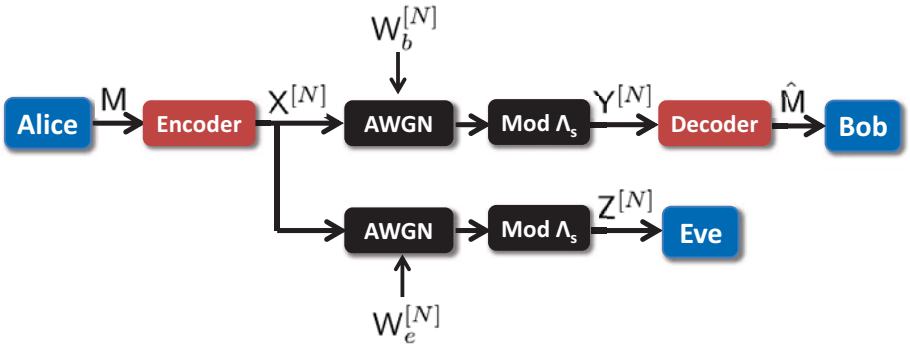



Figure 5.2: The mod-$\Lambda_s$ Gaussian wiretap channel.

become

$$\begin{cases} Y^{[N]} = \left[X^{[N]} + W_b^{[N]}\right] & \mod \Lambda_s, \\ Z^{[N]} = \left[X^{[N]} + W_e^{[N]}\right] & \mod \Lambda_s. \end{cases} \tag{5.3}$$

The idea of wiretap lattice coding over the mod-$\Lambda_s$ GWC [21] can be explained as follows. Let $\Lambda_b$ and $\Lambda_e$ be the AWGN-good lattice and secrecy-good lattice designed for Bob and Eve accordingly. Let $\Lambda_s \subset \Lambda_e \subset \Lambda_b$ be a nested chain of $N$-dimensional lattices in $\mathbb{R}^N$, where $\Lambda_s$ is the shaping lattice. Note that the shaping lattice $\Lambda_s$ here is employed primarily for the convenience of designing the secrecy-good lattice and secondarily for satisfying the power constraint. Consider a one-to-one mapping: $\mathcal{M} \to \Lambda_b/\Lambda_e$ which associates each message $m \in \mathcal{M}$ to a coset $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$. Alice selects a lattice point $\lambda \in \Lambda_e \cap \mathcal{V}(\Lambda_s)$ uniformly at random and transmits $X^{[N]} = \lambda + \lambda_m$, where $\lambda_m$ is the coset representative of $\tilde{\lambda}_m$ in $\mathcal{V}(\Lambda_e)$. This scheme has been proved to achieve both reliability and semantic security in [21] by random lattice codes. We will make it explicit by constructing polar lattice codes in this section.

Let $\Lambda_b$ and $\Lambda_e$ be constructed from a binary partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda_r$, and assume $\Lambda_s \subset \Lambda_r^N$ such that $\Lambda_s \subset \Lambda_r^N \subset \Lambda_e \subset \Lambda_b$[3]. Also, denote by $X_{1:r}^{[N]}$ the bits

[3]This is always possible with sufficient power, since the power constraint is not our primary concern in this section.

encoding $\Lambda^N/\Lambda_r^N$, which include all information bits for message $M$ as a subset. We have that $\left[X^{[N]} + W_e^{[N]}\right] \bmod \Lambda_r^N$ is a sufficient statistic for $X_{1:r}^{[N]}$. This can be seen from [50, Lemma 8], rewritten as follows:

**Lemma 5.2.1** (Sufficiency of mod-$\Lambda$ output [50])**.** *For a partition chain $\Lambda/\Lambda'$ ($\Lambda' \subset \Lambda$), let the input of an AWGN channel be $X = A + B$, where $A \in \mathcal{R}(\Lambda)$ is a random variable, and $B$ is uniformly distributed in $\Lambda \cap \mathcal{R}(\Lambda')$. Reduce the output $Y$ first to $Y' = Y \mod \Lambda'$ and then to $Y'' = Y' \mod \Lambda$. Then the mod-$\Lambda$ map is information-lossless, namely $I(A;Y') = I(A;Y'')$, which means that the output $Y'' = Y' \mod \Lambda$ of mod-$\Lambda$ map is a sufficient statistic for $A$.*

In our context, we identify $\Lambda$ with $\Lambda_r^N$ and $\Lambda'$ with $\Lambda_s$, respectively. Since the bits encoding $\Lambda_r^N/\Lambda_s$ are uniformly distributed[4], the mod-$\Lambda_r^N$ operation is information-lossless in the sense that

$$I\left(X_{1:r}^{[N]}; Z^{[N]}\right) = I\left(X_{1:r}^{[N]}; [X^{[N]} + W_e^{[N]}] \bmod \Lambda_r^N\right). \tag{5.4}$$

As far as mutual information $I\left(X_{1:r}^{[N]}; Z^{[N]}\right)$ is concerned, we can use the mod-$\Lambda_r^N$ operator instead of the mod-$\Lambda_s$ operator here. Under this condition, similarly to the multilevel lattice structure introduced in [50], the mod-$\Lambda_s$ channel can be decomposed into a series of BMS channels according to the partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda_r$. Therefore, the already mentioned polar coding technique for BMS channels can be employed. Moreover, the channel resulted from the lattice partition chain can be proved to be equivalent to that based on the chain rule of mutual information. Following this channel equivalence, we can construct an AWGN-good lattice $\Lambda_b$ and a secrecy-good lattice $\Lambda_e$, using the wiretap coding technique (2.25) at each partition level.

The decomposition into a set of $\Lambda_{\ell-1}/\Lambda_\ell$ channels is used in [50] to construct

[4] In fact, all bits encoding $\Lambda_e/\Lambda_s$ are uniformly distributed in wiretap coding.

AWGN-good lattices. Take the partition chain $\mathbb{Z}/2\mathbb{Z}/\cdots/2^r\mathbb{Z}$ as an example. Given uniform input $X_{1:r}$, let $\mathcal{K}_\ell$ denote the coset indexed by $x_{1:\ell}$, i.e., $\mathcal{K}_\ell = x_1 + \cdots + 2^{\ell-1}x_\ell + 2^\ell\mathbb{Z}$. Given that $X_{1:\ell-1} = x_{1:\ell-1}$, the conditional probability distribution function (PDF) of this channel with binary input $X_\ell$ and output $\bar{Z} = Z \mod \Lambda_\ell$ is

$$f_{\bar{Z}|X_\ell}(\bar{z}|x_\ell) = \frac{1}{\sqrt{2\pi}\sigma_e} \sum_{a\in\mathcal{K}_\ell(x_{1:\ell})} \exp\left(-\frac{1}{2\sigma_e^2}\|\bar{z}-a\|^2\right). \tag{5.5}$$

Since the previous input bits $x_{1:\ell-1}$ cause a shift on $\mathcal{K}_\ell$ and will be removed by the multistage decoder at level $\ell$, the code can be designed according to the channel transition probability (5.5) with $x_{1:\ell-1} = 0$. Following the notation of [50], we use $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ and $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ to denote the $\Lambda_{\ell-1}/\Lambda_\ell$ channel for Bob and Eve respectively. The $\Lambda_{\ell-1}/\Lambda_\ell$ channel can also be used to construct secrecy-good lattices. In order to bound the information leakage of the wiretapper's channel, we firstly express $I(X_{1:r}; Z)$ according to the chain rule of mutual information as

$$I(X_{1:r}; Z) = I(X_1; Z) + I(X_2; Z|X_1) + \cdots + I(X_r; Z|X_{1:r-1}). \tag{5.6}$$

This equation still holds if $Z$ denotes the noisy signal after the mod-$\Lambda_r$ operation, namely, $Z = [X + W_e] \mod \Lambda_r$. We will adopt this notation in the rest of this subsection. We refer to the $\ell$-th channel associated with mutual information $I(X_\ell; Z|X_{1:\ell-1})$ as the equivalent channel denoted by $W'(X_\ell; Z|X_{1:\ell-1})$, which is defined as the channel from $X_\ell$ to $Z$ given the previous $X_{1:\ell-1}$. Then the transition probability distribution of $W'(X_\ell; Z|X_{1:\ell-1})$ is [50, Lemma 6]

$$\begin{aligned} f_{Z|X_\ell}(z|x_\ell) &= \frac{1}{\Pr(\mathcal{K}_\ell(x_{1:\ell}))} \sum_{a\in\mathcal{K}_\ell(x_{1:\ell})} \Pr(a) f_Z(z|a) \\ &= \frac{1}{|\Lambda_\ell/\Lambda_r|} \frac{1}{\sqrt{2\pi}\sigma_e} \sum_{a\in\mathcal{K}_\ell(x_{1:\ell})} \exp\left(-\frac{1}{2\sigma_e^2}\|z-a\|^2\right), \quad z \in \mathcal{V}(\Lambda_r). \end{aligned} \tag{5.7}$$

From (5.5) and (5.7), we can observe that the channel output likelihood ratio (LR) of the $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ channel is equal to that of the $\ell$-th equivalent channel $W'(X_\ell; Z|X_{1:\ell-1})$. Then we have the following channel equivalence lemma.

**Lemma 5.2.2.** *Consider a lattice $L$ constructed by a binary lattice partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda_r$. Constructing a polar code for the $\ell$-th equivalent binary-input channel $W'(X_\ell; Z|X_{1:\ell-1})$ defined by the chain rule (5.6) is equivalent to constructing a polar code for the $\Lambda_{\ell-1}/\Lambda_\ell$ channel $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$.*

*Proof.* See Appendix E. □

Note that another proof based on direct calculation of the mutual information and Bhattacharyya parameters of the subchannels can be found in [103].

*Remark* 5.2.1. Observe that if we define $V'(X_\ell; Y|X_{1:\ell-1})$ as the equivalent channel according to the chain rule expansion of $I(X;Y)$ for the main channel, the same result can be obtained between $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ and $V'(X_\ell; Y|X_{1:\ell-1})$. Moreover, this lemma also holds without the mod-$\Lambda_s$ front-end, i.e., without power constraint. The construction of AWGN-good polar lattices was given in [104], where nested polar codes were constructed based on a set of $\Lambda_{\ell-1}/\Lambda_\ell$ channels. We note that the $\Lambda_{\ell-1}/\Lambda_\ell$ channel is degraded with respect to the $\Lambda_\ell/\Lambda_{\ell+1}$ channel [104, Lemma 3].

Now it is ready to introduce the polar lattice construction for the mod-$\Lambda_s$ GWC shown in Fig. 5.3. A polar lattice $L$ is constructed by a series of nested polar codes $C_1(N, k_1) \subseteq C_2(N, k_2) \subseteq \cdots \subseteq C_r(N, k_r)$ and a binary lattice partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_r$. The block length of polar codes is $N$. Alice splits the message $M$ into $M_1, \cdots, M_r$. We follow the same rule (2.29) to assign bits in the component polar codes to achieve strong secrecy. Note that $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ is degraded with respect to $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ for $1 \le \ell \le r$ because $\sigma_b^2 \le \sigma_e^2$. Treating $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ and $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ as the main channel and wiretapper's channel at each level and using the partition rule (2.28), we can get four sets $\mathcal{A}_\ell$, $\mathcal{B}_\ell$, $\mathcal{C}_\ell$ and $\mathcal{D}_\ell$. Similarly, we

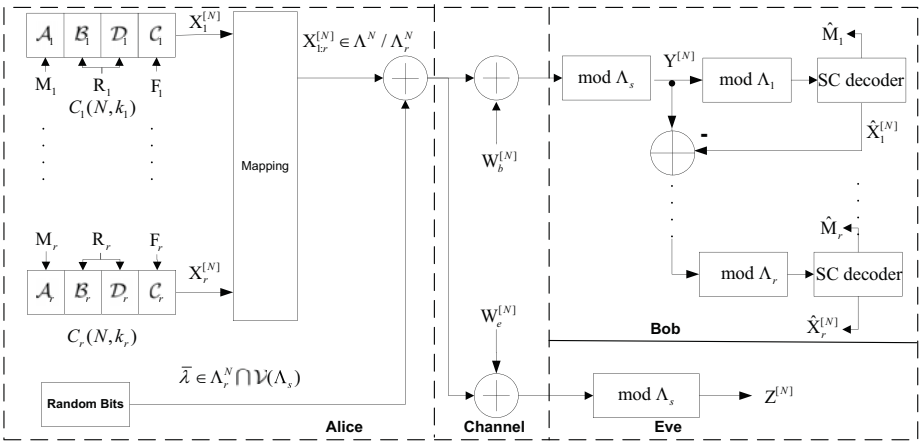



Figure 5.3: The multilevel lattice coding system over the mod-$\Lambda_s$ Gaussian wiretap channel.

assign the bits as follows

$$\begin{aligned} \mathcal{A}_\ell &\leftarrow M_\ell, \quad \mathcal{B}_\ell \leftarrow R_\ell, \\ \mathcal{C}_\ell &\leftarrow F_\ell, \quad \mathcal{D}_\ell \leftarrow R_\ell \end{aligned} \tag{5.8}$$

for each level $\ell$, where $M_\ell$, $F_\ell$ and $R_\ell$ represent message bits, frozen bits (could be set as all zeros) and random bits at level $\ell$. Since the $\Lambda_{\ell-1}/\Lambda_\ell$ channel is degraded with respect to the $\Lambda_\ell/\Lambda_{\ell+1}$ channel, 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. We skip the proof of nested polar codes here. A similar proof can be found in [51].

As a result, the above multilevel construction yields an AWGN-good lattice $\Lambda_b$ and a secrecy-good lattice $\Lambda_e$ simultaneously[5]. More precisely, $\Lambda_b$ is constructed from a set of nested polar codes $C_1(N, |\mathcal{A}_1|+|\mathcal{B}_1|+|\mathcal{D}_1|) \subseteq \cdots \subseteq C_r(N, |\mathcal{A}_r|+|\mathcal{B}_r|+|\mathcal{D}_r|)$, while $\Lambda_e$ is constructed from a set of nested polar codes $C_1(N, |\mathcal{B}_1| + |\mathcal{D}_1|) \subseteq \cdots \subseteq C_r(N, |\mathcal{B}_r| + |\mathcal{D}_r|)$ and with the same lattice partition chain. Note that the random bits in set $\mathcal{D}_\ell$ should be shared to Bob to guarantee the AWGN-goodness of

[5]In this paper, a sequence of lattices $\Lambda_e$ of increasing dimension is called secrecy-good if they achieve the strong secrecy capacity asymptotically. Note that this definition is different from that in [21], which is based on the flatness factor.

$\Lambda_b$. More details are given in the next subsection. It is clear that $\Lambda_e \subset \Lambda_b$. Thus, our proposed coding scheme instantiates the coset coding scheme introduced in [21], where the confidential message is mapped to the coset $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$.

By using the above assignments and Lemma 2.4.2, we have

$$I\left(M_\ell F_\ell; Z_\ell^{[N]}\right) \leq N2^{-N^{\beta'}}, \tag{5.9}$$

where $Z_\ell^{[N]} = Z^{[N]} \bmod \Lambda_\ell$ is the output of the $\Lambda_{\ell-1}/\Lambda_\ell$ channel for Eve. In other words, the employed polar code for the channel $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ can guarantee that the mutual information between the input message and the output is upper bounded by $N2^{-N^{\beta'}}$. Although (5.9) holds for arbitrarily distributed $M_\ell F_\ell$ (see the proof of Lemma 2.4.2), to use the sufficiency of mod-$\Lambda$ proved in Lemma 5.2.2, we firstly assume that $X_\ell$ is uniformly and randomly distributed at each level, and thus we have uniform $M_\ell$ and $F_\ell$. We will further show that the distribution of $M_\ell$ and $F_\ell$ can be made arbitrary in Remark 5.2.2.

According to Lemma 5.2.2, the constructed polar code can also guarantee the same upper-bound on the mutual information between the input message and the output of the channel $W'(X_\ell; Z|X_{1:\ell-1})$, as shown in the following inequality ($X_\ell$ is independent of the previous $X_{1:\ell-1}$):

$$I\left(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}\right) \leq N2^{-N^{\beta'}}. \tag{5.10}$$

Recall that $Z^{[N]}$ is the signal received by Eve after the mod-$\Lambda_r$ operation. From the chain rule of mutual information,

$$\begin{aligned} & I\left(MF; Z^{[N]}\right) \\ &= \sum_{\ell=1}^{r} I\left(Z^{[N]}; M_\ell F_\ell | M_{1:\ell-1} F_{1:\ell-1}\right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\ell=1}^{r} H(M_\ell F_\ell | M_{1:\ell-1} F_{1:\ell-1}) - H\Big(M_\ell F_\ell | Z^{[N]}, M_{1:\ell-1} F_{1:\ell-1}\Big) \\
&\le \sum_{\ell=1}^{r} H(M_\ell F_\ell) - H\Big(M_\ell F_\ell | Z^{[N]}, M_{1:\ell-1} F_{1:\ell-1}\Big) \qquad (5.11) \\
&= \sum_{\ell=1}^{r} I\Big(M_\ell F_\ell ; Z^{[N]}, M_{1:\ell-1} F_{1:\ell-1}\Big) \\
&\le \sum_{\ell=1}^{r} I\Big(M_\ell F_\ell ; Z^{[N]}, X_{1:\ell-1}^{[N]}\Big) \le r N 2^{-N^{\beta'}},
\end{aligned}
$$

where the second inequality holds because $I\Big(M_\ell F_\ell ; Z^{[N]}, X_{1:\ell-1}^{[N]}\Big) = I\Big(M_\ell F_\ell ; Z^{[N]}, U_{1:\ell-1}^{[N]}\Big)$ and adding more variables will not decrease the mutual information. Therefore strong secrecy is achieved since $\lim_{N\to\infty} I\Big(MF; Z^{[N]}\Big) = 0$.

*Remark* 5.2.2. Note that the above analysis actually implies *semantic security*, i.e., (5.11) holds for arbitrarily distributed $M$ and $F$. This is because of the symmetric nature of the $\Lambda_b/\Lambda_e$ channel [50]. We firstly ignore the reliability condition for Bob, and treat $M$ and $F$ together as the confidential message. In this case, $\Lambda_b$ is no longer an AWGN-good lattice, and we will show that $F$ can be further fixed without breaking the security. Since the message $MF$ is drawn from $\mathcal{R}(\Lambda_e)$ and the random bits are drawn from $\Lambda_e \cap \mathcal{R}(\Lambda_s)$, by Lemma 5.2.1, the mod-$\Lambda_e$ map is information lossless and its output is a sufficient statistic for $MF$. In this sense, the channel between the confidential message and the eavesdropper's signal can be viewed as a $\Lambda_b/\Lambda_e$ channel. Because the $\Lambda_b/\Lambda_e$ channel is symmetric, the maximum mutual information is achieved by the uniform input. Consequently, the mutual information corresponding to other input distributions can also be upper-bounded by $rN2^{-N^{\beta'}}$ in (5.11), and we can freeze the bits $F$ for reliability. It is worth mentioning this $\Lambda_b/\Lambda_e$ channel can be seen as the counterpart in lattice coding of the randomness-induced channel defined in [13].

**Theorem 5.2.3** (Achieving secrecy capacity of the mod-$\Lambda_s$ GWC)**.** *Consider a polar lattice $L$ constructed according to* (5.8) *with the binary lattice partition chain*

*$\Lambda/\Lambda_1/\cdots/\Lambda_r$ and $r$ binary nested polar codes with block length $N$. Scale $\Lambda$ and $r$ to satisfy the following conditions:*

*(i)* $h(\Lambda, \sigma_b^2) \to \log(V(\Lambda))$,

*(ii)* $h(\Lambda_r, \sigma_e^2) \to \frac{1}{2}\log(2\pi e\sigma_e^2)$.

*Given $\sigma_e^2 > \sigma_b^2$, all strong secrecy rates $R$ satisfying*

$$R < \frac{1}{2}\log\frac{\sigma_e^2}{\sigma_b^2}$$

*are achievable as $N \to \infty$, using the polar lattice $L$ on the mod-$\Lambda_s$ Gaussian wiretap channel.*

*Proof.* By Lemma 2.4.3 and (5.8),

$$\begin{aligned}
\lim_{N\to\infty} R &= \sum_{\ell=1}^{r} \lim_{N\to\infty} \frac{|\mathcal{A}_\ell|}{N} \\
&= \sum_{\ell=1}^{r} C(V_\ell) - C(W_\ell) \\
&= \sum_{\ell=1}^{r} C(V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)) - C(W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)) \\
&= C(V(\Lambda/\Lambda_r, \sigma_b^2)) - C(W(\Lambda/\Lambda_r, \sigma_e^2)) \\
&= C(\Lambda_r, \sigma_b^2) - C(\Lambda, \sigma_b^2) - C(\Lambda_r, \sigma_e^2) + C(\Lambda, \sigma_e^2) \\
&= h(\Lambda_r, \sigma_e^2) - h(\Lambda_r, \sigma_b^2) + h(\Lambda, \sigma_b^2) - h(\Lambda, \sigma_e^2) \\
&= \frac{1}{2}\log\frac{\sigma_e^2}{\sigma_b^2} - (\epsilon_e - \epsilon_b) - \epsilon_1,
\end{aligned}$$

where

$$\begin{cases}
\epsilon_1 = h(\Lambda, \sigma_e^2) - h(\Lambda, \sigma_b^2) \geq 0, \\
\epsilon_b = h(\sigma_b^2) - h(\Lambda_r, \sigma_b^2) = \frac{1}{2}\log(2\pi e\sigma_b^2) - h(\Lambda_r, \sigma_b^2) \geq 0, \\
\epsilon_e = h(\sigma_e^2) - h(\Lambda_r, \sigma_e^2) = \frac{1}{2}\log(2\pi e\sigma_e^2) - h(\Lambda_r, \sigma_e^2) \geq 0
\end{cases}$$

and $\epsilon_e - \epsilon_b \geq 0$.

By scaling $\Lambda$, we can have $h(\Lambda, \sigma_b^2) \to \log(V(\Lambda))$. Since $\sigma_e^2 > \sigma_b^2$, we also have $h(\Lambda, \sigma_e^2) \to \log(V(\Lambda))$ and thus $\epsilon_1 \approx 0$. The number of levels is also increased until $h(\Lambda_r, \sigma_e^2) \approx \frac{1}{2}\log(2\pi e\sigma_e^2)$, hence $h(\Lambda_r, \sigma_b^2) \approx \frac{1}{2}\log(2\pi e\sigma_e^2)$, such that both $\epsilon_b$ and $\epsilon_e$ are almost 0. Therefore by scaling $\Lambda$ and adjusting $r$, the secrecy rate can get arbitrarily close to $\frac{1}{2}\log\frac{\sigma_e^2}{\sigma_b^2}$. □

*Remark* 5.2.3. The secrecy capacity of the mod-$\Lambda_s$ Gaussian wiretap channel per use is given by

$$\begin{aligned} C_s &= \frac{1}{N}C(\Lambda_s, \sigma_b^2) - \frac{1}{N}C(\Lambda_s, \sigma_e^2) \\ &= \frac{1}{N}h(\Lambda_s, \sigma_e^2) - \frac{1}{N}h(\Lambda_s, \sigma_b^2), \end{aligned} \tag{5.12}$$

since the wiretapper's channel is degraded with respect to the main channel. Because $h(\Lambda_r, \sigma_e^2) \to \frac{1}{2}\log(2\pi e\sigma_e^2)$ and $\Lambda_s \subset \Lambda_r^N$, we have $\frac{1}{N}h(\Lambda_s, \sigma_e^2) \to \frac{1}{2}\log(2\pi e\sigma_e^2)$ and $\frac{1}{N}h(\Lambda_s, \sigma_b^2) \to \frac{1}{2}\log(2\pi e\sigma_b^2)$. Hence $C_s \to \frac{1}{2}\log\frac{\sigma_e^2}{\sigma_b^2}$. It also equals the secrecy capacity of the Gaussian wiretap channel when the signal power goes to infinity. It is noteworthy that we successfully remove the $\frac{1}{2}$-nat gap in the achievable secrecy rate derived in [21] which is caused by the limitation of the $L^\infty$ distance associated with the flatness factor.

*Remark* 5.2.4. We note that conditions (i) and (ii) in the theorem are just mild ones. When the $\sigma_e^2 = 4$ and $\sigma_b^2 = 1$, the gap from $\frac{1}{2}\log\frac{\sigma_e^2}{\sigma_b^2}$ is only $0.05$ when we choose $r = 3$ and partition chain $\eta(\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z})$ with scaling factor $\eta = 2.5$.

*Remark* 5.2.5. From conditions (i) and (ii), we can see that the construction for secrecy-good lattices requires more levels than the construction of AWGN-good lattices. $\epsilon_1$ can be made arbitrarily small by scaling down $\Lambda$ such that both $h(\Lambda, \sigma_e^2)$ and $h(\Lambda, \sigma_b^2)$ are sufficiently close to $\log V(\Lambda)$. For polar lattices for AWGN-goodness [51], we only need $h(\Lambda_{r'}, \sigma_b^2) \approx \frac{1}{2}\log(2\pi e\sigma_b^2)$ for some $r' < r$. S-

ince $\epsilon_b < \epsilon_e$, $\Lambda_{r'}$ may be not enough for the wiretapper's channel. Therefore, more levels are needed in the wiretap coding context. To satisfy the condition $h(\Lambda_r, \sigma_e^2) \to \frac{1}{2}\log(2\pi e\sigma_e^2)$, it is sufficient to guarantee that $P_e(\Lambda_r, \sigma_e^2) \to 0$ by [50, Theorem 13]. When one-dimensional binary partition $\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z}/...$ is used, we have $P_e(\Lambda_r, \sigma_e^2) \leq Q(\frac{2^r}{2\sigma_e}) \leq e^{-\frac{2^{2r}}{8\sigma_e^2}}$, where $Q(\cdot)$ is the Q-function. Letting $r = O(\log N)$, the error probability vanishes as $P_e(\Lambda_r, \sigma_e^2) = e^{-O(N)}$, which implies that $h(\Lambda_r, \sigma_e^2) \to \frac{1}{2}\log(2\pi e\sigma_e^2)$ as $N \to \infty$.

### 5.2.2 Reliability over Mod-$\Lambda_s$ GWC

In the original polar coding scheme for the binary wiretap channel [13], how to assign set $\mathcal{D}$ is a problem. Assigning frozen bits to $\mathcal{D}$ guarantees reliability but only achieves weak secrecy, whereas assigning random bits to $\mathcal{D}$ guarantees strong secrecy but may violate the reliability requirement because $\mathcal{D}$ may be nonempty. In order to ensure strong secrecy, $\mathcal{D}$ is assigned with random bits ($\mathcal{D} \leftarrow \mathsf{R}$), which makes this scheme failed to accomplish the theoretical reliability. For any $\ell$-th level channel $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ at Bob's end, the probability of error is upper bounded by the sum of the Bhattacharyya parameters $\tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$ of subchannels that are not frozen to zero. For each bit-channel index $j$ and $\beta < 0.5$, we have

$$j \in \mathcal{G}(V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)) \cup \mathcal{D}_\ell. \tag{5.13}$$

By the definition (2.25), the sum of $\tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$ over the set $\mathcal{G}(V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ is bounded by $2^{-N^\beta}$, therefore the error probability of the $\ell$-th level channel under the SC decoding, denoted by $P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$, can be upper bounded by

$$P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2) \leq N2^{-N^\beta} + \sum_{j\in\mathcal{D}_\ell} \tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)). \tag{5.14}$$

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

$$\Pr\{\hat{M} \neq M\} \leq \sum_{i=1}^{r} P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2). \tag{5.15}$$

Unfortunately, a proof that this scheme satisfies the reliability condition cannot be attained here because the bound of the sum $\sum_{j\in\mathcal{D}_\ell} \tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$ is not known. Note that significantly low probabilities of error can still be achieved in practice since the size of $\mathcal{D}_\ell$ is very small for sufficiently large $N$.

The reliability problem was recently solved in [44], where a new scheme dividing the information message into several blocks was proposed. For a specific block, $\mathcal{D}_\ell$ is still assigned with random bits and transmitted in advance in the set $\mathcal{A}_\ell$ 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 $\ell$-th level partition $\Lambda_{\ell-1}/\Lambda_\ell$, $P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ vanishes as $N \to \infty$, then the total decoding error probability for Bob can be made arbitrarily small. Consequently, based on this new scheme of assigning the problematic set, the error probability on level $\ell$ can be upper bounded by

$$P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2) \leq \epsilon_{N'}^{\ell} + k_\ell \cdot O(2^{-N'^{\beta}}), \tag{5.16}$$

where $k_\ell$ is the number of information blocks on the $\ell$-th level, $N'$ is the length of each block which satisfies $N' \times k_\ell = N$ and $\epsilon_{N'}^{\ell}$ is caused by the first separate block on the $\ell$-th level consisting of the initial bits in $\mathcal{D}_\ell$. Since $|\mathcal{D}_\ell|$ 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. Meanwhile, by the analysis in [104], when $h(\Lambda, \sigma_b^2) \to \log(V(\Lambda))$, $h(\Lambda_r, \sigma_b^2) \to \frac{1}{2}\log(2\pi e\sigma_b^2)$, and

$R_C \to C(\Lambda/\Lambda_r, \sigma_b^2)$, we have $\gamma_{\Lambda_b}(\sigma_b) \to 2\pi e$. Therefore, $\Lambda_b$ is an AWGN-good lattice[6].

Note that the rate loss incurred by repeatedly transmitted bits in $\mathcal{D}_\ell$ is negligible because of its small size. Specifically, the actual secrecy rate in the $\ell$-th level is given by $\frac{k_\ell}{k_\ell+1}[C(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2) - C(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)]$. Clearly, this rate can be made close to the secrecy capacity by choosing sufficiently large $k_\ell$ as well.

## 5.3 Secrecy-good polar lattices with discrete Gaussian shaping

In this section, we apply Gaussian shaping on the AWGN-good and secrecy-good polar lattices. The idea of lattice Gaussian shaping was proposed in [20] and then implemented in [104] to construct capacity-achieving polar lattices. For wiretap coding, the discrete Gaussian distribution can also be utilized to satisfy the power constraint. In simple terms, after obtaining the AWGN-good lattice $\Lambda_b$ and the secrecy-good lattice $\Lambda_e$, Alice still maps each message $m$ to a coset $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$ as mentioned in Sect. 5.2. However, instead of the mod-$\Lambda_s$ operation, Alice samples the encoded signal $X^N$ from $D_{\Lambda_e+\lambda_m,\sigma_s}$, where $\lambda_m$ is the coset representative of $\tilde{\lambda}_m$ and $\sigma_s^2$ is arbitrarily close to the signal power $P_s$ (see [21] for more details). Based on the lattice Gaussian shaping, we will propose a new partition for the genuine GWC. We will also show that this shaping operation does not hurt the secrecy rate and that the proposed scheme is semantically secure.

[6]More precisely, to make $\Lambda_b$ AWGN-good, we need $P_e(\Lambda_b, \sigma_b^2) \to 0$ by definition. By [104, Theorem 2], $P_e(\Lambda_b, \sigma_b^2) \le rN2^{-N^\beta} + N \cdot P_e(\Lambda_r, \sigma_b^2)$. According to the analysis in Remark 5.2.5, $r = O(\log N)$ is sufficient to guarantee $P_e(\Lambda_r, \sigma_b^2) = e^{-O(N)}$, meaning that a sub-exponentially vanishing $P_e(\Lambda_b, \sigma_b^2)$ can be achieved.

### 5.3.1 Gaussian shaping over polar lattices

As shown in Section 4.3, the shaping scheme is based on the technique of polar codes for asymmetric channels. Recall that the Bhattacharyya parameter for a binary memoryless asymmetric (BMA) channel is defined in Definition 4.3.1. To obtain the desired input distribution of $P_X$ for $W$, the indices with very small $Z(U^i|U^{1:i-1})$ should be removed from the information set of the symmetric channel. Following 4.3, the resultant subset is referred to as the information set $\mathcal{I}$ for the asymmetric channel $W$. For the remaining part $\mathcal{I}^c$, we further find out that there are some bits which can be made independent of the information bits and uniformly distributed. The purpose of extracting such bits is for the interest of our lattice construction. We name the set that includes those independent frozen bits as the independent frozen set $\mathcal{F}$, and 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 set $\mathcal{S}$. The three sets are formally defined as follows:

$$\begin{cases} \text{the independent frozen set: } \mathcal{F} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{[N]}) \geq 1 - 2^{-N^\beta}\} \\ \text{the information set: } \mathcal{I} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{[N]}) \leq 2^{-N^\beta} \text{ and} \\ \qquad\qquad Z(U^i|U^{1:i-1}) \geq 1 - 2^{-N^\beta}\} \\ \text{the shaping set: } \mathcal{S} = (\mathcal{F} \cup \mathcal{I})^c. \end{cases} \tag{5.17}$$

To identify these three sets, one can use Theorem 4.3.5 to calculate $Z(U^i|U^{1:i-1}, Y^{[N]}, X^{[N]})$ using the known constructing techniques for symmetric polar codes [5][6]. We note that $Z(U^i|U^{1:i-1})$ can be computed in a similar way, by constructing a symmetric channel between $\tilde{X}$ and $X \oplus \tilde{X}$. Besides the construction, the decoding process for the asymmetric polar codes can also be converted to the decoding for the symmetric polar codes.

The polar coding scheme according to (5.17), which can be viewed as an exten-

sion of the scheme proposed in [10], has been proved to be capacity-achieving in [104]. Moreover, it can be extended to the construction of multilevel asymmetric polar codes as shown in the previous chapter.

### 5.3.2 Three-dimensional partition

Now we consider the partition of the index set $[N]$ with shaping involved. According to the analysis of asymmetric polar codes, we have to eliminate those indices with small $Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:l-1}^{[N]})$ from the information set of the symmetric channels. Therefore, Alice cannot send message on those subchannels with $Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}$. Note that this part is the same for $\tilde{V}_\ell$ and $\tilde{W}_\ell$, because it only depends on the shaping distribution. At each level, the index set which is used for shaping is given as

$$\mathcal{S}_\ell \triangleq \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}\}, \tag{5.18}$$

and the index set which is not for shaping 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 (2.25)). We rewrite the reliability-good index set $\mathcal{G}_\ell$ and information-bad index set $\mathcal{N}_\ell$ at level $\ell$ as

$$\begin{aligned} \mathcal{G}_\ell &\triangleq \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}, Y^{[N]}) \le 2^{-N^\beta}\}, \\ \mathcal{N}_\ell &\triangleq \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}, Z^{[N]}) \ge 1 - 2^{-N^\beta}\}. \end{aligned} \tag{5.19}$$

Note that $\mathcal{G}_\ell$ and $\mathcal{N}_\ell$ are defined by the asymmetric Bhattacharyya parameters. Nevertheless, by Theorem 4.3.5 and the channel equivalence, we have $\mathcal{G}_\ell = \mathcal{G}(\tilde{V}_\ell)$ and $\mathcal{N}_\ell = \mathcal{N}(\tilde{W}_\ell)$ as defined in (2.25), where $\tilde{V}_\ell$ and $\tilde{W}_\ell$ are the respective symmetric channels or the MMSE-scaled $\Lambda_{\ell-1}/\Lambda_\ell$ channels for Bob and Eve at level $\ell$. The four sets $\mathcal{A}_\ell$, $\mathcal{B}_\ell$, $\mathcal{C}_\ell$, and $\mathcal{D}_\ell$ are defined in the same fashion as (2.28), with $\mathcal{G}_\ell$ and $\mathcal{N}_\ell$

replacing $\mathcal{G}(\tilde{V}_\ell)$ and $\mathcal{N}(\tilde{W}_\ell)$, respectively. Now the whole index set $[N]$ is divided like a cube in three directions, which is shown in Fig. 5.4.

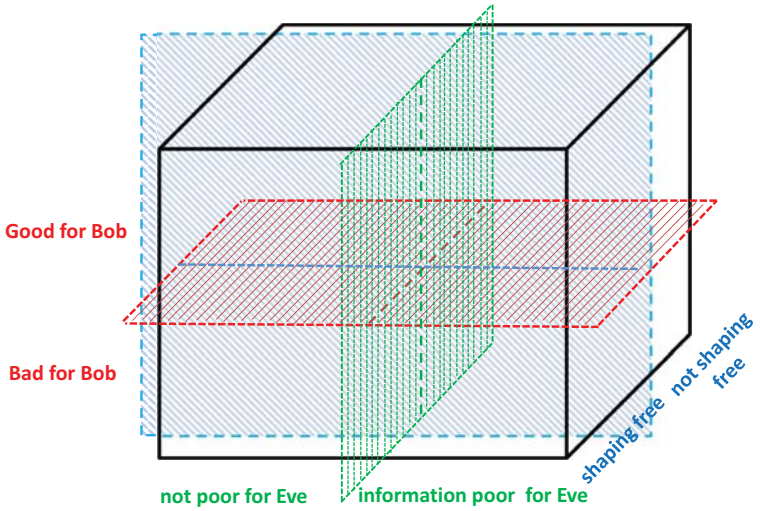



Figure 5.4: Partitions of the index set $[N]$ with shaping.

Clearly, we have eight blocks:

$$\begin{aligned}
\mathcal{A}_\ell^{\mathcal{S}} &= \mathcal{A}_\ell \cap \mathcal{S}_\ell,\ \mathcal{A}_\ell^{\mathcal{S}^c} = \mathcal{A}_\ell \cap \mathcal{S}_\ell^c \\
\mathcal{B}_\ell^{\mathcal{S}} &= \mathcal{B}_\ell \cap \mathcal{S}_\ell,\ \mathcal{B}_\ell^{\mathcal{S}^c} = \mathcal{B}_\ell \cap \mathcal{S}_\ell^c \\
\mathcal{C}_\ell^{\mathcal{S}} &= \mathcal{C}_\ell \cap \mathcal{S}_\ell,\ \mathcal{C}_\ell^{\mathcal{S}^c} = \mathcal{C}_\ell \cap \mathcal{S}_\ell^c \\
\mathcal{D}_\ell^{\mathcal{S}} &= \mathcal{D}_\ell \cap \mathcal{S}_\ell,\ \mathcal{D}_\ell^{\mathcal{S}^c} = \mathcal{D}_\ell \cap \mathcal{S}_\ell^c
\end{aligned} \tag{5.20}$$

By Lemma 4.3.2, we observe that $\mathcal{A}_\ell^{\mathcal{S}} = \mathcal{C}_\ell^{\mathcal{S}} = \emptyset$, $\mathcal{A}_\ell^{\mathcal{S}^c} = \mathcal{A}_\ell$, and $\mathcal{C}_\ell^{\mathcal{S}^c} = \mathcal{C}_\ell$. The shaping set $\mathcal{S}_\ell$ is divided into two sets $\mathcal{B}_\ell^{\mathcal{S}}$ and $\mathcal{D}_\ell^{\mathcal{S}}$. The bits in $\mathcal{S}_\ell$ are determined by the bits in $\mathcal{S}_\ell^c$ according to the mapping. Similarly, $\mathcal{S}_\ell^c$ is divided into the four sets $\mathcal{A}_\ell^{\mathcal{S}^c} = \mathcal{A}_\ell$, $\mathcal{B}_\ell^{\mathcal{S}^c}$, $\mathcal{C}_\ell^{\mathcal{S}^c} = \mathcal{C}_\ell$, and $\mathcal{D}_\ell^{\mathcal{S}^c}$. Note that for wiretap coding, the frozen set becomes $\mathcal{C}_\ell^{\mathcal{S}^c}$, which is slightly different from the frozen set for channel coding. To satisfy the reliability condition, the frozen set $\mathcal{C}_\ell^{\mathcal{S}^c}$ and the problematic set $\mathcal{D}_\ell^{\mathcal{S}^c}$ cannot be set uniformly random any more. Recall that only the independent frozen set $\mathcal{F}_\ell$ at each level, which is defined as $\{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, Y^{[N]}, X_{1:\ell-1}^{[N]}) \geq 1 - 2^{-N^\beta}\}$, can be set uniformly random (which are already shared between Alice

and Bob), and the bits in the unpolarized frozen set $\bar{\mathcal{F}}_\ell$, defined as $\{i \in [N] : 2^{-N^\beta} < Z(U_\ell^i|U_\ell^{1:i-1}, Y^{[N]}, X_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}\}$, should be determined according to the mapping. Moreover, we can observe that $\mathcal{F}_\ell \subset \mathcal{C}_\ell^{\mathcal{S}^c}$ and $\mathcal{D}_\ell^{\mathcal{S}^c} \subset \mathcal{D}_\ell \subset \bar{\mathcal{F}}_\ell$. Here we make the bits in $\mathcal{F}_\ell$ uniformly random and the bits in $\mathcal{C}_\ell^{\mathcal{S}^c} \setminus \mathcal{F}_\ell$ and $\mathcal{D}_\ell^{\mathcal{S}^c}$ determined by the mapping. Therefore, from now on, we adjust the definition of the shaping bits as:

$$\begin{aligned}\mathcal{S}_\ell \triangleq \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta} \text{ or} \\ 2^{-N^\beta} < Z(U_\ell^i|U_\ell^{1:i-1}, Y^{[N]}, X_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}\},\end{aligned} \tag{5.21}$$

which is essentially equivalent to the definition of the shaping set given in Theorem 4.3.9

To sum up, at level $\ell$, we assign the sets $\mathcal{A}_\ell^{\mathcal{S}^c}$, $\mathcal{B}_\ell^{\mathcal{S}^c}$, and $\mathcal{F}_\ell$ with message bits $M_\ell$, uniformly random bits $\mathsf{R}_\ell$, and uniform frozen bits $F_\ell$, respectively. The rest bits $\mathsf{S}_\ell$ (in $\mathcal{S}_\ell$) will be fed with random bits according to $P_{U_\ell^i|U_\ell^{1:i-1}, X_{1:l-1}^{[N]}}$. Clearly, this shaping operation will make the input distribution arbitrarily close to $P_{X_\ell|X_{1:\ell-1}}$. In this case, we can obtain the equality between the Bhattacharyya parameter of asymmetric setting and symmetric setting (see Theorem 4.3.5). This provides us a convenient way to prove the strong secrecy of the wiretap coding scheme with shaping because we have already proved the strong secrecy of a symmetric wiretap coding scheme using the Bhattacharyya parameter of the symmetric setting. A detailed proof will be presented in the following subsection. Before this, we show that the shaping will not change the message rate.

**Lemma 5.3.1.** *For the symmetrized main channel $\tilde{V}_\ell$ and wiretapper's channel $\tilde{W}_\ell$, consider the reliability-good indices set $\mathcal{G}_\ell$ and information-bad indices set $\mathcal{N}_\ell$ defined as in* (5.19)*. By eliminating the shaping set $\mathcal{S}_\ell$ from the original message set defined in* (2.28)*, 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 propor-*

*tion 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* $\frac{1}{2}\log\frac{\tilde{\sigma}_e^2}{\tilde{\sigma}_b^2}$.

