COMPLEXITY AND SECOND MOMENT OF THE MATHEMATICAL THEORY OF COMMUNICATION

BY

HSIN-PO WANG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Assistant Professor Partha S. Dey, Chair
Professor Iwan M. Duursma, Director of Research
Professor József Balogh
Professor Marius Junge

Abstract

The performance of an error correcting code is evaluated by its block error probability, code rate, and encoding and decoding complexity. The performance of a series of codes is evaluated by, as the block lengths approach infinity, whether their block error probabilities decay to zero, whether their code rates converge to channel capacity, and whether their growth in complexities stays under control.

Over any discrete memoryless channel, I build codes such that: for one, their block error probabilities and code rates scale like random codes'; and for two, their encoding and decoding complexities scale like polar codes'. Quantitatively, for any constants $\pi, \rho > 0$ such that $\pi + 2\rho < 1$, I construct a series of error correcting codes with block length N approaching infinity, block error probability $\exp(-N^\pi)$, code rate $N^{-\rho}$ less than the channel capacity, and encoding and decoding complexity $O(N \log N)$ per code block.

Over any discrete memoryless channel, I also build codes such that: for one, they achieve channel capacity rapidly; and for two, their encoding and decoding complexities outperform all known codes over non-BEC channels. Quantitatively, for any constants $\tau, \rho > 0$ such that $2\rho < 1$, I construct a series of error correcting codes with block length N approaching infinity, block error probability $\exp(-(\log N)^\tau)$, code rate $N^{-\rho}$ less than the channel capacity, and encoding and decoding complexity $O(N \log(\log N))$ per code block.

The two aforementioned results are built upon two pillars—a versatile framework that generates codes on the basis of channel polarization, and a calculus–probability machinery that evaluates the performances of codes.

The framework that generates codes and the machinery that evaluates codes can be extended to many other scenarios in network information theory. To name a few: lossless compression with side information, lossy compression, Slepian–Wolf problem, Wyner–Ziv Problem, multiple access channel, wiretap channel of type I, and broadcast channel. In each scenario, the adapted notions of block error probability and code rate approach their limits at the same paces as specified above.

First there is Bo-Le[1]

Then can horses gallop hundreds of miles

Horses capable of galloping far are common

But Bo-Les are scarce

—Han Yu, *On Horses*

[1]The honorific name of Sun Yang, who is a horse tamer in Spring and Autumn period and renowned as a judge of horses; also refers to those who recognize (especially hidden) talent.

Contents

CHAPTER 1

Introduction

Seventy-three years ago, Claude E. Shannon founded the theory of information with an article titled *A Mathematical Theory of Communication*, which was later republished under the name *The Mathematical Theory of Communication* to reflect its omnipotence.

In the eternal work, Shannon explained how to measure the information content of a random variable X and argued that, in the long term, the information content carried by X costs $H(X \mid Y) + \varepsilon$ bits to be remembered, given that we have free access to another random variable Y. Shannon also showed that, if sending X results in the reception of Y, where $X \to Y$ is called a communication channel, then the rate at which information can be transmitted is $I(X\ ;\ Y) - \varepsilon$ bits per usage of channel. These results are now called [Shannon's] source coding theorem and noisy-channel coding theorem, respectively.

The famous article left two loopholes. Loophole one: Shannon's proof involves the existence of certain mathematical objects which, in reality, are next to impossible to find constructively. As a consequence, Shannon's protocol is never utilized beyond academic interest. Loophole two: Whereas Shannon's bound on ε is strong enough to conclude that $\varepsilon \to 0$, which evinces that $H(X \mid Y)$ and $I(X\ ;\ Y)$ are the limits we would like to achieve, we did not know how rapid ε decays to 0. That is, Shannon identified the first order limit of coding, but left the second order limit open.

The current dissertation aims to patch the two said loopholes and succeeds in improving over existing patches. I will present an $\varepsilon \to 0$ coding scheme whose complexity is $O(N \log(\log N))$, where N is the block length, whilst the best known result is $O(N \log N)$. This in turn patches the first loophole further, and is referred to as the complexity paradigm of coding. I will also present an $O(N \log N)$ coding scheme whose ε decays to zero at the pace that is provably optimal, while earlier works only handle the binary case. This in turn fills the second loophole further, and is referred to as the second-moment paradigm of coding. I will then present a joint scheme that achieves both $O(N \log(\log N))$ and the optimal pace of $\varepsilon \to 0$.

My codes are built upon two pillars. Pillar one: The overall code can be seen as a modification of a recently developed code—polar code. Depending on how we modify polar coding, we can inherit its $O(N \log N)$ complexity or reduce it further down to $O(N \log(\log N))$. Pillar two: The polarization kernel can be seen as a flag of the legacy codes—random codes. Since random coding is the only way to achieve the second-moment paradigm, I incorporate it to boost the performance of polar coding to the second-moment paradigm. The two pillars support a coding scheme whose complexity scales like polar coding but performance scales like random coding.

After mastering the complexity and second-moment paradigms of the source coding theorem and the noisy-channel coding theorem, this dissertation moves forward to a network coding scenario called distributed lossless compression. I will then adapt the complexity and second-moment paradigms to these scenario. Using similar techniques, the same result generalizes to even more coding scenarios such as multiple access channels, wiretap channels of type I, and broadcast channels, and is left for future research.

1. Organization of the Dissertation

The remaining sections of the current chapter map one-to-one to the remaining chapters of the current dissertation and serve as their summaries.

Among the chapters: Chapter 2 was presented in the preprint [**WD18b**]. Chapter 4 was presented in the preprints [**WD18a**] and [**WD19a**], the latter of which was later published in IEEE Transactions on Information Theory [**WD21a**]. Chapter 5 was presented in the preprint [**WD18c**]. Chapter 6 was presented in the preprint [**WD19b**], which was later published in IEEE Transactions on Information Theory [**WD21b**].

2. Original Channel Polarization

In Chapter 2, we will revisit Arıkan's original proposal of polar coding that is dedicated to binary-input channels and uses $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$ as the polarization kernel. This chapter serves three purposes: One, the construction by $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$ is simple yet powerful enough to achieve capacity (the first order limit), and is of historical significance. Two, I will provide a complete proof to the strongest version of the statements available in the literature that unifies the techniques spanning across several state-of-the-art works. Three, the main statement and its proof are the starting point of at least three generalizations, and will be referred to as the prototype every now and then.

In the seminal paper, Arıkan started with a symmetric binary-input discrete-output memoryless channel $W(y \mid x)$ and synthesized its children $W^{(1)}(y_1y_2 \mid u_1)$ and $W^{(2)}(y_1y_2u_1 \mid u_2)$ via

$$W^{(1)}(y_1y_2 \mid u_1) \coloneqq \sum_{u_2\in\mathbb{F}_2} \frac{1}{2}W(y_1 \mid u_1+u_2)W(y_2 \mid u_2),$$
$$W^{(2)}(y_1y_2u_1 \mid u_2) \coloneqq \frac{1}{2}W(y_1 \mid u_1+u_2)W(y_2 \mid u_2).$$

Treating $\bullet^{(1)}$ and $\bullet^{(2)}$ as transformations applied to channels, Arıkan continued synthesizing, in a recursive manner, W's grandchildren $(W^{(1)})^{(1)}$, $(W^{(1)})^{(2)}$, $(W^{(2)})^{(1)}$, and $(W^{(2)})^{(2)}$, followed by W's grand-grandchildren $\bigl((W^{(1)})^{(1)}\bigr)^{(1)}$, $\bigl((W^{(1)})^{(1)}\bigr)^{(2)}$, etc, followed by their descendants ad infinitum.

Arıkan observed that a code can be established by selecting a subset of reliable synthetic channels. To evaluate the performance of codes established this way, we proceed to examine the stochastic process $\{\mathsf{W}_n\}$ defined by $\mathsf{W}_0 \coloneqq W$ and $\mathsf{W}_{n+1} \coloneqq (\mathsf{W}_n)^{(1 \text{ or } 2 \text{ with equal chance})}$. The evolution of the synthetic channels can be controlled by $Z(W^{(2)}) = Z(W)^2$ and $Z(W)\sqrt{2-Z(W)^2} \leqslant Z(W^{(1)}) \leqslant 2Z(W) - Z(W)^2$. From that I will prove

$$P\bigl\{Z(\mathsf{W}_n) < e^{-2^{\pi n}}\bigr\} > P\{Z(\mathsf{W}_n) \to 0\} - 2^{-\rho n}, \tag{1.1}$$

where (π, ρ) is any pair of constants that lies in the shaded area in Figure 2.1. This inequality tells us how reliable W_n can be. Thus we learn how the original polar coding performs.

3. Asymmetric Channels

In Chapter 3, I will bring up a "dual picture" of Chapter 2. The dual picture consists of three elements: One, through examining the behavior of W_n when it becomes noisy, i.e., when $Z(W_n) \approx 1$, we have a more complete understanding of $\{W_n\}$ as a stochastic process. Two, the proof of said behavior is the mirror image of the one given in Chapter 2, reinforcing the duality. Three, this result is pivotal to source coding for lossy compressions, to noisy-channel coding over asymmetric channels, and to a pruning technique that will be covered in upcoming chapters.

While inequality (1.1) addresses the behavior of W_n at the reliable end, another parameter T and a stochastic process $\{T(W_n)\}$ are defined to examine the behavior of W_n at the noisy end. It satisfies $T(W^{(1)}) = T(W)^2$ and $T(W^{(2)}) \leqslant 2T(W) - T(W)^2$, and can be used to show that

$$P\{T(W_n) < e^{-2^{\pi n}}\} > P\{T(W_n) \to 0\} - 2^{-\rho n}, \tag{1.2}$$

where (π, ρ) is any pair of constants that lies in the *same* shaded area in Figure 2.1. Note that $T(W_n) \to 0$ iff $Z(W_n) \to 1$, so inequality (1.2) is the mirror image of inequality (1.1), telling us how noisy W_n can be.

Inequality (1.1) alone implies that polar coding is good for error correction over symmetric channels and lossless compression (with or without side information). Inequality (1.2) alone implies that polar coding is good for lossy compression. Together, inequalities (1.1) and (1.2) imply that (a) polar coding is good over asymmetric channels, and (b) polar coding can be modified to attain an even lower complexity. More on (b) in Chapter 4.

The proofs of inequalities (1.1) and (1.2) involve monitoring a "watch list" subset A_m of moderately reliable channels for $m = \sqrt{n}, 2\sqrt{n}, \ldots, n - \sqrt{n}$ for some large perfect square n. When $W_m \in A_m$ is moderately good and $W_{m+\sqrt{n}}$ becomes extraordinary reliable, $W_{m+\sqrt{n}}$ is moved from $A_{m+\sqrt{n}}$ to another "trustworthy" subset $E_{m+\sqrt{n}}$. On the other hand, when $W_{m+\sqrt{n}}$ becomes noisy, it is temporary removed from $A_{m+\sqrt{n}}$, but has some chance to be added back to $A_{m+l\sqrt{n}}$ (for some $l \geqslant 2$) once $W_{m+l\sqrt{n}}$ becomes moderately reliable again. The rest of the proof is to quantify moderate and extraordinary reliabilities and the corresponding frequencies.

4. Pruning Channel Tree

In Chapter 4, I will introduce a novel technique called pruning. Pruning reduces the complexity of encoder and decoder while retaining in part the performance of polar codes. And it is reported prior that pruning reduces the complexity by a constant factor. In this chapter, I will show that pruned polar codes can achieve capacity with encoding and decoding complexity $O(N \log(\log N))$, transcending the old $O(N \log N)$ complexity. We will see in later chapters that pruning is a special case of dynamic kerneling and can be applied to more general polar codes.

To explain pruning, I will establish the trinitarian correspondence among the encoder/decoder, the channel tree, and the channel process $\{W_n\}$. Pruning the channel tree corresponds to trimming the unnecessary part of the encoder and decoder, which reduces complexity. Viewed from the a different perspective, pruning the channel tree corresponds to declaring a stopping time s that is adapted to $\{\mathsf{W}_n\}$ so that the stochastic process $\{\mathsf{W}_{n\wedge\mathsf{s}}\}$ stabilizes whenever channel transformation becomes ineffective.

Now W_s becomes a random variable associated to code performance. For all intents and purposes, it suffices to declare the stopping time s properly and prove inequalities of the form

$$P\big\{Z(\mathsf{W}_\mathsf{s}) < 4^{-n}\big\} > 1 - H(W) - 2^{-\rho n},$$
$$P\big\{T(\mathsf{W}_\mathsf{s}) < 4^{-n}\big\} > H(W) - 2^{-\rho n}$$

and another inequality of the form

$$N\,E[\mathsf{s}] \leqslant O(N\log(\log N)).$$

The first two inequalities imply that the code achieves capacity; the third inequality confirms the complexity being $O(N\log(\log N))$.

5. General Alphabet and Kernel

In Chapter 5, I will investigate thoroughly two known ways to generalize polar coding. One is allowing channels with arbitrary finite input alphabet. This extends polar coding to all channels Shannon had considered in 1948. The other is utilizing arbitrary matrices as polarizing kernels. Doing so provably improves the performance in the long run, and is reportedly improving the performance for moderate block length.

To begin, we will go over four regimes that connect probability theory, random coding theory, and polar coding theory:

- Law of large numbers (LLN) and achieving capacity; this regime concerns whether block error probability P_e decays to 0 while code rate R converges to capacity.
- Large deviation principle (LDP) and error exponent; this regime concerns how fast P_e decays to 0 when an R is fixed.
- Central limit theorem (CLT) and scaling exponent; this regime concerns how fast R approaches capacity a when P_e is fixed.
- Moderate deviation principle (MDP); this regime concerns the general trade-off between P_e and R.

Next, I will go back to prove results regarding polar coding. For any matrix G over any finite field $\mathbb{F}_q$, the LDP data of G include coset distances $D_Z^{(j)} := \operatorname{hdis}(r_j, R_j)$, where hdis is the Hamming distance, r_j is the jth row of G, and R_j is the subspace spanned by the rows below r_j. Coset distances are such that

$$Z(W^{(j)}) \approx Z(W)^{D_Z^{(j)}}.$$

This approximation is used to control small $Z(\mathsf{W}_n)$, which eventually proves a generalization of inequality (1.1). For the dual picture, there is a parameter S generalizing T and satisfying

$$S(W^{(j)}) \approx S(W)^{D_S^{(j)}},$$

where $D_S^{(j)} := \operatorname{hdis}(c_j, C_j)$ is the Hamming distance from c_j the jth column of G^{-1} to C_j the subspace spanned by the columns to the left of c_j. This eventually proves a generalization of inequality (1.2).

The CLT data of an $\ell \times \ell$ matrix G consist of a choice of parameter H, a concave function $h : [0,1] \to [0,1]$ such that $h(0) = h(1) = 0$, and a number ϱ such that

$$\frac{1}{\ell}\sum_{i=1}^{\ell} h(H(W^{(j)})) \leqslant \ell^{-\varrho} h(H(W)),$$

where H could be the conditional entropy or any other handy parameter that maximizes ϱ.

The contribution of Chapter 5 is a calculus–probability machinery that predicts the MDP behavior of polar codes given the LDP and CLT data. The prediction is of the form

$$P\{Z(W_n) < e^{-\ell^{\pi n}}\} > 1 - H(W) - \ell^{-\rho n},$$

where (π, ρ) lies to the left of the convex envelope of $(0, \varrho)$ and the convex conjugate of $t \mapsto \log_\ell\bigl(\ell^{-1}\sum_{j=1}^{\ell}(D_Z^{(j)})^t\bigr)$. This generalizes inequality (1.1). Similarly, the generalization of inequality (1.2) reads

$$P\{S(W_n) < e^{-\ell^{\pi n}}\} > H(W) - \ell^{-\rho n},$$

where $(\pi, \rho) \in [0,1]^2$ to the left of the convex envelope of $(0, \varrho)$ and the convex conjugate of $t \mapsto \log_\ell\bigl(\ell^{-1}\sum_{j=1}^{\ell}(D_S^{(j)})^t\bigr)$.

6. Random dynamic Kerneling

In Chapter 6, two novel ideas are invoked to help achieve the optimal MDP behavior with low complexity. First, the matrix G that induces polarization is not fixed but varying on a channel-by-channel basis. Second, since it is difficult to prove that a specific G is good, a random variable $\mathbb{G}$ is to replace G and I will investigate the typical behavior of $\mathbb{G}$ as a polarizing kernel.

The MDP behavior of random coding, which is provably optimal, reads

$$\frac{-\ln P_{\mathrm{e}}}{N(C-R)^2} \to \frac{1}{2V},$$

where C is channel capacity and V is another intrinsic parameter called channel dispersion or varentropy. Our target behavior is less impressive, yet it is asymptotically optimal in the logarithmic scale:

$$\frac{\ln(-\ln P_{\mathrm{e}})}{\ln(N(C-R)^2)} \approx 1.$$

Or equivalently, for any $\pi + 2\rho < 1$, there are codes such that $P_{\mathrm{e}} < \exp(-N^{\pi})$ and $C - R < N^{-\rho}$.

For the typical LDP behavior of $\mathbb{G}$, we need to understand, for each j, the typical Hamming distance $D_Z^{(j)}$ from its jth row to the subspace spanned by the rows below. This step is essentially the Gilbert–Varshamov bound with slight modifications so that it is easier to manipulate in later steps.

For the typical CLT behavior of $\mathbb{G}$, I choose the concave function $h(x) := \min(x, 1-x)^{\ell/\ln \ell}$. Now we need to understand the typical behavior of

$$\frac{1}{\ell}\sum_{i=1}^{\ell} h(H(W^{(j)})),$$

where $W^{(j)}$ is a random variable depending on $\mathbb{G}$. This boils down to showing that the first few $H(W^{(j)})$ are close to 1, while the last few $H(W^{(j)})$ are close to 0.

To show that $H(W^{(j)}) \approx 1$ and to quantify the approximation, I reduce this to a reliability analysis of noisy-channel coding. To show that $H(W^{(j)}) \approx 0$ and to quantify the approximation, I reduce this to a secrecy analysis of wiretap-channel coding.

7. Joint Pruning and Kerneling

In Chapter 7, I will combine the techniques in Chapters 4 and 5 to depict a trade-off between the complexity—ranging from $O(N \log N)$ to $O(N \log(\log N))$—and the decay of P_{e}—ranging from $\exp(-N^{\pi})$ to $\exp(-(\log N)^{\tau})$.

The main idea is to apply the stopping time analysis to any channel process $\{\mathsf{W}_n\}$ whose MDP behavior is known. It could be a process generated by $\left[\begin{smallmatrix} 1 & 0 \\ c & 1 \end{smallmatrix}\right]$, for which we know that it is guaranteed to have a positive ϱ. It could be a process generated by a large kernel whose ϱ is bounded by some other method. It could also be generated by random dynamic kerneling, for which we know $\varrho \to 1/2$.

The result is that, for any kernel, polar codes have the same gap to capacity before and after pruning; and depending on how aggressively one wants to prune, the complexity per bit is approximately the logarithm of the logarithm of the block error probability.

For example, if the targeted block error probability is $\exp(-N^{\pi})$, then the predicted complexity is $O(N \log N)$. This recovers the result of the previous chapter. On the other hand, if the targeted block error probability is $\exp(-(\log N)^{\tau})$, then the predicted complexity is $O(N \log(\log N))$. This is, by far, the lowest complexity for capacity-achieving codes over generic channels. Plus the gap to capacity decay to 0 optimally fast.

8. Distributed Lossless Compression

In Chapter 8, I will extend the theorems established in the previous chapters to distributed lossless compression problems. A distributed lossless compression problem is a network coding problem where there are m sources, each to be compressed by a compressor that do not talk to each other, and a decompressor that attempt to reconstruct all sources. I will go over the two-source case as a warm up, the three-source case to demonstrate the difficulty, and finally the m-source case for a general result.

I will explain that, modulo the previous chapters, the main challenge is to reduce a multiple-sender problems to several one-sender problems. The reduction consists of two steps. The first step is to demonstrate that a random source X can be "split" into two random fragments $X\langle 1\rangle$ and $X\langle 2\rangle$ such that there is a bijection $X \leftrightarrow (X\langle 1\rangle, X\langle 2\rangle)$ and hence they carry the same amount of information. The second step is to show that, by interleaving the fragments of sources in a way that is related to Gray codes, we can fine-tune the workloads of every sender. That helps us achieve every possible distribution of workloads.

A key to the second step is degree theory, an algebraic topology machinery that determines the surjectivity of a continuous map. The degree theory offers a sufficient condition on whether a map is onto the dominant face of a contra-polymatroid. Here, it is the rate region of a distributed lossless compression problem that is a contra-polymatroid. Dually, the capacity region of a multiple access channel is a polymatroid and a similar argument applies. This fact indicates that, for both distributed lossless compression and multiple access channels, splitting coupled with polar coding achieves the optimal block error probability and the optimal gap to boundary at the cost of $O(N \log N)$ complexity. Or, following the complexity paradigm, one prunes the complexity to $O(N \log(\log N))$ if a slightly higher block error probability is acceptable.

CHAPTER 2

Original Channel Polarization

Fifteen years ago, Erdal Arıkan developed a technique, called *channel combining and splitting*, to combine two identical channels and then split them into two distinct channels [**Ari06**]. At the cost of having to prepare different codes to deal with distinct channels, the two new channels enjoy better metrics. More precisely, the average of the cutoff rates rises. Arıkan then argued that, by recursively synthesizing the children of the children of ... of the channel, the rise in cutoff rates eventually pushes them towards the channel capacity.

As it turns out, after a sufficient amount of recursion, one does not need 2^n different coding schemes to deal with the 2^n descendants of W. This is because most synthetic channels are either satisfactorily reliable—so we just transmit plain messages through these—or desperately noisy—so we just ignore those. The phenomenon is named *channel polarization* and the corresponding coding scheme *polar coding*.

Arıkan showed that this original polar coding achieves channel capacity. That is, if you follow Arıkan's instruction to construct codes, then $P_{\mathrm{e}} \to 0$ and $R \to I(W)$ over any symmetric binary-input discrete-output memoryless channels. This is done via proving, for some functions $\theta(n)$ and $\gamma(n)$, (a) that

$$\mathsf{P}\big\{Z(\mathsf{W}_n) < \theta(n)\big\} > I(W) - \gamma(n),$$

(b) that $2^n\theta(n)$ is an upper bound on P_{e}, and (c) that $\gamma(n)$ is an upper bound on $I(W) - R$.

In this chapter, I will characterize the pace of achieving capacity. We will see that

$$\mathsf{P}\big\{Z(\mathsf{W}_n) < e^{-2^{\pi n}}\big\} > I(W) - 2^{-\rho n}$$

if $(\pi, \rho) \in [0,1]^2$ lies to the left of the convex envelope of $(0, 1/4.714)$ and $1-$ (the binary entropy function). Prior to my work, the largest region of achievable (π, ρ) is considerably smaller and reaches only $(0, 1/5.714)$ [**MHU16**]. See Figure 2.1 for plots.

1. Problem Setup and Primary Definitions

After introducing some definitions, this section describes the main problem to be solved in this chapter. First goes the definition of the channels we want to attack.

Definition 2.1. A *symmetric binary-input discrete-output memoryless channels* (SBDMCs) is a Markov chain $W\colon \mathbb{F}_2 \to \mathcal{Y}$, where

- $\mathbb{F}_2$ is the finite field or order 2;
- $\mathcal{Y}$ is a finite set;

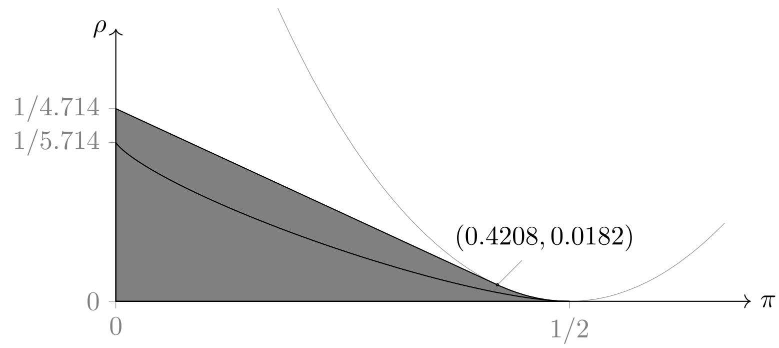

FIGURE 2.1. The achievable region of $(\log(-\log_2 P_e), \log(C-R))$ is shaded. The curve part is 1 minus the binary entropy function. The straight part is the tangent line from $(0, 1/4.714)$ to the curve, the tangent point being $(0.4208, 0, 0182)$. The lower left curve is the previous result [**MHU16**], which attains $(0, 1/5.714)$.

- $W(y \mid x) \in [0,1]$, for $x \in \mathbb{F}_2$ and $y \in \mathcal{Y}$, is an array of transition probabilities satisfying $\sum_{y\in\mathcal{Y}} W(y \mid x) = 1$ for both $x \in \mathbb{F}_2$; and
- there exists an involution $\sigma\colon \mathcal{Y} \to \mathcal{Y}$ such that $W(y \mid 0) = W(\sigma(y) \mid 1)$ for all $y \in \mathcal{Y}$.

Denote by Q the uniform distribution on $\mathbb{F}_2$; treat this as the input distribution of the channel W. Denote by $W(x,y)$ the joint probability $Q(x)W(y \mid x)$. Denote by $W(x \mid y)$ the posterior probability; note that $W(\bullet \mid \bullet)$ assumes two interpretations, depending on whether we want to predict y from x or the other way around. When it is necessary, $W(y) \coloneqq W(y \mid 0)+W(y \mid 1)$ denotes the output probability. Capital variables X and Y, usually with indices, denote the input and output of the channel governed by Q and W.

The definitions of some channel parameters follow.

Definition 2.2. The *conditional entropy* of W is

$$H(W) \coloneqq -\sum_{x\in\mathbb{F}_2}\sum_{y\in\mathcal{Y}} W(x,y)\log_2 W(x \mid y),$$

which is the amount of noise/equivocation/ambiguity/fuzziness caused by W.

Definition 2.3. The *mutual information* of W is

$$I(W) \coloneqq H(Q) - H(W) = \sum_{x\in\mathbb{F}_2}\sum_{y\in\mathcal{Y}} W(x,y)\log_2 \frac{W(x \mid y)}{Q(x)},$$

which is also the channel capacity of W.

Definition 2.4. The *Bhattacharyya parameter* of W is

$$Z(W) \coloneqq 2\sum_{y\in\mathcal{Y}} \sqrt{W(0,y)W(1,y)},$$

which is twice the Bhattacharyya coefficient between the joint distributions $W(0,\bullet)$ and $W(1,\bullet)$.

The overall goal is to construct, for some large N, an encoder $\mathcal{E}\colon \mathbb{F}_2^{RN} \to \mathbb{F}_2^N$ and a decoder $\mathcal{D}\colon \mathcal{Y}^N \to \mathbb{F}_2^{RN}$ such that the composition

$$U_1^{RN} \xmapsto{\mathcal{E}} X_1^N \xmapsto{W^N} Y_1^N \xmapsto{\mathcal{D}} \hat{U}_1^{RN}$$

is the identity map as frequently as possible, and R as close to the channel capacity $I(W)$ as possible.

Definition 2.5. Call N the block length. Call R the code rate. Denote by P_{e}, called the block error probability, the probability that $\hat{U}_1^{RN} \neq U_1^{RN}$.

To reach the overall goal of constructing good error correcting codes, I will introduce the building block of all techniques we are to utilize–channel transformation.

2. Channel Transformation and Tree

This section motivates and defines the channel transformation. For the precise connection between the transformation and the actual encoder/decoder design, please refer to Arıkan's original work [**Ari09**].

Let $G \in \mathbb{F}_2^{2\times 2}$ be the matrix

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Let $U_1, U_2 \in \mathbb{F}_2$ be two uniform random variables. Let $X_1^2 \in \mathbb{F}_2$ be the vector

$$\begin{bmatrix} X_1 & X_2 \end{bmatrix} \coloneqq \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},$$

or $X_1^2 \coloneqq U_1^2 G$ for short. Let $Y_1, Y_2 \in \mathcal{Y}$ be the outputs of two i.i.d. copies of W given the inputs X_1 and X_2, respectively. Then the combination of the two W's is the channel with input U_1^2 and output Y_1^2.

To split the combination of the channels, consider a two-step guessing job:

- Guess U_1 given Y_1^2.
- Guess U_2 given Y_1^2, assuming that the guess $\hat{U}_1$ of U_1 is correct.

Pretend that there is a channel $W^{(1)}$ with input U_1 and output Y_1^2; this channel captures the difficulty of the first step. Pretend also that there is a channel $W^{(2)}$ with input U_2 and output $Y_1^2 U_1$; this channel captures the difficulty of the second step. The precise definitions follows.

Definition 2.6. Define synthetic channels

$$W^{(1)}(y_1^2 \mid u_1) \coloneqq \sum_{u_2 \in \mathbb{F}_2} \frac{1}{2} W(y_1 \mid u_1 + u_2) W(y_2 \mid u_2),$$

$$W^{(2)}(y_1^2 u_1 \mid u_2) \coloneqq \frac{1}{2} W(y_1 \mid u_1 + u_2) W(y_2 \mid u_2).$$

Clearly $W^{(1)}$ and $W^{(2)}$ are of binary input and discrete output. It can be shown that they are symmetric, hence are SBDMCs. Therefore, the channel transformation $W \mapsto (W^{(1)}, W^{(2)})$ maps the set of SBDMCs to the Cartesian square thereof.

Once we accept the idea that the guessing jobs can be modeled as channels, we can talk about manipulating the channels as if they were actual objects instead of describing, abstractly, the change in ways we are guessing the random variables.

Particularly, we can easily imagine that the channel transformation applies recursively and gives birth to descendants $W^{(j_1)}$, $(W^{(j_1)})^{(j_2)}$, $((W^{(j_1)})^{(j_2)})^{(j_3)}$, and so on and so forth. This family tree of synthetic channels rooted at W is called the *channel tree*.

To construct a code, choose a large integer n and synthesize the depth-n descendants of W, which are of the form $\bigl(\cdots((W^{(j_1)})^{(j_2)})\cdots\bigr)^{(j_n)}$. Select a subset of those channels, which is equivalent to selecting a subset of indices $(j_1, j_2, \dots, j_n) \in \{1,2\}^n$. Call this subset $\mathcal{J}$. Then by transmitting messages through the synthetic channels in $\mathcal{J}$, a code is established. This code has block length $N = 2^n$, code rate $R = |\mathcal{J}|/2^n$, and block error probability upper bounded by

$$P_{\mathrm{e}} \leqslant \sum_{j_1^n \in \mathcal{J}} Z\bigl(\bigl(\cdots((W^{(j_1)})^{(j_2)})\cdots\bigr)^{(j_n)}\bigr). \tag{2.1}$$

In all papers I have seen, no upper bound on P_{e} other than inequality (2.1) was used. So we may pretend that the right-hand side of inequality (2.1) is the design block error probability of $\mathcal{J}$.

To construct good codes, it suffices to collect in $\mathcal{J}$ synthetic channels with small Z. But the more we collect, the higher the sum of Z's. This induces a trade-off between P_{e} and R, which is the subject of the current chapter. Let θ be the collecting threshold; that is, $\mathcal{J}$ collects synthetic channels whose Z falls below θ. Then θ parametrizes the trade-off in the sense that $P_{\mathrm{e}} < N\theta$ and R is the density of the synthetic channels whose Z falls below θ.

In the next section, I will introduce some stochastic processes that help us comprehend the trade-off between R and P_{e}.

3. Channel and Parameter Processes

We are to define some stochastic processes whose sample space is independent of those of the channels and user messages. To help distinguish the new source of randomness, I typeset the relevant symbols (such as P, E) in sans serif font.

Definition 2.7. Let $\mathsf{J}_1, \mathsf{J}_2, \dots$ be i.i.d. tosses of a fair coin with sides $\{1,2\}$. That is,

$$\mathsf{J}_n \coloneqq \begin{cases} 1 & \text{w.p. } 1/2, \\ 2 & \text{w.p. } 1/2. \end{cases}$$

Let $\mathsf{W}_0, \mathsf{W}_1, \mathsf{W}_2, \dots$, or $\{\mathsf{W}_n\}$ in short, be a stochastic process of SBDMCs defined as follows:

- $\mathsf{W}_0 \coloneqq W$; and
- $\mathsf{W}_{n+1} \coloneqq \mathsf{W}_n^{(\mathsf{J}_{n+1})}$.

This is called the *channel process*.

Definition 2.8. Let $\{\mathsf{H}_n\}$ be the stochastic process obtained by applying H to $\{\mathsf{W}_n\}$. That is, $\mathsf{H}_n \coloneqq H(\mathsf{W}_n)$. It is called *Arıkan's martingale*.

Definition 2.9. Let $\{\mathsf{Z}_n\}$ be the stochastic process obtained by applying Z to $\{\mathsf{W}_n\}$. That is, $\mathsf{Z}_n \coloneqq Z(\mathsf{W}_n)$. It is called *Bhattacharyya's supermartingale*.

The remainder of this section is devoted to explaining that Arıkan's martingale is a martingale and Bhattacharyya's supermartingale is a supermartingale, as well as other relations among H and Z. It will show that questions regarding the code performance can be passed to questions regarding the processes $\{\mathsf{H}_n\}$ and $\{\mathsf{Z}_n\}$.

Proposition 2.10. *Arıkan's martingale* $\{H_n\}$ *is a martingale.*

Proof. It suffices to check if $H(W^{(1)})+H(W^{(2)}) = 2H(W)$. Recall the inputs and outputs of $W^{(1)}$ and $W^{(2)}$; we have

$$\begin{aligned} H(W^{(1)}) + H(W^{(2)}) &= H(U_1 \mid Y_1^2) + H(U_2 \mid U_1 Y_1^2) = H(U_1^2 \mid Y_1^2) \\ &= H(X_1^2 \mid Y_1^2) = 2H(X \mid Y) = 2H(W). \end{aligned}$$

That finishes the proof. □

Proposition 2.11. *Bhattacharyya's supermartingale* $\{Z_n\}$ *is a supermartingale.*

Proof. It suffices to check if $Z(W^{(1)}) + Z(W^{(2)}) \leqslant 2Z(W)$. But that is the sum of inequalities (2.3) and (2.2) below. □

Lemma 2.12 (Evolution of Z)**.** *The following hold for all SBDMCs* W*:*

$$Z(W^{(2)}) = Z(W)^2, \tag{2.2}$$

$$Z(W^{(1)}) \leqslant 2Z(W) - Z(W)^2, \tag{2.3}$$

$$Z(W^{(1)}) \geqslant Z(W)\sqrt{2 - Z(W)^2}. \tag{2.4}$$

For a proof of the first two inequalities, see [**Ari09**, Proposition 5]. Regarding the third inequality, it is used in [**MHU16**, inequality (5)], wherein the authors cited [**RU08**, Exercise 4.62]. The proofs consist of elementary manipulations of summations and square roots.

The next lemma relates H and Z. Note that any relation automatically applies to $\{H_n\}$ and $\{Z_n\}$.

Lemma 2.13 (Z vs H)**.** *The following hold for all SBDMCs* W*:*

$$\begin{gathered} Z(W) \geqslant H(W), \\ Z(W)^2 \leqslant H(W), \\ 1 - Z(W) \geqslant (1 - H(W)) \ln 2. \end{gathered} \tag{2.5}$$

For proofs, see [**JA18**, Corollary 5]; note that the last two inequalities are specializations of $\phi(Z(W)) \leqslant H(W)$ for a smooth function $\phi(z) \coloneqq h_2((1-\sqrt{1-z^2})/2)$. For a visualization of the region where $(H(W), Z(W))$ could possibly be, see Figure 2.2.

Remark: Imagine that we write down the parameter processes in an infinite array

$$\begin{bmatrix} H_0 & H_1 & H_2 & H_3 & \cdots & \\ Z_0 & Z_1 & Z_2 & Z_3 & \cdots & \end{bmatrix}$$

then Propositions 2.10 and 2.11 and Lemma 2.12 are some horizontal relations, and Lemma 2.13 is some vertical relations. These lemmas help us predict where the processes are going. For example, if we happen to know $H_n \to 0$, then $Z_n \to 0$ because inequality (2.5) says $Z_n \leqslant \sqrt{H_n}$. I call this the *common-fate property*.

The next lemma justifies why we want to predict the processes—because it helps us evaluate code performance.

Lemma 2.14. *Fix an* n*. Declare a code by letting* $\mathcal{J}$ *collect synthetic channels with* Z *less than the threshold* θ*. Then the code rate* R *is* $P\{Z_n < \theta\}$*.*

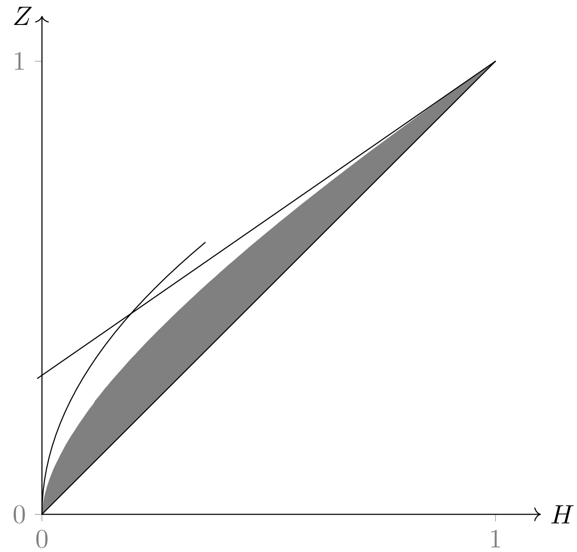

FIGURE 2.2. The possible region where $(H(W), Z(W))$ could lie in. The dark region is the exact one, whereas the outer boundaries are loosened to two pieces so that they are easier to describe—they are a parabola and a line of slope $\ln 2$.

PROOF. Recall that $R = |\mathcal{J}|/2^n$, where $|\mathcal{J}|$ is the number of depth-n synthetic channels of the form $\big(\cdots((W^{(j_1)})^{(j_2)})\cdots\big)^{(j_n)}$ in $\mathcal{J}$. Since W_n assumes each depth-n channel with probability $1/2^n$, the code rate R is the probability that W_n is in $\mathcal{J}$. This quantity, by the definition of $\mathcal{J}$, is the probability that $Z(\mathsf{W}_n) < \theta$; and $Z(\mathsf{W}_n)$ is just Z_n. This finishes the proof. $\square$

Recap: When declaring a code by letting $\mathcal{J}$ collect synthetic channels with Z less than θ, the block error probability has an upper bound $P_\mathrm{e} < N\theta$, and the code rate has an easy expression $R = \mathsf{P}\{\mathsf{Z}_n < \theta\}$. In summary, the following formula depicts the trade-off between P_e and R:

$$\mathsf{P}\{\mathsf{Z}_n < P_\mathrm{e}/N\} \approx R.$$

To rephrase it, we are interested in the cdf of Z_n, especially how close it is to the y-axis.

Our end-of-chapter goal is to characterize the pairs $(\pi, \rho) \in [0,1]^2$ such that

$$\mathsf{P}\big\{\mathsf{Z}_n < e^{-2^{\pi n}}\big\} > I(W) - 2^{-\rho n}.$$

That immediately implies the existence of codes with P_e on the order of $\exp(-2^{\pi n})$ and the gap to capacity $I(W) - R$ on the order of $2^{-\rho n}$. In the next section, we approach this goal via first showing

$$\mathsf{P}\big\{\mathsf{Z}_n < e^{-n^{2/3}}\big\} > I(W) - 2^{-\varrho n + o(n)} \tag{2.6}$$

for some $\varrho > 0$, where $1/\varrho$ is sometimes called the *scaling exponent*.

4. Scaling Exponent Regime

As far as I can tell, the only way to show inequality (2.6) is through the eigen behavior of Z_n. More precisely, I will first declare a concave function $h\colon [0,1] \to [0,1]$ and estimate the supremum

$$2^{-\varrho} \coloneqq \sup_{W\colon \mathrm{SBDMC}} \frac{h(Z(W^{(1)})) + h(Z(W^{(2)}))}{2h(Z(W))}.$$

This is called the *eigen behavior* of Z_n. From that we can infer $E[Z_n] \leqslant Z_0 2^{-\varrho}$, and then $P\{\exp(-n^{2/3}) \leqslant Z_n \leqslant 1-\exp(-n^{2/3})\} < 2^{-\varrho n+o(n)}$, followed by $P\{Z_n \to 0\} = I(W)$, and finally the *en23 behavior* $P\{Z_n < \exp(-n^{2/3})\} > I(W) - 2^{-\varrho n+o(n)}$.

Let us walk through a toy example before we dive into the general case. Consider any binary erasure channel (BEC) W. Then inequality (2.3) assumes equality; that is, $Z(W^{(1)}) = 2Z(W) - Z(W)^2$. Declare an "eigenfunction" $h(z) \coloneqq \sqrt{z(1-z)}$. Then

$$\sup_{W\colon \mathrm{BEC}} \frac{h(Z(W^{(1)})) + h(Z(W^{(2)}))}{2h(Z(W))} = \sup_{0<z<1} \frac{h(2z-z^2) + h(z^2)}{2h(z)} = \frac{\sqrt{3}}{2}. \tag{2.7}$$

This means that $\sqrt{3}/2$ is the "eigenvalue" corresponding to h, hence the name eigen behavior.

We now deduce that $E[h(Z_{n+1})] = E[E[h(Z_{n+1}) \mid Z_n]] \leqslant E[h(Z_n)\sqrt{3}/2]$. Applying this iteratively, we arrive at $E[h(Z_n)] \leqslant h(Z_0)(\sqrt{3}/2)^n$. From there we further deduce that, by Markov's inequality,

$$\begin{aligned} &P\{e^{-n^{2/3}} \leqslant Z_n \leqslant 1 - e^{-n^{2/3}}\} = P\{h(Z_n) \geqslant h(e^{-n^{2/3}})\} \\ &\leqslant \frac{E[h(Z_n)]}{h(\exp(-n^{2/3}))} \leqslant \frac{h(Z_0)(\sqrt{3}/2)^n}{h(\exp(-n^{2/3}))} < \frac{(\sqrt{3}/2)^n}{\exp(-n^{2/3})} < \Big(\frac{\sqrt{3}}{2}\Big)^{n-o(n)}. \end{aligned} \tag{2.8}$$

What we see here is that Z_n refuses to stay around the middle of the interval $[0,1]$, and we can quantify how unwilling Z_n is.

Next, we run the following analysis nonsense to derive that $P\{Z_n \to 0\} = I(W)$: Since $E[h(Z_n)]$ decays exponentially fast in n while h is bounded, $h(Z_n) \to 0$ almost surely. This implies that Z_n might (a) converge to 0, (b) converge to 1, or (c) jump back and forth between 0 and 1. The last one cannot happen to a supermartingale, so there exists a random variable Z_∞ such that $Z_n \to Z_\infty \in \{0,1\}$. By Lemma 2.13, $\{H_n\}$ obeys the same law—there exists a random variable H_∞ such that $H_n \to H_\infty \in \{0,1\}$. Now $P\{Z_n \to 1\} = P\{H_n \to 1\} = P\{H_\infty = 1\} = E[H_\infty] = H_0 = H(W)$, so the complement probability is $P\{Z_n \to 0\} = I(W)$.

Inequality (2.8) and $P\{Z_n \to 0\} = I(W)$ are what we need to derive the en23 behavior of Z_n—the former expels Z_n to the ends of the interval $[0,1]$, and the latter predicates how much Z_n goes to which end. In detail, consider the bad event $B_n \coloneqq \{Z_n \geqslant \exp(-n^{2/3})$ but $Z_m \to 0\}$. This event collects the samples where Z_n used to be moderate or bad but turns out to be good. By Lemma 2.12, there must be some $m \geqslant n$ such that Z_m "visits the middle", i.e., $\exp(-m^{2/3}) \leqslant Z_m \leqslant 1-\exp(-m^{2/3})$. But since the upper bound by inequality (2.8) is a geometric series in m, the probability that some Z_m (for some $m \geqslant n$) visit the middle is at most $(\sqrt{3}/2)^{n-o(n)}$. We conclude that $P\{Z_n < \exp(-n^{2/3})\} \geqslant P\{Z \to 0\} - P(B_n) > I(W) - (\sqrt{3}/2)^{n-o(n)}$.

Remark on the last four paragraphs: We used an eigen-pair $(h, \sqrt{3}/2)$ to quantify the pace of decay of $E[h(Z_n)]$ and further computed how much Z_n visits the middle or close to 0. The general SBDMC version of this argument comes with two modifications. First, the eigenvalue $\sqrt{3}/2$ could be improved if we use a better eigenfunction h. Second, $Z(W^{(1)})$ is not exactly $2Z(W) - Z(W)^2$ but can be as low as $Z(W)\sqrt{2 - Z(W)^2}$ (Lemma 2.12). This has to be taken into consideration for the generalization of supremum (2.7).

The remainder of this section deals with the general SBDMC case.

Theorem 2.15 (SBDMC scaling exponent)**.** *Assume SBDMCs. There exists a concave function* $h : [0,1] \to [0,1]$ *such that* $h(0) = h(1) = 0$ *and*

$$\sup_{W:\,SBDMC} \frac{h(Z(W^{(1)})) + h(Z(W^{(2)}))}{2h(Z(W))} > 2^{-1/4.714}.$$

Proof. Thanks to Lemma 2.12, it suffices to find a function h that minimizes

$$\sup_{0<z<1} \sup_{z\sqrt{2-z^2} \leqslant z' \leqslant 2z-z^2} \frac{h(z') + h(z^2)}{2h(z)}. \tag{2.9}$$

It is not known yet if there is an analytic form for such h, but numerical computation in [**MHU16**, Theorem 2] suggests that $2^{-1/4.717}$ is achievable by some spline h, and I will stick to this number.

Note one: By a compactness argument, supremum (2.9) is strictly less than 1. As long as it is less than 1, the arguments in the remainder of this section apply with $2^{-1/4.717}$ replaced by the weaker supremum. Therefore, readers should not worry about whether $2^{-1/4.714}$ is mathematically sound.

Note two: Despite of potential rounding errors, there is another reason why I think $2^{-1/4.714}$ is not the final value. Recall that we took the supremum over $z\sqrt{2 - z^2} \leqslant z' \leqslant 2z - z^2$; this is a pessimistic estimate and chances are that we will have tighter inequalities to bound $Z(W^{(1)})$. ☐

Now, let us start deriving the en23 behavior of SBDMCs.

Lemma 2.16 (From eigen to en23)**.** *Fix* $\varrho \coloneqq 1/4.714$*. Assume Theorem 2.15. Then*

$$P\{Z_n < e^{-n^{2/3}}\} > I(W) - 2^{-\varrho n + o(n)}. \tag{(2.6)'s copy}$$

Proof. The proof was sketched when we walked through the toy example. First, Theorem 2.15 yields that $E[h(Z_{n+1})] \leqslant E[h(Z_n)2^{-\varrho}]$. Telescoping, we obtain that $E[h(Z_n)] \leqslant Z_0 2^{-\varrho n} < 2^{-\varrho n + o(n)}$. By Markov's inequality, we see

$$P\{e^{-n^{2/3}} \leqslant Z_n \leqslant 1 - e^{-n^{2/3}}\} \leqslant \frac{E[h(Z_n)]}{h(\exp(-n^{2/3}))} < 2^{-\varrho n + o(n)}. \tag{2.10}$$

Next, we recall why $P\{Z_n \to 0\} = I(W)$: By that $\{h(Z_n)\}$ is a bounded supermartingale and decays by a constant factor every time n increases, it converges to 0 almost surely. Since h is concave, $h(0)$ and $h(1)$ are the only places that evaluate to 0, which means that Z_n is getting closer and closer to either 0 or 1. By Lemma 2.12, or that $\{Z_n\}$ is a supermartingale, it cannot jump from the neighborhood of 0 to the neighborhood of 1, so each realization of $\{Z_n\}$ must choose either 0 or 1 and converge to it. By Lemma 2.13, $\{H_n\}$ must converge, and is converging to the same limit $\{Z_n\}$ does. But as $\{H_n\}$ is a bounded martingale, we know that $E[\lim_{n\to\infty} H_n] = \lim_{n\to\infty} E[H_n] = H_0 = H(W)$. Hence $H(W)$ is the probability that $\{H_n\}$ and $\{Z_n\}$ converge to 1. Its complement is that $\{H_n\}$ and $\{Z_n\}$ converge to 0 with probability $1 - H(W) = I(W)$.

Lastly, I explain why $P\{Z_n < \exp(-n^{2/3})\} > P\{Z_n \to 0\} - 2^{-\varrho n+o(n)}$: In general, we would like to believe that if a realization of $\{Z_n\}$ converges to 0, its prefix (the first few terms) would be rather small. But there are exceptions: Let B_n be the exceptional event $\{Z_m \to 0 \text{ but } Z_n \geqslant \exp(-n^{2/3})\}$. We would like to estimate $P(B_n)$. To do so, realize that if $Z_n \geqslant \exp(-n^{2/3})$, then either

- $\exp(-n^{2/3}) \leqslant Z_n \leqslant 1 - \exp(-n^{2/3})$, or
- $1 - \exp(-n^{2/3}) \leqslant Z_n$ and Z_m will visit the closed interval $[\exp(-m^{2/3}), \exp(-m^{2/3})]$ for some later $m > n$.