*Proof.* By Theorem 5.2.3, when shaping is not involved, the message rate can be made arbitrarily close to $\frac{1}{2}\log\frac{\tilde{\sigma}_e^2}{\tilde{\sigma}_b^2}$. By the new definition (5.21) of $\mathcal{S}_\ell$, we still have $\mathcal{A}_\ell^{\mathcal{S}} = \emptyset$, which means the shaping operation will not affect the message rate. □

### 5.3.3 Strong secrecy

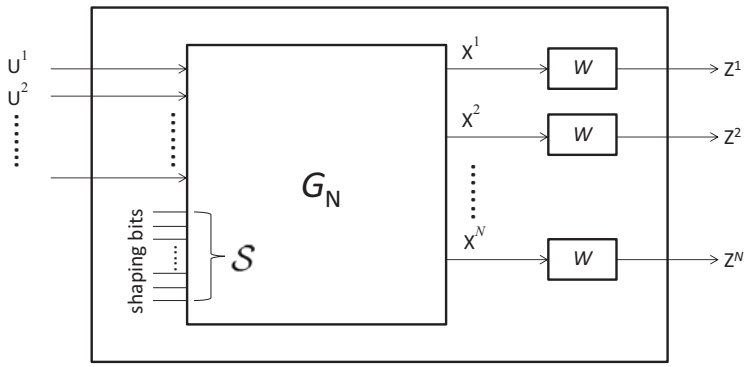



Figure 5.5: Block diagram of the shaping-induced channel $\mathcal{Q}_N(W, \mathcal{S})$.

In [13], an induced channel is defined in order to prove strong secrecy. Here we call it the randomness-induced channel because it is induced by feeding the subchannels in the sets $\mathcal{B}_\ell$ and $\mathcal{D}_\ell$ with uniformly random bits. However, when shaping is involved, the set $\mathcal{B}_\ell$ and $\mathcal{D}_\ell$ are no longer fed with uniformly random bits. In fact, some subchannels (covered by the shaping mapping) should be fed with bits according to a random mapping. We define the channel induced by the shaping bits as the shaping-induced channel.

**Definition 5.3.1** (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 $|\mathcal{S}|$. The input alphabet of $\mathcal{Q}_N(W, \mathcal{S})$ is $\{0,1\}^{N-|\mathcal{S}|}$ and the bits in $\mathcal{S}$ are determined by the input bits according to a random shaping $\Phi_{\mathcal{S}}$. A block diagram of the shaping induced channel is shown in Fig. 5.5.

Based on the shaping-induced channel, we define a new induced channel, which is caused by feeding a part of the input bits of the shaping-induced channel with uniformly random bits.

**Definition 5.3.2** (New induced channel)**.** Based on a shaping induced channel $\mathcal{Q}_N(W, \mathcal{S})$, the new induced channel $\mathcal{Q}_N(W, \mathcal{S}, \mathcal{R})$ is specified in terms of a randomness subset $\mathcal{R}$ of size $|\mathcal{R}|$. The randomness is introduced into the input set of the shaping-induced channel. The input alphabet of $\mathcal{Q}_N(W, \mathcal{S}, \mathcal{R})$ is $\{0,1\}^{N-|\mathcal{S}|-|\mathcal{R}|}$ and the bits in $\mathcal{R}$ are uniformly and independently random. A block diagram of the new induced channel is shown in Fig. 5.6.

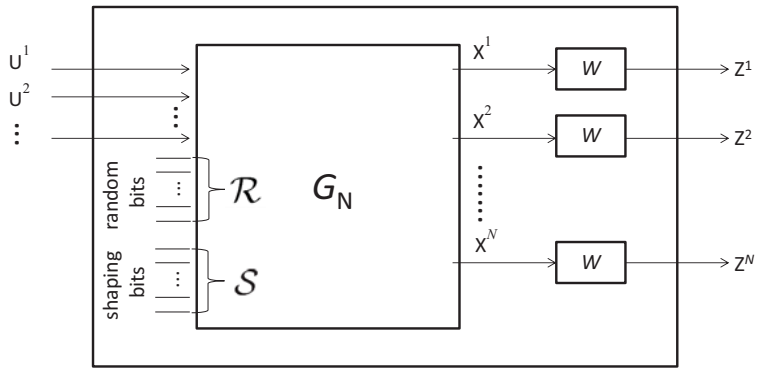



Figure 5.6: Block diagram of the new induced channel $\mathcal{Q}_N(W, \mathcal{S}, \mathcal{R})$.

The new induced channel is a combination of the shaping-induced channel and randomness-induced channel. This is different from the definition given in [13] 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 the random mapping, the new induced channel can still generate $2^N$ possible realizations $x_\ell^{[N]}$ of $X_\ell^{[N]}$ as $N$ goes to infinity, and those $x_\ell^{[N]}$ can be viewed as the output of $N$ i.i.d binary sources with input distribution $P_{X_\ell|X_{1:\ell-1}}$. These are exactly the conditions required by Theorem 4.3.5. Specifically, we have $Z\big(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}, Z^{[N]}\big) = \tilde{Z}\big(\tilde{U}_\ell^i|\tilde{U}_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}, X_\ell^{[N]} \oplus \tilde{X}_\ell^{[N]}, Z^{[N]}\big)$. In simple words, this equation holds when $x_\ell^{[N]}$ and $x_\ell^{[N]} \oplus \tilde{x}_\ell^{[N]}$ are all s-

elected from $\{0,1\}^N$ according to their respective distributions. Then we can exploit the relation between the asymmetric channel and the corresponding symmetric channel 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 random mapping. This is similar to the setting of the randomness induced channel in [13], where the input distribution and the randomness distribution are both set to be uniform. In [13], the randomness-induced channel is further proved to be symmetric; then any other input distribution can also achieve strong secrecy and the symmetry finally results in semantic security. In this work, however, we do not have a proof of the symmetry of the new induced channel. For this reason, we assume for now that the message bits are uniform distributed. To prove semantic security, we will show that the information leakage of the symmetrized version of the new induced channel is vanishing in Sect. 5.3.4.

**Lemma 5.3.2.** *Let $M_\ell$ be the uniformly distributed message bits and $F_\ell$ be the independent frozen bits at the input of the channel at the $\ell$-th level. When shaping bits $\mathsf{S}_\ell$ are selected according to the random mapping $\Phi_{\mathcal{S}_\ell}$* [7] *and $N$ is sufficiently large, the mutual information can be upper-bounded as*

$$I\left(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}\right) \leq O\left(N^2 2^{-N^{\beta'}}\right). \tag{5.22}$$

*Proof.* We firstly assume that $U_\ell^i$ is selected according to the distribution $P_{U_\ell^i|U_\ell^{1:i-1},X_{1:\ell-1}^{[N]}}$ for all $i \in [N]$, i.e.,

$$u_\ell^i = \begin{cases} 0 & \text{with probability } P_{U_\ell^i|U_\ell^{1:i-1},X_{1:\ell-1}^{[N]}}(0|u_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}), \\ 1 & \text{with probability } P_{U_\ell^i|U_\ell^{1:i-1},X_{1:\ell-1}^{[N]}}(1|u_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}). \end{cases} \tag{5.23}$$

[7] As we will see in Sect. 5.3.5, to achieve reliability, $\Phi_{\mathcal{S}_\ell}$ should be shared between Alice and Bob, or we have to use the Markov block coding technique.

for all $i \in [N]$. In this case, the input distribution $P_{X_\ell|X_{1:\ell-1}}$ at each level is exactly the optimal input distribution obtained from the lattice Gaussian distribution. The mutual information between $M_\ell F_\ell$ and $(Z^{[N]}, X_{1:\ell-1}^{[N]})$ in this case is denoted by $I_P(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]})$.

For the shaping induced channel $\mathcal{Q}_N(W_\ell, \mathcal{S}_\ell, \mathcal{R}_\ell)$ ($\mathcal{R}_\ell$ is $\mathcal{B}_\ell^{\mathcal{S}^c}$ according to the above analysis), we write the indices of the input bits $(\mathcal{S}_\ell \cup \mathcal{R}_\ell)^c = [N] \setminus (\mathcal{S}_\ell \cup \mathcal{R}_\ell)$ as $\{i_1, i_2, ..., i_{N-s_\ell-r_\ell}\}$, where $|\mathcal{R}| = r_\ell$ and $|\mathcal{S}_\ell| = s_\ell$, and assume that $i_1 < i_2 < \cdots < i_{N-s_\ell-r_\ell}$. We have

$$\begin{aligned} I_P(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) &= I_P(U_\ell^{(\mathcal{S}_\ell \cup \mathcal{R}_\ell)^c}; Z^{[N]}, X_{1:\ell-1}^{[N]}) \\ &= I_P(U_\ell^{i_1}, U_\ell^{i_2}, ..., U_\ell^{i_{N-r_\ell-s_\ell}}; Z^{[N]}, X_{1:\ell-1}^{[N]}) \\ &= \sum_{j=1}^{N-r_\ell-s_\ell} I_P(U_\ell^{i_j}; Z^{[N]}, X_{1:\ell-1}^{[N]} | U_\ell^{i_1}, U_\ell^{i_2}, ..., U_\ell^{i_{j-1}}) \\ &= \sum_{j=1}^{N-r_\ell-s_\ell} I_P(U_\ell^{i_j}; Z^{[N]}, X_{1:\ell-1}^{[N]}, U_\ell^{i_1}, U_\ell^{i_2}, ..., U_\ell^{i_{j-1}}) \\ &\overset{(a)}{\leq} \sum_{j=1}^{N-r_\ell-s_\ell} I_P(U_\ell^{i_j}; Z^{[N]}, X_{1:\ell-1}^{[N]}, U_\ell^{1}, U_\ell^{2}, ..., U_\ell^{i_j-1}), \end{aligned} \tag{5.24}$$

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:

$$\begin{aligned} &\sum_{j=1}^{N-r_\ell-s_\ell} I_P(U_\ell^{i_j}; Z^{[N]}, X_{1:\ell-1}^{[N]}, U_\ell^{1:i_j-1}) \\ &\quad \overset{(a)}{\leq} \sum_{j=1}^{N-r_\ell-s_\ell} I(\tilde{U}_\ell^{i_j}; Z^{[N]}, X_{1:\ell-1}^{[N]}, \tilde{X}_\ell^{[N]} \oplus X_\ell^{[N]}, \tilde{U}_\ell^{1:i_j-1}) \\ &\qquad + H(\tilde{U}_\ell^{i_j} | Z^{[N]}, X_{1:\ell-1}^{[N]}, \tilde{X}_\ell^{[N]} \oplus X_\ell^{[N]}, \tilde{U}_\ell^{1:i_j-1}) - \sum_{j=1}^{N-r_\ell-s_\ell} H(U_\ell^{i_j} | Z^{[N]}, X_{1:\ell-1}^{[N]}, U_\ell^{1:i_j-1}) \end{aligned}$$

$$\begin{aligned}
&\overset{(b)}{\leq} \sum_{j=1}^{N-r_\ell-s_\ell} I(\tilde{U}_\ell^{i_j}; Z^{[N]}, X_{1:\ell-1}^{[N]}, \tilde{X}_\ell^{[N]} \oplus X_\ell^{[N]}, \tilde{U}_\ell^{1:i_j-1}) \qquad (5.25)\\
&+ \sum_{j=1}^{N-r_\ell-s_\ell} Z(U_\ell^{i_j}|Z^{[N]}, X_{1:\ell-1}^{[N]}, U_\ell^{1:i_j-1}) - (Z(U_\ell^{i_j}|Z^{[N]}, X_{1:\ell-1}^{[N]}, U_\ell^{1:i_j-1}))^2\\
&\overset{(c)}{\leq} \sum_{j=1}^{N-r_\ell-s_\ell} I(\tilde{U}_\ell^{i_j}; Z^{[N]}, X_{1:\ell-1}^{[N]}, \tilde{X}_\ell^{[N]} \oplus X_\ell^{[N]}, \tilde{U}_\ell^{1:i_j-1}) + N2^{-N^\beta}\\
&\overset{(d)}{\leq} N2^{-N^{\beta'}} + N2^{-N^\beta}\\
&\leq 2N2^{-N^{\beta'}}
\end{aligned}$$

for $0 < \beta' < \beta < 0.5$. Inequalities $(a)$-$(d)$ follow from

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

$(b)$ [8, Proposition 2] which gives $H(X|Y) - H(X|Y,Z) \leq Z(X|Y) - (Z(X|Y,Z)^2)$ and Theorem 4.3.5,

$(c)$ our coding scheme guaranteeing that $Z(U_\ell^{i_j}|Z^{[N]}, X_{1:\ell-1}^{[N]}, U_\ell^{1:i_j-1})$ is greater than $1 - 2^{-N^\beta}$ for the frozen bits and information bits,

$(d)$ Lemma 2.4.1.

However, for wiretap coding, the message $M_\ell$, frozen bits $F_\ell$ and random bits $R_\ell$ are all uniformly random, and the shaping bits $S_\ell$ are determined by $S_\ell^c$ according to $\Phi_{\mathcal{S}_\ell}$. Let $Q_{U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]}}$ denote the joint distribution of $(U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]})$ resulted from uniformly distributed $M_\ell F_\ell R_\ell$ and $S_\ell$ according to $\Phi_{\mathcal{S}_\ell}$. By the proofs of [104, Th. 5] and [104, Th. 6], the total variation distance can be bounded as

$$\|Q_{U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]}} - P_{U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]}}\| \leq N2^{-N^{\beta'}} \qquad (5.26)$$

for sufficiently large $N$.

By [11, Proposition 5], the mutual information $I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]})$ caused by

$Q_{U_\ell^{[N]},X_{1:\ell-1}^{[N]},Z^{[N]}}$ satisfies

$$\begin{aligned}|I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) - I_P(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]})| \\ \leq 7N2^{-N^{\beta'}} \log 2^N + h_2(N2^{-N^{\beta'}}) + h_2(4N2^{-N^{\beta'}}) \\ = O(N^2 2^{-N^{\beta'}}),\end{aligned} \tag{5.27}$$

where $h_2(\cdot)$ denotes the binary entropy function.

□

Finally, strong secrecy (for uniform message bits) can be proved in the same fashion as shown in (5.11) as:

$$\begin{aligned}I(M; Z^{[N]}) &\leq \sum_{\ell=1}^{r} I(M_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) \\ &\leq \sum_{\ell=1}^{r} I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) \\ &= O(rN^2 2^{-N^{\beta'}}).\end{aligned} \tag{5.28}$$

Therefore we conclude that the whole shaping scheme is secure in the sense that the mutual information leakage between $M$ and $Z^{[N]}$ vanishes with the block length $N$.

### 5.3.4 Semantic security

In this subsection, we extend strong secrecy of the constructed polar lattices to semantic security, namely the resulted strong secrecy does not rely on the distribution of the message. We take the level-1 wiretapper's channel $W_1$ as an example. Our goal is to show that the maximum mutual information between $M_1 F_1$ and $Z^{[N]}$ is vanishing for any input distribution as $N \to \infty$. Unlike the symmetric randomness induced channel introduced in [13], the new induced channel is generally asymmet-

ric with transition probability

$$Q(z|v) = \frac{1}{2^{r_1}} \sum_{\Phi_{\mathcal{S}_1}} P(\Phi_{\mathcal{S}_1}) \sum_{e \in \{0,1\}^{r_1}} W_1^N(z|(v, e, \Phi_{\mathcal{S}_1}(v, e))G_N),$$

where $\Phi_{\mathcal{S}_1}(v, e)$ represents the shaping bits determined by $v$ (the frozen bits and message bits together) and $e$ (the random bits) according to the random mapping $\Phi_{\mathcal{S}_1}$. It is difficult to find the optimal input distribution to maximize the mutual information for the new induced channel.

To prove the semantic security, we investigate the relationship between the $i$-th subchannel of $W_{1,N}$ and the $i$-th subchannel of its symmetrized version $\tilde{W}_{1,N}$, which are denoted by $W_1^{(i,N)}$ and $\tilde{W}_1^{(i,N)}$, respectively. According to Lemma 4.3.4, the asymmetric wiretap channel $W_1 : X_1 \to Z$ is symmetrized to channel $\tilde{W}_1 : \tilde{X}_1 \to (Z, \tilde{X}_1 \oplus X_1)$. After the $N$-by-$N$ polarization transform, we obtain $W_1^{(i,N)} : U_1^i \to (U_1^{1:i-1}, Z^{[N]})$ and $\tilde{W}_1^{(i,N)} : \tilde{U}_1^i \to (\tilde{U}_1^{1:i-1}, \tilde{X}_1^{[N]} \oplus X_1^{[N]}, Z^{[N]})$. The next lemma shows that if we symmetrize $W_1^{(i,N)}$ directly, i.e., construct a symmetric channel $\widetilde{W_1^{(i,N)}} : \tilde{U}_1^i \to (U_1^{1:i-1}, Z^{[N]}, \tilde{U}_1^i \oplus U_1^i)$ in the sense of Lemma 4.3.4, $\widetilde{W_1^{(i,N)}}$ is degraded with respect to $\tilde{W}_1^{(i,N)}$.

**Lemma 5.3.3.** *The symmetrized channel $\widetilde{W_1^{(i,N)}}$ derived directly from $W_1^{(i,N)}$ is degraded with respect to the $i$-th subchannel $\tilde{W}_1^{(i,N)}$ of $\tilde{W}_1$.*

*Proof.* According to the proof of [10, Theorem 2], we have the relationship

$$\tilde{W}_1^{(i,N)}(\tilde{u}_1^{1:i-1}, \tilde{x}_1^{[N]} \oplus x_1^{[N]}, z^{[N]}|\tilde{u}_1^i) = 2^{-N+1} P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]}).$$

Letting $\tilde{x}_1^{[N]} \oplus x_1^{[N]} = 0^{[N]}$, the equation becomes $\tilde{W}_1^{(i,N)}(u_1^{1:i-1}, 0^{[N]}, z^{[N]}|u_1^i) = 2^{-N+1} P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]})$, which has already been addressed in [10]. However, for a fixed $x_1^{[N]}$ and $\tilde{u}_1^i = u_1^i$, since $G_N$ is full rank, there are $2^{N-1}$ choices of $\tilde{x}_1^{[N]}$ remaining, which means that there exists $2^{N-1}$ outputs symbols of $\tilde{W}_1^{(i,N)}$ having the

same transition probability $2^{-N+1}P_{U_1^{1:i},Z^{[N]}}(u_1^{1:i}, z^{[N]})$. Suppose a middle channel which maps all these output symbols to one single symbol, which is with transition probability $P_{U_1^{1:i},Z^{[N]}}(u_1^{1:i}, z^{[N]})$. The same operation can be done for $\tilde{u}_1^i = u_1^i \oplus 1$, making another symbol with transition probability $P_{U_1^{1:i},Z^{[N]}}(u_1^{1:i}, z^{[N]})$ corresponding to the input $u_1^i \oplus 1$. This is a channel degradation process, and the degraded channel is symmetric.

Then we show that the symmetrized channel $\widetilde{W_1^{(i,N)}}$ is equivalent to the degraded channel mentioned above. By Lemma 4.3.4, the channel transition probability of $\widetilde{W_1^{(i,N)}}$ is

$$\widetilde{W_1^{(i,N)}}(u_1^{1:i-1}, \tilde{u}_1^i \oplus u_1^i, z^{[N]}|\tilde{u}_1^i) = P_{U_1^{1:i},Z^{[N]}}(u_1^{1:i}, z^{[N]}),$$

which is equal to the transition probability of the degraded channel discussed in the previous paragraph. Therefore, $\widetilde{W_1^{(i,N)}}$ is degraded with respect to $\tilde{W}_1^{(i,N)}$. □

*Remark* 5.3.1. In fact, a stronger relationship that $\widetilde{W_1^{(i,N)}}$ is equivalent to $\tilde{W}_1^{(i,N)}$ can be proved. This is because that the output symbols combined in the channel degradation process have the same LR. An evidence of this result can be found in [10, Equation (36)], where $\tilde{Z}(\tilde{W}_1^{(i,N)}) = Z(U_1^i|U_1^{1:i-1}, Z^{[N]}) = \tilde{Z}(\widetilde{W_1^{(i,N)}})$. Nevertheless, the degradation relationship is sufficient for this work. Notice that Lemma 5.3.3 can be generalized to high level $\ell$, with outputs $Z^{[N]}$ replaced by $(Z^{[N]}, X_{1:\ell-1}^{[N]})$.

Illuminated by Lemma 5.3.3, we can also symmetrize the new induced channel at level $\ell$ and show that it is degraded with respect to the randomness-induced channel constructed from $\tilde{W}_\ell$. For simplicity, letting $\ell = 1$, the new induced channel at level 1 is $\mathcal{Q}_N(W_1, \mathcal{S}_1, \mathcal{R}_1) : U_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \to Z^{[N]}$, which is symmetrized to $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1) : \tilde{U}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \to \left(Z^{[N]}, \tilde{U}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \oplus U_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c}\right)$ in the same fashion as in Lemma 4.3.4. Recall that the randomness-induced channel of $\tilde{W}_1$ defined in [13] can be denoted as $\mathcal{Q}_N(\tilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1) : \tilde{U}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \to \left(Z^{[N]}, \tilde{X}_1^{[N]} \oplus X_1^{[N]}\right)$. Note that for the randomness-induced channel $\mathcal{Q}_N(\tilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$, set $\mathcal{R}_1 \cup \mathcal{S}_1$ is fed with

uniformly random bits, which is different from the shaping-induced channel.

**Lemma 5.3.4.** *For an asymmetric channel $W_1 : X_1 \to Z$ and its symmetrized channel $\tilde{W}_1 : \tilde{X}_1 \to (Z, \tilde{X}_1 \oplus X_1)$, the symmetrized version of the new induced channel $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$ is degraded with respect to the randomness-induced channel $\mathcal{Q}_N(\tilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$.*

*Proof.* The proof is similar to that of Lemma 5.3.3. For a fixed realization $x_1^{[N]}$ and input $\tilde{u}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c}$, there are $2^{|\mathcal{S}_1 \cup \mathcal{R}_1|}$ choice of $\tilde{x}_1^{[N]}$ remaining. Since $z^{[N]}$ is only dependent on $x_1^{[N]}$, we can build a middle channel which merges the $2^{|\mathcal{S}_1 \cup \mathcal{R}_1|}$ output symbols of $\mathcal{Q}_N(\tilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$ to one output symbol of $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$, which means that $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$ is degraded with respect to $\mathcal{Q}_N(\tilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$. Again, this result can be generalized to higher levels. □

Finally, we are ready to prove the semantic security of our wiretap coding scheme. For brevity, let $M_\ell F_\ell$ and $\tilde{M}_\ell \tilde{F}_\ell$ denote $U_\ell^{(\mathcal{S}_\ell \cup \mathcal{R}_\ell)^c}$ and $\tilde{U}_\ell^{(\mathcal{S}_\ell \cup \mathcal{R}_\ell)^c}$, respectively. Recall that $M$ is divided into $M_1, ..., M_r$ at each level. We express $MF$ and $\tilde{M}\tilde{F}$ as the collection of message and frozen bits on all levels of the new induced channel and the symmetric randomness-induced channel, respectively. We also define $\tilde{M}\tilde{F} \oplus MF$ as the operation $\tilde{M}_\ell \tilde{F}_\ell \oplus M_\ell F_\ell$ from level 1 to level $r$.

**Theorem 5.3.5** (Semantic security)**.** *For arbitrarily distributed message $M$, the information leakage $I(M; Z^{[N]})$ of the proposed wiretap lattice code is upper-bounded as*

$$I(M; Z^{[N]}) \le I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF) \le rN2^{-N^{\beta'}}, \tag{5.29}$$

*where $I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF)$ is the capacity of the symmetrized channel derived from the non-binary channel $MF \to Z^{[N]}$* [8].

[8] The symmetrization of a non-binary channel is similar to that of a binary channel as shown in Lemma 4.3.4. When $X$ and $\tilde{X}$ are both non-binary, $X \oplus \tilde{X}$ denotes the result of the exclusive or (xor) operation of the binary expressions of $X$ and $\tilde{X}$.

*Proof.* By [13, Proposition 16], the channel capacity of the randomness-induced channel $\mathcal{Q}_N(\tilde{W}_1, \mathcal{S}_1, \mathcal{R}_1)$ is upper-bounded by $N2^{-N^{\beta'}}$ when partition rule (2.25) is used. By channel degradation, the channel capacity of the symmetrized new induced channel $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$ can also be upper-bounded by $N2^{-N^{\beta'}}$. Since this result can be generalized to higher level $\ell$ ($\ell \geq 1$), we obtain $C(\tilde{\mathcal{Q}}_N(W_\ell, \mathcal{S}_\ell, \mathcal{R}_\ell)) \leq N2^{-N^{\beta'}}$, which means $I(\tilde{M}_\ell\tilde{F}_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}, \tilde{M}_\ell\tilde{F}_\ell \oplus M_\ell F_\ell) \leq N2^{-N^{\beta'}}$. Similarly to (5.11), we have

$$
\begin{aligned}
&I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF) \\
&= \sum_{\ell=1}^{r} I(\tilde{M}_\ell\tilde{F}_\ell; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF | \tilde{M}_{1:\ell-1}\tilde{F}_{1:\ell-1}) \\
&= \sum_{\ell=1}^{r} H(\tilde{M}_\ell\tilde{F}_\ell | \tilde{M}_{1:\ell-1}\tilde{F}_{1:\ell-1}) - H(\tilde{M}_\ell\tilde{F}_\ell | Z^{[N]}, \tilde{M}\tilde{F} \oplus MF, \tilde{M}_{1:\ell-1}\tilde{F}_{1:\ell-1}) \\
&\leq \sum_{\ell=1}^{r} H(\tilde{M}_\ell\tilde{F}_\ell) - H(\tilde{M}_\ell\tilde{F}_\ell | Z^{[N]}, \tilde{M}\tilde{F} \oplus MF, \tilde{M}_{1:\ell-1}\tilde{F}_{1:\ell-1}) \\
&= \sum_{\ell=1}^{r} I(\tilde{M}_\ell\tilde{F}_\ell; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF, \tilde{M}_{1:\ell-1}\tilde{F}_{1:\ell-1}) \\
&\overset{(a)}{=} \sum_{\ell=1}^{r} I(\tilde{M}_\ell\tilde{F}_\ell; Z^{[N]}, M_{1:\ell-1}F_{1:\ell-1}, \tilde{M}_\ell\tilde{F}_\ell \oplus M_\ell F_\ell) \\
&\overset{(b)}{\leq} \sum_{\ell=1}^{r} I(\tilde{M}_\ell\tilde{F}_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}, \tilde{M}_\ell\tilde{F}_\ell \oplus M_\ell F_\ell) \\
&\leq rN2^{-N^{\beta'}},
\end{aligned}
\tag{5.30}
$$

where equality $(a)$ holds because $Z^{[N]}$ is determined by $MF\mathsf{R}$ and $\tilde{M}_\ell\tilde{F}_\ell$ is independent of $\tilde{M}_{\ell+1:r}\tilde{F}_{\ell+1:r} \oplus M_{\ell+1:r}F_{\ell+1:r}$, and inequality $(b)$ holds because adding more variables will not decrease the mutual information.

Therefore, we have

$$\begin{aligned} I(M; Z^{[N]}) &\leq I(MF; Z^{[N]}) \\ &\overset{(a)}{\leq} H(\tilde{M}\tilde{F} \oplus MF) - H(MF) + I(MF; Z^{[N]}) \\ &\overset{(b)}{=} I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF) \\ &\leq rN2^{-N^{\beta'}}, \end{aligned} \tag{5.31}$$

where the equality in $(a)$ holds iff $MF$ is also uniform, and $(b)$ is due to the chain rule. □

### 5.3.5 Reliability over GWC

The reliability analysis in Sect. 5.2.2 holds for the wiretap coding without shaping. When shaping is involved, the problematic set $\mathcal{D}_\ell$ at each level is included in the shaping set $\mathcal{S}_\ell$ and hence determined by the random mapping $\Phi_{\mathcal{S}_\ell}$. In this subsection, we propose two decoders to achieve reliability for the shaping case. The first one requires a private link between Alice and Bob to share the random mapping $\Phi_{\mathcal{S}_\ell}$ and the second one uses the Markov block coding technique [44] without sharing the random mapping. Note that in [21], the message is simply recovered by decoding the fine lattice $\Lambda_b$. When instantiated with a polar lattice, the existence of problematic set $\mathcal{D}_\ell$ does not permit decoding in this straightforward way. Yet, this is only a limitation of SC decoding, not that of the proposed coding scheme.

**Decoder 1:** If $\Phi_{\mathcal{S}_\ell}$ is secretly shared between Alice and Bob, the bits in $\mathcal{D}_\ell$ can be recovered by Bob simply by the shared mapping but not requiring the Markov block coding technique. By Theorem 4.3.9, the reliability at each level can be guaranteed by uniformly distributed independent frozen bits and a random mapping $\Phi_{\mathcal{S}_\ell}$ according to $P_{U_\ell^i|U_\ell^{1:i-1},X_{1:\ell-1}^{[N]}}$ at each level. The decoding rule is given as follows.

- Decoding: The decoder receives $y^{[N]}$ and estimates $\hat{u}_\ell^{[N]}$ based on the previ-

ously recovered $x_{1:\ell-1}^{[N]}$ according to the rule

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

Note that probability $P_{U_\ell^i|U_\ell^{1:i-1},X_{1:\ell-1}^{[N]},Y^{[N]}}(u|\hat{u}_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}, y^{[N]})$ can be calculated by the SC decoding algorithm efficiently, treating $Y$ and $X_{1:\ell-1}$ (already decoded by the SC decoder at previous levels) as the outputs of the asymmetric channel. As a result, the expectation of the decoding error probability over the randomized mappings satisfies $E_{\Phi_{\mathcal{S}_\ell}}[P_e(\phi_{\mathcal{S}_\ell})] = O(2^{-N^{\beta'}})$ for any $\beta' < \beta < 0.5$. [9]

Consequently, by the multilevel decoding and union bound, the expectation of the block error probability of our wiretap coding scheme is vanishing as $N \to \infty$. However, this result is based on the assumption that the mapping $\Phi_{\mathcal{S}_\ell}$ is only shared between Alice and Bob. To share this mapping, we can let Alice and Bob have access to the same source of randomness. This means that we may need a private link between Alice and Bob before the described wiretap coding. Fortunately, the rate of this private link can be made vanishing since the proportion of the shaping bits covered by the mapping $\Phi_{\mathcal{S}_\ell}$ can be significantly reduced.

Recall that the shaping set $\mathcal{S}_\ell$ is defined by

$$\begin{aligned} \mathcal{S}_\ell \triangleq \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta} \text{ or} \\ 2^{-N^\beta} < Z(U_\ell^i|U_\ell^{1:i-1}, Y^{[N]}, X_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}\}. \end{aligned} \tag{5.32}$$

[9] Note that it is possible to derandomize $\Phi_{\mathcal{S}_\ell}$ since there exists a deterministic mapping $\phi_{\mathcal{S}_\ell}$ such that $P_e(\phi_{\mathcal{S}_\ell}) = O(2^{-N^{\beta'}})$. Moveover, this deterministic mapping also guarantees vanishing information leakage because the total variation distance $\|Q_{U_\ell^{[N]},X_{1:\ell-1}^{[N]},Z^{[N]}} - P_{U_\ell^{[N]},X_{1:\ell-1}^{[N]},Z^{[N]}}\|$ can still be bounded by $N2^{-N^{\beta'}}$ as long as it is kept secret.

It has been mentioned in [10] that the shaping bits in the subset

$$\{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) \leq 2^{-N^\beta}\}$$

can be recovered according to the rule

$$u_\ell^i = \arg\max_u P_{U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}}(u | u_\ell^{1:i-1}, x_{1:\ell-1}^{1:N}) \quad \text{if} \quad Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) \leq 2^{-N^\beta},$$

instead of mapping. This modification does not change the result of Theorem 4.3.9 and a proof can be found in [105] and [106]. As a result, the deterministic mapping has only to cover the unpolarized set

$$\begin{aligned} d\mathcal{S}_\ell = \{i \in [N] : 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta} \text{ or } \\ 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, Y^{1:N}, X_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta}\}, \end{aligned} \tag{5.33}$$

whose proportion $\frac{|d\mathcal{S}_\ell|}{N} \to 0$ as $N \to \infty$.

*Remark* 5.3.2. By the channel equivalence, when $\Phi_{\mathcal{S}_\ell}$ is shared to Bob, the decoding of $\Lambda_b$ is equivalent to the MMSE lattice decoding proposed in [21] for random lattice codes. When instantiated with a polar lattice, we use multistage lattice decoding. More explicitly, by Lemma 4.3.6, the SC decoding of the asymmetric channel can be converted to the SC decoding of its symmetrized channel, which is equivalent to the MMSE-scaled partition channel in the lattice Gaussian shaping case Lemma 4.3.11.

**Decoder 2:** Alternatively, one can also use the block Markov coding technique [44] to achieve reliability without sharing $\Phi_{\mathcal{S}_\ell}$. As shown in Fig. 5.7, the message at $\ell$-th level is divided into $k_\ell$ blocks. The shaping bits $S_\ell$ for each block is further divided into unpolarized bits $\Delta S_\ell$ and polarized shaping bits $S_\ell \setminus \Delta S_\ell$. As mentioned above, only $\Delta S_\ell$ needs to be covered by mapping and its proportion is vanishing. We

can sacrifice some message bits to convey $\Delta S_\ell$ for the next block without involving significant rate loss. These wasted message bits are denoted by $E_\ell$. For encoding, we start with the last block (Block $k_\ell$). Given $F_\ell$, $M_\ell$ (no $E_\ell$ for the last block) and $R_\ell$, we can obtain $\Delta S_\ell$ according to $\phi_{\mathcal{S}_\ell}$. Then we copy $\Delta S_\ell$ of the last block to the bits $E_\ell$ of its previous block and do encoding to get the $\Delta S_\ell$ of block $k_\ell - 1$. This process ends until we get the $\Delta S_\ell$ of the first block. This scheme is similar with the one we discussed in Sect. 5.2.2. To achieve reliability, we need a secure code with vanishing rate to convey the bits $\Delta S_\ell$ of the first block to Bob. See [107] for an example of such codes. To guarantee an insignificant rate loss, $k_\ell$ is required to be sufficiently large. We may set $k_\ell = O(N^\alpha)$ for some $\alpha > 0$.

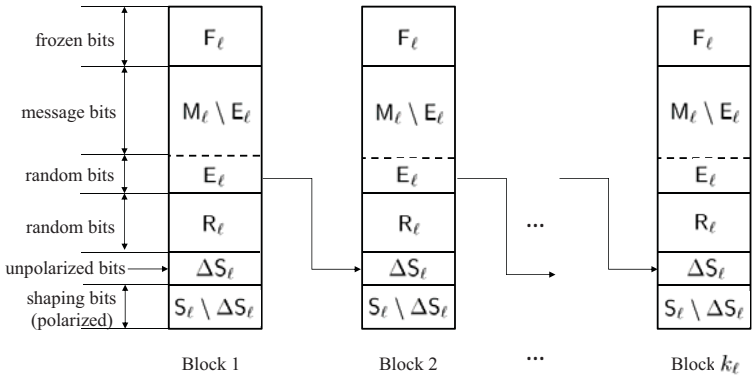



Figure 5.7: Markov block coding scheme without sharing the secret mapping.

We then show that the reliability condition can be satisfied for arbitrarily distributed message bits. By Theorem 4.3.9, we already know that when the frozen bits and information bits are uniformly selected and the shaping bits are selected according to $P_{U_\ell^i|U_\ell^{1:i-1},X_{1:\ell-1}^{[N]}}$, the average block error probability at each level $E_{\Phi_{\mathcal{S}_\ell}}[P_e(\Phi_{\mathcal{S}_\ell})] = O(2^{-N^{\beta'}})$. Note that $E_{\Phi_{\mathcal{S}_\ell}}[P_e(\Phi_{\mathcal{S}_\ell})]$ is the average block error probability over all uniformly distributed frozen bits and information bits. Taking the first level as an example, let $\mathcal{E}_i$ denote the set of pairs of $u_1^{[N]}$ and $y^{[N]}$ such that decoding error occurs at the $i$-th bit, then the block decoding error event is given by $\mathcal{E} \equiv \bigcup_{i\in\mathcal{I}_1} \mathcal{E}_i$. According to our encoding scheme, each codeword $u_1^{[N]}$ occurs with

probability

$$2^{-(|\mathcal{I}_1|+|\mathcal{F}_1|)} \prod_{i\in\mathcal{S}_1} P_{U_1^i|U_1^{1:i-1}}(u_1^i|u_1^{1:i-1}).$$

Then the expectation of decoding error probability over all random mappings is expressed as

$$\begin{aligned} E_{\Phi_{\mathcal{S}_1}}[P_e(\Phi_{\mathcal{S}_1})] = \sum_{u_1^{[N]},y^{[N]}} & 2^{-(|\mathcal{I}_1|+|\mathcal{F}_1|)}\Big(\prod_{i\in\mathcal{S}_1} P_{U_1^i|U_1^{1:i-1}}(u_1^i|u_1^{1:i-1})\Big) \\ & \cdot P_{Y^{[N]}|U_1^{[N]}}(y^{[N]}|u_1^{[N]})\mathbb{1}[(u_1^{[N]},y^{[N]})\in\mathcal{E}]. \end{aligned} \tag{5.34}$$

Therefore, there exists at least one mapping $\phi_{\mathcal{S}_1}$ and a constant $\delta$ such that $P_e(\phi_{\mathcal{S}_1}) = \delta 2^{-N^{\beta'}}$. For this good mapping $\phi_{\mathcal{S}_1}$, the probability of each codeword $u_1^{[N]}$ becomes

$$2^{-(|\mathcal{I}_1|+|\mathcal{F}_1|)}\mathbb{1}\big[\bigcap_{i\in\mathcal{S}_1}\phi_i(u_1^{1:i-1}) = u_1^i\big].$$

For a specific choice of message bits and frozen bits $u_1^{\mathcal{I}_1\cup\mathcal{F}_1}$, define a random variable $\mathfrak{X}$ as

$$\sum_{u_1^{\mathcal{S}_1},y^{[N]}} \mathbb{1}\big[\bigcap_{i\in\mathcal{S}_1}\phi_i(u_1^{1:i-1}) = u_1^i\big]\cdot P_{Y^{[N]}|U_1^{[N]}}(y^{[N]}|u_1^{[N]})\mathbb{1}[(u_1^{[N]},y^{[N]})\in\mathcal{E}].$$

Observe that $0 \le \mathfrak{X} \le 1$ and its expectation over all choices of message bits and frozen bits $E[\mathfrak{X}] = P_e(\phi_{\mathcal{S}_1}) = \delta 2^{-N^{\beta'}}$. Moreover, the variance of $\mathfrak{X}$, denoted by $\sigma^2(\mathfrak{X})$, satisfies $\sigma^2(\mathfrak{X}) \le E[\mathfrak{X}^2] \le E[\mathfrak{X}]$. By the Chebyshev inequality,

$$\Pr(|\mathfrak{X} - \delta 2^{-N^{\beta'}}| \ge 2^{-N^{\beta''}}) \le \frac{\sigma^2(\mathfrak{X})}{2^{-2N^{\beta''}}} \le \delta 2^{-(N^{\beta'}-2N^{\beta''})}.$$

For a sufficient large $N$ and $\beta'' < \beta'$, the probability $\Pr(\mathfrak{X} \ge \delta 2^{-N^{\beta'}} + 2^{-N^{\beta''}}) \to 0$. Therefore, for arbitrarily distributed frozen bits and information bits, the block error

probability under SC decoding is small than $\delta 2^{-N^{\beta'}} + 2^{-N^{\beta''}}$ with probability almost 1. Clearly, $\mathcal{A}_1 \subset \mathcal{I}_1$, meaning that the message bits could be selected according to arbitrary distribution. Note that the above analysis can be generalized to the $\ell$-th level for any $\ell \leq r$.

Now we present the main theorem of this chapter.

**Theorem 5.3.6** (Achieving secrecy capacity of the GWC)**.** *Consider a multilevel lattice code constructed from polar codes based on asymmetric channels and lattice Gaussian shaping $D_{\Lambda,\sigma_s}$. Given $\sigma_e^2 > \sigma_b^2$, let $\epsilon_\Lambda(\tilde{\sigma}_e)$ be negligible and set the number of levels $r = O(\log\log N)$ for $N \to \infty$. Then 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 under semantic security, where $\mathsf{SNR}_b$ and $\mathsf{SNR}_e$ denote the $\mathsf{SNR}$ of the main channel and wiretapper's channel, respectively.*

*Proof.* The reliability condition and the strong secrecy condition are satisfied by Theorem 4.3.9 and Lemma 5.3.2, respectively. It remains to illustrate that the secrecy rate approaches the secrecy capacity. For some $\epsilon' \to 0$, we have

$$\begin{aligned}
\lim_{N\to\infty} R &= \sum_{\ell=1}^{r} \lim_{N\to\infty} \frac{|\mathcal{A}_\ell^{\mathcal{S}^c}|}{N} \\
&= \sum_{\ell=1}^{r} I(X_\ell; Y|X_1,\cdots,X_{\ell-1}) - I(X_\ell; Z|X_1,\cdots,X_{\ell-1}) \\
&\overset{(a)}{=} \frac{1}{2}\log\left(\frac{\tilde{\sigma}_e^2}{\tilde{\sigma}_b^2}\right) - \epsilon' \\
&\overset{(b)}{\geq} \frac{1}{2}\log\left(\frac{1+\mathsf{SNR}_b}{1+\mathsf{SNR}_e}\right) - \epsilon',
\end{aligned} \tag{5.35}$$

where $(a)$ is due to Lemma 5.3.1, and $(b)$ is because the signal power $P_s \leq \sigma_s^2$ [20, Lemma 1][10], respectively. □

[10] Of course, $R$ cannot exceed the secrecy capacity, so this inequality implies that $P_s$ is very close to $\sigma_s^2$.

## 5.4 Summary

We would like to elucidate our coding scheme for the Gaussian wiretap channel in terms of the lattice structure. In Sect. 5.2, we constructed the AWGN-good lattice $\Lambda_b$ and the secrecy-good lattice $\Lambda_e$ without considering the power constraint. When the power constraint is taken into consideration, the lattice Gaussian shaping was implemented in Sect. 5.3. $\Lambda_b$ and $\Lambda_e$ were then constructed according to the MMSE-scaled main channel and wiretapper's channel, respectively. We note that these two lattices themselves are generated only if the independent frozen bits on all levels are $0$s. Since the independent frozen set of the polar codes at each level is filled with random bits, we actually obtain a coset $\Lambda_b + \chi$ of $\Lambda_b$ and a coset $\Lambda_e + \chi$ of $\Lambda_e$ simultaneously, where $\chi$ is a uniformly distributed shift. This is because we can not fix the independent frozen bits $F_\ell$ in our scheme (due to the lack of the proof that the shaping-induced channel is symmetric). By using the lattice Gaussian $D_{\Lambda,\sigma_s}$ as our constellation in each lattice dimension, we would obtain $D_{\Lambda^N,\sigma_s}$ without coding. Since $\Lambda_e+\chi \subset \Lambda_b+\chi \subset \Lambda^N$, we actually implemented the lattice Gaussian shaping over both $\Lambda_b + \chi$ and $\Lambda_e + \chi$. To summarize our coding scheme, Alice firstly assigns each message $m \in \mathcal{M}$ to a coset $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$, then randomly sends a point in the coset $\Lambda_e + \chi + \lambda_m$ ($\lambda_m$ is the coset leader of $\tilde{\lambda}_m$) according to the distribution $D_{\Lambda_e+\chi+\lambda_m,\sigma_s}$ via the shaping operation. This scheme is consistent with the theoretical model proposed in [21].