Either case, Z_n or its descendants will step in the left-hand side of inequality (2.10). Sum inequality (2.10) over $m \geqslant n$ and apply the union bound over $m \geqslant n$; we get an upper bound $P(B_n) < \sum_{m \geqslant n} 2^{-\varrho m+o(m)} < 2^{-\varrho n+o(n)}$. Consequently, $P\{Z_n < \exp(-n^{2/3})\} \geqslant P\{Z_n \to 0\} - P(B_n) > I(W) - 2^{-\varrho n+o(n)}$. That closes the proof. □

So far I have proved $P\{Z_n < \exp(-n^{2/3})\} > I(W) - 2^{-n/4.714+o(n)}$. In the next section, I will prove the same inequality with $\exp\bigl(-e^{n^{1/3}}\bigr)$ in place of $\exp(-n^{2/3})$. After that, I will prove the final goal—$P\{Z_n < \exp(-\ell^{\pi n})\} > I(W) - 2^{-\rho n+o(n)}$ for some pairs (π, ρ) lying in the shaded area in Figure 2.1.

5. Stepping Stone Regime

In this section, we will verify the *een13 behavior* $P\bigl\{Z_n < \exp\bigl(-e^{n^{1/3}}\bigr)\bigr\} > I(W) - 2^{-n/4.714+o(n)}$ on top of the en23 behavior proved in the last section. The idea behind the proof is to keep track of how many Z_m are moderately small, i.e., $Z_m < \exp(-m^{2/3})$ and how many descendants Z_n thereof become even smaller, i.e., $Z_n < \exp\bigl(-e^{n^{1/3}}\bigr)$ for some $n > m$.

To understand the idea better, pretend that we have a Z_n that is moderately small—about $\exp(-n^{2/3})$ small. Then the punchline here is that squaring (Z_n^2) will scale Z_n down rapidly, while doubling $(2Z_n - Z_n^2)$ barely does anything to the order of magnitude of Z_n. So the problem boils down to counting how many times a trajectory of $\{Z_n\}$ undergoes the squaring branches. This number obeys a binomial distribution whose limiting behavior is well-known.

Before the actual proof, let me walk through a tentative strategy to demonstrate what could go wrong. Let $m < n$ be two large numbers, then Lemma 2.16 yields

$$P\bigl\{Z_m < e^{-m^{2/3}}\bigr\} > I(W) - 2^{-\varrho m+o(m)}. \tag{2.11}$$

In order to end up with $Z_n \approx \exp\bigl(-e^{n^{1/3}}\bigr)$ from $Z_m \approx \exp(-m^{2/3})$, it requires, among the remaining $n-m$ Bernoulli trials, $n^{1/3}$ squaring branches. By Hoeffding's inequality, it would not meet the requirement with probability

$$\exp\Bigl(-\Omega\Bigl(\frac{n-m}{2} - n^{1/3}\Bigr)\Bigr). \tag{2.12}$$

Now we see the dilemma: If m is too small compared to n, i.e., $m < n - \Omega(n)$, the right-hand side of inequality (2.11) is $I(W) - 2^{-\varrho n+\Omega(n)}$, which is too far away from the expected code rate $I(W) - 2^{-\varrho n+o(n)}$. If, otherwise, m is comparable to n, i.e., $m = n - o(n)$, the right-hand side of inequality (2.12) is $\exp(-o(n))$, which exceeds the expected gap to capacity $2^{-\varrho n+o(n)}$.

The preceding hand-waving argument demonstrates that no m can settle the argument down once and for all. So the second punchline is to use multiple m's. In my case, I choose $m = \sqrt{n}, 2\sqrt{n}, \ldots, n - \sqrt{n}$ for some perfect square n to apply

Hoeffding's inequality $\sqrt{n}$ times. There are flexibilities in choosing m's; for instance, when n is not a perfect square, using $m = \lceil\sqrt{n}\rceil, 2\lceil\sqrt{n}\rceil, \dots, \lfloor n/\lceil\sqrt{n}\rceil\rfloor\lceil\sqrt{n}\rceil$ would not alter the proof up to some little-o terms. So let us presume that n is always a perfect square.

That could be the end of the proof if it were not for the falling of the assumption: For an m, if the first few branches after Z_m are doubling it, Z_m will soon become so large that the 2 in $2Z_m$ is not negligible. For this concern, Hoeffding's inequality does not help—it does not control whether the $n^{1/3}$ squaring branches take place within the last few branches or spread out evenly. To resolve that, we need to keep an eye on the entire trajectory $Z_m, Z_{m+1}, \dots, Z_n$ to make sure that it stays in the range where doubling is negligible.

The actual proof provided below aims to mimic the tentative argument for multiple m's at once while resolving the issue that doubling too much could break things. I will keep monitoring two conditions—whether a trajectory of Z_n undergoes sufficiently many squaring branches and whether that trajectory of Z_n stays low enough such that doubling is negligible.

Lemma 2.17 (From en23 to een13)**.** *Given Lemma 2.16, that is, given*

$$P\{Z_n < e^{-n^{2/3}}\} > I(W) - 2^{-\varrho n + o(n)},$$

we have

$$P\{Z_n < \exp(-e^{n^{1/3}})\} > I(W) - 2^{-\varrho n + o(n)}. \tag{2.13}$$

PROOF. (Select constants.) Consider the stochastic process $\{19J_n/20\}$. Since $2z - z^2 \leqslant 2z \leqslant z^{19/20}$ whenever $z < 2^{-20}$, we have $Z_{n+1} \leqslant Z_n^{19J_{n+1}/20}$ whenever $Z_n < 2^{-20}$. Also notice that, numerically, $E[J_n^{-1}]6^{1/20} \approx 0.8202 < 2^{-1/3.6}$.

(Define events.) Let n be a perfect square. Let E_0^0 be the empty event. For every $m = \sqrt{n}, 2\sqrt{n}, \dots, n - \sqrt{n}$, we define five series of events A_m, B_m, C_m, E_m, and E_0^m inductively as below: Let A_m be $\{Z_m < \exp(-m^{2/3})\} \setminus E_0^{m-\sqrt{n}}$. Let B_m be a subevent of A_m where $Z_l \geqslant 2^{-20}$ for some $l \geqslant m$. Let C_m a subevent of A_m where

$$J_{m+1}J_{m+2}\cdots J_{m+\sqrt{n}} \leqslant 6^{\sqrt{n}/20}. \tag{2.14}$$

Let E_m be $A_m \setminus (B_m \cup C_m)$. Let E_0^m be $E_0^{m-\sqrt{n}} \cup E_m$. Let a_m, b_m, c_m, e_m, and e_0^m be the probability measures of the corresponding capital letter events. Moreover, let g_m be $I(W) - e_0^m$.

(Bound b_m/a_m from above.) Conditioning on A_m, we want to estimate the probability that $Z_l \geqslant 2^{-20}$ for some $l \geqslant m$. Recall that $\{Z_l\}$ is a supermartingale. Hence by Ville's inequality [**Dur19**, Exercise 4.8.2], $P\{Z_l \geqslant 2^{-20}$ for some $l \geqslant m \mid A_m\} \leqslant 2^{20}Z_m < 2^{20}\exp(-m^{2/3})$. This is an upper bound on b_m/a_m and will be summoned in inequality (2.15).

(Bound c_m/a_m from above.) We want to estimate how often inequality (2.14) happens. That is the probability that $(J_{m+1}J_{m+2}\cdots J_{m+\sqrt{n}})^{-1} \geqslant 6^{-\sqrt{n}/20}$. This probability cannot exceed $E[(J_{m+1}J_{m+2}\cdots J_{m+\sqrt{n}})^{-1}]6^{\sqrt{n}/20} = E[J_1^{-1}]^{\sqrt{n}}6^{\sqrt{n}/20} = (E[J_1^{-1}]6^{1/20})^{\sqrt{n}} < 2^{-\sqrt{n}/3.6}$ by Markov's inequality. This is an upper bound on c_m/a_m and will be summoned in inequality (2.15).

(Bound $(g_{m-\sqrt{n}} - a_m)^+$ from above.) By definitions, $g_{m-\sqrt{n}} - a_m = I(W) - (e_0^{m-\sqrt{n}} + a_m)$. The definition of A_m forces it to be disjoint from $E_0^{m-\sqrt{n}}$, thus

$e_0^{m-\sqrt{n}}+a_m$ is the probability measure of $E_0^{m-\sqrt{n}}\cup A_m$. This union event must contain the event $\{Z_m<\exp(-m^{2/3})\}$ by how A_m was defined. Recall the en23 behavior $P\{Z_m<\exp(-m^{2/3})\}>I(W)-\ell^{-\varrho m+o(m)}$. Chaining all inequalities together, we deduce $g_{m-\sqrt{n}}-a_m<\ell^{-\varrho m+o(m)}$. Let $(g_{m-\sqrt{n}}-a_m)^+$ be $\max(0,g_{m-\sqrt{n}}-a_m)$ so we can write $(g_{m-\sqrt{n}}-a_m)^+<\ell^{-\varrho m+o(m)}$. This upper bound will be summoned in inequality (2.15).

(Bound $e_0^{n-\sqrt{n}}$ from below.) We start rewriting g_m with g_m^+ being $\max(0,g_m)$:

$$\begin{aligned}
g_m &= I(W)-e_0^m = I(W)-(e_0^{m-\sqrt{n}}+e_m) = I(W)-e_0^{m-\sqrt{n}}-e_m\\
&= g_{m-\sqrt{n}}-e_m = g_{m-\sqrt{n}}\Bigl(1-\frac{e_m}{a_m}\Bigr)+\frac{e_m}{a_m}(g_{m-\sqrt{n}}-a_m)\\
&\leqslant g^+_{m-\sqrt{n}}\Bigl(1-\frac{e_m}{a_m}\Bigr)+\frac{e_m}{a_m}(g_{m-\sqrt{n}}-a_m)^+\\
&\leqslant g^+_{m-\sqrt{n}}\Bigl(1-\frac{e_m}{a_m}\Bigr)+(g_{m-\sqrt{n}}-a_m)^+\\
&\leqslant g^+_{m-\sqrt{n}}\Bigl(\frac{b_m}{a_m}+\frac{c_m}{a_m}\Bigr)+(g_{m-\sqrt{n}}-a_m)^+\\
&< g^+_{m-\sqrt{n}}\Bigl(2^{20}e^{-m^{2/3}}+\ell^{-\sqrt{n}/3.6}\Bigr)+\ell^{-\varrho m+o(m)}. \qquad (2.15)
\end{aligned}$$

The first four equalities are by the definitions of g_m and E_0^m. The next equality is simple algebra. The next two inequalities are by $0\leqslant e_m/a_m\leqslant 1$. The next inequality is by the definition of E_m. The last inequality summons upper bounds derived in the last three paragraphs. The last line contains two terms in the big parentheses; between them, $2^{-\sqrt{n}/3.6}$ dominates $2^{20}\exp(-m^{2/3})$ once m is greater than $O(n^{3/4})$. Subsequently, we obtain this recurrence relation:

$$\begin{cases} g_{O(n^{3/4})}\leqslant 1,\\ g_m\leqslant 2g^+_{m-\sqrt{n}}\ell^{-\sqrt{n}/3.6}+\ell^{-\varrho m+o(m)}.\end{cases}$$

Solve it (cf. the master theorem); we get that $g_{n-\sqrt{n}}<\ell^{-\varrho n+o(n)}$. By the definition of $g_{n-\sqrt{n}}$, we immediately get $e_0^{n-\sqrt{n}}>I(W)-\ell^{-\varrho n+o(n)}$.

(Analyze $E_0^{n-\sqrt{n}}$.) We want to estimate H_n when $E_0^{n-\sqrt{n}}$ happens. To be precise, for each $m=\sqrt{n},2\sqrt{n},\dots,n-\sqrt{n}$, we attempt to bound $Z_{m+\sqrt{n}}$ when E_m happens. Fix an m. When E_m happens, its superevent A_m happens, so we know that $Z_m<\exp(-m^{2/3})$. But B_m does not happen, so $Z_l<2^{-20}$ for all $l\geqslant m$. This implies that $Z_{l+1}\leqslant Z_l^{19J_{l+1}/20}$ for those l. Telescope; $Z_{m+\sqrt{n}}$ is less than or equal to Z_m raised to the power of $J_{m+1}J_{m+2}\cdots J_{m+\sqrt{n}}(19/20)^{\sqrt{n}}$. But C_m does not happen, so the product is greater than $6^{\sqrt{n}/20}(19/20)^{\sqrt{n}}=(6(19/20)^{20})^{\sqrt{n}/20}>2^{\sqrt{n}/20}$. Jointly we have $Z_{m+\sqrt{n}}\leqslant Z_m^{2^{\sqrt{n}/20}}<\exp(-m^{2/3}2^{\sqrt{n}/20})$. Recall that $Z_{l+1}\leqslant 2Z_l$ for all $l\geqslant m+\sqrt{n}$. Then telescope again; $Z_n\leqslant 2^{n-m-\sqrt{n}}Z_{m+\sqrt{n}}<2^n\exp(-m^{2/3}2^{\sqrt{n}/20})<\exp\bigl(-e^{n^{1/3}}\bigr)$ provided that n is sufficiently large. In other words, $E_0^{n-\sqrt{n}}$ implies $Z_n<\exp\bigl(-e^{n^{1/3}}\bigr)$.

(Summary.) Now we may conclude, for all perfect squares n, that $P\bigl\{Z_n<\exp\bigl(-e^{n^{1/3}}\bigr)\bigr\}\geqslant P(E_0^{n-\sqrt{n}})=e_0^n>I(W)-\ell^{-\varrho n+o(n)}$. For non-squares, round n down to the nearest square and rerun the whole argument above. We will get $Z_n<2^n\exp(-m^{2/3}2^{\lfloor\sqrt{n}\rfloor/20})$ with probability at least $I(W)-\ell^{-\varrho\lfloor\sqrt{n}\rfloor^2+o(n)}$, which still leads to $\mathsf{P}\{\mathsf{Z}_n < \exp(-e^{n^{1/3}})\} > I(W) - \ell^{-\varrho n+o(n)}$. And hence the proof of the een13 behavior, inequality (2.13), is sound for all n. ☐

This section is parallel to [**WD18c**, section V], to [**WD19b**, appendix C.C], and to [**GRY19**, section 10.2]. Do not confuse this section with the next section. The subtlety is explained in [**WD18c**, section III].

Now we know $\mathsf{P}\{\mathsf{Z}_n < \exp(-e^{n^{1/3}})\} > I(W) - 2^{-\varrho n+o(n)}$. We are ready to learn what (π,ρ) pairs satisfy $\mathsf{P}\{\mathsf{Z}_n < \exp(-2^{\pi n})\} > I(W) - 2^{-\rho n+o(n)}$.

6. Moderate Deviations Regime

I will build upon the een13 behavior and utilize a technique similar to before to answer the following main question: Knowing $\mathsf{P}\{\mathsf{Z}_n < \exp(-2^{0.499n})\} > I(W) - 1/999$ [**AT09**] and $\mathsf{P}\{\mathsf{Z}_n < 1/999\} > I(W) - 2^{-\varrho n+o(n)}$ [**MHU16**], can we find a interpolating result between these two results? This section, finally, offers an answer by characterizing the region of pairs (π,ρ) that satisfy $\mathsf{P}\{\mathsf{Z}_n < \exp(-2^{\pi n})\} > I(W) - 2^{-\rho n+o(n)}$.

Recall $\varrho \coloneqq 1/4.714$. Let $h_2(p) \coloneqq -p\log_2 p - (1-p)\log_2(1-p)$ be the binary entropy function. Let $\mathcal{O} \subseteq [0,1/2]\times[0,\varrho]$ be an open region defined by the following criterion: for any $(\pi,\rho) \in \mathcal{O}$, the ray shooting from (π,ρ) toward the opposite direction of $(0,\varrho)$ does not intersect the function graph of $1-h_2$. See Figure 2.1; this criterion is equivalent to that (π,ρ) lies to the left of the convex envelope of $(0,\varrho)$ and $1-h_2$. This criterion is also equivalent to

$$1 - h_2\Big(\frac{\pi n}{n-m}\Big) > \frac{\rho n - \varrho m}{n-m} \tag{2.16}$$

for all $0 < m < n$. The last criterion is what will be used in the proof.

Theorem 2.18 (From een13 to e2pin)**.** *Fix a pair $(\pi,\rho) \in \mathcal{O}$. Given the conclusion of Lemma 2.17, that is, given*

$$\mathsf{P}\{\mathsf{Z}_n < \exp(-e^{n^{1/3}})\} > I(W) - 2^{-\varrho n+o(n)}, \tag{(2.13)'s copy}$$

then

$$\mathsf{P}\{\mathsf{Z}_n < e^{-2^{\pi n}}\} > I(W) - 2^{-\rho n+o(n)}. \tag{2.17}$$

Proof. (Select constants.) Since inequality (2.16) holds, there exists a small constant $\varepsilon > 0$ such that

$$1 - h_2\Big(\frac{\pi n}{n-m} + 2\varepsilon\Big) > \frac{\rho n - \varrho m}{n-m} \tag{2.18}$$

by the compactness argument. Fix this ε. There exists a small constant $\delta > 0$ such that $\mathsf{Z}_{n+1} \leqslant \mathsf{Z}_n^{\mathsf{J}_{n+1}(1-\varepsilon)}$ whenever $\mathsf{Z}_n < \delta$.

(Define events.) Let n be a perfect square. Let A_0^0 and E_0^0 be the empty event. For every $m = \sqrt{n}, 2\sqrt{n}, \dots, n-\sqrt{n}$, we define six series of events A_m, A_0^m, B_m, C_m, E_m, and E_0^m inductively as follows: Let A_m be $\{\mathsf{Z}_m < \exp(-e^{m^{1/3}})\} \setminus \mathsf{A}_0^{m-\sqrt{n}}$. Let A_0^m be $\mathsf{A}_0^{m-\sqrt{n}} \cup \mathsf{A}_m$. Let B_m be a subevent of A_m where $\mathsf{Z}_l \geqslant \delta$ for some $l \geqslant m$. Let C_m a subevent of A_m where

$$\mathsf{J}_{m+1}\mathsf{J}_{m+2}\cdots\mathsf{J}_n \leqslant 2^{\pi n+2\varepsilon(n-m)}. \tag{2.19}$$

Let E_m be $\mathsf{A}_m \setminus (\mathsf{B}_m \cup \mathsf{C}_m)$. Let E_0^m be $\mathsf{E}_0^{m-\sqrt{n}} \cup \mathsf{E}_m$. Let a_m, a_0^m, b_m, c_m, e_m, and e_0^m be the probability measures of the corresponding capital letter events. Moreover, let f_m be $I(W) - \mathsf{a}_0^m$ and let g_m be $I(W) - \mathsf{e}_0^m$.

(Bound b_m/a_m from above.) Conditioning on A_m, we want to estimate the probability that $Z_l \geqslant \delta$ for some $l \geqslant m$. Recall that Z_l is a supermartingale. Hence by Ville's inequality ([**Dur19**, Exercise 4.8.2]), $P\{Z_l \geqslant \delta \text{ for some } l \geqslant m \mid A_m\} \leqslant Z_m/\delta < \exp(-e^{m^{1/3}})/\delta$. This is an upper bound on b_m/a_m and will be summoned in inequality (2.20).

(Bound c_m/a_m from above.) We want to estimate how often inequality (2.19) happens. This is equivalent to asking how often do $n-m$ fair coin tosses end up with $\pi n + 2\varepsilon(n-m)$ heads. By the large deviations theory, this probability is less than 2 to the power of

$$-(n-m)\Big(1 - h_2\Big(\frac{\pi n}{n-m} + 2\varepsilon\Big)\Big).$$

By inequality (2.18), this exponent is less than $\varrho m - \rho n$. Thus, the probability is less than $2^{\varrho m - \rho n}$. This is an upper bound on c_m/a_m and will be summoned in inequality (2.20).

(Bound f_m^+ from above.) The definition of f_m reads $I(W) - a_0^m$. Here a_0^m is the probability measure of A_0^m, and A_0^m is a superevent of A_m by how the former is defined. Event A_0^m must contain $\big\{Z_m < \exp(-e^{m^{1/3}})\big\}$ by how A_m was defined. By the een13 behavior, $P\big\{Z_m < \exp(-e^{m^{1/3}})\big\} > I(W) - \ell^{-\varrho m + o(m)}$. Chaining all inequalities together, we infer that $f_m < \ell^{-\varrho m + o(m)}$. Let f_m^+ be $\max(0, f_m)$ so we can write $f_m^+ < \ell^{-\varrho m + o(m)}$. This upper bound will be summoned in inequality (2.20).

(Bound $e_0^{n-\sqrt{n}}$ from below.) We start rewriting $g_m - f_m^+$ with $(f_{m-\sqrt{n}} - a_m)^+$ being $\max(0, f_{m-\sqrt{n}} - a_m)$:

$$\begin{aligned}
g_m - f_m^+ &= I(W) - e_0^m - (I(W) - a_0^m)^+ \\
&= I(W) - e_0^{m-\sqrt{n}} - e_m - (I(W) - a_0^{m-\sqrt{n}} - a_m)^+ \\
&= g_{m-\sqrt{n}} - e_m - (f_{m-\sqrt{n}} - a_m)^+ \\
&\leqslant g_{m-\sqrt{n}} - e_m - \frac{e_m}{a_m}(f_{m-\sqrt{n}} - a_m)^+ \\
&\leqslant g_{m-\sqrt{n}} - e_m - \frac{e_m}{a_m}(f_{m-\sqrt{n}}^+ - a_m) \\
&= g_{m-\sqrt{n}} - f_{m-\sqrt{n}}^+ + f_{m-\sqrt{n}}^+\Big(1 - \frac{e_m}{a_m}\Big) \\
&\leqslant g_{m-\sqrt{n}} - f_{m-\sqrt{n}}^+ + f_{m-\sqrt{n}}^+\Big(\frac{b_m}{a_m} + \frac{c_m}{a_m}\Big) \\
&< g_{m-\sqrt{n}} - f_{m-\sqrt{n}}^+ + \ell^{-\varrho(m-\sqrt{n}) + o(m-\sqrt{n})}\Big(\exp(-e^{m^{1/3}})/\delta + 2^{\varrho m - \rho n}\Big). \qquad (2.20)
\end{aligned}$$

The first three equalities are by the definitions of g_m and f_m. The next inequality is by $0 \leqslant e_m/a_m \leqslant 1$. The next inequality is by $\max(0, f-a) = \max(a, f) - a \geqslant \max(0, f) - a$. The next equality is simple algebra. The next inequality is by the definition of E_m. The last inequality summons upper bounds derived in the last three paragraphs. Now the last line contains two terms in the big parentheses; between them, $2^{\varrho m - \rho n}$ dominates $\exp(-e^{m^{1/3}})/\delta$ once $n \to \infty$. Subsequently, we obtain this recurrence relation

$$\begin{cases} g_0 - f_0^+ = 0; \\ g_m - f_m^+ \leqslant g_{m-\sqrt{n}} - f_{m-\sqrt{n}}^+ + 2\ell^{-\rho n + o(n)}. \end{cases}$$

Solve it (cf. the Cesàro summation); we get that $g_{n-\sqrt{n}} - f^+_{n-\sqrt{n}} < \ell^{-\rho n+o(n)}$. Once again we summon $f^+_{n-\sqrt{n}} < \ell^{-\varrho(n-\sqrt{n})+o(n-\sqrt{n})} < \ell^{-\varrho n+o(n)}$; therefore $g_{n-\sqrt{n}} < \ell^{-\rho n+o(n)}$. Based on the definition of $g_{n-\sqrt{n}}$ we immediately get $e_0^{n-\sqrt{n}} > I(W) - \ell^{-\rho n+o(n)}$.

(Analyze $E_0^{n-\sqrt{n}}$.) We want to estimate Z_n when $E_0^{n-\sqrt{n}}$ happens. To be precise, for each $m = \sqrt{n}, 2\sqrt{n}, \dots, n-\sqrt{n}$, we attempt to bound Z_n when E_m happens . Fix an m. When E_m happens, its superevent A_m happens, so we know that $Z_m < \exp\bigl(-e^{m^{1/3}}\bigr)$. But B_m does not happen, so $Z_l < \delta$ for all $l \geqslant m$. This implies $Z_{l+1} \leqslant Z_l^{J_{l+1}(1-\varepsilon)}$ for those l. Telescope; Z_n is less than or equal to Z_m raised to the power of $J_{m+1}J_{m+2}\cdots J_n(1-\varepsilon)^{n-m}$. But C_m does not happen, so the product is greater than $2^{\pi n+2\varepsilon(n-m)}(1-\varepsilon)^{n-m}$, which is greater than $2^{\pi n}$ granted that $\varepsilon < 1/2$. Jointly we have $Z_n \leqslant Z_m^{2^{\pi n}} < \exp\bigl(-e^{m^{1/3}}2^{\pi n}\bigr) < \exp(-2^{\pi n})$. In other words, $E_0^{n-\sqrt{n}}$ implies $Z_n < \exp(-\ell^{\pi n})$.

(Summary.) Now we may conclude, for all perfect squares n, that $P\{Z_n < \exp(-2^{\pi n})\} \geqslant P(E_0^{n-\sqrt{n}}) = e_0^n > I(W) - 2^{-\rho n+o(n)}$. For non-squares, round n down to the nearest square and rerun the whole argument above. We will get $Z_n < 2^n \exp\bigl(-e^{m^{1/3}}2^{\pi\lfloor\sqrt{n}\rfloor^2}\bigr)$ with probability at least $I(W) - \ell^{-\rho\lfloor\sqrt{n}\rfloor^2+o(n)}$, which still leads to $P\{Z_n < \exp(-2^{\pi n})\} > I(W) - 2^{-\rho n+o(n)}$. And hence the proof of the moderate deviations behavior, inequality (2.17), is sound for all n. □

This section is parallel to [**WD18b**, section V], to [**WD18c**, section VI], to [**WD19b**, appendix C.D], and to [**GRY19**, section 10.3]. Do not confuse this section with the previous. The subtlety is explained in [**WD18c**, section III].

7. Chapter Summary

In this chapter, we defined SBDMC and set a goal—that we want to construct error correcting codes and characterize their block error probabilities P_e and codes rates R. We succeed in proving that Z_n is such that

$$P\{Z_n < \exp(-2^{\pi n})\} > I(W) - 2^{-\rho n+o(n)},$$

which implies that there are codes with $P_\mathrm{e} < 2^n \exp(-2^{\pi n})$ and $I(W) - R < 2^{-\rho n+o(n)}$. By a topological argument that fluctuates π, we get $P_\mathrm{e} < \exp(-2^{\pi n})$ for n very large. By fluctuating ρ, similarly, we get $I(W) - R < 2^{-\rho n}$ granted that n is astronomically large.

Corollary 2.19 (Good code for SBDMC)**.** *Over SBDMCs, polar coding as constructed by Arıkan enjoys, for any $(\pi, \rho) \in \mathcal{O}$, block error probability $\exp(-N^\pi)$ and code rate $I(W) - N^{-\rho}$ for large N.*

Over BEC, we have a better estimate of $\varrho = 1/3.627$ [**HAU14**]. In this case, the region $\mathcal{O}$ is with $(0, 1/3.627)$ in place of $(0, 1/4.714)$. See also Figure 2.3.

Corollary 2.20 (Good code for BEC)**.** *Over BECs, polar coding as constructed by Arıkan enjoys, for any (π, ρ) lying to the left of the convex envelope of $(0, 1/3.627)$ and $1 - h_2$, block error probability $\exp(-N^\pi)$ and code rate $I(W) - N^{-\rho}$ for large N.*

In the next chapter, I will present the dual picture to this chapter. The duality stands for three aspects: The scenario is in duality, the theorem statement is in duality, and the proof technique is in duality.

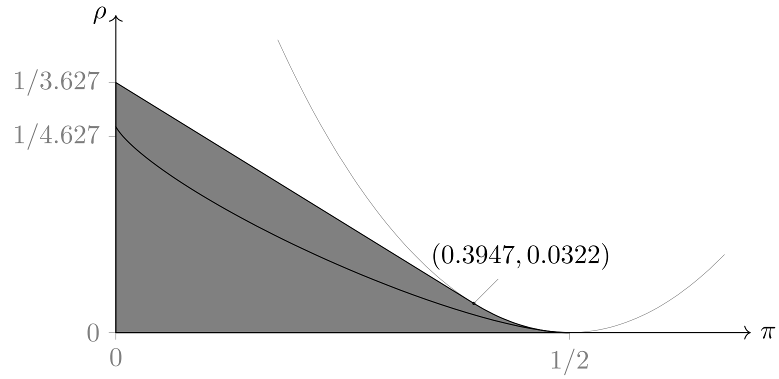

FIGURE 2.3. BEC special case of $(\log(-\log_2 P_\mathrm{e}), \log(C-R))$. The tangent line is from $(0, 1/3.6227)$ to $(0.4208, 0, 0182)$. The lower left curve is the previous result [**MHU16**], which attains $(0, 1/4.627)$.

CHAPTER 3

Asymmetric Channels

Researchers in the polar coding field, knowing the SBDMC results, had looked forward to applying polar coding to more channels or sources. As it turns out, polar coding also applies to source coding for lossy compression and noisy-channel coding over asymmetric binary-input discrete-output channels. But for polarization to work in those scenarios, the original analysis is subject to some modifications.

First and foremost, let us review lossy compression. In lossy compression, the compressor is presented a random variable Y and wants to send some messages to the decompressor so that the latter can generate a random variable X that is close enough to Y under a certain distance metric. A trade-off emerges as the compressor wants to send as few messages as possible while the X generated from those messages should be as close to Y as possible. This subject is usually referred to as the *rate–distortion theory*.

Polar coding applies to lossy compression by pretending that X and Y are the input and output of an abstract channel, called the *test channel*. Once there is a channel, it is polarized, a subset $\mathcal{J}$ of synthetic channels is selected, and synthetic channels are handled in two ways depending on whether they are in selected $\mathcal{J}$ or not. In this particular case, synthetic channels that are reliable correspond to the bits that snapshot the essence of Y, and hence is worth messaging to the decompressor. On the other hand, synthetic channels that are noisy correspond to the randomness that separates X from Y, which can be simulated by a pseudo random number generator on the decompressor side. By examining the pace the test channel polarizes, we gain control of the rate–distortion trade-off.

Secondly, let me elaborate on asymmetric channels. Asymmetric channels differ from symmetric ones by the fact that the uniform input distribution does not necessarily achieve capacity. As a result, a code designer needs to spend extra resources on shaping the input distribution apart from the usual anti-error routine. This shaping component of coding shares common elements with generating X from the messages the compressor sends as in the lossy compression scenario.

Polar coding applies to asymmetric channels by using a specialized decoder as an encoder. The new encoder polarizes the input distribution Q as if Q were a channel with constant output. Now the reliable descendants of Q are those who make Q in the shape of Q; they are inflexible, deducible form the other descendants, and unable to carry new information. On the other hand, the noisy descendants of Q are the source of the randomness of Q and can carry user messages. Meanwhile, the actual channel W through which we transmit messages is polarized as usual. It suffices to select a subset $\mathcal{J}$ of indices that correspond to, simultaneously, the noisy descendants of Q and the reliable descendants of W.

It happens that the performances of lossy compression and coding over asymmetric channels are both controlled by a stochastic process $\{T(\mathsf{W}_n)\}$. By the end of this chapter, I will prove

$$P\{T(W_n) < e^{-2^{\pi n}}\} > H(W) - 2^{-\rho n}$$

in order to describe the performances of polar coding in those scenarios.

1. Problem Setup—Lossy Compression

Let $\mathbb{F}_2$ be the finite field of order 2. Let $\mathcal{Y}$ be any finite set equipped with a probability measure $W(y)$. Let $\operatorname{dist}: \mathbb{F}_2 \times \mathcal{Y} \to [0,1]$ be a bounded distortion function; this function quantifies how well an $x \in \mathbb{F}_2$ represents/approximates a $y \in \mathcal{Y}$. For instance, we can have $\mathcal{Y} \subseteq [0,1]^2$ as a set of pairs and $\operatorname{dist}(x,(y_0,y_1)) \coloneqq y_x$.

The overall goal is to construct, for some large block length N, a compressor $\mathcal{C}: \mathcal{Y}^N \to \mathbb{F}_2^{RN}$ and a decompressor $\mathcal{D}: \mathbb{F}_2^{RN} \to \mathbb{F}_2^N$ such that the composition

$$Y_1^N \xmapsto{\mathcal{C}} U_1^{RN} \xmapsto{\mathcal{D}} X_1^N$$

minimizes $D \coloneqq E\left[\sum_{j=1}^N \operatorname{dist}(X_j, Y_j)/N\right]$, the long-term average of the point-wise distortion between X_1^N and Y_1^N, for a given code rate R. Or conversely, we want to minimizes R for a given D. The standard result follows.

Theorem 3.1 (Rate–distortion trade-off)**.** *Allow arbitrarily large N. Then the infimum of code rates R that are achievable within a fixed distortion Δ is*

$$R(\Delta) \coloneqq \min_{W(x|y)} I(X\,;Y), \tag{3.1}$$

where the minimum is taken over all transition arrays $W(x \mid y)$ such that, when X is governed by W, the expected distortion is bounded as $E[\operatorname{dist}(X,Y)] \leqslant \Delta$.

For a proof, see standard textbooks, e.g., [**Bla87**]. See Figure 3.1 for an example of the function $R(\Delta)$.

To apply polar coding, we fix beforehand a transition array $W(x \mid y)$ that minimizes formula (3.1). Let X and Y be governed by $W(x \mid y)$ and $W(y)$. And then pretend that W is a channel as if X were the input and Y were the output. (Although in reality, it is Y that is given to the compressor and the decompressor outputs X.) This W is called the test channel and we can apply the channel transformation to it. It remains to specify what to do to the descendants of W and explain how the behavior of $\{W_n\}$ relates to the compression metrics D and R.

Note that a test channel is not a priori an SBDMC—it satisfies all conditions of being an SBDMC but not the symmetry one. Plus, we sometimes want to communicate over asymmetric channels. So we have to deal with asymmetric channels sooner or later. Or now.

2. Problem Setup—Asymmetric Channel

A BDMC (binary-input discrete-output memoryless channel) is an SBDMC except that the involution $\sigma: \mathcal{Y} \to \mathcal{Y}$ is not mandatory. The main aftermath after taking away the symmetry is that the uniform distribution is not guaranteed to be optimal in terms of achievable code rates. That being the case, there always exists another input distribution that achieves the optimal rate.

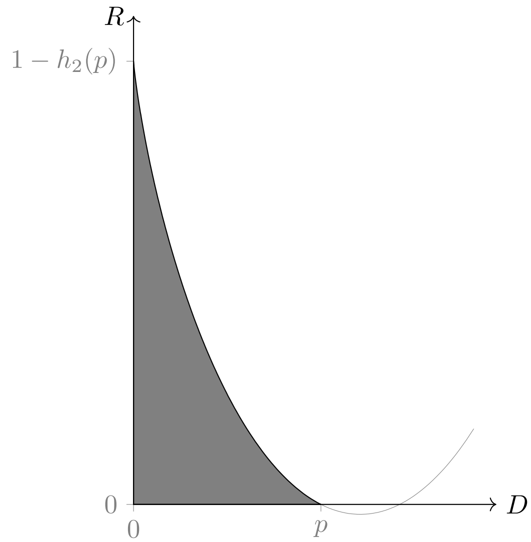

FIGURE 3.1. Assume lossy compression for a binary source of mean p and Hamming distortion function $\mathrm{hdis}(x, y) := \mathbb{I}\{x \neq y\}$. The shaded area is where (R, D) is not possible. The curve part is $1 - h_2$ shifted.

Theorem 3.2 (Asymmetric chapacity)**.** *The channel capacity of a BDMC W, the supremum of codes rates at which reliable communication can happen, is*

$$\max_{Q(x)} I(X\,;Y),$$

where the maximum is taken over all input distributions $Q(x)$ on $\mathbb{F}_2$.

For a proof, see standard textbooks, e.g., [Bla87].

Fix this Q from now on. Then we can apply channel transformation following the same philosophy. In detail, we define a vector $U_1^2 \in \mathbb{F}_2^2$ by

$$\begin{bmatrix} U_1 & U_2 \end{bmatrix} := \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \tag{3.2}$$

or $U_1^2 := X_1^2 G^{-1}$ for short. Note that this implicitly assigns a non-uniform, non-product distribution to U_1^2 if Q is not uniform to begin with. Ignoring that, we proceed to define $W^{(1)}$ to be an abstract channel with input U_1 and output Y_1^2, and $W^{(2)}$ a channel with input U_2 and output $Y_1^2 U_1$.

The question to answer is, How do asymmetric channels evolve under channel transformation? It turns out that nothing really changes; the old theory extends to the new channels seamlessly. And it is due to a symmetrization technique.

3. Channel Symmetrization

Given a BDMC W, we want to find an SBDMC $\tilde{W}$ such that any meaningful property concerning the descendants of $\tilde{W}$ automatically applies to those of W. To that end, a strategy is to define an equivalence relation $\cong$ such that (a) a BDMC is equivalent to at least one SBDMC, (b) channel parameters such as H and Z are functions in classes, meaning that $H(W) = H(\tilde{W})$ if $W \cong \tilde{W}$, and (c) the channel transformation respects the equivalence relation, meaning that $W^{(j)} \cong \tilde{W}^{(j)}$ if $W \cong \tilde{W}$. If such relation can be found, than almost all questions we want to ask about W have answers when we, instead, ask an SBDMC in the same class as W.

Definition 3.3. Two BDMCs, W and $\tilde{W}$, are said to be equivalent, denoted by $W \cong \tilde{W}$, if $\{W(0 \mid Y), W(1 \mid Y)\}$ and $\{\tilde{W}(0 \mid \tilde{Y}), \tilde{W}(1 \mid \tilde{Y})\}$ obey the same distribution on the power set of $[0,1]$.

Let me briefly remark what $\cong$ identifies. For one, the labeling on $\mathcal{Y}$ is not important; after all, the decoder only care about the posterior probabilities. For two, if two outputs y, y' have the same posterior probabilities, that is, $W(x \mid y) = W(x \mid y')$ for both $x \in \mathbb{F}_2$, the decoder might as well identify y and y'. For three, relabeling the input $\mathbb{F}_2$ does not matter; the decoder just cares about how biased $\{W(0 \mid Y), W(1 \mid Y)\}$ is, but not about toward which way it biases.

Lemma 3.4 (Reduction to symmetry)**.** *For any BDMC W, it is equivalent to at least one SBDMC.*

PROOF. Let $F \in \mathbb{F}_2$ be a "flag" that obeys an independent uniform distribution on $\mathbb{F}_2$. Let $\tilde{W}$ be a channel with input $X - F$ and output (F, Y). Intuition: When the encoder attempts to input X into a channel, it sees the flag F and input $X - F$ instead; the decoder also sees the flag F so it would simply add that back after all the decoding jobs.

This $\tilde{W}$ is a SBDMC because adding F to X will turn the input into a uniform random variable. This $\tilde{W}$ is equivalent to W because

$$\{\tilde{W}(0 \mid fy), \tilde{W}(1 \mid fy)\} = \{\tilde{W}(0 - f \mid fy), \tilde{W}(1 - f \mid fy)\} = \{W(0 \mid y), W(1 \mid y)\}.$$

That is, the random sets $\{\tilde{W}(0 \mid fy), \tilde{W}(1 \mid fy)\}$ and $\{W(0 \mid y), W(1 \mid y)\}$ coincide, and hence obey the same distribution. □

Before I state (b) that channel parameters are function in classes, one more parameter is defined to be utilized in the remainder of this chapter. Note that the definitions of H and Z automatically apply to the asymmetric case.

Definition 3.5. Define the *total variation norm* of W to be

$$T(W) := \sum_{y \in \mathcal{Y}} W(y) \sum_{x \in \mathbb{F}_2} \left| W(x \mid y) - \frac{1}{2} \right|,$$

which is the total variation distance from $W(\bullet \mid y)$ to the uniform distribution, weighted by the frequency each $y \in \mathcal{Y}$ appears.

Lemma 3.6 (Parameters in class)**.** *For any two equivalent BDMCs, W and $\tilde{W}$,*

$$H(W) = H(\tilde{W}), \quad Z(W) = Z(\tilde{W}), \quad T(W) = T(\tilde{W}).$$

PROOF. By definition. □

Remark: When W is BDMC and $\tilde{W}$ is SBDMC, $I(W) = H(Q) - H(W) \neq 1 - H(\tilde{W}) = I(\tilde{W})$ unless $H(Q) = 1$ (i.e., Q is uniform). What used to be $I(\tilde{W})$ in the last chapter becomes $1 - H(W)$ in this chapter, In particular, results of the form $P\{Z_n < \theta(n)\} > I(\tilde{W}) - \gamma(n)$ are becoming $P\{Z_n < \theta(n)\} > 1 - H(W) - \gamma(n)$ when cited.

Lemma 3.7 (Transformation in class)**.** *For any two equivalent BDMCs, W and $\tilde{W}$,*

$$W^{(1)} \cong \tilde{W}^{(1)}, \qquad W^{(2)} \cong \tilde{W}^{(2)}.$$

PROOF. By definition. □

We therefore conclude that, when it comes to channel transformation and channel processes, reasoning about W is logically equivalent to reasoning about $\tilde{W}$.

Being able to control the descendants of a BDMC, we will connect the performance of coding to channel parameters/processes in the next section.

4. Problem Reduction to Process

Now we can reduce lossy compression to operations on synthetic channels. Let W be a test channel of a lossy compression problem. Then, for any descendant W_n (including W itself), we want a compressor to observe W_n's output and send a message to the decompressor so that the latter can generate W_n's input. By that $H(\mathsf{W}_n) \to \mathsf{H}_\infty \in \{0, 1\}$, the descendants of W are either extremely reliable or completely noisy, and hence assume easy treatments.

When a realization of W_n is completely noisy, it means that its input X_n is almost irrelevant to its output Y_n. In this case, the decompressor does not need to know anything about Y_n and can query a pseudo random number generator to simulate X_n. When a realization of W_n is (extremely) reliable, on the other hand, its input X_n depends (heavily) on its output Y_n. In this case, the compressor is suggested to send X_n to the decompressor so that the latter can simply output X_n.

The last paragraph can actually be translated into a lossy compression scheme but I decided to omit the details as they can be found in past works, e.g., [**KU10**]. I claim without a proof that the excess of distortion of this coding scheme is bounded by the average of total variation norms of the noisy descendants. More symbolically,

$$D - \Delta \leqslant \frac{1}{N} \sum_{j_1^n \in \mathcal{J}} T\Big(\big(\cdots((W^{(j_1)})^{(j_2)})\cdots\big)^{(j_n)}\Big). \tag{3.3}$$

It remains to define $\mathcal{J}$. Motivated by inequality (3.3), we simply let $\mathcal{J}$ collect depth-n synthetic channels whose T is less than a threshold θ. Then, similar to Chapter 2, we have that $D - \Delta < \theta$ and $R = 1 - \mathsf{P}\{T(\mathsf{W}_n) < \theta\}$. To rephrase it, the trade-off between P_e and R for lossy compression is

$$\mathsf{P}\{T(\mathsf{W}_n) < D - \Delta\} \approx 1 - R.$$

Let $\{\mathsf{T}_n\}$ be the total variation process defined by $\mathsf{T}_n \coloneqq T(\mathsf{W}_n)$.

The last paragraph is not the only driving force for learning the total variation process $\{\mathsf{T}_n\}$; channel coding over asymmetric channels enjoys a similar reduction.

Recall that $Q(x)$, presumably non-uniform, is the capacity-achieving input distribution w.r.t. an asymmetric W. Pretend that $Q(x, \clubsuit)$ models a channel with a constant output $\clubsuit$. Then we can polarize Q and talk about its descendants. Details omitted, the block error probability of polar coding over W is [**HY13**]

$$P_\mathrm{e} \leqslant \sum_{j_1^n \in \mathcal{J}} Z\Big(\big(\cdots((W^{(j_1)})^{(j_2)})\cdots\big)^{(j_n)}\Big) + T\Big(\big(\cdots((Q^{(j_1)})^{(j_2)})\cdots\big)^{(j_n)}\Big). \tag{3.4}$$

We know how to bound the sum of Z's, so it remains to bound the sum of T's.

All in all, we now want to show

$$\mathsf{P}\big\{\mathsf{T}_n < e^{-2^{\pi n}}\big\} > H(W) - 2^{-\rho n} \tag{3.5}$$

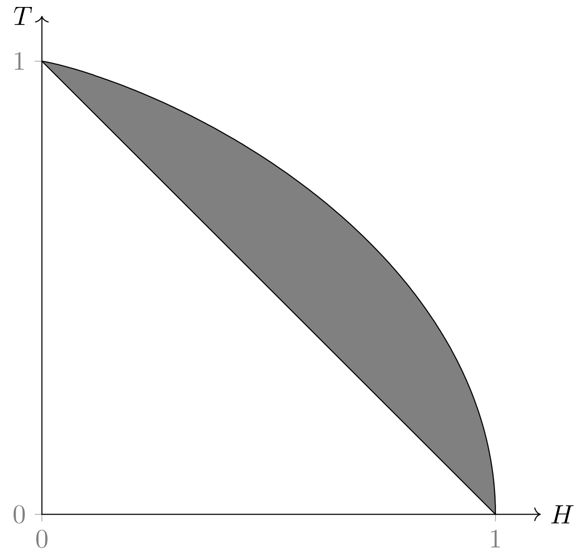

FIGURE 3.2. The possible region where $(T(W), Z(W))$ could lie in. The curve part is half of h_2 after rotating and rescaling.

for (π, ρ) lying in the same region as in Chapter 2. The following section argues that we can reuse the same the proof.

5. Stochastic Process Nonsense

In this section, I will show that since the total variation process $\{\mathsf{T}_n\}$ satisfies almost all properties satisfied by $\{\mathsf{Z}_n\}$, the majority of the proof of inequality (2.17) applies to inequality (3.5).

Let us start with an counterpart to Lemma 2.12.

Lemma 3.8 (Evolution of T)**.** [**Mur21**, Lemma 3] *The following hold for all SB-DMCs W:*

$$T(W^{(1)}) = T(W)^2, \tag{3.6}$$

$$T(W^{(2)}) \leqslant 2T(W) - T(W). \tag{3.7}$$

This implies a nice consequence—that there is another supermartingale that will dominate the behavior of W_n at the noisy end.

Lemma 3.9. *The process of total variation norms, $\{\mathsf{T}_n\}$, is a supermartingale.*

Note that we do not have a counterpart to inequality (2.4). This will affect how we deal with the en23 behavior of $\{\mathsf{T}_n\}$. The counterpart of Lemma 2.13 follows.

Lemma 3.10 (T vs H)**.** [**Mur21**, Lemma 4] *The following holds for all SBDMCs W:*

$$1 - T(W) \leqslant H(W) \leqslant h_2\Big(\frac{1 - T(W)}{2}\Big)$$

where h_2 is the binary entropy function.

See Figure 3.2 for a visualization.

The counterpart to Lemma 2.16 follows. Note that we skipped Theorem 2.15, but the following proof shows that we need not duplicate Theorem 2.15.

Lemma 3.11 (From eigen to en23)**.** *Fix $\varrho \coloneqq 1/4.714$. Assume Theorem 2.15. Then*

$$P\{T_n < e^{-n^{2/3}}\} > H(W) - 2^{-\varrho n + o(n)}.$$

Proof. By Lemmas 2.13 and 3.10, $T(W) \to 0$ iff $H(W) \to 1$ iff $Z(W) \to 1$ on the noisy end and $T(W) \to 1$ iff $H(W) \to 0$ iff $Z(W) \to 0$ on the reliable end. More strongly, each "iff" can be stated as two Hölder conditions in two directions. That is to say, there exist constants $c, d > 0$ such that $T(W) \leqslant c(1 - H(W))^d$ and $1 - H(W) \leqslant cT(W)^d$, as well as $1 - H(W) \leqslant c(1 - Z(W))^d$ and $1 - Z(W) \leqslant c(1 - H(W))^d$, as well as the other four inequalities on the reliable end.

This common-fate property implies that, if n is sufficiently large,

$$\begin{aligned} P\{e^{-n^{2/3}} \leqslant T_n \leqslant 1 - e^{-n^{2/3}}\} &< P\{e^{-n^{3/4}} \leqslant Z_n \leqslant 1 - e^{-n^{3/4}}\} \\ &\leqslant \frac{E[h(Z_n)]}{h(\exp(-n^{3/4}))} \leqslant \frac{h(Z_0)2^{-\varrho}}{\exp(-n^{3/4})} < 2^{-\varrho n - o(n)}. \end{aligned}$$

This is the counterpart to inequality (2.10).

It remains to show that $P\{T_n \to 0\} = H(W)$ and that the bad event $B_n \coloneqq \{T_n \to 0 \text{ but } T_n \geqslant \exp(-n^{2/3})\}$ is exponentially rare, i.e., $2^{-\varrho(n)+o(n)}$-rare. The former is again by the common-fate property $P\{T_n \to 0\} = P\{H_n \to 1\} = E[H_\infty] = E[H_0] = H(W)$. The latter is by that $\{T_n\}$, a supermartingale, cannot jump back and forth between the neighborhood of 0 and the neighborhood of 1, so $P(B_n)$ is bounded from above by $\sum_{m \geqslant n} 2^{-\varrho m + o(m)} < 2^{-\varrho n + o(n)}$.

We can finally conclude that

$$P\{T_n < e^{-n^{2/3}}\} > P\{T_n \to 0\} - P(B_n) > H(W) - 2^{-\varrho n + o(n)}.$$

This calls the end of the proof. ☐

The last lemma is the most technical one in this chapter. It uses the common-fate property to show that not only do $\{Z_n\}$, $\{H_n\}$, and $\{T_n\}$ control each other's limit, but they also control each other's pace of convergence. In particular, we can also show that $P\{H_n < \exp(-n^{2/3})\} > I(W) - 2^{-\varrho n + o(n)}$ and that $P\{1 - H_n < \exp(-n^{2/3})\} > H(W) - 2^{-\varrho n + o(n)}$ although it is not useful here.