For semantic security, a symmetrized new induced channel from $\tilde{M}\tilde{F}$ to $(Z^{[N]}, \tilde{M}\tilde{F} \oplus MF)$ was constructed to upper-bound the information leakage. This channel is directly derived from the new induced channel from $MF$ to $Z^{[N]}$. According to Lemma 5.3.3, this symmetrized new induced channel is degraded with respect to the symmetric randomness-induced channel from $\tilde{M}\tilde{F}$ to $(Z^{[N]}, \tilde{X}_{1:r}^{[N]} \oplus X_{1:r}^{[N]})$. Moreover, when $\tilde{F}$ is frozen, the randomness-induced channel from $\tilde{M}$ to $(Z^{[N]}, \tilde{X}_{1:r}^{[N]} \oplus X_{1:r}^{[N]})$ corresponds to the $\Lambda_b/\Lambda_e$ channel given in Sect. 5.2 (with

MMSE scaling).

CHAPTER 6

# Polar Lattices for Quantization

## 6.1 Introduction

In Chapter 4, polar lattices have been proved to be able to achieve the capacity of AWGN channels. In this chapter, we propose a new construction of polar lattices to solve the dual problem, i.e., achieving the rate-distortion bound of a memoryless Gaussian source, which means that polar lattices can also be good for the lossy compression of continuous sources. The structure of the proposed polar lattices enables us to integrate the post-entropy coding process into the lattice quantizer, which simplifies the quantization process. The overall complexity of encoding and decoding process is $O(N \log^2 N)$ for a sub-exponentially decaying excess distortion. Moreover, the nesting structure of polar lattices further provides solutions for some multi-terminal coding problems. The Wyner-Ziv coding problem for a Gaussian source can be solved by an AWGN capacity-achieving polar lattice nested in a rate-distortion bound achieving one, and the Gelfand-Pinsker problem can be solved in a reversed manner.

Vector quantization (VQ) [108] has been widely used for source coding of image and speech data since the 1980s. Compared with scalar quantization, the advantage of VQ, guaranteed by Shannon's rate-distortion theory, is that better performance can always be achieved by coding vectors instead of scalars, even in the case of memoryless sources. However, the Shannon theory does not provide us any constructive VQ design scheme. During the past several decades, many practical VQ techniques with

relatively low complexity have been proposed, such as lattice VQ [25], multistage VQ [109], tree-structured VQ [110], gain-shape VQ [111], etc. Among them, lattice VQ is of particular interest because its highly regular structure makes compact storage and fast quantization possible.

We present an explicit construction of polar lattices for quantization, which achieves the rate-distortion bound of the continuous Gaussian source. It is well known that the optimal output alphabet size is infinite for continuous-amplitude sources. Particularly, the rate distortion function for the Gaussian source of variance $\sigma_s^2$ under the squared-error distortion measure $d(x,y) = \|x-y\|^2$ is given by

$$R(\Delta) = \max\left\{\frac{1}{2}\log\left(\frac{\sigma_s^2}{\Delta}\right), 0\right\}, \tag{6.1}$$

where $\Delta$ and $R$ denote the average distortion and rate per symbol, respectively. However, in practice, the size of the reconstruction alphabet needs to be finite. Unconstructively, [112, Theorem 9.6.2] shows the existence of a block code with finite number of output letters that achieves performance arbitrarily close to the rate-distortion bound. Then a size-constrained output alphabet rate-distortion function $R_M(\Delta)$ was defined in [113] with $M$ denoting the size of the output alphabet. The well-known trellis coded quantization (TCQ) [114] was motivated by this alphabet constrained rate-distortion theory. It was shown that for a given encoding rate of $R$ bits per symbol, the rate-distortion function $R(\Delta)$ can be approached by using a TCQ encoder with rate $R+1$ after an initial Lloyd-Max quantization. It is equivalent to the trellis coded modulation (TCM) in the sense that $m$ information bits are transmitted using $2^{m+1}$ constellation points. A near-optimum lattice quantization scheme based on tailbiting convolutional codes was introduced in [115]. Despite enjoying a good practical performance, a theoretical proof of the rate-distortion bound achiev-

ing TCQ with low complexity is still missing. More recently, a scheme based on low density Construction-A (LDA) lattices [116] was proved to be quantization-good (defined in Sect. 6.2) using the minimum-distance lattice decoder. However, in practice the ideal performance cannot be realized by the suboptimal belief-propagation decoding algorithm.

Polar lattices have the potential in solving this problem with low complexity. As shown in Chapter 4, this class of lattices allows us to employ the discrete Gaussian distribution for lattice shaping. This distribution shares many similar properties to the continuous Gaussian distribution and obtains the optimal shaping gain when its associated flatness factor is negligible. We may use the discrete Gaussian distribution instead of the continuous one as the distribution of the reconstruction alphabet. This idea has already been proposed in [117] for random lattice quantization. It is also shown in [104] that even using binary lattice partition, the number of the partition levels $r$ does not need to be very large ($O(\log\log N)$) to achieve the capacity $\frac{1}{2}\log(1+\mathsf{SNR})$ of the additive white Gaussian noise (AWGN) channel, where $\mathsf{SNR}$ denotes the signal noise ratio. By the duality between source coding and channel coding, the quantization lattices can be roughly viewed as a channel coding lattice constructed on the test channel. For a Gaussian source with variance $\sigma_s^2$ and an average distortion $\Delta$, the test channel is actually an AWGN channel with noise variance $\Delta$. In this case, the "$SNR$" of the test channel is $\frac{\sigma_s^2-\Delta}{\Delta}$, and its "capacity" is $\frac{1}{2}\log(\frac{\sigma_s^2}{\Delta})$, which implies that the rate of the polar lattice quantizer can be made arbitrarily close to $\frac{1}{2}\log(\frac{\sigma_s^2}{\Delta})$. Therefore, based on this idea, we propose the construction of polar lattices which are good for quantization in this work. We note that the difference between the quantization polar lattices and the AWGN channel coding polar lattices not only lies in the construction of their component polar codes, but also in the role of their associate flatness factors. For the AWGN channel coding polar lattices, the flatness factor is required to be negligible to ensure a coding rate

close to the AWGN capacity and it has no influence on the error correction performance. For the quantization polar lattices, however, the flatness factor affects both the compression rate and the distortion performance. This is also the reason why the lattice Gaussian distribution can be optimal for both channel coding and quantization simultaneously (see Remark 6.2.1), and consequently be utilized for Gaussian Wyner-Ziv and Gelfand-Pinsker coding.

The novel technical ingredients of this work are the following:

- The construction of polar lattices for the Gaussian source and the proof of their rate-distortion bound achieving. This is a dual work of capacity-achieving polar lattices for the AWGN channel, and it can also be considered as an extension of binary polar lossy coding to the multilevel coding scenario. Compared with traditional lattice quantization schemes [118, 1], which generally require a separate entropy encoding process after obtaining the quantized lattice points, our scheme naturally integrates these two processes together. The analysis of these quantization polar lattices prepares us for the further discussion of Gaussian Wyner-Ziv and Gelfand-Pinsker problems.

- The solutions of the Gaussian Wyner-Ziv and Gelfand-Pinsker problems, which consist of two nested polar lattices. One is AWGN capacity-achieving and the other is Gaussian rate-distortion bound achieving. The two lattices are simultaneously shaped according to a proper lattice Gaussian distribution. Note that the Wyner-Ziv and Gelfand-Pinsker problems for the binary case have been solved by Korada and Urbanke [9] using nested polar codes. However, in the Gaussian case, the problems turn out to be more complicated as the Wyner-Ziv bound becomes lower (the Gelfand-Pinsker capacity becomes larger by duality). As mentioned in [119], the extremely severe conditions [119, eq. (12)] and [119, eq. (18)] for the bound, which corresponds to the scenario where both encoder and decoder know the side information, can be satisfied in

the Gaussian case rather than the binary case because of infinite alphabet size, meaning that more effort should be made for the Gaussian case. As a result, our polar lattice coding scheme achieves the whole region of the Wyner-Ziv bound and has no requirement on the signal noise ratio for the Gelfand-Pinsker capacity.

As mentioned above, although the TCQ technique performs well in practice, its theoretical limit is still unclear, to the best of our knowledge. Polar lattices, as we will see, can be theoretically proved to be able to achieve the rate-distortion bound. Moreover, thanks to their low complexity, considerably high-dimensional polar lattices are available in practice, providing a quantization performance with gap less than $0.2$ dB to the achievable bound when the lattice dimension $N = 2^{18}$.

The sparse regression codes were also proved to achieve the the optimal rate-distortion bound of i.i.d Gaussian sources with polynomial complexity [120, 121]. In fact, there exists a trade-off between the distortion performance and encoding complexity. For a block length $N$, typical encoding complexity of this kind of codes is $O((N/\log N)^2)$ for an exponentially decaying excess distortion with exponent $O(N/\log N)$, and their designed random matrix incurs $N \times O(N^2)$ storage complexity. In comparison, the construction of polar lattices is as explicit as that of polar codes themselves, and the complexity is quasi-linear $O(N \log^2 N)$ for a sub-exponentially decaying excess distortion with exponent roughly $O(\sqrt{N})$.

The saliently nesting structure of polar lattices also gives us solutions to the Gaussian Wyner-Ziv and Gelfand-Pinsker problems. According to the prior work by Zamir, Shamai and Erez [122, 123], the two problems can be solved by nested quantization-good and AWGN-good lattices. However, due to the lack of explicit construction of such good lattices, no explicit solution was addressed. A practical scheme based on multidimensional nested lattice codes for the Gaussian Wyner-Ziv problem was also proposed in [124]. The performance of this scheme can be very

close to the Wyner-Ziv bound but a theoretical proof is still missing. A lattice-based Gelfand-Pinsker coding scheme using repeat-accumulate codes, which were concatenated with trellis shaping, was also presented in [125]. This scheme was shown to be able to obtain a very close-to-capacity performance. Unfortunately, the complexity grows exponentially to achieve the shaping gain and a theoretical proof for the Gelfand-Pinsker capacity-achieving is also missing. In this work, we solve these problems by combining the AWGN capacity achieving polar lattices proposed in Chapter 4 and the rate-distortion bound achieving ones.

## 6.2 Polar Lattices for Gaussian Sources

Let $Y \sim N(0, \sigma_s^2)$ denote a one dimensional Gaussian source with zero mean and variance $\sigma_s^2$. Let $Y^{1:N}$ ($\underline{Y}$) be $N$ independent copies of $Y$ and $y^{1:N}$ ($\underline{y}$) be a realization of $Y^{1:N}$. The PDF of $\underline{Y}$ is given by $f_{\underline{Y}}(\underline{y}) = f_{\sigma_s}(\underline{y})$. For an $N$-dimensional polar lattice $L$ and its associated quantizer $Q_L(\cdot)$, the average distortion $\Delta$ after quantization is given by

$$\Delta = \frac{1}{N} \int_{\mathbb{R}^N} \| \underline{y} - Q_L(\underline{y}) \|^2 \, f_{\underline{Y}}(\underline{y}) d\underline{y}. \tag{6.2}$$

The normalized second moment (NSM) of a quantization lattice $L$ is defined as

$$G(L) = \frac{\frac{1}{N} \int_{\mathcal{V}(L)} \|v\|^2 dv}{V(L)^{1+2/N}}, \tag{6.3}$$

where vector $v$ is uniformly distributed in $\mathcal{V}(L)$.

**Definition 6.2.1.** An $N$-dimensional lattice $L$ is called quantization-good [118] if

$$\lim_{N \to \infty} G(L) = \frac{1}{2\pi e}. \tag{6.4}$$

In [1], an entropy-coded dithered quantization (ECDQ) scheme based on quantization-good lattices was proposed to achieve the rate-distortion bound (6.1). This scheme requires a pre-shared dither which is uniformly distributed in the Voronoi region of a quantization-good lattice and an entropy encoder after lattice quantization. For our quantization scheme, we will show that dither is not necessary and the entropy encoder can be integrated in the lattice quantization process, which brings much convenience for practical application.

Our task is to construct a polar lattice which achieves the rate distortion bound of the Gaussian source with reconstruction distribution $D_{\Lambda,\sqrt{\sigma_s^2-\Delta}}$. Following the notation of AWGN-good polar lattices in Chapter 4, we use $X$ to denote the reconstruction alphabet. Firstly, we prove that the rate achieved by $D_{\Lambda,\sqrt{\sigma_s^2-\Delta}}$ can be arbitrarily close to $\frac{1}{2}\log(\frac{\sigma_s^2}{\Delta})$. Note that the following theorem is essentially the same as Theorem 3.2.1. Here we just reexpress it in the source coding formulation.

**Theorem 6.2.1** ([20]). *Consider a Gaussian test channel where the reconstruction constellation $X$ has a discrete Gaussian distribution $D_{\Lambda-\mathrm{c},\sigma_r}$ for arbitrary $\mathrm{c}\in\mathbb{R}^n$, and where $\sigma_r^2=\sigma_s^2-\Delta$ with $\Delta$ being the average distortion. Let $\tilde{\sigma}_\Delta \triangleq \frac{\sigma_r\sqrt{\Delta}}{\sigma_s}$. Then, if $\epsilon=\epsilon_\Lambda(\tilde{\sigma}_\Delta)<\frac{1}{2}$ and $\frac{\pi\epsilon_t}{1-\epsilon_t}<\epsilon$ where*

$$\epsilon_t \triangleq \begin{cases} \epsilon_\Lambda(\sigma_r/\sqrt{\frac{\pi}{\pi-t}}), & t\geq 1/e \\ (t^{-4}+1)\epsilon_\Lambda(\sigma_r/\sqrt{\frac{\pi}{\pi-t}}), & 0<t<1/e \end{cases} \tag{6.5}$$

*the discrete Gaussian constellation results in mutual information $I_\Delta \geq \frac{1}{2}\log(\frac{\sigma_s^2}{\Delta})-\frac{5\epsilon}{n}$ per channel use.*

The statement of Theorem 6.2.1 is non-asymptotical, i.e., it can hold even if $n=1$. Therefore, it is possible to construct a good polar lattice over one-dimensional lattice partition such as $\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z}$.... The flatness factor $\epsilon$ can be made negligible by scaling this binary partition. This technique has already been used to construct

AWGN capacity-achieving polar lattices in Chapter 4.

Note that when the test channel is chosen to be an AWGN channel with noise variance $\Delta$ and the reconstruction alphabet is discrete Gaussian distributed, the source distribution is not exactly a continuous Gaussian distribution. In fact, it is a distribution obtained by adding a continuous Gaussian of variance $\Delta$ to a discrete Gaussian $D_{\Lambda-c,\sigma_r}$, which is expressed as the following convolution

$$f_{Y'}(y') = \frac{1}{f_{\sigma_r}(\Lambda - c)} \sum_{\lambda \in \Lambda - c} f_{\sigma_r}(\lambda) f_{\sigma}(y' - \lambda), y' \in \mathbb{R}^n, \tag{6.6}$$

where $\sigma = \sqrt{\Delta}$ and $Y'$ denotes the new source. For simplicity, in this work we only consider one dimensional binary partition chain ($n = 1$) and $Y'$ is also a one dimensional source.

Therefore, we are actually quantizing source $Y'$ instead of $Y$ using the discrete Gaussian distribution. However, when the flatness factor $\epsilon_\Lambda(\tilde{\sigma}_\Delta)$ is small, a good quantizer constructed from polar lattices for the source $Y'$ is also good for source $Y$ because of the following lemma. The relationship between the quantization of source $Y'$ and $Y$ is shown in Fig. 6.1.

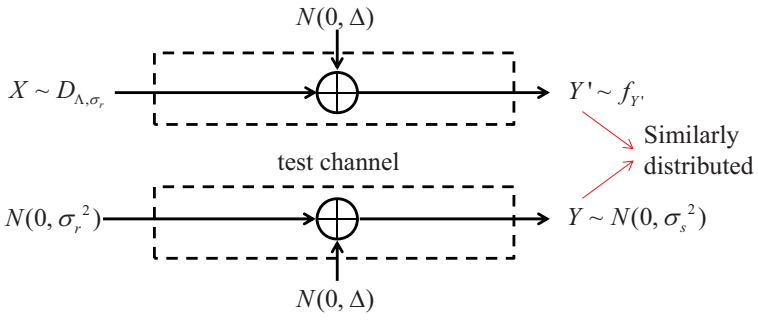



Figure 6.1: The relationship between the quantization of source $Y'$ and $Y$.

**Lemma 6.2.2** ([20]). *If $\epsilon = \epsilon_\Lambda(\tilde{\sigma}_\Delta) < \frac{1}{2}$, the variational distance between the density $f_{Y'}$ of source $Y'$ defined in (6.6) and the Gaussian density $f_Y$ satisfies $\mathbb{V}(f_{Y'}, f_Y) \leq 4\epsilon$.*

Now we consider the construction of polar lattices for quantization. Firstly, we consider the quantization of source $Y'$ using the reconstruction distribution $D_{\Lambda,\sigma_r}$. Since binary partition is used, $X$ can be represented by a binary string $X_{1:r}$, and we have $\lim_{r\to\infty} P_{X_{1:r}} = P_X = D_{\Lambda,\sigma_r}$. Because the polar lattices are constructed by "Construction D", we are interested in the test channel at each level. Similar to the setting of shaping for AWGN-good polar lattices in Section 4.3, given the previous $x_{1:\ell-1}$ and the coset $\mathcal{A}_\ell$ determined by $x_{1:\ell}$, the channel transition PDF at level $\ell$ is

$$\begin{aligned} &P_{Y'|X_\ell,X_{1:\ell-1}}(y'|x_\ell,x_{1:\ell-1}) \\ &= \frac{\sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} P(a)P_{Y'|A}(y'|a)}{P\{\mathcal{A}_\ell(x_{1:\ell})\}} \\ &= \exp\left(-\frac{y'^2}{2(\sigma_s^2+\Delta)}\right)\frac{1}{f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))}\frac{1}{2\pi\sqrt{\Delta}\sigma_s}\sum_{a\in\mathcal{A}_\ell(x_{1:\ell})}\exp\left(-\frac{1}{2\tilde{\sigma}_\Delta^2}\left(|\alpha y'-a|^2\right)\right), \end{aligned} \tag{6.7}$$

where $\alpha = \frac{\sigma_r^2}{\sigma_r^2+\Delta}$ is equal to the MMSE coefficient and $\tilde{\sigma}_\Delta = \frac{\sigma_r\sqrt{\Delta}}{\sigma_s}$. Consequently, using $D_{\Lambda,\sigma_r}$ as the constellation, the $\ell$-th channel is generally asymmetric with the input distribution $P_{X_\ell|X_{1:\ell-1}}$ $(\ell \le r)$, which can be calculated according to the definition of $D_{\Lambda,\sigma_r}$.

The lattice quantization can be viewed as lossy compression for all binary testing channels from level 1 to $r$. Here we start with the first level. Let $y'^{1:N}$ denotes the realization of $N$ i.i.d copies of source $Y'$. Although $Y'$ is a continuous source with density given by (6.6) and $y'$ is drawn from $\mathbb{R}$, to keep the notations consistent (the definition of the Bhattacharyya parameter [4] is given by a summation), from now on we will express the distortion measurement as well as the variantional distance in the form of summation instead of integral.

Since the test channel at each level is not necessarily symmetric and the reconstruction constellation is not uniformly distributed, we have to consider the lossy compression for nonuniform source and asymmetric distortion measure. The solution of this problem has already been introduced in [10], and it turns out to be similar

to the construction of capacity-achieving polar codes for asymmetric channels.

For the first level, letting $U_1^{1:N} = X_1^{1:N} G_N$, where $G_N$ is the $N \times N$ generator matrix of polar codes, we define the information set $\mathcal{I}_1$, frozen set $\mathcal{F}_1$ and shaping set $\mathcal{S}_1$ based on the Bhattacharyya parameter as follows:

$$\begin{cases} \mathcal{F}_1 = \{i \in [N] : Z(U_1^i | U_1^{1:i-1}, Y'^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \mathcal{I}_1 = \{i \in [N] : Z(U_1^i | U_1^{1:i-1}) > 2^{-N^\beta} \text{ and } Z(U_1^i | U_1^{1:i-1}, Y'^{1:N}) < 1 - 2^{-N^\beta}\} \\ \mathcal{S}_1 = \{i \in [N] : Z(U_1^i | U_1^{1:i-1}) \leq 2^{-N^\beta}\}. \end{cases} \tag{6.8}$$

The shaping set $\mathcal{S}_1$ is determined by the distribution $P_{X_1}$. Note that this definition is similar to that in (5.17). The difference is that $\mathcal{I}_1$ is designed to be slightly larger to guarantee a desired distortion level. The asymmetric Bhattacharyya parameter $Z(U_1^i | U_1^{1:i-1}, Y'^{1:N})$ and $Z(U_1^i | U_1^{1:i-1})$ can be efficiently calculated by symmetric Bhattacharyya parameter $\tilde{Z}(\tilde{U}_1^i | \tilde{U}_1^{1:i-1}, X_1^{1:N} \oplus \tilde{X}_1^{1:N}, Y'^{1:N})$ and $\tilde{Z}(\tilde{U}_1^i | \tilde{U}_1^{1:i-1}, X_1^{1:N} \oplus \tilde{X}_1^{1:N})$, respectively (see Section 4.3 for more details). According to Theorem 4.3.7, the proportion of set $\mathcal{I}_1$ approaches $I(X_1; Y')$ when $N \to \infty$.

After getting $\mathcal{F}_1$, $\mathcal{I}_1$ and $\mathcal{S}_1$, for a source sequence $y'^{1:N}$, the encoder determines $u_1^{1:N}$ according to the following rule:

$$u_1^i = \begin{cases} 0 \text{ w. p. } P_{U_1^i | U_1^{1:i-1}, Y'^{1:N}}(0 | u_1^{1:i-1}, y'^{1:N}) \\ 1 \text{ w. p. } P_{U_1^i | U_1^{1:i-1}, Y'^{1:N}}(1 | u_1^{1:i-1}, y'^{1:N}) \end{cases} \text{ if } i \in \mathcal{I}_1, \tag{6.9}$$

and

$$u_1^i = \begin{cases} \bar{u}_1^i & \text{if } \ i \in \mathcal{F}_1 \\ \arg\max\limits_u P_{U_1^i | U_1^{1:i-1}}(u | u_1^{1:i-1}) & \text{if } \ i \in \mathcal{S}_1. \end{cases} \tag{6.10}$$

Here $\bar{u}_1^i$ is a uniformly random bit determined before lossy compression. The output of the encoder at level 1 is $u_1^{\mathcal{I}_1} = \{u_1^i, i \in \mathcal{I}_1\}$. To reconstruct $x_1^{1:N}$,

the decoder uses the shared $u_1^{\mathcal{F}_1}$ and the received $u_1^{\mathcal{I}_1}$ to recover $u_1^{\mathcal{S}_1}$ according to $\arg\max_u P_{U_1^i|U_1^{1:i-1}}(u|u_1^{1:i-1})$ and then $x_1^{1:N} = u_1^{1:N}G_N$. The probability $P_{U_1^i|U_1^{1:i-1}}$ and $P_{U_1^i|U_1^{1:i-1},Y'^{1:N}}$ can both be calculated efficiently by the successive cancellation algorithm with complexity $O(N\log N)$.

**Theorem 6.2.3.** *Let $Q_{U_1^{1:N},Y'^{1:N}}(u_1^{1:N}, y'^{1:N})$ denote the joint distribution for $U_1^{1:N}$ and $Y'^{1:N}$ according to the encoding rule described in (6.9) and (6.10). Consider another encoder using the encoding rule (6.9) for all $i \in [N]$ and let $P_{U_1^{1:N},Y'^{1:N}}(u_1^{1:N}, y'^{1:N})$ denote the resulted joint distribution. For any $\beta' < \beta < 1/2$ satisfying (6.8) and $R_1 = \frac{|\mathcal{I}_1|}{N} > I(X_1;Y')$,*

$$\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) = O(2^{-N^{\beta'}}). \tag{6.11}$$

The same statement has been given in [10] yet without proof. Here we prove the theorem in Appendix F for completeness.

Now we introduce the construction for higher levels. Taking the second level as an example, to make up the reconstruction constellation distribution, the input distribution at level 2 should be $P_{X_2|X_1}$. Based on the quantization results $(U_1^{1:N}, Y'^{1:N})$ given by the encoder at level 1, some $U_2^i$ ($U_2^{1:N} = X_2^{1:N}G_N$) is almost deterministic given $(U_2^{1:i-1}, U_1^{1:N})$. Since there is a one-to-one mapping between $X_1^{1:N}$ and $U_1^{1:N}$, given $(U_2^{1:i-1}, U_1^{1:N})$ is the same as given $(U_2^{1:i-1}, X_1^{1:N})$. We define the information set $\mathcal{I}_2$, frozen set $\mathcal{F}_2$ and shaping set $\mathcal{S}_2$ as follows:

$$\begin{cases} \mathcal{F}_2 = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}, Y'^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \mathcal{I}_2 = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}) > 2^{-N^\beta} \text{ and} \\ \qquad\qquad Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}, Y'^{1:N}) < 1 - 2^{-N^\beta}\} \\ \mathcal{S}_2 = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}) \leq 2^{-N^\beta}\}. \end{cases} \tag{6.12}$$

The proportion of $\mathcal{I}_2$ approaches $I(X_2;Y|X_1)$ when $N$ is sufficiently large [104].

For a given source sequence pair $(u_1^{1:N}, y'^{1:N})$ or $(x_1^{1:N}, y'^{1:N})$, the encoder at level 2 determines $u_2^{1:N}$ according to the following rule:

$$u_2^i = \begin{cases} 0 \text{ w.p. } P_{U_2^i|U_2^{1:i-1},X_1^{1:N},Y'^{1:N}}(0|u_2^{1:i-1}, x_1^{1:N}, y'^{1:N}) \\ 1 \text{ w.p. } P_{U_2^i|U_2^{1:i-1},X_1^{1:N},Y'^{1:N}}(1|u_2^{1:i-1}, x_1^{1:N}, y'^{1:N}) \end{cases} \quad \text{if } i \in \mathcal{I}_2, \tag{6.13}$$

and

$$u_2^i = \begin{cases} \bar{u}_2^i \quad \text{if} \quad i \in \mathcal{F}_2 \\ \arg\max_u P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(u|u_2^{1:i-1}, x_1^{1:N}) \quad \text{if} \quad i \in \mathcal{S}_2. \end{cases} \tag{6.14}$$

We further extend Theorem 6.2.3 to the second level.

**Theorem 6.2.4.** *Let $Q_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}(u_2^{1:N}, u_1^{1:N}, y'^{1:N})$ denote the joint distribution for $U_2^{1:N}$ and $(U_1^{1:N}, Y'^{1:N})$ according to the encoding rule described in (6.13) and (6.14). Consider another encoder using the encoding rule (6.13) for all $i \in [N]$ and let $P_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}(u_2^{1:N}, u_1^{1:N}, y'^{1:N})$ denote the resulted joint distribution. For any $\beta' < \beta < 1/2$ satisfying (6.12) and $R_2 = \frac{|\mathcal{I}_2|}{N} > I(X_2; Y'|X_1)$,*

$$\mathbb{V}(P_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}, Q_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}) = O(2^{-N^{\beta'}}). \tag{6.15}$$

We present the proof of Theorem 6.2.4 in Appendix G.

Note that Theorem 6.2.4 is based on the assumption that $\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) = O(2^{-N^{\beta'}})$, which means that we also need $R_1 > I(X_1; Y')$. Therefore, we have $\sum_{i=1}^2 R_i > I(X_1X_2; Y')$.

By induction, for level $\ell$ ($\ell \leq r$), we define the three sets $\mathcal{F}_\ell$, $\mathcal{I}_\ell$ and $\mathcal{S}_\ell$ in the same form as (6.12) with $X_{1:\ell-1}^{1:N}$ replacing $X_1^{1:N}$ and $U_\ell$ replacing $U_2$. Similarly, the encoder determines $u_\ell^{1:N}(u_\ell^{1:N} = x_\ell^{1:N}G_N)$ according to the rule given by (6.13) and (6.14), with $X_{1:\ell-1}^{1:N}$ and $x_{1:\ell-1}^{1:N}$ replacing $X_1^{1:N}$ and $x_1^{1:N}$, respectively. Let

$Q_{U_{1:\ell}^{1:N},Y'^{1:N}}(u_{1:\ell}^{1:N}, y'^{1:N})$ denote the associate joint distribution resulted from this encoder and $P_{U_{1:\ell}^{1:N},Y'^{1:N}}(u_{1:\ell}^{1:N}, y'^{1:N})$ denote the one that resulted from an encoder only using (6.13) for all $i \in [N]$. We have $\mathbb{V}(P_{U_{1:\ell}^{1:N},Y'^{1:N}}, Q_{U_{1:\ell}^{1:N},Y'^{1:N}}) = O(\ell \cdot 2^{-N^{\beta'}})$ for any rate $R_\ell = \frac{|\mathcal{I}_\ell|}{N} > I(X_\ell; Y'|X_{1:\ell-1})$. Specifically, at level $r$, for any rate $R_r > I(X_r; Y'|X_{1:r-1})$ and $\sum_{i=1}^{r} R_i > I(X_{1:r}; Y')$, we have

$$\mathbb{V}(P_{U_{1:r}^{1:N},Y'^{1:N}}, Q_{U_{1:r}^{1:N},Y'^{1:N}}) = O(r \cdot 2^{-N^{\beta'}}). \tag{6.16}$$

By Lemma 4.3.1, $I(X_{1:r}; Y')$ is arbitrarily close to $I(X; Y')$ when $N$ is sufficiently large and $r = O(\log\log N)$, which gives us $\mathbb{V}(P_{U_{1:r}^{1:N},Y'^{1:N}}, Q_{U_{1:r}^{1:N},Y'^{1:N}}) = O(2^{-N^{\beta'}})$.

Now we present the main theorem of this section. The proof is given in Appendix H.

**Theorem 6.2.5.** *Given a Gaussian source $Y$ with variance $\sigma_s^2$ and an average distortion $\Delta \le \sigma_s^2$, for any rate $R > \frac{1}{2}\log(\frac{\sigma_s^2}{\Delta})$, there exists a multilevel polar code with rate $R$ such that the distortion is arbitrarily close to $\Delta$ when $N \to \infty$ and $r = O(\log\log N)$. This multilevel polar code is actually a shifted polar lattice $L + \mathrm{c}$ constructed from the lattice partition $\Lambda/\Lambda'$ with a shaping according to the discrete Gaussian distribution $D_{\Lambda,\sigma_r}$, where $\sigma_r = \sqrt{\sigma_s^2 - \Delta}$ and the partition chain is scaled to make $\epsilon_\Lambda(\frac{\sigma_r\sqrt{D}}{\sigma_s}) \to 0$.*

*Remark* 6.2.1. From the proof of Theorem 6.2.5, it seems that $R$ could be slightly smaller than $\frac{1}{2}\log\frac{\sigma_s^2}{\Delta}$ (Since $R > I(X; Y') \ge \frac{1}{2}\log\frac{\sigma_s^2}{\Delta} - \frac{5\epsilon_\Lambda(\tilde{\sigma}_\Delta)}{n}$.) to reach an average distortion $\Delta$, which would contradict Shannon's rate-distortion theory. However, this is not the case. When $R < \frac{1}{2}\log\frac{\sigma_s^2}{\Delta}$, an arbitrarily small $\epsilon_\Lambda(\tilde{\sigma}_\Delta)$ cannot be guaranteed, which means that the resulted distortion cannot be arbitrarily close to $\Delta$. To achieve the desired distortion, we need $R > \frac{1}{2}\log\frac{\sigma_s^2}{\Delta} - \frac{5\epsilon_\Lambda(\tilde{\sigma}_\Delta)}{n}$ for all possibly small $\epsilon_\Lambda(\tilde{\sigma}_\Delta)$, which leads to $R > \frac{1}{2}\log\frac{\sigma_s^2}{\Delta}$.

## 6.3 Simulation Results

The quantization performance of polar lattices for a Gaussian source with standard deviation $\sigma_s = 3$ and target distortion from $0.1$ to $2.5$ is shown in Fig. 6.2. It reveals that the rate-distortion bound is approached as the dimension of polar lattices increases from $N = 2^{10}$ to $N = 2^{18}$. Particularly, when $N = 2^{18}$, the gap to the rate-distortion bound is less than $0.2$ dB.

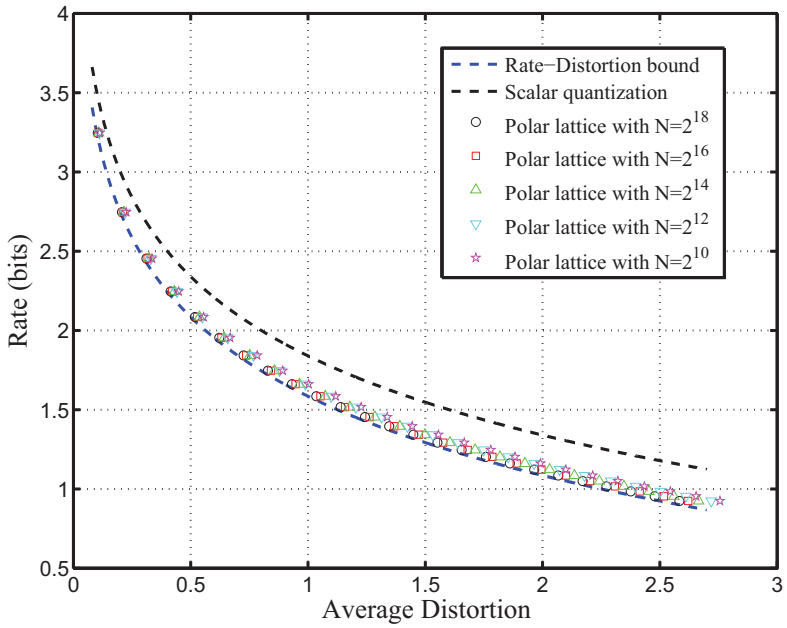
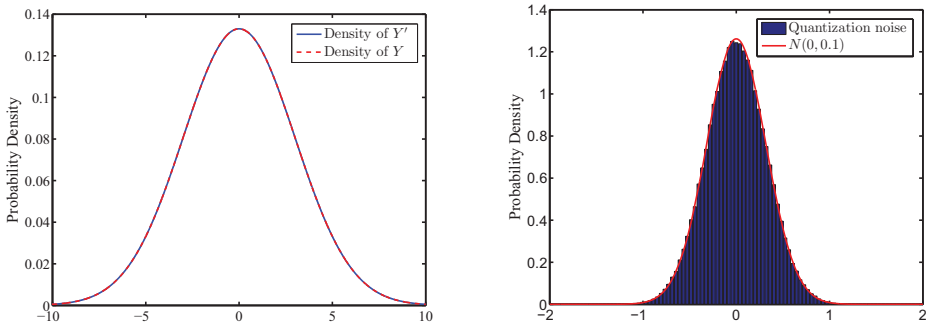



Figure 6.2: Quantization performance of polar lattices for the Gaussian source with $\sigma_s = 3$.

In this work, the number of levels is chosen to be $6$ to guarantee a negligible variational distance $\mathbb{V}(f_{Y'}, f_Y)$ for all target distortions. For a target distortion $\Delta = 0.5$, the two densities of $Y'$ and $Y$ are compared in Fig. 6.3a, where negligible difference between $f_Y$ and $f_{Y'}$ is found since $\mathbb{V}(f_{Y'}, f_Y) \approx 1.1 \times 10^{-7}$. Moreover, the quantization noise behaves similarly to a Gaussian noise as shown in Fig. 6.3b, which will be useful to understand the idea of Gaussian Wyner-Ziv coding and Gelfand-Pinsker coding in the next section.

A performance comparison between the TCQ and polar lattice for quantization




(a) $\mathbb{V}(f_{Y'}, f_Y)$ when $r = 6$ and $\Delta = 0.5$. (b) Quantization noise v.s. Gaussian noise when $\Delta = 0.1$.

Figure 6.3: Variational distance $\mathbb{V}(f_{Y'}, f_Y)$ and quantization noise for polar lattice quantizer.

is shown in Table 6.1. The $\mathsf{SNR} = \frac{\sigma_s^2}{\Delta}$ is shown in dB. For the TCQ, the dimension is $1000$ and the number of states is $256$. For the quantization polar lattice, the dimension is $1024$. It can be observed that the performance of the polar lattice is superior to that of the TCQ with roughly same block length (especially for higher rate). The performance of the Lloyd-Max scalar quantizer is also shown.

Table 6.1: Performance comparison with TCQ for Gaussian source ($\mathsf{SNR}$ in dB).

| Rate (bits) | TCQ | Polar Lattice Quantizer | Lloyd-Max Quantizer | Rate-Distortion Bound |
|---|---|---|---|---|
| 1 | 5.56 | 5.59 | 4.40 | 6.02 |
| 2 | 11.04 | 11.55 | 9.30 | 12.04 |
| 3 | 16.64 | 17.57 | 14.62 | 18.06 |

## 6.4 Gaussian Wyner-Ziv Coding

### 6.4.1 System model

For the Wyner-Ziv problem, let $X, Y$ be two joint Gaussian source and $X = Y + Z$, where $Z$ is a Gaussian noise independent of $Y$ with variance $\sigma_z^2$.[1] A typical system

[1] For a more general Wyner-Ziv model in the Gaussian case, the relationship between the two joint source can also be $Y = X + Z$, where $Z \sim N(0, \sigma_z^2)$ is a Gaussian noise independent of $X$. In this case, we can perform the MMSE rescaling on $Y$ to make $X = \dot{\alpha}Y + \dot{Z}$, where $\dot{\alpha} = \frac{\sigma_x^2}{\sigma_y^2}$ and $\dot{Z}$ is with

model of Wyner-Ziv coding for the Gaussian case is shown in Fig. 6.4. Given the side information $Y$, which is only available at the decoder's side, the Wyner-Ziv rate-distortion bound on source $X$ for a target average distortion $\Delta$ between $X$ and its reconstruction $\hat{X}$ is given by

$$R_{WZ}(\Delta) = \max\left\{\frac{1}{2}\log\left(\frac{\sigma_z^2}{\Delta}\right), 0\right\}. \tag{6.17}$$

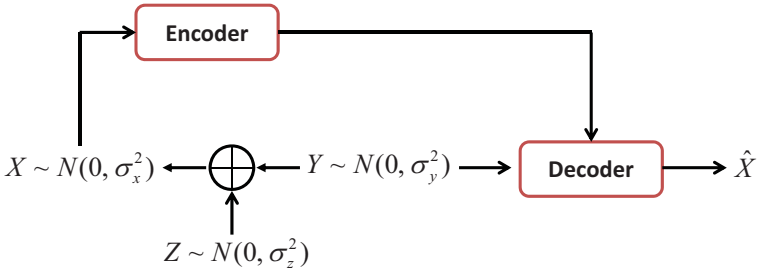



Figure 6.4: Wyner-Ziv coding for the Gaussian case. The variances of $Z$, $Y$ and $X$ are given by $\sigma_z^2$, $\sigma_y^2$ and $\sigma_x^2 = \sigma_y^2 + \sigma_z^2$, respectively.

### 6.4.2 A solution using continuous auxiliary variable

To achieve this bound, we assume a continuous auxiliary Gaussian random variable $X'$ which has an average distortion $\Delta'$ with source $X$, i.e., $X' = X + N(0, \Delta')$. Then we can also obtain that $X' = Y + N(0, \Delta' + \sigma_z^2)$. Letting $\sigma_{x'}^2$ be the variance of $X'$, the difference between the mutual information $I(X'; X)$ and $I(X'; Y)$ is given by

$$\begin{aligned} I(X'; X) - I(X'; Y) &= \frac{1}{2}\log\frac{\sigma_{x'}^2}{\Delta'} - \frac{1}{2}\log\frac{\sigma_{x'}^2}{\Delta' + \sigma_z^2} \\ &= \frac{1}{2}\log\frac{\Delta' + \sigma_z^2}{\Delta'}. \end{aligned} \tag{6.18}$$

variance $\frac{\sigma_z^2\sigma_x^2}{\sigma_z^2+\sigma_x^2}$. Then the Wyner-Ziv bound is given by $R_{WZ}(\Delta) = \max\left\{\frac{1}{2}\log\left(\frac{\sigma_z^2\sigma_x^2}{(\sigma_z^2+\sigma_x^2)\Delta}\right), 0\right\}$. Therefore, the system model can still be described by Fig. 6.4, with $Y$ and $Z$ being replaced by $\dot{\alpha}Y$ and $\dot{Z}$, respectively.

Let $I(X';X) - I(X';Y) = R_{WZ}(\Delta)$ and assume $\Delta \leq \sigma_z^2$. Then we have

$$\Delta' = \eta\Delta, \tag{6.19}$$

where $\eta = \frac{\sigma_z^2}{\sigma_z^2 - \Delta}$. Note that $\eta$ is the reciprocal of the MMSE rescaling parameter in the scenario of quantizing a Gaussian source with variance $\sigma_z^2$ for a target average distortion $\Delta$.

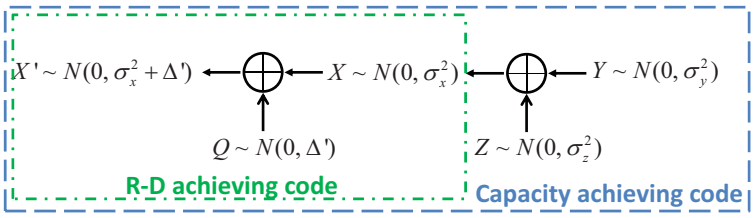
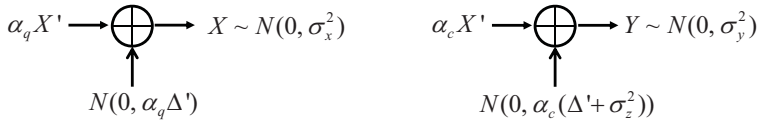



Figure 6.5: A solution of the Gaussian Wyner-Ziv problem using a continuous Gaussian random variable $X'$.