The een13 behavior follows.

Lemma 3.12 (From en23 to een13)**.** *Given Lemma 3.11, that is, given*

$$P\{T_n < e^{-n^{2/3}}\} > H(W) - 2^{-\varrho n + o(n)},$$

we have

$$P\{T_n < \exp(-e^{n^{1/3}})\} > H(W) - 2^{-\varrho n + o(n)}.$$

Proof. The proof is just a copy of that of Lemma 2.17. All we did there was by that Z_n is squared or doubled with equal probability. Since T_n is squared or doubled with equal chance, the conclusion follows. ☐

The ultimate behavior follows.

Theorem 3.13 (From een13 to e2pin)**.** *Fix a pair* $(\pi, \rho) \in \mathcal{O}$*. Given the conclusion of Lemma 3.12, that is, given*

$$P\{T_n < \exp(-e^{n^{1/3}})\} > H(W) - 2^{-\varrho n + o(n)},$$

then

$$P\{T_n < e^{-2^{\pi n}}\} > H(W) - 2^{-\rho n + o(n)}.$$

Proof. The proof follows the same type of argument as in Theorem 2.18. All we need is based on the fact/axiom that the process is squared or doubled with equal chance. □

6. Chapter Wrap up

In this chapter, we first see two coding problems—source coding for lossy compression and noisy channel coding over asymmetric channels. I then argued that polar coding applies to those scenarios with an extension to asymmetric (test) channels. To deal with asymmetric channels, we prove a series of lemmas whose takeaway is that we only have to consider the SBDMC $\tilde{W}$ that lies in the same equivalence class. To characterize the performance of polar coding, it remains to understand the behavior of $\{T_n\}$. And then we move on to proving the behavior of $\{T_n\}$ using the same techniques we used to handle $\{Z_n\}$.

Now that we have proved $P\{Z_n < \exp(-2^{\pi n})\} > I(W) - 2^{-\rho n + o(n)}$ (from the previous chapter) and $P\{T_n < \exp(-2^{\pi n})\} > H(W) - 2^{-\rho n + o(n)}$ (in this chapter), plug them into inequalities (3.3) and (3.4). For lossy compression, my result implies that lossy compression via polar coding enjoys code rate $I(W) + 2^{-\rho n + o(n)}$ and distortion $\Delta + \exp(-2^{\pi n})$. Eliminate the $o(n)$ term with a topological argument.

Corollary 3.14 (good code for lossy compression)**.** *For any lossy compression problem whose test channel* W *is a BDMC, using polar coding yields excess of distortion* $D - \Delta < \exp(-N^{\pi})$ *and gap to capacity* $R - I(W) < N^{-\rho}$ *for any* $(\pi, \rho) \in \mathcal{O}$ *and big* N*.*

The asymmetric channel case is more involved. We actually have, and need, four inequalities

$$P\{Z(W_n) < \theta\} > 1 - H(W) + \gamma,$$
$$P\{T(W_n) < \theta\} > H(W) + \gamma,$$
$$P\{Z(Q_n) < \theta\} > 1 - H(Q) + \gamma,$$
$$P\{T(Q_n) < \theta\} > H(Q) + \gamma,$$

and basic facts

$$\{Z(W_n) < \theta\} \cap \{T(W_n) < \theta\} = 0,$$
$$\{Z(Q_n) < \theta\} \cap \{T(Q_n) < \theta\} = 0,$$
$$\{Z(Q_n) < \theta\} \cap \{T(W_n) < \theta\} = 0.$$

Here, $\{Q_n\}$ is the channel process grown from Q, and θ (threshold) and γ (gap) are the shorthands of the complicated functions. The basic facts are consequences of the common-fate property and the monotonicity property $H(W) = H(X \mid Y) \leqslant H(X) = H(X \mid \clubsuit) = H(Q)$.

Applying the inclusion–exclusion principle, we conclude that $P\{T(Q_n) < \theta$ and $Z(W_n) < \theta\} > H(Q) - H(W) - 6\gamma = I(W) - 6\gamma$. This implies the existence of codes with $P_{\mathrm{e}} < \exp(-2^{\pi n})$ and $R > I(W) - 2^{-\rho n}$.

Corollary 3.15 (Good code for BDMC)**.** *For any BDMC, polar coding yields block error probability* $P_{\mathrm{e}} < \exp(-N^{\pi})$ *and gap to capacity* $I(W) - R < N^{-\rho}$ *for any* $(\pi, \rho) \in \mathcal{O}$ *and big* N*.*

The next chapter contains yet another application of Theorems 2.18 and 3.13. In brief, the idea is to prune the channel process when W_n becomes too reliable or too noisy. And it is only after learning how reliable/noisy W_n is that we know where to prune.

7. A Side Note on Lossless Compression

Lossless compression with side information at the decoder side can be solved by polar coding as well but assumes a different format [**Ari10, CK10**]. In this scenario, the random variable to be compressed is denoted by X, and will be treated as the input of an abstract channel W; the side information accessible by the decoder is denoted by Y and treated as the output of W.

W is then polarized. For if a descendant of W is reliable, it means that its input is mostly determined by its output, and hence needs no extra action. If, otherwise, a descendant of W is noisy, its input is largely independent from its output, and we should record this input. The code rate is thus the density of the noisy descendants. There will be a block error if the decoder fails to recover the input of a descendant that was not recorded; the block error probability is thus the sum of the Z's of the reliable descendants.

This come down to the following corollary.

Corollary 3.16 (Good code for lossless compression)**.** *For any binary source* X *to be compressed losslessly and side information* Y*, polar coding yields block error probability* $P_{\mathrm{e}} < \exp(-N^{\pi})$ *and gap to entropy* $R - H(X \mid Y) < N^{-\rho}$*, for any* $(\pi, \rho) \in \mathcal{O}$ *and big* N*.*

CHAPTER 4

Pruning Channel Tree

Complexity of encoder and decoder influences, sometimes dominates, practicality of a code. Throughout the history of coding—Hamming, Reed–Muller, turbo, LDPC, polar, et seq—real world codes are always easy to implement in the first place regardless of achieving capacity or not (usually not). This is why, apart from N, P, and R, we should care about the complexities of codes.

Polar code, an outlier in said list, is the first code to achieve capacity in combination with a very low complexity. Considering that the decoder should at least read in all N symbols in a code block, which costs $O(N)$ resources, polar coding's complexity of $O(N \log N)$ is impressive, if not surprising.

In this chapter, a modification is made to polar coding. It aims to eliminate the inefficient components in the encoder/decoder to reduce the complexity even further. By doing so carefully, the new complexity is $O(N \log(\log N))$, the block error probability decays quasi-polynomially fast to 0, and the gap to capacity decays as fast as before. For comparison, existing works like [**AYK11, EKMF$^+$17, MHCG20**] did not break the $O(N \log N)$ barrier, and some latest work [**HMF$^+$20**] achieves $N \log(\log N)$ *latency* only in the fully-parallel mode.

Being called pruning, this technique is inspired by a trinitarian correspondence among the encoder/decoder, the channel tree, and the channel process. The argument start from the channel process side: Since $H_n \to H_\infty \in \{0, 1\}$, the increments $H_{n+1} - H_n$ converge to 0. By the correspondence, that means that on the channel tree side, branches that are deep enough are purposeless—applying channel transformation barely changes anything. Now turn our focus to the encoder/decoder side; the said observation implies that some components in the encoder/decoder are consuming resources without helping the code become better, and we should have removed them.

The goal of the current chapter is to make precise the last paragraph.

1. Encoder/Decoder vs Tree vs Process

The design of the encoder and decoder of polar coding is best described by figures. There is a device called EU (encoding unit) and another device called DU (decoding unit); their actual implementations are not of interest here. But the devices are such that, when you wrap two copies of W with one EU–DU pair, like Figure 4.1 does, the top pair of pins (A and B) behaves like $W^{(1)}$, and the bottom pair of pins (C and D) behaves like $W^{(2)}$. We say that Figure 4.1 corresponds to a tree with three vertices: W the root and $W^{(1)}$ and $W^{(2)}$ its children.

To construct the grandchildren of W, wrap more EU–DU pairs around two copies of Figure 4.1. For instance, in Figure 4.2, the two copies of pin A are connected to another EU, the two copies of pin B to another DU. Recall that pin A–pin B behaves like the input–output of $W^{(1)}$; so wrapping one more layer of

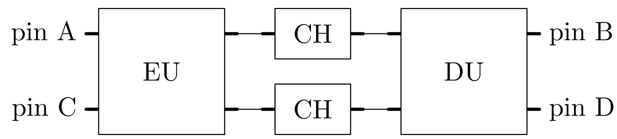

FIGURE 4.1. The design of encoder and decoder—level 1.

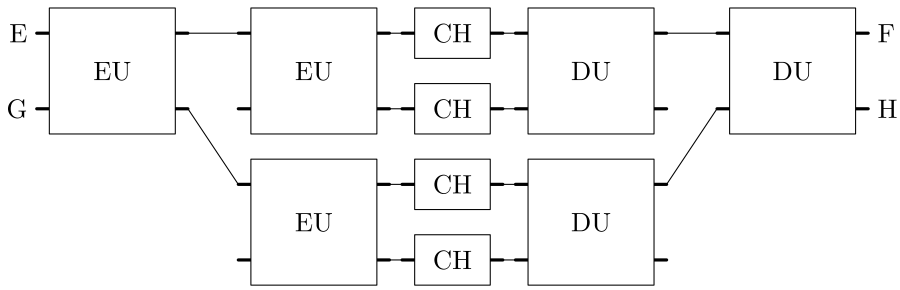

FIGURE 4.2. The design of encoder and decoder—transforming $W^{(1)}$ further at level 2.

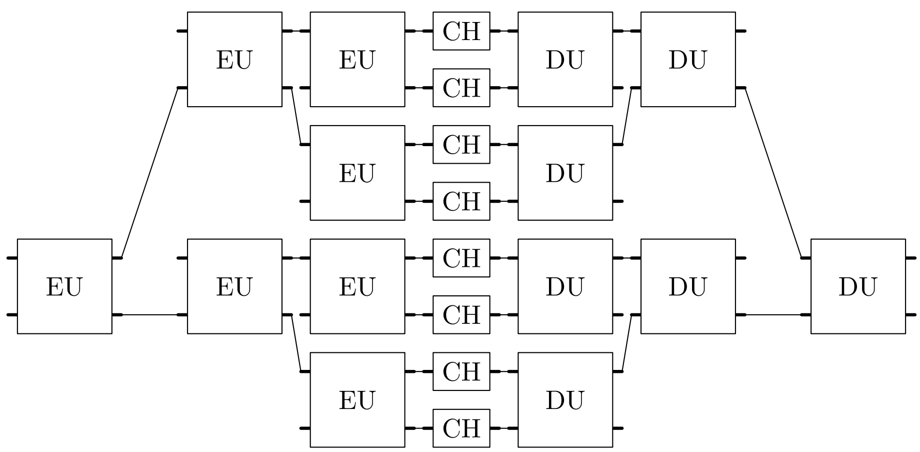

FIGURE 4.3. The design of encoder and decoder—transforming $(W^{(1)})^{(2)}$ further at level 3.

EU–DU transforms it further into $(W^{(1)})^{(1)}$ (from pin E to pin F) and $(W^{(1)})^{(2)}$ (from pin G to pin H). Note that the two copies of pin C and pin D are naked (not connecting to anything); this represents the fact that there are two copies of $W^{(2)}$ that are not (yet) transformed into $(W^{(2)})^{(1)}$ and $(W^{(2)})^{(2)}$. Now Figure 4.2 corresponds to a channel tree with five vertices: W the root, $W^{(1)}$ and $W^{(2)}$ its children, plus $(W^{(1)})^{(1)}$ and $(W^{(1)})^{(2)}$ the children of the elder sibling.

Duplicate Figure 4.2. This time, wrap around the two copies of pin G and H as shown in Figure 4.3. Then we are effectively transforming $(W^{(1)})^{(2)}$ into $((W^{(1)})^{(2)})^{(1)}$ and $((W^{(1)})^{(2)})^{(2)}$. Now Figure 4.3 corresponds to a channel tree with seven vertices: W, $W^{(1)}$, $W^{(2)}$, $(W^{(1)})^{(1)}$, and $(W^{(1)})^{(2)}$ and its children.

Lesson: Each pair of pins corresponds to a (synthetic) channel; one channel may have multiple copies that each corresponds to a pair. For any synthetic channel, we wrap another layer of EU–DU pair around the corresponding pins if we want to transform it further. We leave the pins corresponding to a channel naked if we do not want to transform it anymore.

Rephrase in terms of channel process: The channel process $\{\mathsf{W}_n\}$ is defined prior to whether we want to transform channels or not. But we can choose "not to look at it". More rigorously, we look at the endless sequence $\mathsf{J}_1, \mathsf{J}_2, \dots$ and determine the least n such that $(\mathsf{J}_1, \mathsf{J}_2, \dots, \mathsf{J}_n)$ points to a channel we do not want to transform anymore. Let s be that n in this case. That is, s is a random variable depending on $\{\mathsf{J}_n\}$; moreover, whether or not $\mathsf{s} = n$ is determined by the first n terms of $\{\mathsf{J}_n\}$. This makes s a stopping time adapted to $\{\mathsf{J}_n\}$. Consequently, $\{\mathsf{W}_{n\wedge\mathsf{s}}\}$ is a stopped process that evolves like $\{\mathsf{W}_n\}$ at the beginning but halts at W_s, a channel we are satisfied with.

It remains to do the following three things: (a) Show how s is related to the complexity, (b) show how W_s is related to the code metrics P_e and R, and (c) define a good stopping time s to optimize (a) and (b).

2. Stopping Time vs Complexity

Fix an SBDMC W. Fix an n; we are to construct a low-complexity code of block length $N = 2^n$. Pretend that the complexity is the number of EU and DU devices used. Doing so is backed up by the fact that, in reality, EU and DU cost bounded amount of arithmetic and memory.

Let s be any stopping time adapted to $\{\mathsf{J}_n\}$, and assume $\mathsf{s} \leqslant n$. Having $\mathsf{s} \leqslant n$ is to make sure that we do not transform any depth-n channel further. We have the following lemma concerning s and the complexity.

Lemma 4.1 (Complexity in terms of s)**.** *Use s to generate a code. Then the encoding and decoding complexity is $O(N\mathsf{E}[\mathsf{s}])$ per code block, or $O(\mathsf{E}[\mathsf{s}])$ per channel usage.*

Proof. Let us begin with $N = 2^n$ copies of the channel W. If $\mathsf{s} > 0$, then there are 2^{n-1} EU–DU pairs that wrap around W's to synthesize 2^{n-1} copies of $W^{(1)}$ and $W^{(2)}$.

Next, let us consider the $\mathsf{J}_1 = 1$ case. If $\mathsf{s} > 1$ in this case, it means that we want to transform $W^{(1)}$. We need 2^{n-2} EU–DU pairs to wrap around the 2^{n-1} copies of $W^{(1)}$. Similarly, if $\mathsf{s} > 1$ in the $\mathsf{J}_1 = 2$ case, then we are transforming 2^{n-1} copies of $W^{(2)}$ using another 2^{n-2} EU–DU pairs.

Next, consider the depth-2 branching of $(\mathsf{J}_1, \mathsf{J}_2)$. For each of the four possible realizations of the pair $(\mathsf{J}_1, \mathsf{J}_2) = (j_1, j_2)$, if $\mathsf{s} > 2$, then we are to transform the 2^{n-2} copies of $(W^{(j_1)})^{(j_2)}$ with 2^{n-3} EU–DU pairs.

Finally, consider the general case at arbitrary depth: If $\mathsf{s} > m$ for a certain tuple $(\mathsf{J}_1, \mathsf{J}_2, \dots, \mathsf{J}_m) = (j_1, j_2, \dots, j_m)$, then we will transform 2^{n-m} copies of $\mathsf{W}_m = \big(\cdots((W^{(j_1)})^{(j_2)})\cdots\big)^{(j_m)}$ further with 2^{n-m-1} EU–DU pairs. The total number of EU–DU pairs is thus

$$\sum_{m=0}^{n}\sum_{j_1^m} 2^{n-m-1}\cdot \mathsf{I}\{\mathsf{J}_1^m = j_1^m \text{ and } \mathsf{s} > m\} = \sum_{m=0}^{n} 2^{n-1}\mathsf{P}\{\mathsf{s} > m\} = \frac{N}{2}\mathsf{E}[\mathsf{s}].$$

Here I is the indicator of an event.

Recall that I claimed without a proof that EU and DU devices cost constant resources. Thus, the complexity of a code is proportional to the number of such devices used. So a code generated by s has complexity $O(N\mathsf{E}[\mathsf{s}])$ per code block or $O(\mathsf{E}[\mathsf{s}])$ per channel usage. The end. $\square$

The last lemma connects a stopping time s to the cost of its code. Intuitively speaking, it attributes to the fact that preparing a synthetic channel at depth m costs m layers of EU–DU. Hence preparing W_s costs s layers, and preparing all instances of W_s costs $N\mathsf{E}[\mathsf{s}]$.

In the next section, we will see how W_s connects to the code performance in a very similar way, i.e., $P_\mathrm{e} \leqslant N\mathsf{E}[\mathsf{Z}_\mathsf{s} \cdot \mathsf{I}\{\mathsf{Z}_\mathsf{s} < \theta\}]$ and $R = \mathsf{P}\{\mathsf{Z}_\mathsf{s} < \theta\}$, due to very similar reasons.

3. Stopped Process vs Performance

Thanks to the stopping time s, the channel tree is pruned, or "harvested", before it reaches depth n. This implies that the leafs W_s are not as polarized as before. We therefore are liable to rebound the block error probability and code rate.

Let $\mathcal{J}$ be some set of indices $(\mathsf{J}_1, \mathsf{J}_2, \ldots, \mathsf{J}_\mathsf{s})$ that point to the synthetic channels we are to use to send plain messages. Lemmas follow.

Lemma 4.2 (R in terms of $\mathcal{J}$)**.** *The code rate*

$$R = \mathsf{P}\{(\mathsf{J}_1, \mathsf{J}_2, \ldots, \mathsf{J}_\mathsf{s}) \in \mathcal{J}\}$$

is the probability that the prefix J_1^s is selected in $\mathcal{J}$.

Proof. I claim without a proof that the code rate is the density of the naked pins that correspond to channels in $\mathcal{J}$.

Assuming that, we see that each W_s corresponds to $2^{n-\mathsf{s}}$ pairs of naked pins. That is to say, each W_s assumes $2^{n-\mathsf{s}}$ copies in the EU–DU circuit. Since there are always 2^n pairs of naked pins (adding ED/DU does not alter the number of naked pins), a pair of naked pins possesses probability measure 2^{-n}. Thus each W_s possesses probability measure $2^{-\mathsf{s}}$, the same probability measure possessed by $(\mathsf{J}_1, \mathsf{J}_2, \ldots, \mathsf{J}_\mathsf{s})$. This finishes the proof. $\square$

Lemma 4.3 (P_e in terms of $\mathcal{J}$)**.** *The block error probability*

$$P_\mathrm{e} \leqslant N\mathsf{E}[\mathsf{Z}_\mathsf{s} \cdot \mathsf{I}\{(\mathsf{J}_1, \mathsf{J}_2, \ldots, \mathsf{J}_\mathsf{s}) \in \mathcal{J}\}]$$

is bounded by the sum of Z_s that are selected in $\mathcal{J}$, weighted by multiplicity.

Proof. I claim without a proof that the block error probability is bounded from above by the sum of Z's of the synthetic channels selected in $\mathcal{J}$, multiplicity included.

Assuming that, and knowing that each W_s assumes $2^{n-\mathsf{s}}$ copies and possesses probability measure $2^{-\mathsf{s}}$, we infer that each W_s contributes $N\mathsf{P}\{\mathsf{J}_1^\mathsf{s} = j_1^\mathsf{s}\} \cdot Z(\mathsf{W}_\mathsf{s})$ to the upper bound if it is selected, otherwise it would have contributed 0. Each W_s contributing $N\mathsf{P}\{\mathsf{J}_1^\mathsf{s} = j_1^\mathsf{s} \in \mathcal{J}\} \cdot Z(\mathsf{W}_\mathsf{s})$, their sum is clearly $N\mathsf{E}[\mathsf{Z}_\mathsf{s} \cdot \mathsf{I}\{\mathsf{J}_1^\mathsf{s} \in \mathcal{J}\}]$. This is the upper bound we want to prove. $\square$

It is time to declare an s and compute the induced R, P_e, and complexity. The basic strategy, like what we did in Chapter 2, is to set a threshold θ and hope that W_s is "at least θ-good", or $\mathsf{Z}_\mathsf{s} \leqslant \theta$ to be rigorous. In contrast to Chapter 2, we can now control s, so it is actually more efficient if we incorporate θ into the declaration of s.

4. Actual Code Construction

Let θ be 4^{-n}. Define the stopping time s as below

$$\mathsf{s} \coloneqq n \wedge \min\{m : \mathsf{T}_m < \theta \text{ or } \mathsf{Z}_m < \theta\}. \tag{4.1}$$

That is, we look for the first m such that either T_m or Z_m are small enough; but if we cannot find such m before reaching depth n, let s default to n. Let $\mathcal{J}$ be the indices with $\mathsf{Z}_{\mathsf{s}} < \theta$. That is, if s is set to be m because $\mathsf{Z}_m < \theta$, then the corresponding $\mathsf{J}_1^{\mathsf{s}}$ are included in $\mathcal{J}$. If s is set to be m because $\mathsf{T}_m < \theta$ or because we reach $m = n$, then the corresponding $\mathsf{J}_1^{\mathsf{s}}$ are not included in $\mathcal{J}$.

The block error probability incurred by this choice of s and $\mathcal{J}$ is bounded by the weighted sum of Z's that are all smaller than $\theta \coloneqq 4^{-n}$. Hence $P_{\mathrm{e}} < N\theta = 1/N$ by Lemma 4.3.

The complexity and code rate are more complicated. See the next two theorems.

Theorem 4.4. *Given that s is defined as in the first paragraph of this section, an upper bound on the complexity is*

$$O(N\mathsf{E}[\mathsf{s}]) = O(N\log(\log N)).$$

Proof. By how s is defined, we have

$$\mathsf{E}[\mathsf{s}] = \sum_{m=0}^{n-1} \mathsf{P}\{\mathsf{s} > m\} = \sum_{m=0}^{n-1} \mathsf{P}\{\mathsf{T}_m \geqslant 4^{-n} \text{ and } \mathsf{Z}_m \geqslant 4^{-n}\}. \tag{4.2}$$

It remains to understand what is going on in $\{\mathsf{T}_m \geqslant 4^{-n} \text{ and } \mathsf{Z}_m \geqslant 4^{-n}\}$. For that, recall the last two chapters:

$$\mathsf{P}\{\mathsf{Z}_m < \exp(-2^{m/40})\} > I(W) + 2^{-m/5+o(m)},$$
$$\mathsf{P}\{\mathsf{T}_m < \exp(-2^{m/40})\} > H(W) + 2^{-m/5+o(m)}.$$

Here I choose $(\pi, \rho) = (1/40, 1/5) \in \mathcal{O}$. In words, I showed that either Z_m or T_m, exclusively, becomes small with high probability. Hence it is unlikely that $\mathsf{T}_m \geqslant 4^{-n}$ and $\mathsf{Z}_m \geqslant 4^{-n}$ at once unless m is small, in which case the proved threshold $\exp(-2^{m/40})$ is less than the demanded 4^{-n}.

We now classify m into two classes: Those such that $\exp(-2^{m/40}) \geqslant 4^{-n}$ are called small m. Those such that $\exp(-2^{m/40}) < 4^{-n}$ are called large m. For small m, we have nothing but $\mathsf{P}\{\text{both } \mathsf{T}_m, \mathsf{Z}_m \geqslant 4^{-n}\} \leqslant 1$. For large m,

$$\mathsf{P}\{\text{both } \mathsf{T}_m, \mathsf{Z}_m \geqslant 4^{-n}\} \leqslant \mathsf{P}\{\text{both } \mathsf{T}_m, \mathsf{Z}_m \geqslant \exp(-2^{m/40})\} < 2^{-m/5+o(m)}.$$

Continue bounding inequality (4.2):

$$\begin{aligned}
&\sum_{m=0}^{n-1} \mathsf{P}\{\text{both } \mathsf{T}_m, \mathsf{Z}_m \geqslant 4^{-n}\} \\
&\quad = \sum_{\text{small } m} \mathsf{P}\{\text{both } \mathsf{T}_m, \mathsf{Z}_m \geqslant 4^{-n}\} + \sum_{\text{large } m} \mathsf{P}\{\text{both } \mathsf{T}_m, \mathsf{Z}_m \geqslant 4^{-n}\} \\
&\quad = \sum_{\text{small } m} 1 + \sum_{\text{large } m} 2^{-m/5+o(m)} = \#\{\text{small } m\text{'s}\} + O(1).
\end{aligned}$$

The number of small m's is the root of the euqation $\exp(-2^{m/40}) = 4^{-n}$. The root is $m = O(\log n)$. Hence the complexity $\mathsf{E}[\mathsf{s}] < O(\log n) = O(\log(\log N))$, as desired. $\square$

Theorem 4.5. *Given that* s *and* $\mathcal{J}$ *are defined as in the first paragraph of this section, a lower bound on the code rate* R *is*

$$R \geqslant I(W) - 2^{-n/5+o(n)}.$$

PROOF. Just to clarify the whole picture, let me claim the following trichotomy:

- The frequency that s is set to m due to $\mathsf{T}_m < \theta$ is $H(W) - 2^{-n/5+o(n)}$.
- The frequency that s is set to m due to reaching n is $2^{-n/5+o(n)}$.
- The frequency that s is set to m due to $\mathsf{Z}_m < \theta$ is $I(W) - 2^{-n/5+o(n)}$.

$\mathcal{J}$ collects those $\mathsf{J}_1^{\mathsf{s}}$ when $\mathsf{Z}_m < \theta$ is the case. So the theorem statement is implied by the last bullet point. I will show the last bullet point; the proof thereof applies to the first bullet point, so I am effectively showing all three bullet points at once.

In order to show the last bullet point, it suffices to understand the bad event $B \coloneqq \{\mathsf{Z}_{\mathsf{s}} \geqslant \theta \text{ but } \mathsf{Z}_m \to 0\}$. Once we know how to bound $P(B)$, we will conclude that, with Lemma 4.2, the code rate is $R = P\{\mathsf{Z}_{\mathsf{s}} < \theta\} \geqslant P\{\mathsf{Z}_m \to 0\} - P(B)$, which will be $I(W) - 2^{-n/5+o(n)}$. To bound $P(B)$, two cases will be discussed—$\mathsf{Z}_{\mathsf{s}} \geqslant \theta$ because $\mathsf{T}_m < \theta$ happens first, and $\mathsf{Z}_{\mathsf{s}} \geqslant \theta$ because we reach $m = n$.

Here goes the rigorous bound on $P(B)$: If $\mathsf{Z}_{\mathsf{s}} \geqslant \theta$, then s is not set to the current value because $\mathsf{Z}_m < \theta$; it must be the case that the other criterion $\mathsf{T}_{\mathsf{s}} < \theta$ holds, or that we reach $\mathsf{s} = n$. For the former case, $P\{\mathsf{T}_{\mathsf{s}} < \theta \text{ but } \mathsf{Z}_m \to 0\} = P\{\mathsf{T}_{\mathsf{s}} < \theta \text{ but } \mathsf{T}_m \to 1\} \leqslant \theta = 2^{-n}$ by the common-fate property and $\{\mathsf{T}_n\}$ being a martingale. For the latter case, $P\{\mathsf{s} = n \text{ and } \mathsf{Z}_m \to 0\} < P\{\mathsf{s} = n\} = P\{\text{both } \mathsf{T}_m, \mathsf{Z}_m \geqslant \theta \text{ for all } m < n\} < P\{\text{both } \mathsf{T}_{n-1}, \mathsf{Z}_{n-1} \geqslant \theta\} = 2^{-(n-1)/5+o(n-1)}$. Sum the upper bounds of the two cases; it is $2^{-n/5+o(n)}$.

Now that we get $P(B) < 2^{-n/5+o(n)}$, deduce the code rate $R = P\{\mathsf{Z}_{\mathsf{s}} < \theta\} = P\{\mathsf{Z}_m \to 0\} - P(B) = I(W) - 2^{-n/5+o(n)}$; that is what we want to prove. □

Recap: So far we had seen that, when a code is defined by my choice of s and $\mathcal{J}$, (a) the block error probability is $1/N$, (b) the complexity is $O(N \log(\log N))$ per code block or $O(\log(\log N))$ per channel usage, and (c) the code rate is $I(W) - 2^{-n/5+o(n)}$. The minor term $o(n)$ within R can be eliminated by a topological argument. We summarize the chapter with the following corollary.

Corollary 4.6 (Log-log code for SBDMC)**.** *Over any SBDMC, there exist capacity-achieving codes with encoding and decoding complexity* $O(N \log(\log N))$ *per code block or* $O(\log(\log N))$ *per channel usage. Moreover, the said codes have gaps to capacity* $I(W) - R < N^{-1/5}$.

In one sentence, *log-logarithmic complexity achieves capacity.*

5. A Note on Better Pace

The only place we used the pair $(1/40, 1/5) \in \mathcal{O}$ explicitly is when we were solving $\exp(-2^{m/40}) = 4^{-n}$ for m and obtain that $m = O(\log n)$. We could, instead, choose another pair such as $(1/150, 1/4.8) \in \mathcal{O}$. Then, we still get to keep $m = O(\log n)$ but the gap to capacity is $I(W) - R < N^{-1/4.8}$. By a topological argument, any $\rho < 1/4.714$ works. This is why I claimed that the pruned polar code has the same gap to capacity as the original version.

Concerning the equation $\exp(-2^{m/40}) = 4^{-n}$, we can also replace 4^{-n} with $\exp(-n^{\tau})$ for arbitrarily large τ. The root is then $m = O(\tau \log n)$, which is $O(\log n)$ if τ is fixed. That means the block error probability can be as low as $\exp(-n^{\tau})$

while keeping the log-logarithmic complexity. This asymptote lies below $1/\operatorname{poly}(N)$ and belongs to $\exp(-1\operatorname{poly}(\log N))$. This is why I claimed that the pruned polar code has quasi-polynomial block error probability.

6. A Note on Asymmetric Case

For if the concerned channel W is an asymmetric BDMC, we will need to apply a similar argument to $\{\mathsf{Q}_n\}$, the channel process grown from the non-uniform input distribution Q treated as a channel with constant output.

In the asymmetric case, the stopping time will be defined as

$$\mathsf{s} \coloneqq n \wedge \min\{m : \min(Z(\mathsf{W}_m), T(\mathsf{W}_m)) < 4^{-n} \text{ and } \min(Z(\mathsf{Q}_m), T(\mathsf{Q}_m)) < 4^{-n}\}.$$

To put in another way, we are looking for the minimal m that satisfies any of the following:

- Both $Z(\mathsf{W}_m), Z(\mathsf{Q}_m)$ are small.
- Both $T(\mathsf{W}_m), T(\mathsf{Q}_m)$ are small.
- Or, both $Z(\mathsf{W}_m), T(\mathsf{Q}_m)$ are small.
- (Both $T(\mathsf{W}_m), Z(\mathsf{Q}_m)$ being small is not possible.)

And if no m meets the requirement before depth n, the stopping time defaults to n. Furthermore, $\mathcal{J}$ will be the indices $\mathsf{J}_1^{\mathsf{s}}$ with small $Z(\mathsf{W}_m)$ and small $T(\mathsf{Q}_{\mathsf{s}})$: that is, the indices where s is set to m due to the third bullet point.

The block error probability is upper bounded similarly to inequality (3.4):

$$P_{\mathrm{e}} \leqslant N\mathbb{E}[Z(W_{\mathsf{s}})\mathsf{I}(\mathcal{J})] + N\mathbb{E}[T(\mathsf{Q}_{\mathsf{s}})\mathsf{I}(\mathcal{J})] \leqslant 2N \cdot 4^{-n} = 2^{-n+1}.$$

The complexity is upper bounded similarly to the symmetric case

$$\begin{aligned}
\mathsf{E}[\mathsf{s}] &= \sum_{m=0}^{n-1} \mathsf{P}\{\mathsf{s} > m\} \\
&= \sum_{m=0}^{n-1} \mathsf{P}\{\max(Z(\mathsf{W}_m), Z(\mathsf{W}_m)) \geqslant 4^{-n} \text{ or } \max(Z(\mathsf{Q}_m), Z(\mathsf{Q}_m)) \geqslant 4^{-n}\} \\
&\leqslant \sum_{m=0}^{n-1} \mathsf{P}\{\max(Z(\mathsf{W}_m), Z(\mathsf{W}_m)) \geqslant 4^{-n}\} + \mathsf{P}\{\max(Z(\mathsf{Q}_m), Z(\mathsf{Q}_m)) \geqslant 4^{-n}\} \\
&= O(\log n) + O(1) + O(\log n) + O(1) = O(\log n).
\end{aligned}$$

The code rate is lower bounded similarly to the symmetric case

$$\begin{aligned}
R = \mathsf{P}\{\mathsf{J}_1^{\mathsf{s}} \in \mathcal{J}\} &= \mathsf{P}\{\text{both } Z(\mathsf{W}_{\mathsf{s}}), T(\mathsf{Q}_{\mathsf{s}}) < 4^{-n}\} \\
&\geqslant \mathsf{P}\{Z(\mathsf{W}_m), T(\mathsf{Q}_m) \to 0\} - \mathsf{P}\{Z(\mathsf{W}_m) \to 0 \text{ but } Z(\mathsf{W}_{\mathsf{s}}) \geqslant 4^{-n}\} \\
&\qquad - \mathsf{P}\{T(\mathsf{Q}_m) \to 0 \text{ but } T(\mathsf{Q}_{\mathsf{s}}) \geqslant 4^{-n}\} \\
&\geqslant \mathsf{P}\{Z(\mathsf{W}_m), T(\mathsf{Q}_m) \to 0\} - \mathsf{P}\{Z(\mathsf{W}_m) \to 0 \text{ but } T(\mathsf{W}_{\mathsf{s}}) \leqslant 4^{-n}\} \\
&\qquad - \mathsf{P}\{T(\mathsf{Q}_m) \to 0 \text{ but } Z(\mathsf{Q}_{\mathsf{s}}) \leqslant 4^{-n}\} - \mathsf{P}\{\mathsf{s} = n\} \\
&\geqslant I(W) - 4^{-n} - 4^{-n} - 2^{-\rho n + o(n)} = I(W) - 2^{-\rho n + o(n)}.
\end{aligned}$$

Hence the following corollary.

Corollary 4.7 (Log-log code for BDMC)**.** *Over any BDMC, there exist capacity-achieving codes with encoding and decoding complexity* $O(N\log(\log N))$ *per code block or* $O(\log(\log N))$ *per channel usage. Moreover, the said codes have gaps to capacity* $I(W) - R < N^{-1/5}$*.*

The same statement can be made to lossless and lossy compression.

Theorem 4.8 (Log-log code for lossless compression). *If X is a binary source and Y is the side information, then lossless compression can be done with encoding and decoding complexity $O(N\log(\log N))$ per code block or $O(\log(\log N))$ per source observation, block error probability $P_{\mathrm{e}} < 1/N$, and gap to entropy $H(X \mid Y) - R < N^{-1/5}$.*

Corollary 4.9 (Log-log code for lossy compression). *If $Y \in \mathcal{Y}$ is a random source and $\mathrm{dist}\colon \mathbb{F}_2 \times \mathcal{Y} \to [0,1]$ is the distortion function, then lossy compression can be performed with encoding and decoding complexity $O(N\log(\log N))$ per code block or $O(\log(\log N))$ per source observation. excess of distortion $D - \Delta < 1/N^2$, and gap to capacity $R - I(W) < N^{-1/5}$.*

The discussion in the previous section applies. That is to say, the gap to capacity can be made as low as $N^{-1/4.8}$, the block error probability (or the excess of distortion) as low as $\exp(-n^\tau)$, while retaining the same log-logarithmic complexity.

7. Prospective Pruning

The pruning technique generalizes to arbitrary input (non-binary) and arbitrary matrix (no longer $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$). The next two chapters establish the theory of $\{\mathsf{W}_n\}$ for the general scenario. After that we can prune, yielding W_{s}, etc.

CHAPTER 5

General Alphabet and Kernel

Binary channel models, be it BEC, BSC, BDMC, or BI-AWGN, are favored in theory for their simplicity and for real world communications. But the real world is not always binary; we might want to compress losslessly a non-binary source (yes, no, or unanswered); or we might want to approximate colors with a 256-color palette. This is one of the two generalizations I want to make in this chapter—to enable all input alphabets of finite sizes [**ŞTA09**].

Enabling more input alphabets creates new challenges. The root of all challenges is that, to synthesize $W^{(1)}$ and $W^{(2)}$, we need to talk about the random variables $U_1^2 \coloneqq X_1^2 G^{-1}$ (c.f. formula (3.2)). This does not a priori make sense as the addition structure on the input alphabet $\mathcal{X}$ does not uniquely exist. Plus, even if we equip $\mathcal{X}$ with some group structure (abelian or not), there are cases where the descendants W_n are not polarized. This chapter addresses this challenge—how to equip $\mathcal{X}$ with a proper algebraic structure to facilitate polarization, even if it means adding dummy symbols into $\mathcal{X}$.

Specking of the root of the definition $U_1^2 \coloneqq X_1^2 G^{-1}$, it is not hard to imagine that, if we substitute G with a general $\ell \times \ell$ matrix, the overall performance of polar coding may differ [**KSU10**]. To tell if it becomes better, worse, or not working at all, we need a machinery that predicts the performance of each matrix. This is the second generalization I want to make in this chapter—to judge all matrices as the polarizing kernel.

The judgement of a matrix will be very similar to those in Chapters 2 and 3. We will need to prove (or assume) an eigen behavior. From that we can derive the en23 behavior (named after $\exp(-n^{2/3})$), the een13 behavior (named after $\exp\bigl(-e^{n^{1/3}}\bigr)$), and finally the elpin behavior (named after $\exp(-\ell^{\pi n})$). Along the way, I will explain how these behaviors relate to LLN, LDP, CLT, and MDP in probability theory.

This is a long chapter. Here goes the actual plan for this chapter.

1. Chapter Organization

I will define the most general channel (section 2). Following that is a reduction of input alphabet to prime size or prime-power size (section 3). Even more definitions are then made, including channel parameters P_{e}, Z, Z_{mxd}, T, S, and $S_{\max}$ (section 4) and how to use an invertible $\ell \times \ell$ matrix to synthesize $W^{(1)}, W^{(2)}, \ldots, W^{(\ell)}$ (section 5). Inserted here is a briff review of LLN, LDP, CLT, and MDP in probability theory and how they relate to coding theory (section 6). I will then clarify that only a subset of matrices are qualified to polarize channels (section 7). After that is a compactness argument that shows $\varrho > 0$ for those qualified matrices (section 9). Finally, the region of (π, ρ) where $\mathsf{P}\{\mathsf{Z}_n < \exp(-\ell^{\pi n})\} > 1 - H(W) + \ell^{-\rho n}$ will be pictured in two steps (sections 10 and 11).

2. General Problem Setup

The following channel is what was considered by Shannon in the eternal paper, followed by other big scholars who generalized Shannon. It is of historical interest to prove theorems over the same class of channels.

Definition 5.1. A *discrete memoryless channel* (DMC) is a Markov chain $W\colon \mathcal{X} \to \mathcal{Y}$ such that

- the input alphabet $\mathcal{X}$ is a finite set,
- the output alphabet $\mathcal{Y}$ is a finite set, and
- $W(y \mid x)$ is the array of transition probabilities satisfying $\sum_{y\in\mathcal{Y}} W(y \mid x) = 1$ for all $x \in \mathcal{X}$.

Denote by $Q(x)$ an input distribution, by $W(x, y)$ the joint distribution, by $W(x \mid y)$ the posterior distribution, and by $W(y)$ the output distribution.

Definition 5.2. The *conditional entropy* of W is

$$H(W) := -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} W(x, y) \log_{|\mathcal{X}|} W(x \mid y),$$

which is the amount of noise/equivocation/ambiguity/fuzziness caused by W.

Definition 5.3. The *mutual information* of W is

$$I(W) := H(Q) - H(W) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} W(x, y) \log_{|\mathcal{X}|} \frac{W(x \mid y)}{Q(x)},$$

Fix a Q that maximizes $I(W)$. Call this Q a *capacity-achieving input distribution.* Call this maximal $I(W)$ the *channel capacity* of W.

Let $\mathcal{U}$ be a finite set called the user alphabet. Messages to be transmitted over W will be pre-encoded into $\mathcal{U}$ and distributed uniformly. The overall goal is to construct, for some large N, an encoder $\mathcal{E}\colon \mathcal{U}^{RN} \to \mathcal{X}^N$ and a decoder $\mathcal{D}\colon \mathcal{Y}^N \to \mathcal{U}^{RN}$ such that the composition

$$U_1^{RN} \xmapsto{\mathcal{E}} X_1^N \xmapsto{W^N} Y_1^N \xmapsto{\mathcal{D}} \hat{U}_1^{RN}$$

is the identity map as frequently as possible, and R as close to the channel capacity $I(W)$ as possible.

Definition 5.4. Call N the block length. Call R the code rate. Denote by P_{e}, called the block error probability, the probability that $\hat{U}_1^{RN} \neq U_1^{RN}$.

In the next section, I will argue that we can reduce the problem of noisy-channel coding over arbitrary DMCs to $\mathcal{U}$ and $\mathcal{X}$ having finite field structure.

3. Pay Asymmetry to Buy Non-Field

Reducing the size of user alphabet is easier and will be handled before the reduction of input alphabet: Consider integer factorization $|\mathcal{U}| = p_1^{k_1} p_2^{k_2} \cdots p_\omega^{k_\omega}$. We can then choose a bijection

$$\mathcal{U} \cong \mathbb{F}_{p_1^{k_1}} \times \mathbb{F}_{p_2^{k_2}} \times \cdots \times \mathbb{F}_{p_\omega^{k_\omega}}.$$

To transmit a symbol $u \in \mathcal{U}$, it suffices to transmit its projections onto each of the constituent fields $\mathbb{F}_{p_1^{k_1}}, \ldots, \mathbb{F}_{p_\omega^{k_\omega}}$. It remains to show, for each finite filed $\mathbb{F}_{p^k}$, how to design encoder/decoder for the user alphabet $\mathcal{U} = \mathbb{F}_{p^k}$.

Moreover, each finite field admits a vector space structure

$$\mathbb{F}_{p^k} \cong \mathbb{F}_p^k.$$

Therefore, when necessary, each $u \in \mathbb{F}_{p^k}$ is considered as a vector in $\mathbb{F}_p^k$, with each of its coordinates treated as a standalone symbol in $\mathbb{F}_p$. That is to say, when necessary, we can even assume $\mathcal{U} = \mathbb{F}_p$ is of prime order.

Reducing the size of user alphabet takes effort and will be handled in the sequel: Recall that $|\mathcal{U}|$ is either a prime or a prime power. Let q be any power of $|\mathcal{U}|$ greater than or equal to $|\mathcal{X}|$. Consider an embedding $\mathcal{X} \subseteq \mathbb{F}_q$; call an $x \in \mathcal{X}$ an actual symbol; call a $v \in \mathbb{F}_q \backslash \mathcal{X}$ a virtual symbol. Define a preprocessor channel $\natural\colon \mathbb{F}_q \to \mathcal{X}$ that sends an actual symbol to itself, and a virtual symbol to a fixed actual symbol $x_1 \in \mathcal{X}$. In terms of transition probabilities:

$$\natural(y \mid x) = \begin{cases} 1 & \text{if } y = x \in \mathcal{X}, \\ 1 & \text{if } y = x_1 \text{ and } x \notin \mathcal{X}, \\ 0 & \text{otherwise.} \end{cases}$$

That is to say, all virtual symbols are semantic copies of x_1; they make no difference to the decoder whatsoever.

Now build the following degraded channel

$$\mathbb{F}_q \xrightarrow{\ \natural\ } \mathcal{X} \xrightarrow{\ W\ } \mathcal{Y}.$$

By the data processing inequality, the channel capacity of $W \circ \natural$ is at most the channel capacity of W. Hence using $\mathbb{F}_q$ as the input alphabet does no harm at best, and handicaps ourselves at worst. The former is the case: If we take any capacity-achieving input distribution Q of W and apply it directly to $W \circ \natural$, then we can transmit information at the same rate. In conclusion, W and $W \circ \natural$ share the same channel capacity. It suffices to consider $W \circ \natural$—channels with prime-power sized input alphabet—for the remainder of this chapter.

Remark: $\natural \circ W$ is not a symmetric channel; in particular, the uniform distribution does not achieve the best transmission rate. (Because that means x_1 is input more frequently than it should have been.) As a consequence, this technique of allowing dummy symbols, albeit trivial, does not apply until [**HY13**] taught us how to deal with asymmetric channels. Hence the section title *pay asymmetry to buy non-field*.

We hereby assume that $\mathcal{X} = \mathbb{F}_q$ is a finite field. If necessary, p is the characteristic of $\mathbb{F}_q$ and $\mathbb{F}_p$ denotes the ground field of $\mathbb{F}_q$. I am going to define more channel parameters in the next section; some of them require the finite field structure on $\mathcal{X}$ in a way that is not required/obvious when $q = 2$.

4. Parameters and Hölder Tolls

Let W be any DMC with input alphabet $\mathcal{X} = \mathbb{F}_q$ for some prime power q. To emphasize that the input alphabet was reduced to q, we call W a *q-ary channel*. The parameters below generalize H, I, and Z defined in Chapter 2 and T defined in Chapter 3.

Definition 5.5. The *bit error probability* of a q-ary channel W is

$$P_{\mathrm{e}}(W) \coloneqq \sum_{y \in \mathcal{Y}} W(y)\Big(1 - \max_{x \in \mathcal{X}} W(x \mid y)\Big),$$

which is how often the maximum a posteriori estimator $y \mapsto \arg\max_{x\in\mathcal{X}} W(x \mid y)$ makes a mistake.

Definition 5.6. The *Bhattacharyya parameter* of a q-ary channel W is

$$Z(W) \coloneqq \frac{1}{q-1} \sum_{\substack{x,x'\in\mathbb{F}_q \\ x\neq x'}} \sum_{y\in\mathcal{Y}} \sqrt{W(x,y)W(x',y)}.$$

In addition, define

$$Z_{\mathrm{mxd}}(W) \coloneqq \max_{0\neq d\in\mathbb{F}_q} \sum_{x\in\mathbb{F}_q} \sum_{y\in\mathcal{Y}} \sqrt{W(x,y)W(x+d,y)}.$$

The parameter below is not used directly in this work, but plays a central role in a theorem I cite. I include it for completeness:

$$Z_d(W) \coloneqq \sum_{x\in\mathbb{F}_q} \sum_{y\in\mathcal{Y}} \sqrt{W(x,y)W(x+d,y)}.$$

Remarks: Z is the average of Z_d over $d \in \mathbb{F}_q^\times$ and Z_{mxd} is the maximum thereof. The rescaling is such that $0 \leqslant Z \leqslant Z_{\mathrm{mxd}} \leqslant (q-1)Z \leqslant q-1$ and that $0 \leqslant Z_d \leqslant 1$. When $q=2$, i.e., for the binary case, $Z = Z_{\mathrm{mxd}} = Z_1 =$ the Z-parameter defined in Definition 2.4. Notice how the summation over distinct $x, x' \in \mathbb{F}_2$ turns into twice the case of $(x,x') = (0,1)$.

Definition 5.7. The *total variation norm* of a q-ary channel W is

$$T(W) \coloneqq \sum_{y\in\mathcal{Y}} W(y) \sum_{x\in\mathcal{X}} \left| W(x \mid y) - \frac{1}{q} \right|.$$

which is the total variation distance from $W(\bullet \mid y)$ to the uniform distribution, weighted by the frequency each $y \in \mathcal{Y}$ appears.

Definition 5.8. The *Fourier coefficients* of a q-ary channel W is

$$M(w \mid y) \coloneqq \sum_{z\in\mathbb{F}_q} W(z \mid y)\chi(wz),$$

where $\chi\colon \mathbb{F}_q \to \mathbb{C}$ is an additive character defined as $\chi(x) \coloneqq \exp(2\pi i \operatorname{tr}(x)/p)$, and $\operatorname{tr}\colon \mathbb{F}_q \to \mathbb{F}_p$ is the field trace onto the ground field.

Definition 5.9. The *Fourier ℓ^1 norm* of a q-ary channel W are

$$S(W) \coloneqq \frac{1}{q-1} \sum_{0\neq w\in\mathbb{F}_q} \sum_{y\in\mathcal{Y}} W(y) \cdot \Big| M(w \mid y) \Big|$$

In addition, define

$$S_{\max}(W) \coloneqq \max_{0\neq w\in\mathbb{F}_q} \sum_{y\in\mathcal{Y}} W(y) \cdot \Big| M(w \mid y) \Big|.$$

Remarks: The rescaling is such that $0 \leqslant S \leqslant S_{\max} \leqslant (q-1)S \leqslant q-1$. When $q=2$, the parameters collapse—$S = S_{\max} = T = 1 - 2P_{\mathrm{e}} =$ the T-parameter defined in Definition 3.5.

I borrow some lemmas to relate these parameters.

Lemma 5.10 (P vs Z)**.** [**MT14**, Lemma 22 with $k = 1$] *For any q-ary channel W,*

$$\frac{q-1}{q^2}\Big(\sqrt{1+(q-1)Z(W)}-\sqrt{1-Z(W)}\Big)^2 \leqslant P_{\mathrm{e}}(W) \leqslant \frac{q-1}{2}Z(W).$$

Lemma 5.11 (P vs T)**.** [**MT14**, Lemma 23 with $k = q-1$] *For any q-ary channel W,*

$$\frac{q-1}{q} - P_{\mathrm{e}}(W) \leqslant \frac{T(W)}{2} \leqslant \frac{q-1}{q} - \frac{1}{q}\Big((q-1)qP_{\mathrm{e}}(W) - (q-1)(q-2)\Big).$$

Lemma 5.12 (P vs S)**.** [**MT14**, Lemma 26 with $k = q-1$] *For any q-ary channel W,*

$$1 - \frac{q}{q-1}P_{\mathrm{e}}(W) \leqslant S(W) \leqslant (q-1)q\Big(\frac{q-1}{q} - P_{\mathrm{e}}(W)\Big)\sqrt{1-\frac{q}{q-1}\frac{q-2}{q-1}}.$$

Lemma 5.13 (P vs H)**.** [**FM94**, Theorem 1] *For any q-ary channel W,*

$$h_2(P_{\mathrm{e}}(W)) + P_{\mathrm{e}}(W)\log_2(q-1) \geqslant H(W)\log_2 q \geqslant 2P_{\mathrm{e}}(W),$$

$$H(W)\log_2 q \geqslant (q-1)q\log_2\frac{q}{q-1}\Big(P_{\mathrm{e}}(W) - \frac{q-2}{q-1}\Big) + \log_2(q-1).$$

Here, h_2 is the binary entropy function. The upper bound is Fano's inequality. The first lower bound is useful when $H(W)$ and $P_{\mathrm{e}}(W)$ are small; the second lower bound is useful when $H(W)$ and $P_{\mathrm{e}}(W)$ are close to 1.

The phenomenon described by the cited lemmas is distilled to form the following definition. It is the quantified version of the common-fate property.

Definition 5.14. Fix a q. Let A and B be two channel parameters. Say A and B are *bi-Hölder at* (a, b) if there exist constants $c, d > 0$ such that, for all q-ary channels, $|A(W) - a| \leqslant c \cdot |B(W) - b|^d$ and $|B(W) - b| \leqslant c \cdot |A(W) - a|^d$.

Bi-Hölder-ness is clearly an equivalence relation. In particular, we will make use of its transitivity property—if A and B are bi-Hölder at (a, b) and B and C are bi-Hölder at (b, c), then A and C are bi-Hölder at (a, c). In this case, we say A, B, and C are bi-Hölder at (a, b, c). This notion generalizes to arbitrarily many parameters. What Lemmas 5.10 to 5.13 tell us is where the parameters are bi-Hölder at.

Proposition 5.15 (Implicit bi-Hölder tolls)**.** *Fix a prime power q. Then channel parameters H, P_{e}, Z, and Z_{mxd} are bi-Hölder at $(0, 0, 0, 0)$. Channel parameters H, P_{e}, T, S, and $S_{\max}$ are bi-Hölder at $(1, 1-1/q, 0, 0, 0)$.*

PROOF. As for the first statement: Z, Z_{mxd} are bi-Hölder at $(0, 0)$ since $Z \leqslant Z_{\mathrm{mxd}} \leqslant (q-1)Z$. Lemma 5.10 implies that P_{e} and Z are bi-Hölder at $(0, 0)$. Lemma 5.13 (with the first lower bound) implies that P_{e} and H are bi-Hölder at $(0, 0)$. Now apply the transitivity to conclude the first statement.

As for the second statement: S and $S_{\max}$ are bi-Hölder at $(0, 0)$ since $S \leqslant S_{\max} \leqslant (q-1)S$. Lemma 5.12 implies that P_{e} and S are bi-Hölder at $(1-1/q, 0)$. Lemma 5.11 implies that P_{e} and T are bi-Hölder at $(1-1/q, 0)$. Lemma 5.13 (with the second lower bound) implies that P_{e} and H are bi-Hölder at $(1-1/q, 1)$. Now apply the transitivity to conclude the second statement. □

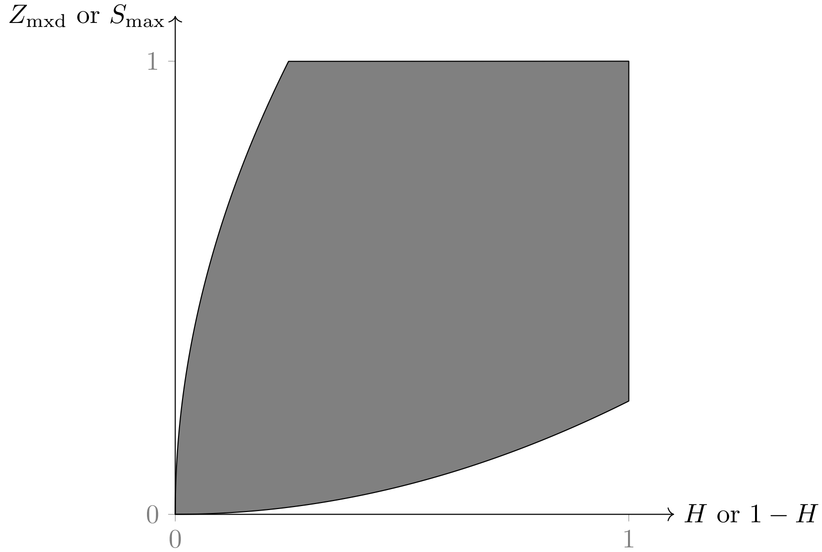

FIGURE 5.1. A figurative example of bi-Hölder relation. This plot assumes $c = 2$ and $d = 1/2$. But c is in general greater, which makes the dark region larger.

In short, when a channel is reliable, $H, P_{\mathrm{e}}, Z, Z_{\mathrm{mxd}}$ are small and $T, S, S_{\max}$ are large. When a channel is noisy, $H, P_{\mathrm{e}}, Z, Z_{\mathrm{mxd}}$ are large and $T, S, S_{\max}$ are small. See Figure 5.1 for an example of the possible region.

Sometimes, explicit bounds are desired as it saves some epsilon–delta notations. The bounds do not have to be tight, but they should be as easy to manipulate as possible. The next proposition serves exactly that purpose.

Proposition 5.16 (Explicit Hölder tolls)**.** *For all q-ary channels W, it holds that*

$$Z_{\mathrm{mxd}}(W) \leqslant q^3\sqrt{H(W)}, \tag{5.1}$$

$$H(W) \leqslant q^3\sqrt{Z_{\mathrm{mxd}}(W)}, \tag{5.2}$$

$$S_{\max}(W) \leqslant q^3\sqrt{1 - H(W)}, \tag{5.3}$$

$$1 - H(W) \leqslant q^3\sqrt{S_{\max}(W)}. \tag{5.4}$$

PROOF. The proof is nothing but working out Lemmas 5.10 to 5.13 very carefully. By "working out", I mean to Taylor expand every inequality at a proper point and keep track of how the constants c, d accumulate as transitivity applies.

In the upcoming arguments, H, P_{e}, Z, Z_{mxd}, S, and $S_{\max}$ mean $H(W)$, $P_{\mathrm{e}}(W)$, $Z(W)$, $Z_{\mathrm{mxd}}(W)$, $S(W)$, and $S_{\max}(W)$, respectively. Also q' means $q-1$, and q'' means $q-2$. Furthermore, lg means the base-2 logarithm; this is handy when we jump back and forth between nats, bits, and q-bits.

First we show inequality (5.1). Start from Z_{mxd}: By definition $Z_{\mathrm{mxd}} \leqslant q'Z$. Move on to Z: By Lemma 5.10, $q'q^{-2}(\sqrt{1+q'Z} - \sqrt{1-Z})^2 \leqslant P_{\mathrm{e}}$, hence $\sqrt{1+q'Z} - \sqrt{1-Z} \leqslant q\sqrt{P_{\mathrm{e}}/q'}$. Multiplying both sides by the conjugate yields $(1+q'Z) - (1-Z) \leqslant q\sqrt{P_{\mathrm{e}}/q'}(\sqrt{1+q'Z} + \sqrt{1-Z})$. The left-hand side is qZ; in the right-hand side $\sqrt{1+q'z} + \sqrt{1-z}$ assumes the maximum $q/\sqrt{q'}$ at $z = q''/q'$ by taking derivative. So $Z \leqslant \sqrt{P_{\mathrm{e}}/q'}(q/\sqrt{q'}) = q\sqrt{P_{\mathrm{e}}}/q'$. Move on to P_{e}: By Lemma 5.13 (the first lower bound), $2P_{\mathrm{e}} \leqslant H \lg q$ or equivalently $P_{\mathrm{e}} \leqslant H \log_4 q$. Now we chain the inequalities $Z_{\mathrm{mxd}} \leqslant q'Z \leqslant q\sqrt{P_{\mathrm{e}}} \leqslant q\sqrt{H \log_4 q}$. This completes inequality (5.1) as $q\sqrt{\log_4 q} < q^3$.

Second we show inequality (5.2). Start from H: By Lemma 5.13 (the upper bound, Fano's inequality), $H \lg q \leqslant h_2(P_{\mathrm{e}}) + P_{\mathrm{e}} \lg q'$. By Figure 5.2, $h_2(P_{\mathrm{e}}) + P_{\mathrm{e}} \lg q' \leqslant \sqrt{eP_{\mathrm{e}}} + P_{\mathrm{e}} \lg q' = \sqrt{P_{\mathrm{e}}}(\sqrt{e} + \sqrt{P_{\mathrm{e}}} \lg q')$. What is inside parentheses is less than $\sqrt{e} + \sqrt{q'/q} \lg q'$. Hence $H \leqslant \sqrt{P_{\mathrm{e}}}(\sqrt{e} + \sqrt{q'/q} \lg q')/\lg q$. Focus on the scalar—$(\sqrt{e} + \sqrt{q'/q} \lg q')/\lg q$ has maximum $\sqrt{e}$ at $q = 2$ (remember that $q \geqslant 2$). So $H \leqslant \sqrt{eP_{\mathrm{e}}}$. Move on to P_{e}: By Lemma 5.10, $P_{\mathrm{e}} \leqslant q'Z/2$. Move on to Z: By definition $Z \leqslant Z_{\mathrm{mxd}}$. Now we chain the inequalities $H \leqslant \sqrt{eP_{\mathrm{e}}} \leqslant \sqrt{eq'Z/2} \leqslant \sqrt{eq'Z_{\mathrm{mxd}}/2}$. This completes inequality (5.2) as $\sqrt{eq'/2} < q^3$.

Third we show inequality (5.3). Start from $S_{\max}$: By definition $S_{\max} \leqslant q'S$. Move on to S: By Lemma 5.12, $S \leqslant q'q(q'/q - P_{\mathrm{e}})\sqrt{1 - \frac{q}{q'}\frac{q''}{q'}}$. The square root simplifies to $\sqrt{1/(q')^2} = 1/q'$ as $qq'' = (q')^2 - 1$. So $S \leqslant q' - qP_{\mathrm{e}}$. Move on to $q' - qP_{\mathrm{e}}$: By Lemma 5.13 (the upper bound, Fano's inequality), $H \lg q \leqslant h_2(P_{\mathrm{e}}) + P_{\mathrm{e}} \lg q'$. We claim that $h_2(P_{\mathrm{e}}) + P_{\mathrm{e}} \lg q' \leqslant \lg q - 2(q'/q - P_{\mathrm{e}})^2/\ln 2$. To prove the claim, Taylor expand both sides at $P_{\mathrm{e}} = q'/q$. Verify that both evaluate to $\lg q$ at $P_{\mathrm{e}} = q'/q$; verify that both have derivative 0 at $P_{\mathrm{e}} = q'/q$; and verify that the acceleration of the left-hand side, $-1/(P_{\mathrm{e}}(1-P_{\mathrm{e}}) \ln 2)$, is more negative than the acceleration of the right-hand side, $-4/\ln 2$. By Taylor's theorem, mean value theorem, or the Euler method, the function with greater acceleration is greater; hence the claim. See also [**FM94**, Fig. 1]; the Φ-curve seems parabolic at the upper right corner. Now we have $H \lg q \leqslant \lg q - 2(q'/q - P_{\mathrm{e}})^2/\ln 2$, which is equivalent to $2(q'/q - P_{\mathrm{e}})^2/\ln q \leqslant 1 - H$ and to $q' - qP_{\mathrm{e}} \leqslant q\sqrt{(1-H)\ln(q)/2}$. Now we chain the inequalities $S_{\max} \leqslant q'S \leqslant q'(q' - qP_{\mathrm{e}}) \leqslant q'q\sqrt{(1-H)\ln(q)/2}$. This completes inequality (5.3) as $q'q\sqrt{\ln(q)/2} < q^3$.

Fourth we show inequality (5.4). Start from $1 - H$: By Lemma 5.13 (the second lower bound), $H \lg q \geqslant q'q \lg(q/q')(P_{\mathrm{e}} - q''/q') + \lg q'$. The right-hand side is $\lg q - q' \lg(q/q')(q' - qP_{\mathrm{e}})$ by matching the (rational) coefficients of $P_{\mathrm{e}} \lg q$, $P_{\mathrm{e}} \lg q'$, $\lg q$, and $\lg q'$, respectively. As $H \lg q \geqslant \lg q - q' \lg(q/q')(q' - qP_{\mathrm{e}})$ we bound $\lg(q/q') = \lg(1 + 1/q') \leqslant 1/q'$ by the tangent line at $1/q' = 0$. So $H \lg q \geqslant \lg q - (q' - qP_{\mathrm{e}})$ and hence $1 - H \leqslant (q' - qP_{\mathrm{e}})/\lg q$. Move on to $q' - qP_{\mathrm{e}}$: By Lemma 5.13, $1 - qP_{\mathrm{e}}/q' \leqslant S$ so $q' - qP_{\mathrm{e}} \leqslant q'S$. Move on to S: By definition $S \leqslant S_{\max}$. Now we chain the inequalities $1 - H \leqslant (q' - qP_{\mathrm{e}})/\lg q \leqslant q'S/\lg q \leqslant q'S_{\max}/\lg q$. This completes inequality (5.4) as $q'/\lg q < q^3$. □

Hölder-ness is a "toll" because, while Propositions 5.15 and 5.16 are the bridges that connect channel parameters, it feels like we pay fees, or taxes, to translate bounds on channels; and the fee/tax is charged regardless we go one way or another. For example, say we start with $H(W_n) < \exp(-n^{2/3})$, then $Z_{\mathrm{mxd}}(W_n) < q^3 \exp(-n^{2/3}/2)$, which contains annoying minor terms that harden the proof. Once we succeed in proving $Z_{\mathrm{mxd}}(W_n) < \exp\bigl(-e^{n^{1/3}}\bigr)$, we pay the price again when translating it back to $H(W_n) < q^3 \exp\bigl(-e^{n^{1/3}}/2\bigr)$.

The next section synthesizes channels on the basis of general matrices, and shows how synthetic channels relate to parameters defined here.

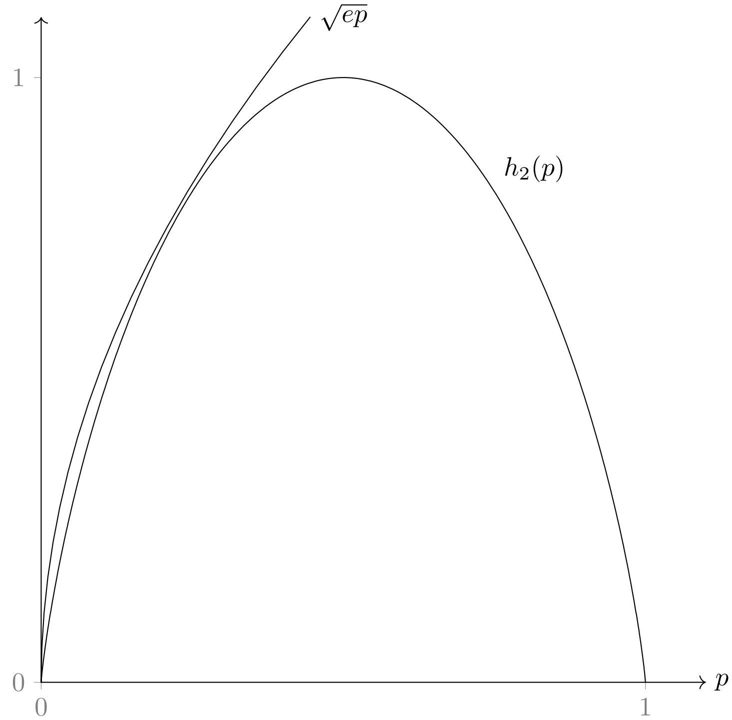

FIGURE 5.2. The binary entropy function and an upper bound of $\sqrt{ep}$.

5. Kernel and Fundamental Theorems

Fix a prime power q. Let $\ell \geqslant 2$. Let $G \in \mathbb{F}_q^{\ell\times\ell}$ be an invertible $\ell \times \ell$ matrix over $\mathbb{F}_q$. This matrix is to replace $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$ that was used to polarize channels in Chapters 2 to 4, and is called a *kernel*. See creffig:nine for an illustration of the EU–DU pairs that corresponds to a 3×3 kernel and how to wrap them around (synthetic) channels.

Let X_1^ℓ be ℓ i.i.d. random variables that follow Q, the capacity-achieving input distribution of W. Define $U_1^\ell \coloneqq X_1^\ell G^{-1}$. Let Y_1^ℓ be the outputs of ℓ i.i.d. copies of W given the inputs X_1^ℓ. That is to say, each (X_j, Y_j) is governed by $Q(x)$ and $W(y \mid x)$. Now consider the following guessing job:

- Guess U_1 given Y_1^ℓ.
- Guess U_2 given Y_1^ℓ, assuming that the guess $\hat{U}_1$ of U_1 is correct.
- Guess U_3 given Y_1^ℓ, assuming that the guesses $\hat{U}_1^2$ of U_1^2 are correct.

$$\vdots$$

- Guess U_ℓ given Y_1^ℓ, assuming that the guesses $\hat{U}_1^{\ell-1}$ of $U_1^{\ell-1}$ are correct.

For each $j = 1, 2, \ldots, \ell$, pretend that $W^{(j)}$ is a channel whose input is U_j and output is $Y_1^\ell U_1^{j-1}$. This channel captures the difficulty of the jth guessing job. The precise definition follows.

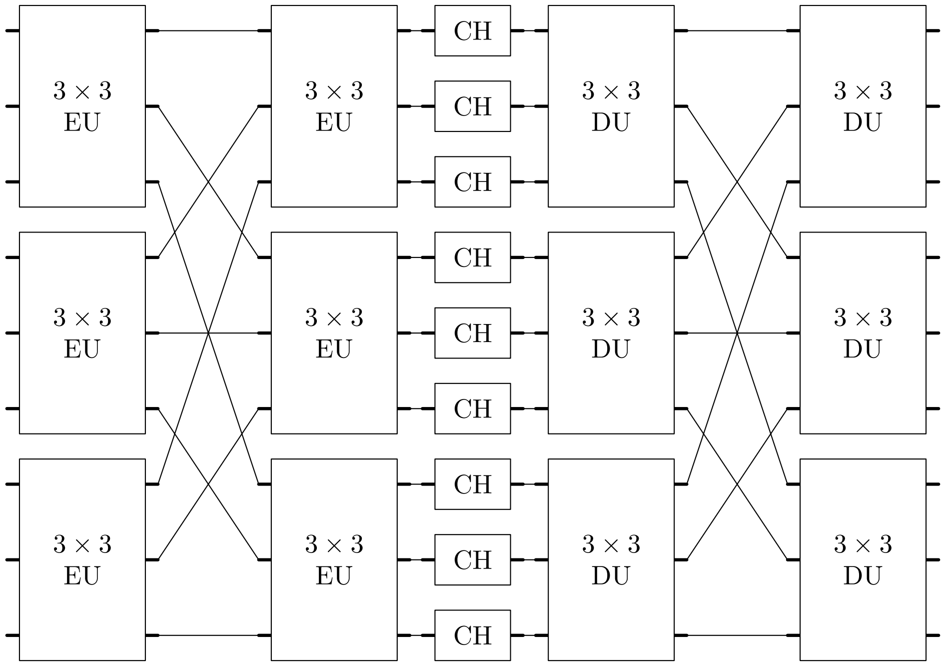

FIGURE 5.3. The design of encoder and decoder when $\ell = 3$.

Definition 5.17. For each $j = 1, 2, \ldots, \ell$, define a synthetic channel

$$W^{(j)}(y_1^\ell u_1^{j-1} \mid u_j) \coloneqq \frac{P\{Y_1^\ell U_1^{j-1} = y_1^\ell u_1^{j-1}\}}{P\{U_j = u_j\}} = \frac{\sum_{u_{j+1}^\ell \in \mathbb{F}_q^\ell} W(u_1^\ell G, y_k)}{\sum_{v_1^{j-1} u_{j+1}^\ell \in \mathbb{F}_q^{\ell-1}} Q^\ell(v_1^{j-1} u_j^\ell G)},$$

where Q^ℓ is the product measure and W^ℓ is the product channel, i.e.,

$$Q^\ell(x_1^\ell) = \prod_{j=1}^{\ell} Q(x_j), \qquad W^\ell(x_1^\ell, y_1^\ell) \coloneqq \prod_{j=1}^{\ell} W(x_j, y_j),$$

and $v_1^{j-1} u_j^\ell G$ is vector concatenation before vector-matrix multiplication.

There are three results that control $W^{(j)}$ via the parameters $H, Z_{\text{mxd}}, S_{\max}$ defined in the last section. These results play central roles in the theory of polar coding and is bestowed the title of fundamental theorems.

Theorem 5.18 (Fundamental theorem of polar coding—H version (FTPCH))**.** *For any q-ary channel W and any invertible matrix kernel $G \in \mathbb{F}_q^{\ell \times \ell}$,*

$$\sum_{j=1}^{\ell} H(W^{(j)}) = \ell H(W).$$

PROOF. We derive that

$$\sum_{j=1}^{\ell} H(W^{(j)}) = \sum_{j=1}^{\ell} H(U_j \mid Y_1^\ell U_1^{j-1}) = H(U_1^\ell \mid Y_1^\ell) = H(X_1^\ell \mid Y_1^\ell) = \ell H(X \mid Y).$$

The first equality is by definition. The next equality is by the chain rule of conditional entropy. The next equality is by G being invertible. The last equality is because (X_1^ℓ, Y_1^ℓ) are i.i.d. copies of (X, Y). $\square$

Fundamental theorem stated and proved, the other two fundamental theorems require more details about G to be stated.

Let $0_1^{j-1} 1_j u_{j+1}^\ell \in \mathbb{F}_q^\ell$ be a vector of 0 repeated $j-1$ times followed by a 1 and $\ell - j$ arbitrary symbols. A *coset code* is a subset of codewords of the form $\{0_1^{j-1} 1_j u_{j+1}^\ell G : u_{j+1}^\ell \in \mathbb{F}_q^{\ell-j}\} \subseteq \mathbb{F}_q^\ell$. The coset codes have weight distributions just like every other code does. Let $\operatorname{hwt}(x_1^\ell)$ be the Hamming weight of x_1^ℓ. The weight enumerator of the jth coset code is this one-variable polynomial over the integers

$$f_Z^{(j)}(z) \coloneqq \sum_{u_{j+1}^\ell} z^{\operatorname{hwt}(0_1^{j-1} 1_j u_{j+1}^\ell G)} \in \mathbb{Z}[z].$$

Note that this coincides with the distance enumerator from the jth row of G to the span of the rows beneath.

Theorem 5.19 (Fundamental theorem of polar coding—Z end (FTPCZ))**.** *For any q-ary channel W and any invertible matrix kernel $G \in \mathbb{F}_q^{\ell \times \ell}$,*

$$Z_{\mathrm{mxd}}(W^{(j)}) \leqslant f_Z^{(j)}(Z_{\mathrm{mxd}}(W)).$$

PROOF. By the definition of the synthetic channel $W^{(j)}$ and that of the Bhattacharyya parameter, $Z_{\mathrm{mxd}}(W^{(j)})$ is

$$\max_{0 \neq d_j \in \mathbb{F}_q} \sum_{u_j \in \mathbb{F}_q} \sum_{u_1^{j-1} y_1^\ell \in \mathbb{F}_q^j \times \mathcal{Y}^\ell} \sqrt{W^{(j)}(u_j, y_1^\ell u_1^{j-1}) W^{(j)}(u_j + d_j, y_1^\ell u_1^{j-1})}.$$

By the nature of $\max_{0 \neq d_j \in \mathbb{F}_q}$, it suffices to show that the double sum within is at most $f_Z^{(j)}(Z_{\mathrm{mxd}}(W))$ for arbitrary nonzero d_j.

In the upcoming argument, vector concatenation takes precedence over vector-matrix multiplication and vector addition. Fix a $d_j \in \mathbb{F}_q^\times$, we argue that

$$\begin{aligned}
&\sum_{u_j \in \mathbb{F}_q} \sum_{y_1^\ell u_1^{j-1} \in \mathcal{Y}^\ell \times \mathbb{F}_q^j} \sqrt{W^{(j)}(u_j, y_1^\ell u_1^{j-1}) W^{(j)}(u_j + d_j, y_1^\ell u_1^{j-1})} \\
&\quad = \sum_{u_1^j y_1^\ell} \sqrt{W^{(j)}(u_j, y_1^\ell u_1^{j-1}) W^{(j)}(u_j + d_j, y_1^\ell u_1^{j-1})} \\
&\quad = \sum_{u_1^j y_1^\ell} \sqrt{\sum_{u_{j+1}^\ell \in \mathbb{F}_q^{\ell-j}} W^\ell(u_1^j u_{j+1}^\ell G, y_1^\ell) \sum_{v_{j+1}^\ell \in \mathbb{F}_q^{\ell-j}} W^\ell(u_1^{j-1}(u_j + d_j) v_{j+1}^\ell G, y_1^\ell)} \\
&\quad \leqslant \sum_{u_1^j y_1^\ell} \sum_{u_{j+1}^\ell} \sum_{v_{j+1}^\ell} \sqrt{W^\ell(u_1^j u_{j+1}^\ell G, y_1^\ell) W^\ell(u_1^{j-1}(u_j + d_j) v_{j+1}^\ell G, y_1^\ell)} \\
&\quad = \sum_{y_1^\ell} \sum_{u_1^\ell} \sum_{d_{j+1}^\ell \in \mathbb{F}_q^{\ell-j}} \sqrt{W^\ell(u_1^\ell G, y_1^\ell) W^\ell(u_1^{j-1}(u_j^\ell + d_j^\ell) G, y_1^\ell)} \\
&\quad = \sum_{y_1^\ell} \sum_{x_1^\ell \in \mathbb{F}_q^\ell} \sum_{d_{j+1}^\ell} \sqrt{W^\ell(x_1^\ell, y_1^\ell) W^\ell(x_1^\ell + 0_1^{j-1} d_j^\ell G, y_1^\ell)}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{d_{j+1}^\ell} \sum_{y_1^\ell} \sum_{x_1^\ell} \sqrt{W^\ell(x_1^\ell, y_1^\ell) W^\ell(x_1^\ell + e_1^\ell, y_1^\ell)} \\
&= \sum_{d_{j+1}^\ell} \sum_{y_1^\ell} \sum_{x_1^\ell} \prod_{k\in[\ell]} \sqrt{W(x_k, y_k) W(x_k + e_k, y_k)} \\
&= \sum_{d_{j+1}^\ell} \sum_{y_1^\ell} \sum_{x_1^\ell} \prod_{k\in K} \sqrt{W(x_k, y_k) W(x_k + e_k, y_k)} \prod_{k\notin K} W(x_k, y_k) \\
&= \sum_{d_{j+1}^\ell} \prod_{k\in K} \Bigl(\sum_{x_k y_k} \sqrt{W(x_k, y_k) W(x_k + e_k, y_k)} \Bigr) \prod_{k\notin K} \Bigl(\sum_{x_k y_k} W(x_k, y_k) \Bigr) \\
&= \sum_{d_{j+1}^\ell} \prod_{k\in K} \Bigl(\sum_{x_k y_k} \sqrt{W(x_k, y_k) W(x_k + e_k, y_k)} \Bigr) \\
&\leqslant \sum_{d_{j+1}^\ell} \prod_{k\in K} \max_{0\neq e_k\in\mathbb{F}_q} \Bigl(\sum_{x_k y_k} \sqrt{W(x_k, y_k) W(x_k + e_k, y_k)} \Bigr) \\
&= \sum_{d_{j+1}^\ell} \prod_{k\in k} Z_{\mathrm{mxd}}(W) = \sum_{d_{j+1}^\ell} Z_{\mathrm{mxd}}(W)^{|K|} = \sum_{d_{j+1}^\ell} Z_{\mathrm{mxd}}(W)^{\mathrm{hwt}(0_1^{j-1} d_j d_{j+1}^\ell G)} \\
&= \sum_{d_{j+1}^\ell} Z_{\mathrm{mxd}}(W)^{\mathrm{hwt}(0_1^{j-1} 1_j d_{j+1}^\ell G)} = f_Z^{(j)}(Z_{\mathrm{mxd}}(W)).
\end{aligned}
$$

The first equality abbreviates the summation. The next equality expands $W^{(j)}$ by the very definition, where u_{j+1}^ℓ and v_{j+1}^ℓ are free variables in $\mathbb{F}_q$. The next inequality is by sub-additivity of square root. In the next equality we define $d_{j+1}^\ell \coloneqq v_{j+1}^\ell - u_{j+1}^\ell$; so summing over v_{j+1}^ℓ is equivalent to summing over d_{j+1}^ℓ. In the next equality we define $x_1^\ell \coloneqq u_1^\ell G$; so summing over u_1^ℓ is equivalent to summing over x_1^ℓ as G is invertible. In the next equality we substitute $e_1^\ell \coloneqq 0_1^{j-1} d_j^\ell G$ and reorder the summation. The next equality expands the product of the memoryless channels. The next equality classifies the indices into two classes—$k \in K$ are those such that $e_k \neq 0$ and $k \notin K$ are those such that $e_k = 0$. The next equality is the distributive law $ax+ay+bx+by = (a+b)(x+y)$. The next equality uses the fact that $W(x,y)$ sum to 1. In the next inequality we replace e_k by a nonzero element that maximizes the sum in the parentheses. In the next equality we realize that the maximum is the Bhattacharyya parameter (surprisingly). The second last equality uses the fact that multiplying a vector by a scalar preserves its Hamming weight. And quod erat demonstrandum. □

The last fundamental theorem lives in the dual picture. Let $u_1^{j-1} 1_j 0_{j+1}^\ell \in \mathbb{F}_q^\ell$ be a vector of $j-1$ arbitrary symbols followed by a 1 and 0 repeated $\ell - j$ times. Let $G^{-\top}$ be the inverse transpose of G. The jth dual coset code of G is a subset of codewords of the form $\{u_1^{j-1} 1_j 0_{j+1}^\ell G^{-\top} : u_1^{j-1} \in \mathbb{F}_q^{j-1}\} \subseteq \mathbb{F}_q^\ell$. The weight enumerator of the jth dual coset code is defined to be this one-variable polynomial over the integers

$$f_S^{(j)}(s) \coloneqq \sum_{u_1^{j-1}} s^{\mathrm{hwt}(u_1^{j-1} 1_j 0_{j+1}^\ell G^{-\top})} \in \mathbb{Z}[s].$$

We can now state the dual of the second fundamental theorem.

Theorem 5.20 (Fundamental theorem of polar coding—S-end (FTPCS))**.** *For any q-ary channel W and any invertible matrix kernel $G \in \mathbb{F}_q^{\ell\times\ell}$,*

$$S_{\max}(W^{(j)}) \leqslant f_Z^{(j)}(S_{\max}(W)).$$

PROOF. (Reminder: For those who attempt to skip the proof, the proof spans about 2.2 pages.)

Recall the character $\chi(x) \coloneqq \exp(2\pi i \operatorname{tr}(x)/p)$. We need these properties of χ: (xa) $\chi(0) = 1$; (xb) $|\chi(x)| = 1$ for all $x \in \mathbb{F}_q$; (xc) $\chi(x)\chi(z) = \chi(x+z)$ for all $x, z \in \mathbb{F}_q$; and (xd) $\sum_{x\in\mathbb{F}_q} \chi(x) = 0$. See also [**MT14**, Definition 24] or a dedicated book [**Ter99**]. To prove the theorem, we first verify that the Fourier coefficients recover the origin: Let $M(w, y) \coloneqq W(y)M(w \mid y) = \sum_{z\in\mathbb{F}_q} W(z, y)\chi(wz)$, then

$$\sum_{w\in\mathbb{F}_q} M(w,y)\chi(-xw) = \sum_{w\in\mathbb{F}_q}\sum_{z\in\mathbb{F}_q} W(z,y)\chi(wz)\chi(-xw)$$
$$= \sum_{z\in\mathbb{F}_q} W(z,y) \sum_{w\in\mathbb{F}_q} \chi(w(z-x)) = \sum_{z\in\mathbb{F}_q} W(z,y) q\mathbb{I}\{z-x=0\} = qW(x,y).$$

The first equality expands $M(w, y)$ by the definition. The next equality uses that χ is an additive character (xc), and reorders the summation. The next equality uses $\sum_{w\in\mathbb{F}_q} \chi(w) = 0$ (xd) and $\sum_{w\in\mathbb{F}_q} \chi(0) = q$ (xa); and $\mathbb{I}$ is the indicator function.

Knowing $W(x_j, y_j) = q^{-1} \sum_{w_j\in\mathbb{F}_q} M(w_j, y_j)\chi(-x_j w_j)$, we proceed to

$$W^{(j)}(u_j, y_1^\ell u_1^{j-1}) = \sum_{u_{j+1}^\ell} W^\ell(u_1^\ell G, y_1^\ell) = \sum_{u_{j+1}^\ell \in\mathbb{F}_q^{\ell-j}} W^\ell(x_1^\ell, y_1^\ell) = \sum_{u_{j+1}^\ell} \prod_{k\in[\ell]} W(x_k, y_k)$$
$$= \sum_{u_{j+1}^\ell} \prod_{k\in[\ell]} \Big(\frac{1}{q}\sum_{w_k\in\mathbb{F}_q} M(w_k,y_k)\chi(-x_k w_k)\Big) = \frac{1}{q^\ell}\sum_{u_{j+1}^\ell}\sum_{w_1^\ell}\prod_{k\in[\ell]} M(w_k,y_k)\chi(-x_k w_k)$$
$$= \frac{1}{q^\ell}\sum_{u_{j+1}^\ell}\sum_{w_1^\ell} \chi(-x_1^\ell (w_1^\ell)^\top) \prod_{k\in[\ell]} M(w_k,y_k) = \frac{1}{q^\ell}\sum_{u_{j+1}^\ell}\sum_{w_1^\ell} \chi(-x_1^\ell (w_1^\ell)^\top) M^\ell(w_1^\ell, y_1^\ell)$$
$$= \frac{1}{q^\ell}\sum_{u_{j+1}^\ell}\sum_{w_1^\ell} \chi(-u_1^\ell G (w_1^\ell)^\top) M^\ell(w_1^\ell, y_1^\ell) = \frac{1}{q^\ell}\sum_{u_{j+1}^\ell}\sum_{w_1^\ell} \chi(-u_1^\ell (w_1^\ell G^\top)^\top) M^\ell(w_1^\ell, y_1^\ell)$$
$$= \frac{1}{q^\ell}\sum_{u_{j+1}^\ell}\sum_{v_1^\ell} \chi(-u_1^\ell (v_1^\ell)^\top) M^\ell(v_1^\ell G^{-\top}, y_1^\ell)$$
$$= \frac{1}{q^\ell}\sum_{v_1^\ell} \chi(-u_1^j (v_1^j)^\top) M^\ell(v_1^\ell G^{-\top}, y_1^\ell) \sum_{u_{j+1}^\ell} \chi(-u_{j+1}^\ell (v_{j+1}^\ell)^\top)$$
$$= \frac{1}{q^\ell}\sum_{v_1^\ell} \chi(-u_1^j (v_1^j)^\top) M^\ell(v_1^\ell G^{-\top}, y_1^\ell) q^{\ell-j}\mathbb{I}\{v_{j+1}^\ell = 0\}$$
$$= \frac{1}{q^j}\sum_{v_1^j} \chi(-u_1^j (v_1^j)^\top) M^\ell(v_1^j 0_{j+1}^\ell G^{-\top}, y_1^\ell).$$

The first equality expands the definition of $W^{(j)}$. In the next equality, we substitute $x_1^\ell \coloneqq u_1^\ell G$. The next equality expands the definition of W^ℓ down to W. The next two equalities Fourier expand W and reorder the operators. The next equality merges all $\chi(-x_k w_k)$ into one inner product by additivity (xc). In the next equality we define $M^\ell(w_1^\ell, y_1^\ell)$ to be the product of all $M(w_k, y_k)$. The next two equalities use $x_1^\ell(w_1^\ell)^\top = u_1^\ell G(w_1^\ell)^\top = u_1^\ell(w_1^\ell G^\top)^\top$. In the next equality we define $v_1^\ell \coloneqq w_1^\ell G^\top$; so summing over w_1^ℓ is equivalent to summing over v_1^ℓ. The last three equalities sum over u_{j+1}^ℓ to force $v_{j+1}^\ell = 0$.

Having that $W^{(j)}(u_j, y_1^\ell u_1^{j-1}) = q^{-j}\sum_{v_1^j}\chi(-u_1^j(v_1^j)^\top)M^\ell(v_1^j 0_{j+1}^\ell G^{-\top}, y_1^\ell)$ in mind, we move on to

$$\begin{aligned}
M^{(j)}(\omega_j, u_1^{j-1}y_1^\ell) &\coloneqq \sum_{z_j\in\mathbb{F}_q} W^{(j)}(z_j, y_1^\ell u_1^{j-1})\chi(\omega_j z_j)\\
&= \sum_{z_j\in\mathbb{F}_q}\frac{1}{q^j}\sum_{v_1^j}\chi(-u_1^{j-1}z_j(v_1^j)^\top)M^\ell(v_1^j0_{j+1}^\ell G^{-\top}, y_1^\ell)\chi(\omega_j z_j)\\
&= \frac{1}{q^j}\sum_{v_1^j}\chi(-u_1^{j-1}(v_1^{j-1})^\top)M^\ell(v_1^j0_{j+1}^\ell G^{-\top}, y_1^\ell)\sum_{z_j\in\mathbb{F}_q}\chi(z_j(\omega_j - v_j))\\
&= \frac{1}{q^j}\sum_{v_1^j}\chi(-u_1^{j-1}(v_1^{j-1})^\top)M^\ell(v_1^j0_{j+1}^\ell G^{-\top}, y_1^\ell)q\mathbb{I}\{\omega_j = v_j\}\\
&= \frac{q}{q^j}\sum_{v_1^{j-1}}\chi(-u_1^{j-1}(v_1^{j-1})^\top)M^\ell(v_1^{j-1}\omega_j0_{j+1}^\ell G^{-\top}, y_1^\ell).
\end{aligned}$$

In the first line we let $M^{(j)}$ be the Fourier coefficients of $W^{(j)}$. The next equality plugs in what we have in mind about $W^{(j)}$. The next three equalities sum over z_j to force $v_j = \omega_j$.

With the fact that $M^{(j)}(\omega_j, u_1^{j-1}y_1^\ell)$ is equal to $q^{1-j}\sum_{v_1^{j-1}}\chi(-u_1^{j-1}(v_1^{j-1})^\top)\times M^\ell(v_1^{j-1}\omega_j 0_{j+1}^\ell G^{-\top}, y_1^\ell)$, we obtain that with arbitrary $0 \neq \omega_j \in \mathbb{F}_q$,

$$\begin{aligned}
&\sum_{u_1^{j-1}y_1^\ell\in\mathbb{F}^{j-1}\times\mathcal{Y}^\ell}|M^{(j)}(\omega_j, u_1^{j-1}y_1^\ell)| \qquad (5.5)\\
&= \sum_{u_1^{j-1}y_1^\ell}\left|\frac{q}{q^j}\sum_{v_1^{j-1}}\chi(-u_1^{j-1}(v_1^{j-1})^\top)M^\ell(v_1^{j-1}\omega_j0_{j+1}^\ell G^{-\top}, y_1^\ell)\right|\\
&\leqslant \sum_{u_1^{j-1}y_1^\ell}\frac{q}{q^j}\sum_{v_1^{j-1}}|M^\ell(v_1^{j-1}\omega_j0_{j+1}^\ell G^{-\top}, y_1^\ell)|\\
&= \sum_{y_1^\ell}\sum_{v_1^{j-1}}|M^\ell(v_1^{j-1}\omega_j0_{j+1}^\ell G^{-\top}, y_1^\ell)| = \sum_{y_1^\ell}\sum_{v_1^{j-1}}\prod_{k\in[\ell]}|M(w_k, y_k)|\\
&= \sum_{y_1^\ell}\sum_{v_1^{j-1}}\prod_{k\in K}|M(w_k, y_k)|\prod_{k\notin K}|M(w_k, y_k)|\\
&= \sum_{v_1^{j-1}}\prod_{k\in K}\Big(\sum_{y_k}|M(w_k, y_k)|\Big)\prod_{k\notin K}\Big(\sum_{y_k}|M(w_k, y_k)|\Big)\\
&= \sum_{v_1^{j-1}}\prod_{k\in K}\Big(\sum_{y_k}|M(w_k, y_k)|\Big) \leqslant \sum_{v_1^{j-1}}\prod_{k\in K}S_{\max}(W)\\
&= \sum_{v_1^{j-1}}S_{\max}(W)^{|K|} = \sum_{v_1^{j-1}}S_{\max}(W)^{\operatorname{hwt}(v_1^{j-1}\omega_j0_{j+1}^\ell G^{-\top})}
\end{aligned}$$

$$= \sum_{v_1^{j-1}} S_{\max}(W)^{\mathrm{hwt}(v_1^{j-1}1_j0_{j+1}^{\ell}G^{-\top})} = f_S^{(j)}(S_{\max}(W)).$$

The first equality expands the Fourier coefficients. The next inequality is triangle inequality plus (xb). The next equality cancels the summation over u_1^{j-1} with q^{1-j}. In the next equality we substitute $w_1^\ell \coloneqq v_1^{j-1}\omega_j 0_{j+1}^\ell G^{-\top}$; slightly different from the free w_1^ℓ before, they are now restricted to a proper subspace. The next equality classifies the indices into two classes—$j \in K$ are those such that $w_j \neq 0$ and $k \notin K$ are such that $w_k = 0$. The next two equalities reorder the operators and simplify $\sum_{y_k} |M(0, y_k)| = \sum_{y_k} W(y_k) = 1$. The next inequality replaces w_j by one that maximizes $\sum_{y_k} |M(w_j, y_k)|$. The rest is trivial.

Theorem 5.20 claims that $S_{\max}(W^{(j)}) \leqslant f_S^{(j)}(S_{\max}(W))$. Since $S_{\max}(W^{(j)})$ is merely the maximum of formula (5.5) over $0 \neq \omega_j \in \mathbb{F}_q$, we arrive at $S_{\max}(W^{(j)}) \leqslant f_S^{(j)}(S_{\max}(W))$. And quod erat demonstrandum. □

Sometimes, a simpler bound is enough for the analysis—instead of the exact weight enumerator, we use (#codewords)$z^{\text{minimum distance}}$ as an upper bound. The number of codewords is easy to predict (it is some power of 2). So one only needs to record the minimum distances.

Definition 5.21. For any $G \in \mathbb{F}_q^{\ell\times\ell}$, define *coset distance*

$$D_Z^{(j)} \coloneqq \mathrm{hdis}(r_j, R_j),$$

where hdis is the Hamming distance, r_j is the jth row of G, and R_j is the subspace spanned by the rows beneath r_j.

Then FTPCZ reads $Z_{\mathrm{mxd}}(W^{(j)}) \leqslant q^{\ell-j} Z_{\mathrm{mxd}}(W)^{D_Z^{(j)}}$. The dual picture obeys the same logic.

Definition 5.22. For any $G \in \mathbb{F}_q^{\ell\times\ell}$, define *dual coset distance*

$$D_S^{(j)} \coloneqq \mathrm{hdis}(c_j, C_j),$$

where hdis is the Hamming distance, c_j is the jth column of G^{-1}, and C_j is the subspace spanned by the columns before c_j.

Then FTPCS reads $S_{\max}(W^{(j)}) \leqslant q^{j-1} S_{\max}(W)^{D_Z^{(j)}}$.

The fundamental theorems introduced here are the “local relations” that control how one iteration of the channel transformation manipulates W. But in the end, we want to talk about all of its descendants at once, hence the introduction of the stochastic processes.

Definition 5.23. Let $\mathsf{J}_1, \mathsf{J}_2, \ldots$ be i.i.d. uniform random variables taking values in $\{1, 2, \ldots, \ell\}$. Fix a q-ary channel W and an invertible matrix $G \in \mathbb{F}_q^{\ell\times\ell}$. Let $\mathsf{W}_9, \mathsf{W}_1, \mathsf{W}_2, \ldots$, or $\{\mathsf{W}_n\}$ in short, be a stochastic process of DMCs defined as follows:

- $\mathsf{W}_0 \coloneqq W$; and
- $\mathsf{W}_{n+1} \coloneqq \mathsf{W}_n^{(\mathsf{J}_{n+1})}$.

This is called the *channel process*.

Imagine the following infinite array:

$$\begin{bmatrix} S_{\max}(W_0) & S_{\max}(W_1) & S_{\max}(W_2) & S_{\max}(W_3) & \cdots \\ T(W_0) & T(W_1) & T(W_2) & T(W_3) & \cdots \\ H(W_0) & H(W_1) & H(W_2) & H(W_3) & \cdots \\ P_{\mathrm{e}}(W_0) & P_{\mathrm{e}}(W_1) & P_{\mathrm{e}}(W_2) & P_{\mathrm{e}}(W_3) & \cdots \\ Z_{\mathrm{mxd}}(W_0) & Z_{\mathrm{mxd}}(W_1) & Z_{\mathrm{mxd}}(W_2) & Z_{\mathrm{mxd}}(W_3) & \cdots \end{bmatrix}$$

The Hölder tolls, Lemmas 5.10 to 5.13 and Propositions 5.15 and 5.16, are vertical relations. The fundamental theorems, Theorems 5.18 to 5.20, are horizontal relations. The top two rows are related to the distortion of lossy compression and shaping the input distributions for asymmetric channels. The bottom two rows are related to the block error probability of lossless compression and the anti-error part of noisy-channel coding. The middle row controls the code rate. In particular, Theorem 5.18 implies the following generalization of Proposition 2.10.

Proposition 5.24. $\{H(W_n)\}$ *is a martingale.*

Propositions 5.15 and 5.16 and Theorems 5.18 to 5.20 are all we need to control the behavior of $\{W_n\}$. But before we make use of these tools to examine the performance of polar coding, let us walk through some terminologies to see the big picture and what to expect.

6. Probability Theory Regimes

There is an analogy between coding theory and probability theory that connects the results from both sides and the proofs thereof. This analogy constituents the picture of the expected performance of coding. This section is a brief introduction to that and is inspired by [**AW14**].

Consider i.i.d. copies of some bounded random variable $X_1, X_2, \ldots, X_N$ and their average $\bar{X}_N$. We want to understand the distribution of $\bar{X}_N$, i.e., we want to understand $P\{\bar{X}_N \leqslant x\}$ for various x. The key is to identify the following

- the mean $\mu \coloneqq E[X_1]$ with channel capacity C,
- the number of copies N with the block length N,
- the cutoff x with the code rate R
- the cumulative probability $P\{\bar{X}_N \leqslant x\}$ with the block error probability P_{e}, and
- the variance σ^2 with another channel parameter called V.

For example, when Shannon said there exist error correcting codes with code rate $R < C$ close to channel capacity and block error probability $P_{\mathrm{e}} \to 0$, this translates into when the cutoff $x < \mu$ is close to the mean, the cumulative probability $P\{\bar{X}_N \leqslant x\}$ converges to 0 as $N \to \infty$. The latter is the law of large numbers (LLN). That establishes the first analogue.

Later, Gallager said that the block error probability scales like $\exp(-E_{\mathrm{r}}(R)N)$ for a fixed $R < C$. It translates into that the cumulative probability scales like $P\{\bar{X}_N \leqslant x\} \approx \exp(-L(x)N)$ for a fixed cutoff $z < \mu$ [**Gal68**]. The former is called the error exponent regime; the latter is called the *large deviation principle* (LDP). Moreover, Gallager's error exponent function E_{r} is analogous to Cramér's rate function $L(x)$. (To avoid abusing the word rate, I will referred to this as the Cramér function.) The former is the convex conjugate of Gallager's E_0 function; the latter is the convex conjugate of the cumulant generating function $t \mapsto \ln E[\exp(tX_1)]$. That establishes the second analogue.

TABLE 5.1. An analogy among probability theory, random coding theory, and polar coding theory. All $\delta > 0$ can be made arbitrarily close to 0.