The above-mentioned solution for the Gaussian Wyner-Ziv problem is depicted by Fig. 6.5. Firstly we design a lossy compression code for source $X$ with Gaussian reconstruction alphabet $X'$. The average distortion between $X'$ and $X$ is $\Delta' = \eta\Delta$. Then we construct an AWGN capacity achieving code from $Y$ and $X'$. The final reconstruction of $X$ is given by $\hat{X} = Y + \frac{1}{\eta}(X' - Y)$. Clearly $\frac{1}{\eta}(X' - Y)$ is a scaled version of the Gaussian noise, which is independent of $X'$. The variance of $\frac{1}{\eta}(X' - Y)$ is

$$\begin{aligned} \frac{1}{\eta^2}(\Delta' + \sigma_z^2) &= \frac{\Delta}{\eta} + \frac{\sigma_z^2}{\eta^2} \\ &= \frac{\sigma_z^2 - \Delta}{\sigma_z^2}\Delta + \frac{\sigma_z^2 - \Delta}{\sigma_z^2}(\sigma_z^2 - \Delta) \\ &= \sigma_z^2 - \Delta. \end{aligned} \tag{6.20}$$

Then we can check that $X = \hat{X} + N(0, \Delta)$, which corresponds to the desired distortion, and the required data rate is $I(X';X) - I(X';Y) = \frac{1}{2}\log(\frac{\sigma_z^2}{\Delta})$.




(*a*) (*b*)

Figure 6.6: The MMSE rescaled channel blocks $(a)$ and $(b)$ for the Gaussian channels $X \to X'$ and $Y \to X'$, respectively.

### 6.4.3 A practical solution using lattice Gaussian distribution

The problem of the above-mentioned solution is that $X'$ is a continuous Gaussian random variable, which is impractical for the design of lattice codes. In order to utilize the proposed polar lattice coding technique, $X'$ is expected to obey a lattice Gaussian distribution. To this end, we perform MMSE rescaling on $X'$ for the AWGN channels $X \to X'$ and $Y \to X'$, respectively. The rescaled channels are shown in Fig. 6.6, where

$$\alpha_q = \frac{\sigma_x^2}{\sigma_x^2 + \Delta'} = \frac{\sigma_x^2(\sigma_z^2 - \Delta)}{\sigma_x^2(\sigma_z^2 - \Delta) + \sigma_z^2 \Delta}, \tag{6.21}$$

and

$$\alpha_c = \frac{\sigma_y^2}{\sigma_x^2 + \Delta'} = \frac{(\sigma_x^2 - \sigma_z^2)(\sigma_z^2 - \Delta)}{\sigma_x^2(\sigma_z^2 - \Delta) + \sigma_z^2 \Delta}. \tag{6.22}$$

Clearly, $\alpha_c < \alpha_q$. To combine the two blocks in Fig. 6.6 together, block $(b)$ is scaled by $\frac{\alpha_q}{\alpha_c}$. Consequently, a reversed version of the solution illustrated in Fig. 6.5 is obtained and shown in Fig. 6.7. For the reconstruction of $X$, we have the following proposition.

**Proposition 6.4.1.** *To achieve the $R_{WZ}(\Delta)$ bound by the reversed structure shown in Fig. 6.7, the reconstruction of $X$ is given by*

$$\hat{X} = \alpha_q X' + \gamma(\frac{\alpha_q}{\alpha_c} Y - \alpha_q X'), \quad \gamma = \frac{\sigma_y^2 \Delta}{\sigma_x^2 \sigma_z^2}. \tag{6.23}$$

*Proof.* It suffices to prove that $X = \hat{X} + N(0, \Delta)$. According to Fig. 6.7, we have $X = \alpha_q X' + N(0, \alpha_q \Delta')$, meaning that showing $\hat{X} = \alpha_q X' + N(0, \alpha_q \Delta' - \Delta)$ would complete this proof.

Clearly, $\frac{\alpha_q}{\alpha_c} Y - \alpha_q X'$ is a Gaussian random variable with $0$ mean and variance $\alpha_q \Delta' + \frac{\alpha_q}{\alpha_c}\sigma_z^2$, and it is independent of $X'$. By substituting the parameters $\Delta'$, $\alpha_q$ and $\alpha_c$, we have

$$\begin{aligned} \alpha_q \Delta' - \Delta &= \frac{\sigma_x^2(\sigma_z^2 - \Delta)}{\sigma_x^2(\sigma_z^2 - \Delta) + \sigma_z^2 \Delta} \frac{\sigma_z^2}{\sigma_z^2 - \Delta} \Delta - \Delta \\ &= \frac{\sigma_y^2 \Delta^2}{\sigma_x^2 \sigma_z^2 - \sigma_y^2 \Delta}, \end{aligned} \tag{6.24}$$

and

$$\begin{aligned} \gamma^2(\alpha_q \Delta' + \frac{\alpha_q}{\alpha_c}\sigma_z^2) &= \Big(\frac{\sigma_y^2 \Delta}{\sigma_x^2 \sigma_z^2}\Big)^2 \Big( \frac{\sigma_x^2(\sigma_z^2 - \Delta)}{\sigma_x^2(\sigma_z^2 - \Delta) + \sigma_z^2 \Delta} \frac{\sigma_z^2}{\sigma_z^2 - \Delta} \Delta + \frac{\sigma_x^2}{\sigma_y^2}\sigma_z^2 \Big), \\ &= \Big(\frac{\sigma_y^2 \Delta}{\sigma_x^2 \sigma_z^2}\Big)^2 \Big[ \sigma_x^2 \sigma_z^2 \Big( \frac{\Delta}{\sigma_x^2 \sigma_z^2 - \sigma_y^2 \Delta} + \frac{1}{\sigma_y^2} \Big) \Big] \\ &= \frac{\sigma_y^2 \Delta^2}{\sigma_x^2 \sigma_z^2 - \sigma_y^2 \Delta}, \end{aligned} \tag{6.25}$$

as we desired. □

Now the continuous Gaussian random variable $\alpha_q X'$ can be replaced by a lattice Gaussian distributed variable $A \sim D_{\Lambda, \sigma_a^2}$, where $\sigma_a^2 = \alpha_q^2 \sigma_{x'}^2$. Let $\bar{X} = A + N(0, \alpha_q \Delta')$ and $\frac{\alpha_q}{\alpha_c}\bar{Y} = \bar{X} + N(0, \frac{\alpha_q}{\alpha_c}\sigma_z^2)$. Let $\bar{B} = \frac{\alpha_q}{\alpha_c}\bar{Y}$ and $\sigma_b^2 = \frac{\alpha_q^2}{\alpha_c^2}\sigma_y^2$ for convenience. By Lemma 6.2.2, the distributions of $\bar{X}$ and $\bar{Y}$ can be made arbitrarily close to those of $X$ and $Y$, respectively. Then the polar lattices are designed by treating $\bar{X}$ as the source and $\bar{Y}$ as its side information. A rate-distortion bound achieving polar lattice $L_1$ is constructed for source $\bar{X}$ with target distortion $\alpha_q \Delta'$, and an AWGN capacity-achieving polar lattice $L_2$ is constructed to help the decoder extract some information from $\bar{Y}$, as shown in Fig. 6.7. Finally, the decoder recon-

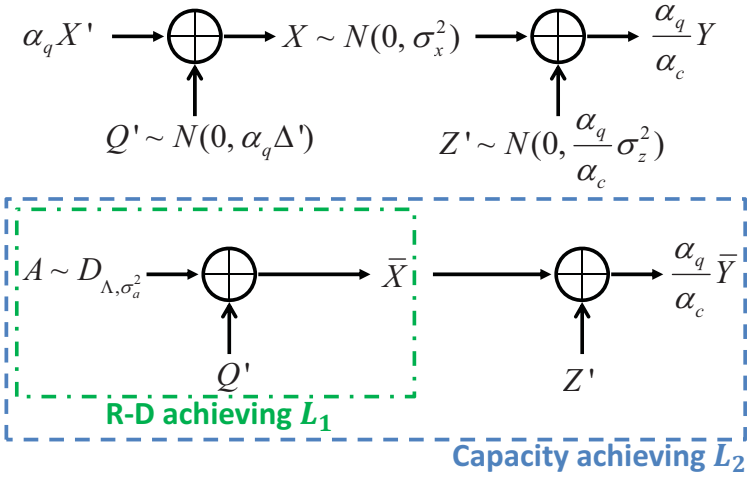



Figure 6.7: A reverse solution of the Gaussian Wyner-Ziv problem, which is more compatible with lattice Gaussian distribution.

structs $\check{X} = A + \gamma(\bar{B} - A)$. Conceptually, $\bar{B} - A$ is a Gaussian noise which is independent of $A$. [2] Recall that $\gamma = \frac{\sigma_y^2 \Delta}{\sigma_x^2 \sigma_z^2}$ scales $\bar{B} - A$ to $N(0, \alpha_q \Delta' - \Delta)$. By Lemma 6.2.2 again, the distributions of $\check{X}$ and $\hat{X}$ can be very close, resulting in an average distortion close to $\Delta$.

When lattice Gaussian distribution is utilized, by [104, Lemma 10], $L_1$ and $L_2$ are accordingly constructed for the MMSE-rescaled Gaussian noise variance $\tilde{\sigma}_q^2$ and $\tilde{\sigma}_c^2$, where

$$\tilde{\sigma}_q^2 = \frac{\sigma_a^2}{\sigma_x^2} \alpha_q \Delta' = \frac{\sigma_a^2 \sigma_z^2 \Delta}{\sigma_x^2 \sigma_z^2 - \sigma_y^2 \Delta}, \tag{6.26}$$

and

$$\tilde{\sigma}_c^2 = \frac{\sigma_a^2}{\sigma_b^2} (\alpha_q \Delta' + \frac{\alpha_q}{\alpha_c} \sigma_z^2) = \frac{\sigma_a^2 \sigma_z^4}{\sigma_x^2 \sigma_z^2 - \sigma_y^2 \Delta}. \tag{6.27}$$

Since $\Delta \leq \sigma_z^2$, we also have $\tilde{\sigma}_q^2 \leq \tilde{\sigma}_c^2$.

Now we are ready to introduce the polar lattice coding strategy. The idea is

[2] In fact, when $A$ is reconstructed by the decoder, $\bar{B} - A$ is not exactly a Gaussian noise $N(0, \alpha_q \Delta' + \frac{\alpha_q}{\alpha_c} \sigma_z^2)$, since the quantization noise of $L_1$ is not exactly Gaussian distributed. However, according to Theorem 6.2.5, the two distributions can be made arbitrarily close when $N$ is sufficiently large. See Fig. 6.3b for an example.

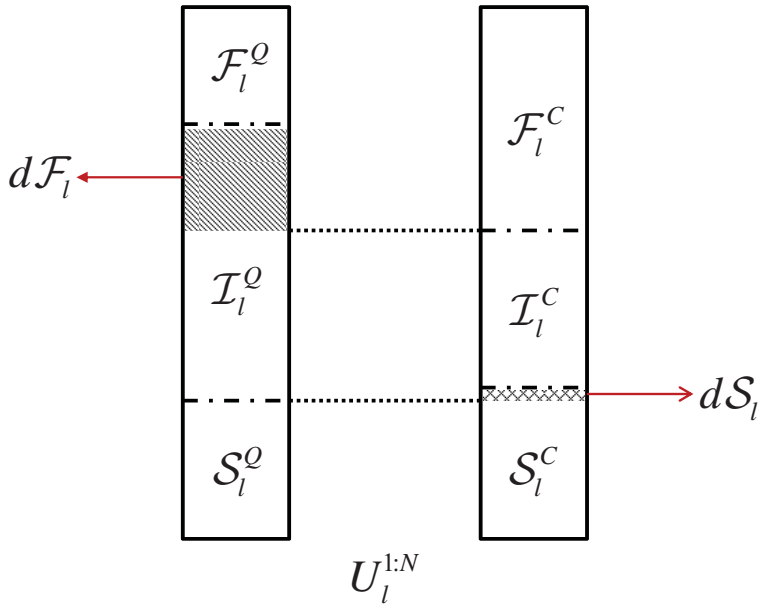



Figure 6.8: The partitions of $U_\ell^{1:N}$ for quantization lattice $L_1$ (left) and channel coding lattice $L_2$ (right). $\mathcal{F}_\ell^Q \subseteq \mathcal{F}_\ell^C$, $\mathcal{I}_\ell^C \subseteq \mathcal{I}_\ell^Q$, and $\mathcal{S}_\ell^Q \subseteq \mathcal{S}_\ell^C$. Without the side information, $U_\ell^{\mathcal{I}_\ell^Q}$ should be sent to achieve the target distortion. With the side information, however, $U_\ell^{\mathcal{I}_\ell^C}$ can be decoded and hence only $U_\ell^{d\mathcal{I}_\ell}$ need to be sent.

similar to the one mentioned in [9]. We choose a good constellation $D_{\Lambda,\sigma_a^2}$ such that the flatness factor $\epsilon_\Lambda(\tilde{\sigma}_q)$ is negligible. Let $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'/\cdots$ be a one-dimensional binary partition chain labeled by bits $A_1/A_2/\cdots/A_{r-1}/A_r/\cdots$. Then $P_{A_{1:r}}$ and $A_{1:r}$ approach $D_{\Lambda,\sigma_a^2}$ and $A$, respectively, as $r \to \infty$. Consider $N$ i.i.d. copies of $A$. Let $U_\ell^{1:N} = A_\ell^{1:N} G_N$ for each $1 \le \ell \le r$. The partitions of $U_\ell^{1:N}$ for both $L_1$ and $L_2$ are shown in Fig. 6.8, where the left block is for quantization lattice $L_1$ and the right one for channel coding lattice $L_2$. According to Section 6.2 and [104], for $0 < \beta < 0.5$, the frozen set $\mathcal{F}_\ell^Q$ ($\mathcal{F}_\ell^C$), information set $\mathcal{I}_\ell^Q$ ($\mathcal{I}_\ell^C$) and the shaping set $\mathcal{S}_\ell^Q$ ($\mathcal{S}_\ell^C$) for lattice $L_1$ ($L_2$) are given by

$$\begin{cases} \mathcal{F}_\ell^Q = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{X}^{1:N}) \ge 1 - 2^{-N^\beta}\} \\ \mathcal{I}_\ell^Q = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) > 2^{-N^\beta} \text{ and} \\ \qquad\qquad Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{X}^{1:N}) < 1 - 2^{-N^\beta}\} \\ \mathcal{S}_\ell^Q = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) \le 2^{-N^\beta}\}, \end{cases} \tag{6.28}$$

and

$$\begin{cases} \mathcal{F}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \mathcal{I}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) \geq 1 - 2^{-N^\beta} \text{and} \\ \qquad Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) \leq 2^{-N^\beta}\} \\ \mathcal{S}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta} \text{ or} \\ \qquad 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) < 1 - 2^{-N^\beta}\}. \end{cases} \tag{6.29}$$

For these two partitions, we have the following lemma.

**Lemma 6.4.2.** *Let $L_1$ and $L_2$ be two polar lattices constructed according to the above two partition rules respectively. $L_2$ is nested within $L_1$, i.e., $L_2 \subseteq L_1$.*

*Proof.* Both $L_1$ and $L_2$ follow the multilevel lattice structure (3.46). Let $\{C_1^q, ..., C_\ell^q, ..., C_r^q\}$ and $\{C_1^c, ..., C_\ell^c, ..., C_r^c\}$ denote the multilevel codes for $L_1$ and $L_2$, respectively. When shaping is not involved, the generator matrixes of $C_\ell^q$ and $C_\ell^c$ correspond to the sets of row indices $\mathcal{I}_\ell^Q \cup \mathcal{S}_\ell^Q$ and $\mathcal{I}_\ell^C \cup \mathcal{S}_\ell^Q$, respectively. By the relationship $\tilde{\sigma}_q^2 \leq \tilde{\sigma}_c^2$ and Lemma 4.2.2, the partition channel $C(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_c^2)$ is degraded with respect to $C(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_q^2)$. Then by the equivalence lemma Lemma 4.3.11, we have $\mathcal{I}_\ell^C \subseteq \mathcal{I}_\ell^Q$, meaning that $C_\ell^c \subseteq C_\ell^q$ for $1 \leq \ell \leq r$. As a result, $L_2 \subseteq L_1$. □

By channel degradation, we have $\mathcal{F}_\ell^Q \subseteq \mathcal{F}_\ell^C$. Let $d\mathcal{F}_\ell$ denote the set $\mathcal{F}_\ell^C \setminus \mathcal{F}_\ell^Q$. Meanwhile, we have $\mathcal{S}_\ell^Q \subseteq \mathcal{S}_\ell^C$ by definition. Denoting by $d\mathcal{S}_\ell$ the set $\mathcal{S}_\ell^C \setminus \mathcal{S}_\ell^Q$, $d\mathcal{S}_\ell$ can be written as

$$\begin{aligned} d\mathcal{S}_\ell = \{i \in [N] : 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta} \text{ or} \\ 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) < 1 - 2^{-N^\beta}\}, \end{aligned} \tag{6.30}$$

and the proportion $\frac{|d\mathcal{S}_\ell|}{N} \to 0$ as $N \to \infty$. Also observe that $d\mathcal{I}_\ell = \mathcal{I}_\ell^Q \setminus \mathcal{I}_\ell^C =$

$d\mathcal{F}_\ell \cup d\mathcal{S}_\ell$.

Given an $N$-dimensional realization vector $x^{1:N}$ of $X^{1:N}$, the encoder evaluates $u_\ell^{1:N}$ from level 1 to level $r$ successively according to the random rounding quantization rules given in Section 6.2. (See (6.9), (6.10), (6.13) and (6.14).) Recall that treating $x^{1:N}$ as a realization of $\bar{X}^{1:N}$ is safe because $X$ and $\bar{X}$ are similarly distributed. Then $u_\ell^{d\mathcal{I}_\ell}$ is sent to the decoder for each level. For the decoder, the realization vector $y^{1:N}$ of $Y^{1:N}$ is scaled to $b^{1:N} = \frac{\alpha_q}{\alpha_c} y^{1:N}$. Since $u_\ell^{\mathcal{F}_\ell^Q}$ is shared between the encoder and decoder before transmission, after receiving $u_\ell^{d\mathcal{I}_\ell}$, $u_\ell^{\mathcal{I}_\ell^C}$ and $u_\ell^{\mathcal{S}_\ell^Q}$ can be decoded with vanishing error probability since their associate Bhattacharyya parameters are arbitrarily small when $N \to \infty$. The details of SC decoding for Gaussian channels have been discussed in Chapter 4. According to Lemma 4.3.6, probabilities $P_{U_\ell^i|U_\ell^{1:i-1},A_{1:\ell-1}^{1:N}}$, $P_{U_\ell^i|U_\ell^{1:i-1},A_{1:\ell-1}^{1:N},\bar{X}^{1:N}}$ and $P_{U_\ell^i|U_\ell^{1:i-1},A_{1:\ell-1}^{1:N},\bar{B}^{1:N}}$ can be evaluated with $O(N \log N)$ complexity. It is worth mentioning that $u_\ell^{\mathcal{S}_\ell^C}$ is covered by a pre-shared random mapping in Section 4.3.2. However, as shown in Theorem 6.2.3, replacing the random mapping with MAP decision for $u_\ell^{\mathcal{S}_\ell^Q}$ will not change the results of Theorem 4.3.7 and Theorem 4.3.9. The proof is similar to that of Theorem 6.2.3 and omitted for brevity. Then the whole vector $u_\ell^{1:N}$ can be recovered with high probability. After obtaining the $u_\ell^{1:N}$ for $1 \le \ell \le r$, the realization $a^{1:N}$ of $A^{1:N}$ can be recovered from $u_\ell^{1:N}$ according to the following equation

$$\chi = \sum_{\ell=1}^{r} 2^{\ell-1} \left[ \sum_{i \in \mathcal{I}_\ell} u_\ell^i \psi(\mathbf{g}_i) + \sum_{i \in \mathcal{S}_\ell} u_\ell^i \psi(\mathbf{g}_i) + \sum_{i \in \mathcal{F}_\ell} u_\ell^i \psi(\mathbf{g}_i) \right], \tag{6.31}$$

where $\mathbf{g}_i$ denotes the $i$-th row of the polarization matrix $G_N$ and $\psi$ is the natural embedding. Clear $a^{1:r}$ is drawn from $D_{2^r\mathbb{Z}^N+\chi,\sigma_a}$. For each dimension, when $r$ is sufficiently large, the probability of choosing a constellation point outside the interval $[-2^{r-1}, 2^{r-1})$ is negligible (see Lemma 4.3.1 for more detail). Therefore, there exists only one point within $[-2^{r-1}, 2^{r-1})$ with probability close to 1 and

$a^{1:N}$ can be recovered by $\chi \mod 2^r$. Finally, the reconstruction of $x^{1:N}$ is given by $\check{x}^{1:N} = a^{1:N} + \gamma(b^{1:N} - a^{1:N})$.

To sum up, we have the following Wyner-Ziv coding scheme.

- Encoding: For the $N$-dimensional i.i.d. source vector $X^{1:N}$, the encoder evaluates $U^{\mathcal{I}_\ell^Q}$ by random rounding, and then sends $U_\ell^{d\mathcal{I}_\ell}$ to the decoder.
- Decoding: Using the pre-shared $U^{\mathcal{F}_\ell^Q}$ and the received $U_\ell^{d\mathcal{I}_\ell}$, the decoder recovers $U^{\mathcal{I}_\ell^C}$ and $U^{\mathcal{S}_\ell^Q}$ from the side information $B^{1:N}$. For each level the decoder obtains $U_\ell^{1:N}$, then $A^{1:N}$ can be recovered according to (6.31).
- Reconstruction: $\check{X}^{1:N} = A^{1:N} + \gamma(B^{1:N} - A^{1:N})$.

With regard to the design rate, by Theorem 6.2.5, the rate $R_{L_1}$ of $L_1$ can be arbitrarily close to $\frac{1}{2}\log\frac{\sigma_x^2}{\alpha_q\Delta'}$. However, the encoder does not need to send that much information to the decoder because of the side information. By Theorem 4.3.10, the rate $R_{L_2}$ of $L_2$ can be arbitrarily close to $\frac{1}{2}\log\left(\frac{\sigma_b^2}{\alpha_q\Delta' + \frac{\alpha_q}{\alpha_c}\sigma_z^2}\right)$. After some tedious calculation, we have

$$R_{L_1} \to \frac{1}{2}\log\left(\frac{\sigma_x^2\sigma_z^2 - \sigma_y^2\Delta}{\sigma_z^2\Delta}\right)^+, \tag{6.32}$$

and

$$R_{L_2} \to \frac{1}{2}\log\left(\frac{\sigma_x^2\sigma_z^2 - \sigma_y^2\Delta}{\sigma_z^4}\right)^-, \tag{6.33}$$

meaning that the transmission rate $R_{L_1} - R_{L_2} \to \frac{1}{2}\log(\frac{\sigma_z^2}{\Delta})^+$.

Before presenting the main theorem of the Gaussian Wyner-Ziv coding, we need a more stringent requirement on the flatness factor. This requirement is to guarantee a sub-exponentially decaying error probability for our lattice coding scheme.

**Proposition 6.4.3.** *For a one-dimensional binary partition chain* $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'/\cdots$ *and any given* $\tilde{\sigma}$, $r = O(\log N)$ *is sufficient to guarantee a sub-exponentially vanishing flatness factor* $\epsilon_\Lambda(\tilde{\sigma}) = O(2^{-\sqrt{N}})$. *Moreover, the mutual information of the bottom level* $I(\bar{X}; A_r|A_{1:r-1}) \to 0$ *and using the first* $r$ *levels only incurs a capacity loss* $\sum_{\ell>r} I(\bar{X}; A_\ell|A_{1:\ell-1}) \leq O(\frac{1}{N})$.

*Proof.* Since the partition is with dimension one, we can assume that $\Lambda = \eta\mathbb{Z}$. Let $\Lambda^* = \frac{1}{\eta}\mathbb{Z}$ be the dual lattice of $\Lambda$. By [21, Corollary 1], we have

$$\begin{aligned}
\epsilon_\Lambda(\tilde{\sigma}) &= \Theta_{\Lambda^*}(2\pi\tilde{\sigma}^2) - 1 \\
&= \sum_{\lambda\in\Lambda^*} \exp(-2\pi^2\tilde{\sigma}^2\|\lambda\|^2) - 1 \\
&= 2\sum_{\lambda\in\frac{1}{\eta}\mathbb{Z}_+} \exp(-2\pi^2\tilde{\sigma}^2\|\lambda\|^2) \qquad (6.34)\\
&\leq \frac{2\exp(-2\pi^2\tilde{\sigma}^2\frac{1}{\eta^2})}{1-\exp(-2\pi^2\tilde{\sigma}^2\frac{3}{\eta^2})} \\
&\leq 4\exp(-2\pi^2\tilde{\sigma}^2\frac{1}{\eta^2}),
\end{aligned}$$

where $Z_+$ denotes positive integers and the last inequality satisfies for sufficiently small $\eta$. Let $\frac{1}{\eta^2} = O(\sqrt{N})$ and hence $\epsilon_\Lambda(\tilde{\sigma}) = O(2^{-\sqrt{N}})$. In addition, by Lemma 4.3.1, a number of levels $r_1 = O(\log\log N)$ is needed to guarantee a vanishing mutual information at the bottom level. Let $\delta\mathbb{Z}/\cdots/2^{r_1}\mathbb{Z}$ be a partition such that $I(\bar{X}; A_r|A_{1:r-1}) \to 0$ for a constant $\delta$. Finally, the number of levels for partition $\eta\mathbb{Z}/\cdots/\delta\mathbb{Z}/\cdots/2^{r_1}\mathbb{Z}$ satisfies $r = \log(\frac{2^{r_1}}{\eta}) = O(\log N)$. □

The following theorem is proved in Appendix I.

**Theorem 6.4.4.** *Let* $X$ *be a Gaussian source and* $Y$ *be another Gaussian source correlated to* $X$ *as* $X = Y + Z$, *where* $Z \sim N(0, \sigma_z^2)$ *is an independent Gaussian noise. Consider a target distortion* $0 \leq \Delta \leq \sigma_z^2$ *for source* $X$ *when* $Y$ *is only available for the decoder. Let* $\Lambda/\Lambda_1/\cdots/\Lambda_r$ *be a one-dimensional binary partition*

*chain such that* $\epsilon_\Lambda(\tilde{\sigma}_q) = O(2^{-\sqrt{N}})$ *and* $r = O(\log N)$. *For any* $0 < \beta' < \beta < 0.5$, *there exists two nested polar lattices* $L_1$ *and* $L_2$ *with a differential rate* $R = R_{L_1} - R_{L_2}$ *arbitrarily close to* $\frac{1}{2}\log(\frac{\sigma_z^2}{\Delta})$ *such that the expect distortion* $\Delta_Q$ *satisfies*

$$\Delta_Q \leq \Delta + O(2^{-N^{\beta'}}), \tag{6.35}$$

*and the block error probability satisfies*

$$P_e^{WZ} \leq O(2^{-N^{\beta'}}). \tag{6.36}$$

## 6.5 Gaussian Gelfand-Pinsker coding

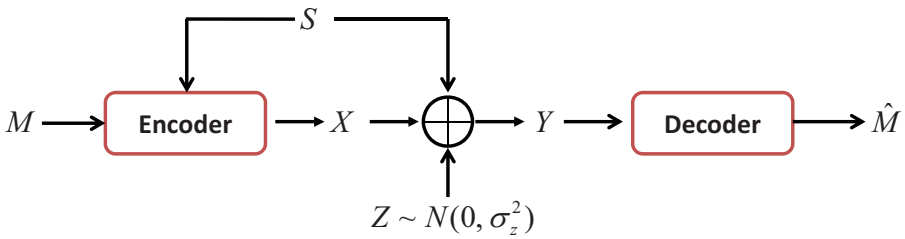



Figure 6.9: Gelfand-Pinsker coding for the Gaussian case.

For the Gelfand-Pinsker problem, with some abuse of notations, consider the channel described by $Y = X + S + Z$, where $X$ and $Y$ are the channel input and output, respectively, $Z$ is an unknown additive Gaussian noise with variance $\sigma_z^2$ and $S$ is an interference Gaussian signal with variance $\sigma_i^2$ known only to the encoder. A diagram of Gelfand-Pinsker coding is shown in Fig. 6.9. Message $M$ is encoded into $X$ which satisfies the power constraint $\frac{1}{N}E[\|X^{1:N}\|^2] \leq P$. The channel capacity of this Gaussian Gelfand-Pinsker model [126, 127] is given by

$$C_{GP} = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_z^2}\right). \tag{6.37}$$

To achieve this capacity, the roles of quantization lattice and channel coding

lattice are reversed. To see this, we still start with a continuous auxiliary variable and then replace it with a discrete Gaussian distributed one. Letting $\rho = \frac{P}{P+\sigma_z^2}$, we firstly design a lossy compression code for $\rho S$ with Gaussian reconstruction alphabet $S'$. The distortion between $S'$ and $\rho S$ is targeted to be $P$, i.e., $S' = \rho S + N(0, P)$. Then the encoder transmits $X = S' - \rho S$ ($X$ is independent of $S$), which satisfies the power constraint. Moreover, the relationship between $Y$ and $S'$ is given by

$$\begin{aligned} S' &= X + \rho S \\ &= X + \rho(Y - X - Z) \\ &= \rho Y + (1-\rho)X - \rho Z. \end{aligned} \tag{6.38}$$

Then the expectation

$$E\big[Y \cdot [(1-\rho)X - \rho Z]\big] = (1-\rho)E[X^2] - \rho E[Z^2] = 0, \tag{6.39}$$

meaning that $(1-\rho)X - \rho Z$ is independent of $Y$, which gives $S' = \rho Y + N(0, \frac{P\sigma_z^2}{P+\sigma_z^2})$. Then we construct an AWGN capacity-achieving code to recover $S'$ from $\rho Y$. Without the power constraint, the maximum data rate that can be sent is actually $I(S'; \rho Y)$. However, when power constraint is taken into consideration, some bits should be selected according to the realization of $S$ since $S'$ and $S$ are related. The maximum data rate becomes $I(S'; \rho Y) - I(S'; \rho S) = \frac{1}{2}\log(1 + \frac{P}{\sigma_z^2}) = C_{GP}$. A diagram of this solution is shown in Fig. 6.10, where

$$\sigma_{s'}^2 = \rho^2 \sigma_i^2 + P, \tag{6.40}$$

and

$$\sigma_y^2 = \frac{1}{\rho^2}\frac{P^2}{P+\sigma_z^2} + \sigma_i^2 = \sigma_i^2 + P + \sigma_z^2 \tag{6.41}$$

are the variances of $S'$ and $Y$, respectively.

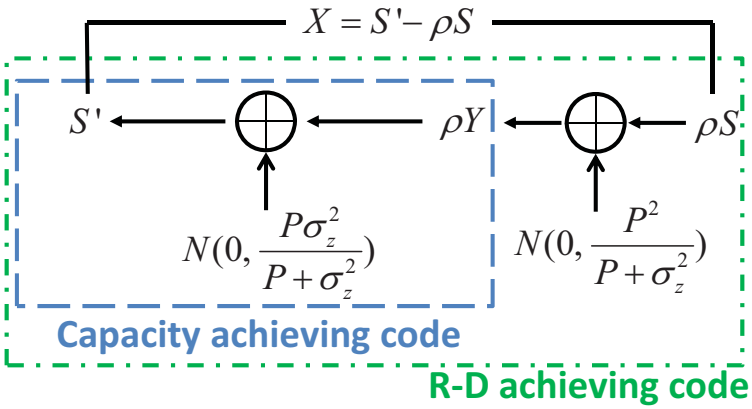



Figure 6.10: A solution of the Gaussian Gelfand-Pinsker problem using continuous Gaussian random variable $S'$.

Similarly, we prefer to use a discrete lattice Gaussian distributed version of $S'$ to approach this capacity. The idea is to perform MMSE rescaling on $S'$ to get a reversed version of the model shown in Fig. 6.10. The analysis is similar to that presented in Section 6.4 and is omitted here for brevity. Finally, the reversed solution is given in Fig. 6.11.

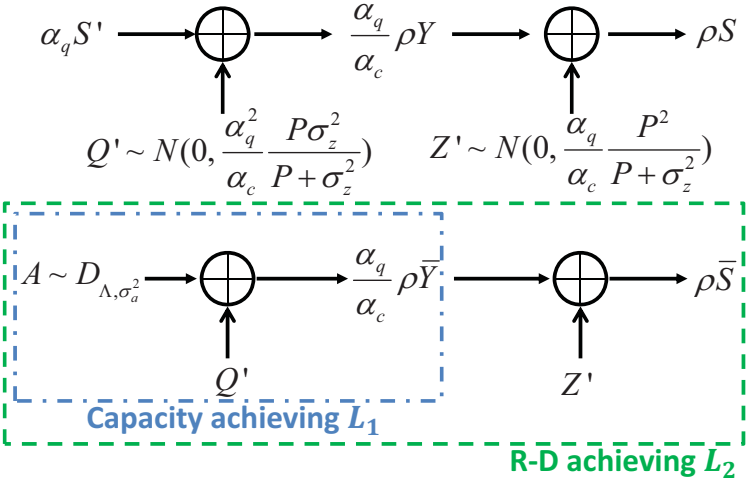



Figure 6.11: A reverse solution of the Gaussian Gelfand-Pinsker problem.

With some abuse of notation, let $A$ denote the discrete version of $\alpha_q S'$. The

MMSE rescaling factor $\alpha_c$ for channel coding and $\alpha_q$ for quantization are given by

$$\alpha_c = \frac{P\sigma_y^2}{P\sigma_i^2 + (P+\sigma_z^2)^2}, \tag{6.42}$$

and

$$\alpha_q = \frac{P\sigma_i^2}{P\sigma_i^2 + (P+\sigma_z^2)^2}, \tag{6.43}$$

respectively. The variance $\sigma_a^2$ for $D_{\Lambda,\sigma_a^2}$ is chosen to be $\alpha_q^2\sigma_{s'}^2$. Polar lattices $L_1$ and $L_2$ are accordingly constructed for Gaussian noise variance $\tilde{\sigma}_c^2$ and $\tilde{\sigma}_q^2$, where

$$\tilde{\sigma}_c^2 = \frac{\alpha_q^2\alpha_c}{\rho} \cdot \frac{\sigma_z^2\sigma_{s'}^2}{\sigma_y^2}, \tag{6.44}$$

and

$$\tilde{\sigma}_q^2 = \frac{\alpha_q^3}{\rho^2} \cdot \frac{P\sigma_{s'}^2}{\sigma_i^2}. \tag{6.45}$$

Check that $\frac{\tilde{\sigma}_c^2}{\tilde{\sigma}_q^2} = \frac{\sigma_z^2}{P+\sigma_z^2} \leq 1$. Recall that $X = S' - \rho S$=$(1-\alpha_q)S' + \alpha_q S' - \rho S$. When $\alpha_q S'$ is replaced by $A$, the encoded signal, denoted by $\bar{X}$, is given by

$$\bar{X} = \frac{1-\alpha_q}{\alpha_q}A + A - \rho S. \tag{6.46}$$

Note that the distributions of $S$ and $\bar{S}$ can be arbitrarily close when $\epsilon_\Lambda(\tilde{\sigma}_c) \to 0$. Clearly, $A - \rho\bar{S}$ is a Gaussian random variable independent of $A$ with distribution $N(0, \alpha_q P)$. By Lemma 6.2.2, $\bar{X}$ can be very close to a Gaussian random variable with distribution $N(0,P)$. [3] Thus, the power constraint can be satisfied.

With some abuse of notations, let $B = \frac{\alpha_q}{\alpha_c}\rho Y$, $\bar{B} = \frac{\alpha_q}{\alpha_c}\rho\bar{Y}$, $T = \rho S$, and $\bar{T} = \rho\bar{S}$ for convenience. We choose a good constellation $D_{\Lambda,\sigma_a^2}$ such that the flatness fac-

[3] Check that $\left(\frac{1-\alpha_q}{\alpha_q}\right)^2\sigma_a^2 = (1-\alpha_q)P$.

tor $\epsilon_\Lambda(\tilde{\sigma}_c)$ is negligible. Let $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'/\cdots$ be a one-dimensional binary partition chain labeled by bits $A_1/A_2/\cdots/A_{r-1}/A_r/\cdots$. Then $P_{A_{1:r}}$ and $A_{1:r}$ approach $D_{\Lambda,\sigma_a^2}$ and $A$, respectively, as $r \to \infty$. Consider $N$ i.i.d. copies of $A$. Let $U_\ell^{1:N} = A_\ell^{1:N} G_N$ for each $1 \le \ell \le r$. The partition of $U_\ell^{1:N}$ is shown in Fig. 6.12, where the left block is for quantization lattice $L_2$ and the right one for channel coding lattice $L_1$. For $0 < \beta < 0.5$, the frozen set $\mathcal{F}_\ell^Q$ ($\mathcal{F}_\ell^C$), information set $\mathcal{I}_\ell^Q$ ($\mathcal{I}_\ell^C$) and the shaping set $\mathcal{S}_\ell^Q$ ($\mathcal{S}_\ell^C$) for lattice $L_2$ ($L_1$) are given by

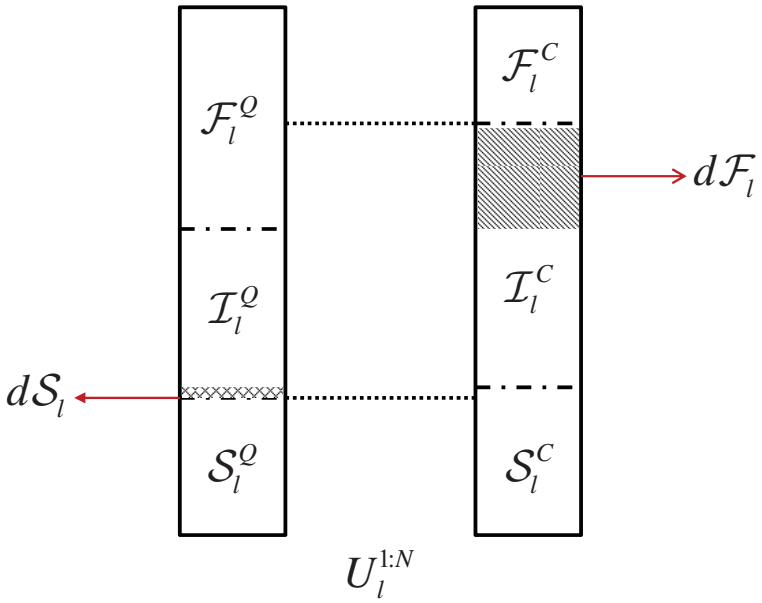



Figure 6.12: The partitions of $U_\ell^{1:N}$ for quantization lattice $L_2$ (left) and channel coding lattice $L_1$ (right). $\mathcal{F}_\ell^C \subseteq \mathcal{F}_\ell^Q$ and $\mathcal{S}_\ell^Q \subseteq \mathcal{S}_\ell^C$. Without the power constraint, $U_\ell^{\mathcal{I}_\ell^C}$ can be sent as message bits. With the power constraint, however, $U_\ell^{\mathcal{I}_\ell^Q}$ should be selected according to the interference $S^{1:N}$ and hence only $U_\ell^{d\mathcal{F}_\ell}$ can be fed with the message bits.

$$\begin{cases} \mathcal{F}_\ell^Q = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{T}^{1:N}) \ge 1 - 2^{-N^\beta}\} \\ \mathcal{I}_\ell^Q = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) > 2^{-N^\beta} \text{ and} \\ \qquad\qquad Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{T}^{1:N}) < 1 - 2^{-N^\beta}\} \\ \mathcal{S}_\ell^Q = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) \le 2^{-N^\beta}\}, \end{cases} \tag{6.47}$$

and

$$\begin{cases} \mathcal{F}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \mathcal{I}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) \geq 1 - 2^{-N^\beta} \text{ and} \\ \qquad Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) \leq 2^{-N^\beta}\} \\ \mathcal{S}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta} \text{ or} \\ \qquad 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) < 1 - 2^{-N^\beta}\}. \end{cases} \tag{6.48}$$

By channel degradation, we have $\mathcal{F}_\ell^C \subseteq \mathcal{F}_\ell^Q$. Let $d\mathcal{F}_\ell$ denote the set $\mathcal{F}_\ell^Q \setminus \mathcal{F}_\ell^C$. Meanwhile, we also have $\mathcal{S}_\ell^Q \subseteq \mathcal{S}_\ell^C$. The difference $d\mathcal{S}_\ell = \mathcal{S}_\ell^C \setminus \mathcal{S}_\ell^Q$ can also be written as (6.30), and the proportion $\frac{|d\mathcal{S}_\ell|}{N} \to 0$ as $N \to \infty$.

Given an $N$-dimensional realization vector $s^{1:N}$ of $S^{1:N}$, the encoder scales $s^{1:N}$ to $t^{1:N} = \rho s^{1:N}$ and evaluates $u_\ell^{1:N}$ from level 1 to level $r$ successively according to the random rounding quantization rules. Note that $u_\ell^{\mathcal{F}_\ell^C}$ is uniformly random and known to the decoder, and $u_\ell^{d\mathcal{F}_\ell}$ is fed with message bits which are also uniform. Recall that treating $t^{1:N}$ as a realization of $\bar{T}^{1:N}$ is reasonable because $T$ and $\bar{T}$ are similarly distributed. Then $u_\ell^{1:N}$ can be obtained for $1 \leq \ell \leq r$. When $r$ is sufficiently large, the lattice points outside $[-2^{r-1}, 2^{r-1})$ occur with almost 0 probability, the realization $a^{1:N}$ of $A^{1:N}$ can be determined by $u_{1:r}^{1:N}$ according to (6.31). Then $x^{1:N} = \frac{1}{\alpha_q} a^{1:N} - \rho s^{1:N}$ is the encoded signal as discussed in (6.46).

For the encoder, the realization vector $y^{1:N}$ of $Y^{1:N}$ is scaled to $b^{1:N} = \frac{\alpha_q}{\alpha_c} \rho y^{1:N}$. The task is to recover $u_\ell^{\mathcal{I}_\ell^C}$ at each level and hence message $u_\ell^{d\mathcal{F}_\ell}$ can be obtained. Note that $u_\ell^{\mathcal{F}_\ell^Q}$ is shared between the encoder and decoder before transmission, and $u_\ell^{\mathcal{S}_\ell^Q}$ can be decoded with vanishing error probability using the bit-wise MAP rule. The decoder still needs to know the unpolarized bits $u_\ell^{d\mathcal{S}_\ell}$ since the Bhattacharyya parameters of those indices are not necessarily vanishing. Therefore, a code with negligible rate is needed to send $u_\ell^{\mathcal{S}_\ell^Q}$ to the decoder in advance at each level. This

two phases transmission method has already been used in [9]. In this sense, $L_2$ is not exactly nested within $L_1$ because of those unpolarized indices. When $u_\ell^{d\mathcal{S}_\ell}$ is also available, $u_\ell^{\mathcal{I}_\ell^C}$ can be decoded with very small error probability Theorem with $O(N \log N)$ complexity (See Theorem 4.3.10.).

The Gaussian Gelfand-Pinsker coding scheme is summarized as follows.