	Random variables	Random codes	Polar codes
LLN	$\bar{X} \to \mu$	$(P_{\mathrm{e}}, R) \to (0, C)$	$(P_{\mathrm{e}}, R) \to (0, C)$
LDP	$\mathbb{P}\{\bar{X} - \mu > x\} \approx e^{-NL(x)}$	$P_{\mathrm{e}} \approx e^{-E_{\mathrm{r}}(R)N}$	$P_{\mathrm{e}} \approx e^{-N^{1-\delta}}$
CLT	$\bar{X} - \mu \sim \mathcal{N}(0, \frac{\sigma}{\sqrt{N}})$	$C - R \approx Q^{-1}(P_{\mathrm{e}})\sqrt{\frac{V}{N}}$	$C - R \approx N^{-1/2+\delta}$
MDP	$\frac{-\ln P\{\bar{X}-\mu>\gamma(N)x\}}{\gamma(N)^2} \approx NL(x)$	$\frac{-\ln P_{\mathrm{e}}}{(C-R)^2} \approx \frac{N}{2V}$	$\frac{-\ln P_{\mathrm{e}}}{(C-R)^2} \approx N^{1-\delta}$

On a parallel track, Strassen said that the code rate scales like $C + \Phi^{-1}(P_{\mathrm{e}}) \times \sqrt{V/N}$, where Φ is the cdf of the standard normal distribution, and V is another intrinsic parameter called channel dispersion or varentropy [**Str62**]. It translates into that the cutoff scales like $x \approx \mu + \Phi^{-1}(p)\sqrt{\sigma^2/N}$ in order to attain a certain cumulative probability p. The former is called the finite block length regime (although, in fact, both this and the error exponent regime has finite N); the latter is called the *central limit theorem* (CLT). That establishes the third analogue. Moreover V is the analog of variance in coding theory.

It is clear that LDP and CLT generalize LLN in ways that fix a variable and inspect the asymptote of the other variable. A third regime varies both. By [**AW14, PV10**],

$$\frac{-\ln P_{\mathrm{e}}}{N(C-R)^2} \longrightarrow \frac{1}{2V}.$$

And it specializes to $P_{\mathrm{e}} \approx \exp(-\Omega(N))$ for a fixed $R < C$ and $R \approx C - O(1)/\sqrt{N}$ for a fixed P_{e}. Similarly, in probability theory, there is

$$\frac{\ln P\{\bar{X} - \mu > \gamma(N)x\}}{N\gamma(N)^2} \longrightarrow L(x),$$

where $\gamma(N)$ is some appropriate re-scalars, and $L(x) = 1/\sigma^2 x^2$ is another Cramér function . This is called the *moderate deviation principle* (MDP). That establishes the fourth analogue.

For their achievability bounds, the aforementioned results use random coding whose complexity is out of control. Beside random coding, polar coding is the only low-complexity code that is strong enough to approach the LDP, CLT, and MDP regimes. Some history is briefed below. See also Table 5.1 for a comparison among probability theory and random and polar coding.

Arıkan's original works on channel polarization [**Ari09**] established the foundation of polar coding, placing polar coding in the LLN paradigm. Arıkan–Telatar [**AT09**] characterized the LDP behavior of polar coding, showing that P_{e} scales like $\exp(-\sqrt{N})$ when an $R < I$ is fixed. Later, Korada–Şaşoğlu–Urbanke [**KSU10**] generalized polar codes from Arıkan's kernel $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$ to any invertible $\ell \times \ell$ matrix G over $\mathbb{F}_2$, granted that $\ell \geqslant 2$ and G is not column-equivalent to a lower triangular matrix. And then they showed that the LDP behavior is $P_{\mathrm{e}} \approx \exp(-N^{\mathsf{E}_{\mathrm{c}}(G)})$ where $\mathsf{E}_{\mathrm{c}}(G)$ is a constant depending on the kernel matrix G. In fact, $\mathsf{E}_{\mathrm{c}}(G)$ is the expectation of $-\log_\ell \mathsf{D}_1$. The notation $\mathsf{E}_{\mathrm{c}}(G)$ is meant to resemble Gallager's error exponent $E_{\mathrm{r}}(R)$; but be aware of that the former is inserted at $\exp(-N^{\text{this}})$ place while the latter is inserted at $\exp(-\text{this}N)$ place. The LDP behavior of polar codes is then refined in [**HMTU13**]. Therein, P_{e} is approximated by $\exp(-\ell^{\mathfrak{E}})$ where $\mathfrak{E} := \mathsf{E}_{\mathrm{c}}(G)n - \sqrt{\mathsf{V}_{\mathrm{c}}(G)n}Q^{-1}(R/I) + o(\sqrt{n})$ is a more accurate exponent, ℓ is the matrix dimension, n is the depth of the code, and $\mathsf{V}_{\mathrm{c}}(G)$ is another constant depending on G. The notation $\mathsf{V}_{\mathrm{c}}(G)$ is meant to resemble the channel dispersion V. Appearing to be a CLT behavior, this result lies in the corner of the LDP paradigm that touches the MDP paradigm. Finally, Mori–Tanaka [**MT14**] generalized everything above from binary input to channels with prime-power input. Over arbitrary input alphabets, [**ŞTA09, Sas11**] showed the equivalence of [**Ari09, AT09**]. Over binary but asymmetric channels, [**SRDR12, HY13**] showed the counterpart of [**Ari09, AT09**] with the channel capacity in place of the symmetric capacity I. No further result on the LDP side, e.g. over non-binary asymmetric channels, is known. The technique in [**HY13**] (inequality (3.4)) and section 3 fill the gap.

The CLT behavior of polar codes turns out to be difficult to characterize. It was Korada–Montanari–Telatar–Urbanke [**KMTU10**] who came up with the idea that approximating an eigenfunction tightly bounds the eigenvalue $\ell^{-\varrho}$. Here $\varrho > 0$ will become a number such that R scales like $I - N^{-\varrho}$ with a fixed P_{e}. It is called the scaling exponent because it controls the scaling of N as a function of the gap to capacity $I - R$. (Although the LDP regime can also be rephrased as the scaling of N as a function of $\ln P_{\mathrm{e}}$, it was named error exponent regime beforehand. Thus the name scaling exponent [regime] is dedicated to the CLT regime). They had $0.2669 \leqslant \varrho \leqslant 0.2841$ over binary erasure channels (BECs). The upper bound was brought down to $3.553\varrho \leqslant 1$ over binary-input discrete-output memoryless channels (BDMCs) [**GHU12**]. Hassini–Alishahi–Urbanke [**HAU14**] moved down the upper bound to $3.627\varrho \leqslant 1$ over BECs and $3.579\varrho \leqslant 1$ over BDMCs. They also proved a lower bound $1 \leqslant 6\varrho$ over BDMCs. The latter is suboptimal and [**GB14, MHU16**] improved the bound to $1 \leqslant 5.702\varrho$ and to $1 \leqslant 4.714\varrho$. Additive white Gaussian noise channles (AWGNCs) have continuous output alphabet, but [**FT17**] show that they have $1 \leqslant 4.714\varrho$ too. Over BECs particularly, [**FV14, YFV19**] examined a series of larger kernels; the current record is a 64×64 kernel believed to have $1 \leqslant 2.9\varrho$. Near the end of the road to $2\varrho < 1$, [**PU16**] showed that by allowing $q \to \infty$, Reed–Solomon kernels achieve $2\varrho < 1$ over q-ary channels. This does not really prove that polar codes achieve $2\varrho < 1$ over any specific channel, but gave hopes. Fazeli–Hassani–Mondelli–Vardy [**FHMV17, FHMV18, FHMV20**], eventually, showed that large random kernels achieve $2\varrho < 1$ over BECs, breaking the barrier. Guruswami–Riazanov–Ye [**GRY19, GRY20**] extended their result to all BDMCs utilizing the dynamic kernel technique. Our paper [**WD21b**] fills the gap.

Between LDP and CLT is polar coding's MDP behavior. First, Guruswami–Xia [**GX13, GX15**] showed that there exists $\rho > 0$ such that P_{e} scales like $\exp(-N^{0.49})$ while R scales like $I - N^{-\rho}$ over BDMCs. This raised a question about what are the possible pairs (π, ρ) such that (P_{e}, R) scales like $(\exp(-N^{\pi}), I - N^{-\rho})$. Mondelli–Hassani–Urbanke [**MHU16**] answered this, partially, in the same paper they bounded ρ. They showed that under a certain curve connecting $(0, 1/5.714)$ and $(1/2, 0)$ all (π, ρ) are achievable over BDMCs. For BECs the upper left corner is $(0, 1/4.627)$. A straightforward generalization to AWGNCs was also given in [**FT17**]. We in [**WD18b, WD18c**] improved their result, suggesting that via a combinatorial trick the upper left corner of the curve is $(0, \varrho)$ for any ϱ that is

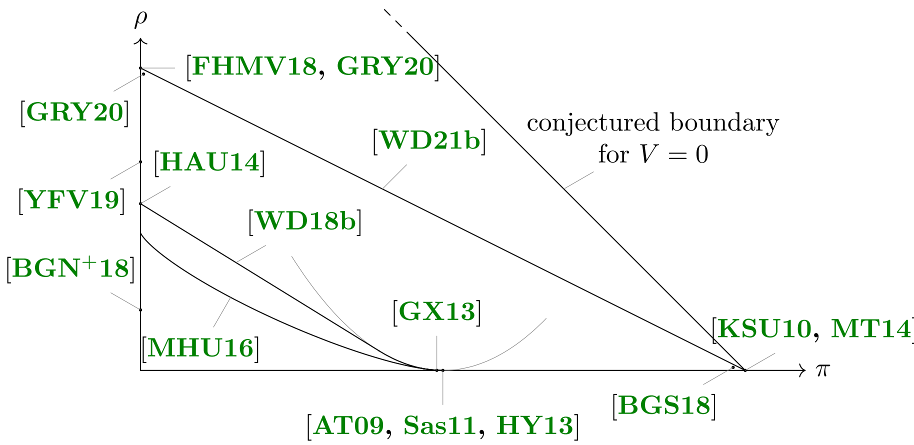

FIGURE 5.4. Recent works on polar coding arranged on a ρ–π plot. Note that results utilizing different kernels over various channels are mixed. The higher ρ, π, the better performance.

valid in the CLT regime. The same trick also implicated that over BECs all (π, ρ) such that $\pi + 2\rho < 1$ are achievable, which is mainly owing to [**FHMV17**]'s result that $2\rho < 1$ over BECs is achievable. Meanwhile, [**BGN$^+$18**] made the first step to investigate the general kernel matrices over general prime-ary channels. They showed that it is possible to achieve $\rho > 0$ with $P_\mathrm{e} \approx N^{-\Omega(1)}$. This is, strictly speaking, "only" a CLT behavior as the desired block error probability in the MDP world is $\exp(-N^\pi)$. Later, Błasiok–Guruswami–Sudan [**BGS18**] were able to show that for all $\pi < \mathsf{E}_\mathrm{c}(G)$ there exists $\rho > 0$ such that (π, ρ) is achievable. This makes it a direct generalization of [**GX13**] to all polarizing kernel matrices G over all prime-ary channels. The preprint [**GRY20**] contains a section that pushes the conference abstract [**GRY19**] to positive π while maintaining $\rho \approx 1/2$. Over the general DMCs, our [**WD21b**] fills the gap.

See Figure 5.4 for a comprehensive plot of all these results.

In Chapter 2, I presented the main contribution of [**WD18b**] which, to be more precise, is an interpolating result $(P_\mathrm{e}, R) \approx (\exp(-N^\pi), C - N^{-\rho})$ for pairs (π, ρ) lying in the region $\mathcal{O}$ that touches $(0, 1/4.714)$ and $(1/2, 0)$. This is the result I want to generalize in this chapter. Id est, I want to characterize the region of π, ρ for any invertible matrix G over any finite field $\mathbb{F}_q$. And this region will be determined by the best $\varrho > 0$ one can find (or believe in) plus the coset distance profile $D_Z^{(j)}$ and $D_S^{(j)}$.

I will do this step-by-step. First, I will show that most kernels enjoy a (very weak) CLT behavior. More precisely, kernels that satisfy a certain ergodic precondition will enjoy an eigen behavior with positive ϱ. After that, we either stick to the weak but provable ϱ or assume a higher ϱ based on experiments, simulations, and/or heuristics. And then we go through an ergodic–eigen–en23–een13–elpin that resembles the eigen–en23–een13–elpin chain in Chapters 2 and 3.

7. An Ergodicity Precondition

Before we board the long proof train eigen–en23–een13–elpin, I want to recall the classification of matrix kernels into two groups. The bad group consists of matrices that do nothing to the channels, and hence get no chance to polarize channels, let alone enjoying any LDP, CLT, or MDP behavior. The good group consists of matrices that can polarize channels in a calculus regard. And then in the next section I will show that all of them enjoy some LDP, CLT, and MDP behaviors. This strengthens the dichotomy even further—a matrix is either not altering the channels at all, or polarizing the channels exponentially fast.

To motivate the classification, recall that channels $W^{(j)}$ are synthesized based on the matrix-multiplication $X_1^\ell = U_1^\ell G$. The kernel $G \in \mathbb{F}_q^{\ell\times\ell}$ is said to be ergodic if it mixes/blends the content U_1^ℓ such that there are nontrivial relations between each U_j and all of Y_1^ℓ. The following two counterexamples demonstrate the necessity of this condition.

Example 5.25. Consider any prime power q and any $\ell \geqslant 2$. If $G \in \mathbb{F}_q^{\ell\times\ell}$ is an upper triangular matrix with 1's on the diagonal, then $X_1^\ell = U_1^\ell G$ is such that $X_j - U_j$ is a linear combination of U_1^{j-1}. Therefore, when we want to guess U_j given $Y_1^\ell U_1^{j-1}$, it suffices to guess X_j based on Y_j and then subtract the correction term $X_j - U_j = U_1^\ell G = U_1^{j-1} 0_j^\ell G$. That is to say, we are facing essentially the same guessing job as if we were facing W; the channel transformation does no benefit to us. In this case, $H(\mathsf{W}_n)$ stays where $H(W)$ is and does not polarize to $\{0, 1\}$. Lesson: G must not be upper triangular. In fact, it can be shown that if two matrices G and $\tilde{G}$ differ by some upper triangular row-operations, i.e., $\tilde{G} = \urcorner G$, then they share the same polarizing ability.

Example 5.26. Consider $q = 4$ and $G = \left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$. Let W be the independent product of BEC(1/3) and BEC(2/3). That is, W takes a pair $(x', x'') \in \mathbb{F}_2^2$ as an input and then outputs

$$\begin{cases} (x', x'') & \text{w.p. } 2/9, \\ (?, x'') & \text{w.p. } 1/9, \\ (x', ?) & \text{w.p. } 4/9, \\ (?, ?) & \text{w.p. } 2/9. \end{cases}$$

Let $c \in \mathbb{F}_4 \setminus \mathbb{F}_2$ be a non-binary element, and identify each $(x', x'') \in \mathbb{F}_2^2$ with $x \coloneqq x' + cx'' \in \mathbb{F}_4$, then W behaves like a channel with input alphabet $\mathbb{F}_4$, and

$$\begin{bmatrix} x_1' & x_2' \end{bmatrix} + c \begin{bmatrix} x_1'' & x_2'' \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 \end{bmatrix} G = \begin{bmatrix} u_1' & u_2' \end{bmatrix} G + c \begin{bmatrix} u_1'' & u_2'' \end{bmatrix} G.$$

In plain English, G multiplies the prime component and the double-prime component separately. Now we attempt to use G to polarize W. Doing that is equivalent to polarizing BEC(1/3) and BEC(2/3) separately. Then, with probability 1/3, the entropy process $\{H(\mathsf{W}_n)\}$ converges to 1/2 because the prime component converges to the entirely noisy channel but the double-prime component converges to the completely reliable channel. Lesson: G needs to bring interaction to the vector space substructure within a prime-power finite field.

The two examples motivate the following definition and theorem for classifying and prejudging matrices.

Definition 5.27. For any invertible matrix $G \in \mathbb{F}_q^{\ell\times\ell}$, the *lowered form* of G is the lower triangular matrix $\tilde{G}$ with 1's on the diagonal and such that $\tilde{G}G^{-1}$ is upper triangular. A matrix G is said to be *ergodic* if the off-diagonal entries of $\tilde{G}$ generate $\mathbb{F}_q$ as an $\mathbb{F}_p$-algebra; or $\mathbb{F}_q = \mathbb{F}_p[\tilde{G}]$ for short.

Theorem 5.28 (Ergodic kernel polarizes)**.** [**MT14**, Theorem 14] *Let W be a q-ary channel. The matrix kernel $G \in \mathbb{F}_q^{\ell\times\ell}$ is ergodic iff $\{H(\mathsf{W}_n)\}$ converges to 0 or 1 almost surely.*

Sketch of the proof in [**MT14**]. The stated theorem is strong and general, handling all edge cases. The majority of its proof involves reducing some general situation (e.g., $\ell \geqslant 2$) to a special case (e.g., $\ell = 2$) to ease notational burdens. I pick out what I think is the key part of the proof.

First, we know that a bounded martingale will almost always converge, which implies $H(\mathsf{W}_n) - H(\mathsf{W}_{n+1}) \to 0$. So all we need to show is that when $H(\mathsf{W}_n)$ starts slowing down, i.e., when $|H(\mathsf{W}_n) - H(\mathsf{W}_{n+1})|$ is small, W_n will be either very reliable or very noisy. To put in another way, our goal is to prove $H(\mathsf{W}_n)$ only slows down when it is reaching 0 or 1. Then we can conclude that the limit of $H(\mathsf{W}_n)$ is either 0 or 1.

To show that W_n is not mediocre when $|H(\mathsf{W}_n) - H(\mathsf{W}_{n+1})|$ is small, the subscript Bhattacharyya parameter $Z_d(W)$ is introduced to inspect the relation between $W(x \mid y)$ and $W(x + d \mid y)$, for any $d \in \mathbb{F}_q^\times$. Now follow the recipe below to show W_n is extreme:

- Show that if $|H(W) - H(W^{(j)})|$ is small, then $|H(W) - H(W^{[c]})|$ is small for some simpler channel transformation $\bullet^{[c]}$, where c is any nonzero entry of $\tilde{G}$.
- Show that if $|H(W) - H(W^{[c]})|$ is small, then $Z_{cd}(W)(1 - Z_d(W))$ is small for all $d \in \mathbb{F}_q^\times$.
- Show that if $Z_d(W)$ is small, then $Z_{cd}(W)$ is small for any nonzero entry c of $\tilde{G}$. Or, if $Z_d(W)$ is close to 1, then $Z_{cd}(W)$ is close to 1 for any nonzero entry c of $\tilde{G}$. This step is the scalar-multiplication part of $\mathbb{F}_p[\tilde{G}]$ as an $\mathbb{F}_p$-algebra.
- Show that if both $Z_c(W)$ and $Z_d(W)$ are small, then $Z_{c+d}(W)$ is small for any entries c and d of $\tilde{G}$. Similarly, if both $Z_c(W)$ and $Z_d(W)$ are close to 1, then $Z_{c+d}(W)$ is close to 1 for any $c, d \in \mathbb{F}_q^\times$. This step is the addition part of $\mathbb{F}_p[\tilde{G}]$ as an $\mathbb{F}_p$-algebra.
- Show that if $\{Z_d(W) : d \in \mathbb{F}_q^\times\}$ are all small, then W is very reliable. Otherwise, if $\{Z_d(W) : d \in \mathbb{F}_q^\times\}$ are all close to 1, then W is very noisy.

The key to the first $\bullet$ is to simplify the given premise $|H(W) - H(W^{(j)})| < \varepsilon$ concerning $\tilde{G}$ to a condition $|H(W) - H(W^{[c]})| < \ell\varepsilon$ concerning some 2×2 matrix, where

$$W^{[c]}(y_1y_2u_1 \mid u_2) \coloneqq \frac{1}{2}W(y_1 \mid u_1 + cu_2)W(y_2 \mid u_2),$$

for any nonzero entry c of $\tilde{G}$. Note that $W^{[c]}$ represents the guessing job of U_2 given $Y_1Y_2U_1$ and the matrix kernel $\left[\begin{smallmatrix}1&0\\c&1\end{smallmatrix}\right]$. To simplify the premise, let (j, i) points to a non-zero entry c of $\tilde{G}$. By how U_i, U_j, X_i, X_j are related by c, one can prove

$$0 \leqslant H(W) - H(W^{[c]}) \leqslant (\ell - j + 1)H(W) - \sum_{k=j}^{\ell} H(W^{(k)}).$$

In layman's terms, the extra amount of information $W^{[c]}$ can steal from W is at most the amount that was stolen by $W^{(j)}, W^{(j+1)}, \ldots, W^{(\ell)}$ from $W^{\ell-j+1}$. Therefore, the difference $|H(\mathsf{W}_n) - H(\mathsf{W}_n^{[c]})|$ converges to 0.

The key to the second and third • is to show $f(Z_{cd}(W)(1-Z_d(W))) \leqslant H(W) - H(W^{[c]})$ for all $d \in \mathbb{F}_q^\times$, where f is some monotonically increasing function that passes $(0,0)$. Once this is done, we have that either $Z_{cd}(\mathsf{W}_n)$ is small or $Z_d(\mathsf{W}_n)$ is close to 1. If small $Z_{cd}(\mathsf{W}_n)$ is the case, then we obtain small $Z_{c^2d}(\mathsf{W}_n)$ when plugging in $d = cd$; we further obtain small $Z_{c^3d}(\mathsf{W}_n)$ when plugging in $d = c^2d$; and so on. A similar argument applies if we choose close-to-1 $Z_d(\mathsf{W}_n)$—we will obtain close-to-1 $Z_{d/c}(\mathsf{W}_n)$, $Z_{d/c^2}(\mathsf{W}_n)$, etc.

The fourth • is by an independent inequality. And it implies that "being small" and "being close to 1" are properties that can propagate among $Z_d(\mathsf{W}_n)$ for distinct d's. The first four •'s together imply that either all of $\{Z_d(W) : d \in \mathbb{F}_q^\times\}$ are small or all of them are close to 1. The last • is just some Hölder tolls that connect the fact that all $Z_d(W)$ are small with the fact that $H(W)$ is small, and vice versa. And the Hölder tolls imply that W_n is extreme. This finishes the sketch of the proof. □

Remark on the theorem: Knowing that $\{H(\mathsf{W}_n)\}$ converges to 0 or 1 says nothing about the pace of convergence. In particular, we do not even know if at least one of $H(W^{(j)})$ is not equal to $H(W)$—it could be that $H(\mathsf{W}_n)$ stays unchanged for many n's and then moves a little bit before another long relaxing. For instance, $1/\lfloor \ln n \rfloor$ eventually converges to 0 but it moves occasionally.

Our next goal in this section is to extract, from the proof of Theorem 5.28, a lemma that some $H(W^{(j)})$ is different from $H(W)$. This lemma will evince that the pace of convergence is exponential in n in the next section.

Lemma 5.29 (Ergodic kernel perturbs). *Let W be a q-ary channel. Let $G \in \mathbb{F}_q^{\ell\times\ell}$ be an ergodic kernel. Then $H(W^{(j)}) \neq H(W)$ for some $1 \leqslant j \leqslant \ell$ unless $H(W) \in \{0,1\}$.*

Proof. First and foremost, assume that W is a symmetric channel with uniform input and that $G = \tilde{G}$ is a lower triangular matrix. The former is due to a symmetrization technique that identifies the conditional entropy of a q-ary channel W and its symmetric sibling $\tilde{W}$. The latter is by that upper triangular row-operations do not alter the synthetic channels up to some equivalence relation. See [MT14] for more details about these two reductions.

Now we assume the opposite of the conclusion, that $H(W^{(j)}) = H(W)$ for all $1 \leqslant j \leqslant \ell$. Then, for any $1 \leqslant j \leqslant \ell$,

$$\begin{aligned}(\ell - j + 1)H(W) &= \sum_{k=j}^{\ell} H(W^{(k)}) = \sum_{k=j}^{\ell} H(U_k \mid Y_1^\ell U_1^{k-1}) = H(U_j^\ell \mid Y_1^\ell U_1^{j-1}) \\ &= \sum_{k=j}^{\ell} H(U_k \mid Y_1^\ell U_1^{j-1} U_{k+1}^\ell) \leqslant \sum_{k=j}^{\ell} H(W) = (\ell - j + 1)H(W).\end{aligned}$$

The equality that starts the second line changes the order we guess U_k—the new order is $k = 1, 2, \ldots, j-1, \ell, \ell-1, \ldots, j$. The inequality in the second line is by that $\tilde{G}$ is lower triangular, and hence guessing U_k is no harder than guessing X_k from Y_k follow by the subtraction of the correction term $X_k - U_k = U_{k+1}^\ell \tilde{G}$. Now the inequality squeezes, so all $H(U_k \mid Y_1^\ell U_1^{j-1} U_{k+1}^\ell)$ are equal to $H(W)$.

Plugging in $k = j$ yields, particularly, $H(U_j \mid Y_1^\ell U_1^{j-1} U_{j+1}^\ell) = H(W)$. That is, the last guessing job, which is supposedly the easiest, turns out to be as difficult as the others. Fix any pair $i < j$ such that the (j, i)th entry of $\tilde{G}$ is $c \neq 0$. Then

$$H(W) = H(U_j \mid Y_1^\ell U_1^{j-1} U_{j+1}^\ell) \leqslant H(U_j \mid Y_i Y_j U_1^{j-1} U_{j+1}^\ell) \leqslant H(W).$$

The first inequality is by monotonicity. The second inequality is by that guessing U_j is no harder than guessing X_j up to the subtraction of some correction term. Now the inequalities squeeze and $H(U_j \mid Y_i Y_j U_1^{j-1} U_{j+1}^\ell) = H(W)$.

Now look closer at the "channel" $U_j \mid Y_i Y_j U_1^{j-1} U_{j+1}^\ell$. We get a second chance to learn more about U_j—since X_i is a linear combination of U_i^ℓ, wherein the coefficient of U_j is c, the output Y_i corresponding to the input X_i also carries information about U_j. The way this information is carried is the same as the outputs corresponding to the inputs U_i and cU_i. This comes down to the second synthetic channel of the 2×2 kernel $\left[\begin{smallmatrix} 1 & 0 \\ c & 1 \end{smallmatrix}\right]$, i.e.

$$W^{[c]}(y_1 y_2 u_1 \mid u_2) \coloneqq \frac{1}{2} W(y_1 \mid u_1 + c u_2) W(y_2 \mid u_2).$$

Hence $H(U_j \mid Y_i Y_j U_1^{j-1} U_{j+1}^\ell) \leqslant H(W^{[c]}) \leqslant H(W)$. The inequalities squeeze again, $H(W^{[c]}) = H(W)$.

Now we cite a difficult inequality in [**MT14**, Appendix A]:

$$H(W) - H(W^{[c]}) \geqslant -\ln\Bigl(1 - \frac{1}{q} \sum_{d \in \mathbb{F}^\times} Z_{cd}(W)^2 (1 - Z_d(W))\Bigr) \geqslant 0.$$

Squeeze one more time; we end up with $Z_{cd}^2(1 - Z_d) = 0$, for all $d \in \mathbb{F}_q^\times$. (We omit the argument "(W)" from now on.) If it is the case that $Z_{cd} = 0$, then the next factor $1 - Z_{cd}$ around the corner is 1, which forces $Z_{c^2 d}$ to be 0. The latter in turns forces $Z_{c^3 d}$ to be 0. This argument propagates throughout the multiplicative orbit $\langle c \rangle d \subseteq \mathbb{F}_q^\times$. Similarly, if it is the case that $Z_d = 1$, then $Z_{d/c}, Z_{d/c^2}, \ldots, Z_d$ are, forcably, all 1. In summary, all Z_d in the orbit $\langle c \rangle d$ share a common fate. In fact, since c can be any nonzero off-diagonal entry of $\tilde{G}$, we deduce that all Z_d in the big orbit $\langle \tilde{G} \rangle d$ share a common fate.

The last piece of the jigsaw puzzle is to show that all Z_d, no matter which orbit they lie in, share the same common fate. We hereby cite [**MT14**, Lemma 21]: If $Z_d = Z_e = 1$ and $d + e \neq 0$, then $Z_{d+e} = 1$. Therefore, if $Z_d = 1$ for some $d \in \mathbb{F}_p$ in the ground field, then $Z_d = 1$ for all $d \in \mathbb{F}_p$. By that off-diagonal entries of $\tilde{G}$ generate $\mathbb{F}_q$ as an $\mathbb{F}_p$-algebra and that Z_d's lying in the same $\tilde{G}$-orbit share the common fate, we conclude that $Z_d = 1$ for all $d \in \mathbb{F}_q$. If, otherwise, no d in the ground field has $Z_d = 1$, then all d have $Z_d = 0$.

Now we have that all $d \in \mathbb{F}_q$ share the common fate—either all $Z_d(W)$ are 1 or all $Z_d(W)$ are 0. So $Z(W)$ is either 1 or 0. Thus $H(W)$ is either 1 or 0. This completes the proof. $\square$

We just see that either G does nothing to the channel W or G will synthesize a $W^{(j)}$ that has a distinct conditional entropy. In other words, $H(\mathsf{W}_1)$ is not a constant but a true random variable. In the next section, I will leverage the fact that $H(\mathsf{W}_1)$ is not constant, regardless how small its variance is, to show that $H(\mathsf{W}_0), H(\mathsf{W}_1), H(\mathsf{W}_2), \ldots$ polarizes in an exponential pace.

8. Eigen Behavior

What constitutes this section is a compactness argument that aims to show that every ergodic kernel has a positive ϱ as in the eigenvalue formula

$$\ell^{-\varrho} \coloneqq \sup_{W\colon q\text{-ary}} \frac{h(H(W^{(1)})) + h(H(W^{(2)})) + \cdots + h(H(W^{(\ell)}))}{\ell h(H(W))}, \tag{5.6}$$

where h is an easy-to-handle eigenfunciton. From that we can conclude $\mathsf{E}[h(\mathsf{W}_n)] < \ell^{-\varrho n}$ and then move on to the en23 behavior.

We can almost see how supremum (5.6) can be estimated: Take a (strictly) concave h. By the last section, $H(W^{(j)})$ are not equal to each other whilst their average is $H(W)$. By Jensen's inequality, $\mathsf{E}[h(H(\mathsf{W}_1))] \leqslant h(H(\mathsf{W}_0))$, and the equality cannot hold. This means that the fraction within the supremum is strictly less than 1. Should the supremum be less than 1, a positive ϱ exists. Now we see why supremum (5.6) is nontrivial: The supremum of some less-than-1 numbers can be 1, especially when the domain is not compact.

The remainder of this section finds a compact subset of q-ary channels, on which the supremum is strictly less than 1, and handles the supremum over the complement set separately.

Lemma 5.30 (Eigen for mediocre channels)**.** *Fix a $G \in \mathbb{F}_q^{\ell\times\ell}$. Fix an $h(z) \coloneqq \sqrt{\min(z, 1-z)}$. For any $\delta > 0$,*

$$\sup_{\delta\leqslant H(W)\leqslant 1-\delta} \frac{h(H(W^{(1)})) + h(H(W^{(2)})) + \cdots + h(H(W^{(\ell)}))}{\ell h(H(W))} < 1, \tag{5.7}$$

where the supremum is taken over all q-ary (Q, W)-pairs whose conditional entropy lies within $[\delta, 1-\delta]$.

Proof. Approach one: All delta–epsilon arguments in Lemma 5.29 (which is inspired by Theorem 5.28) can be made explicit. That will give an upper bound on supremum (5.7).

Approach two: It suffices to show that the space of q-ary channels with mediocre conditional entropy is sequentially compact, plus $\bullet^{(j)}$ and H are continuous w.r.t. the same topology. Once this is done, any sequence $W_1, W_2, \ldots$ whose corresponding fractions converge to 1 must converge to the corresponding fraction of some W_∞, which is strictly less than 1 and leads to a contradiction.

Let the simplex $\Delta^{\mathcal{X}}$ be the closed subset of $[0,1]^{\mathcal{X}}$ constrained by that the sum of coordinates is 1. This is the set of all probability distributions on $\mathcal{X}$. Let $\mathcal{P}(\Delta^{\mathcal{X}})$ be the set of probability distributions on $\Delta^{\mathcal{X}}$. A pair (Q, W) of a DMC W with an input distribution Q corresponds to a distribution on $\Delta^{\mathcal{X}}$, i.e. an element of $\mathcal{P}(\Delta^{\mathcal{X}})$, through the posterior probabilities seen by the decoder. In details, whenever the decoder sees $Y = y$, it looks up the symbol y in the table of posterior probabilities and learns a tuple

$$(W(x_1 \mid y), W(x_2 \mid y), \ldots, W(x_q \mid y)),$$

where $x_1, x_2, \ldots, x_q$ enumerate the symbols of $\mathcal{X}$. This tuple is an element of $\Delta^{\mathcal{X}}$. Now that the channel output Y is a random variable, the tuple of posterior probabilities

$$(W(x_1 \mid Y), W(x_2 \mid Y), \ldots, W(x_q \mid Y))$$

is itself random. This tuple is a $\Delta^{\mathcal{X}}$-valued random variable and obeys some distribution in $\mathcal{P}(\Delta^{\mathcal{X}})$. This distribution of a random tuple is the representative of (Q, W) in $\mathcal{P}(\Delta^{\mathcal{X}})$.

Here comes the measure theory nonsense: Since $\mathcal{X}$ is finite, $[0,1]^{\mathcal{X}}$ and $\Delta^{\mathcal{X}}$ are compact w.r.t. the Euclidean topology. So $\mathcal{P}(\Delta^{\mathcal{X}})$, the set of all distributions on $\Delta^{\mathcal{X}}$, is tight. By Prokhorov's theorem, $\mathcal{P}(\Delta^{\mathcal{X}})$ is sequentially compact w.r.t. the topology of weak convergence. Now notice that $H(W)$ is just the expectation/integral

$$E\Big[\sum_{x\in\mathcal{X}} -W(x \mid Y)\log_q W(x \mid Y)\Big] = \int \sum_{x\in\mathcal{X}} -W(x \mid y)\log_q W(x \mid y)\,\mathrm{d}y.$$

Note that the "integratee" $\sum_{x\in\mathcal{X}} -W(x \mid y)\log_q W(x \mid y) \leqslant q$ is bounded and continuous in the tuple. So H can be extended to a continuous map from $\mathcal{P}(\Delta^{\mathcal{X}})$ to $[0,1]$. By extension I mean that H is now defined over all q-ary input channels regardless of whether the output alphabet is discrete or continuous. Similarly, all $H(W^{(j)})$ can be written as more complicated expectations/integrals of $W(x \mid Y)$. They are all continuous w.r.t. the topology of weak convergence.

Finally, this is what happens if there exists a sequence of channels $\mathsf{W}_1, \mathsf{W}_2, \ldots,$ (notice the font) whose corresponding fractions

$$\frac{h(H(\bullet^{(1)})) + h(H(\bullet^{(2)})) + \cdots + h(H(\bullet^{(\ell)}))}{\ell h(H(\bullet))}$$

converge to 1: Map these channels to $\mathcal{P}(\Delta^{\mathcal{X}})$. A sequence in $\mathcal{P}(\Delta^{\mathcal{X}})$ must contain a convergent subsequence. Let W_∞ be the limit of any such subsequence. Then, as W_∞ also satisfies Lemma 5.29,

$$\frac{h(H(\mathsf{W}_\infty^{(1)})) + h(H(\mathsf{W}_\infty^{(2)})) + \cdots + h(H(\mathsf{W}_\infty^{(\ell)}))}{\ell h(H(\mathsf{W}_\infty))} < 1.$$

(In particular, the denominator is $\geqslant h(\delta) = \sqrt{\delta}$.) A contradiction. This means the supremum is strictly less than 1 and is what I calimed.

This proof is partially inspired by [**Nas18a**, **Nas18b**]. □

The case of mediocre channels is done. It remains to upper bound the supremum for channels in the neighborhoods of $H = 0$ and $H = 1$. First goes the neighborhood of $H = 0$.

Lemma 5.31 (Eigen for reliable channels). *Fix a $G \in \mathbb{F}_q^{\ell\times\ell}$ with coset distance $D_Z^{(j)} \geqslant 5$ for some j, that is, the Hamming distance from some row to the subspace spanned by the rows below is 5 or farther. Fix an $h(z) \coloneqq \sqrt{\min(z, 1-z)}$. Then there exists an δ such that*

$$\sup_{0<H(W)<\delta} \frac{h(H(W^{(1)})) + h(H(W^{(2)})) + \cdots + h(H(W^{(\ell)}))}{\ell h(H(W))} < 1, \tag{5.8}$$

where the supremum is taken over all q-ary (Q, W)-pairs whose conditional entropy lies within the interval $(0, \delta)$.

Proof. Let us assume that it is the last row of G that has Hamming weight 5. For if $\operatorname{hdis}(r_j, R_j) \geqslant 5$ is satisfied with other index $j < \ell$, the following proof works with minor modifications.

Temporarily let δ be $H(W)$ and assume that this is really small. This paragraph bounds $\varepsilon =: H(W^{(\ell)})$ from above: Pay the explicit Hölder toll (Proposition 5.16), we obtain $Z_{\text{mxd}}(W) < q^3\sqrt{H(W)} = q^3\sqrt{\delta}$. Apply FTPC$Z$ (Theorem 5.19), we see that $Z_{\text{mxd}}(W^{(\ell)}) \leqslant f_Z^{(\ell)}(Z_{\text{mxd}}(W)) = Z_{\text{mxd}}(W)^5 < q^{15}\delta^{2.5}$. Pay the Hölder toll for the return-trip (Proposition 5.16), we arrive at $\varepsilon = H(W^{(\ell)}) < q^3\sqrt{Z_{\text{mxd}}(W^{(\ell)})} < q^{10.5}\delta^{1.25}$, or $\varepsilon < q^{10.5}\delta^{1.25}$ for short. This is the only $H(W^{(j)})$ I know how to estimate.

Look at the fraction in supremum (5.8). The last term in the numerator is $h(\varepsilon) \leqslant \sqrt{\varepsilon} < \sqrt{q^{10.5}\delta^{1.25}}$. We do not know about the other terms in the numerator; but at least we can apply Jenson's inequality (i.e., we assume they are all equal to minimize the extent of polarization)

$$\begin{aligned}
&h(H(W^{(1)})) + h(H(W^{(2)})) + \cdots + h(H(W^{(\ell-1)}))\\
&\qquad \leqslant (\ell-1)h\Big(\frac{H(W^{(1)}) + H(W^{(2)}) + \cdots + H(W^{(\ell-1)})}{\ell-1}\Big)\\
&\qquad = (\ell-1)h\Big(\frac{\ell\delta-\varepsilon}{\ell-1}\Big) \leqslant (\ell-1)\sqrt{\frac{\ell\delta-\varepsilon}{\ell-1}} = \sqrt{\ell-1}\sqrt{\ell\delta-\varepsilon} < \sqrt{(\ell-1)\ell\delta}.
\end{aligned}$$

The first equality is by FTPCH (Theorem 5.18).

Now the fraction in supremum (5.8) has its numerator simplified down to two terms:

$$\text{supremum (5.8)} \leqslant \frac{\sqrt{(\ell-1)\ell\delta} + \sqrt{\varepsilon}}{\ell\sqrt{\delta}} = \sqrt{\frac{\ell-1}{\ell}} + \frac{1}{\ell}\sqrt{\frac{\varepsilon}{\delta}} \tag{5.9}$$

If we let $\delta \to 0$, then $\varepsilon/\delta \leqslant q^{10.5}\delta^{1.25}/\delta = q^{10.5}\delta^{0.25} \to 0$. So there is a positive δ such that the right-hand side of inequality (5.9) is strictly less than 1. This completes the proof. $\square$

Next goes the neighborhood of $H = 1$. But it is just the dual of the previous lemma.

Lemma 5.32 (Eigen for noisy channels)**.** *Fix a $G \in \mathbb{F}_q^{\ell\times\ell}$ with dual coset distance $D_S^{(j)} \geqslant 5$ for some j, that is, the Hamming distance from some column of G^{-1} to the subspace spanned by the column to the left is 5 or farther. Fix $h(z) := \sqrt{\min(z, 1-z)}$. Then there exists an δ such that*

$$\sup_{1-\delta<H(W)<1} \frac{h(H(W^{(1)})) + h(H(W^{(2)})) + \cdots + h(H(W^{(\ell)}))}{\ell h(H(W))} < 1,$$

where the supremum is taken over all q-ary (Q, W)-pairs whose conditional entropy lies within the interval $(1-\delta, 1)$.

Proof. The proof is merely the dual of the proof of Lemma 5.31. We shall not repeat. $\square$

Lemmas 5.30 to 5.32 together imply that, if G and G^{-1} have large Hamming distances among their rows and columns, respectively, then supremum (5.6) is strictly less than 1. However, this does not imply anything about matrices with shorter distances, especially when G is 4×4 or smaller. Thankfully, Kronecker product (tensor product) leverages the Hamming distances.

Lemma 5.33 (Thress steps as one big step)**.** *Fix any ergodic matrix $K \in \mathbb{F}_q^{\ell\times\ell}$. Its cubic Kronecker power $G \coloneqq K^{\otimes 3}$ (a) is ergodic, (b) has some $D_Z^{(j)} \geqslant 8$, and (c) has some $D_S^{(j)} \geqslant 8$.*

Proof. For (a): It suffices to consider the Kronecker power of the lowered form $\tilde{K}$. Since the diagonal of $\tilde{K}$ is all 1, any Kronecker power $\tilde{K}^{\otimes n}$ keeps a copy of the original $\tilde{K}$ around the diagonal. Use the original copies to generate $\mathbb{F}_q$ as an $\mathbb{F}_p$-algebra.

For (b): The proof is the combination of two simple facts:

- An ergodic matrix must have some coset distance $\geqslant 2$.
- Let (u, U) be a vector–subspace pair of some ambient vector space $\mathbb{U}$ and (v, V) a vector–subspace pair of another ambient vector space $\mathbb{V}$, then $\operatorname{hdis}(u \otimes v, u \otimes V + U \otimes \mathbb{V}) \geqslant \operatorname{hdis}(u, U)\operatorname{hdis}(v, V)$.

The first bullet point is again a consequence of Theorem 5.28. To elaborate, note that K and and its lowered form $\tilde{K}$ share the same coset distance profile. Then note that $\tilde{K}$ has at least one off-diagonal entry that is nonzero. Let j be the last row with nonzero off-diagonal entry, then $D_Z^{(j)} \geqslant 2$. Hence K has some coset distance $\geqslant 2$. The second bullet point is worth a standalone lemma and will be proved as Lemma 5.34. The two bullet points imply that the $(j\ell^2 + j\ell + j)$th coset distance of $G \coloneqq K^{\otimes 3}$ is at least 8.

For (c): It is the dual of (b). The proof ends here. □

Lemma 5.34. *Let (u, U) be a vector–subspace pair of some ambient vector space $\mathbb{U}$ and (v, V) a vector–subspace pair of another ambient vector space $\mathbb{V}$, then it holds that $\operatorname{hdis}(u \otimes v, u \otimes V + U \otimes \mathbb{V}) \geqslant \operatorname{hdis}(u, U)\operatorname{hdis}(v, V)$.*

Proof. View u, v and $u \otimes v$ as a column vector, a row vector, and a rank-1 matrix, respectively. Now we want to compute the Hamming distance from the matrix $u \otimes v$ to a subset of matrices $u \otimes V + U \otimes \mathbb{V}$. This is equal to the Hamming distance from $u \otimes (v + V)$ to $U \otimes \mathbb{V}$. Let $v' \in v + V$ be any row vector (whose Hamming weight is at least $\operatorname{hdis}(v, V)$). It suffices to prove $\operatorname{hdis}(u \otimes v', U \otimes \mathbb{V}) \geqslant \operatorname{hdis}(u, U)\operatorname{hwt}(v')$.

Without loss of generality, assume that the first coordinate of v' is nonzero. Then the first column of $u \otimes v'$ is a nonzero multiple of u, whilst the first column of any matrix in $U \otimes \mathbb{V}$ is a column vector in U. The number of mismatching entries in the first column is thus $\geqslant \operatorname{hdis}(u, U)$. Repeat this argument for all columns where v' is nonzero, then the total number of mismatching entries is $\geqslant \operatorname{hdis}(u, U)\operatorname{hwt}(v') \geqslant \operatorname{hdis}(u, U)\operatorname{hdis}(v, V)$. This is what we want to show. □

Remark: There is no practical reason to use a large matrix with poor coset distance profile to polarize channels. What Lemma 5.33 is good for is to prove that easy-to-implement 2×2 matrices such as

$$\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$$

polarize channels, given that c generates $\mathbb{F}_q$ over $\mathbb{F}_p$.

Recap of this section: Up to this point, we have seen that an ergodic kernel G will make $H(W_1)$ non-constant, that the eigenvalue is < 1 for mediocre channels, that the eigenvalue is < 1 for reliable channels given $D_Z^{(j)} \geqslant 5$, that the eigenvalue is < 1 for noisy channels given $D_S^{(j)} \geqslant 5$, and that the third Kronecker power

$K^{\otimes 3}$ has sufficient coset distances. These result in $K^{\otimes 3}$ having eigenvalue (i.e., supremum (5.6)) < 1. The next theorem summarizes the eigen behavior of any ergodic kernel.

Theorem 5.35 (From ergodic to eigen)**.** *If $K \in \mathbb{F}_q^{\ell\times\ell}$ is an ergodic matrix, then $G \coloneqq K^{\otimes 3}$ is a kernel such that*

$$\sup_{0<H(W)<1} \frac{h(H(W^{(1)})) + h(H(W^{(2)})) + \cdots + h(H(W^{(\ell^3)}))}{\ell^3 h(H(W))} < 1,$$

where the supremum is taken over all q-ary channels that are not completely noisy or entirely reliable.

The next section will take advantage of the eigen theorem to show that every ergodic kernel enjoys an en23 behavior with positive ϱ.

9. En23 Behavior

Fix a q-ary W and an ergodic $G \in \mathbb{F}_q^{\ell\times\ell}$. The latter could be a large matrix with sufficient coset distances or the cubic power of any ergodic matrix. Let the entropy process $\{H_n\}$ be defined by $H_n \coloneqq H(W_n)$, where $\{W_n\}$ is the channel process grown from W via G. Let the Bhattacharyya process $\{Z_n\}$ be defined by $Z_n \coloneqq Z_{\mathrm{mxd}}(W_n)$, We want to understand the asymptotic behavior of $\{H_n\}$ and $\{Z_n\}$.

To get the asymptotic behavior, Lemma 2.16 provides a template to translate an eigen behavior into an en23 behavior. The only difference is that there, the eigen behavior was given in terms of Z, and here, the eigen behavior is given in terms of H. This turns out to be ineffective to the proof; in fact, an eigen behavior can be given in terms of any parameter that is bi-Hölder to H and Z. Beyond this, there is nothing new to comment on. Let us go straight toward the lemma.

Lemma 5.36 (From eigen to en23)**.** *If a kernel $G \in \mathbb{F}_q^{\ell\times\ell}$ and a concave function h are such that $h(0) = h(1) = 0$ and*

$$\sup_{0<H(W)<1} \frac{h(H(W^{(1)})) + h(H(W^{(2)})) + \cdots + h(H(W^{(\ell)}))}{\ell h(H(W))} = \ell^{-\varrho},$$

then

$$P\big\{Z_n < e^{-n^{2/3}}\big\} > 1 - H(W) - \ell^{-\varrho n + o(n)}.$$

Proof. Telescope $E[h(H_n)] \leqslant E[E[h(H_n) \mid J_1 J_2 \cdots J_{n-1}]] \leqslant E[h(H_{n-1})]\ell^{-\varrho} \leqslant \ell^{-\varrho n}$. So H_n refuses to stay around the middle

$$\begin{aligned} P\big\{e^{-n^{3/4}} \leqslant H_n \leqslant 1 - e^{-n^{3/4}}\big\} &= P\big\{h(H_n) \geqslant h\big(e^{-n^{3/4}}\big)\big\} \\ &\leqslant \frac{E[h(H_n)]}{h(\exp(-n^{3/4}))} \leqslant \frac{h(H_0)\ell^{-\varrho n}}{h(\exp(-n^{3/4}))} < \frac{\ell^{-\varrho n}}{\exp(-n^{3/4})} < \ell^{-\varrho n + o(n)}. \end{aligned}$$

Next, recall that $H_n \to H_\infty \in \{0, 1\}$ and $P\{H_n \to 0\} = 1 - H_0 = 1 - H(W)$. So

$$\begin{aligned} P\big\{H_n < e^{-n^{3/4}}\big\} &\geqslant P\{H_n \to 0\} - P\big\{H_m \to 0 \text{ but } H_n \geqslant e^{-n^{3/4}}\big\} \\ &= 1 - H(W) - P\big\{H_m \text{ will visit } \big[e^{-m^{3/4}}, 1 - e^{-m^{3/4}}\big] \text{ for some } m \geqslant n\big\} \\ &> 1 - H(W) - \sum_{m=n}^{\infty} \ell^{-\varrho m + o(m)} = 1 - H(W) - \ell^{-\varrho n + o(n)}. \end{aligned}$$

Now pay the Hölder toll to translate the inequality into one about Z_n. That is, $P\{Z_n < \exp(-n^{2/3})\} > 1 - H(W) - \ell^{\varrho n - o(n)}$. That is exactly the inequality we want. □

Before the section ends, I want to comment on possible sources of the eigen behavior. As one can see in Chapter 2, the eigen behavior for BDMC and $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$ was given in terms of Z because it is easier to bound Z back then (cf. Lemma 2.12). Here, the eigen behavior is given in terms of H because it forms a martingale, and Jensen's inequality indicates that the eigenvalue is $\leqslant 1$. In the next chapter, readers will see H again for the same reason. In general, eigen behavior can be stated in any parameter such that $\mathsf{YourParameter}_n \leqslant \exp(-n^{3/4})$ implies $Z_n \leqslant \exp(-n^{2/3})$; because then the proof given above works.

In the next section, we will go one step further to the een13 behavior that will be built on top of this section's en23 behavior, which was built on top of last section's ergodic behavior. The proof in the next section assumes the same format as that in Lemma 2.17.

10. Een13 Behavior

The derivation of the een13 behavior from the en23 behavior uses the fact that $Z_{\mathrm{mxd}}(W^{(j)})$ is about $Z_{\mathrm{mxd}}(W)$ to the power of $D_Z^{(j)}$. By that an ergodic kernel has nontrivial coset distance profile, at least one $D_Z^{(j)}$ will square the $Z_{\mathrm{mxd}}(W)$. Now it is a matter of Hoeffding's inequality to control the frequency that a trajectory of $\{Z_n\}$ undergoes squaring.

To facilitate the reasoning about coset distances, define the distance process $\{D_n\}$ via $D_n \coloneqq D_Z^{(J_n)}$. Then, by FTPC$Z$, $Z_{n+1} \leqslant q^{\ell - J_{n+1}} Z_n^{D_{n+1}} \leqslant q^\ell Z_n^{D_{n+1}} \approx Z_n^{D_{n+1}}$. And now we can telescope and talk about the product $D_1 D_2 \cdots D_n$. For instance, the product is 1 with probability $P\{D_1 = 1\}^n$. This is the probability that Z_0–Z_n do not undergo any squaring or higher powering, and we have no control on such Z_n (except the trivial one $Z_n \leqslant q^{(\ell-1)^n} Z_0$). Hence, I hope that $P\{D_1 = 1\}^n$ is dominated by the desired gap to capacity $\ell^{-\varrho n}$. Equivalently, I hope that $P\{D_1 = 1\} \leqslant \ell^{-\varrho}$.

Conjecture 5.37 (Good D implies good ϱ)**.** *For any kernel $G \in \mathbb{F}_q^{\ell \times \ell}$,*

$$P\{D_1 = 1\} < \ell^{-\varrho}.$$

Even if $P\{D_1 = 1\} \geqslant \ell^{-\varrho}$, we can still re-choose a lower $\varrho > 0$ to satisfy $P\{D_1 = 1\} < \ell^{-\varrho}$. This is one of the two new issues, in contrast to Chapter 2, that we need to take care of in this section. The other new issue is that $\{Z_n\}$ is no longer a supermartingale.

Lemma 5.38 (Artificial supermartingale)**.** *For any $\varepsilon > 0$, there exist a smaller $\varepsilon > 0$ and a small $\delta > 0$ such that $\{Z_n^\varepsilon \wedge \delta\}$ is a supermartingale. Here $Z_n^\varepsilon \wedge \delta$ is a shorthand for* $\min(Z_n^\varepsilon, \delta)$.

Proof. It suffices to pick ε and δ such that $E[Z_1^\varepsilon \wedge \delta] \leqslant Z_0^\varepsilon \wedge \delta$ for any W. If $Z_0^\varepsilon \geqslant \delta$, there is nothing to prove; thus we may assume that $Z_0^\varepsilon < \delta$. Start from $P\{D_1 = 1\} < \ell^{-\varrho} < 1$. Then either $q^{\ell\varepsilon} P\{D_1 = 1\} < 1$ or we can reselect a smaller $\varepsilon > 0$ to make it true. We then choose a small $\delta > 0$ such that $q^{\ell\varepsilon} \times P\{D_1 = 1\} + q^{\ell\varepsilon}\delta P\{D_1 \geqslant 2\} \leqslant 1$. Now upper bound the conditional expectation by handling the two cases:

$$E[Z_1^\varepsilon \wedge \delta] \leqslant (q^\ell Z_0)^\varepsilon P\{D_1 = 1\} + (q^\ell Z_0^2)^\varepsilon P\{D_1 \geqslant 2\}$$

$$= Z_0^\varepsilon(q^{\ell\varepsilon} P\{D_1 = 1\} + q^{\ell\varepsilon} Z_0^\varepsilon P\{D_1 \geqslant 2\}) \leqslant Z_0^\varepsilon \cdot 1 = Z_0^\varepsilon \wedge \delta.$$

This finishes the proof of $E[Z_1^\varepsilon \wedge \delta] \leqslant Z_0$. By the tree structure of the process, $E[Z_{n+1} \mid J_1 J_2 \cdots J_n]$ versus Z_n is just $E[Z_1]$ vs Z_0 with $W \leftarrow W_n$. This completes the proof of the whole lemma. □

Remark: The choice of δ in the lemma implies that $q^{\ell\varepsilon}\delta$ is less than $(1 - q^{\ell\varepsilon} P\{D_1 = 1\})/P\{D_1 \geqslant 2\} < 1$, or equivalently $q^{\ell\varepsilon} \leqslant \delta^{-1}$. Thus $Z_{n+1} \leqslant q^\ell Z_n^{D_{n+1}} \leqslant \delta^{-1/\varepsilon} Z_n^{D_{n+1}} \leqslant Z_n^{D_{n+1}-\varepsilon}$ whenever $Z_n^\varepsilon < \delta$. In other words, when we look at Z_n^ε in the "safe zone" $[0, \delta]$, not only is Z_n^ε a supermartingale, but FTPCZ also takes a simpler form $Z_{n+1} \leqslant Z_n^{D_{n+1}-\varepsilon} \leqslant Z_n^{D_{n+1}(1-\varepsilon)}$ that facilitates telescoping.

The main statement of the een13 behavior follows.

Lemma 5.39 (From en23 to een13)**.** *Given $P\{D_1 = 1\} < \ell^{-\varrho}$ and Lemma 5.36, that is, given*

$$P\{Z_n < e^{-n^{2/3}}\} > 1 - H(W) - \ell^{-\varrho n + o(n)},$$

we have

$$P\{Z_n < \exp(-e^{n^{1/3}})\} > 1 - H(W) - \ell^{-\varrho n + o(n)}. \tag{5.10}$$

Proof. (Select constants.) Since $P\{D_1 = 1\} < \ell^{-\varrho}$, there exists a large $\lambda > 1$ such that $E[D_1^{-\lambda}] < \ell^{-\varrho}$. Pick a small $\varepsilon > 0$ such that $E[D_1^{-\lambda}]8^{\varepsilon\lambda} < \ell^{-\varrho}$. Invoke Lemma 5.38; pick a smaller $\varepsilon > 0$ and a small $\delta > 0$ such that $\{Z_n^\varepsilon \wedge \delta\}$ is a supermartingale. According to the proof of Lemma 5.38 (and the remark underneath), this δ is such that $Z_{n+1} \leqslant Z^{D_{n+1}-\varepsilon}$ whenever $Z_n^\varepsilon < \delta$.

(Define events.) Consider only perfect square n. (If it is not the case, follow the workaround in Lemma 2.16.) Let E_0^0 be the empty event. For every $m = \sqrt{n}, 2\sqrt{n}, \ldots, n - \sqrt{n}$, we define five series of events A_m, B_m, C_m, E_m, and E_0^m inductively as below: Let A_m be $\{Z_m < \exp(-m^{2/3})\} \setminus E_0^{m-\sqrt{n}}$. Let B_m be a subevent of A_m where $Z_l^\varepsilon \geqslant \delta$ for some $l \geqslant m$. Let C_m a subevent of A_m where

$$D_{m+1} D_{m+2} \cdots D_{m+\sqrt{n}} \leqslant 8^{\varepsilon\sqrt{n}}. \tag{5.11}$$

Let E_m be $A_m \setminus (B_m \cup C_m)$. Let E_0^m be $E_0^{m-\sqrt{n}} \cup E_m$. Let a_m, b_m, c_m, e_m, and e_0^m be the probability measures of the corresponding capital letter events. Moreover, let g_m be $1 - H(W) - e_0^m$.

(Bound b_m/a_m from above.) Conditioning on A_m, I want to estimate the probability that $Z_l^\varepsilon \geqslant \delta$ for some $l \geqslant m$. Recall that $\{Z_l^\varepsilon \wedge \delta\}$ is made a supermartingale. By Ville's inequality, $P\{Z_l^\varepsilon \geqslant \delta$ for some $l \geqslant m \mid A_m\} \leqslant Z_m^\varepsilon/\delta < \exp(-m^{2/3}\varepsilon)/\delta$. This is an upper bound on b_m/a_m and will be summoned in inequality (5.12).

(Bound c_m/a_m from above.) I am to estimate how frequently inequality (5.11) happens. It is the probability of $(D_{m+1} D_{m+2} \cdots D_{m+\sqrt{n}})^{-\lambda} \geqslant 8^{-\varepsilon\lambda\sqrt{n}}$. This probability must not exceed $E[(D_{m+1} D_{m+2} \cdots D_{m+\sqrt{n}})^{-\lambda}]8^{\varepsilon\lambda\sqrt{n}} = E[D_1^{-\lambda}]^{\sqrt{n}} 8^{\varepsilon\lambda\sqrt{n}} = (E[D_1^{-\lambda}]8^{\varepsilon\lambda})^{\sqrt{n}} < \ell^{-\varrho\sqrt{n}}$ by Markov's inequality. This is an upper bound on c_m/a_m and will be summoned in inequality (5.12).

(Bound $(g_{m-\sqrt{n}} - a_m)^+$ from above.) By definitions, $g_{m-\sqrt{n}} - a_m = 1 - H(W) - (e_0^{m-\sqrt{n}} + a_m)$. The definition of A_m forces it to be disjoint from $E_0^{m-\sqrt{n}}$, thus $e_0^{m-\sqrt{n}} + a_m$ is the probability measure of $E_0^{m-\sqrt{n}} \cup A_m$. This union event must contain the event $\{Z_m < \exp(-m^{2/3})\}$ by how A_m was defined. Recall the en23 behavior $P\{Z_m < \exp(-m^{2/3})\} > 1 - H(W) - \ell^{-\varrho m + o(m)}$. Chaining all inequalities together, we deduce $g_{m-\sqrt{n}} - a_m < \ell^{-\varrho m+o(m)}$. Let $(g_{m-\sqrt{n}} - a_m)^+$ be $\max(0, g_{m-\sqrt{n}} - a_m)$ so we can write $(g_{m-\sqrt{n}} - a_m)^+ < \ell^{-\varrho m+o(m)}$. This upper bound will be summoned in inequality (5.12).

(Bound $e_0^{n-\sqrt{n}}$ from below.) We start rewriting g_m with g_m^+ being $\max(0, g_m)$:

$$\begin{aligned}
g_m &= 1 - H(W) - e_0^m = 1 - H(W) - (e_0^{m-\sqrt{n}} + e_m) \\
&= 1 - H(W) - e_0^{m-\sqrt{n}} - e_m = g_{m-\sqrt{n}} - e_m \\
&= g_{m-\sqrt{n}}\left(1 - \frac{e_m}{a_m}\right) + \frac{e_m}{a_m}(g_{m-\sqrt{n}} - a_m) \\
&\leqslant g_{m-\sqrt{n}}^+\left(1 - \frac{e_m}{a_m}\right) + \frac{e_m}{a_m}(g_{m-\sqrt{n}} - a_m)^+ \\
&\leqslant g_{m-\sqrt{n}}^+\left(1 - \frac{e_m}{a_m}\right) + (g_{m-\sqrt{n}} - a_m)^+ \\
&\leqslant g_{m-\sqrt{n}}^+\left(\frac{b_m}{a_m} + \frac{c_m}{a_m}\right) + (g_{m-\sqrt{n}} - a_m)^+ \\
&< g_{m-\sqrt{n}}^+\left(e^{-m^{2/3}\varepsilon}/\delta + \ell^{-\varrho\sqrt{n}}\right) + \ell^{-\varrho m+o(m)}. \qquad (5.12)
\end{aligned}$$

The first four equalities are by the definitions of g_m and E_0^m. The next equality is by elementary algebra. The next two inequalities are by $0 \leqslant e_m/a_m \leqslant 1$. The next inequality is by the definition of E_m. The last inequality summons upper bounds derived in the last three paragraphs. The last line contains two terms in the big parentheses; between them, $\ell^{-\varrho\sqrt{n}}$ dominates $\exp(-m^{2/3}\varepsilon)/\delta$ once m is greater than $O(n^{3/4})$. Subsequently, we obtain this recurrence relation:

$$\begin{cases} g_{O(n^{3/4})} \leqslant 1, \\ g_m \leqslant 2g_{m-\sqrt{n}}^+ \ell^{-\varrho\sqrt{n}} + \ell^{-\varrho m+o(m)}. \end{cases}$$

Solve it (cf. the master theorem); we get that $g_{n-\sqrt{n}} < \ell^{-\varrho n+o(n)}$. By the definition of $g_{n-\sqrt{n}}$, we immediately get $e_0^{n-\sqrt{n}} > 1 - H(W) - \ell^{-\varrho n+o(n)}$.

(Analyze $E_0^{n-\sqrt{n}}$.) We want to estimate H_n when $E_0^{n-\sqrt{n}}$ happens. To be precise, for each $m = \sqrt{n}, 2\sqrt{n}, \ldots, n-\sqrt{n}$, we attempt to bound $Z_{m+\sqrt{n}}$ when E_m happens. Fix an m. When E_m happens, its superevent A_m happens, so we know that $Z_m < \exp(-m^{2/3})$. But B_m does not happen, so $Z_l^\varepsilon < \delta$ for all $l \geqslant m$. This implies that $Z_{l+1} \leqslant Z_l^{D_{l+1}(1-\varepsilon)}$ for those l. Telescope; $Z_{m+\sqrt{n}}$ is less than or equal to Z_m raised to the power of $D_{m+1}D_{m+2}\cdots D_{m+\sqrt{n}}(1-\varepsilon)^{\sqrt{n}}$. But C_m does not happen, so the product is greater than $8^{\varepsilon\sqrt{n}}(1-\varepsilon)^{\sqrt{n}} = (8\sqrt[\varepsilon]{1-\varepsilon})^{\varepsilon\sqrt{n}}$. The latter is greater than $2^{\varepsilon\sqrt{n}}$ granted that $\varepsilon < 1/2$. Jointly we have $Z_{m+\sqrt{n}} \leqslant Z_m^{2^{\varepsilon\sqrt{n}}} < \exp(-m^{2/3}2^{\varepsilon\sqrt{n}})$. Recall that $Z_{l+1} \leqslant q^\ell Z_l$ for all $l \geqslant m+\sqrt{n}$. Then telescope again; $Z_n \leqslant q^{\ell(n-m-\sqrt{n})}Z_{m+\sqrt{n}} < q^{\ell n}\exp(-m^{2/3}2^{\varepsilon\sqrt{n}}) < \exp\left(-e^{n^{1/3}}\right)$ provided that n is sufficiently large. In other words, $E_0^{n-\sqrt{n}}$ implies $Z_n < \exp\left(-e^{n^{1/3}}\right)$.

(Summary.) Now we may conclude $P\left\{Z_n < \exp\left(-e^{n^{1/3}}\right)\right\} \geqslant P(E_0^{n-\sqrt{n}}) = e_0^n > 1 - H(W) - \ell^{-\varrho n+o(n)}$. And hence the proof of the een13 behavior is sound. □

In the next section, I will show the elpin behavior of $\{Z_n\}$. That will imply the MDP behavior of polar coding and the latter will specialize to the LDP and CLT behaviors of polar coding. In contract to this section, where we used $P\{D_n = 1\}$

and $P\{D_n \geqslant 2\}$ to determine the behavior of $\{Z_n\}$, we will have to use the "full power" of $\{D_n\}$ in what follows.

11. Elpin Behavior

As the een13 behavior lowers Z_m to the order of $\exp\bigl(-e^{m^{1/3}}\bigr)$, we can afford telescoping $Z_{l+1} \leqslant Z_l^{D_{l+1}-\varepsilon}$ for more l's without having to worry about $Z_l^\varepsilon \geqslant \delta$. So the trade-off between the block error probability ($\approx \sum Z_n$) and the code rate ($\approx P\{Z_n < \cdots\}$) amounts to the behavior of the product $D_{m+1}D_{m+2}\cdots D_n$ or, as probability theorists may prefer, the behavior of the i.i.d. sum $\log_\ell D_{m+1} + \log_\ell D_{m+2} + \cdots + \log_\ell D_n$.

As section 6 suggests, the limiting behavior of i.i.d. sums is well known. In particular, I will borrow the LDP result that concerns the limiting behavior of $P\{\bar{X}_n < x\}$ when x is fixed and n increases. First goes some definitions.

Definition 5.40. The *cumulant generating function* of $\log_\ell D_1$ is defined to be the logarithm of its moment generating function or, more precisely,

$$K(t) := \log_\ell(E[D_1^t]) = \frac{1}{\ln \ell} \ln \sum_{j=1}^{\ell} \frac{D_1^t}{\ell}.$$

Definition 5.41. The *Cramér function* of $\log_\ell D_1$ is defined to be the convex conjugate of the cumulant generating function or, more precisely,

$$L(s) := \sup_{t \leqslant 0} st - K(t).$$

$L(s)$ controls the limiting law of the sum of $\log_\ell D_l$ or the product of D_l. An example (which we have seen) is that, when $D_l \in \{1, 2\}$, its logarithm is just a coin toss, and $1 - h_2$ controls the limiting law.

Theorem 5.42 (Cramér's theorem)**.** *For any n,*

$$P\{D_1 D_2 \cdots D_n \leqslant \ell^{sn}\} \leqslant \ell^{-nL(s)}.$$

Furthermore, $L(s)$ is the greatest value satisfying that, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \log_\ell P\{D_1 D_2 \cdots D_n < \ell^{sn}\} = -L(s).$$

Proof. What I actually need in the sequel is the upper bound on P. The fact that $L(s)$ gives the best upper bound is irrelevant to the validity of the theorems in this chapter, but indicates that the theorems are somewhat optimal. The proof of the latter is thus omitted; see standard textbooks instead (e.g., [**DZ10**]).

To prove the upper bound, recall that $L(s)$ is a supremum. This supremum, by some compactness argument, turns out to be a maximum. Let t be the argument that maximizes $st - K(t)$. Then $D_1 D_2 \cdots D_n \leqslant \ell^{sn}$ is equivalent to $(D_1 D_2 \cdots D_n)^t \geqslant \ell^{stn}$. Apply Markov's inequality, the probability that $D_1^t D_2^t \cdots D_n^t \geqslant \ell^{stn}$ is at most $E[D_1^t D_2^t \cdots D_n^t]\ell^{-stn} = (E[D_1^t]\ell^{-st})^n = \ell^{(K(t)-st)n} = \ell^{-nL(s)}$ ∎

Now we define the achievable region in terms of $L(s)$; it ought to be the set of (π, ρ) pairs such that $P\{Z_n < \exp(-\ell^{\pi n})\} > 1 - H(W) + \ell^{-\rho n}$. Consider ϖ being the mean of $\log_\ell D_1$. We know that, most of the times, i.i.d. averages of $\log_\ell D_l$ concentrate around the mean so we should not expect that $\pi > \varpi$. Similarly, we do not expect that $\rho > \varrho$. The remaining characterization of (π, ρ) is in terms of $L(s)$.

Definition 5.43. Let $\mathcal{O} \subseteq [0, \varpi] \times [0, \varrho]$ be an open region defined by the following criterion: for any $(\pi, \rho) \in \mathcal{O}$, the ray shooting from (π, ρ) toward the opposite direction of $(0, \varrho)$ does not intersect the function graph of $\rho = \mathsf{L}(\pi)$.

Note that the criterion is equivalent to, geometrically, that (π, ρ) lies to the left of the convex envelope of $(0, \varrho)$ and $\mathsf{L}(s)$. So is it equivalent to

$$\mathsf{L}\Big(\frac{\pi n}{n-m}\Big) > \frac{\rho n - \varrho m}{n-m} \tag{5.13}$$

for all $0 < n < m$. The main theorem of this chapter can now be stated and proved.

Theorem 5.44 (From een13 to elpin)**.** *Fix a pair $(\pi, \rho) \in \mathcal{O}$. Given Lemma 5.39, that is, given*

$$\mathsf{P}\big\{\mathsf{T}_n < \exp\big(-e^{n^{1/3}}\big)\big\} > H(W) - \ell^{-\varrho n + o(n)},$$

we have

$$\mathsf{P}\big\{\mathsf{T}_n < e^{-\ell^{\pi n}}\big\} > H(W) - \ell^{-\rho n + o(n)}.$$

Proof. (Select constants.) Since inequality (5.13) holds, there exists a small constant $\varepsilon > 0$ such that

$$\mathsf{L}\Big(\frac{\pi n}{n-m} + 2\varepsilon\Big) > \frac{\rho n - \varrho m}{n-m} \tag{5.14}$$

by the compactness argument. Pass this ε to Lemma 5.38: There exists a smaller $\varepsilon > 0$ and a small $\delta > 0$ such that $\mathsf{Z}_n^\varepsilon \wedge \delta$ is a supermartingale and $\mathsf{Z}_{n+1} \leqslant \mathsf{Z}_n^{\mathsf{D}_{n+1}(1-\varepsilon)}$ whenever $\mathsf{Z}_n < \delta$.

(Define events.) Let n be a perfect square. (If it is not the case, see how Theorem 2.18 bypasses.) Let A_0^0 and E_0^0 be the empty event. For every $m = \sqrt{n}, 2\sqrt{n}, \ldots, n - \sqrt{n}$, we define six series of events A_m, A_0^m, B_m, C_m, E_m, and E_0^m inductively as follows: Let A_m be $\big\{\mathsf{Z}_m < \exp(-e^{m^{1/3}})\big\} \setminus \mathsf{A}_0^{m-\sqrt{n}}$. Let A_0^m be $\mathsf{A}_0^{m-\sqrt{n}} \cup \mathsf{A}_m$. Let B_m be a subevent of A_m where $\mathsf{Z}_l^\varepsilon \geqslant \delta$ for some $l \geqslant m$. Let C_m a subevent of A_m where

$$\mathsf{D}_{m+1}\mathsf{D}_{m+2}\cdots\mathsf{D}_n \leqslant \ell^{\pi n + 2\varepsilon(n-m)}. \tag{5.15}$$

Let E_m be $\mathsf{A}_m \backslash (\mathsf{B}_m \cup \mathsf{C}_m)$. Let E_0^m be $\mathsf{E}_0^{m-\sqrt{n}} \cup \mathsf{E}_m$. Let a_m, a_0^m, b_m, c_m, e_m, and e_0^m be the probability measures of the corresponding capital letter events. Moreover, let f_m be $1 - H(W) - \mathsf{a}_0^m$ and let g_m be $1 - H(W) - \mathsf{e}_0^m$.

(Bound $\mathsf{b}_m/\mathsf{a}_m$ from above.) Conditioning on A_m, we want to estimate the probability that $\mathsf{Z}_l \geqslant \delta$ for some $l \geqslant m$. Recall that Z_l is a supermartingale. Hence by Ville's inequality, $\mathsf{P}\{\mathsf{Z}_l^\varepsilon \geqslant \delta \text{ for some } l \geqslant m \mid \mathsf{A}_m\} \leqslant \mathsf{Z}_m^\varepsilon/\delta < \exp\big(-e^{m^{1/3}}\varepsilon\big)/\delta$. This is an upper bound on $\mathsf{b}_m/\mathsf{a}_m$ and will be summoned in inequality (5.16).

(Bound $\mathsf{c}_m/\mathsf{a}_m$ from above.) We want to estimate how often inequality (5.15) happens. This is equivalent to asking how often do $n - m$ fair coin tosses end up with $\pi n + 2\varepsilon(n-m)$ heads. By Theorem 5.42, this probability is less than ℓ to the power of

$$-(n-m)\mathsf{L}\Big(\frac{\pi n}{n-m} + 2\varepsilon\Big).$$

By inequality (5.14), this exponent is less than $\varrho m - \rho n$. Thus, the probability is less than $\ell^{\varrho m - \rho n}$. This is an upper bound on $\mathsf{c}_m/\mathsf{a}_m$ and will be summoned in inequality (5.16).

(Bound f_m^+ from above.) The definition of f_m reads $1 - H(W) - \mathsf{a}_0^m$. Here a_0^m is the probability measure of A_0^m, and A_0^m is a superevent of A_m by how the former is defined. Event A_0^m must contain $\{Z_m < \exp(-e^{m^{1/3}})\}$ by how A_m was defined. By the een13 behavior, $P\{Z_m < \exp(-e^{m^{1/3}})\} > 1 - H(W) - \ell^{-\varrho m + o(m)}$. Chaining all inequalities together, we infer that $f_m < \ell^{-\varrho m + o(m)}$. Let f_m^+ be $\max(0, f_m)$ so we can write $f_m^+ < \ell^{-\varrho m + o(m)}$. This upper bound will be summoned in inequality (5.16).

(Bound $e_0^{n-\sqrt{n}}$ from below.) We start rewriting $g_m - f_m^+$ with $(f_{m-\sqrt{n}} - a_m)^+$ being $\max(0, f_{m-\sqrt{n}} - a_m)$:

$$\begin{aligned}
g_m - f_m^+ &= 1 - H(W) - e_0^m - (1 - H(W) - a_0^m)^+ \\
&= 1 - H(W) - e_0^{m-\sqrt{n}} - e_m - (1 - H(W) - a_0^{m-\sqrt{n}} - a_m)^+ \\
&= g_{m-\sqrt{n}} - e_m - (f_{m-\sqrt{n}} - a_m)^+ \\
&\leqslant g_{m-\sqrt{n}} - e_m - \frac{e_m}{a_m}(f_{m-\sqrt{n}} - a_m)^+ \\
&\leqslant g_{m-\sqrt{n}} - e_m - \frac{e_m}{a_m}(f_{m-\sqrt{n}}^+ - a_m) \\
&= g_{m-\sqrt{n}} - f_{m-\sqrt{n}}^+ + f_{m-\sqrt{n}}^+\Big(1 - \frac{e_m}{a_m}\Big) \\
&\leqslant g_{m-\sqrt{n}} - f_{m-\sqrt{n}}^+ + f_{m-\sqrt{n}}^+\Big(\frac{b_m}{a_m} + \frac{c_m}{a_m}\Big) \\
&< g_{m-\sqrt{n}} - f_{m-\sqrt{n}}^+ + \ell^{-\varrho(m-\sqrt{n}) + o(m-\sqrt{n})}\Big(\exp\big(-e^{m^{1/3}}\varepsilon\big)/\delta + \ell^{\varrho m - \rho n}\Big). \qquad (5.16)
\end{aligned}$$

The first three equalities are by the definitions of g_m and f_m. The next inequality is by $0 \leqslant e_m/a_m \leqslant 1$. The next inequality is by $\max(0, f - a) = \max(a, f) - a \geqslant \max(0, f) - a$. The next equality is by elementary algebra. The next inequality is by the definition of E_m. The last inequality summons upper bounds derived in the last three paragraphs. Now the last line contains two terms in the big parentheses; between them, $\ell^{\varrho m - \rho n}$ dominates $\exp(-e^{m^{1/3}}\varepsilon)/\delta$ once $n \to \infty$. Subsequently, we obtain this recurrence relation

$$\begin{cases} g_0 - f_0^+ = 0; \\ g_m - f_m^+ \leqslant g_{m-\sqrt{n}} - f_{m-\sqrt{n}}^+ + 2\ell^{-\rho n + o(n)}. \end{cases}$$

Solve it (cf. the Cesàro summation); we get that $g_{n-\sqrt{n}} - f_{n-\sqrt{n}}^+ < \ell^{-\rho n + o(n)}$. Once again we summon $f_{n-\sqrt{n}}^+ < \ell^{-\varrho(n-\sqrt{n}) + o(n-\sqrt{n})} < \ell^{-\varrho n + o(n)}$; therefore $g_{n-\sqrt{n}} < \ell^{-\rho n + o(n)}$. Based on the definition of $g_{n-\sqrt{n}}$ we immediately get $e_0^{n-\sqrt{n}} > 1 - H(W) - \ell^{-\rho n + o(n)}$.

(Analyze $E_0^{n-\sqrt{n}}$.) We want to estimate Z_n when $E_0^{n-\sqrt{n}}$ happens. To be precise, for each $m = \sqrt{n}, 2\sqrt{n}, \ldots, n - \sqrt{n}$, we attempt to bound Z_n when E_m happens . Fix an m. When E_m happens, its superevent A_m happens, so we know that $Z_m < \exp(-e^{m^{1/3}})$. But B_m does not happen, so $Z_l < \delta$ for all $l \geqslant m$. This implies $Z_{l+1} \leqslant Z_l^{D_{l+1}(1-\varepsilon)}$ for those l. Telescope; Z_n is less than or equal to Z_m raised to the power of $D_{m+1}D_{m+2}\cdots D_n(1-\varepsilon)^{n-m}$. But C_m does not happen, so the product is greater than $\ell^{\pi n + 2\varepsilon(n-m)}(1-\varepsilon)^{n-m}$, which is greater than $\ell^{\pi n}$ granted that $\varepsilon < 1/2$. Jointly we have $Z_n \leqslant Z_m^{\ell^{\pi n}} < \exp(-e^{m^{1/3}}\ell^{\pi n}) < \exp(-\ell^{\pi n})$. In other words, $E_0^{n-\sqrt{n}}$ implies $Z_n < \exp(-\ell^{\pi n})$.

(Summary.) Now we may conclude $P\{Z_n < \exp(-\ell^{\pi n})\} \geqslant P(E_0^{n-\sqrt{n}}) = e_0^n > 1 - H(W) - \ell^{-\rho n + o(n)}$. And hence the proof of the elpin behavior is sound. $\square$

12. Consequences, Dual Picture Included

That Z_n can be proved low implies that we have good codes over symmetric channels of prime-power input alphabet. To code over the most general DMCs, we need to talk about $\{S_n\}$—it is the process of $S_{\max}(W_n)$ that controls the behavior of W_n when it becomes noisy.

$S_{\max}$ (and $\{S_n\}$) really is the dual counterpart of Z_{mxd} (and $\{Z_n\}$). The Hölder tolls we pay to translate between H and Z_{mxd} is the same amount as the tolls we pay to translate between $1 - H$ and $S_{\max}$. Plus, FTPCZ (Theorem 5.19) and FTPCS (Theorem 5.20) state almost the same phenomenon in terms of Z_{mxd} and $S_{\max}$, respectively. By the process nonsense that was once used to show Theorem 3.13 out of Theorem 2.18, Theorem 5.44 assumes an S-version.

Theorem 5.45 (From eigen to elpin)**.** *Fix a pair (π, ρ) lying to the left of the convex envelope of $(0, \varrho)$ and the Cramér function of $\log_\ell D_S^{(J_1)}$. Given the premise of Lemma 5.36, that is, given*

$$\sup_{0<H(W)<1} \frac{h(H(W^{(1)})) + h(H(W^{(2)})) + \cdots + h(H(W^{(\ell)}))}{\ell h(H(W))} = \ell^{-\varrho n},$$

then

$$P\{S_n < e^{-\ell^{\pi n}}\} > H(W) - \ell^{-\rho n + o(n)}.$$

This theorem together with Theorem 5.44 implies good codes over all DMCs. This is because the block error probability is bounded from above by the sum of small $P_{\mathrm{e}}(W_n)$ and small $T(W_n)$, and they are further bounded by Z_n and S_n.

Corollary 5.46 (Good code for DMC)**.** *Fix a q-ary DMC. Fix an ergodic $G \in \mathbb{F}_q^{\ell \times \ell}$ with a positive $\varrho > 0$. Fix a pair (π, ρ) lying to the left of (a) the convex envelope of $(0, \varrho)$ and the Cramér function of $\log_\ell D_Z^{(J_1)}$ and (b) the convex envelope of $(0, \varrho)$ and the Cramér function of $\log_\ell D_S^{(J_1)}$. Then polar coding with kernel G enjoys block error probability $\exp(-\ell^{\pi n})$ and code rate $N^{-\rho}$ less than the channel capacity.*

The same can be stated for lossless and lossy compression. For lossless compression, if the source alphabet $\mathcal{X}$ is not a prime power, add dummy symbols till it is. Then the block error probability is the sum of small $P_{\mathrm{e}}(W_n)$. For lossy compression, if the palette $\mathcal{X}$ in the distortion function $\mathrm{dist}\colon \mathcal{X} \times \mathcal{Y} \to [0, 1]$ is not prime power, add dummy symbols and penalize dummy symbols with distortion 1 (i.e., $\mathrm{dist}(x, y) = 1$ if x is dummy). Then the excess of distortion is the average of small $T(W_n)$.

Corollary 5.47 (Good code for lossless compression)**.** *Fix a q-ary source X and side information Y. Fix an ergodic $G \in \mathbb{F}_q^{\ell \times \ell}$ (whose ϱ is guaranteed to be positive). Fix a pair (π, ρ) lying to the left of the convex envelope of $(0, \varrho)$ and the Cramér function of $\log_\ell D_Z^{(J_1)}$. Then polar coding with kernel G enjoys block error probability $\exp(-\ell^{\pi n})$ and code rate $N^{-\rho}$ plus the conditional entropy.*

Corollary 5.48 (Good code for lossy compression)**.** *Fix a random source $Y \in \mathcal{Y}$ and a distortion function $\mathrm{dist}\colon \mathbb{F}_q \times \mathcal{Y} \to [0, 1]$. Fix an ergodic $G \in \mathbb{F}_q^{\ell \times \ell}$ (whose ϱ is guaranteed to be positive). Fix a pair (π, ρ) lying to the left of the convex envelope of $(0, \varrho)$ and the Cramér function of $\log_\ell D_S^{(J_1)}$. Then polar coding with kernel G enjoys excess of distortion $\exp(-\ell^{\pi n})$ and code rate $N^{-\rho}$ plus the test channel capacity.*

There are questions to be answered. One, how can we predict the best ϱ for a specific G? Two, what is the best ϱ among all G? Can we achieve the optimal exponent $\varrho = 1/2$? Three, Can we reduce the complexity? The first question is open. The next chapter answers the second question. The chapter after answers the third question.

CHAPTER 6

Random dynamic Kerneling

SIGNIFICANT is the fact that random coding achieves capacity. Researchers have been using random coding to prove the first achievability bounds for various coding scenarios. This evinces that random coding is the very reason why capacity, or capacity region when there are more than two users, is what it is. This is the first moment of coding theory. In addition, random coding achieves capacity at an unbeatable pace—the block error probability decays exponentially fast in the block length, and the block length grows inverse-quadratically in the gap to capacity. This sets up a holy grail that is left to code designers to chase after, which is the second-moment paradigm of coding theory.

The question is simple, Can polar coding touch the second-moment paradigm, i.e., achieve capacity at a pace that is comparable to random coding?

This chapter gives an affirmative answer. But it requires one to rethink the polarizing kernel from the bottom up. As mentioned in Chapter 5, a general kernel G can be implemented as an EU–DU pair, and some copies of EU–DU pairs will wrap around the parent channels to synthesize the child channels. In doing so, not all channels need to be wrapped. As demonstrated in Chapter 4, A channel (a pair of pins) is left naked if an algorithm decides that the channel is sufficiently polarized and not worth more EU–DU pairs. The contribution made in Chapter 4 is that this reduces complexity. Stretching this idea a bit, we see that a channel (a pair of pins) can also be wrapped by an EU–DU pair that corresponds to a different kernel [**YB15**]. Once we accept the idea that G can vary on a channel-by-channel basis, the problem becomes how to find the best G that fits a given channel.

So the simple question becomes, Can we find a good kernel for each and every channel to help polar coding touch the edge of the second-moment paradigm, i.e., to achieve capacity at a nearly-optimal pace?

Now we rethink what exactly is needed here. Do we need an algorithm to generate, for each and every channel, a kernel plus a certificate that this particular kernel is good at polarizing? Or, only do we need to prove that, for each and every channel, there exists at least one kernel that polarizes decently. The latter is considerably easier than the former because we do not specify *which* kernel is good—only that it exists. And we know how this could be done: random coding. Use a random matrix $\mathbb{G}$ to polarize a channel and show that, on average, the channel is polarized to a satisfactory extent. Then we are done; the rest is the pigeonhole principle.

I call selecting a kernel for each channel *dynamic kerneling*, This chapter in its entirety is a quantization of random dynamic kerneling.

1. The Holy Grail

Fix a DMC W and a capacity-achieving input distribution Q. The following optimality bound is borrowed.

Theorem 6.1 (Optimal codes)**.** [**AW14**, Theorem 2], [**PV10**, Theorem 6]. *Fix constants $\pi, \rho > 0$ such that $\pi + 2\rho > 1$. Assume $V > 0$, that is, assume*

$$V \coloneqq \operatorname{Var}\Bigl[\ln \frac{W(X \mid Y)}{Q(X)}\Bigr] > 0.$$

(Remark: This is an easy assumption to meet; it mostly excludes trivial channels such as those with $H = 0, 1$.) Then no block code assumes $P_{\mathrm{e}} < \exp(-\ell^{\pi n})$ and $R > C - N^{-\rho}$ except for sufficiently small N.

Fix exponents $\pi, \rho > 0$ such that $\pi + 2\rho < 1$. I am to construct a series of codes with error probability $\exp(-N^{\pi})$ and code rate $N^{-\rho}$ less than the channel capacity. The construction, based on polar coding, will enjoy encoding and decoding complexity $O(N \log N)$.

2. Why Dynamic Kernels

Readers may prefer numerical evidence for why dynamic kerneling is superior. Since there is essentially one matrix of 2×2 size, I will take 3×3 for demonstration. Over $\mathbb{F}_2$, there are essentially two matrices of 3×3 size:

$$G_{\mathrm{Ye}} \coloneqq \begin{bmatrix} 1 & & \\ 0 & 1 & \\ 1 & 1 & 1 \end{bmatrix} \qquad G_{\mathrm{Barg}} \coloneqq \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 0 & 1 \end{bmatrix}$$

G_{Ye} has coset distances $\{1, 1, 3\}$ and dual coset distances $\{1, 2, 2\}$. And G_{Barg} is its dual: G_{Barg} has coset distances $\{1, 2, 2\}$ and dual coset distance $\{1, 1, 3\}$. Both G_{Ye} and G_{Barg} have $\varrho = 1/4.938$ over BECs. Their regions of realizable (π, ρ)-pairs over BECs are plotted in Figure 6.1.

Now consider this variant of polar coding: When $H(W) \leqslant 1/2$, wrap the channel with G_{Barg}; when $H(W) > 1/2$, wrap the channel with G_{Ye}. Let $\mathcal{G}$ denote an amoebic kernel that becomes G_{Barg} when it sees a reliable channel and becomes G_{Ye} when it sees a noisy one. Then two things change. Firstly, the estimate of ϱ rises from $1/4.938$ to $1/4.183$. This is because G_{Barg} is better at polarizing reliable channels and G_{Ye} is better at polarizing noisy ones. This is the division of labor. Secondly, the coset distances are now $\{1, 2, 2\}$ on both reliable and noisy ends. Comparing to $\{1, 1, 3\}$, which reward you more (3) with a lower probability (1/3), distances $\{1, 2, 2\}$ reward you less (2) but with a higher probability (2/3). And the random variable that rewards you less, but steadily, is preferred. Now $\mathcal{G}$'s region of (π, ρ)-pairs becomes Figure 6.2.

For if we allow larger matrices, then it is not hard to imagine that the space of DMCs is partitioned into several territories, each is "ruled by" the kernel that is best at polarizing the channels living within. These kernels will be collectively called $\mathcal{G}$, and then we can talk about the (π, ρ)-pairs of $\mathcal{G}$.

3. Why Random Kernels

Before the construction begins, let me elaborate on why random matrices are preferred. As the last chapter paved the way to the determination of the MDP region $\mathcal{O}$ of a kernel, we know $\mathcal{G}$ should meet the following three requirements:

- LDP requirement: The coset distances of $\mathcal{G}$ should be large. Namely, the matrix that $\mathcal{G}$ becomes when it sees reliable channels should have large coset distances. In particular, the geometric mean of $D_Z^{(j)}$ should be $\ell - o(\ell)$.

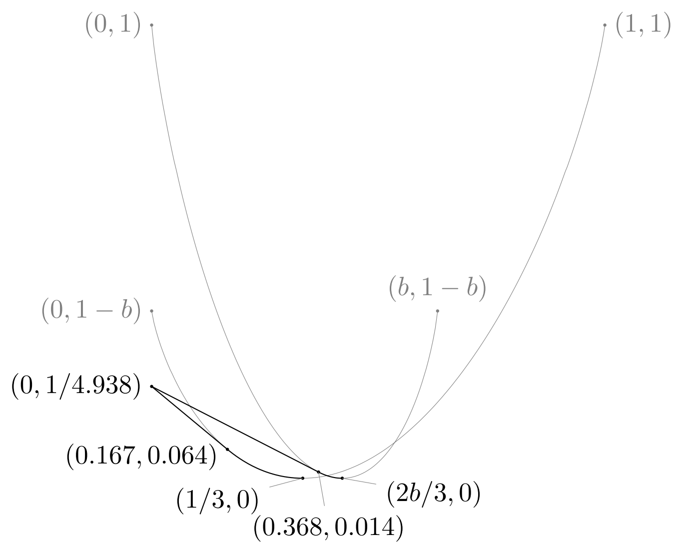

FIGURE 6.1. The two regions and Cramér functions of G_{Ye} over BECs. Here $b = \log_3 2$. The curve passing $(0, 1-b)$–$(0.167, 0.064)$–$(1/3, 0)$–$(1, 1)$ is the Cramér function of the uniform distribution on $\{0, 0, 1\}$ and also the KL divergence of $p+(1-p)$ w.r.t. $1/3+2/3$. The curve passing $(0, 1/4.938)$–$(0, 1-b)$–$(0.167, 0.064)$–$(1/3, 0)$ is the boundary of realizable (π, ρ)-pairs in the estimate of Z_n. The curve passing $(0, 1)$–$(0.368, 0.014)$–$(2b/3, 0)$–$(b, 1-b)$ is the Cramér function of the uniform distribution on $\{b, b, 0\}$ and also the KL divergence of $p/b + (1 - p/b)$ w.r.t. $2/3 + 1/3$. The curve passing $(0, 1/4.938)$–$(0.368, 0.014)$–$(2b/3, 0)$–$(1, 0)$ is the boundary of realizable (π, ρ)-pairs in the estimate of S_n. For G_{Barg} over BECs, swap the two curves.

- Dual LDP requirement: The geometric mean of $\log_\ell D_S^{(j)}$ should be $\ell - o(\ell)$. Namely, the matrix that $\mathcal{G}$ becomes when it sees noisy channels should have large coset distances.
- CLT requirement: The performance of $\mathcal{G}$ measured by the eigenvalue $\ell^{-\varrho}$ should be good. Namely, $\varrho = 1/2 - o(1)$ as $\ell \to \infty$.

It is not easy to meet any of the requirements, let alone all of them at once. As a comparison, classical coding theory usually concerns how to construct a rectangular matrix (a code) with high Hamming weights. Less frequently it concerns finding nested matrices (aka. a flag of codes) with high Hamming weights. In this viewpoint, the LDP requirement is equivalent to finding a maximal chain of nested matrices (a full flag of codes) with high Hamming weight. The CLT requirement, on the other hand, does not seem to have classical counterpart.

Polarizing W is just side of the story. On the other side, we have to polarize Q, the input distribution seen as a channel, for asymmetric channels and lossy compression. Thus we have four requirements to meet—LDP, dual LDP, W's CLT, and Q's CLT.

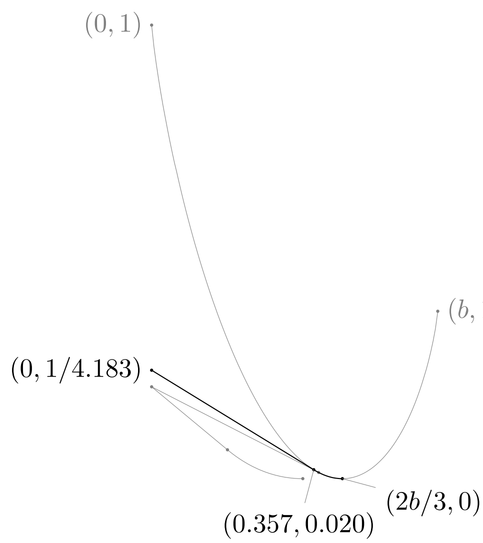

FIGURE 6.2. The region and Cramér function of $\mathcal{G}$, the compound strategy, over BECs. The old regions and Cramér functions are in gray for the ease of comparison.

Random coding is the perfect remedy for this because it measures the density of the "good objects" that meet various conditions. Take the LDP requirement as an example. If $\mathbb{G}$ is a random matrix that is drawn uniformly, then any rectangular sub-matrix is uniformly random. By classical coding theory, we know how to compute the Hamming weights of a random rectangular matrix. To be more precise, we know how to compute the probability that the minimum distance of a random linear code is at least a certain threshold. We just have to repeat the argument for all rectangular sub-matrices, and then apply the union bound.

Random matrices have more uses than that. If we draw $\mathbb{G}$ from the uniform ensemble of invertible matrices, then its Hamming distances are still under control. (In fact, excluding singular matrices avoids low Hamming distances in a helpful manner, so we are actually doing a favor for the random coding argument.) Now that the inverse $\mathbb{G}^{-1}$ follows the uniform distribution on invertible matrices, bounding the dual coset distances of $\mathbb{G}$ is as easy as bounding the primary coset distances of $\mathbb{G}$—just multiply the union bound by 2.

Below goes the actual plan to attack the problem: In section 4, I will show that a random matrix $\mathbb{G}$ has sufficient distances, and meet the LDP requirement (and its dual). In section 5, I will attempt to bound the eigenvalue of $\mathbb{G}$; this bound then comes down to four bounds:

- a bound derived from the LDP requirement via Hölder tolls,
- a bound derived from the dual LDP requirement via Hölder tolls,
- a bound for noisy-channel coding via random linear codes, and
- a bound for wiretap-channel coding via random linear codes.

Section 6 prepares some materials for the last two •. In section 7, the noisy-channel coding bound will be proved. In section 8, the wiretap-channel coding bound will be proved.

Notice that the random matrix $\mathbb{G}$ is typeset in the blackboard bold font. So will the related notations (such as $\mathbb{P}, \mathbb{E}$) be typeset in the blackboard bold font.

4. LDP Data of Random Matrix

This section is dedicated to showing that a random matrix $\mathbb{G} \in \mathbb{F}_q^{\ell\times\ell}$ drawn uniformly from the general linear group $\mathrm{GL}(\ell, q)$ has good distances $D_Z^{(j)}$ anf $D_S^{(j)}$. But before that, What is "good"? Let us consider $q = 2$ as an example. The following hand-waving argument is borrowed from [**BF02**].

To understand the typical behavior of $D_Z^{(j)}$ when $\mathbb{G}$ varies, it is easier if we compute the average of $f_Z^{(j)}(z)$ over $\mathbb{G}$, where $f_Z^{(j)}(z)$ is the weight enumerator of the coset code $\{0_1^{j-1}1_j u_{j+1}^\ell \mathbb{G} : u_{j+1}^\ell \in \mathbb{F}_q^{\ell-j}\} \subseteq \mathbb{F}_q^\ell$. Since $\mathbb{G}$ is invertible and distributed uniformly, $0_1^{j-1}1_j u_{j+1}^\ell \mathbb{G}$ is just a random nonzero vector drawn from $\mathbb{F}_q^\ell$. So every message u_{j+1}^ℓ contributed $((1+z)^\ell - 1)/(2^\ell - 1)$ to the average of $f_Z^{(j)}$, where -1 is to exclude the all-zero vector. Overall, the average of $f_Z^{(j)}(z)$ over $\mathbb{G}$ is $((1+z)^\ell - 1)/2^j$. Forget about the -1; the z^k coefficient of $(1+z)^\ell/2^j$ is $\binom{\ell}{k}/2^j$. By the large deviations theory, $[z^k]f_Z^{(j)} = \binom{\ell}{k}/2^j \approx 2^{\ell h_2(k/\ell)-j}$. Roughly speaking, the coefficient would (very likely) be 0 for a realization of $\mathbb{G}$ if the average is less than 1 (by a lot). So we conclude a rule of thumb: $D_Z^{(j)} > k$ iff $[z^k]f_Z^{(j)} < 1$ iff $h_2(k/\ell) < j/\ell$. See Figure 6.3 for an illustration.

Let us not forget that we are to bound the distances with high probability. Therefore, there is not a clear cut at $h_2(k/\ell) < j/\ell$; we have to leave a gap between $h_2(k/\ell)$ and j/ℓ to facilitate Markov's inequality. Here, I will use the upper bound $\sqrt{ep} \geqslant h_2(p)$ shown in Figure 5.2, and will leave a gap by loosening the bound to $\sqrt{3p} \geqslant h_2(p)$. The formal statement and proof is below.

Theorem 6.2 (Typical LDP behavior)**.** *Let $z \coloneqq Z_{\mathrm{mxd}}(W)$. Fix an $\ell \geqslant 30$. Draw a $\mathbb{G} \in \mathrm{GL}(\ell, q)$ uniformly at random. Fix a $j \in [\ell]$. Then, with probability less than $3q^{-\sqrt{\ell}/13}$, the following fails:*

$$f_Z^{(j)}(z) \leqslant \ell(1+q'z)^{\ell-1}(q'z)^{\lceil j^2/3\ell\rceil}, \tag{6.1}$$

where $q' \coloneqq q - 1$. In particular, we have $D_Z^{(j)} \geqslant \lceil j^2/3\ell\rceil$.

Proof. Divide j into two cases: $1 \leqslant j \leqslant \sqrt{3\ell}$ and $\sqrt{3\ell} < j \leqslant \ell$. For $j = 1, 2, \dots, \sqrt{3\ell}$, the exponent $\lceil j^2/3\ell\rceil$ is nothing but 1, thus the inequality to be proved reads $f_Z^{(j)}(z) \leqslant \ell(1+q'z)^{\ell-1}q'z$, coefficient-wisely. The right-hand side overcounts all nonzero codewords by choosing a position (ℓ), assigning a nonzero symbol ($q'z$), and arbitrarily filling in the rest of $\ell-1$ blanks ($(1+q'z)^{\ell-1}$). On the left-hand side, $f_Z^{(j)}(z)$ enumerates only codewords of the form $0_1^{j-1}1_j u_{j+1}^\ell \mathbb{G}$, which are all nonzero as $\mathbb{G}$ is invertible. Hence inequality (6.1) holds for $j \leqslant \sqrt{3\ell}$ and nonnegative z regardless of what kernel $\mathbb{G}$ is in effect.