- Encoding: According to the $N$-dimensional i.i.d. interference vector $S^{1:N}$, the encoder evaluates $U^{\mathcal{I}_\ell^Q}$ by random rounding, and then feeds $U_\ell^{d\mathcal{F}_\ell}$ with message bits. $U^{\mathcal{F}_\ell^C}$ is pre-shared and $U^{\mathcal{S}_\ell^Q}$ is determined by other bits according to $D_{\Lambda,\sigma_a^2}$. For each level, the encoder obtains $U_\ell^{1:N}$ for $1 \le \ell \le r$, then $A^{1:N}$ is recovered from $U_{1:r}^{1:N}$. The encoded signal is given by

$$\bar{X}^{1:N} = \frac{1}{\alpha_q} A^{1:N} - \rho S^{1:N}. \tag{6.49}$$

- Decoding: Using the pre-shared $U^{\mathcal{F}_\ell^Q}$ and the bits $U_\ell^{d\mathcal{S}_\ell}$ by the two phases transmission, the decoder recovers $U^{\mathcal{I}_\ell^C}$ including the message bits and $U^{\mathcal{S}_\ell^Q}$ from the received signal.

For the rate of lattice codes, we have

$$R_{L_1} \to \frac{1}{2} \log \left( \frac{P\sigma_i^2 + (P+\sigma_z^2)^2}{\sigma_z^2 (P+\sigma_z^2)} \right)^-, \tag{6.50}$$

and

$$R_{L_2} \to \frac{1}{2} \log \left( \frac{P\sigma_i^2 + (P+\sigma_z^2)^2}{(P+\sigma_z^2)^2} \right)^+, \tag{6.51}$$

indicating that the $R_{L_1} - R_{L_2} \to C_{GP}$.

The proof of the following theorem is given in Appendix J.

**Theorem 6.5.1.** *Let $S$ be a Gaussian noise known to the encoder, and $Z$ be anoth-*

*er independent and unknown Gaussian noise with variance $\sigma_z^2$. Consider a power constraint $P$ for the encoded signal. Let $\Lambda/\Lambda_1/\cdots/\Lambda_r$ be a one-dimensional binary partition chain such that $\epsilon_\Lambda(\tilde{\sigma}_c) = O(2^{-\sqrt{N}})$ and $r = O(\log N)$. For any $0 < \beta' < \beta < 0.5$, there exists two nested polar lattices $L_1$ and $L_2$ with a differential rate $R = R_{L_1} - R_{L_2}$ arbitrarily close to $\frac{1}{2}\log(1 + \frac{P}{\sigma_z^2})$ such that the expect transmit power $P_T$ satisfies*

$$P_T \leq P + O(2^{-N^{\beta'}}), \tag{6.52}$$

*and the block error probability satisfies*

$$P_e^{GP} \leq O(2^{-N^{\beta'}}). \tag{6.53}$$

## 6.6 Polar Lattices for Wyner's Common Information

In this section, we show that polar lattices can be utilized to extract Wyner's common information for joint Gaussian sources. More explicitly, we prove that extracting Wyner's common information is equivalent to lossy compression for a single Gaussian source. Before that, we firstly present some background of the Gray-Wyner network and Wyner's common information.

### 6.6.1 Gray-Wyner Network and Common Information

There are different ways to characterize the amount of common information in literature. Apart from Shannon's mutual information and Gács and Körner's common randomness [128], Wyner proposed an alternative definition to quantify the common information of $(X, Y)$ with finite alphabet [129] as $C(X, Y) = \inf_{X-W-Y} I(X, Y; W)$, where the infimum is taken over all $W$, such that $X - W - Y$

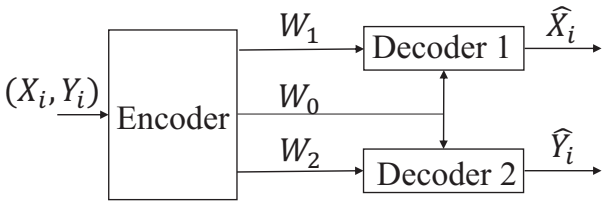



Figure 6.13: Gray-Wyner source coding network.

forms a Markov chain.

Wyner's definition is originated from his earlier work on the Gray-Wyner network [130], as depicted in Fig. 6.13. It demonstrates Wyner's first interpretation [129] of $C(X,Y)$. This network model contains an encoder that observes a pair of sequences $(X_1^N, Y_1^N)$ and outputs three messages $W_0$, $W_1$ and $W_2$ with rate $R_0, R_1, R_2$, respectively. Decoder 1 reconstructs $X_1^N$ by observing $(W_0, W_1)$ and decoder 2 reconstructs $Y_1^N$ from $(W_0, W_2)$. Wyner also gave a second interpretation of the common information. In that model, a common message $W$ is sent to two independent processors. The processors generate output sequences separately according to distributions $P_{X|W}(x|w)$ and $P_{Y|W}(y|w)$. The output sequences $\hat{X}_1^N$ and $\hat{Y}_1^N$ have joint probability $P_{\hat{X}_1^N\hat{Y}_1^N}(\hat{x}_1^N, \hat{y}_1^N) = \sum_{w\in\mathcal{W}} \frac{1}{|\mathcal{W}|} P_{X_1^N|W}(\hat{x}_1^N|w) P_{Y_1^N|W}(\hat{y}_1^N|w)$. Wyner showed that $C(X,Y)$ is equal to the minimum rate on the shared message, with constrains that the sum of rates equals the joint entropy or that the joint distribution $P_{\hat{X}_1^N\hat{Y}_1^N}(\hat{x}_1^N, \hat{y}_1^N)$ is arbitrarily close to $P_{X_1^N Y_1^N}(x_1^N, y_1^N)$.

Wyner and Gács-Körner's works on common information can be considered two different viewpoints of the lossless Gray-Wyner region. Their works were then extended by [131, 132] to the lossy case, where the output sequences $(\hat{X}_1^N, \hat{Y}_1^N)$ have certain distortion.

### 6.6.2 Common Information for Joint Gaussian Sources

The common information was generalized to the joint Gaussian sources in [132]. Let $X$, $Y$ be two bivariate Gaussian RVs with zero mean and covariance matrix $\left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]$.

The common RV of $(X, Y)$ is described by a Gaussian RV $W$ with mean $0$ and variance $\rho$ $(0 < \rho < 1)$ such that

$$\begin{cases} X = W + \sqrt{1-\rho}N_1 \\ Y = W + \sqrt{1-\rho}N_2 \end{cases}, \tag{6.54}$$

where $N_1$ and $N_2$ are standard Gaussian RVs and $N_1, N_2, W$ are independent of each other. Clearly, the common information is given by $I(X, Y; W) = \frac{1}{2}\log\frac{1+\rho}{1-\rho}$.

However, extracting a continuous common message from $(X, Y)$ is difficult. We show that a discrete version of $W$ is also eligible to convey the common message of two joint Gaussian RVs, according to Wyner's second approach to the characterization of common information [129].

**Lemma 6.6.1.** *Let $\bar{W}$ be a RV which follows a discrete Gaussian distribution $D_{\Lambda,\sqrt{\rho}}$. Consider two continuous RVs $\bar{X}$ and $\bar{Y}$ such that $\bar{X} = \bar{W} + \sqrt{1-\rho}N_1$ and $\bar{Y} = \bar{W} + \sqrt{1-\rho}N_2$, where $N_1$ and $N_2$ are the same as that given in (6.54). Let $f_{\bar{X},\bar{Y}}(x, y)$ and $f_{X,Y}(x, y)$ denote the joint PDF of $(\bar{X}, \bar{Y})$ and $(X, Y)$, respectively. If $\epsilon = \epsilon_\Lambda\left(\sqrt{\frac{\rho(1-\rho)}{1+\rho}}\right) < \frac{1}{2}$, the statistical distance between $f_{\bar{X},\bar{Y}}(x, y)$ and $f_{X,Y}(x, y)$ is upper-bounded by $\int_{\mathbb{R}^2}\left|f_{\bar{X},\bar{Y}}(x, y) - f_{X,Y}(x, y)\right| dxdy \leq 4\epsilon$, and the mutual information $I(\bar{X}, \bar{Y}; \bar{W})$ satisfies*

$$I(\bar{X}, \bar{Y}; \bar{W}) \geq \frac{1}{2}\log\frac{1+\rho}{1-\rho} - 5\epsilon\log(e).$$

*Proof.* Since $\bar{X} - \bar{W} - \bar{Y}$ is a Markov chain, we have

$$\begin{aligned} f_{\bar{X},\bar{Y}}(x,y) &= \sum_{a\in\Lambda} f_{\bar{X},\bar{Y},\bar{W}}(x,y,a) \\ &= \sum_{a\in\Lambda} f_{\bar{W}}(a) f_{\bar{X}|\bar{W}}(x|a) f_{\bar{Y}|\bar{W}}(y|a) \\ &= \frac{1}{f_{\sqrt{\rho}}(\Lambda)} \sum_{a\in\Lambda} \frac{1}{\sqrt{2\pi\rho}} \exp\left(-\frac{a^2}{2\rho}\right) \frac{1}{\sqrt{2\pi(1-\rho)}} \\ &\quad \cdot \exp\left(-\frac{(x-a)^2}{2(1-\rho)}\right) \frac{1}{\sqrt{2\pi(1-\rho)}} \exp\left(-\frac{(y-a)^2}{2(1-\rho)}\right) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right) \\ &\quad \cdot \frac{\sum_{a\in\Lambda} \frac{1}{\sqrt{2\pi\frac{\rho(1-\rho)}{1+\rho}}} \exp\left(-\frac{(a-\frac{\rho(x+y)}{1+\rho})^2}{2\cdot\frac{\rho(1-\rho)}{1+\rho}}\right)}{f_{\sqrt{\rho}}(\Lambda)}, \end{aligned} \tag{6.55}$$

where $\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right) = f_{X,Y}(x,y)$ is the PDF of two joint Gaussian variables. By the definition of the flatness factor, we have

$$\begin{aligned} &|V(\Lambda) \sum_{a\in\Lambda} \frac{1}{\sqrt{2\pi\frac{\rho(1-\rho)}{1+\rho}}} \exp\left(-\frac{(a-\frac{\rho(x+y)}{1+\rho})^2}{2\frac{\rho(1-\rho)}{1+\rho}}\right) - 1| \\ &\qquad \le \epsilon_\Lambda\left(\sqrt{\frac{\rho(1-\rho)}{1+\rho}}\right) = \epsilon. \end{aligned} \tag{6.56}$$

Since $\epsilon_\Lambda(\sigma)$ is a monotonically decreasing function of $\sigma$ (see [21, Remark 2]), we have $\epsilon_(\sqrt{\rho}) \le \epsilon$ and hence

$$|V(\Lambda) f_{\sqrt{\rho}}(\Lambda) - 1| \le \epsilon. \tag{6.57}$$

Combining (6.55), (6.56) and (6.57) gives us

$$f_{X,Y}(x,y)(1-2\epsilon) \le f_{X,Y}(x,y) \cdot \frac{1-\epsilon}{1+\epsilon} \le f_{\bar{X},\bar{Y}}(x,y), \tag{6.58}$$

and

$$f_{\bar{X},\bar{Y}}(x,y) \leq f_{X,Y}(x,y) \cdot \frac{1+\epsilon}{1-\epsilon} \leq f_{X,Y}(x,y)(1+4\epsilon), \tag{6.59}$$

when $\epsilon < \frac{1}{2}$.

Finally,

$$\begin{aligned} &\int_{\mathbb{R}^2} |f_{\bar{X},\bar{Y}}(x,y) - f_{X,Y}(x,y)| dxdy \\ &\quad \leq 4\epsilon \int_{\mathbb{R}^2} f_{X,Y}(x,y) dxdy \\ &\quad = 4\epsilon. \end{aligned} \tag{6.60}$$

Similarly, the Kullback-Leibler divergence between $f_{\bar{X},\bar{Y}}(x,y)$ and $f_{X,Y}(x,y)$ can be upper-bounded as

$$\begin{aligned} \mathbb{D}(f_{\bar{X},\bar{Y}} \| f_{X,Y}) &= \int_{\mathbb{R}^2} f_{\bar{X},\bar{Y}}(x,y) \log \frac{f_{\bar{X},\bar{Y}}(x,y)}{f_{X,Y}(x,y)} dxdy \\ &\leq \int_{\mathbb{R}^2} f_{\bar{X},\bar{Y}}(x,y) \log(1+4\epsilon) dxdy \\ &= \log(1+4\epsilon). \end{aligned} \tag{6.61}$$

For any $\sqrt{\frac{\rho(1-\rho)}{1+\rho}} > 0$, $\epsilon$ can be made arbitrarily small by scaling $\Lambda$. Therefore, when $\epsilon \to 0$, $\bar{W}$ can be viewed as the common message according to Wyner's second approach. To see that $I(\bar{X},\bar{Y};\bar{W})$ can be arbitrarily close to the common information,

we rewrite $\mathbb{D}(f_{\bar{X},\bar{Y}}\|f_{X,Y})$ as

$$\begin{aligned}\mathbb{D}(f_{\bar{X},\bar{Y}}\|f_{X,Y}) &= \int_{\mathbb{R}^2} f_{\bar{X},\bar{Y}}(x,y)\log\frac{f_{\bar{X},\bar{Y}}(x,y)}{f_{X,Y}(x,y)}dxdy \\ &= -\int_{\mathbb{R}^2} f_{\bar{X},\bar{Y}}(x,y)\log f_{X,Y}(x,y)dxdy - h(\bar{X},\bar{Y}) \\ &= -\int_{\mathbb{R}^2} f_{\bar{X},\bar{Y}}(x,y)\log\Bigg(\frac{1}{2\pi\sqrt{1-\rho^2}} \\ &\qquad \cdot\exp\bigg(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\bigg)\Bigg)dxdy - h(\bar{X},\bar{Y}) \\ &= \log\left(2\pi\sqrt{1-\rho^2}\right) + \frac{E_{\bar{X},\bar{Y}}[x^2+y^2-2\rho xy]}{2(1-\rho^2)} - h(\bar{X},\bar{Y}) \\ &= \log\left(2\pi\sqrt{1-\rho^2}\right) + \frac{1+E_{\bar{W}}[w^2]}{1+\rho} - h(\bar{X},\bar{Y}).\end{aligned} \tag{6.62}$$

Note that $\epsilon_\Lambda(\sqrt{\rho}) \le \epsilon$. By [20, Lemma 9] and [20, Remark 3], it is easy to make $E_{\bar{W}}[w^2] \ge \rho(1-2\epsilon)$. Then we have

$$\begin{aligned}\mathbb{D}(f_{\bar{X},\bar{Y}}\|f_{X,Y}) &\ge \log\left(2\pi\sqrt{1-\rho^2}\right) + 1 - \epsilon - h(\bar{X},\bar{Y}) \\ &= h(X,Y) - h(\bar{X},\bar{Y}) - \epsilon.\end{aligned} \tag{6.63}$$

Using (6.61), we obtain

$$\begin{aligned}I(X,Y;W) - I(\bar{X},\bar{Y};\bar{W}) &= h(X,Y) - h(\bar{X},\bar{Y}) \\ &\le \log(1+4\epsilon) + \epsilon \\ &\le 5\epsilon.\end{aligned} \tag{6.64}$$

Therefore, according to Wyner's second approach, $\bar{W}$ is an eligible candidate of the common message of $(X,Y)$ when $\epsilon \to 0$. □

Similar to the quantization for Gaussian sources in Section 6.2, using $D_{\Lambda,\sqrt{\rho}}$ as the reconstruction distribution, we can design a quantization polar lattice from Construction D to extract the common randomness. The only difference is that the

size of the source alphabet is doubled in this work. The next theorem shows that the design of polar lattices for extracting the common randomness of a pair of joint Gaussian sources is exactly the same as that for quantizing a single Gaussian source, which means that the technique proposed in [105] can be directly employed to our work.

**Theorem 6.6.2.** *The construction of a polar lattice for extracting the common randomness of a pair of joint Gaussian sources $(X, Y)$ is equivalent to the construction of a rate-distortion bound achieving polar lattice for a Gaussian source $\frac{X+Y}{2}$.*

*Proof.* Let $\bar{W}$ be labelled by bits $\bar{W}_1, \cdots, \bar{W}_r$ ($\bar{W}_{1:r}$) according to a binary partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'(\Lambda_r)$. Then, $D_{\Lambda,\sqrt{\rho}}$ induces a distribution $P_{\bar{W}_{1:r}}$ whose limit corresponds to $D_{\Lambda,\sqrt{\rho}}$ as $r \to \infty$.

By the chain rule of mutual information

$$I(\bar{X}, \bar{Y}; \bar{W}_{1:r}) = \sum_{\ell=1}^{r} I(\bar{X}, \bar{Y}; \bar{W}_\ell | \bar{W}_{1:\ell-1}), \tag{6.65}$$

we obtain $r$ binary-input channels $V_\ell$ for $1 \le \ell \le r$. Given the realization $w_{1:\ell-1}$ of $\bar{W}_{1:\ell-1}$, denote by $\mathcal{A}_\ell(w_{1:\ell})$ the coset of $\Lambda_\ell$ indexed by $w_{1:\ell-1}$ and $w_\ell$. According to

[79], the channel transition PDF of the $\ell$-th channel $V_\ell$ is given by

$$\begin{aligned}
&f_{\bar{X},\bar{Y}|\bar{W}_\ell,\bar{W}_{1:\ell-1}}(x,y|w_\ell,w_{1:\ell-1}) \\
&\quad= \frac{1}{f_{\sqrt{\rho}}(\mathcal{A}_\ell(w_{1:\ell}))} \sum_{a\in\mathcal{A}_\ell(w_{1:\ell})} f_{\sqrt{\rho}}(a) f_{\bar{X},\bar{Y}|\bar{W}}(x,y|a) \\
&\quad= \frac{1}{f_{\sqrt{\rho}}(\mathcal{A}_\ell(w_{1:\ell}))} \sum_{a\in\mathcal{A}_\ell(w_{1:\ell})} \frac{1}{\sqrt{2\pi\rho}} \exp\left(-\frac{a^2}{2\rho}\right) \frac{1}{\sqrt{2\pi(1-\rho)}} \\
&\qquad\cdot \exp\left(-\frac{(x-a)^2}{2(1-\rho)}\right) \frac{1}{\sqrt{2\pi(1-\rho)}} \exp\left(-\frac{(y-a)^2}{2(1-\rho)}\right) \\
&\quad= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right) \frac{1}{f_{\sqrt{\rho}}(\mathcal{A}_\ell(w_{1:\ell}))} \\
&\qquad\cdot \sum_{a\in\mathcal{A}_\ell(w_{1:\ell})} \frac{1}{\sqrt{2\pi\frac{\rho(1-\rho)}{1+\rho}}} \exp\left(-\frac{(a-\frac{\rho(x+y)}{1+\rho})^2}{2\cdot\frac{\rho(1-\rho)}{1+\rho}}\right),
\end{aligned} \tag{6.66}$$

Using the channel symmetrization technique for asymmetric channels, the channel transition pdf of the symmetrized channel $\tilde{V}_\ell$ is

$$\begin{aligned}
&f_{\tilde{V}_\ell}(x,y,w_{1:\ell-1},w_\ell\oplus\tilde{w}_\ell|\tilde{w}_\ell) \\
&\quad= f_{\bar{X},\bar{Y},\bar{W}_{1:\ell}}(x,y,w_{1:\ell}) \\
&\quad= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right) \frac{1}{f_{\sqrt{\rho}}(\Lambda)} \\
&\qquad\cdot \sum_{a\in\mathcal{A}_\ell(w_{1:\ell})} \frac{1}{\sqrt{2\pi\frac{\rho(1-\rho)}{1+\rho}}} \exp\left(-\frac{(a-\frac{\rho(x+y)}{1+\rho})^2}{2\cdot\frac{\rho(1-\rho)}{1+\rho}}\right).
\end{aligned} \tag{6.67}$$

Comparing with the $\Lambda_{\ell-1}/\Lambda_\ell$ channel [104, eqn. (13)] or equivalently the symmetrized channel (4.34), we see that the symmetrized channel (6.67) is equivalent to a $\Lambda_{\ell-1}/\Lambda_\ell$ channel with noise variance $\frac{\rho(1-\rho)}{1+\rho}$.

Recall that $X, Y$ are two bivariate Gaussian variables with zero mean and covariance matrix $\left[\begin{smallmatrix}1 & \rho\\ \rho & 1\end{smallmatrix}\right]$. It is well known that $\frac{X+Y}{2}$ is a Gaussian variable with zero mean and variance $\frac{1+\rho}{2}$. Now consider the construction of polar lattices to quantize $\frac{X+Y}{2}$ using the reconstruction distribution $D_{\Lambda,\sqrt{\rho}}$. The MMSE re-scaled coefficient and

noise variance are given by $\frac{2\rho}{1+\rho}$ and $\frac{\rho(1-\rho)}{1+\rho}$, which is the same as that in (6.67). □

We note that the result of Theorem 6.6.2 can be further extended into multiple Gaussian RVs.

## 6.7 Summary

In this chapter, we present an explicit construction of polar lattices which are good for lossy compression. They are further utilized to resolve the Gaussian version of the Wyner-Ziv and Gelfand-Pinsker problems. Compared with the original idea given in [122], dither is not necessary in our scheme due to the property of discrete lattice Gaussian distribution [20], and the entropy encoder is already integrated in our lattice quantization process. The complexity of encoding and decoding complexity is $O(N \log^2 N)$ for a sub-exponentially decaying excess distortion. The quantization polar lattices can also be used to extract Wyner's common information of joint Gaussian sources.

CHAPTER 7

# Polar Codes and Polar Lattices for Independent Fading Channels

## 7.1 Introduction

Real-world wireless channels are generally modeled as time-varying fading channels due to multiple signal paths and user mobility. Compared with time-invariant channel models, the wireless fading channel models allow the channel gain to change randomly over time. In practice, we usually consider block fading channel models, where channel gain varies at a longer time scale than symbol transmission time. In the fast fading case, the code block length typically spans a large number of coherence time intervals and the channel is ergodic with a well-defined Shannon capacity. In this chapter we study the fast block fading channel with stationary ergodic channel gains. By assuming a perfect interleaving/de-interleaving on symbols, the channel can be considered to be memoryless, which offers much convenience for coding design. We further assume that channel state information (CSI) is available to receiver through training sequences, and the transmitter only has the channel distribution information (CDI).

Quasi-static fading channel with two states was studied in [133]. Construction of polar codes for block Rayleigh fading channels when CSI or CDI is available to both transmitter and receiver was considered in [134]. To achieve capacity, a series of polar codes are designed for different fading states, meaning that the scheme on-

ly works for finite fading states. In this work, we consider the case in which CSI is available to the receiver (the transmitter only knows CDI), and the fading states can be infinite. This is the case when a communication system is operated in the open-loop mode. By introducing the fading channel statistic into the polarization process, we construct one polar code and show that the same channel capacity can be achieved as in the case where CSI is available to both. The previous work [135] of polar codes for fading channels does not require CSI for the transmitter either. The authors proposed a novel hierarchic scheme to construct polar codes through two phases of polarization. The channel state is assumed to be constant over each coherence interval and the channel is modeled as a mixture of BSCs. The first phase of polarization is to get each BSC polarized into a set of extremal subchannels (ignoring the unpolarized part), which is treated as a set of realizations of BECs. Then, the second phase of polarization is to get the synthesized BECs polarized. This scheme achieves the ergodic capacity of binary input fading channels with finite states when the two phases are both sufficiently polarized. As a result, much longer block length than standard polar codes is needed to achieve channel capacity. According to [135, Tm. 2], the block length of polar codes is required to be $NB$, where $B$ is the number of blocks, and $N$ is the block length of standard polar codes as introduced in [4]. To achieve capacity, both $B$ and $N$ are required to be sufficiently large. In this chapter, we propose a new scheme with one-phase polarization to achieve the ergodic capacity by treating the channel gain as part of channel outputs. Consequently, we manage to get rid of the second phase of polarization, and the capacity is achieved by polar codes with block length $N$. We also note that the encoding and decoding complexity is reduced from $O(NB\log(NB))$ in [135] to $O(N\log N)$.

With regard to lattice codes for fading channels, algebraic tools [136] play an important role in explicit coding design. It was shown in [137] that lattice codes constructed from algebraic number field can achieve full diversity over fading channels,

which results in better error probability performance. A more recent work showed that number field lattices are able to achieve Gaussian and Rayleigh channel capacity within a constant gap [138]. This scheme was universal and further extended to the multiple-input and multiple-output (MIMO) context [139]. It is still an open question whether this gap can be removed. More recently, lattice codes were investigated in ergodic fading channels [140, 141] and proved to be capacity-achieving under the ambiguity decoding proposed in [142]. However, the construction of such lattice codes for ergodic fading channels is still implicit. In this work, we will resolve this problem using polar lattices for the i.i.d. fading case.

The technical ingredients of this work are the following:

- Explicit construction of polar codes for binary-input fading channels and their proof of capacity achieving. This work can be viewed as an extension of the work in [134], where CSI is required for transmitter to construct a series of polar codes for different channel gains to achieve the ergodic capacity. In our work, however, CSI is unnecessary for transmitter and only one single polar code is sufficient to achieve the ergodic capacity, which simplifies the coding design.

- Explicit construction of polar lattices for fading channels with certain input power constraint. As predicted in [141], there exist good lattice codes to approach the ergodic capacity of fading channels. Our work provides an explicit construction of such lattice codes. The work of lattice codes [50, 20, 104] for the AWGN channel can be considered as a specific case of our work, because the AWGN channel can be viewed as a special fading channel with a constant channel gain. The concept of AWGN-goodness or Poltyrev capacity is generalized to the fading case. Similar to Chapter 4, we firstly investigate the construction of polar lattices without power constraint and then apply lattice Gaussian shaping over them to satisfy the power constraint.

## 7.2 Polar Codes for Binary-Input Fading Channels

Consider the binary-input i.i.d. ergodic fading channel model

$$Y = HX + Z, \tag{7.1}$$

where $X \in \{-1, +1\}$ is the binary signal after BPSK modulation, $Y$ is the channel output, $Z$ is a zero mean independent Gaussian noise with variance $\sigma^2$, and $H$ is the channel gain. In this work, for convenience, we assume that $H$ follows Rayleigh distribution with probability density function (PDF)

$$P_H(h) = \frac{h}{\sigma_h^2} e^{-\frac{h^2}{2\sigma_h^2}}, \tag{7.2}$$

where $\sigma_h = \sqrt{\frac{2}{\pi}} \cdot E[H]$. As we will see later, our work can be easily generalized to other regular fading distributions [143].

Since we assume that $H$ is available to the receiver, the fading channel can be modeled as a channel with input $X$ and outputs $(Y, H)$, as shown in Fig. 7.1.

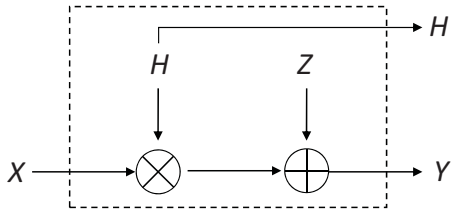



Figure 7.1: Binary-input ergodic fading channel with CSI available at the end of receiver.

We firstly show that the channel $\tilde{W} : X \to (Y, H)$ is symmetric. To see this, we

check the channel transition PDF of $\tilde{W}$, which is given by

$$\begin{aligned} P_{Y,H|X}(y,h|x) = &\quad P_H(h)P_{Y|X,H}(y|x,h) \\ = &\quad P_H(h)P_Z(z=y-xh) \\ = &\quad P_H(h)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-xh)^2}{2\sigma^2}}. \end{aligned} \tag{7.3}$$

We define a permutation $\phi$ over the outputs $(y,h)$ such that $\phi(y,h) = (-y,h)$. Check that $P_{Y,H|X}(y,h|+1) = P_{Y,H|X}(\phi(y,h)|-1)$ and hence $\tilde{W}$ is symmetric. It is well-known that uniform input distribution achieves the capacity of symmetric channels. Therefore, letting $X$ be uniform, the capacity of $\tilde{W}$ is given by

$$\begin{aligned} C(\tilde{W}) &= I(X;Y,H) \\ &= I(X;Y|H) \\ &= \sum_x \int_0^\infty \frac{h}{\sigma_h^2} e^{-\frac{h^2}{2\sigma_h^2}} dh \int_{-\infty}^\infty \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-xh)^2}{2\sigma^2}} \log\Big(\frac{e^{-\frac{(y-xh)^2}{2\sigma^2}}}{\frac{1}{2}e^{-\frac{(y-h)^2}{2\sigma^2}} + \frac{1}{2}e^{-\frac{(y+h)^2}{2\sigma^2}}}\Big) dy \\ &= 1 - \frac{1}{\sqrt{2\pi}\sigma\sigma_h^2}\int_0^\infty h e^{-\frac{h^2}{2\sigma_h^2}} dh \int_{-\infty}^\infty \big(1 - \log(1+e^{-\frac{2yh}{\sigma^2}})\big) dy, \end{aligned} \tag{7.4}$$

which is the same as the capacity when CSI is avaliable to both transmitter and receiver [134].

To achieve $C(\tilde{W})$, we combine $N$ independent copies of $\tilde{W}$ to $\tilde{W}_N$ and split it to obtain subchannel $\tilde{W}_N^{(i)}$ for $1 \le i \le N$. Let $U^{1:N} = X^{1:N}G_N$, $\tilde{W}_N^{(i)}$ has input $U^i$ and outputs $(U^{1:i-1}, Y^{1:N}, H^{1:N})$. Since $\tilde{W}$ is symmetric, $\tilde{W}_N^{(i)}$ is symmetric as well [4]. We can identify the information set $\mathcal{I}$ according to the Bhattacharyya parameter $\tilde{Z}(\tilde{W}_N^{(i)})$. Treating $(Y,H)$ as the outputs, by Definition 2.1.2,

$$\tilde{Z}(\tilde{W}) = \sum_{y,h} \sqrt{P_{Y,H|X}(y,h|+1)P_{Y,H|X}(y,h|-1)}. \tag{7.5}$$

Note that $\tilde{Z}(\tilde{W}_N^{(i)})$ can be evaluated recursively for BECs, starting with the initial

Bhattacharyya parameter $\tilde{Z}(\tilde{W})$ (see [4, eqn. (38)]). For general BMS channels, it is difficult to calculate $\tilde{Z}(\tilde{W}_N^{(i)})$ directly because of the exponentially increasing size of the output alphabet of $\tilde{W}_N^{(i)}$. Fortunately, we can apply the degrading and upgrading merging algorithms [5, 7] to estimate $\tilde{Z}(\tilde{W}_N^{(i)})$ within acceptable accuracy.

In practice, the two approximations caused by the degrading and upgrading processes are typically close. Therefore, we focus on the degrading transform for brevity.

Define the likelihood ratio (LR) of $(y, h)$ as

$$LR(y,h) \triangleq \frac{P_{Y,H|X}(y,h|+1)}{P_{Y,H|X}(y,h|-1)}. \tag{7.6}$$

By (7.3), we have $LR(y,h) = e^{\frac{2yh}{\sigma^2}}$ for the fading case. Clearly, $LR(y,h) \geq 1$ for any $y \geq 0$. Each $LR(y,h)$ corresponds to a BSC with crossover probability $\frac{1}{LR(y,h)+1}$ and its capacity is

$$C[LR(y,h)] = 1 - \mathfrak{h}_2\Big(\frac{1}{LR(y,h)+1}\Big), \tag{7.7}$$

where $\mathfrak{h}_2(\cdot)$ is the binary entropy function.

The fading channel $\tilde{W}$ is then quantized according to $C[LR(y,h)]$. Let $\mu = 2Q$ be the size of degraded channel output alphabet. The region $\{y \geq 0, h \geq 0\}$ is divided into $L$ sets

$$A_i = \left\{ y \geq 0, h \geq 0 : \frac{i-1}{Q} \leq C[LR(y,h)] < \frac{i}{Q} \right\}, \tag{7.8}$$

for $1 \leq i \leq Q$. The outputs in $A_i$ are mapped to one symbol, and $\tilde{W}$ is quantized to a mixture of $Q$ BSCs with the crossover probability

$$p_i = \frac{\int_{A_i} P_{Y,H|X}(y,h|-1)dydh}{\int_{A_i} P_{Y,H|X}(y,h|+1)dydh + \int_{A_i} P_{Y,H|X}(y,h|-1)dydh}. \tag{7.9}$$

Note that $p_i$ can be numerically evaluated. Since $LR(y,h) = e^{\frac{2yh}{\sigma^2}}$, $A_i$ is rewritten as

$$A_i = \left\{ y \geq 0, h \geq 0 : \frac{\sigma^2}{2} \ln \left( \frac{1}{\mathfrak{h}_2^{-1}(\frac{Q-i+1}{Q})} - 1 \right) \leq yh < \frac{\sigma^2}{2} \ln \left( \frac{1}{\mathfrak{h}_2^{-1}(\frac{Q-i}{Q})} - 1 \right) \right\} \quad (7.10)$$

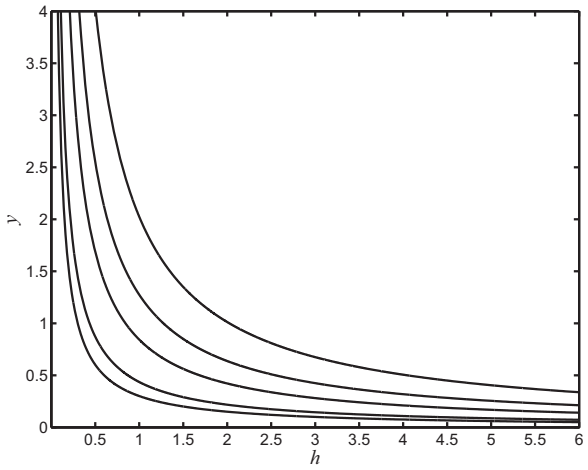



Figure 7.2: Typical boundaries of $A_i$ for channel quantization.

Let $\delta_1$ and $\delta_2$ denote the two constants in (7.10). We have

$$\int_{A_i} P_{Y,H|X}(y,h|+1)dydh = \int_0^\infty \frac{h}{\sigma_h^2} e^{-\frac{h^2}{2\sigma_h^2}} dh \int_{\frac{\delta_1}{h}}^{\frac{\delta_2}{h}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-h)^2}{2\sigma^2}} dy, \quad (7.11)$$

and $\int_{A_i} P_{Y,H|X}(y,h|-1)dydh$ is calculated similarly. Typical boundaries of $A_i$ are depicted in Fig. 7.2.

Let $\tilde{W}_Q$ denote the quantized channel from $\tilde{W}$ after the degrading transform. By [5, Lemma 13], the difference between the two channel capacities is upper-bounded by $\frac{1}{Q}$. A comparison between $C(\tilde{W}_Q)$ and $C(\tilde{W})$ for different $SNR = \frac{\sigma_h^2}{\sigma^2}$ when $Q = 128$ is shown in Fig. 7.3. When $Q$ is sufficiently large, we can use $\tilde{W}_Q$ to approximate $\tilde{W}$ to construct polar codes. The size of the subchannel output alphabet after degrading merging is no more than $2Q$.

The proof of the following theorem is standard and fully given in [5]. We omit it

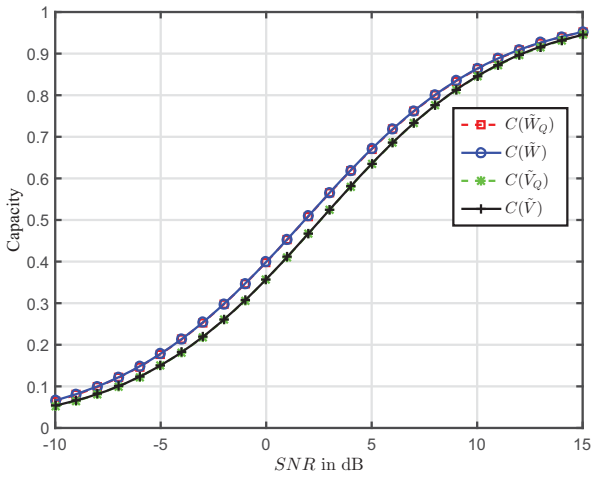



Figure 7.3: Comparisons between $C(\tilde{W}_Q)$ and $C(\tilde{W})$, and between $C(\tilde{V}_Q)$ and $C(\tilde{V})$, when $Q = 128$. Here $\tilde{W}$ denotes the channel $X \to (Y, H)$, and $\tilde{V}$ denotes the channel $X \to Y$, i.e., the two channel models when the receiver knows the CSI and the CDI, respectively. $\tilde{W}_Q$ and $\tilde{V}_Q$ denote the quantized version of $\tilde{W}$ and $\tilde{V}$, respectively.

for brevity.

**Theorem 7.2.1.** *Let $\tilde{W} : X \to (Y, H)$ be a binary-input i.i.d. fading channel. Let $N$ denote the block length and $\mu = 2Q$ denote the limit of the size of output alphabet. A polar code constructed by the degrading merging algorithm achieves the capacity $C(\tilde{W})$ when $N$ and $\mu$ are both sufficiently large. The block error probability under SC decoding is upper-bounded by $N2^{-N^\beta}$ for $0 < \beta < \frac{1}{2}$.*

Simulation result of polar codes with different block length for the binary-input Rayleigh fading channel with CSI available to the receiver are shown in Fig. 7.4, where $SNR = 5$ dB and $C(\tilde{W}) = 0.6709$. The performance can be further improved by using more sophisticated decoding algorithms [144, 39].

*Remark* 7.2.1. It has been pointed out in [134] that polar codes for the Rayleigh fading channel with known CDI suffer a penalty for not having complete information. The statement can be seen clearly from our construction. Treating $H$ as part of channel outputs, the binary channel $X \to Y$ is degraded with respect to the channel $X \to (Y, H)$, and $I(X; Y, H) \geq I(X; Y)$. Let $\tilde{V}$ denote the channel $X \to Y$. The

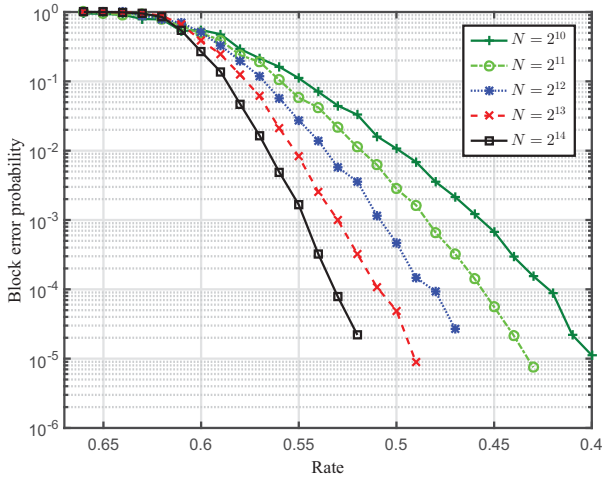



Figure 7.4: Performance of polar codes for the Rayleigh fading channel with CSI available to the receiver when $N = 2^{10}, 2^{11}, ..., 2^{14}$.

channel transition PDF of $\tilde{V}$ is written as

$$P_{Y|X}(y|x) = \int_h P_{Y,H|X}(y,h|x)dh, \tag{7.12}$$

where $P_{Y,H|X}(y,h|x)$ is given by (7.3). It is clear that $\tilde{V}$ is a BMS channel. Therefore, the degrading and upgrading merging algorithms can also be applied to construct polar codes for $\tilde{V}$. A comparison between $C(\tilde{W})$ and $C(\tilde{V})$ is shown in Fig. 7.3. By Remark 2.1.1, the polar code constructed when the receiver only knows the CDI is a subcode of that when the receiver knows the CSI. Simulation results of polar codes for the binary-input Rayleigh fading channel with CDI available to the receiver are shown in Fig. 7.5, where $SNR = 5$ dB and $C(\tilde{V}) = 0.6352$.

*Remark* 7.2.2. Our construction method can be generalized to other regular fading distributions such as the Rician distribution, the lognormal distribution, and the Nakagami distribution [143]. $P_{Y,H|X}(y,h|x)$, $C(\tilde{W})$ and $LR(y,h)$ can be calculated similarly, with $P_H(h)$ being replaced by other fading distributions. We apply the same channel quantization method to construct polar codes and achieve the capacity.

*Remark* 7.2.3. We note that the work in this part can be viewed as a component of the

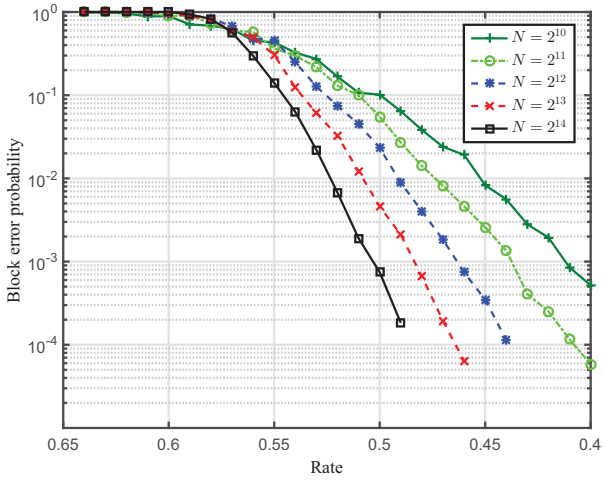



Figure 7.5: Performance of polar codes for the Rayleigh fading channel with CDI available to the receiver when $N = 2^{10}, 2^{11}, ..., 2^{14}$.

work in the next section, i.e., designing polar lattices for i.i.d. fading channels under certain power constraint. As we will see in the following content, a polar lattice is constructed from a series of polar codes according to a binary lattice partition chain. The fading $\Lambda/\Lambda'$ channel introduced by the partition chain at each level is a BMS channel with outputs $(\bar{Y}, H)$, where $H$ is the fading coefficient and $\bar{Y}$ is the result of $\frac{Y}{H} \mod \Lambda'$. The only difference between this fading $\Lambda/\Lambda'$ channel and the binary-input fading channel discussed in this section is the mod-$\Lambda'$ front-end. Treating $\bar{Y}$ as a modified version of $Y$, the construction method of polar codes given in this section will be used for fading $\Lambda/\Lambda'$ channels.

## 7.3 Polar Lattices for i.i.d. Fading Channels

In this section, we extend polar codes to polar lattices for i.i.d. fading channels. The reason for this extension is that the input of fading channels is not necessarily limited to be binary. In general, the input $X$ is subject to a power constraint $P$, i.e.,

$$E[X^2] \leq P. \tag{7.13}$$

In this case, lattice codes offer more choices of input constellation. As introduced in Chapter 4, polar lattices are originally designed for AWGN channels, and proved to be capable of achieving the AWGN capacity under lattice Gaussian shaping. However, for fading channels, which can be considered as AWGN channels with varying signal noise ratio, the multilevel lattice design method cannot be directly adapted. We may follow the idea of [134] and [135] to design a set of polar lattices for different signal noise ratio, assuming that CSI is known to the transmitter or using a second phase of polarization. As we have seen in Sect. 7.2, a more natural way is to treat fading states as a part of channel outputs and design polar lattices directly for this channel with combined outputs. To this end, we will generalize the definitions of mod-$\Lambda$ channel and $\Lambda/\Lambda'$ channel to the fading scenario. The design criteria of lattice codes for AWGN channels are also generalized to i.i.d. fading channels, making it possible to use other capacity-achieving codes to construct lattices for fading channels. Again we will show that the generalized $\Lambda/\Lambda'$ channels are sequentially degraded and polar lattices fit themselves naturally to this new scenario. We firstly construct polar lattices which achieve the Poltyrev capacity of i.i.d. fading channels and then perform lattice Gaussian shaping to achieve the ergodic capacity.