For $j = \lfloor\sqrt{3\ell}\rfloor + 1, \lfloor\sqrt{3\ell}\rfloor + 2, \dots, \ell$, let $d \coloneqq j^2/3\ell$. To make inequality (6.1) hold, we execute a two-phase procedure to avoid all codewords of weight less than d and to eliminate kernels with poor overall score. In further detail, we will reject a kernel $\mathbb{G}$ if there exists u_{j+1}^ℓ such that $\mathrm{hwt}(0_1^{j-1}1_j u_{j+1}^\ell \mathbb{G}) < d$ and call it phase I. Afterwards, among surviving kernels with only *heavy* (high weight) codewords, we will reject a kernel if its overall score $f_Z^{(j)}(z)$ is too low and call it phase II. The

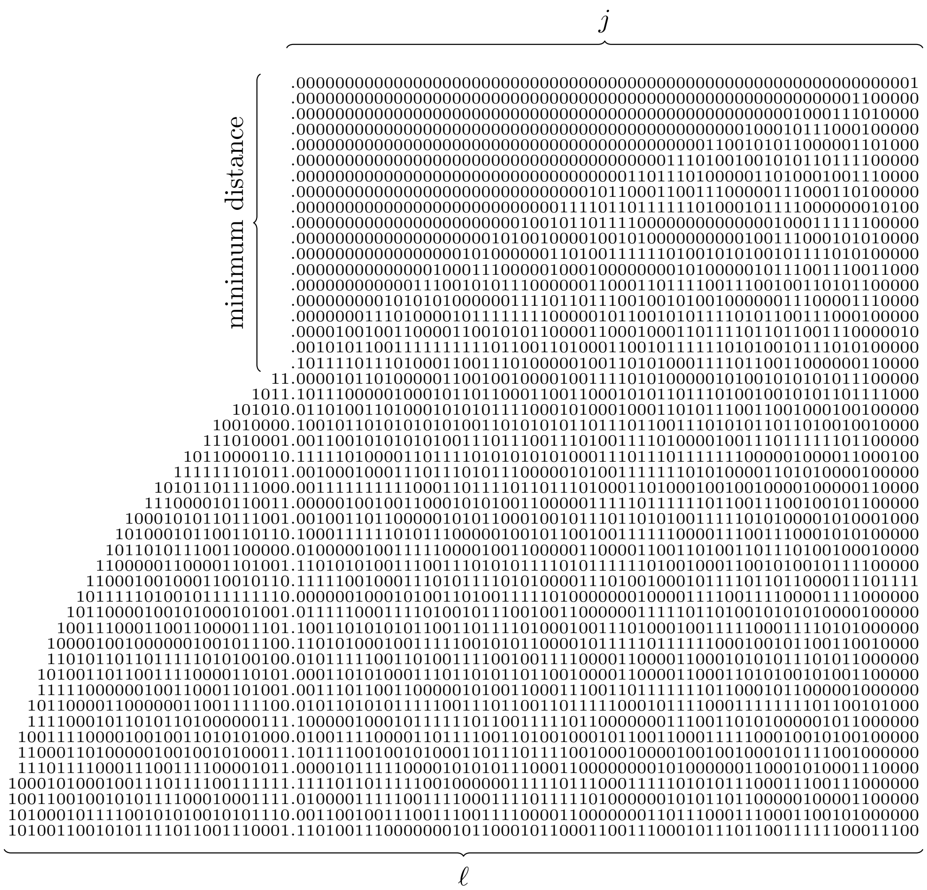

FIGURE 6.3. The binary representations of $\binom{96}{k}/2^{64}$ for $k = 0$ (top) to $k = 48$ (bottom). Note how the leading digits depict the rotation of h_2. (Zoom out if it is not clear).

failing probability $3q^{-\sqrt{\ell}/13}$ is the price we pay for rejecting. Up to this point, two things remain to be analyzed: how much probability we pay for rejecting *light* (low weight) codewords in phase I (answer: $q^{-\sqrt{\ell}/13}$), and what is the Markov cutoff that honors inequality (6.1) in phase II (answer: $2q^{-\sqrt{\ell}/13}$).

Phase I analysis is as follows: Fix u_{j+1}^{ℓ} and vary $\mathbb{G} \in \mathrm{GL}(\ell, q)$; the codeword $\mathbb{X}_1^{\ell} \coloneqq 0_1^{j-1} 1_j u_{j+1}^{\ell} \mathbb{G}$ is a nonzero vector distributed uniformly on $\mathbb{F}_q^{\ell} \setminus \{0_1^{\ell}\}$. This distribution is almost identical to the uniform distribution on $\mathbb{F}_q^{\ell}$. Assume $\mathbb{X}_1^{\ell}$ follows the latter; this makes $\mathbb{X}_1^{\ell}$ lighter, which is compatible with the direction of the inequalities we want. Then the probability that $\mathbb{X}_1^{\ell}$ has weight less than d is the probability that ℓ Bernoulli trials—each $\mathbb{X}_j$ is "zero" with probability $1/q$ and "nonzero" with probability q'/q—result in less than d "nonzero"s. By the large deviations theory [DZ10, Exercise 2.2.23(b)], $\mathrm{hwt}(\mathbb{X}_1^{\ell}) < d$ holds with probability less than

$$\exp\Big(-\ell \mathbb{D}\Big(\frac{d}{\ell} \,\Big\|\, \frac{1}{2}\Big)\Big) = 2^{-\ell(1-h_2(d/\ell))}$$

for the $q = 2$ case, where $\mathbb{D}$ is the Kullback–Leibler divergence. For general q, similarly, $\operatorname{hwt}(\mathbb{X}_1^\ell) < d$ holds with probability less than

$$\exp\Big(-\ell\mathbb{D}\Big(\frac{d}{\ell}\ \Big\|\ 1 - \frac{1}{q}\Big)\Big) \leqslant q^{-\ell(1-h_2(d/\ell))}.$$

It is less than $q^{-\ell(1-h_2(d/\ell))}$ by the comparison made in Figure 6.4 (meaning that $q = 2$ is the most difficult case). Now we obtain $h_2(d/\ell) < \sqrt{ed/\ell} = \sqrt{ej^2/3\ell^2} = (\sqrt{e/3})j/\ell < 0.952j/\ell$. Therefore, the single-word rejecting probability is less than $q^{-\ell(1-h_2(d/\ell))} < q^{-\ell+0.952j}$. Take into account that there are $q^{\ell-j}$ codewords, one for each u_{j+1}^ℓ. The union bound yields $q^{\ell-j}q^{-\ell+0.952j} = q^{-0.048j} < q^{-0.048\sqrt{3\ell}} < q^{-\sqrt{\ell}/13}$. Therefore, the total rejecting probability is $q^{-\sqrt{\ell}/13}$. Phase I ends here.

Phase II analysis is as follows: After we reject some $\mathbb{G}$ in phase I, some codewords will disappear; particularly, this includes all light codewords. Therefore, the expectation of $f_Z^{(j)}(z)$ is bounded by the weight enumerator of all heavy codewords rescaled by the number of codewords. In detail, start from

$$\begin{aligned}\mathbb{E}[f_Z^{(j)}(z) \mid \mathbb{G} \text{ survives phase I}] &= \mathbb{E}[f_Z^{(j)}(z) \cdot \mathbb{I}\{\mathbb{G} \text{ survives}\}]/\mathbb{P}\{\mathbb{G} \text{ survives}\} \\ &\leqslant \mathbb{E}[f_Z^{(j)}(z) \cdot \mathbb{I}\{\mathbb{G} \text{ survives}\}]/(1 - q^{-\sqrt{\ell}/13}).\end{aligned} \tag{6.2}$$

$\mathbb{I}$ is the indicator function. In the denominator, $1 - q^{-\sqrt{\ell}/13} > 1/4$ as $\ell \geqslant 30$. Put that aside and redefine $d := \lceil j^2/3\ell\rceil$. The expected value part is bounded from above by

$$\begin{aligned}\mathbb{E}[f_Z^{(j)}(z) \cdot \mathbb{I}\{\mathbb{G} \text{ survives}\}] &= \mathbb{E}\Big[\sum_{u_{j+1}^\ell} z^{\operatorname{hwt}(0_1^{j-1}1_j u_{j+1}^\ell \mathbb{G})} \cdot \mathbb{I}\{\mathbb{G} \text{ survives}\}\Big] \\ &\leqslant \mathbb{E}\Big[\sum_{u_{j+1}^\ell} z^{\operatorname{hwt}(0_1^{j-1}1_j u_{j+1}^\ell \mathbb{G})} \cdot \mathbb{I}\{\operatorname{hwt}(0_1^{j-1}1_j u_{j+1}^\ell \mathbb{G}) \geqslant d\}\Big] \\ &= \sum_{u_{j+1}^\ell} \mathbb{E}[z^{\operatorname{hwt}(0_1^{j-1}1_j u_{j+1}^\ell \mathbb{G})} \cdot \mathbb{I}\{\operatorname{hwt}(0_1^{j-1}1_j u_{j+1}^\ell \mathbb{G}) \geqslant d\}] \\ &\leqslant q^{\ell-j}\mathbb{E}[z^{\operatorname{hwt}(\mathbb{X}_1^\ell)} \cdot \mathbb{I}\{\operatorname{hwt}(\mathbb{X}_1^\ell) \geqslant d\}] = q^{\ell-j}q^{-\ell}\sum_{x_1^\ell} z^{\operatorname{hwt}(x_1^\ell)} \cdot \mathbb{I}\{\operatorname{hwt}(x_1^\ell) \geqslant d\} \\ &= q^{-j}\sum_{w\geqslant d}\tbinom{\ell}{w}(q'z)^w \leqslant q^{-j}\sum_{w\geqslant d}\tbinom{\ell}{d}\tbinom{\ell-d}{w-d}(q'z)^w \\ &= q^{-j}\tbinom{\ell}{d}\sum_{w\geqslant d}\tbinom{\ell-d}{w-d}(q'z)^{w-d}(q'z)^d = q^{-j}\tbinom{\ell}{d}(1+q'z)^{\ell-d}(q'z)^d\end{aligned}$$

(overestimate the scalar $q^{-j}\binom{\ell}{d}$)

$$\leqslant (q^{-\sqrt{\ell}/13}\ell/2)(1+q'z)^{\ell-d}(q'z)^d.$$

The first equality expands the definition. The next inequality replaces $\mathbb{G}$ surviving phase I by a weaker condition. The next equality swaps $\mathbb{E}$ and $\sum$. The next inequality replaces the ensemble of $0_1^{j-1}1_j u_{j+1}^\ell \mathbb{G}$ by a uniform $\mathbb{X}_1^\ell \in \mathbb{F}_q^\ell$. The next equality expands the definition of the expectation over $\mathbb{X}_1^\ell$. The next equality counts codewords. The next inequality selects w positions by first selecting d and then selecting $w - d$. The next two equalities factor and apply the binomial theorem. The rest is by a series of inequalities that overestimate the scalar: $q^{-j}\binom{\ell}{d} =$

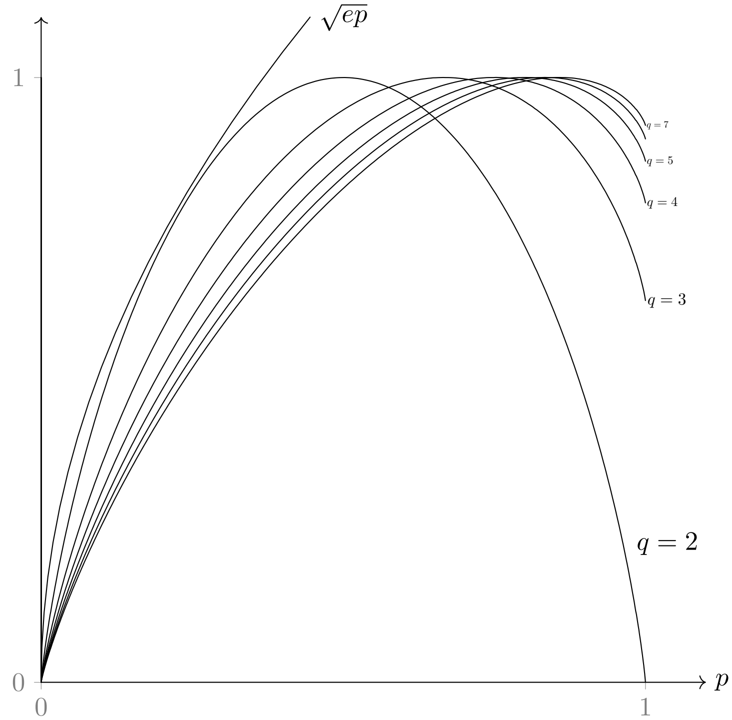

FIGURE 6.4. One minus the Kullback–Leibler divergences $1 - \mathbb{D}(p \parallel 1-1/q)/\ln(q)$ for $q = 2,3,5,7$ and an upper bound of $\sqrt{ep}$.

$q^{-j}\binom{\ell}{\lceil j^2/3\ell\rceil} < q^{-j}\binom{\ell}{j^2/3\ell}\ell/2 < q^{-j}2^{\ell h_2(j^2/3\ell^2)}\ell/2 \leqslant q^{-j+\ell h_2(j^2/3\ell^2)}\ell/2$. Similar to the end of phase I, the exponent part is $-j+\ell h_2(j^2/3\ell^2) < -j+\ell\sqrt{ej^2/3\ell^2} = -j+j\sqrt{e/3} < -0.048j < -0.048\sqrt{3\ell} < -\sqrt{\ell}/13$. Hence the scalar part is less than $q^{-\sqrt{\ell}/13}\ell/2$. Put $1-q^{-\sqrt{\ell}/13} > 1/4$ back to the denominator as in inequality (6.2); $\mathbb{E}[f_Z^{(j)}(z) \mid \mathbb{G}$ survives phase I] has an upper bound of

$$2q^{-\sqrt{\ell}/13}\ell(1+q'z)^{\ell-d}(q'z)^d.$$

By Markov's inequality, inequality (6.1) holds with probability $1-2q^{-\sqrt{\ell}/13}$, i.e., the rejecting probability is $2q^{-\sqrt{\ell}/13}$. Phase II ends here.

The sum of the two rejecting probabilities is $3q^{-\sqrt{\ell}/13}$ as claimed in the theorem statement, hence settles the proof of Theorem 6.2. □

The bound I just proved,

$$f_Z^{(j)}(z) \leqslant \ell(1+q'z)^{\ell-1}(q'z)^{\lceil j^2/3\ell\rceil}, \tag{(6.1)'s copy}$$

implies its dual siblings because $\mathbb{G}^{-1}$ is uniform on $\mathrm{GL}(\ell,q)$.

Corollary 6.3. *Let $s \coloneqq S_{\max}(W)$. Then, with probability $1-3q^{-\sqrt{\ell}/13}$, the following holds for each $j \in [\ell]$:*

$$f_S^{(\ell-j+1)}(s) \leqslant \ell(1+q'z)^{\ell-1}(q's)^{\lceil j^2/3\ell\rceil}. \tag{6.3}$$

Together, we have that with failing probability at most $6\ell q^{-\sqrt{\ell}/13}$, both inequalities (6.1) and (6.3) hold for all $j \in [\ell]$. In particular, we have that $D_Z^{(j)} \geqslant \lceil j^2/3\ell \rceil$ and $D_Z^{(\ell-j+1)} \geqslant \lceil j^2/3\ell \rceil$.

Does the preceding result allow the possibility that $\pi \to 1$? That is to say, Does inequality (6.1) imply that the average of $\log_\ell D_Z^{(j)}$ is $1 - o(1)$ as $\ell \to \infty$? The answer is

$$\frac{1}{\ell}\sum_{j=1}^{\ell} \log_\ell D_Z^{(j)} \geqslant \frac{1}{\ell}\sum_{j=\sqrt{3\ell}}^{\ell} \log_\ell \frac{j^2}{3\ell} \approx \frac{2}{\ell \ln \ell}\int_{\sqrt{3\ell}}^{\ell} \ln j \,\mathrm{d}j - \log_\ell 3\ell \approx 1 - \frac{3.1}{\ln \ell} - \frac{\sqrt{3}}{\sqrt{\ell}}.$$

Now readers can see why I insist on loosening h_2 to the square root—I do not want to integrate $\ln(h_2^{-1}(p))$.

The next section estimates the eigenvalue, $\ell^{-\varrho}$, of random kernels.

5. CLT Data of Random Matrix

This section is devoted to proving the eigen behavior of random kernels. To that end, we first need a concave eigenfunction $h\colon [0,1] \to [0,1]$. Let $\alpha \coloneqq \ln(\ln \ell)/\ln \ell$ and define

$$h(z) =: \min(z, 1-z)^{\alpha}.$$

From our experience in Lemma 5.31, the eigenvalue for reliable channels is a supremum taken over an open condition $0 < H(W) < \delta$, and hence requires explicit bounds coming from FTPCs and Hölder tolls. The next lemma does that and shows that h has a good eigenvalue if $H(W) < \ell^{-2}$.

Lemma 6.4 (Typical CLT—reliable case)**.** *Let $\ell \geqslant \max(3^q, e^4, q^5)$. Assume that $0 < H(W) < \ell^{-2}$ and that $\mathbb{G} \in \mathbb{F}_q^{\ell\times\ell}$ satisfies inequality (6.1), then*

$$\frac{h(H(W^{(1)})) + h(H(W^{(2)})) + \cdots + h(H(W^{(\ell)}))}{\ell h(H(W))} < 2\ell^{-1/2+3\alpha}. \tag{6.4}$$

Proof. In the proof of Lemma 5.31, we classify $h(H(W^{(j)}))$ into two cases—those with coset distance treated as 1 and those with nontrivial ($\geqslant 5$) coset distance. Now that we have a spectrum of coset distances ($\lceil j^2/3\ell \rceil$), there will be as many cases as there are coset distances.

Consider first $j \leqslant k \coloneqq \lfloor \ell^{1/2+5\alpha/2} \rfloor$. This is the case when H_1 goes up but how far it can go is limited by FTPCH (Theorem 5.18). More precisely,

$$\begin{aligned} h(H(W^{(1)})) + h(H(W^{(2)})) + \cdots + h(H(W^{(k)})) &\leqslant kh\Big(\frac{H(W^{(1)}) + \cdots + H(W^{(k)})}{k}\Big) \\ \leqslant kh\Big(\frac{\ell H(W)}{k}\Big) \leqslant k\ell^{\alpha} h(H(W)) k^{-\alpha} &\leqslant \ell^{\alpha}(\ell^{1/2+5\alpha/2})^{1-\alpha} h(H(W)) \\ \leqslant \ell^{\alpha+1/2+5\alpha/2-\alpha/2-5\alpha^2/2} h(H(W)) &\leqslant \ell^{1/2+3\alpha} h(H(W)). \end{aligned}$$

That is to say, the terms $h(H(W^{(j)}))$ with $j \leqslant k \coloneqq \lfloor \ell^{1/2+5\alpha/2} \rfloor$ contribute $\ell^{1/2+3\alpha}$ to the right-hand side of inequality (6.4). We now have another $\ell^{1/2+3\alpha}$ to spare for larger j.

Consider next $j \geqslant k \coloneqq \lceil \ell^{1/2+5\alpha/2} \rceil$. This is the case where we can invoke FTPCZ (Theorem 5.19) to show that $Z_{\mathrm{mxd}}(W^{(j)})$ is significantly smaller than

$Z_{\mathrm{mxd}}(W)$. So we just sandwich FTPCZ by Hölder tolls to translate it into control on H. In greater detail, with $z \coloneqq Z_{\mathrm{mxd}}(W)$,

$$\begin{aligned}
h(H(W^{(k)})) + H(H(W^{(k+1)})) + \cdots + h(H(W^{(\ell)})) &\leqslant \ell \max_{j \geqslant k} H(W^{(j)})^{\alpha} \\
&\leqslant \ell \max_{j \geqslant k} q^{3\alpha} Z_{\mathrm{mxd}}(W^{(j)})^{\alpha/2} \leqslant \ell q^{3\alpha} \max_{j \geqslant k} f_Z^{(j)}(z)^{\alpha/2} \\
&\leqslant \ell q^{3\alpha} \max_{j \geqslant k} \big(\ell(1+(q-1)z)^{\ell-1}((q-1)z)^{\lceil j^2/3\ell \rceil}\big)^{\alpha/2} \\
&\leqslant \ell q^{3\alpha} \max_{j \geqslant k} \ell^{\alpha/2} e^{qz\ell\alpha/2} (qz)^{j^2\alpha/6\ell} \leqslant \ell q^{3\alpha} \ell^{\alpha/2} e^{qz\ell\alpha/2} (qz)^{k^2\alpha/6\ell} \\
&\leqslant \ell q^{3\alpha} \ell^{\alpha/2} e^{qz\ell\alpha/2} (qz)^{\ln(\ell)^5\alpha/6} \leqslant \ell q^{3\alpha} \ell^{\alpha/2} e^{q^4\alpha/2} (qz)^{\ln(\ell)^5\alpha/6} \\
&\leqslant \ell q^{3\alpha} \ell^{\alpha/2} e^{q^4\alpha/2} (q^8 H(W))^{\ln(\ell)^5\alpha/12}. \qquad (6.5)
\end{aligned}$$

Here, the first inequality uses the largest $H(W^{(j)})$ to bound the rest. The next inequality pays the Hölder toll. The next inequality applies FTPCZ. The next inequality applies the assumption that inequality (6.1) holds. The next inequality simplifies $q-1$ and $\ell-1$ and the ceiling function. The next inequality knows that the maximum happens at $j = k$. The next inequality uses $k^2/\ell = \ell^{1-5\alpha}/\ell = \ell^{5\ln(\ln \ell)/\ln \ell} = e^{5\ln(\ln \ell)} = \ln(\ell)^5$. The next inequality pays the Hölder toll for the return trip to simplify $z \leqslant q^3\sqrt{H(W)} \leqslant q^3\sqrt{H(W)} \leqslant q^3/\ell$ in the exponent. The next inequality pays the Hölder toll for the other z in the base.

We are half way to the goal. To show that formula (6.5) $\leqslant \ell^{1/2+3\alpha} h(H(W))$, raise them to the power of $12/\alpha$ and take the quotient:

$$\begin{aligned}
&\big(\ell q^{3\alpha} \ell^{\alpha/2} e^{q^4\alpha/2} (q^8 H(W))^{\ln(\ell)^5\alpha/12}\big)^{12/\alpha} \div \big(\ell^{1/2+3\alpha} h(H(W))\big)^{12/\alpha} \\
&\quad = \ell^{12/\alpha} q^{36} \ell^6 e^{6q^4} (q^8 H(W))^{\ln(\ell)^5} \ell^{-6/\alpha-36} H(W)^{-12} \\
&\quad = e^{6q^4} \ell^{6\ln \ell - 30} H(W)^{\ln(\ell)^5 - 12} q^{8\ln(\ell)^5+36} < e^{6q^4} \ell^{6\ln \ell - 30} \ell^{-2\ln(\ell)^5+24} q^{8\ln(\ell)^5+36} \\
&\quad < e^{6q^4} \ell^{6\ln \ell - 2\ln(\ell)^5} q^{8\ln(\ell)^5+36} = e^{6q^4} \ell^{6\ln \ell - 0.4\ln(\ell)^5} \ell^{-1.6\ln(\ell)^5} q^{8\ln(\ell)^5+36} \\
&\quad \leqslant e^{6q^4} \ell^{6\ln \ell - 0.4\ln(\ell)^5} q^{-8\ln(\ell)^5} q^{8\ln(\ell)^5+36} = e^{6q^4} \ell^{6\ln \ell - 0.4\ln(\ell)^5} q^{36} \\
&\quad = e^{6q^4} \ell^{-0.3\ln(\ell)^5} \ell^{6\ln \ell - 0.1\ln(\ell)^5} q^{36} < e^{6q^4} e^{-0.3\ln(41)^2(q\ln 3)^4} \ell^{6\ln \ell - 0.1\ln(\ell)^5} q^{36} \\
&\quad < e^{6q^4} e^{-6.02q^4} \ell^{6\ln \ell - 0.1\ln(\ell)^5} q^{36} < \ell^{6\ln \ell - 0.1\ln(\ell)^5} q^{36} \\
&\quad = \ell^{6\ln \ell - \ln(\ell)^5/15} \ell^{-\ln(\ell)^5/30} q^{36} = \ell^{6\ln \ell - \ln(\ell)^5/15} q^{-5\ln(19)^5/30} q^{36} \\
&\quad < \ell^{6\ln \ell - \ln(\ell)^5/15} q^{-36.8} q^{36} < \ell^{6\ln \ell - \ln(\ell)^5/15} \leqslant \ell^{6\ln \ell - \ln(22)^4\ln(\ell)/15} \\
&\quad < \ell^{6\ln \ell - 6.08} < \ell^0 \leqslant 1.
\end{aligned}$$

The inequality involving 1.6 uses $\ell \geqslant q^5$. The inequality involving 0.3 uses $\ell \geqslant \max(41, 3^q)$. The inequality involving /30 uses $\ell \geqslant \max(19, q^5)$. The inequality involving /15 uses $\ell \geqslant 22$. I have just showed that formula (6.5)$/\ell^{-1/2+3\alpha} h(\mathsf{H}_0)$ is less than 1, with and hence without the power of $12/\alpha$.

To summarize, we saw that $h(H(W^{(k)})) + H(H(W^{(k+1)})) + \cdots + h(H(W^{(\ell)})) \leqslant$ formula (6.5) $\leqslant \ell^{-1/2+3\alpha} h(\mathsf{H}_0)$. Hence the summands $h(H(W^{(j)}))$ with $j \geqslant k \coloneqq \lceil \ell^{1/2+5\alpha/2} \rceil$ contribute $\ell^{1/2+3\alpha}$ to the right-hand side of inequality (6.4). Both small j case and large j case together contribute $2\ell^{1/2+3\alpha}$, as desired. This is the end of Lemma 6.4. $\square$

The following is automatic by duality given Lemma 6.4.

Lemma 6.5 (Typical CLT—noisy case)**.** *Assume $1-\ell^{-2}<H(W)<1$ and that $\mathbb{G}\in\mathbb{F}_q^{\ell\times\ell}$ satisfies inequality (6.3). Then*

$$\frac{\ell h(H(W^{(1)}))+\ell h(H(W^{(2)}))+\cdots+\ell h(H(W^{(\ell)}))}{\ell h(H(W))}<2\ell^{-1/2+3\alpha}.$$

We are left with the case when $\ell^{-2}<H(W)<1-\ell^{-2}$. In this zone, neither of FTPCZ and FTPCS can help because the Hölder tolls do not yield any meaningful estimate on $Z_{\mathrm{mxd}}(W)$ or $S_{\max}(W)$ to begin with. The contribution I make here is to reduce the estimate of $H(W^{(j)})$, with large j, to noisy-channel coding and to reduce the large j cases to wiretap-channel coding.

Lemma 6.6 (Typical CLT—mediocre case)**.** *Assume $\ell^{-2}\leqslant H(W)\leqslant 1-\ell^{-2}$ and that $\mathbb{G}\in\mathrm{GL}(\ell,q)$ is drawn uniformly at random. Then, with exceptional probability $2\ell^{-\ln(\ell)/20}$,*

$$\frac{h(H(W^{(1)}))+h(H(W^{(2)}))+\cdots+h(H(W^{(\ell)}))}{\ell h(H(W))}<4\ell^{-1/2+3\alpha}.$$

Proof. The desired inequality is the sum of the following three inequalities:

$$\sum_{j=\lceil H(W)\ell+\ell^{1/2+\alpha}\rceil+1}^{\ell} h(H(W^{(j)}))<\ell^{1/2+\alpha}, \tag{6.6}$$

$$\sum_{j=\lfloor H(W)\ell-\ell^{1/2+\alpha}\rfloor+1}^{\lceil H(W)\ell+\ell^{1/2+\alpha}\rceil} h(H(W^{(j)}))<2\ell^{1/2+\alpha},$$

$$\sum_{j=1}^{\lfloor H(W)\ell-\ell^{1/2+\alpha}\rfloor} h(H(W^{(j)}))<\ell^{1/2+\alpha}. \tag{6.7}$$

Here $\ell^{-2}\leqslant H(W)\leqslant 1-\ell^{-2}$ is used to rewrite the denominator $h(H(W))\geqslant\ell^{2\alpha}$. The middle one is trivial because $h\leqslant 2^{-\alpha}$. Inequality (6.6) will be proved in section 7. Inequality (6.7) will be proved in section 8. ☐

Let me comment on the heuristic behind those inequalities: To show that $h(H(W^{(j)}))$ is in general small, we first have to identify at which end each $H(W^{(j)})$ will be. We learned from the $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$ case that smaller indices usually imply noisier synthetic channels. So we believe that for $j\ll\ell H(W)$, the conditional entropy $H(W^{(j)})$ is high, and for $j\gg\ell H(W)$, the conditional entropy $H(W^{(j)})$ is low. We also believe that there is a ambiguous zone $j\approx H(W^{(j)})$ where the conditional entropy can be anywhere. From the CLT regime of random coding, we believe that the width of the ambiguous zone should be on the order of $\sqrt{\ell}$, hence the partition.

Once we have Lemmas 6.4 and 6.5 (proved above) and Lemma 6.6 (part of whose proof is in the next two sections and no later), we can conclude the eigen behavior of a random kernel $\mathbb{G}$.

Theorem 6.7 (Typical CLT behavior)**.** *Assume W is a q-ary channel. Assume $\ell\geqslant\max(3^q,e^4,q^5)$. Assume $\mathbb{G}\in\mathrm{GL}(\ell,q)$ is drawn uniformly at random. Then, with failing probability at most $2\ell^{-\ln(\ell)/20}+6\ell q^{-\sqrt{\ell}/13}$,*

$$\frac{h(H(W^{(1)}))+h(H(W^{(2)}))+\cdots+h(H(W^{(\ell)}))}{\ell h(H(W))}<4\ell^{-1/2+3\alpha}.$$

This is equivalent to saying that $4\ell^{-1/2+3\alpha}$ is the eigenvalue, or $\varrho \approx 1/2 - 4\alpha$. This exponent approaches $1/2$ as $\ell \to \infty$. Hence we know, at least, that there is a chance to achieve $\rho + 2\pi \to 1$. How exactly Theorems 6.2 and 6.7 imply that all $\rho + 2\pi < 1$ are possible will be explained in section 9.

6. Symmetrization and Universal Bound

Before the actual proof of inequalities (6.6) and (6.7), There are two more tools to be developed. First is a reduction borrowed from [**MT14**] that says, instead of considering general q-ary channel W, it suffices to consider a symmetric one $\tilde{W}$. A convenient consequence is that the uniform input distribution will achieve capacity. This helps simplify the inequalities further.

Lemma 6.8 (Symmetrization)**.** *For any q-ary channel, there is a symmetric q-ary channel $\tilde{W}$ such that $H(\tilde{W}) = H(W)$ and $H(\tilde{W}^{(j)}) = H(W^{(j)})$ for all $j \in [\ell]$.*

Proof. The strategy here is close to what we did in Chapter 3—one can establish an equivalence relation $W \cong \tilde{W}$ on channels and show that channel parameters and channel transformations respect the equivalence relation.

Please see [**MT14**, Definition 6 and Lemmas 7 and 8] plus the arguments in between for the formal treatment. □

The second tool, built on top of the first one, is an inequality concerning Gallager's E_0 function. Let us start from the definition.

Definition 6.9. Define *Gallager's E-null function*

$$E_0(t) \coloneqq -\ln \sum_{y\in\mathcal{Y}} \Bigl(\sum_{x\in\mathcal{X}} Q(x) W(y \mid x)^{\frac{1}{1+t}}\Bigr)^{1+t}$$

and its complement

$$\bar{E}_0(t) \coloneqq \ln \sum_{y\in\mathcal{Y}} \Bigl(\sum_{x\in\mathcal{X}} W(x,y)^{\frac{1}{1+t}}\Bigr)^{1+t}.$$

$\bar{E}_0$ is said to be the complement of E_0 as

$$\begin{aligned}\bar{E}_0(t) &= \ln \sum_{y\in\mathcal{Y}} \Bigl(\sum_{x\in\mathcal{X}} (q^{-1} W(y \mid x))^{\frac{1}{1+t}}\Bigr)^{1+t} \\ &= t\ln q + \ln \sum_{y\in\mathcal{Y}} \Bigl(\sum_{x\in\mathcal{X}} q^{-1} W(y \mid x)^{\frac{1}{1+t}}\Bigr)^{1+t} = t\ln q - E_0(t).\end{aligned}$$

Or $E_0(t) + \bar{E}_0(t) = t\ln q$ in short. That Q is uniform is used, otherwise a non-constant cannot penetrate the summations. The E-null function and its complement are deeply connected to the following family of measures.

Definition 6.10. For any $t \in [-2/5, 1]$, define the *t-tilted probability mass function* $W^t \colon \mathcal{X} \times \mathcal{Y} \to [0,1]$ as in [**CS07**, Definition 1]:

$$W^t(x,y) \coloneqq \frac{\bigl(\sum_{\xi\in\mathcal{X}} W(\xi,y)^{\frac{1}{1+t}}\bigr)^{1+t}}{\sum_{\eta\in\mathcal{Y}} \bigl(\sum_{\xi\in\mathcal{X}} W(\xi,\eta)^{\frac{1}{1+t}}\bigr)^{1+t}} \times \frac{W(x,y)^{\frac{1}{1+t}}}{\sum_{\xi\in\mathcal{X}} W(\xi,y)^{\frac{1}{1+t}}}$$

When $t = 0$, the tilted $W^t(x, y)$ falls back to its italic origin $W(x, y)$. These measures can be interpreted as follows: W^t behaves like a channel with a dedicated input distribution. The first fraction in the definition specifies the output distribution $W^t(y)$. The second fraction specifies the posterior distribution $W^t(x \mid y)$ when y is known. As W^t is not an actual channel, it is not meaningful to alter the input distribution and ask for the corresponding output. Like the symmetrization technique, all that matters is that we can compute the conditional entropies as if they were real channels. Quantities we are interested in are listed below.

Definition 6.11. Let H_e be the base-e entropy. Let $H_e(W^t)$ be $H_e(X^t \mid Y^t)$ where (X^t, Y^t) is a random tuple that obeys W^t. Let $H_e(X^t \upharpoonright y)$ be the entropy of the posterior distribution of X^t given $Y^t = y$; to be specific, $H_e(X^t \upharpoonright y) = \sum_{x \in X} W^t(x \mid y) \ln W^t(x \mid y)$.

Then $\bar{E}_0$ and W^t are connected in the following manner—W^t is a family of channels that "lives along the path" $\bar{E}_0(t)$.

Lemma 6.12 (Second derivative)**.** [**CS07**, Formula (13) and (19)] *For* $t \in [0, 1]$*,*

$$\frac{\mathrm{d}}{\mathrm{d}t} \bar{E}_0(t) = \bar{E}_0{}'(t) = H_e(W^t)$$

and

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}t^2} \bar{E}_0(t) = \bar{E}_0{}''(t) = \frac{\mathrm{d}}{\mathrm{d}t} H_e(W^t) &= \frac{1}{1+t} \sum_{y \in \mathcal{Y}} W^t(y) \sum_{x \in \mathcal{X}} W^t(x \mid y) \ln(W^t(x \mid y))^2 \\ &+ \frac{t}{1+t} \sum_{y \in \mathcal{Y}} W^t(y) H_e(X^t \upharpoonright y)^2 - H_e(W^t)^2. \end{aligned} \tag{6.8}$$

Since $\bar{E}_0{}''(t)$ and every other term in equation (6.8) is holomorphic in t, the equation holds in any region that assumes no poles. In particular, $-2/5 \leqslant t \leqslant 1$ is such a region. In that region, the next lemma helps bounding the terms in $\bar{E}_0{}''(t)$.

Lemma 6.13 (Second moment)**.** *If* $w_1, w_2, , \ldots, w_q$ *are positive numbers that totals to* 1*, then*

$$\sum_i w_i \ln(w_i)^2 \leqslant \begin{Bmatrix} \ln(q)^2 & \textit{for } q \geqslant 3 \\ 0.563 & \textit{for } q = 2 \end{Bmatrix} \leqslant 1.2 \ln(q)^2.$$

Lemma 6.12 (with the holomorphic continuation to $t = -2/5$) and Lemma 6.13 jointly imply the following universal quadratic bound.

Lemma 6.14 (Universal quadratic bound)**.** [**CS07**, Theorem 2] *Let* W *be a* q*-ary channel. Assume the uniform input distribution. Then Gallager's E-null function satisfies*

$$\begin{aligned} E_0(0) &= 0, \\ E_0'(0) &= I(W) \ln q, \\ E_0''(t) &\geqslant -2 \ln(q)^2 \end{aligned}$$

for all $t \in [-2/5, 1]$*. In particular, it satisfies*

$$E_0(t) \geqslant I(W) t \ln q - t^2 \ln(q)^2.$$

Proof. With Lemma 6.13, $\sum_{x\in\mathcal{X}} W^t(x \mid y)(\ln W^t(x \mid y))^2 \leqslant 1.2\ln(q)^2$ can be stated. Now equation (6.8) becomes

$$\begin{aligned}\bar{E_0}''(t) &\leqslant \frac{1}{1+t}\sum_{y\in\mathcal{Y}} W^t(y)\cdot 1.2\ln(q)^2 + \frac{\max(0,t)}{1+t}\sum_{y\in\mathcal{Y}} W^t(y)\ln(q)^2 \\ &\leqslant \frac{1}{1+t}\cdot 1.2\ln(q)^2 + \frac{\max(0,t)}{1+t}\ln(q)^2 \leqslant 2\ln(q)^2\end{aligned}$$

for all $t \in [-2/5, 1]$.

Since $E_0(t)$ is a linear function $t\ln q$ minus $\bar{E_0}(t)$, their first derivatives sum to $\ln q$ while their second derivatives are opposite. That means that $E_0(t) = E_0(0) + E_0(0)'t + E_0''(\tau)t^2/2$, for some $\tau \in [-2/5, 1]$, and therefore $\geqslant 0 + I(W)t\ln q - t^2\ln(q)^2$. □

After seeing that it suffices to consider symmetric channels and that $E(t)$ has a universal quadratic bound, we are now ready to prove inequalities (6.6) and (6.7).

7. Noisy-Channel Random Coding

This section and the next take advantage of the universal bound developed three lines ago and continues proving Theorem 6.7. This section deals with

$$\text{((6.6)'s copy)} \qquad \sum_{i=\lceil H(W)\ell+\ell^{1/2+\alpha}\rceil+1}^{\ell} h(H(W^{(i)})) < \ell^{-1/2+\alpha}$$

by passing it to an estimate that captures the performance of noisy-channel coding.

Proof of inequality (6.6). Owing to h's concavity, the left-hand side of inequality (6.6) is first simplified into

$$\sum_{i=j+1}^{\ell} h(H(W^{(k)})) \leqslant (\ell-j)h\Big(\frac{1}{\ell-j}\sum_{i=j+1}^{\ell} H(W^{(i)})\Big), \tag{6.9}$$

where $j \coloneqq \lceil H(W)\ell + \ell^{1/2+\alpha}\rceil$ for short. It suffices to prove that the right-hand side is less than $\ell^{-1/2+\alpha}$. But what lies inside the h on the right-hand side is a sum of $H(W^{(i)})$, which is equal to, by the chain rule, $H(U_{j+1}^{\ell} \mid U_1^j Y_1^{\ell})$. In order to prove inequality (6.9), I will then show

$$(\ell-j)h\Big(\frac{1}{\ell-j}H(U_{j+1}^{\ell} \mid Y_1^{\ell}U_1^j)\Big) < \ell^{-1/2+\alpha}. \tag{6.10}$$

But what is $H(U_{j+1}^{\ell} \mid Y_1^{\ell}U_1^j)$? It measures the equivocation at Bob's end when U_1^j is known to Bob. In other words, we may as well pretend that

- there is a random rectangular full-rank matrix $\mathbb{G}'$ with ℓ columns and only $k \coloneqq \ell - j = \lfloor I(W)\ell - \ell^{1/2+\alpha}\rfloor$ rows,
- Alice computes and sends $X_1^{\ell} \coloneqq U_{j+1}^{\ell}\mathbb{G}'$ to Bob, and
- Bob attempts to decode $\hat{U}_{j+1}^{\ell}$ upon receiving Y_1^{ℓ} using the maximum a posteriori decoder.

The equivocation is thus, by Fano's inequality, bounded in terms of the probability that Bob fails to decode U_{j+1}^{ℓ}:

$$H(U_{j+1}^{\ell} \mid Y_1^{\ell}U_1^j) \leqslant -P_{\mathrm{e}}\ln_q P_{\mathrm{e}} - (1-P_{\mathrm{e}})\ln_q(1-P_{\mathrm{e}}) + P_{\mathrm{e}}\ln_q(q^k-1)$$

$$\leqslant -P_{\mathrm{e}} \ln_q P_{\mathrm{e}} + \frac{P_{\mathrm{e}}}{\ln q} + P_{\mathrm{e}} = P_{\mathrm{e}} \cdot \Big(\frac{1 - \ln P_{\mathrm{e}}}{\ln q} + k \Big). \tag{6.11}$$

Here P_{e} is the probability that Bob fails to decode, $\hat{U}^\ell_{j+1} \neq U^\ell_{j+1}$.

In what follows is how to compute Bob's block error probability. The generator matrix $\mathbb{G}'$ used by Alice is selected uniformly from the ensemble of full-rank k-by-ℓ matrices. The difference of every pair of codewords distributes uniformly on $\mathbb{F}_q^\ell \setminus \{0_1^\ell\}$. Over symmetric channels, the difference alone determines the difficulty of decoding because $W^\ell(y_1^\ell \mid \xi_1^\ell + x_1^\ell) = W^\ell(\sigma_1^\ell(y_1^\ell) \mid x_1^\ell)$ for some component-wise involution σ_1^ℓ on $\mathcal{Y}^\ell$ depending on ξ_1^ℓ. Therefore, Gallager's bound applies. To elaborate, let $t \in [0, 1]$. Then Bob's average error probability satisfies [**Gal68**, Inequalities (5.6.2) to (5.6.14)]

$$\begin{aligned}
\mathbb{E}P_{\mathrm{e}} &= \mathbb{E}EP\{\text{Bob fails to decode } U^\ell_{j+1} \text{ given } \mathbb{G}', Y_1^\ell\} \\
&= \mathbb{E}\sum_{u_1^k} \frac{1}{q^k} \sum_{y_1^\ell} W^\ell(y_1^\ell \mid u_1^k\mathbb{G}')\mathbb{I}\{\text{Bob has } \hat{U}^\ell_{j+1} \neq u_1^k \text{ given } \mathbb{G}', y_1^\ell \mid U^\ell_{j+1} = u_1^k\} \\
&= \mathbb{E}\sum_{y_1^\ell} W^\ell(y_1^\ell \mid 0_1^\ell)\mathbb{I}\{\text{Bob has } \hat{U}^\ell_{j+1} \neq 0_1^k \text{ given } \mathbb{G}', y_1^\ell \mid U^\ell_{j+1} = 0_1^k\} \\
&\leqslant \mathbb{E}\sum_{y_1^\ell} W^\ell(y_1^\ell \mid 0_1^\ell)\Big(\sum_{v_1^k \neq 0_1^k} \mathbb{I}\{\text{Bob prefers } v_1^k \text{ over } 0_1^k \text{ given } \mathbb{G}', y_1^\ell\}\Big)^t \\
&\leqslant \mathbb{E}\sum_{y_1^\ell} W^\ell(y_1^\ell \mid 0_1^\ell)\Big(\sum_{v_1^k \neq 0_1^k} \frac{W^\ell(y_1^\ell \mid v_1^k\mathbb{G}')^{\frac{1}{1+t}}}{W^\ell(y_1^\ell \mid 0_1^\ell)^{\frac{1}{1+t}}}\Big)^t \\
&= \mathbb{E}\sum_{y_1^\ell} W^\ell(y_1^\ell \mid 0_1^\ell)^{\frac{1}{1+t}}\Big(\sum_{v_1^k \neq 0_1^k} W^\ell(y_1^\ell \mid v_1^k\mathbb{G}')^{\frac{1}{1+t}}\Big)^t \\
&\leqslant \sum_{y_1^\ell} W^\ell(y_1^\ell \mid 0_1^\ell)^{\frac{1}{1+t}}\Big(\mathbb{E}\sum_{v_1^k \neq 0_1^k} W^\ell(y_1^\ell \mid v_1^k\mathbb{G}')^{\frac{1}{1+t}}\Big)^t \\
&= \sum_{y_1^\ell} W^\ell(y_1^\ell \mid 0_1^\ell)^{\frac{1}{1+t}}\Big(\sum_{x_1^\ell \neq 0_1^\ell} \frac{q^k - 1}{q^\ell - 1} W^\ell(y_1^\ell \mid x_1^\ell)^{\frac{1}{1+t}}\Big)^t \\
&\leqslant q^{kt}\sum_{y_1^\ell} W^\ell(y_1^\ell \mid 0_1^\ell)^{\frac{1}{1+t}}\Big(\sum_{x_1^\ell \neq 0_1^\ell} \frac{1}{q^\ell} W^\ell(y_1^\ell \mid x_1^\ell)^{\frac{1}{1+t}}\Big)^t \\
&\leqslant q^{kt}\sum_{y_1^\ell} W^\ell(y_1^\ell \mid 0_1^\ell)^{\frac{1}{1+t}}\Big(\sum_{x_1^\ell} \frac{1}{q^\ell} W^\ell(y_1^\ell \mid x_1^\ell)^{\frac{1}{1+t}}\Big)^t \\
&= q^{kt}\sum_{y_1^\ell} \Big(\sum_{x_1^\ell} \frac{1}{q^\ell} W^\ell(y_1^\ell \mid x_1^\ell)^{\frac{1}{1+t}}\Big)\Big(\sum_{x_1^\ell} \frac{1}{q^\ell} W^\ell(y_1^\ell \mid x_1^\ell)^{\frac{1}{1+t}}\Big)^t \\
&= q^{kt}\sum_{y_1^\ell} \Big(\sum_{x_1^\ell} \frac{1}{q^\ell} W^\ell(y_1^\ell \mid x_1^\ell)^{\frac{1}{1+t}}\Big)^{1+t} \\
&= q^{kt}\sum_{y_1^\ell} \Big(\sum_{x_1^\ell} Q^\ell(x_1^\ell) W^\ell(y_1^\ell \mid x_1^\ell)^{\frac{1}{1+t}}\Big)^{1+t} \\
&= \exp(kt\ln q - (\textit{the E-null function of } W^\ell)(t))
\end{aligned}$$

$$= \exp(kt\ln q - \ell E_0(t)).$$

In summary, the average block error probability $\mathbb{E}P_{\mathrm{e}} = \mathbb{E}EP\{$Bob fails to decode U_{j+1}^{ℓ} given $\mathbb{G}'\}$ is no more than $\exp(kt\ln q - \ell E_0(t))$ whenever $0 \leqslant t \leqslant 1$. Recall the universal quadratic bound developed in Lemma 6.14: $E_0(t) \geqslant I(W)t\ln q - t^2\ln(q)^2$. We obtain that the exponent is

$$\begin{aligned}
kt\ln q - \ell E_0(t) &\leqslant (I(W)\ell - \ell^{1/2+\alpha})t\ln q - \ell E_0(t) \\
&\leqslant (I(W)\ell - \ell^{1/2+\alpha})t\ln q - \ell(I(W)t\ln q - t^2\ln(q)^2) \\
&= (\ell t\ln q - \ell^{1/2+\alpha})t\ln q
\end{aligned}$$

(redeem the inequality at $t = \ell^{-1/2+\alpha}/2\ln q$)

$$\begin{aligned}
&\mapsto (\ell\ell^{-1/2+\alpha}/2 - \ell^{1/2+\alpha})\ell^{-1/2+\alpha}/2 \\
&= -\ell^{2\alpha}/4 = -\ell^{2\ln(\ln\ell)/\ln\ell}/4 = -\ln(\ell)^2/4.
\end{aligned}$$

So far, the average error probability $\mathbb{E}P_{\mathrm{e}}$ is shown to be less than $\exp(-\ln(\ell)^2/4) = \ell^{-\ln(\ell)/4}$.