### 7.3.1 Polar Lattices for Fading Channels without Power Constraint

For i.i.d. fading channels, the channel gain varies. The above analysis for AWGN channels need to be generalized. Since the receiver knows the CSI, the fading effect on input $X$ can be canceled by scaling $Y$ with $\frac{1}{H}$, and meanwhile the equivalent additive Gaussian noise variance on $X$ is scaled by $\frac{1}{H^2}$. We define the fading mod-$\Lambda$ channel as follows.

**Definition 7.3.1** (fading mod-$\Lambda$ channel)**.** A fading mod-$\Lambda$ channel is a fading chan-

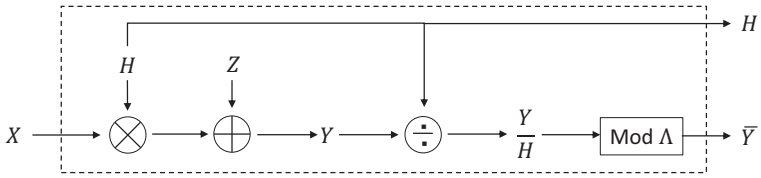



Figure 7.6: A block diagram of the fading mod-$\Lambda$ channel.

nel with an input in $\mathcal{V}(\Lambda)$, and an output being scaled by $\frac{1}{H}$ before the mod-$\mathcal{V}(\Lambda)$ operation. A block diagram of this model is shown in Fig. 7.6.

Note that our work can be extended to independent block fading channels where the fading coefficient remains the same during $n$ transmission symbols. We can choose a proper $n$-dimensional lattice partition chain according to the coherence time of the fading channel. The fading coefficient is also assumed to be independent between different blocks. The fading mod-$\Lambda$ channel is closely related to a mod-$\Lambda$ channel with noise variance $\frac{\sigma^2}{h^2}$. For i.i.d. fading channels considered in this work, we set $n = 1$ and use one-dimensional lattice partition chain. The channel transition PDF of the fading mod-$\Lambda$ channel is given by

$$\begin{aligned} P_{\tilde{Y},H|X}(\tilde{y},h|x) &= P_{Y,H|X}(y=\tilde{y}h+h\cdot\Lambda,h|x)\frac{dy}{d\tilde{y}} \\ &= h\cdot P_H(h)\sum_{\lambda\in\Lambda}P_{Y|X,H}(y=\tilde{y}h+\lambda h|x,h) \\ &= h\cdot P_H(h)\sum_{\lambda\in\Lambda}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\tilde{y}h+\lambda h-xh)^2}{2\sigma^2}} \\ &= P_H(h)\sum_{\lambda\in\Lambda}\frac{1}{\sqrt{2\pi}\frac{\sigma}{h}}e^{-\frac{(\tilde{y}+\lambda-x)^2}{2\left(\frac{\sigma}{h}\right)^2}}, \end{aligned} \tag{7.14}$$

where the second term in the last equation is the channel transition PDF of a mod-$\Lambda$ channel with noise variance $\frac{\sigma^2}{h^2}$. The channel transition PDF for higher dimension $n$ can be derived similarly. As a result, the fading mod-$\Lambda$ channel can be viewed as an independent combination of a Rayleigh distributed variable $H$ and a mod-$\Lambda$ channel

with noise variance $\frac{\sigma^2}{H^2}$. The capacity of the fading mod-$\Lambda$ channel is

$$\begin{aligned} C_H(\Lambda, \sigma^2) &= C(X; \tilde{Y}, H) \\ &= C(X; \tilde{Y}|H) \\ &= \int_h P_H(h) C(X; \tilde{Y}|h) dh \\ &= E_h\left[C\left(\Lambda, \frac{\sigma^2}{h^2}\right)\right] \\ &= \log V(\Lambda) - E_h\left[\mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h^2}\right)\right]. \end{aligned} \tag{7.15}$$

Similarly, a fading $\Lambda/\Lambda'$ channel is a fading mod-$\Lambda'$ channel whose input is restricted to discrete lattice points in $(\Lambda + a) \cap \mathcal{R}(\Lambda')$ for some translate $a$. By the same argument of (7.14), it can be viewed as an independent combination of a Rayleigh distributed variable $H$ and a $\Lambda/\Lambda'$ channel with noise variance $\frac{\sigma^2}{H^2}$. The capacity of the fading $\Lambda/\Lambda'$ channel is given by

$$\begin{aligned} C_H(\Lambda/\Lambda', \sigma^2) &= E_h\left[C\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right] - E_h\left[C\left(\Lambda, \frac{\sigma^2}{h^2}\right)\right] \\ &= E_h\left[\mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h^2}\right)\right] - E_h\left[\mathfrak{h}\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right] + \log\big(V(\Lambda')/V(\Lambda)\big). \end{aligned} \tag{7.16}$$

For a self-similar partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'$, we have

$$C_H(\Lambda/\Lambda', \sigma^2) = C_H(\Lambda/\Lambda_1, \sigma^2) + \cdots + C_H(\Lambda_{r-1}/\Lambda', \sigma^2). \tag{7.17}$$

Since the $\Lambda/\Lambda'$ channel is symmetric, it is easy to check that the $\Lambda/\Lambda'$ fading channel is symmetric as well. Moreover, if $|\Lambda/\Lambda'| = 2$, the $\Lambda/\Lambda'$ fading channel is a BMS channel. Taking the binary partition $\mathbb{Z}/2\mathbb{Z}$ as an example, the input of the $\mathbb{Z}/2\mathbb{Z}$ fading channel is $X \in \{0, 1\}$, and a permutation $\phi$ over the outputs $(\tilde{y}, h)$ is defined such that $\phi(\tilde{y}, h) = ([\tilde{y} - 1] \mod 2\mathbb{Z}, h)$. Check that $P_{\tilde{Y},H|X}(\tilde{y}, h|0) = P_{\tilde{Y},H|X}(\phi(\tilde{y}, h)|1)$.

It is now clear that polar lattices can be constructed to achieve the (ergodic) Poltyrev capacity of the i.i.d. fading channels, as we did for the AWGN channel in Chapter 4. Recall that the Poltyrev capacity $C_\infty$ of a general additive-noise channel is defined as the capacity per unit volume in Chapter 3. For the independent AWGN channel, we have

$$C_\infty = -\mathfrak{h}(\sigma^2) = \frac{1}{2}\log\left(\frac{1}{2\pi e\sigma^2}\right), \tag{7.18}$$

where $\mathfrak{h}(\sigma^2)$ denotes the differential entropy of a Gaussian random variable with variance $\sigma^2$.

For independent fading channels, $C_\infty$ is generalized as [143]

$$C_\infty = -E_h\left[\mathfrak{h}\left(\frac{\sigma^2}{h^2}\right)\right] = E_h\left[\frac{1}{2}\log\left(\frac{h^2}{2\pi e\sigma^2}\right)\right]. \tag{7.19}$$

In the special case of Rayleigh fading,

$$\begin{aligned} C_\infty &= -\int_h \frac{h}{\sigma_h^2} e^{-\frac{h^2}{2\sigma_h^2}} \frac{1}{2}\log\left(\frac{2\pi e\sigma^2}{h^2}\right) dh \\ &\underset{t=\frac{h^2}{2\sigma_h^2}}{=} -\frac{1}{2}\int_t e^{-t}\left(\log\left(\frac{2\pi e\sigma^2}{2\sigma_h^2}\right) - \log t\right) dt \\ &= -\frac{1}{2}\log\left(2\pi e\sigma^2 \cdot \frac{e^\zeta}{2\sigma_h^2}\right), \end{aligned} \tag{7.20}$$

where $\zeta = -\int_0^\infty e^{-x}\ln x dx$ is the Euler-Mascheroni constant.

To approach the Poltyrev capacity $-\frac{1}{2}\log\left(2\pi e\sigma^2 \cdot \frac{e^\zeta}{2\sigma_h^2}\right)$, we construct polar lattices according to the following three design criteria:

(a) The top lattice $\Lambda$ gives negligible capacity $E_h\left[C\left(\Lambda, \frac{\sigma^2}{h^2}\right)\right]$.

(b) The bottom lattice $\Lambda'$ has a small error probability $E_h\left[P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right]$.

(c) Each component polar code $\mathcal{C}_\ell$ is a capacity-approaching code for the $\Lambda_{\ell-1}/\Lambda_\ell$

fading channel.

For criterion (a), we pick a top lattice $\Lambda$ for a large channel gain $h_l$ such that $\mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h_l^2}\right) \approx \log V(\Lambda)$. By Remark 4.2.1, $\mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h^2}\right) \geq \mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h_l^2}\right)$ for $0 \leq h \leq h_l$.

$$\begin{aligned} E_h\left[\mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h^2}\right)\right] &= \int_0^{h_l} P_H(h)\mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h^2}\right)dh + \int_{h_l}^{\infty} P_H(h)\mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h^2}\right)dh \\ &\gtrapprox \mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h_l^2}\right)\int_0^{h_l} P_H(h)dh + n\int_{h_l}^{\infty} P_H(h)\mathfrak{h}\left(\frac{\sigma^2}{h^2}\right)dh \\ &= \mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h_l^2}\right)\left(1 - e^{-\frac{h_l^2}{2\sigma_h^2}}\right) + \frac{n}{2}\log\left(\frac{2\pi e\sigma^2}{h_l^2}\right)e^{-\frac{h_l^2}{2\sigma_h^2}} \\ &\quad - \frac{n}{2}\log e \cdot E_1\left(\frac{h_l^2}{2\sigma_h^2}\right), \end{aligned} \tag{7.21}$$

where $E_1(x) = \int_x^{\infty} \frac{e^{-t}}{t}dt$ is the exponential integral, and $E_1(x) \to 0$ for $x \to \infty$. The approximation is due to the fact $\mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h^2}\right) \to n\mathfrak{h}\left(\frac{\sigma^2}{h^2}\right)$ as $h \to \infty$. Let $h_l = O(N)$. We have $E_h\left[\mathfrak{h}\left(\Lambda, \frac{\sigma^2}{h^2}\right)\right] \approx \log V(\Lambda)$, and $E_h\left[C\left(\Lambda, \frac{\sigma^2}{h^2}\right)\right] \approx 0$ as $N \to \infty$ according to (7.15).

For criterion (b), we pick a bottom lattice $\Lambda'$ for a small channel gain $h_s$ such that $P_e\left(\Lambda', \frac{\sigma^2}{h_s^2}\right) \to 0$. Since $P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right) \leq P_e\left(\Lambda', \frac{\sigma^2}{h_s^2}\right)$ for $h \geq h_s$,

$$\begin{aligned} E_h\left[P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right] &= \int_0^{h_s} P_H(h)P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right)dh + \int_{h_s}^{\infty} P_H(h)P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right)dh \\ &\leq \left(1 - e^{-\frac{h_s^2}{2\sigma_h^2}}\right) + P_e\left(\Lambda', \frac{\sigma^2}{h_s^2}\right) \cdot e^{-\frac{h_s^2}{2\sigma_h^2}}. \end{aligned} \tag{7.22}$$

Let $h_s = O\left(\frac{1}{N^{\delta}}\right)$ for some constant $\delta \geq 1$. We have $E_h\left[P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right] \to 0$ as $N \to \infty$. Since the volume $V(\Lambda')$ is sufficiently large to cover almost all of the noised signal, by [50], we have $E_h\left[\mathfrak{h}\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right] \approx nE_h\left[\mathfrak{h}\left(\frac{\sigma^2}{h^2}\right)\right]$ when $E_h\left[P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right] \to 0$. Note that $\delta$ is required to be lager than 1 here to guarantee that $E_h\left[P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right]$ vanishes polynomially (see the proof of Theorem 7.3.1).

For criterion (c), we choose a binary partition chain and construct binary polar codes to achieve the capacity of the $\Lambda_{\ell-1}/\Lambda_{\ell}$ fading channel for $1 \leq \ell \leq r$. Since

the $\Lambda_{\ell-1}/\Lambda_\ell$ fading channel is a BMS channel, treating $(\tilde{Y}, H)$ as the outputs, the construction method proposed in Sect. 7.2 can be used. It remains to verify $\mathcal{C}_{\ell-1} \subseteq \mathcal{C}_\ell$. Since the the $\Lambda/\Lambda'$ fading channel can be viewed as an independent combination of a Rayleigh distributed variable $H$ and a $\Lambda/\Lambda'$ channel with noise variance $\frac{\sigma^2}{H^2}$, by Lemma 4.2.2 and Remark 2.1.1, we immediately have $\mathcal{C}_{\ell-1} \subseteq \mathcal{C}_\ell$. Simulation results of polar codes for the one-dimensional binary partition chain $\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z}/8\mathbb{Z}/16\mathbb{Z}$ with $\sigma = 1$, $\sigma_h = 1.2575$ and block length $N = 2^{14}$ are shown in Fig. 7.7.

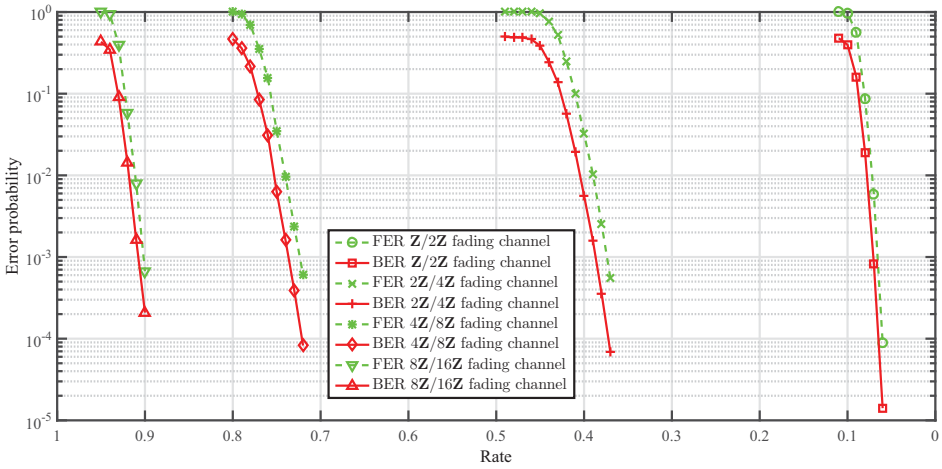



Figure 7.7: Performance of polar codes for the $\mathbb{Z}/2\mathbb{Z}$, $2\mathbb{Z}/4\mathbb{Z}$, $4\mathbb{Z}/8\mathbb{Z}$ and $8\mathbb{Z}/16\mathbb{Z}$ fading channels with $\sigma = 1$, $\sigma_h = 1.2575$ and $N = 2^{14}$. The capacities of these four channels are about 0.1172, 0.4929, 0.8200 and 0.9500, respectively. FER denotes the frame (block) error probability, and BER denotes the bit error probability.

**Theorem 7.3.1** (Good polar lattices for fading channel)**.** *For an independent Rayleigh fading channel with given $\sigma_h^2$ and $\sigma^2$, select an $n$-dimensional binary lattice partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'$ such that both the criterion (a) and (b) are satisfied. Construct a polar lattice $L$ from this partition chain and $r$ nested polar codes with block length $N$. Let $r = n\delta O(\log N)$ for a fixed dimension $n$ and some constant $\delta \geq 1$. $L$ can achieve the Poltyrev capacity of the i.i.d. fading channel, i.e., $\gamma_L(\sigma) \to 2\pi e \cdot \frac{e^\zeta}{2\sigma_h^2}$ and $P_e(L, \sigma^2) = O\big(\frac{1}{N^{2\delta-1}}\big) \to 0$, as $N \to \infty$.*

*Proof.* By the the union bound of the error probability under the multi-stage lattice

decoding [50], $P_e(L,\sigma^2)$ is upper-bounded by

$$P_e(L,\sigma^2) \le rN2^{-N^\beta} + N \cdot E_h\left[P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right]. \tag{7.23}$$

Let $h_s = O\big(\frac{1}{N^\delta}\big)$ for some constant $\delta \ge 1$ be a small channel gain and let $h_l = O(N)$ be a large channel gain. Consider a fine lattice $\Lambda_f$ and a coarse lattice $\Lambda_c$ in the lattice partition chain such that $\mathfrak{h}(\Lambda_f,\sigma^2) \approx \log V(\Lambda_f)$ and $P_e(\Lambda_c,\sigma^2) \to 0$. Let $d$ be the minimum distance of $\Lambda_c$. By the Chernoff bound, we have

$$P_e(\Lambda_c,\sigma^2) \le nQ\left(\frac{d}{2\sigma}\right) \le n\exp\left(-\frac{d^2}{8\sigma^2}\right), \tag{7.24}$$

when $Q(\cdot)$ denotes the Q-function. Let $d = O(\sqrt{N})$. For a fixed $n$, $P_e(\Lambda_c,\sigma^2)$ decays exponentially. In this case, the number of partition levels between $\Lambda_f$ and $\Lambda_c$ is $nO(\log N)$. We further let $\Lambda = \frac{1}{h_l}\Lambda_f$ and $\Lambda' = \frac{1}{h_s}\Lambda_c$. Check that $\mathfrak{h}(\Lambda_f,\sigma^2) = \mathfrak{h}\Big(\Lambda, \frac{\sigma^2}{h_l^2}\Big) + \log\big(V(\Lambda_f)/V(\Lambda)\big)$ and $P_e\Big(\Lambda', \frac{\sigma^2}{h_s^2}\Big) = P_e(\Lambda_c,\sigma^2)$, which means $\mathfrak{h}\Big(\Lambda, \frac{\sigma^2}{h_l^2}\Big) \approx \log V(\Lambda)$ and $P_e\Big(\Lambda', \frac{\sigma^2}{h_s^2}\Big) = e^{-O(N)}$. Therefore, criteria (a) and (b) are satisfied when $N \to \infty$. The number $r$ of levels between $\Lambda$ and $\Lambda'$ is given by

$$\begin{aligned} r &= \log\big(V(\Lambda')/V(\Lambda)\big) \\ &= \log\big(V(\Lambda_f)/V(\Lambda)\big) + \log\big(V(\Lambda_c)/V(\Lambda_f)\big) + \log\big(V(\Lambda')/V(\Lambda_c)\big) \\ &= n\log(h_l/h_s) + \log\big(V(\Lambda_c)/V(\Lambda_f)\big) \\ &= n\delta O(\log N). \end{aligned} \tag{7.25}$$

Moreover, according to (7.22), $E_h\big[P_e\big(\Lambda', \frac{\sigma^2}{h^2}\big)\big] = O\big(\frac{1}{N^{2\delta}}\big)$, and then $P_e(L,\sigma^2) = O\big(\frac{1}{N^{2\delta-1}}\big)$.

Let $R_{\mathcal{C}} = \sum_{\ell=1}^{r} R_\ell$ be the total rate of polar codes from level 1 to level $r$. Since

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

$$\begin{aligned} \log\left(\frac{\gamma_L(\sigma)}{2\pi e}\cdot\frac{2\sigma_h^2}{e^{\zeta}}\right) &= \log\left(\frac{V(L)^{\frac{2}{nN}}}{2\pi e\sigma^2}\cdot\frac{2\sigma_h^2}{e^{\zeta}}\right) \\ &= \log\left(\frac{2^{-\frac{2}{n}R_{\mathcal{C}}}V(\Lambda')^{\frac{2}{n}}}{2\pi e\sigma^2}\cdot\frac{2\sigma_h^2}{e^{\zeta}}\right) \\ &= -\frac{2}{n}R_{\mathcal{C}} + \frac{2}{n}\log V(\Lambda') - \log\left(2\pi e\sigma^2\frac{e^{\zeta}}{2\sigma_h^2}\right). \end{aligned} \tag{7.26}$$

Define

$$\begin{cases} \epsilon_1 = C_H(\Lambda, \sigma^2), \\ \epsilon_2 = nE_h\left[\mathfrak{h}\left(\frac{\sigma^2}{h^2}\right)\right] - E_h\left[\mathfrak{h}\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right], \\ \epsilon_3 = C_H(\Lambda/\Lambda', \sigma^2) - R_{\mathcal{C}} = \sum_{\ell=1}^{r} C_H(\Lambda_{\ell-1}/\Lambda_\ell, \sigma^2) - R_\ell, \end{cases} \tag{7.27}$$

We note that, $\epsilon_1 \geq 0$ represents the capacity of the mod-$\Lambda$ fading channel, $\epsilon_2 \geq 0$ due to the data processing theorem, and $\epsilon_3 \geq 0$ is the total capacity loss of component codes.

Then, we have

$$\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\cdot\frac{2\sigma_h^2}{e^{\zeta}}\right) = \frac{2}{n}(\epsilon_1 - \epsilon_2 + \epsilon_3). \tag{7.28}$$

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

$$\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\cdot\frac{2\sigma_h^2}{e^{\zeta}}\right) \leq \frac{2}{n}(\epsilon_1 + \epsilon_3). \tag{7.29}$$

By the design criteria (a)-(c), we have $\epsilon_1 \to 0$ and $\epsilon_3 \to 0$. Therefore, $\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\cdot\frac{2\sigma_h^2}{e^{\zeta}}\right) \to 0$, which represents the Poltyrev capacity. The right hand side of (7.29) gives an upper bound on the gap to the Poltyrev capacity of the ergodic fading channel. □

*Remark* 7.3.1. The slowly vanishing error probability $P_e(L,\sigma^2) = O\big(\frac{1}{N^{2\delta-1}}\big)$ is mainly caused by the uncoded error probability $E_h\left[P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right]$ associated with the bottom lattice $\Lambda'$. As we will see in the next section, a sub-exponentially vanishing error probability can be achieved when the power constraint is taken into consideration, because the probability of choosing a non-zero lattice point from $\Lambda'$ vanishes exponentially in the lattice Gaussian distribution.

### 7.3.2 Polar Lattices with Gaussian Shaping

In this subsection, we discuss the lattice Gaussian shaping for the polar lattices constructed for fading channels. It is well known that shaping is a source coding problem merely related to the chosen input distribution. For the case in which only the receiver knows CSI, the optimal input distribution for fading channels is the continuous Gaussian distribution [145], which is the same as that for AWGN channels. It has been shown in [20] that lattice Gaussian distribution preserves many properties of the continuous Gaussian distribution, including the ability of achieving the AWGN capacity. Therefore, the lattice Gaussian shaping technique proposed for the AWGN-good polar lattices in Chapter 4 can be applied to the fading channel with minor modification.

It has been proved in [20] that the lattice Gaussian distribution preserves the capacity of the AWGN channel when the associated flatness factor is negligible.

Motivated by Theorem 3.2.1, one may choose a low-dimensional $\Lambda$ such as $\mathbb{Z}$ and $\mathbb{Z}^2$ whose mutual information has a negligible gap to the AWGN channel capacity, and then construct polar lattices to achieve the capacity.

For the ergodic fading channel with power constraint $P$, letting the input $X$ be

Gaussian, the ergodic channel capacity is given by [145]

$$\begin{aligned} I(X;Y,H) &= E_h\left[\frac{1}{2}\log\left(1+\frac{Ph^2}{\sigma^2}\right)\right] \\ &= \frac{1}{2}\int_0^\infty \frac{h}{\sigma_h^2} e^{-\frac{h^2}{2\sigma_h^2}} \log\left(1+\frac{Ph^2}{\sigma^2}\right) dh \\ &= \frac{1}{2}\log e \int_{t=0}^\infty e^{-t} \ln\left(1+\frac{2\sigma_h^2 P}{\sigma^2}t\right) dt \\ &= \frac{1}{2}\log e \cdot \exp\left(\frac{\sigma^2}{2\sigma_h^2 P}\right) E_1\left(\frac{\sigma^2}{2\sigma_h^2 P}\right), \end{aligned} \tag{7.30}$$

where $\frac{1}{2}\log\left(1+\frac{Ph^2}{\sigma^2}\right)$ is the capacity of an AWGN channel with noise variance $\frac{\sigma^2}{h^2}$ and the same power constraint. To achieve the ergodic capacity, our strategy is to pick a lattice Gaussian distribution which is able to achieve the AWGN capacity $\frac{1}{2}\log\left(1+\frac{Ph^2}{\sigma^2}\right)$ for almost all possible $h$. For an instant Gaussian noise variance $\frac{\sigma^2}{h^2}$, the MMSE re-scaled noise in Theorem 3.2.1 is now a function of $h$ and has standard deviation $\tilde{\sigma}(h) = \frac{\sigma_s \sigma}{\sqrt{h^2\sigma_s^2+\sigma^2}}$. For a component lattice $\Lambda$, by Remark 3.2.1, $\epsilon_\Lambda\left(\tilde{\sigma}(h)\right)$ increases as $h$ grows. We can choose such that $\varepsilon = \epsilon_\Lambda\left(\tilde{\sigma}(h_l)\right) \to 0$ for a large $h_l$, then the resulted mutual information by $D_{\Lambda-c,\sigma_s}$ is lower-bounded as

$$\begin{aligned} E_h[I_D(h)] &= \int_{h=0}^{h_l} P_H(h) I_D(h) dh + \int_{h_l}^\infty P_H(h) I_D(h) dh \\ &\geq \int_{h=0}^{h_l} P_H(h)\left(\frac{1}{2}\log\left(1+\frac{Ph^2}{\sigma^2}\right) - \frac{5\varepsilon}{n}\right) dh \\ &\geq E_h\left[\frac{1}{2}\log\left(1+\frac{Ph^2}{\sigma^2}\right)\right] - \int_{h_l}^\infty \frac{1}{2} P_H(h)\frac{Ph^2}{\sigma^2} dh - \frac{5\varepsilon}{n}, \\ &= E_h\left[\frac{1}{2}\log\left(1+\frac{Ph^2}{\sigma^2}\right)\right] - \frac{1}{2}\left(\frac{h_l^2}{\sigma_h^2}+2\right)\frac{P\sigma_h^2}{\sigma^2} e^{-\frac{h_l^2}{2\sigma_h^2}} - \frac{5\varepsilon}{n}. \end{aligned} \tag{7.31}$$

Let $c = 0$ for simplicity. For sufficiently large $h_l$ and small $\varepsilon$, $D_{\Lambda,\sigma_s}$ is able to approach the ergodic capacity. Let the binary partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'/\cdots$ be labelled by bits $X_1, \cdots, X_r, \cdots$. Then, $D_{\Lambda,\sigma_s}$ induces a distribution $P_{X_{1:r}}$ whose limit corresponds to $D_{\Lambda,\sigma_s}$ as $r \to \infty$.

By the chain rule of mutual information

$$I(Y,H;X_{1:r})=\sum_{\ell=1}^{r}I(Y,H;X_\ell|X_{1:\ell-1}), \tag{7.32}$$

we obtain $r$ binary-input channels $W_\ell$ for $1\le\ell\le r$. Given $x_{1:\ell-1}$, denote by $\mathcal{A}_\ell(x_{1:\ell})$ the coset of $\Lambda_\ell$ indexed by $x_{1:\ell-1}$ and $x_\ell$. Similar to (4.16), the channel transition PDF of the $\ell$-th channel $W_\ell$ is written as

$$\begin{aligned}&P_{Y,H|X_\ell,X_{1:\ell-1}}(y,h|x_\ell,x_{1:\ell-1})\\&\qquad=\exp\left(-\frac{\|\frac{y}{h}\|^2}{2(\sigma_s^2+\frac{\sigma^2}{h^2})}\right)\frac{P_H(h)}{f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))}\frac{1}{2\pi\sigma\sigma_s}\\&\qquad\quad\sum_{a\in\mathcal{A}_\ell(x_{1:\ell})}\exp\left(-\frac{\|\alpha(h)y-a\|^2}{2\tilde{\sigma}^2(h)}\right),\end{aligned} \tag{7.33}$$

where $\alpha(h)=\frac{h\sigma_s^2}{h^2\sigma_s^2+\sigma^2}$ and $\tilde{\sigma}(h)=\frac{\sigma_s\sigma}{\sqrt{h^2\sigma_s^2+\sigma^2}}$ are the generalized MMSE coefficient and noise standard deviation. In general, $W_\ell$ is asymmetric, and we have to employ the polar coding technique for asymmetric channels [10] to achieve the capacity $I(Y,H;X_\ell|X_{1:\ell-1})$ of each level.

As shown in Chapter 4, the construction as well as the decoding of polar codes for a BMAC can be converted to that for a BMS channel by channel symmetrization (see [104, Lemma 7]). By replacing $Y$ with $(Y,H)$, Theorem 4.3.7 and Theorem 4.3.9 can be easily extended to our work. Therefore, the construction method of multilevel polar codes in Chapter 4 works for the fading case as well. Besides information bits and frozen bits at each level, we have shaping bits which are determined by the former two according to the lattice Gaussian distribution. Applying a similar argument as in Lemma 4.3.11, the symmetrized channel of $W_\ell$ at each level is equivalent to a $\Lambda_{\ell-1}/\Lambda_\ell$ fading channel. Consequently, the resultant polar codes for the symmetrized channels are sequentially nested by the analysis in Sect. 7.3.1, and hence we obtain a polar lattice $L$ which is Poltyrev capacity-achieving for the i.i.d.

fading channel. Moreover, the multistage decoding is performed on the MMSE-scaled signal $\alpha(h)y$ (cf. [104, Lemma 8]). Since the frozen sets of the polar codes are filled with random bits (but shared with the receiver), we actually obtain a coset $L + c'$ of the polar lattice, where the shift $c'$ accounts for the effects of all random frozen bits. Finally, since we start from $D_{\Lambda,\sigma_s}$, we would obtain $D_{\Lambda^N,\sigma_s}$ without coding; since $L + c' \subset \Lambda^N$ by construction, we obtain a discrete Gaussian distribution $D_{L+c',\sigma_s}$.

With regard to the number of partition levels, the same analysis given in Sect. 7.3.1 can be applied. By setting $h_l = O(N)$ and $\Lambda = \frac{1}{h_l}\Lambda_f$ for a fine lattice $\Lambda_f$, we have $\mathfrak{h}\big(\Lambda, \tilde{\sigma}^2(h_l)\big) \approx \log V(\Lambda)$ and hence $\epsilon_\Lambda(\tilde{\sigma}(h_l)) \to 0$ as $N \to \infty$ by the same argument of (7.21). Note that $\tilde{\sigma}(h_l) \to \frac{\sigma}{h_l}$ for large $h_l$. However, for the small channel gain $h_s$, we do not need $h_s = O\big(\frac{1}{N^\delta}\big)$ because of the lattice Gaussian shaping. To see this, let $h_s = 1$ and the bottom lattice $\Lambda' = \Lambda_c$ for a coarse lattice $\Lambda_c$. By the definition (3.19) of lattice Gaussian distribution, the probability of choosing a lattice point which is outside of $\mathcal{V}(\Lambda')$ is given by

$$\begin{aligned}
\sum_{\lambda\in\Lambda'\setminus\{0\}} D_{\Lambda',\sigma_s}(\lambda) &= \sum_{\lambda\in\Lambda_c\setminus\{0\}} D_{\Lambda_c,\sigma_s}(\lambda) \\
&= \sum_{\lambda\in\Lambda_c\setminus\{0\}} \frac{f_{\sigma_s}(\lambda)}{\sum_{\lambda'\in\Lambda_c} f_{\sigma_s}(\lambda')} \\
&\le \frac{\sum_{\lambda\in\Lambda_c\setminus\{0\}} f_{\sigma_s}(\lambda)}{f_{\sigma_s}(\lambda'=0)} \\
&\le (\sqrt{2\pi}\sigma_s)^n P_e(\Lambda_c, \sigma_s^2) \\
&\le n(\sqrt{2\pi}\sigma_s)^n Q\big(\frac{d}{2\sigma_s}\big) \\
&\le n(\sqrt{2\pi}\sigma_s)^n \exp\big(-\frac{d^2}{8\sigma_s^2}\big).
\end{aligned} \tag{7.34}$$

Recall that the minimum distance $d$ of $\Lambda_c$ scales as $d = O(\sqrt{N})$, and the second inequality satisfies for sufficiently large $d$ [1]. Then, $\sum_{\lambda\in\Lambda'\setminus\{0\}} D_{\Lambda',\sigma_s}(\lambda)$ vanishes

[1] Taking $n = 1$ for an example, it is easy to check that $\sum_{\lambda\in\Lambda_c\setminus\{0\}} f_{\sigma_s}(\lambda) \le 2\int_{d-1}^{\infty} f_{\sigma_s}(x)dx \le$

exponentially for a fixed $n$ and a sufficiently large $N$, which means that only one lattice point in $\mathcal{V}(\Lambda')$ is chosen with probability close to 1, and the lattice point from $\Lambda'$ can be directly decoded according to the lattice Gaussian distribution. Therefore, the uncoded error probability $E_h\left[P_e\left(\Lambda', \frac{\sigma^2}{h^2}\right)\right]$ associated with the bottom lattice $\Lambda'$ vanishes exponentially, and it can be ignored since the error probability of polar codes for each partition channel vanishes sub-exponentially. By the same argument of (7.25), the number of levels is given by

$$\begin{aligned} r &= \log\left(V(\Lambda')/V(\Lambda)\right) \\ &= n\log(h_l) + \log\left(V(\Lambda_c)/V(\Lambda_f)\right) \\ &= nO(\log N), \end{aligned} \tag{7.35}$$

which is sufficient to achieve the ergodic capacity.

We summarize our main result in the following theorem:

**Theorem 7.3.2.** *For a sufficiently large channel gain $h_l = O(N)$, choose a good constellation with negligible flatness factor $\epsilon_\Lambda(\tilde{\sigma}(h_l))$ and negligible $\epsilon_t$ as in Theorem 3.2.1, and construct a polar lattice with $r = nO(\log N)$ levels. Then, for i.i.d. fading channels, the message rate approaches the ergodic capacity $E_h\left[\frac{1}{2}\log\left(1+\frac{Ph^2}{\sigma^2}\right)\right]$, while the error probability under the multi-stage decoding is bounded by*

$$P_e \leq rN2^{-N^{\beta'}}, \quad 0 < \beta' < 0.5, \tag{7.36}$$

*as $N \to \infty$.*

*Proof.* The proof of Theorem 7.3.2 can be adapted from the proofs of Theorem 4.3.7 and Theorem 4.3.9 by replacing $Y$ with $(Y, H)$. □

$2\int_{\frac{d}{2}}^{\infty} f_{\sigma_s}(x)dx = P_e(\Lambda_c, \sigma_s^2)$ for a sufficiently large $d$. A similar argument holds for higher dimensions.

Basing on the union bound, the upper-bounds of the block error probability of polar lattices under the SC decoding are plotted in Fig. 7.8, where $\sigma_s = 3$, $\sigma = 1$, $\sigma_h = 1.2575$ and $N = 2^{10}, 2^{12}, ..., 2^{20}$. Here we choose the binary-partition chain $\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z}/8\mathbb{Z}/16\mathbb{Z}/32\mathbb{Z}$, and let $r = 5$. In this case, the ergodic capacity is $2.0967$, and the channel capacities from level $1$ to level $5$ are given by $0.1213$, $0.5105$, $0.8437$, $0.5859$ and $0.0307$, respectively. Note that the gap between the achievable rate and the ergodic capacity is smaller than 0.2 for a block error probability $10^{-5}$ when $N = 2^{20}$.

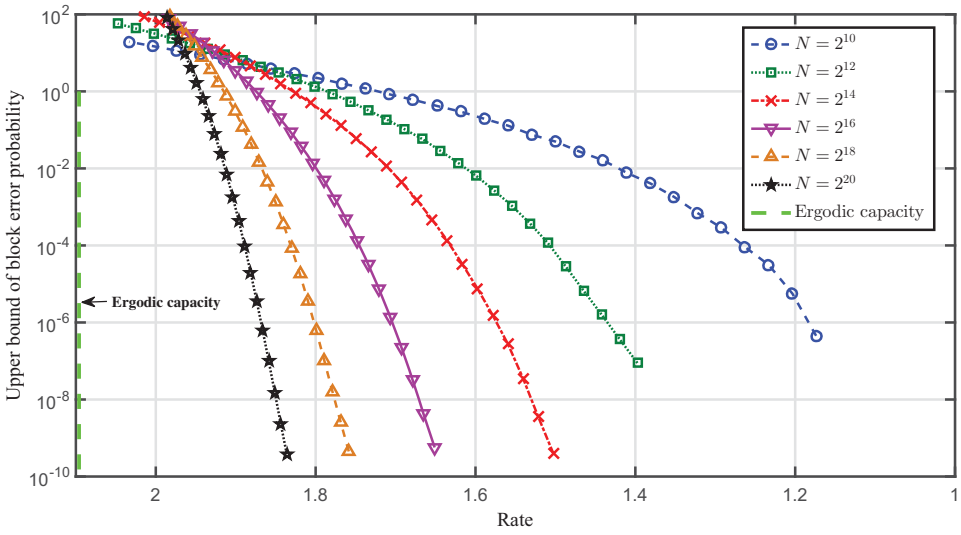



Figure 7.8: The upper-bounds of the block error probability of polar lattices under the SC decoding when $\sigma_s = 3$, $\sigma = 1$, $\sigma_h = 1.2575$ and $N = 2^{10}, 2^{12}, ..., 2^{20}$.

## 7.4 Summary

Explicit construction of polar codes and polar lattices for i.i.d. fading channels is proposed in this paper. By treating the channel gain as part of channel outputs, the work of polar codes and polar lattices for time-invariant channels is generalized to fading channels. We propose a simple construction of polar codes to achieve the ergodic capacity of binary-input i.i.d. fading channels when the CSI is not available

to the transmitter. Furthermore, polar codes are extended to polar lattices to achieve the ergodic capacity of i.i.d. fading channels with certain power constraint.

CHAPTER 8

# Conclusions and Future Work

Inheriting the versatility of polar codes, polar lattices show their capability of achieving the channel capacity over AWGN channels, the strong secrecy capacity over Gaussian wiretap channels and the rate-distortion bound for continuous Gaussian sources. The explicit structure of polar lattices provides us better understanding and more insights into the construction of good lattices. We summarize our points as follows.

① MMSE estimation: The construction of the capacity-good or quantization-good polar lattices gives us an evidence of how indispensable the MMSE estimation is for achieving the AWGN capacity or the Gaussian rate-distortion bound. Taking the capacity-good polar lattice as an example, for a given power $P$ and noise variance $\sigma_w^2$, we firstly construct an AWGN-good lattice for the MMSE-scaled noise variance $\tilde{\sigma}_w^2$ and then implement the shaping to achieve the capacity $\frac{1}{2}\log(1 + P/\sigma_w^2)$. From the perspective of the multilevel lattice construction, the MMSE estimation actually symmetrizes the channel at each level and the transition probability of the symmetrized channel is based on both the input distribution and the transition probability of the channel before symmetrization.

② Coding gain and shaping gain: From the work of Forney [50], we know that the AWGN-good lattices achieve the coding gain. In fact, the coding gain is achieved by maximizing the minimum distance of the lattice points. As to

the polar lattices investigated in this report, the frozen set $|\mathcal{F}_\ell|$ is playing the role of ensuring the coding gain. After obtaining the AWGN-good lattices, we still have to set up a boundary to satisfy the power constraint. The shaping set $\mathcal{S}_\ell$ is doing such job. Thanks to the work of [20], we are provided with the discrete lattice Gaussian distribution which can achieve the shaping gain and the AWGN capacity. By designing $\mathcal{S}_\ell$ according to the discrete Gaussian distribution, we find a practical and efficient way to implement the optimal shaping.

③ Lattice decoding: To decrease the decoding complexity, lattice decoding is usually used. This kind of decoding can be described as follows. Given a received vector, the decoder firstly finds its nearest lattice point over the AWGN-good lattice (see [50] and checks this can be done efficiently). Then the decoder checks whether this lattice point is in the shaping region. If it is, then this point is eligible; otherwise, we have to map this point into the shaping region by a modulo operation. It has been proved in [20] that the lattice decoding with MMSE scaling is equivalent to the MAP decoding. Note that our polar lattice decoding method can be viewed as a special MMSE lattice decoding which performs the mentioned two decoding steps simultaneously. In fact, some bits in $\mathcal{S}_\ell$ might be decoded earlier than those in $\mathcal{I}_\ell$.

④ Dither: It has been proved that dither is unnecessary to achieve the AWGN capacity or Gaussian rate distortion bound [20, 146]. Particularly, for the polar lattices, there always exists a specific choice on the frozen bits $u^{\mathcal{F}_\ell}$ at each level, making the resulted coset of polar lattice good for the AWGN capacity or quantization. The consequence of all these fixed frozen bits can be viewed as a certain shift on the polar lattice, which means that dither is theoretically not necessary. However, it could still be helpful in practice because we currently

cannot find out those $u^{\mathcal{F}_\ell}$. A uniformly random selection over $u^{\mathcal{F}_\ell}$ would still guarantee an asymptotically good performance on average.

⑤ Lattice wiretap coding: The work of polar lattices for Gaussian wiretap channels can be summarized as follows. Two polar lattices are constructed for Bob and Eve, respectively. The polar lattice for Bob is aiming to guarantee reliable transition, and the one for Eve is to obtain security, or to confuse Eve with random lattice points. Since the wiretapper's channel is degraded with respect to the main channel, the two polar lattices are nested, and they result in a lattice partition. Only the coset of this lattice partition is used to carry the secure message. Moreover, the nesting property allows the transmitter to perform lattice shaping of these two lattices simultaneously.

⑥ Lattice quantization: Similar to the construction of polar lattices for channel coding, the polar lattices for source quantization are designed level by level. However, the selections of the frozen set $|\mathcal{F}_\ell|$, the information set $|\mathcal{I}_\ell|$, and the shaping set $|\mathcal{S}_\ell|$ are different due to the unpolarized bits. The rate-approaching property still holds since the proportion of those unpolarized bits vanishes. Let $X$ and $Y$ be the input (reconstruction) and output (source) of the AWGN channel (test channel), respectively. The according rates $R_c$ and $R_q$ of polar lattices for channel coding and quantization satisfy $R_c \leq I(X;Y) \leq R_q$. When the lattice dimension goes to infinity, we have both limitations $R_c \to I(X;Y)$ and $R_q \to I(X;Y)$, which coincide with Shannon's channel coding and source coding theorems.

The future work will focus on the following directions.

- The practical performance of the capacity-good polar lattices (AWGN-good polar lattices with optimal shaping) and polar lattices for quantization: Recall that we have proved the asymptotical goodness of polar lattices for AWGN

channels and Gaussian sources, by combining the multi-stage decoding for lattice codes and the SC decoding for the component polar codes. To make polar lattices more competitive in practice, it is important to improve their finite-length performance. One way is to use more sophisticated decoding algorithms for the component polar code at each level. It has been reported in [144] that for short to moderate block lengths, the frame error rate performance of polar codes under SC decoding does not compete with other families of codes such as LDPC or Turbo codes. The authors then proposed a successive cancellation list-decoding (SCL), which significantly outperforms the simple SC decoding, and approaches the Maximum-Likelihood (ML) performance at high SNR region. Furthermore, when an outer cyclic redundancy check (CRC) code is concatenated, the SCL decoder can compete with LDPC codes. There are other promising decoding methods of polar codes, such as the BP decoding [39] and the linear programming (LP) decoding [40].

- Polar lattices for general continuous channels and sources: The majority of this thesis discusses the construction of polar lattices for AWGN channels or i.i.d. Gaussian sources. Although the application of polar lattices for i.i.d. fading channels is also discussed in Chapter 7, it can be viewed as a generalization of polar lattices to channels with additive white Gaussian noise whose variance is varying. A more interesting question is whether polar lattices can be generalized to channels with colored-Gaussian noise or continuous sources with memory. In [1, Chaptor 6], lattice codes with mismatched decoding and non-equiprobable signaling were proposed, providing a non-constructive solution to this problem. Finding an explicit construction of such lattices remains an open problem.

- The application of polar lattices for quantum information and quantum key

distribution (QKD): Since polar codes have already been generalized to quantum channels, the extension of polar lattices to quantum cases seems to be natural. In fact, many applications of classical polar codes have been generalized to their quantum cases in [147], such as the classical-quantum asymmetric channels, classical-quantum MAC channels, classical-quantum interference channels and classical-quantum broadcast channels. Compared with polar codes, polar lattices are able to provide more freedom of modulations in these quantum cases, as we have observed in classical AWGN channels. Another promising application of polar lattices is the quantum key distribution problem, which refers to secret key generation. Since polar lattices are able to achieve the secrecy capacity of Gaussian wiretap channels, we may also use them to solve the dual problem, i.e., secret key generation from Gaussian sources in the presence of an eavesdropper.

APPENDIX A

# Proof of Lemma 4.2.2

*Proof.* By the self-similarity of the lattice partition chain, we can scale a $\Lambda_{\ell-1}/\Lambda_\ell$ channel to a $\Lambda_\ell/\Lambda_{\ell+1}$ channel by multiplying the output of a $\Lambda_{\ell-1}/\Lambda_\ell$ channel with $T$. Since $T = \alpha V$ for some scale factor $\alpha > 0$ and orthogonal matrix $V$, the Gaussian noise for each dimension is still independent of each other and the noise variance per dimension is increased after the scaling. Therefore, a $\Lambda_{\ell-1}/\Lambda_\ell$ channel is stochastically equivalent to a $\Lambda_\ell/\Lambda_{\ell+1}$ channel with a larger Gaussian noise variance per dimension. For our design examples, a $\mathbb{Z}/2\mathbb{Z}$ channel with Gaussian noise variance $\sigma^2$ is equivalent to a $2\mathbb{Z}/4\mathbb{Z}$ channel with Gaussian noise variance $4\sigma^2$, and a $\mathbb{Z}^2/R\mathbb{Z}^2$ channel with noise variance $\sigma^2$ per dimension is equivalent to a $R\mathbb{Z}^2/2\mathbb{Z}^2$ channel with noise variance $2\sigma^2$ per dimension. Then our task is to prove that a $\Lambda_\ell/\Lambda_{\ell+1}$ channel with noise variance $\sigma_2^2$ is degraded with respect to a $\Lambda_\ell/\Lambda_{\ell+1}$ channel with noise variance $\sigma_1^2$ if $\sigma_1^2 \leq \sigma_2^2$.

To see the channel degradation, we construct an intermediate channel with input in $\mathcal{R}(\Lambda_{\ell+1})$ and a mod-$\Lambda_{\ell+1}$ operation at the receiver's front end. The noise variance of this mod-$\Lambda_{\ell+1}$ channel is given by $\sigma_2^2 - \sigma_1^2$ per dimension. By the property $[X+Y] \mod \Lambda_{\ell+1} = \Big[X \mod \Lambda_{\ell+1} + Y\Big] \mod \Lambda_{\ell+1}$, we can find that the concatenated channel that consists of a $\Lambda_\ell/\Lambda_{\ell+1}$ channel with noise variance $\sigma_1^2$ followed by the mentioned intermediate channel is stochastically equivalent to a $\Lambda_\ell/\Lambda_{\ell+1}$ channel with noise variance $\sigma_2^2$. This relationship is depicted in Fig. A.1. According to Definition 2.1.3, the proof is completed. Note that a more detailed proof can be found in [78]. □

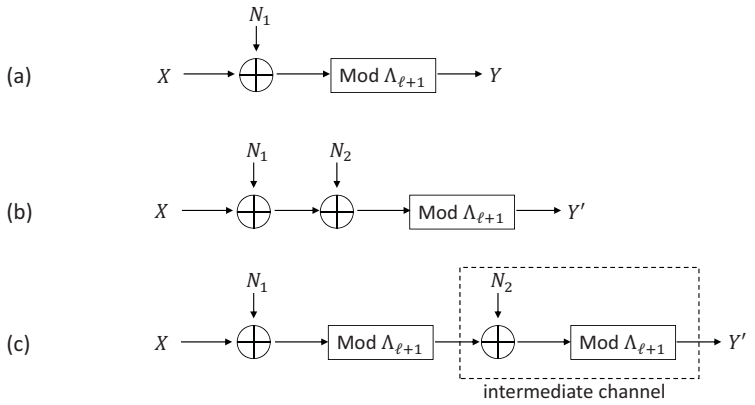



Figure A.1: Let $X \in \mathcal{R}(\Lambda_{\ell+1})$ denote the channel input. Let $N_1$ and $N_2$ denote two independent additive Gaussian noise with variances $\sigma_1^2$ and $\sigma_2^2 - \sigma_1^2$, respectively. Clearly, the two $\Lambda_\ell/\Lambda_{\ell+1}$ channels with noise variances $\sigma_1^2$ and $\sigma_2^2$ can be described by channel (a) and (b), respectively. By the property of modulo operation, channel (b) is equivalent to channel (c), which is a concatenated channel made by concatenating channel (a) with an intermediate mod-$\Lambda_{\ell+1}$ channel.

APPENDIX B

# Proof of Lemma 4.3.1

*Proof.* For convenience we consider a one-dimensional partition chain $\mathbb{Z}/2\mathbb{Z}/\cdots$. The proof can be extended to the multi-dimensional case by sandwiching the partition in $\mathbb{Z}^n/2\mathbb{Z}^n/\cdots$, which reduces to the one-dimensional case.

For level $r$, the selected coset $\mathcal{A}_r$ can be written 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 $\lambda_1$ and $\lambda_2$ 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 $\sigma_s^2$, we can find a positive integer $T$, making the probability

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

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 $\delta$, then $\lambda_1$ and $\lambda_2$ 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

$$\begin{aligned} P(\mathcal{A}_{r-1}) = \frac{f_{\sigma_s}(\mathcal{A}_{r-1}(x_{1:r-1}))}{f_{\sigma_s}(\mathbb{Z})} &< \frac{\frac{1}{\sqrt{2\pi}\sigma_s}\sum_{x\in 2^{r-1}\mathbb{Z}} \exp(-\frac{(x+\lambda_1)^2}{2\sigma_s^2})}{\frac{1}{\sqrt{2\pi}\sigma_s}} \\ &\leq 2\sum_{x\in 2^{r-1}\mathbb{Z}_-} \exp(-\frac{(x+\lambda_1)^2}{2\sigma_s^2}) \\ &\leq 2\frac{\exp(-\frac{\lambda_1^2}{2\sigma_s^2})}{1-\exp(-\frac{(2^{r-1})^2}{2\sigma_s^2})} \leq 2\frac{\exp(-\frac{T^2}{2})}{1-\exp(-\frac{9T^2}{2})}, \end{aligned}$$

where $\mathbb{Z}_-$ represents all non-positive integers. This means the probability of choosing $\mathcal{A}_{r-1}$ goes to zero when $T$ (or $N$) is large. Therefore, we have that the point $\lambda_1$ is in the interval $[-T\sigma_s, T\sigma_s]$ and $\lambda_2$ lies outside. Without loss of generality, we assume that the two cosets corresponding to $x_r = 0$ and $x_r = 1$ are $\lambda_1 + 2^r\mathbb{Z}$ and $\lambda_2 + 2^r\mathbb{Z}$, respectively. Then we have

$$\begin{aligned}\frac{P(x_r = 0|x_{1:r-1})}{P(x_r = 1|x_{1:r-1})} &= \frac{\sum\limits_{x\in 2^r\mathbb{Z}} \exp(-\frac{(x+\lambda_1)^2}{2\sigma_s^2})}{\sum\limits_{x\in 2^r\mathbb{Z}} \exp(-\frac{(x+\lambda_2)^2}{2\sigma_s^2})} \\ &\geq \frac{\exp(-\frac{\lambda_1^2}{2\sigma_s^2})}{2\sum\limits_{x\in 2^r\mathbb{Z}_+} \exp(-\frac{(x+\lambda_2)^2}{2\sigma_s^2})} \\ &\geq \frac{\exp(-\frac{\lambda_1^2}{2\sigma_s^2})}{2\cdot\exp(-\frac{\lambda_2^2}{2\sigma_s^2})}\left(1-\exp\left(-\frac{2^{2r}}{2\sigma_s^2}\right)\right),\end{aligned}$$

where $\mathbb{Z}_+$ represents all non-negative integers. Since $\lambda_2 - \lambda_1 = 2^{r-1} = 3T\sigma_s$ and $\lambda_2 + \lambda_1 \geq T\sigma_s$, for any $T > 1$, we can obtain

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

Assume that $\frac{1}{4}\exp(\frac{3}{2}\delta^2N^2) = M$, 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,

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

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

$$I(Y; X_r|X_{1:r-1}) \leq \log(e)\left(\frac{1}{\sqrt{M}} + \frac{1}{M}\right) = \log(e)\left(\frac{2}{\exp(\delta_1 2^{2r})} + \frac{4}{\exp(\delta_2 2^{2r})}\right),$$

where $\delta_1$ and $\delta_2$ are two positive constants. Therefore, when $r = O(\log\log N)$, we have $I(Y; X_r|X_{1:r-1}) \to 0$, and $\sum_{\ell \geq r} I(Y; X_\ell|X_{1:\ell-1}) \leq O(\frac{1}{N})$. □

APPENDIX C

# Proof of Theorem 4.3.7

*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 $i$-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 $u^{1:N}$ 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

$$\begin{aligned} E[P_e] = \sum_{u^{1:N},y^{1:N}} & 2^{-(|\mathcal{I}|+|\mathcal{F}|)}(\prod_{i\in\mathcal{S}} P_{U^i|U^{1:i-1}}(u^i|u^{1:i-1})) \\ & \cdot P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N})\mathbb{1}[(u^{1:N},y^{1:N})\in\mathcal{E}]. \end{aligned}$$

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

$$
\begin{aligned}
&2\|Q_{U^{1:N},Y^{1:N}} - P_{U^{1:N},Y^{1:N}}\| \\
&= \sum_{u^{1:N},y^{1:N}} |Q(u^{1:N},y^{1:N}) - P(u^{1:N},y^{1:N})| \\
&\overset{(a)}{=} \sum_{u^{1:N},y^{1:N}} |\sum_i (Q(u^i|u^{1:i-1}) - P(u^i|u^{1:i-1}))(\prod_{j=1}^{i-1} P(u^i|u^{1:i-1})) \\
&\qquad\qquad (\prod_{j=i+1}^{N} Q(u^i|u^{1:i-1}))Q(y^{1:N}|u^{1:N})| \\
&\leq \sum_{i\in\mathcal{I}\cup\mathcal{F}} \sum_{u^{1:N},y^{1:N}} |Q(u^i|u^{1:i-1}) - P(u^i|u^{1:i-1})|(\prod_{j=1}^{i-1} P(u^i|u^{1:i-1})) \\
&\qquad\qquad (\prod_{j=i+1}^{N} Q(u^i|u^{1:i-1}))Q(y^{1:N}|u^{1:N}) \\
&= \sum_{i\in\mathcal{I}\cup\mathcal{F}} \sum_{u^{1:i-1}} 2P(u^{1:i-1})\|Q_{U^i|U^{1:i-1}=u^{1:i-1}} - P_{U^i|U^{1:i-1}=u^{1:i-1}}\| \\
&\overset{(b)}{\leq} \sum_{i\in\mathcal{I}\cup\mathcal{F}} \sum_{u^{1:i-1}} P(u^{1:i-1})\sqrt{2\ln 2 D(P_{U^i|U^{1:i-1}=u^{1:i-1}}\|Q_{U^i|U^{1:i-1}=u^{1:i-1}})} \\
&\leq \sum_{i\in\mathcal{I}\cup\mathcal{F}} \sqrt{2\ln 2 \sum_{u^{1;i-1}} P(u^{1:i-1}) D(P_{U^i|U^{1:i-1}=u^{1:i-1}}\|Q_{U^i|U^{1:i-1}=u^{1:i-1}})} \\
&\leq \sum_{i\in\mathcal{I}\cup\mathcal{F}} \sqrt{2\ln 2 D(P_{U^i|U^{1:i-1}}||Q_{U^i|U^{1:i-1}})} \\
&\leq \sum_{i\in\mathcal{I}} \sqrt{2\ln 2(1 - H(U^i|U^{1:i-1}))} + \sum_{i\in\mathcal{F}} \sqrt{2\ln 2(1 - H(U^i|U^{1:i-1}))} \\
&\leq \sum_{i\in\mathcal{I}} \sqrt{2\ln 2(1 - Z(U^i|U^{1:i-1})^2)} + \sum_{i\in\mathcal{F}} \sqrt{2\ln 2(1 - Z(U^i|U^{1:i-1},Y^{1:N})^2)} \\
&\leq 2N\sqrt{4\ln 2 \cdot 2^{-N\beta}},
\end{aligned}
\tag{C.1}
$$

where equality $(a)$ follows from [10, 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

$$\begin{aligned} E[P_e] &= Q_{U^{1:N},Y^{1^N}}(\mathcal{E}) \\ &\leq \|Q_{U^{1:N},Y^{1:N}} - P_{U^{1:N},Y^{1:N}}\| + P_{U^{1:N},Y^{1:N}}(\mathcal{E}) \\ &\leq \|Q_{U^{1:N},Y^{1:N}} - P_{U^{1:N},Y^{1:N}}\| + \sum_{i\in\mathcal{I}} P_{U^{1:N},Y^{1:N}}(\mathcal{E}_i), \end{aligned} \tag{C.2}$$

where

$$\begin{aligned} P_{U^{1:N},Y^{1:N}}(\mathcal{E}_i) &\leq \sum_{u^{1:N},y^{1:N}} P(u^{1;i-1},y^{1:N})P(u^i|u^{1:i-1},y^{1:N}) \cdot \mathbb{1}[P(u^i|u^{1:i-1},y^{1:N}) \\ &\leq P(u^i\oplus 1|u^{1:i-1},y^{1:N})] \\ &\leq \sum_{u^{1:N},y^{1:N}} P(u^{1;i-1},y^{1:N})P(u^i|u^{1:i-1},y^{1:N})\sqrt{\frac{P(u^i\oplus 1|u^{1:i-1},y^{1:N})}{P(u^i|u^{1:i-1},y^{1:N})}} \\ &= Z(U^i|U^{1:i-1},Y^{1:N}) \leq 2^{-N^\beta}. \end{aligned} \tag{C.3}$$

From (C.1) and (C.2), we have $E[P_e] \leq 2N\sqrt{4\ln 2 \cdot 2^{-N^\beta}} + N2^{-N^\beta} = N2^{-N^{\beta'}}$ for any $\beta' < \beta < 0.5$.

□

APPENDIX D

# Proof of Theorem 4.3.9

*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

$$\begin{aligned} E[P_e] = \sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} & 2^{-(|\mathcal{I}_2|+|\mathcal{F}_2|)}\Big(\prod_{i\in\mathcal{S}_2} P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(u_2^i|u_2^{1:i-1}, x_1^{1:N})\Big) \\ & \cdot P_{Y^{1:N},X_1^{1:N}|U_2^{1:N}}(y^{1:N}, x_1^{1:N}|u_2^{1:N}) \mathbb{1}[(u_2^{1:N}, x_1^{1:N}, y^{1:N}) \in \mathcal{E}]. \end{aligned}$$

Now we define the probability distribution $Q_{U_2^{1:N},X_1^{1:N},Y^{1:N}}$ as

$$\begin{aligned} & Q_{U_2^{1:N},X_1^{1:N},Y^{1:N}}(u_2^{1:N}, x_1^{1:N}, y^{1:N}) \\ & \quad = 2^{-(|\mathcal{I}_2|+|\mathcal{F}_2|)} \cdot Q_{X_1^{1:N}}(x_1^{1:N})\Big(\prod_{i\in\mathcal{S}_2} P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(u_2^i|u_2^{1:i-1}, x_1^{1:N})\Big) \\ & \quad \cdot P_{Y^{1:N}|X_1^{1:N},U_2^{1:N}}(y^{1:N}|u_2^{1:N}, x_1^{1:N}). \end{aligned}$$

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

$$
\begin{aligned}
&2\|Q_{U_2^{1:N},X_1^{1:N},Y^{1:N}} - P_{U_2^{1:N},X_1^{1:N},Y^{1:N}}\| \\
&= \sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(u_2^{1:N},x_1^{1:N},y^{1:N}) - P(u_2^{1:N},x_1^{1:N},y^{1:N})| \\
&= \sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(u_2^{1:N}|x_1^{1:N})Q(x_1^{1:N})Q(y^{1:N}|u_2^{1:N},x_1^{1:N}) \\
&\quad - P(u_2^{1:N}|x_1^{1:N})P(x_1^{1:N})P(y^{1:N}|u_2^{1:N},x_1^{1:N})| \\
&\overset{(a)}{\leq} \sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(u_2^{1:N}|x_1^{1:N}) - P(u_2^{1:N}|x_1^{1:N})|P(x_1^{1:N})P(y^{1:N}|u_2^{1:N},x_1^{1:N}) \\
&\quad + \sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(x_1^{1:N}) - P(x_1^{1:N})|Q(u_2^{1:N}|x_1^{1:N})P(y^{1:N}|u_2^{1:N},x_1^{1:N})
\end{aligned}
$$

where inequation $(a)$ follows from [10, 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 4.3.7, we can prove that

$$
\begin{aligned}
&\sum_{u_2^{1:N},x_1^{1:N},y^{1:N}} |Q(u_2^{1:N}|x_1^{1:N}) - P(u_2^{1:N}|x_1^{1:N})|P(x_1^{1:N})P(y^{1:N}|u_2^{1:N},x_1^{1:N}) \\
&\qquad \leq 2N\sqrt{4\ln 2 \cdot 2^{-N\beta}}.
\end{aligned} \tag{D.1}
$$

According to the result of the coding scheme for level 1, we already have

$$
2\|Q_{U_1^{1:N},Y^{1:N}} - P_{U_1^{1:N},Y^{1:N}}\| \leq 2N\sqrt{4\ln 2 \cdot 2^{-N\beta}}. \tag{D.2}
$$

Since we have $P_{Y^{1:N}|U_1^{1:N}} = Q_{Y^{1:N}|U_1^{1:N}}$, we can write

$$
2\|Q_{U_1^{1:N}} - P_{U_1^{1:N}}\| \leq 2N\sqrt{4\ln 2 \cdot 2^{-N\beta}}. \tag{D.3}
$$

Clearly, there is a one to one mapping between $U_1^{1:N}$ and $X_1^{1:N}$, then we immediately have $2\|Q_{X_1^{1:N}} - P_{X_1^{1:N}}\| \leq 2N\sqrt{4\ln 2 \cdot 2^{-N\beta}}$. Therefore, for the second summation,

$$\begin{aligned}\sum_{u_2^{1:N}, x_1^{1:N}, y^{1:N}} & |Q(x_1^{1:N}) - P(x_1^{1:N})| Q(u_2^{1:N}|x_1^{1:N}) P(y^{1:N}|u_2^{1:N}, x_1^{1:N}) \\ &= \sum_{x_1^{1:N}} |Q(x_1^{1:N}) - P(x_1^{1:N})| \\ &\leq 2N\sqrt{4\ln 2 \cdot 2^{-N\beta}}.\end{aligned} \tag{D.4}$$

Then we have $||Q_{U_2^{1:N}, X_1^{1:N}, Y^{1^N}} - P_{U_2^{1:N}, X_1^{1:N}, Y^{1^N}}|| \leq 4N\sqrt{4\ln 2 \cdot 2^{-N\beta}}$, and

$$\begin{aligned}E[P_e] &= Q_{U_2^{1:N}, X_1^{1:N}, Y^{1^N}}(\mathcal{E}) \\ &\leq \|Q_{U_2^{1:N}, X_1^{1:N}, Y^{1^N}} - P_{U_2^{1:N}, X_1^{1:N}, Y^{1^N}}\| + P_{U_2^{1:N}, X_1^{1:N}, Y^{1^N}}(\mathcal{E}) \\ &\leq \|Q_{U_2^{1:N}, X_1^{1:N}, Y^{1^N}} - P_{U_2^{1:N}, X_1^{1:N}, Y^{1^N}}\| + \sum_{i \in \mathcal{I}} P_{U_2^{1:N}, X_1^{1:N}, Y^{1^N}}(\mathcal{E}_i),\end{aligned} \tag{D.5}$$

The rest part of the proof follows the same fashion of the proof of Theorem 4.3.7. Finally we have $E[P_e] \leq N2^{-N^{\beta'}}$ for any $\beta' < \beta < 0.5$. □

APPENDIX E

# Proof of Lemma 5.2.2

*Proof.* It is sufficient to demonstrate that channel $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ is degraded with respect to $W'(X_\ell; Z|X_{1:\ell-1})$ and $W'(X_\ell; Z|X_{1:\ell-1})$ is degraded with respect to $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ as well. To see this, we firstly construct a middle channel $\hat{W}$ from $Z \in \mathcal{V}(\Lambda_r)$ to $\bar{Z} \in \mathcal{V}(\Lambda_\ell)$. For a specific realization $\bar{z}$ of $\bar{Z}$, this $\hat{W}$ maps $\bar{z} + [\Lambda_\ell/\Lambda_r]$ to $\bar{z}$ with probability 1, where $[\Lambda_\ell/\Lambda_r]$ represents the set of the coset leaders of the partition $\Lambda_\ell/\Lambda_r$. Then we obtain channel $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ by concatenating $W'(X_\ell; Z|X_{1:\ell-1})$ and $\hat{W}$, which means $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ is degraded to $W'(X_\ell; Z|X_{1:\ell-1})$. Similarly, we can also construct a middle channel $\check{W}$ from $\bar{Z}$ to $Z$. For a specific realization $\bar{z}$ of $\bar{Z}$, this $\hat{W}$ maps $\bar{z}$ to $\bar{z} + [\Lambda_\ell/\Lambda_r]$ with probability $\frac{1}{|\Lambda_\ell/\Lambda_r|}$, where $|\Lambda_\ell/\Lambda_r|$ is the order of this partition. This means that $W'(X_\ell; Z|X_{1:\ell-1})$ is also degraded to $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$.

By channel degradation and [5, Lemma 1], letting channel $W$ and $W'$ denote $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ and $W'(X_\ell; Z|X_{1:\ell-1})$ for short, we have

$$\tilde{Z}(W_N^{(i)}) \leq \tilde{Z}(W'^{(i)}_N) \text{ and } \tilde{Z}(W_N^{(i)}) \geq \tilde{Z}(W'^{(i)}_N),$$
$$I(W_N^{(i)}) \leq I(W'^{(i)}_N) \quad \text{and } I(W_N^{(i)}) \geq I(W'^{(i)}_N),$$

meaning that $\tilde{Z}(W_N^{(i)}) = \tilde{Z}(W'^{(i)}_N)$ and $I(W_N^{(i)}) = I(W'^{(i)}_N)$. □

APPENDIX F

# Proof of Theorem 6.2.3

*Proof.* Firstly, we change the encoding rule for the $u_1^i$ in $i \in \mathcal{S}_1$ and (6.10) is modified to

$$u_1^i = \begin{cases} \bar{u}_1^i \quad \text{if} \quad i \in \mathcal{F}_1 \\ \begin{cases} 0 \quad \text{w. p.} \quad P_{U_1^i|U_1^{1:i-1}}(0|u_1^{1:i-1}) \\ 1 \quad \text{w. p} \quad P_{U_1^i|U_1^{1:i-1}}(1|u_1^{1:i-1}) \end{cases} \quad \text{if} \quad i \in \mathcal{S}_1. \end{cases} \tag{F.1}$$

Let $Q'_{U_1^{1:N},Y'^{1:N}}(u_1^{1:N}, y'^{1:N})$ denote the associate joint distribution for $U_1^{1:N}$ and $Y_1^{1:N}$ according to the encoding rule described in (6.9) and (F.1). Then the variational

distance between $P_{U^{1:N},Y^{1:N}}$ and $Q'_{U^{1:N},Y^{1:N}}$ can be bounded as follows.

$$
\begin{aligned}
&2\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q'_{U_1^{1:N},Y'^{1:N}}) \\
&= \sum_{u_1^{1:N},y'^{1:N}} |Q'(u_1^{1:N}, y'^{1:N}) - P(u_1^{1:N}, y'^{1:N})| \\
&\overset{(a)}{=} \sum_{u_1^{1:N},y'^{1:N}} |\sum_i (Q'(u_1^i|u_1^{1:i-1}, y'^{1:N}) - P(u_1^i|u_1^{1:i-1}, y'^{1:N}))(\prod_{j=1}^{i-1} P(u_1^i|u_1^{1:i-1}, y'^{1:N})) \\
&\qquad (\prod_{j=i+1}^{N} Q(u_1^i|u_1^{1:i-1}, y'^{1:N}))P(y'^{1:N})| \\
&\overset{(b)}{\leq} \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sum_{u_1^{1:N},y'^{1:N}} |Q(u_1^i|u_1^{1:i-1}, y'^{1:N}) - P(u_1^i|u_1^{1:i-1}, y'^{1:N})|(\prod_{j=1}^{i-1} P(u_1^i|u_1^{1:i-1}, y'^{1:N})) \\
&\qquad (\prod_{j=i+1}^{N} Q(u_1^i|u_1^{1:i-1}, y'^{1:N}))P(y'^{1:N}) \\
&= \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sum_{u_1^{1:i},y'^{1:N}} |Q(u_1^i|u_1^{1:i-1}, y'^{1:N}) - P(u_1^i|u_1^{1:i-1}, y'^{1:N})|(\prod_{j=1}^{i-1} P(u_1^i|u_1^{1:i-1}, y'^{1:N}))P(y'^{1:N}) \\
&= \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sum_{u_1^{1:i-1},y'^{1:N}} 2P(u_1^{1:i-1}, y'^{1:N})\mathbb{V}(Q_{U^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}}, P_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}}) \\
&\overset{(c)}{\leq} \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sum_{u_1^{1:i-1},y'^{1:N}} P(u_1^{1:i-1}, y'^{1:N})\sqrt{2\ln 2 D(P_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}}||Q_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}})} \\
&\overset{(d)}{\leq} \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sqrt{2\ln 2 \sum_{u_1^{1;i-1},y'^{1:N}} P(u_1^{1:i-1}, y'^{1:N}) D(P_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}}||Q_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}})} \\
&\leq \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sqrt{2\ln 2 D(P_{U_1^i}||Q_{U_1^i}|U_1^{1:i-1}, Y'^{1:N})} \\
&\overset{(e)}{\leq} \sum_{i\in\mathcal{F}_1} \sqrt{2\ln 2(1 - H(U_1^i|U_1^{1:i-1}, Y'^{1:N}))} + \sum_{i\in\mathcal{S}_1} \sqrt{2\ln 2(H(U_1^i|U_1^{1:i-1}) - H(U_1^i|U_1^{1:i-1}, Y'^{1:N}))} \\
&\overset{(f)}{\leq} \sum_{i\in\mathcal{F}_1} \sqrt{2\ln 2(1 - Z(U_1^i|U_1^{1:i-1}, Y'^{1:N})^2)} + \sum_{i\in\mathcal{S}_1} \sqrt{2\ln 2(Z(U_1^i|U_1^{1:i-1}) - Z(U_1^i|U_1^{1:i-1}, Y^{1:N})^2)} \\
&\overset{(g)}{\leq} 2N\sqrt{4\ln 2 \cdot 2^{-N^\beta}} = O(2^{-N^{\beta'}}),
\end{aligned}
$$

where $D(\cdot||\cdot)$ is the relative entropy, and the equalities and the inequalities follow from

$(a)$ The telescoping expansion [10].

$(b)$ $Q'(u_1^i|u_1^{1:i-1}, y'^{1:N}) = P(u_1^i|u_1^{1:i-1}, y'^{1:N})$ for $i \in \mathcal{I}_1$.

$(c)$ Pinsker's inequality.

$(d)$ Jensen's inequality.

$(e)$ $Q'(u_1^i|u_1^{1:i-1}) = \frac{1}{2}$ ($\bar{u}_1^i$ is uniformly random) for $i \in \mathcal{F}_1$ and $Q'_{U_1^i|U_1^{1:i-1},Y'^{1:N}} = P_{U_1^i|U_1^{1:i-1}}$

for $i \in \mathcal{S}_1$.

$(f)$ $Z(X|Y)^2 < H(X|Y) < Z(X|Y)$.

$(g)$ (6.8).

Following the same fashion,

$$\begin{aligned}
&2\mathbb{V}(Q'_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) \\
&\leq \sum_{i\in\mathcal{S}_1} \sqrt{2\ln 2 D(Q_{U_1^i}||Q'_{U^i}|U_1^{1:i-1}, Y'^{1:N})} \\
&\overset{(h)}{=} \sum_{i\in\mathcal{S}_1} \sqrt{2\ln 2 (H(U_1^i|U_1^{1:i-1}) - 0)} \\
&\leq \sum_{i\in\mathcal{S}_1} \sqrt{2\ln 2 Z(U_1^i|U_1^{1:i-1})} \leq N\sqrt{2\ln 2 \cdot 2^{-N^\beta}} = O(2^{-N^{\beta'}}),
\end{aligned}$$

where inequality $(h)$ follows from the MAP decision in (6.10) for $i \in \mathcal{S}_1$.

Finally, we have

$$\begin{aligned}
&\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) \\
&\leq \mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q'_{U_1^{1:N},Y'^{1:N}}) + \mathbb{V}(Q'_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) \\
&= O(2^{-N^{\beta'}}).
\end{aligned}$$

Clearly, when $N$ goes to infinity, for any $R > \frac{|\mathcal{I}_1|}{N} = I(X_1;Y')$, $\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}})$ is arbitrarily small. □

APPENDIX G

# Proof of Theorem 6.2.4

*Proof.* The variational distance can be upper bounded as follows.

$$
\begin{aligned}
&2\mathbb{V}(P_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}, Q_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}) \\
&= \sum_{u_2^{1:N},u_1^{1:N},y'^{1:N}} |Q(u_2^{1:N},u_1^{1:N},y'^{1:N}) - P(u_2^{1:N},u_1^{1:N},y'^{1:N})| \\
&= \sum_{u_2^{1:N},u_1^{1:N},y'^{1:N}} |P(u_2^{1:N}|u_1^{1:N},y'^{1:N})P(u_1^{1:N},y'^{1:N}) - Q(u_2^{1:N}|u_1^{1:N},y'^{1:N})Q(u_1^{1:N},y'^{1:N})| \\
&\le \sum_{u_2^{1:N},u_1^{1:N},y'^{1:N}} |P(u_2^{1:N}|u_1^{1:N},y'^{1:N}) - Q(u_2^{1:N}|u_1^{1:N},y'^{1:N})|P(u_1^{1:N},y'^{1:N}) \\
&\qquad + \sum_{u_2^{1:N},u_1^{1:N},y'^{1:N}} |P(u_1^{1:N},y'^{1:N}) - Q(u_1^{1:N},y'^{1:N})|Q(u_2^{1:N}|u_1^{1:N},y'^{1:N}).
\end{aligned} \tag{G.1}
$$

Treating $(U_1^{1:N}, Y'^{1:N})$ as a new source with distribution $P(u_1^{1:N}, y'^{1:N})$, the first summation can be proved to be $O(2^{-N^{\beta'}})$ in the same fashion as the proof of Theorem 6.2.3. For the second summation, we have

$$
\begin{aligned}
&\sum_{u_{1:2}^{1:N},y'^{1:N}} |P(u_1^{1:N},y'^{1:N}) - Q(u_1^{1:N},y'^{1:N})|Q(u_2^{1:N}|u_1^{1:N},y'^{1:N}) \\
&= \sum_{u_1^{1:N},y'^{1:N}} |P(u_1^{1:N},y'^{1:N}) - Q(u_1^{1:N},y'^{1:N})| \\
&= 2\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) = O(2^{-N^{\beta'}}).
\end{aligned} \tag{G.2}
$$

Finally,

$$
\mathbb{V}(P_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}, Q_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}) = O(2 \cdot 2^{-N^{\beta'}}).
$$

□

APPENDIX H

# Proof of Theorem 6.2.5

*Proof.* Firstly, for the source $Y'$, we consider the average performance of the multi-level polar codes with all possible choice of $u_\ell^{\mathcal{F}_\ell}$ at each level. If the encoding rule described in the form of (6.13) is used for all $i \in [N]$ at each level, the resulted average distortion is given by

$$D_{P,Y'} = \frac{1}{N} \sum_{u_{1:r}^{1:N}, y'^{1:N}} P_{U_{1:r}^{1:N}, Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N}) d(y'^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)),$$

where $\mathcal{M}(u_{1:r}^{1:N} G_N)$ denotes a mapping from $u_{1:r}^{1:N}$ to $x^{1:N}$ according to (6.31) (remind that $x$ is drawn from $\Lambda$ according to $D_{\Lambda,\sigma_r}$). For instance, let $\Lambda = \mathbb{Z}$ and the partition is given by $\mathbb{Z}/2\mathbb{Z}/...2^r\mathbb{Z}$, then $x^{1:N} = x_1^{1:N} + 2x_2^{1:N} + ... + 2^{r-1}x_r^{1:N}$ and $x_\ell^{1:N} = u_\ell^{1:N} G_N$. When $r \to \infty$, there exists a one-to-one mapping from $u_{1:r}^{1:N}$ to $x^{1:N}$. Then we have

$$\begin{aligned}
D_{P,Y'} &= \frac{1}{N} \sum_{x^{1:N}, y'^{1:N}} P_{X^{1:N}, Y'^{1:N}}(x^{1:N}, y'^{1:N}) d(y'^{1:N}, x^{1:N}) \\
&= \frac{1}{N} \cdot N \sum_{x,y'} P_{X,Y'}(x, y') d(x, y') \\
&= \sum_{x \in \Lambda} P_X(x) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Delta}} \exp(-\frac{(y'-x)^2}{2\Delta})(y'-x)^2 dy' \\
&= \Delta.
\end{aligned}$$

The result $D_{P,Y'} = \Delta$ is reasonable since the encoder does not do any compression. If we replace $P_{U_{1:r}^{1:N}, Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N})$ with $Q_{U_{1:r}^{1:N}, Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N})$ and com-

press $y'^{1:N}$ to $u_\ell^{\mathcal{I}_\ell}$ at each level, the resulted average distortion $D_{Q,Y'}$ can be bounded as

$$\begin{aligned} D_{Q,Y'} &= \frac{1}{N} \sum_{u_{1:r}^{1:N}, y'^{1:N}} Q_{U_{1:r}^{1:N}, Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N}) d(y'^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)) \\ &\leq \frac{1}{N} \sum_{u_{1:r}^{1:N}, y'^{1:N}} P_{U_{1:r}^{1:N}, Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N}) d(y'^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)) \\ &\quad + \frac{1}{N} \sum_{u_{1:r}^{1:N}, y'^{1:N}} |P_{U_{1:r}^{1:N}, Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N}) - Q_{U_{1:r}^{1:N}, Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N})| d(y'^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)) \end{aligned}$$

Since the densities of both $Y'$ and $X$ decrease exponentially to their square norms, the distortion caused by large $x$ or $y'$ is negligible, we can always assume a maximum distortion $d_{\max}$ between $y'$ and $x$. Then we have

$$\begin{aligned} D_{Q,Y'} &\leq D_{P,Y'} + \frac{2}{N} \mathbb{V}(P_{U_{1:r}^{1:N}, Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N}), Q_{U_{1:r}^{1:N}, Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N})) \cdot N d_{\max} \\ &= \Delta + O(2^{-N^{\beta'}}), \end{aligned}$$

where the equation follows from (6.16) and $r = O(\log \log N)$ [104, Lemma 5].

Now we consider using the same encoder to quantize the Gaussian source $Y$. The resulted average distortion $D_{Q,Y}$ can be written as

$$\begin{aligned} D_{Q,Y} &= \frac{1}{N} \sum_{u_{1:r}^{1:N}, y^{1:N}} Q_{U_{1:r}^{1:N}, Y^{1:N}}(u_{1:r}^{1:N}, y^{1:N}) d(y^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)) \\ &= \frac{1}{N} \sum_{u_{1:r}^{1:N}, y^{1:N}} P_{Y^{1:N}}(y^{1:N}) Q_{U_{1:r}^{1:N} | Y^{1:N}}(u_{1:r}^{1:N} | y^{1:N}) \cdot d(y^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)). \end{aligned}$$

Since the same encoder is used, for a same realization $y^{1:N}$, we have

$Q_{U_{1:r}^{1:N}|Y^{1:N}}(u_{1:r}^{1:N}|y^{1:N}) = Q_{U_{1:r}^{1:N}|Y'^{1:N}}(u_{1:r}^{1:N}|y^{1:N})$, and hence

$$\begin{aligned}
&D_{Q,Y} - D_{Q,Y'} \\
&= \frac{1}{N} \sum_{u_{1:r}^{1:N}, y^{1:N}} (P_{Y^{1:N}}(y^{1:N}) - P_{Y'^{1:N}}(y^{1:N})) \cdot Q_{U_{1:r}^{1:N}|Y^{1:N}}(u_{1:r}^{1:N}|y^{1:N}) d(y^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)) \\
&\leq \frac{1}{N} N d_{\max} \sum_{y^{1:N}} |P_{Y^{1:N}}(y^{1:N}) - P_{Y'^{1:N}}(y^{1:N})|.
\end{aligned}$$

Again, by the telescoping expansion,

$$\begin{aligned}
&\sum_{y^{1:N}} |P_{Y^{1:N}}(y^{1:N}) - P_{Y'^{1:N}}(y^{1:N})| \\
&= \sum_{y^{1:N}} \sum_{i=1}^{N} |P_{Y^i}(y^i) - P_{Y'^i}(y^i)| P_{Y^{1:i-1}}(y^{1:i-1}) P_{Y'^{i+1:N}}(y^{i+1:N}) \\
&= \sum_{i=1}^{N} \sum_{y^i} |P_{Y^i}(y^i) - P_{Y'^i}(y^i)| \\
&\underset{Lemma\,6.2.2}{\leq} N \cdot 8\epsilon_\Lambda(\tilde{\sigma}_\Delta).
\end{aligned}$$

As a result,

$$D_{Q,Y} \leq \Delta + O(2^{-N^{\beta'}}) + 8\epsilon_\Lambda(\tilde{\sigma}_\Delta) d_{\max} N. \tag{H.1}$$

By scaling $\Lambda$, we can make $\epsilon_\Lambda(\tilde{\sigma}_\Delta) \ll \frac{1}{8 d_{\max} N}$, and $D_{Q,Y}$ can be arbitrarily close to $\Delta$ with $R > I(X;Y') \geq \frac{1}{2}\log\frac{\sigma_s^2}{\Delta} - \frac{5\epsilon_\Lambda(\tilde{\sigma}_\Delta)}{n}$ ($n$ could be 1). When $\epsilon_\Lambda(\tilde{\sigma}_\Delta) \to 0$, we have $I(X;Y') \to \frac{1}{2}\log\frac{\sigma_s^2}{\Delta}$ and $R > \frac{1}{2}\log\frac{\sigma_s^2}{\Delta}$.

Now it is ready to explain the lattice structure. From the definition of $\mathcal{F}_\ell$ and [104, Lemma 6], it is easy to find that $\mathcal{F}_\ell \subseteq \mathcal{F}_{\ell-1}$ for $1 < \ell \leq r$. When $u_\ell^{\mathcal{S}_\ell}$ is uniformly selected and $u_\ell^{\mathcal{F}_\ell} = 0$ at each level, the constructed polar code at level $\ell - 1$ is a subset of the polar code at level $\ell$. Therefore, the resulted multilevel code is actually a polar lattice and the MAP decision on the bits in $\mathcal{S}_\ell$ is a shaping operation

according to $D_{\Lambda,\sigma_r}$. Moreover, since $D_{Q,Y}$ is an average distortion over all random choices of $u_\ell^{\mathcal{F}_\ell}$, there exists at least one specific choice of $u_\ell^{\mathcal{F}_\ell}$ at each level making the average distortion satisfying (H.1). This is exactly a shift on the constructed polar lattice. Consequently, the shifted polar lattice achieves the rate-distortion bound of the Gaussian source. □

APPENDIX I

# Proof of Theorem 6.4.4

*Proof.* We firstly show that the target distortion can be achieved. Recall $U_\ell^{1:N} = A_\ell^{1:N} G_N$ for each level $\ell$. Let $P_{U_{1:r}^{1:N}, \bar{X}^{1:N}}$ denote the joint distribution between $U_{1:r}^{1:N}$ and $\bar{X}_{1:r}^{1:N}$ when the encoder performs no compression at each level, i.e., the encoder applies encoding rule (6.9) for all indices $i \in [N]$ at level 1, encoding rule (6.13) for all $i \in [N]$ at level 2 and similar rules for higher levels, with the notation $X$ and $Y'$ being replaced by $A$ and $\bar{X}$, respectively. Let $Q_{U_{1:r}^{1:N}, \bar{X}^{1:N}}$ denote the joint distribution when only $U_\ell^{\mathcal{I}_\ell^Q}$ is recorded following the random rounding rule at each level. $U_\ell^{\mathcal{F}_\ell^Q}$ is a uniformly random sequence shared between the encoder and decoder, and $U_\ell^{\mathcal{S}_\ell^Q}$ is determined according to the MAP rule (see (6.10) and (6.14)). As illustrated in (6.16),

$$\mathbb{V}(P_{U_{1:r}^{1:N}, \bar{X}^{1:N}}, Q_{U_{1:r}^{1:N}, \bar{X}^{1:N}}) = O(r \cdot 2^{-N^{\beta'}}). \tag{I.1}$$

Since $r = O(\log N)$, we can write $\mathbb{V}(P_{U_{1:r}^{1:N}, \bar{X}^{1:N}}, Q_{U_{1:r}^{1:N}, \bar{X}^{1:N}}) = O(2^{-N^{\beta'}})$. When quantization is performed for the source $X$, let $Q_{U_{1:r}^{1:N}, X^{1:N}}$ denote the resulted joint distribution. By Lemma 6.2.2 again,

$$\begin{aligned}
&\sum_{u_{1:r}^{1:N}, x^{1:N}} |Q_{U_{1:r}^{1:N}, \bar{X}^{1:N}}(u_{1:r}^{1:N}, x^{1:N}) - Q_{U_{1:r}^{1:N}, X^{1:N}}(u_{1:r}^{1:N}, x^{1:N})| \\
&= \sum_{x^{1:N}} |P_{\bar{X}^{1:N}}(x^{1:N}) - P_{X^{1:N}}(x^{1:N})| \sum_{u_{1:r}^{1:N}} Q_{U_{1:r}^{1:N}|X^{1:N}}(u_{1:r}^{1:N}|x^{1:N}) \\
&= \sum_{x^{1:N}} |P_{\bar{X}^{1:N}}(x^{1:N}) - P_{X^{1:N}}(x^{1:N})| \leq N \cdot 8\epsilon_\Lambda(\tilde{\sigma}_q).
\end{aligned}$$

It has been shown that $\epsilon_\Lambda(\tilde{\sigma}_q) = O(2^{-\sqrt{N}})$, we further have

$$\begin{aligned}
&\mathbb{V}(P_{U_{1:r}^{1:N},\bar{X}^{1:N}}, Q_{U_{1:r}^{1:N},X^{1:N}}) \\
&\qquad \leq \mathbb{V}(P_{U_{1:r}^{1:N},\bar{X}^{1:N}}, Q_{U_{1:r}^{1:N},\bar{X}^{1:N}}) + \mathbb{V}(Q_{U_{1:r}^{1:N},\bar{X}^{1:N}}, Q_{U_{1:r}^{1:N},X^{1:N}}) \\
&\qquad = O(2^{-N^{\beta'}}) + O(2^{-\sqrt{N}}) \\
&\qquad = O(2^{-N^{\beta'}}).
\end{aligned} \tag{I.2}$$

As mentioned in Section 6.4, the encoder only sends $U_\ell^{d\mathcal{I}_\ell}$ to the decoder, which then utilizes the side information to recover $U_\ell^{\mathcal{I}_\ell^C}$. Here we assume that $U_\ell^{\mathcal{I}_\ell^C}$ can be correctly decoded and $U_\ell^{\mathcal{S}_\ell^Q}$ is recovered according to the MAP rule. In this case, the decoder enjoys the same joint distribution $Q_{U_{1:r}^{1:N},X^{1:N}}$ as the encoder does. Recall that $B = \frac{\alpha_q}{\alpha_c} Y$ and $\bar{B} = \frac{\alpha_q}{\alpha_c}\bar{Y}$. Let $Q_{U_{1:r}^{1:N},X^{1:N},B^{1:N}}$ denote the resulted joint distribution of $U_{1:r}^{1:N}$, $X^{1:N}$, and $B^{1:N}$ when the encoder performs compression, i.e., compresses $X^{1:N}$ to $U_\ell^{\mathcal{I}_\ell^Q}$ at each level. Let $P_{U_{1:r}^{1:N},\bar{X}^{1:N},\bar{B}^{1:N}}$ denote the resulted joint distribution of $U_{1:r}^{1:N}$, $\bar{X}^{1:N}$, and $\bar{B}^{1:N}$ when the encoder performs no compression for $\bar{X}^{1:N}$.

$$\begin{aligned}
&2\mathbb{V}(P_{U_{1:r}^{1:N},\bar{X}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},X^{1:N},B^{1:N}}) \\
&= \sum_{u_{1:r}^{1:N},x^{1:N},b^{1:N}} |P(u_{1:r}^{1:N}, x_{1:r}^{1:N}, b^{1:N}) - Q(u_{1:r}^{1:N}, x_{1:r}^{1:N}, b^{1:N})| \\
&= \sum_{u_{1:r}^{1:N},x^{1:N},b^{1:N}} |P(u_{1:r}^{1:N}, x_{1:r}^{1:N}) P(b^{1:N}|u_{1:r}^{1:N}, x_{1:r}^{1:N}) - Q(u_{1:r}^{1:N}, x_{1:r}^{1:N}) Q(b^{1:N}|u_{1:r}^{1:N}, x_{1:r}^{1:N})|.
\end{aligned} \tag{I.3}$$

According to Fig. 6.7, $\alpha_q X' \to X \to B$ and $A \to \bar{X} \to \bar{B}$ are two Markov chains. We have

$$P(b^{1:N}|u_{1:r}^{1:N}, x_{1:r}^{1:N}) = Q(b^{1:N}|u_{1:r}^{1:N}, x_{1:r}^{1:N}) \tag{I.4}$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_{z'}^2}} \exp(-\frac{(b^i - x^i)^2}{2\sigma_{z'}^2}). \tag{I.5}$$

Therefore,

$$\mathbb{V}(P_{U_{1:r}^{1:N},\bar{X}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},X^{1:N},B^{1:N}}) = \mathbb{V}(P_{U_{1:r}^{1:N},\bar{X}^{1:N}}, Q_{U_{1:r}^{1:N},X^{1:N}}) = O(2^{-N^{\beta'}}). \quad \text{(I.6)}$$

Recall that the reconstruction of $\bar{X}$ is given by $\check{X} = A + \gamma(\bar{B} - A)$. The average distortion $\Delta_P$ caused by $P_{U_{1:r}^{1:N},\bar{X}^{1:N},\bar{B}^{1:N}}$ can be expressed as

$$\Delta_P = \frac{1}{N} \sum_{u_{1:r}^{1:N},x^{1:N},b^{1:N}} P_{U_{1:r}^{1:N},\bar{X}^{1:N},\bar{B}^{1:N}}(u_{1:r}^{1:N}, x^{1:N}, b^{1:N}) d(x^{1:N}, \check{x}^{1:N}), \quad \text{(I.7)}$$

where $\check{x}^{1:N} = \gamma b^{1:N} + (1-\gamma)\mathcal{M}(u_{1:r}^{1:N} G_N)$, where $\mathcal{M}(u_{1:r}^{1:N} G_N)$ is a mapping from $u_{1:r}^{1:N}$ to $a^{1:N}$ according to the lattice Gaussian distribution. Clearly, given $u_{1:r}^{1:N}$, there is a one-to-one mapping between $b^{1:N}$ and $\check{x}^{1:N}$ when $r$ is sufficiently large. Thus, $\Delta_P$ can be written as

$$\begin{aligned}
\Delta_P &= \frac{1}{N} \sum_{u_{1:r}^{1:N},\check{x}^{1:N},x^{1:N}} P_{U_{1:r}^{1:N},\check{X}^{1:N},\bar{X}^{1:N}}(u_{1:r}^{1:N}, \check{x}^{1:N}, x^{1:N},) d(\check{x}^{1:N}, x^{1:N}) \\
&= \frac{1}{N} \sum_{\check{x}^{1:N},x^{1:N}} P_{\check{X}^{1:N},\bar{X}^{1:N}}(\check{x}^{1:N}, x^{1:N}) d(\check{x}^{1:N}, x^{1:N}) \\
&= \frac{1}{N} \cdot N \sum_{\check{x},x} P_{\check{X},\bar{X}}(\check{x}, x) d(\check{x}, x) \\
&= \int_{\check{x}} f_{\check{X}}(\check{x}) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Delta}} \exp(-\frac{(x-\check{x})^2}{2\Delta})(x-\check{x})^2 dx d\check{x} \\
&= \Delta.
\end{aligned}$$

The expected distortion $\Delta_Q$ achieved by $Q_{U_{1:r}^{1:N},X^{1:N},B^{1:N}}$ satisfies

$$\begin{aligned}
\Delta_Q - \Delta_P &= \frac{1}{N} \sum_{u_{1:r}^{1:N},x^{1:N},b^{1:N}} (P_{U_{1:r}^{1:N},\bar{X}^{1:N},\bar{B}^{1:N}} - Q_{U_{1:r}^{1:N},X^{1:N},B^{1:N}}) d(\check{x}^{1:N}, x^{1:N}) \\
&\leq \frac{1}{N} N d_{\max} \sum_{y^{1:N}} |P_{U_{1:r}^{1:N},\bar{X}^{1:N},\bar{B}^{1:N}} - Q_{U_{1:r}^{1:N},X^{1:N},B^{1:N}}| \\
&= O(2^{-N^{\beta'}}).
\end{aligned}$$

Now we show that the decoder is able to decode $U_\ell^{\mathcal{I}_\ell^C}$ with vanishing error probability.

$$\begin{aligned}
2\mathbb{V}(P_{U_{1:r}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},B^{1:N}}) &= \sum_{u_{1:r}^{1:N},b^{1:N}} |P(u_{1:r}^{1:N},b^{1:N}) - Q(u_{1:r}^{1:N},b^{1:N})| \\
&= \sum_{u_{1:r}^{1:N},b^{1:N}} |\sum_{x^{1:N}} \left[P(u_{1:r}^{1:N},x^{1:N},b^{1:N}) - Q(u_{1:r}^{1:N},x^{1:N},b^{1:N})\right]| \\
&\leq \sum_{u_{1:r}^{1:N},b^{1:N}} \sum_{x^{1:N}} |P_{U_{1:r}^{1:N},\bar{X}^{1:N},\bar{B}^{1:N}} - Q_{U_{1:r}^{1:N},X^{1:N},B^{1:N}}| \\
&= O(2^{-N^{\beta'}}). \qquad \text{(I.8)}
\end{aligned}$$

By the result of [104], $P_{U_{1:r}^{1:N},\bar{B}^{1:N}}$ results in an expectation of error probability $E_P^\ell[P_e]$ at each level such that $E_P^\ell[P_e] = O(2^{-N^{\beta'}})$. To see this, let $\mathcal{E}_i$ denote the set of pairs of $u_\ell^{1:N}$ and $b^{1:N}$ such that the SC decoding error occurs at the $i$th bit for level $\ell$, then the block decoding error event is given by $\mathcal{E}^\ell \equiv \bigcup_{i\in\mathcal{I}_\ell^C} \mathcal{E}_i$. Then the expectation of decoding error probability over all random mapping is expressed as

$$\begin{aligned}
E_P^\ell[P_e] &= \sum_{u_{1:\ell}^{1:N},b^{1:N}} P_{U_{1:\ell}^{1:N},\bar{B}^{1:N}}(u_{1:\ell}^{1:N},b^{1:N}) \mathbb{1}[(u_\ell^{1:N},b^{1:N}) \in \mathcal{E}^\ell] \\
&\leq \sum_{i\in\mathcal{I}_\ell^C\cup\mathcal{S}_\ell^Q} \sum_{u_{1:\ell}^{1:N},b^{1:N}} P_{U_{1:\ell}^{1:N},\bar{B}^{1:N}}(u_{1:\ell}^{1:N},b^{1:N}) \mathbb{1}[(u_\ell^{1:N},b^{1:N}) \in \mathcal{E}_i] \\
&\leq \sum_{i\in\mathcal{I}_\ell^C\cup\mathcal{S}_\ell^Q} \sum_{u_\ell^{1:i},u_{1:\ell-1}^{1:N},b^{1:N}} P(u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N}) P(u_\ell^i|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N}) \\
&\qquad \cdot \mathbb{1}[P(u_\ell^i|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N}) \leq P(u_\ell^i\oplus 1|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N})] \qquad \text{(I.9)} \\
&\leq \sum_{i\in\mathcal{I}_\ell^C\cup\mathcal{S}_\ell^Q} \sum_{u_\ell^{1:i},u_{1:\ell-1}^{1:N},b^{1:N}} P(u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N}) P(u_\ell^i|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N}) \\
&\qquad \sqrt{\frac{P(u_\ell^i\oplus 1|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N})}{P(u_\ell^i|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N})}} \\
&\leq N\cdot Z(U_\ell^i|U_\ell^{1:i-1},A_{1:\ell-1}^{1:N},\bar{B}^{1:N}) \\
&= O(2^{-N^{\beta'}}).
\end{aligned}$$

Then by the union bound, we immediately obtain that the expectation of multi-stage decoding error probability for the polar lattice $E_P[P_e] = O(2^{-N^{\beta'}})$. Let $P_e^{WZ}$ denote the expectation of error probability caused by $Q_{U_{1:r}^{1:N},B^{1:N}}$, i.e., it is an average error probability over all choices of the frozen bits $U_\ell^{\mathcal{F}_\ell^C}$ and shaping bits $U_\ell^{d\mathcal{S}_\ell}$ for each level. Let $\mathcal{E}$ denote the set of the pairs $(u_{1:r}^{1:N}, b^{1:N})$ such that a lattice decoding error occurs. We have

$$\begin{aligned} P_e^{WZ} - E_P[P_e] &= \sum_{u_{1:\ell}^{1:N},b^{1:N}} (P(u_{1:\ell}^{1:N}, b^{1:N}) - Q(u_{1:\ell}^{1:N}, b^{1:N})) \cdot \mathbb{1}[(u_{1:r}^{1:N}, b^{1:N}) \in \mathcal{E}] \\ &\le 2\mathbb{V}(P_{U_{1:r}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},B^{1:N}}) \qquad \text{(I.10)} \\ &\le O(2^{-N^{\beta'}}). \end{aligned}$$

With regard to the data rate, we have

$$\sum_{\ell=1}^{r} \frac{|\mathcal{I}_\ell^Q|}{N} \to \frac{1}{2}\log\left(\frac{\sigma_x^2\sigma_z^2 - \sigma_y^2\Delta}{\sigma_z^2\Delta}\right)^+, \qquad \text{(I.11)}$$

and

$$\sum_{\ell=1}^{r} \frac{|\mathcal{I}_\ell^C|}{N} \to \frac{1}{2}\log\left(\frac{\sigma_x^2\sigma_z^2 - \sigma_y^2\Delta}{\sigma_z^4}\right)^-. \qquad \text{(I.12)}$$

Finally,

$$R = \sum_{\ell=1}^{r} \frac{|\mathcal{I}_\ell^Q|}{N} - \frac{|\mathcal{I}_\ell^C|}{N} \to \frac{1}{2}\log(\frac{\sigma_z^2}{\Delta})^+. \qquad \text{(I.13)}$$

□

APPENDIX J

# Proof of Theorem 6.5.1

*Proof.* We firstly show the power constraint $P$ can be satisfied. Recall $U_\ell^{1:N} = A_\ell^{1:N} G_N$ for each level $\ell$. Similar to the previous proof, denote by $P_{U_{1:r}^{1:N},\bar{T}^{1:N}}$ the joint distribution between $U_{1:r}^{1:N}$ and $\bar{T}_{1:r}^{1:N}$ when the encoder applies random rounding rule for all indices $i \in [N]$ at level $\ell$. Denote by $Q_{U_{1:r}^{1:N},T^{1:N}}$ the joint distribution when $U_\ell^{\mathcal{I}_\ell^Q}$ and $U_\ell^{d\mathcal{S}_\ell}$ are encoded following the random rounding rule at each level. $U_\ell^{\mathcal{F}_\ell^C}$ is a uniformly random sequence shared between the encoder and decoder, $U_\ell^{d\mathcal{F}_\ell}$ is a uniform message sequence and $U_\ell^{\mathcal{S}_\ell^C}$ is determined according to the MAP rule.

Notice that $\epsilon_\Lambda(\tilde{\sigma}_q) \leq \epsilon_\Lambda(\tilde{\sigma}_c) = O(2^{-\sqrt{N}})$. Similar to the previous proof, we have

$$\mathbb{V}(P_{U_{1:r}^{1:N},\bar{T}^{1:N}}, Q_{U_{1:r}^{1:N},T^{1:N}}) \leq O(2^{-N^{\beta'}}) + O(2^{-\sqrt{N}}) = O(2^{-N^{\beta'}}). \tag{J.1}$$

Thus, the average transmit power realized by $Q_{U_{1:r}^{1:N},T^{1:N}}$ can be arbitrarily close to that realized by $P_{U_{1:r}^{1:N},\bar{T}^{1:N}}$, i.e.,

$$\begin{aligned}\frac{1}{N}\sum_{u_{1:r}^{1:N},t^{1:N}} &[Q_{U_{1:r}^{1:N},T^{1:N}}(u_{1:r}^{1:N},t^{1:N}) - P_{U_{1:r}^{1:N},\bar{T}^{1:N}}(u_{1:r}^{1:N},t^{1:N})]d(\frac{1}{\alpha_q}\mathcal{M}(u_{1:r}^{1:N}G_N),t^{1:N}) \\ &\leq \frac{2}{N}Nd_{max}\mathbb{V}(P_{U_{1:r}^{1:N},\bar{T}^{1:N}}, Q_{U_{1:r}^{1:N},T^{1:N}}), \\ &\leq O(2^{-N^{\beta'}}),\end{aligned} \tag{J.2}$$

where $\mathcal{M}(u_{1:r}^{1:N}G_N)$ denotes a mapping from $u_{1:r}^{1:N}$ to $t^{1:N}$ according to the lattice Gaussian distribution. Note that for a constant $\sigma_a^2$ and $\sigma_s^2$, the assumption of a maximum distortion $d_{max}$ is reasonable since the Gaussian distribution decays exponen-

tially with large distortion.

Now we show that the average transmit power realized by $P_{U_{1:r}^{1:N},\bar{T}^{1:N}}$ is arbitrarily close to $P$. When $r$ is sufficiently large, $P_{A_{1:r}} \to D_{\Lambda,\sigma_a^2}$, and $\bar{T} = A + N(0, \alpha_q P)$ as shown in Fig. 6.11. Then the variable $\frac{1-\alpha_q}{\alpha_q}A + (A - \bar{T}) = \frac{1-\alpha_q}{\alpha_q}A + N(0, \alpha_q P)$ corresponds to a variable resulted from adding a lattice Gaussian distributed variable to an independent Gaussian noise. Notice that $\frac{1-\alpha_q}{\alpha_q}$ only involves a scale on $D_{\Lambda,\sigma_a^2}$. When $\epsilon_\Lambda(\tilde{\sigma}_q) \leq O(2^{-\sqrt{N}})$, the flatness factor associated with the AWGN channel from $\frac{1-\alpha_q}{\alpha_q}A$ to $\frac{1-\alpha_q}{\alpha_q}A + N(0, \alpha_q P)$ is also upper bounded by $O(2^{-\sqrt{N}})$.

Check that $\left(\frac{1-\alpha_q}{\alpha_q}\right)^2 \sigma_a^2 = (1 - \alpha_q)P$. Let $\dot{X}$ and $\ddot{X}$ denote $\frac{1-\alpha_q}{\alpha_q}A + N(0, \alpha_q P)$ and Gaussian random variable with distribution $N(0, P)$, respectively. By Lemma 6.2.2, $\mathbb{V}(P_{\dot{X}}, P_{\ddot{X}}) \leq O(2^{-\sqrt{N}})$. Let $x^{1:N} = \frac{1}{\alpha_q}\mathcal{M}(u_{1:r}^{1:N}G_N) - t^{1:N}$, we have

$$\begin{aligned}
&\frac{1}{N}\sum_{u_{1:r}^{1:N},t^{1:N}} P_{U_{1:r}^{1:N},\bar{T}^{1:N}}(u_{1:r}^{1:N}, t^{1:N}) d(\frac{1}{\alpha_q}u_{1:r}^{1:N}G_N, t^{1:N}) \\
&\qquad = \frac{1}{N}\sum_{x^{1:N}} P_{\dot{X}^{1:N}} d(x^{1:N}, 0) \qquad \text{(J.3)} \\
&\qquad = E_{\dot{X}}[x^2].
\end{aligned}$$

Since $\mathbb{V}(P_{\dot{X}}, P_{\ddot{X}}) \leq O(2^{-\sqrt{N}})$, $E_{\dot{X}}[x^2] - E_{\ddot{X}}[x^2] \leq O(2^{-\sqrt{N}})$. Consequently,

$$\begin{aligned}
P_T &= \frac{1}{N}\sum_{u_{1:r}^{1:N},t^{1:N}} Q_{U_{1:r}^{1:N},T^{1:N}}(u_{1:r}^{1:N}, t^{1:N}) d(\frac{1}{\alpha_q}\mathcal{M}(u_{1:r}^{1:N}G_N), t^{1:N}) && \text{(J.4)} \\
&\leq \frac{1}{N}\sum_{u_{1:r}^{1:N},t^{1:N}} P_{U_{1:r}^{1:N},\bar{T}^{1:N}}(u_{1:r}^{1:N}, t^{1:N}) d(\frac{1}{\alpha_q}\mathcal{M}(u_{1:r}^{1:N}G_N), t^{1:N}) + O(2^{-N^{\beta'}}) && \text{(J.5)} \\
&\leq E_{\ddot{X}}[x^2] + O(2^{-\sqrt{N}}) + O(2^{-N^{\beta'}}) && \text{(J.6)} \\
&= P + O(2^{-N^{\beta'}}). && \text{(J.7)}
\end{aligned}$$

Now we prove the reliability. Recall that $Y = S + X + Z$, where $X = S' - \rho S$

is independent of $S$. Scaling $Y$ by $\rho$ gives us

$$\begin{aligned} \rho Y &= \rho S' + \rho(1-\rho)S + \rho Z \quad \text{(J.8)} \\ &= \alpha_c S' + \rho(1-\rho)S - (\alpha_c - \rho)S' + \rho Z. \quad \text{(J.9)} \end{aligned}$$

Check that

$$\alpha_c - \alpha_q = \frac{P(P+\sigma_z^2)}{P\sigma_i^2 + (P+\sigma_z^2)^2} = (1-\alpha_q)\rho, \quad \text{(J.10)}$$

leaving us $\alpha_c - \rho = \alpha_q(1-\rho)$. Scale $\rho Y$ by $\frac{\alpha_q}{\alpha_c}$. Then

$$\frac{\alpha_q}{\alpha_c}\rho Y = \alpha_q S' + \frac{\alpha_q}{\alpha_c}(1-\rho)(\rho S - \alpha_q S') + \frac{\alpha_q}{\alpha_c}\rho Z. \quad \text{(J.11)}$$

Note that both $\rho S - \alpha_q S'$ and $Z$ are independent of $S'$. Replacing $\alpha_q S'$ with $A$, we have

$$\frac{\alpha_q}{\alpha_c}\rho \dot{Y} = A + \frac{\alpha_q}{\alpha_c}(1-\rho)(\rho S - A) + \frac{\alpha_q}{\alpha_c}\rho Z, \quad \text{(J.12)}$$

which corresponds to the reverse solution shown in Fig. 6.11. Let $\bar{Y}$ denote the channel output when $S$ is replaced by $\bar{S}$, i.e.,

$$\frac{\alpha_q}{\alpha_c}\rho \bar{Y} = A + \frac{\alpha_q}{\alpha_c}(1-\rho)(\rho \bar{S} - A) + \frac{\alpha_q}{\alpha_c}\rho Z. \quad \text{(J.13)}$$

Recall $\bar{T} = \rho\bar{S}$, $T = \rho S$, and $\bar{B} = \frac{\alpha_q}{\alpha_c}\rho\bar{Y}$. Also let $\dot{B} = \frac{\alpha_q}{\alpha_c}\rho\dot{Y}$. According to the previous proof, we already have $\mathbb{V}(P_{U_{1:r}^{1:N},\bar{T}^{1:N}}, Q_{U_{1:r}^{1:N},T^{1:N}}) \leq O(2^{-N^{\beta'}})$. Note that $Z$ is an independent Gaussian noise, it is not difficult to obtain that

$$\mathbb{V}(P_{U_{1:r}^{1:N},\bar{T}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},T^{1:N},\dot{B}^{1:N}}) = \mathbb{V}(P_{U_{1:r}^{1:N},\bar{T}^{1:N}}, Q_{U_{1:r}^{1:N},T^{1:N}}) \leq O(2^{-N^{\beta'}}), \quad \text{(J.14)}$$

since $P_{\bar{B}^{1:N}|U_{1:r}^{1:N},\bar{T}^{1:N}} = Q_{\dot{B}^{1:N}|U_{1:r}^{1:N},T^{1:N}}$. By (I.8),

$$\mathbb{V}(P_{U_{1:r}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},\dot{B}^{1:N}}) \leq \mathbb{V}(P_{U_{1:r}^{1:N},\bar{T}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},T^{1:N},\dot{B}^{1:N}}) \leq O(2^{-N^{\beta'}}) \quad \text{(J.15)}$$

Note that $\bar{B}$ is a variable obtained by adding $A$ to a Gaussian noise, and $\dot{B}$ is the real signal received because the side information $S$ is Gaussian distributed. Similar to (I.9), the expectation of error probability $E_P[P_e]$ caused by $P_{U_{1:r}^{1:N},\bar{B}^{1:N}}$ can be unpper bounded as $E_P[P_e] \leq O(2^{-N^{\beta'}})$. Finally, the expectation of error probability $P_e^{GP}$ caused by $Q_{U_{1:r}^{1:N},\dot{B}^{1:N}}$ satisfies

$$\begin{aligned} P_e^{GP} &\leq E_P[P_e] + 2\mathbb{V}(P_{U_{1:r}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},\dot{B}^{1:N}}) && \text{(J.16)} \\ &\leq O(2^{-N^{\beta'}}). && \text{(J.17)} \end{aligned}$$

□

# Bibliography


[1] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge, UK: Cambridge University Press, book in preparation. 4, 15, 2, 29, 34, 35, 39, 40, 92, 93, 132, 135, 198

[2] H. S. Cronie and S. B. Korada, "Lossless source coding with polar codes," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, TX, June 2010, pp. 904–908. 9, 13, 14, 15, 16

[3] Y. Yan, "Construction of lattices for communication and security," Ph.D. dissertation, Imperical College London, London, United Kingdom, 2015. 19, 43

[4] 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, July 2009. 1, 6, 7, 12, 25, 91, 137, 171, 174, 175

[5] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013. 1, 71, 107, 175, 176, 211

[6] 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, July 2009. 1, 6, 13, 23, 107

[7] R. Pedarsani, S. Hassani, I. Tal, and I. Telatar, "On the construction of polar codes," in *Proc. 2011 IEEE Int. Symp. Inform. Theory*, July 2011, pp. 11–15. 1, 175

[8] E. Arıkan, "Source polarization," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, TX, 2010, pp. 899–903. 1, 67, 68, 115

[9] S. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, April 2010. 1, 132, 149, 160

[10] 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. 1, 54, 64, 67, 68, 70, 72, 108, 117, 118, 123, 137, 139, 190, 206, 209, 213

[11] M. Mondelli, S. H. Hassani, and R. Urbanke, "How to achieve the capacity of asymmetric channels," Sep. 2014. [Online]. Available: http://arxiv.org/abs/1103.4086 1, 115

[12] D. Sutter, J. Renes, F. Dupuis, and R. Renner, "Achieving the capacity of any DMC using only polar codes," in *Proc. 2012 IEEE Inform. Theory Workshop*, Sept. 2012, pp. 114–118. 1

[13] 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. 1, 91, 92, 101, 104, 111, 112, 113, 116, 118, 120

[14] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," Jan. 2013. [Online]. Available: http://arxiv.org/abs/1301.6150 1

[15] E. Abbe and I. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5437–5448, Aug. 2012. 1

[16] S. Hassani and R. Urbanke, "Universal polar codes," in *Proc. 2014 IEEE Int. Symp. Inform. Theory*, June 2014, pp. 1451–1455. 1

[17] E. Sasoglu and L. Wang, "Universal polarization," Jul. 2013. [Online]. Available: http://arxiv.org/abs/1307.7495 1

[18] M. Wilde and S. Guha, "Polar codes for classical-quantum channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 1175–1187, Feb. 2013. 1

[19] 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. 2, 43, 54, 55

[20] C. Ling and J. Belfiore, "Achieving AWGN channel capacity with lattice gaussian coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014. 2, 30, 34, 37, 38, 55, 56, 80, 81, 106, 126, 135, 136, 166, 169, 172, 188, 196

[21] C. Ling, L. Luzzi, J. Belfiore, and D. Stehle, "Semantically secure lattice codes for the Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014. 2, 29, 38, 90, 91, 92, 93, 95, 99, 100, 103, 106, 121, 123, 127, 153, 164

[22] 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. 2

[23] 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, June 2002. 2

[24] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian k -user interference channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3450–3482, June 2014. 2

[25] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*. New York: Springer, 1993. 2, 29, 31, 36, 43, 44, 49, 130

[26] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, July 1948. 6

[27] N. Stolte, "Rekursive codes mit der plotkin-konstruktion und ihre decodierung," Ph.D. dissertation, Technische Universität Darmstadt, 2002. 6

[28] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013. 6, 13, 23

[29] S. B. Korada, E. Şaşoğlu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6253–6264, Dec. 2010. 6

[30] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2009. 6, 11, 21, 61, 62

[31] N. Hussami, S. B. Korada, and R. Urbanke, "Performance of polar codes for channel and source coding," in *Proc. 2009 IEEE Int. Symp. Inform. Theory*, Seoul, South Korea, June 2009, pp. 1488–1492. 6, 12

[32] S. Guha and M. Wilde, "Polar coding to achieve the Holevo capacity of a pure-loss optical channel," in *Proc. 2012 IEEE Int. Symp. Inform. Theory*, Cambridge, MA, July 2012, pp. 546–550. 7

[33] M. Wilde and S. Guha, "Polar codes for degradable quantum channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4718–4729, July 2013. 7

[34] J. Renes, F. Dupuis, and R. Renner, "Efficient quantum polar coding," *Phys. Rev. Lett.*, vol. 109, p. 050504, Sep. 2011. 7

[35] E. Arıkan and I. Telatar, "On the rate of channel polarization," in *Proc. 2009 IEEE Int. Symp. Inform. Theory*. Seoul, South Korea: IEEE, June 2009, pp. 1493–1495. 11

[36] I. Tal and A. Vardy, "List decoding of polar codes," in *Proc. 2011 IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011, pp. 1–5. 12

[37] E. Arıkan, H. Kim, G. Markarian, U. Ozgür, and E. Poyraz, "Performance of short polar codes under ML decoding," in *ICT-MobileSummit 2009 Conf. Proc.*, 2009. 12

[38] E. Arıkan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Commun. Lett.*, vol. 12, no. 6, pp. 447–449, June 2008. 12

[39] A. Eslami and H. Pishro-Nik, "On finite-length performance of polar codes: Stopping sets, error floor, and concatenated design," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 919–929, Mar. 2013. 12, 177, 198

[40] N. Goela, S. B. Korada, and M. Gastpar, "On LP decoding of polar codes," in *Proc. 2010 IEEE Inform. Theory Workshop*, Dublin, Ireland, Aug. 2010, pp. 1–5. 12, 198

[41] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975. 23, 89

[42] I. Csiszár, "Almost independence and secrecy capacity," *Probl. of Inform. Transmission*, vol. 32, pp. 48–57, 1996. 23, 90

[43] 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. 24, 25, 26, 28

[44] E. Şaşoğlu and A. Vardy, "A new polar coding scheme for strong security on wiretap channels," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, Istanbul, Turkey, July 2013, pp. 1117–1121. 24, 27, 28, 91, 105, 121, 123

[45] G. Böcherer and R. Mathar, "Operating ldpc codes with zero shaping gap," in *Proc. 2011 IEEE Inform. Theory Workshop*, Oct. 2011, pp. 330–334. 28

[46] 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. 29

[47] 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. 30, 43, 45

[48] 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. 30

[49] G. Poltyrev, "On coding without restictions for the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 40, pp. 409–417, Mar. 1994. 30, 35, 39, 40, 41, 54

[50] 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. 32, 35, 37, 47, 55, 56, 57, 58, 59, 60, 61, 64, 82, 94, 96, 97, 101, 104, 172, 184, 186, 195, 196

[51] Y. Yan, C. Ling, and X. Wu, "Polar lattices: Where Arıkan meets Forney," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, Istanbul, Turkey, July 2013, pp. 1292–1296. 32, 43, 92, 99, 103

[52] 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. 36

[53] H. A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1767–1773, Nov. 1997. 39, 42, 43, 44

[54] A. Ingber, R. Zamir, and M. Feder, "Finite-dimensional infinite constellations," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1630–1656, March 2013. 41

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

[56] 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, 45, 51, 54, 84

[57] E. S. Barnes and N. J. A. Sloane, "New lattice packings of spheres," *Can. J. Math.*, vol. 35, pp. 117–130, 1983. 43, 46

[58] 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, 50, 51

[59] 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, 49, 51, 54, 55, 84

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

[61] 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 (I-TA)*, Feb. 2013, pp. 1–6. 45

[62] 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. 45

[63] 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. 47, 54, 61, 69

[64] 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. 47

[65] 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. 48

[66] 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. 49

[67] N. Sommer, M. Feder, and O. Shalvi, "Low-density lattice codes," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1561–1585, Apr. 2008. 51, 52, 54, 84

[68] 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. 52

[69] 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. 54

[70] 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. 54

[71] 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. 54

[72] 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. 54

[73] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "How to achieve the capacity of asymmetric channels," *CoRR*, vol. abs/1406.7373, 2014. [Online]. Available: http://arxiv.org/abs/1406.7373 54, 64, 67

[74] 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. 54

[75] 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. 54

[76] N. di Pietro, G. Zémor, and J. J. Boutros, "New results on Construction A lattices based on very sparse parity-check matrices," in *IEEE Int. Symp. Inform. Theory (ISIT)*, July 2013, pp. 1675–1679. 54, 84

[77] D. Micciancio and O. Regev, "Worst-case to average-case reductions based on Gaussian measures," in *Proc. Ann. Symp. Found. Computer Science*, Oct. 2004, pp. 372–381. 56

[78] Y. Yan, L. Liu, C. Ling, and X. Wu, "Construction of capacity-achieving lattice codes: Polar lattices," Nov. 2014. [Online]. Available: http://arxiv.org/abs/1411.0187 59, 200

[79] 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, July 1999. 60, 61, 65, 168

[80] 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. 63

[81] 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. 63

[82] D. Goldin and D. Burshtein, "Improved bounds on the finite length scaling of polar codes," *IEEE Trans. Inf. Theory*, pp. 6966–6978, Sep. 2014. 63

[83] A. Ingber, R. Zamir, and M. Feder, "Finite dimensional infinite constellations," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1630–1656, Mar. 2013. 64, 84

[84] 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. 71, 83

[85] Y. Yan and C. Ling, "A construction of lattices from polar codes," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2012, pp. 124–128. 84

[86] S. Leung-Yan-Cheong, "On a special class of wiretap channels," *IEEE Trans. Inf. Theory*, vol. 23, no. 5, pp. 625–627, Sep. 1977. 90

[87] S. Goldwasser and S. Micali, "Probabilistic encryption," *J. Comput. Syst. Sci.*, vol. 28, no. 2, pp. 270–299, 1984. 90

[88] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Proc. CRYPTO 2012*, ser. Lecture Notes in Computer Science, vol. 7417. Springer-Verlag, 2012, pp. 294–311. 90, 93

[89] T. C. Gulcu and A. Barg, "Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component," Oct. 2014. [Online]. Available: http://arxiv.org/abs/1410.3422 91

[90] Y.-P. Wei and S. Ulukus, "Polar coding for the general wiretap channel," in *Proc. 2015 IEEE Inform. Theory Workshop*, Jerusalem, Israel, April 2015, pp. 1–5. 91

[91] F. Oggier, P. Solé, and J.-C. Belfiore, "Lattice codes for the wiretap Gaussian channel: Construction and analysis," Mar. 2011. [Online]. Available: http://arxiv.org/abs/1103.4086 92

[92] A. Ernvall-Hytonen and C. Hollanti, "On the eavesdropper's correct decision in Gaussian and fading wiretap channels using lattice codes," in *Proc. 2011 IEEE Inform. Theory Workshop*, Paraty, Brazil, Oct. 2011, pp. 210–214. 92

[93] E. Abbe and A. Barron, "Polar coding schemes for the AWGN channel," in *Proc. 2011 IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011. 92

[94] A. Joseph and A. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2541–2557, May 2012. 92

[95] C. Ling, L. Luzzi, and M. Bloch, "Secret key generation from Gaussian sources using lattice hashing," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, Istanbul, Turkey, July 2013, pp. 2621–2625. 92

[96] M. Hayashi and R. Matsumoto, "Construction of wiretap codes from ordinary channel codes," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, USA, June 2010, pp. 2538–2542. 93

[97] M. Cheraghchi, F. Didier, and A. Shokrollahi, "Invertible extractors and wiretap protocols," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1254–1274, Feb 2012. 93

[98] H. Tyagi and A. Vardy, "Explicit capacity-achieving coding scheme for the Gaussian wiretap channel," in *Proc. 2014 IEEE Int. Symp. Inform. Theory*, Honolulu, USA, June 2014, pp. 956–960. 93

[99] M. Bloch and J. Laneman, "Strong secrecy from channel resolvability," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8077–8098, Dec. 2013. 93

[100] R. A. Chou, M. R. Bloch, and J. Kliewer, "Low-complexity channel resolvability codes for the symmetric multiple-access channel," in *Proc. 2014 IEEE Inform. Theory Workshop*, Hobart, Australia, Nov. 2014, pp. 466–470. 93

[101] 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. 93

[102] Y. Liang, H. Vincent, and S. Shamai, "Information theoretic security," in *Found. Trends Commun. Inf. Theory*. Norwell, MA, USA: Now Publishers, 2009. 93

[103] 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*, Honolulu, HI, June 2014. 98

[104] Y. Yan, L. Liu, C. Ling, and X. Wu, "Construction of capacity-achieving lattice codes: Polar lattices," Nov. 2014. [Online]. Available: http://arxiv.org/abs/1411.0187 98, 105, 106, 108, 115, 131, 139, 148, 149, 168, 172, 190, 191, 218, 219, 224

[105] L. Liu and C. Ling, "Polar lattices for lossy compression," Jan. 2015. [Online]. Available: http://arxiv.org/abs/1501.05683 123, 167

[106] E. E. Gad, Y. Li, J. Kliewer, M. Langberg, A. Jiang, and J. Bruck, "Asymmetric error correction and flash-memory rewriting using polar codes," Oct. 2014. [Online]. Available: http://arxiv.org/abs/1410.3542 123

[107] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," Mar. 2015. [Online]. Available: http://arxiv.org/abs/1503.08778 124

[108] R. Gray, "Vector quantization," *IEEE ASSP Mag.*, vol. 1, no. 2, pp. 4–29, April 1984. 129

[109] B. Juang and J. Gray, A., "Multiple stage vector quantization for speech coding," in *IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, vol. 7, May 1982, pp. 597–600. 130

[110] A. Gersho and V. Cuperman, "Vector quantization: A pattern-matching technique for speech coding," *IEEE Commun. Mag*, vol. 21, no. 9, pp. 15–21, Dec. 1983. 130

[111] M. Sabin and R. Gray, "Product code vector quantizers for waveform and voice coding," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 3, pp. 474–488, June 1984. 130

[112] R. G. Gallager, *Information theory and reliable communication*. Springer, 1968, vol. 2. 130

[113] W. Finamore and W. Pearlman, "Optimal encoding of discrete-time continuous-amplitude memoryless sources with finite output alphabets," *IEEE Trans. Inf. Theory*, vol. 26, no. 2, pp. 144–155, Mar. 1980. 130

[114] M. Marcellin and T. Fischer, "Trellis coded quantization of memoryless and Gauss-markov sources," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 82–93, Jan. 1990. 130

[115] B. Kudryashov and K. Yurkov, "Near-optimum low-complexity lattice quantization," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, USA, June 2010, pp. 1032–1036. 130

[116] S. Vatedka and N. Kashyap, "Some "goodness" properties of LDA lattices," Oct. 2014. [Online]. Available: http://arxiv.org/abs/1410.7619 131

[117] C. Ling and L. Gan, "Lattice quantization noise revisited," in *Proc. 2013 IEEE Inform. Theory Workshop*, Sept. 2013, pp. 1–5. 131

[118] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, July 1996. 132, 134

[119] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976. 132

[120] R. Venkataramanan, T. Sarkar, and S. Tatikonda, "Lossy compression via sparse linear regression: Computationally efficient encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3265–3278, June 2014. 133

[121] R. Venkataramanan, A. Joseph, and S. Tatikonda, "Lossy compression via sparse linear regression: Performance under minimum-distance encoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3254–3264, June 2014. 133

[122] 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, June 2002. 133, 169

[123] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005. 133

[124] C. Ling, S. Gao, and J.-C. Belfiore, "Wyner-ziv coding based on multidimensional nested lattices," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1328–1335, May 2012. 133

[125] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, Oct. 2005. 134

[126] M. H. M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983. 154

[127] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Inform. Transm*, vol. 9, no. 1, pp. 19–31, 1980. 154

[128] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems Contr. Inform. Theory*, vol. 2, no. 2, pp. 149–162, 1973. 161

[129] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, March 1975. 161, 162, 163

[130] R. Gray and A. Wyner, "Source coding for a simple network," *Bell System Technical Journal*, vol. 53, no. 9, pp. 1681–1721, 1974. 162

[131] K. B. Viswanatha, E. Akyol, and K. Rose, "The lossy common information of correlated sources," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3238–3253, 2014. 162

[132] G. Xu, W. Liu, and B. Chen, "A lossy source coding interpretation of Wyner's common information," *IEEE Trans. Inf. Theory*, vol. 62, pp. 754 – 768, 2016. 162

[133] J. Boutros and E. Biglieri, "Polarization of quasi-static fading channels," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, July 2013, pp. 769–773. 170

[134] A. Bravo-Santos, "Polar codes for the Rayleigh fading channel," *IEEE Commun. Lett.*, vol. 17, no. 12, pp. 2352–2355, Dec. 2013. 170, 172, 174, 177, 180

[135] H. Si, O. Koyluoglu, and S. Vishwanath, "Polar coding for fading channels: Binary and exponential channel cases," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2638–2650, Aug. 2014. 171, 180

[136] F. Oggier and E. Viterbo, *Algebraic number theory and code design for Rayleigh fading channels*. The Netherlands: Now publishers inc, 2004. 171

[137] X. Giraud, E. Boutillon, and J. Belfiore, "Algebraic tools to build modulation schemes for fading channels," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 938–952, May 1997. 171

[138] R. Vehkalahti and L. Luzzi, "Number field lattices achieve Gaussian and Rayleigh channel capacity within a constant gap," in *Proc. 2015 IEEE Int. Symp. Inform. Theory*, June 2015, pp. 436–440. 172

[139] L. Luzzi and R. Vehkalahti, "Almost universal codes achieving ergodic MIMO capacity within a constant gap," July 2015. [Online]. Available: http://arxiv.org/abs/1507.07395 172

[140] A. Hindy and A. Nosratinia, "Approaching the ergodic capacity with lattice coding," in *Proc. 2014 IEEE Global Commun. Conf.*, Dec. 2014, pp. 1492–1496. 172

[141] ——, "Achieving the ergodic capacity with lattice codes," in *Proc. 2015 IEEE Int. Symp. Inform. Theory*, June 2015, pp. 441–445. 172

[142] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1767–1773, Nov. 1997. 172

[143] S. Vituri and M. Feder, "Dispersion of infinite constellations in fast fading channels," April 2014. [Online]. Available: http://arxiv.org/abs/1206.5401 173, 178, 183

[144] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015. 177, 198

[145] A. El Gamal and Y. H. Kim, *Network information theory*. Cambridge university press, 2011. 188, 189

[146] G. D. Forney, "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proc. Annu. Allerton Conf. on Commun. Control and Computing*, vol. 41, no. 1, Oct. 2003, pp. 430–439. 196

[147] C. Hirche, “Polar codes in quantum information theory,” Jan. 2015. [Online]. Available: https://arxiv.org/abs/1501.03737 199