Run Markov's inequality with cutoff $\ell^{-\ln(\ell)/20}$. To put it another way, we sample a random full-rank matrix $\mathbb{G}' \in \mathbb{F}_q^{k\times\ell}$ and reject it if $P\{$Bob fails to decode U_{j+1}^{ℓ} given $\mathbb{G}'\} \geqslant \ell^{-\ln(\ell)/5}$. Then the rejecting probability is $\ell^{-\ln(\ell)/20}$ because $1/20+1/5 = 1/4$. An upper bound on Bob's error probability being $P_{\mathrm{e}} < \ell^{-\ln(\ell)/5}$, an upper bound on Bob's equivocation is

$$H(U_{j+1}^{\ell} \mid Y_1^{\ell}U_1^{j}) \leqslant \ell^{-\ln(\ell)/5}\Big(\frac{1-\ln\ell^{-\ln(\ell)/5}}{\ln q} + k\Big) = \ell^{-\ln(\ell)/5}\Big(\frac{1+\ln(\ell)^2/5}{\ln q} + k\Big)$$

by inequality (6.11). Plugging the right-hand side into $kh(\text{this place}/k)$, we derive that the left-hand side of inequality (6.10) is less than

$$\begin{aligned}
kh\Big(\frac{\ell^{-\ln(\ell)/5}}{k}\Big(\frac{1+\ln(\ell)^2/5}{\ln q} + k\Big)\Big) &= k\cdot\Big(\ell^{-\ln(\ell)/5}\Big(\frac{1+\ln(\ell)^2/5}{k\ln q} + 1\Big)\Big)^{\alpha} \\
= \ell^{-\alpha\ln(\ell)/5}k\cdot\Big(\frac{1+\ln(\ell)^2/5}{k\ln q} + 1\Big)^{\alpha} &< \ell^{-\alpha\ln(\ell)/5}\ell\cdot\Big(\frac{1+\ln(\ell)^2/5}{\ell\ln q} + 1\Big)^{\alpha} \\
< \ell^{-\alpha\ln(\ell)/5}\cdot\ell\cdot 2^{\alpha} = 2^{\alpha}\ell\ln(\ell)^{-\ln(\ell)/5}.
\end{aligned}$$

The first inequality uses that the left-hand side increases monotonically in k and k is $\ell - j = \lfloor I(W)\ell - \ell^{1/2+\alpha}\rfloor < \ell$. The second inequality uses the assumption $\ell \geqslant 2$. The quantity at the end of the chain of inequalities decays to 0 as $\ell \to \infty$, so eventually it becomes less than $\ell^{1/2+\alpha}$, the right-hand side of inequality (6.10). This proves that inequalities (6.6) and (6.9) hold with failing probability $\ell^{-\ln(\ell)/20}$ as soon as ℓ is large enough.

The lower bound on ℓ in the statement of Lemma 6.6 is large enough ($\ell > 20$). Hence inequality (6.6), the first half of Lemma 6.6, is settled. ☐

That random kernels make $h(H(W^{(j)}))$ small for large $j \gg \ell H(W)$ is the first half; the next section settles the second half of Lemma 6.6, making $h(H(W^{(j)}))$ small for small $j \ll \ell H(W)$.

8. Wiretap-Channel Random Coding

This subsection contains the very last ingredient of the proof of Lemma 6.6 and Theorem 6.7. We dealt with inequality (6.6) in the last subsection. We now deal with

$$\text{((6.7)'s copy)} \qquad \sum_{i=1}^{\lfloor H(W)\ell-\ell^{1/2+\alpha}\rfloor} h(H(W^{(i)})) < \ell^{1/2+\alpha}.$$

by passing it to an estimate that captures the performance of wiretap-channel coding.

Proof of inequality (6.7). Similar to how we motivated inequality (6.10), we hereby apply Jensen's inequality and the chain rule of conditional entropy to simplify inequality (6.7). The left-hand side becomes $jh(H(U_1^j \mid Y_1^\ell)/j)$ where $j \coloneqq \lfloor H(W)\ell - \ell^{1/2+\alpha}\rfloor$ for short. (This is not the same j as in the last subsection.) The input being uniform, the argument of h is $H(U_1^j \mid Y_1^\ell)/j = 1 - I(U_1^j\,;Y_1^\ell)/j$, which can be replaced by $I(U_1^j \mid Y_1^\ell)/j$ thanks to the symmetry $h(1-z) = h(z)$. We will show

$$jh\Big(\frac{1}{j}I(U_1^j\,;Y_1^\ell)\Big) < \ell^{1/2+\alpha}. \tag{6.12}$$

But what is $I(U_1^j\,;Y_1^\ell)$? It is the amount of information Eve learns from wiretapping Y_1^ℓ if Eve knows that U_{j+1}^ℓ are junk. In other words, we may pretend that

- Alice transmits $X_1^\ell \coloneqq U_1^j V_{j+1}^\ell \mathbb{G}$, wherein U_1^j are the confidential bits and V_{j+1}^ℓ are the obfuscating bits,
- Bob receives X_1^ℓ in full, and
- Eve learns Y_1^ℓ.

This context falls back to (a special case of) the traditional setup of wiretap channels [**Wyn75**] where various bounds are studied, some in terms of Gallager's E-null function.

Here are some preliminaries to control the information leaked to Eve. We follow the blueprint of how Hayashi derived the secrecy exponent in [**Hay06**, Inequality (21)]. Consider the communication protocol depicted in Figure 6.5: Karl fixes a kernel $\mathbb{G} \in \mathrm{GL}(\ell, q)$ and everyone knows $\mathbb{G}$. Alice chooses the confidential message U_1^ℓ. Vincent chooses the obfuscating bits V_{j+1}^ℓ. Charlie generates Y_1^ℓ by plugging $X_1^\ell \coloneqq U_1^j V_{j+1}^\ell \mathbb{G}$ into a simulator of W^ℓ. Eve learns Y_1^ℓ and is interested in knowing U_1^j alone. So the channel on topic is the composition of Vincent and Charlie. Notation: Running out of symbols, we all use $\mathbb{P}$ with proper subscripts to indicate the corresponding probability measures. That said, indices in the subscript will be omitted. As Eve is interested in the relation between U_1^j and Y_1^ℓ, let $Y_1^\ell \restriction Gu_1^j$ be the r.v. that follows the posterior distribution of Y_1^ℓ given $\mathbb{G} = G$ and $U_1^j = u_1^j$. More formally, $\mathbb{P}_{Y\restriction Gu}(y_1^\ell) = \mathbb{P}_{Y|\mathbb{G}U}(y_1^\ell \mid G, u_1^\ell) = \mathbb{P}_{\mathbb{G}UY}(G, u_1^j, y_1^\ell)/\mathbb{P}_{\mathbb{G}U}(G, u_1^j)$. We could have defined $Y_1^\ell \restriction G$ to be the posterior distribution of Y_1^ℓ given $\mathbb{G} = G$; but it is simply the same distribution as Y_1^ℓ since $U_1^j V_{j+1}^\ell G$ traverses all inputs uniformly regardless of the choice of G. That is, $\mathbb{P}_{Y|\mathbb{G}}(y_1^\ell \mid G) = \mathbb{P}_Y(y_1^\ell)$ for all $y_1^\ell \in \mathcal{Y}^\ell$.

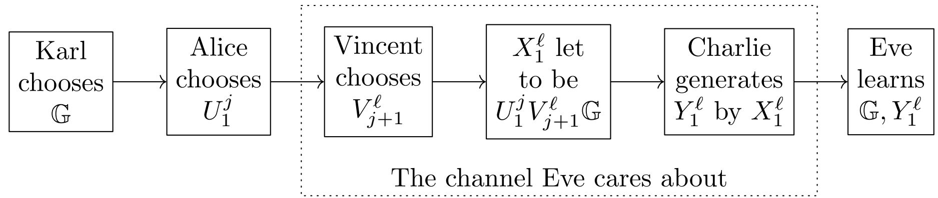

FIGURE 6.5. A finer setup for Hayashi's secrecy exponent. Charlie generates Y_1^ℓ such that $X_1^\ell := U_1^j V_{j+1}^\ell \mathbb{G}$ and Y_1^ℓ follow W^ℓ. Despite of the seemingly sequential structure, Karl, Alice, and Vincent work independently.

Fix G as an instance of $\mathbb{G}$. Let I_e be the base-e mutual information. The channel Eve cares about leaks information of this amount:

$$I_e(U_1^j\,;Y_1^\ell \mid G) = \sum_{u_1^j y_1^\ell} \mathbb{P}_{UY|\mathbb{G}}(u_1^j, y_1^\ell \mid G) \ln \frac{\mathbb{P}_{Y|\mathbb{G}U}(y_1^\ell \mid G, u_1^j)}{\mathbb{P}_{Y|\mathbb{G}}(y_1^\ell \mid G)}$$

$$= \sum_{u_1^j} \mathbb{P}_U(u_1^j) \sum_{y_1^\ell} \mathbb{P}_{Y|\mathbb{G}U}(y_1^\ell \mid G, u_1^j) \ln \frac{\mathbb{P}_{Y|\mathbb{G}U}(y_1^\ell \mid G, u_1^j)}{\mathbb{P}_{Y|\mathbb{G}}(y_1^\ell \mid G)}$$

$$= \sum_{u_1^j} \mathbb{P}_U(u_1^j) \sum_{y_1^\ell} \mathbb{P}_{Y\restriction Gu}(y_1^\ell) \ln \frac{\mathbb{P}_{Y\restriction Gu}(y_1^\ell)}{\mathbb{P}_Y(y_1^\ell)} = \sum_{u_1^j} \mathbb{P}_U(u_1^j) \mathbb{D}(Y_1^\ell \restriction Gu_1^j \parallel Y_1). \tag{6.13}$$

$\mathbb{D}(Y_1^\ell \restriction Gu_1^j \parallel Y_1^\ell)$ is the Kullback–Leibler divergence from the posterior distribution of Y_1^ℓ given G, u_1^j to the coarsest distribution Y_1^ℓ. We are to take expectation over $\mathbb{G}$ to find the average information leak since we are interested in Markov's inequality. Formula (6.13) gives rise to

$$\mathbb{E}I_e(U_1^j\,;Y_1^\ell \mid \mathbb{G}) = \sum_G \mathbb{P}_{\mathbb{G}}(G) I_e(U_1^j\,;Y_1^\ell \mid G)$$

$$= \sum_G \mathbb{P}_{\mathbb{G}}(G) \sum_{u_1^j} \mathbb{P}_U(u_1^j) \mathbb{D}(Y_1^\ell \restriction Gu_1^j \parallel Y_1^\ell). \tag{6.14}$$

We now discover that there are redundancies in traversing all G and u_1^ℓ: After all, X_1^j is $u_1^j V_{j+1}^\ell G = u_1^j 0_{j+1}^\ell G + 0_1^j V_{j+1}^\ell G$, which is a fixed linear combination of the first j rows plus a random vector from the span of the bottom $\ell - j$ rows. When V_1^ℓ varies, the track of X_1^ℓ forms an affine subspace of $\mathbb{F}_q^\ell$, a *coset code* as in the context of the fundamental theorems. So what matters is the distribution of this coset code.

In the aforementioned manner, we replace the uniform ensemble of $(\mathbb{G}, U_1^j)$ by the uniform ensemble of $\mathbb{K}$, a rank-$(\ell - j)$ affine subspace of $\mathbb{F}_q^\ell$, where $j := \lfloor H(W)\ell - \ell^{1/2+\alpha} \rfloor$. Karl and Alice together choose $\mathbb{K}$ uniformly. Vincent chooses $X_1^\ell \in \mathbb{K}$ uniformly. Charlie generates Y_1^ℓ by entering X_1^ℓ into a simulator of W^ℓ. See Figure 6.6 for the depiction of the new scheme. Hence formula (6.14) induces

$$\mathbb{E}I_e(U_1^j\,;Y_1^\ell \mid \mathbb{G}) = \sum_K \mathbb{P}_{\mathbb{K}}(K) \mathbb{D}(Y_1^\ell \restriction K \parallel Y_1^\ell)$$

where $Y_1^\ell \restriction K$ is the a posteriori distribution of Y_1^ℓ given $\mathbb{K} = K$. Suddenly, the quantity $\mathbb{E}I_e(U_1^j\,;Y_1^\ell \mid \mathbb{G})$ we are interested in turns into the mutual information

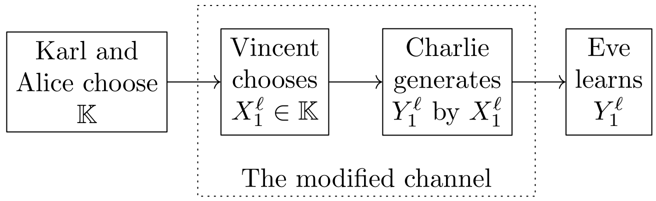

FIGURE 6.6. A simplified setup for Hayashi's secrecy exponent. Charlie generates Y_1^ℓ such that X_1^ℓ and Y_1^ℓ follow W^ℓ.

$I_e(\mathbb{K}\,;Y_1^\ell)$ between $\mathbb{K}$ and Y_1^ℓ as $\mathbb{K}$ replaces the role of U_1^j in formula (6.13). Recall that in Lemma 6.14 the mutual information is the derivative of Gallager's E-null function. We exploit this. Define the double-stroke E-null function for $(\mathbb{K}, Y_1^\ell)$ as follows

$$\mathbb{E}_0(t) \coloneqq -\ln \sum_{y_1^\ell} \Bigl(\sum_K \mathbb{P}_{\mathbb{K}}(K) \mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)^{\frac{1}{1+t}} \Bigr)^{1+t}.$$

Then $\mathbb{E}_0'(0) = I_e(\mathbb{K}\,;Y_1^\ell) = \mathbb{E} I_e(U_1^j\,;Y_1^\ell \mid \mathbb{G})$. Owing to the concavity of the E-null function, $\mathbb{E}_0'(0) \leqslant \mathbb{E}_0(t)/t$ whenever $-2/5 \leqslant t < 0$. Recap: To bound the average leaked information $\mathbb{E} I_e(U_1^j\,;Y_1^\ell \mid \mathbb{G})$ it suffices to bound $I_e(\mathbb{K}\,;Y_1^\ell)$, which is then morphed to bound $\mathbb{E}_0'(0)$ from above and to bound $\mathbb{E}_0(t)$ from below.

The double-stroke E-null function is bounded as below. Assume $-2/5 \leqslant t < 0$. Let s be $-t/(1+t)$; so $0 < s \leqslant 2/3$ and $(1+s)(1+t) = 1$. For any fixed K and fixed $x_1^\ell \in K$, the base of the $(1+t)$-th root in the definition of the double-stroke E-null function is

$$\begin{aligned}
\mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K) &= \sum_{\xi_1^\ell \in K} \mathbb{P}_{X|\mathbb{K}}(\xi_1^\ell \mid K) \mathbb{P}_{Y|X}(y_1^\ell \mid \xi_1^\ell) \\
&= \sum_{\xi_1^\ell \in K} q^j \mathbb{P}_X(\xi_1^\ell) \mathbb{P}_{Y|X}(y_1^\ell \mid \xi_1^\ell) = \sum_{\xi_1^\ell \in K} q^j \mathbb{P}_{XY}(\xi_1^\ell, y_1^\ell) \\
&= q^j \Bigl(\mathbb{P}_{XY}(x_1^\ell, y_1^\ell) + \sum_{x_1^\ell \neq \xi_1^\ell \in K} \mathbb{P}_{XY}(\xi_1^\ell, y_1^\ell) \Bigr) \\
&= q^j \Bigl(\mathbb{P}_{XY}(x_1^\ell, y_1^\ell) + \mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell) \Bigr).
\end{aligned}$$

Here $\mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)$ is a temporary shorthand for the summation of $\mathbb{P}_{XY}(\xi_1^\ell, y_1^\ell)$ over $\xi_1^\ell \in K$ that excludes x_1^ℓ. Raise $\mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)$ to the power of s; it becomes $q^{js}(\mathbb{P}_{XY}(x_1^\ell, y_1^\ell) + \mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell))^s \leqslant q^{js}(\mathbb{P}_{XY}(x_1^\ell, y_1^\ell)^s + \mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)^s)$ by sub-additivity. Put that aside; raise $\mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)$ to the power of $1 + s = 1/(1+t)$:

$$\begin{aligned}
\mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)^{1+s} &= \mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K) \mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)^s = \sum_{x_1^\ell \in K} q^j \mathbb{P}_{XY}(x_1^\ell, y_1^\ell) \mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)^s \\
&\leqslant \sum_{x_1^\ell \in K} q^j \mathbb{P}_{XY}(x_1^\ell, y_1^\ell) q^{js} \Bigl(\mathbb{P}_{XY}(x_1^\ell, y_1^\ell)^s + \mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)^s \Bigr) \\
&= q^{j+js} \Bigl(\sum_{x_1^\ell \in K} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)^{1+s} + \sum_{x_1^\ell \in K} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell) \mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)^s \Bigr).
\end{aligned}$$

The inequality rewrites the sth power of $\mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)$. Then the inner sum of the E-null function morphs as follows

$$\begin{aligned}
\sum_K \mathbb{P}_{\mathbb{K}}(K)\mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)^{1+s} &\leqslant \sum_K \mathbb{P}_{\mathbb{K}}(K) q^{j+js}\Big(\sum_{x_1^\ell \in K} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)^{1+s} \\
&\qquad + \sum_{x_1^\ell \in K} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)\mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)^s\Big) \\
\text{(diagonal arc)} \qquad &= q^{j+js}\sum_K \mathbb{P}_{\mathbb{K}}(K)\sum_{x_1^\ell \in K} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)^{1+s} \\
\text{(off-diagonal arc)} \qquad &\quad + q^{j+js}\sum_K \mathbb{P}_{\mathbb{K}}(K)\sum_{x_1^\ell \in K} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)\mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)^s.
\end{aligned}$$

The inequality rewrites the $(s+1)$th power of $\mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)$. Divide and conquer—the inner sum of the double-stroke E-null function is split into two arcs as labeled.

The diagonal arc is exactly

$$\begin{aligned}
q^{j+js}\sum_K \mathbb{P}_{\mathbb{K}}(K)\sum_{x_1^\ell \in K} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)^{1+s} &= q^{j+js}\frac{1}{q^j}\sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)^{1+s} \\
= q^{js}\sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_X(x_1^\ell)^{1+s}\mathbb{P}_{Y|X}(y_1^\ell \mid x_1^\ell)^{1+s} &= q^{js-\ell s}\sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_X(x_1^\ell)\mathbb{P}_{Y|X}(y_1^\ell \mid x_1^\ell)^{1+s}.
\end{aligned}$$

The off-diagonal arc is

$$\begin{aligned}
&q^{j+js}\sum_K \mathbb{P}_{\mathbb{K}}(K)\sum_{x_1^\ell \in K} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)\mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)^s \\
&\quad = q^{j+js}\sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)\sum_{K \ni x_1^\ell} \mathbb{P}_{\mathbb{K}}(K)\mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)^s.
\end{aligned}$$

The inner sum is loosened to

$$\begin{aligned}
\sum_{K \ni x_1^\ell} \mathbb{P}_{\mathbb{K}}(K)\mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)^s &= \frac{1}{q^j}\sum_{K \ni x_1^\ell} \mathbb{P}_{\mathbb{K}|X}(K \mid x_1^\ell)\mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)^s \\
&\leqslant \frac{1}{q^j}\Big(\sum_{K \ni x_1^\ell} \mathbb{P}_{\mathbb{K}|X}(K \mid x_1^\ell)\mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)\Big)^s \\
&= \frac{1}{q^j}\Big(\sum_{K \ni x_1^\ell} \mathbb{P}_{\mathbb{K}|X}(K \mid x_1^\ell)\sum_{x_1^\ell \neq \xi_1^\ell \in K} \mathbb{P}_{XY}(\xi_1^\ell, y_1^\ell)\Big)^s \\
&= \frac{1}{q^j}\Big(\frac{q^{\ell-j}-1}{q^\ell-1}\sum_{x_1^\ell \neq \xi_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_{XY}(\xi_1^\ell, y_1^\ell)\Big)^s \leqslant \frac{1}{q^{j+js}}\Big(\sum_{x_1^\ell \neq \xi_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_{XY}(\xi_1^\ell, y_1^\ell)\Big)^s
\end{aligned}$$

The last equality counts the multiplicity of ξ_1^ℓ. So the off-diagonal arc is loosened to

$$\begin{aligned}
&q^{j+js}\sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)\sum_{K \ni x_1^\ell} \mathbb{P}_{\mathbb{K}}(K)\mathbb{P}_{XY}(K \backslash x_1^\ell, y_1^\ell)^s \\
&\quad \leqslant \sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)\Big(\sum_{x_1^\ell \neq \xi_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_{XY}(\xi_1^\ell, y_1^\ell)\Big)^s
\end{aligned}$$

$$\begin{aligned}
&\leqslant \sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)\Big(\sum_{\xi_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_{XY}(\xi_1^\ell, y_1^\ell)\Big)^s \\
&= \sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_{XY}(x_1^\ell, y_1^\ell)\mathbb{P}_Y(y_1^\ell)^s = \mathbb{P}_Y(y_1^\ell)\mathbb{P}_Y(y_1^\ell)^s = \mathbb{P}_Y(y_1^\ell)^{1+s}.
\end{aligned}$$

Both the diagonal and off-diagonal arcs being conquered, merge them and raise to the $(1+t)$-th power. The summand for any fixed y_1^ℓ in the definition of the double-stroke E-null function is

$$\begin{aligned}
&\Big(\sum_K \mathbb{P}_{\mathbb{K}}(K)\mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)^{\frac{1}{1+t}}\Big)^{1+t} = (\text{off-diagonal} + \text{diagonal})^{1+t} \\
&\quad \leqslant \text{off-diagonal}^{1+t} + \text{diagonal}^{1+t} \leqslant \Big(\mathbb{P}_Y(y_1^\ell)^{1+s}\Big)^{1+t} + \text{diagonal}^{1+t} \\
&\quad = \mathbb{P}_Y(y_1^\ell) + \text{diagonal}^{1+t} \leqslant \mathbb{P}_Y(y_1^\ell) + \Big(q^{js-\ell s} \sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_X(x_1^\ell)\mathbb{P}_{Y|X}(y_1^\ell \mid x_1^\ell)^{1+s}\Big)^{1+t} \\
&\quad = \mathbb{P}_Y(y_1^\ell) + q^{\ell t - jt}\Big(\sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_X(x_1^\ell)\mathbb{P}_{Y|X}(y_1^\ell \mid x_1^\ell)^{1+s}\Big)^{1+t}
\end{aligned}$$

The first equality divides. The next inequality applies the sub-additivity of $(1+t)$th power (note that $t < 0$). We can finally bound the double-stroke E-null function per se:

$$\begin{aligned}
\exp(-\mathbb{E}_0(t)) &= \sum_{y_1^\ell}\Big(\sum_K \mathbb{P}_{\mathbb{K}}(K)\mathbb{P}_{Y|\mathbb{K}}(y_1^\ell \mid K)^{\frac{1}{1+t}}\Big)^{1+t} \\
&\leqslant \sum_{y_1^\ell} \mathbb{P}_Y(y_1^\ell) + q^{\ell t - jt}\Big(\sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_X(x_1^\ell)\mathbb{P}_{Y|X}(y_1^\ell \mid x_1^\ell)^{1+s}\Big)^{1+t} \\
&= 1 + q^{\ell t - jt} \sum_{y_1^\ell}\Big(\sum_{x_1^\ell \in \mathbb{F}_q^\ell} \mathbb{P}_X(x_1^\ell)\mathbb{P}_{Y|X}(y_1^\ell \mid x_1^\ell)^{1+s}\Big)^{1+t} \\
&= 1 + q^{\ell t - jt}\exp(-(\textit{the E-null function of } W^\ell)(t)) \\
&= 1 + q^{\ell t - jt}\exp(-\ell E_0(t)).
\end{aligned}$$

All efforts we spent on bounding $I_e(U_1^j \,;\, Y_1^\ell)$ are for three creeds: First, it demonstrates that Gallager's bounds via E-null functions (which behaves like cumulant generating functions) is a powerful tool that can be useful to the dual case. Second, it fits the paradigm that solving the primary (noisy channel) and the dual (wiretap channel) problems as a whole is easier than solving the primary problem alone. Third, the universal quadratic bound can be used to further bound the E-null function.

We infer that

$$\begin{aligned}
\mathbb{E} I_e(U_1^j \,;\, Y_1^\ell \mid \mathbb{G}) &= I_e(\mathbb{K} \,;\, Y_1^\ell) = \mathbb{E}_0'(0) \leqslant \frac{1}{t}\mathbb{E}_0(t) = \frac{1}{-t}\ln\Big(\exp(-\mathbb{E}_0(t))\Big) \\
&\leqslant \frac{1}{-t}\ln\Big(1 + q^{\ell t - jt}\exp(-\ell E_0(t))\Big) < \frac{1}{-t}q^{\ell t - jt}\exp(-\ell E_0(t)) \\
&= \exp(-\ln(-t) + (\ell - j)t\ln q - \ell E_0(t)).
\end{aligned}$$

Recall the universal quadratic bound $E_0(t) \geqslant I(W)t\ln q - t^2\ln(q)^2$ as stated in Lemma 6.14 and used in the previous subsection. But this time $-2/5 \leqslant t < 0$. We obtain that the exponent is

$$\begin{aligned}
&-\ln(-t) + (\ell - j)t\ln q - \ell E_0(t)\\
&\quad= -\ln(-t) + (\ell - H(W)\ell + \ell^{1/2+\alpha})t\ln q - \ell E_0(t)\\
&\quad= -\ln(-t) + (I(W)\ell + \ell^{1/2+\alpha})t\ln q - \ell E_0(t)\\
&\quad\leqslant -\ln(-t) + (I(W)\ell + \ell^{\frac{1}{2}+\alpha})t\ln q - \ell(I(W)t\ln q - t^2\ln(q)^2)\\
&\quad= -\ln(-t) + (\ell t\ln q + \ell^{1/2+\alpha})t\ln q
\end{aligned}$$

(redeem the inequality at $t = -\ell^{-1/2+\alpha}/2\ln q$)

$$\begin{aligned}
&\mapsto -\ln\Big(\frac{\ell^{-1/2+\alpha}}{2\ln q}\Big) - \Big(-\frac{\ell\ell^{-1/2+\alpha}}{2} + \ell^{1/2+\alpha}\Big)\frac{\ell^{-1/2+\alpha}}{2}\\
&= \frac{\ln\ell}{2} - \alpha\ln\ell + \ln 2 + \ln(\ln q) - \frac{\ell^{2\alpha}}{4}\\
&= \frac{\ln\ell}{2} - \ln(\ln\ell) + \ln 2 + \ln(\ln q) - \frac{\ell^{2\ln(\ln\ell)/\ln\ell}}{4}\\
&< \frac{\ln\ell}{2} + \ln(\ln q) - \frac{\ln(\ell)^2}{4}.
\end{aligned}$$

The first inequality uses $\ell - j = \ell - H(W)\ell + \ell^{1/2+\alpha}$. The last inequality uses the assumption $\ell > e^2$. With the last line we conclude that $\mathbb{E}I_e(U_1^j\,;\,Y_1^\ell \mid \mathbb{G}) < \exp(\ln(\ell)/2 + \ln\ln q - \ln(\ell)^2/4) = \ell^{1/2-\ln(\ell)/4}\ln q$. Switch back to the base-q mutual information $\mathbb{E}I(U_1^j\,;\,Y_1^\ell \mid \mathbb{G}) < \ell^{1/2-\ln(\ell)/4}$.

We now reject kernels $\mathbb{G}$ such that $I(U_1^j\,;\,Y_1^\ell \mid \mathbb{G}) \geqslant \ell^{1/2-\ln(\ell)/5}$. By Markov's inequality, the opposite direction ($<$) holds with probability $1 - \ell^{-\ln(\ell)/20}$ because $1/5 + 1/20 = 1/4$. Plug this upper bound into h. The left-hand side of inequality (6.12) is less than

$$\begin{aligned}
jh\Big(\frac{1}{j}\ell^{1/2-\ln(\ell)/5}\Big) &= jj^{-\alpha}\ell^{\alpha/2-\alpha\ln(\ell)/5} < \ell^{1-\alpha}\ell^{\alpha/2-\alpha\ln(\ell)/5}\\
&= \ell^{1-\alpha/2-\alpha\ln(\ell)/5} = \ell\ln(\ell)^{-1/2-\ln(\ell)/5}.
\end{aligned}$$

The inequality uses that the left-hand side increases monotonically in j and $j \coloneqq H(W)\ell - \ell^{1/2+\alpha} < \ell$. The quantity at the end of the inequalities decays to 0 as $\ell \to \infty$, so eventually it becomes less than $\ell^{1/2+\alpha}$, the right-hand side of inequality (6.12). This proves that inequality (6.7) holds with failing probability $\ell^{-\ln(\ell)/20}$ as soon as ℓ is large enough.

The lower bound on ℓ in the statement of Lemma 6.6 is large enough, hence inequality (6.7), the second half of Lemma 6.6 settled. That means the proof of the whole Theorem 6.7 is complete. ∎

We just finished the last piece of the proof of Theorem 6.7, which states that random kernels possesses good ϱ with high probability. The next section combines this fact with the coset distance profile $\lceil j^2/\ell\rceil$ to conclude that low-complexity nearly-optimal polar codes exist when $\pi + 2\rho \to 1$.

9. Chapter Conclusion

Theorem 6.2 shows that, with high probability, a random kernel enjoys coset distance profile $D_Z^{(j)} \geqslant \lceil j^2/\ell \rceil$. And its dual $D_S^{(\ell-j+1)} \geqslant \lceil j^2/\ell \rceil$ is immediate. Theorem 6.7 shows that, with high probability, a random kernel enjoys eigenvalue $\ell^{-1/2+4\alpha}$, which means $\rho = 1/2 - 4\alpha$, where $\alpha \coloneqq \ln \ell / \ln(\ln \ell)$. Now, Does any pair (π, ρ) lying under the line $\pi + 2\rho = 1$ lie to the left of the convex envelope of $(0, 1/2 - 4\alpha)$ and the Cramér function of $\log_\ell \lceil J_1^2/\ell \rceil$ for some large ℓ?

Let us not go all the way down to a derivation of the Cramér function of $\log_\ell \lceil J_1^2/\ell \rceil$; a one-sided bound suffices. Consider the moment generating function evaluated at slope $-1/2$ (that is the slope of $\pi + 2\rho = 1$):

$$\begin{aligned}
E[D_1^{-1/2}] &= \frac{1}{\ell}\sum_{j=1}^{\ell}\Bigl(\Bigl\lceil \frac{j^2}{3\ell}\Bigr\rceil\Bigr)^{-1/2} < \frac{1}{\ell}\sum_{j=1}^{\lfloor\sqrt{3\ell}\rfloor} 1^{-1/2} + \frac{1}{\ell}\sum_{j=\lfloor\sqrt{3\ell}\rfloor+1}^{\ell}\Bigl(\frac{j^2}{3\ell}\Bigr)^{-1/2} \\
&< \frac{1}{\ell}\sqrt{3\ell} + \frac{1}{\ell}\sqrt{3\ell}\int_{\sqrt{3\ell}}^{\ell}\frac{\mathrm{d}j}{j} = \sqrt{3\ell} + \frac{\sqrt{3}}{\sqrt{\ell}}\ln j\Big|_{\sqrt{3\ell}}^{\ell} < \sqrt{3\ell} + \sqrt{3\ell}\ln \ell \\
&= \ell^{-1/2} + 2\ell^{-1/2+\alpha} < 4\ell^{-1/2+\alpha} < \ell^{-1/2+2\alpha}.
\end{aligned}$$

So the cumulant generating function is bounded as

$$K(-1/2) = \log_\ell E[D_1^{-1/2}] < -1/2 + 2\alpha.$$

The Cramér function as a supremum is then bounded by

$$L(s) \geqslant s \cdot (-1/2) - K(-1/2) \geqslant -s/2 + 1/2 - 2\alpha.$$

The convex envelope of $(0, 1/2 - 4\alpha)$ and the segment $\rho = -\pi/2 + 1/2 - 2\alpha$ (note that only the part with $\rho \geqslant 0$ counts) is a straight line connecting $(0, 1/2 - 4\alpha)$ and $(1 - 4\alpha, 0)$. As ℓ goes to infinity, α goes to 0, hence the convex envelope reveals the right triangle $(0, 1/2)$–$(0, 0)$–$(1, 0)$.

By Chapter 5, any (π, ρ) in this right triangle is realizable by some polar codes, with complexity $O(N \log N)$. Let me state the full theorem here.

Theorem 6.15 (Hypotenuse). *Fix a q-ary channel W. Fix exponents $\pi + 2\rho < 1$. Then there exists a large ℓ and an amoebic kernel (a strategy to assign kernels to synthetic channels) $\mathcal{G}$ such that*

$$P\{Z_n < \exp(-\ell^{\pi n})\} > 1 - H(W) - \ell^{-\rho n},$$
$$P\{S_n < \exp(-\ell^{\pi n})\} > H(W) - \ell^{-\rho n}$$

for large n.

Recall that at the beginning of Chapter 5, I have demonstrated how to reduce arbitrary input alphabet to prime-power input alphabet. Hence the theorem applies to all DMCs, and we have reached the holy grail at the beginning of the chapter.

Corollary 6.16 (Random codes' durability, polar codes' simplicity). *Over any discrete memoryless channel, for any constants $\pi, \rho > 0$ such that $\pi + 2\rho < 1$, there is a series of error correcting codes with block length N approaching infinity, block error probability $\exp(-N^\pi)$, code rate $N^{-\rho}$ less than the channel capacity, and encoding and decoding complexity $O(N \log N)$ per code block.*

TABLE 6.1. The references to the various performances of polar codes over various channels. The starred behaviors are weak—only requiring π or ρ or both to be positive.

	Symmetric					Asymmetric	
	BEC	SBDMC	p-ary	q-ary	finite	BDMC	finite
LLN	[Ari09]	[Ari09]	[ŞTA09]	[ŞTA09]	[ŞTA09]	[SRDR12]	Cor 6.16
LDP*	[AT09]	[AT09]	[ŞTA09]	[MT10]	[Sas11]	[HY13]	Cor 6.16
CLT*	[KMTU10]	[HAU14]	[BGN+18]	Cor 6.16	Cor 6.16	Cor 6.16	Cor 6.16
MDP*	[GX15]	[GX15]	[BGS18]	Cor 6.16	Cor 6.16	Cor 6.16	Cor 6.16
LDP	[KSU10]	[KSU10]	Cor 6.16	Cor 6.16	Cor 6.16	Cor 6.16	Cor 6.16
CLT	[FHMV18]	[GRY20]	Cor 6.16	Cor 6.16	Cor 6.16	Cor 6.16	Cor 6.16
MDP	Cor 6.16	Cor 6.16	Cor 6.16	Cor 6.16	Cor 6.16	Cor 6.16	Cor 6.16

This is the second-moment paradigm code that was promised in the abstract. Table 6.1 compares this corollary to past works. The same can be stated for lossless compression and lossy compression.

Corollary 6.17 (Very good code for lossless compression)**.** *For any lossless compression problem, for any constants $\pi, \rho > 0$ such that $\pi + 2\rho < 1$, there is a series of source codes with block length N approaching infinity, block error probability $\exp(-N^\pi)$, code rate $N^{-\rho}$ plus the conditional entropy, and encoding and decoding complexity $O(N \log N)$ per code block.*

Corollary 6.18 (Very good code for lossy compression)**.** *For any lossy compression problem, for any constants $\pi, \rho > 0$ such that $\pi + 2\rho < 1$, there is a series of source codes with block length N approaching infinity, block error probability $\exp(-N^\pi)$, code rate $N^{-\rho}$ plus the test channel capacity, and encoding and decoding complexity $O(N \log N)$ per code block.*

The next chapter prunes.

CHAPTER 7

Joint Pruning and Kerneling

COMBINATION of two techniques, when executed properly, gives results that inherit advantages from the two individual techniques. In this chapter, I want to combine pruning from Chapter 4 and random dynamic kerneling from Chapter 6, and will verify that the chimera codes have the optimal gap to capacity and log-logarithmic complexity.

Let me elaborate. First, we do not expect the resulting codes to have elpin error, that is, block error probability $\exp(-\ell^{\pi n})$. This is because we set the threshold $\theta \coloneqq N^{-2}$ and prune the channel tree whenever Z_m reaches θ. Should we wait for Z_m to become as small as $\theta \coloneqq \exp(-\ell^{\pi n})$, pruning will not take place at $O(\log n)$ depth, and there will be little to no savings on EU–DU pairs. In conclusion, there is a conflict between elpin error and log-logarithmic complexity, and I will give up the optimal decay of error to favor low complexity.

That being said, theer is no obvious conflict between complexity and code rate, and it is very likely that we can retain both from Chapter 4 and Chapter 6, respectively. Hence this constituents the goal for this section—Construct error correcting codes with gap to capacity close to $N^{-1/2}$, encoding and decoding complexity $O(N \log(\log N))$ per block, and block error probability as small as possible.

1. Toolbox Checklist

From Chapter 6, it is possible to construct a channel process $\{\mathsf{W}_m\}$ such that, for any $\pi + 2\rho < 1$ and large m (depending on π, ρ),

$$P\{\mathsf{Z}_m < e^{-\ell^{\pi m}}\} > 1 - H(W) - \ell^{-\rho m}, \tag{7.1}$$

$$P\{\mathsf{S}_m < e^{-\ell^{\pi m}}\} > H(W) - \ell^{-\rho m}. \tag{7.2}$$

From Chapter 5, we know that every ergodic kernel assumes a positive ϱ, so every kernel assumes at least a pair $\pi, \rho > 0$ such that inequalities (7.1) and (7.2) hold.

From Chapter 4, we can set a threshold θ, which serves two purposes: One, we prune the channel tree at the point where Z_m or S_m falls below θ. This in turn defines the stopping time

$$\mathsf{s} \coloneqq n \wedge \min\{m : \min(Z_{\text{mxd}}(\mathsf{W}_m), S_{\max}(\mathsf{W}_m)) < \theta\}.$$

Two, we collect in $\mathcal{J}$ indices that point to synthetic channels W_m whose Z_{mxd}-parameter reaches θ, which is the cause why s is set to this depth m.

For the asymmetric case, θ defines the stopping time

$$\mathsf{s} \coloneqq n \wedge \min\left\{m : \begin{matrix}\min(Z_{\text{mxd}}(\mathsf{W}_m), S_{\max}(\mathsf{W}_m)) < \theta \text{ and}\\ \min(Z_{\text{mxd}}(\mathsf{Q}_m), S_{\max}(\mathsf{Q}_m)) < \theta\end{matrix}\right\}.$$

$\mathcal{J}$ will then collect indices pointing to W_m whose Z_{mxd}-value reaches θ and to Q_m whose $S_{\max}$-value reaches θ. To put it differently, $\mathcal{J}$ is the event where both $Z_{\text{mxd}}(\mathsf{W}_\mathsf{s}), S_{\max}(\mathsf{Q}_\mathsf{s}) < \theta$.

θ and s determine a code. We have three lemmas that generalize Lemmas 4.1 to 4.3 to arbitrary matrix kernels. Their proofs are straightforward (perhaps tautological) and will be sketchy.

Lemma 7.1 (Complexity in terms of s). *The encoding and decoding complexity is $O(\mathsf{E}[\mathsf{s}])$ per channel usage, or $O(N\mathsf{E}[\mathsf{s}])$ per code block.*

Proof. Similar to Lemma 4.1, I claim without a proof that the encoding and decoding complexity is proportional to the number of EU–DU devices in the circuit and to the number of synthetic channels (multiplicity included) that undergo channel transformation.

Since a trajectory $\mathsf{W}_0, \mathsf{W}_1, \ldots, \mathsf{W}_\mathsf{s}$ undergoes the transformation s times, the average number of transformations is $\mathsf{E}[\mathsf{s}]$, and hence the total number of transformations is $N\mathsf{E}[\mathsf{s}]$. □

Lemma 7.2 (R in terms of $\mathcal{J}$). *The code rate is $\mathsf{P}\{\mathsf{J}_1^\mathsf{s} \in \mathcal{J}\}$, or $\mathsf{P}(\mathcal{J})$ for short.*

Proof. Similar to Lemma 4.2, I claim without a proof that the code rate is the density of naked pins that are selected in $\mathcal{J}$.

Every pair of naked pins possesses probability measure $1/N$ because there are always N pairs of naked pins. Every synthetic channel W_s assumes $N/2^\mathsf{s}$ copies in the circuit, hence possesses probability measure $2^{-\mathsf{s}}$. All indices J_1^s in $\mathcal{J}$ possesses probability measure 2^{-m}, which coincides with the measure of pins. Thus the density of selected pins is $\mathsf{P}\{\mathsf{J}_1^\mathsf{s} \in \mathcal{J}\}$. □

Lemma 7.3 (P_e in terms of $\mathcal{J}$). *The block error probability is $N\mathsf{E}[P_\mathrm{e}(\mathsf{W}_\mathsf{s}) \cdot \mathsf{I}\{\mathsf{J}_1^\mathsf{s} \in \mathcal{J}\}] \leqslant qN\theta$ for the symmetric case. For the asymmetric case, the block error probability is $N\mathsf{E}[P_\mathrm{e}(\mathsf{W}_\mathsf{s}) \cdot \mathsf{I}\{\mathsf{J}_1^\mathsf{s} \in \mathcal{J}\}] \leqslant qN\theta + N\mathsf{E}[T(\mathsf{W}_\mathsf{s}) \cdot \mathsf{I}\{\mathsf{J}_1^\mathsf{s} \in \mathcal{J}\}]$, witch is at most $3qN\theta$ in total.*

Proof. Similar to Lemma 4.3, I claim without a proof that, for the symmetric case, the block error probability is the sum of the P_e-values of the W_s in $\mathcal{J}$ (multiplicity included). And for the symmetric case, the block error probability is the said sum plus the total of the T-values of the Q_s in $\mathcal{J}$ (multiplicity included).

For the sum of $P_\mathrm{e}(\mathsf{W}_\mathsf{s})$, we learn from Lemma 5.10 that $P_\mathrm{e}(\mathsf{W}_\mathsf{s}) \leqslant qZ(\mathsf{W}_\mathsf{s})/2 \leqslant qZ_\mathrm{mxd}(\mathsf{W}_\mathsf{s})/2 \leqslant q\theta/2$. Hence a sum of at most N items, each at most $q\theta/2$, is at most $qN\theta$.

For the sum of $T(\mathsf{Q}_\mathsf{s})$, we learn form Lemma 5.11 that

$$\begin{aligned} T(\mathsf{Q}_\mathsf{s}) &\leqslant \frac{2(q-1)}{q} - \frac{2}{q}\Big((q-1)qP_\mathrm{e}(W) - (q-1)(q-2)\Big) \\ &= \frac{2(q-1)}{q}\Big(1 - \big(qP_\mathrm{e}(\mathsf{Q}_\mathsf{s}) - (q-2)\big)\Big) = \frac{2(q-1)^2}{q}\Big(1 - \frac{q}{q-1}P_\mathrm{e}(\mathsf{Q}_\mathsf{s})\Big) \end{aligned}$$

and then from Lemma 5.12 that

$$\leqslant \frac{2(q-1)^2}{q} S(\mathsf{Q}_n) \leqslant 2qS(\mathsf{Q}_n) \leqslant 2qS_\mathrm{max}(\mathsf{Q}_n) \leqslant 2q\theta.$$

Hence a sum of at most N items, each at most $2q\theta$, is at most $2qN\theta$. The contributions of $P_\mathrm{e}(\mathsf{W}_\mathsf{s})$ and $T(\mathsf{Q}_n)$ amount to $3qN\theta$; that finishes the proof. □

Now I can state and prove the main theorem in this chapter. My contribution is twofold. First contribution: If you insist on using a certain matrix G to construct polar codes, then either G is not ergodic (in which case there is no polarization at all), or you can construct log-logarithm codes with some positive ρ. Second contribution: If you allow dynamic kerneling with very large ℓ, then you can construct log-logarithm codes whose ρ is arbitrarily close to $1/2$, the optimal exponent.

2. Log-logarithmic Codes

The first case I want to discuss here is when a fixed kernel G is selected prior. In this case, the best ρ we can hope for is the ϱ in the eigen/en23/een13 behavior of G. (After all, pruning should not improve the extent of polarization.)

Theorem 7.4 (Log-log for asymmetric q-ary)**.** *Fix a q-ary channel W. Fix a kernel $G \in \mathbb{F}_q^{\ell\times\ell}$ and a pair (π,ρ) that satisfy the two-sided elpin behavior*

$$P\{Z_m < e^{-\ell^{\pi m}}\} > 1 - H(W) - \ell^{-\rho m + o(m)},$$
$$P\{S_m < e^{-\ell^{\pi m}}\} > H(W) - \ell^{-\rho m + o(m)}.$$

Then pruning the channel tree with threshold $\theta \coloneqq 1/3qN^2$ yields

$$E[s] = O(\log(\log N)),$$
$$P(\mathcal{J}) = I(W) - N^{-\rho+o(1)},$$
$$NE[P_{\mathrm{e}}(W_s)\cdot I(\mathcal{J})] + NE[T(Q_s)\cdot I(\mathcal{J})] \leqslant 1/N.$$

Proof. To estimate $E[s] = \sum_{m=1}^{n-1} P\{s > m\}$, we mimic Theorem 4.4. Consider small m and large m. Those with $\exp(-\ell^{\pi m}) \geqslant \theta$ are called small m. Those with $\exp(-\ell^{\pi m}) < \theta$ are called large m. For small m, we do not expect decent polarization and assume $s \geqslant m$. That is, we upper bound $P\{s > m\} \leqslant 1$ by a pessimistic value. For large m, we argue that

$$\begin{aligned}
P\{s > m\} &\leqslant P\{Z_{\mathrm{mxd}}(W_m) \vee S_{\max}(W_m)) \geqslant \theta \text{ or } Z_{\mathrm{mxd}}(Q_m) \vee S_{\max}(Q_m) \geqslant \theta\} \\
&\leqslant P\{\max(Z_{\mathrm{mxd}}(W_m), S_{\max}(W_m)) \geqslant \theta\} + P\{\max(Z_{\mathrm{mxd}}(Q_m), S_{\max}(Q_m)) \geqslant \theta\} \\
&\leqslant P\{Z_{\mathrm{mxd}}(W_m) \vee Z_{\mathrm{mxd}}(W_m) \geqslant e^{-\ell^{\pi m}}\} + P\{S_{\max}(Q_m) \vee S_{\max}(Q_m) \geqslant e^{-\ell^{\pi m}}\} \\
&\leqslant 2\ell^{-\rho m + o(m)} + 2\ell^{-\rho m + o(m)} = \ell^{-\rho m + o(m)}.
\end{aligned}$$

As a result, the complexity is

$$\begin{aligned}
E[s] &= \sum_{m=0}^{n-1} P\{s > m\} = \#\{\text{small } m\} + \sum_{\text{large m}} \ell^{-\rho m + o(m)} \\
&= O(\log n) + O(1) = O(\log(\log N)).
\end{aligned}$$

Here we use the fact that the number of small m's is the root of the equation $\exp(-\ell^{\pi m}) = \theta = 1/3qN^2 = 1/3q\ell^{2n}$, which is $\log(n)$ (note that $N = \ell^n$).

To estimate $R = P(\mathcal{J})$, we mimic Theorem 4.5. It is the same as estimating the frequency that s is set to m due to $Z_{\mathrm{mxd}}(W_m) < \theta$ and $S_{\max}(Q_m) < \theta$. This frequency is

$$\begin{aligned}
R = P(\mathcal{J}) &= P\{Z_{\mathrm{mxd}}(W_s) \leqslant \theta \text{ and } S_{\max}(Q_s) \leqslant \theta\} \\
&\geqslant P\{Z_{\mathrm{mxd}}(W_m) \to 0 \text{ and } S_{\max}(Q_m) \to 0\} \\
&\qquad - P\{Z_{\mathrm{mxd}}(W_m) \to 0 \text{ but } Z_{\mathrm{mxd}}(W_s) \geqslant \theta\} \\
&\qquad - P\{S_{\max}(Q_m) \to 0 \text{ but } S_{\max}(Q_s) \geqslant \theta\} \\
&\geqslant I(W) - P\{Z_{\mathrm{mxd}}(W_m) \to 0 \text{ but } S_{\max}(W_s) < \theta\} \\
&\qquad - P\{S_{\max}(Q_m) \to 0 \text{ but } Z_{\mathrm{mxd}}(Q_s) < \theta\} - P\{s = n\}.
\end{aligned}$$

Here we use the fact that if $Z_{\text{mxd}}(W_{\mathsf{s}}) \geqslant \theta$, then s is set to its current value due to $S_{\max}(W_{\mathsf{s}}) < \theta$ or, otherwise, due to hitting n. It remains to estimate the three minus terms on the right-hand side.

To bound $P\{Z_{\text{mxd}}(W_m) \to 0 \text{ but } S_{\max}(W_{\mathsf{s}}) < \theta\}$, notice that $1 - H(W_{\mathsf{s}}) \leqslant q^3\sqrt{S_{\max}(W_{\mathsf{s}})} < q^3/N\sqrt{3q} \leqslant N^{-1+o(1)}$. Now the probability that $H(W_m) \to 0$ given $H(W_{\mathsf{s}}) \geqslant 1 - N^{-1+o(1)}$ is $N^{-1+o(1)}$ by the martingale property. For a similar reason, $P\{S_{\max}(Q_m) \to 0 \text{ but } Z_{\text{mxd}}(Q_{\mathsf{s}}) < \theta\}$ is no greater than the probability that $H(Q_m) \to 1$ given $H(Q_{\mathsf{s}}) < N^{-1+o(1)}$, which is $N^{-1+o(1)}$ by the martingale property again. Last is to bound $P\{\mathsf{s} = n\}$, but that is just $P\{\mathsf{s} > n-1\} - P\{\mathsf{s} > n\}$, and is thus $\ell^{-\rho(n-1)-o(n-1)} = N^{-\rho+o(1)}$. We conclude

$$R \geqslant I(W) - N^{-1+o(1)} - N^{-1+o(1)} - N^{-\rho+o(1)} = I(W) - N^{-\rho+o(1)}.$$

The bound $NE[P_{\text{e}}(W_{\mathsf{s}}) \cdot I(\mathcal{J})] + NE[T(Q_{\mathsf{s}}) \cdot I(\mathcal{J})] \leqslant 1/N$ is tautological. And the proof ends here. □

When allowing dynamic kerneling, random or not, the proof of the last theorem does not change—we still set a threshold θ and prune the channel tree by θ. The only difference is that ρ can now be arbitrarily close to $1/2$.

To put it another way, like I once commented under Lemma 5.34, there is no point to use a large kernel without knowing that it has a good (π, ρ) pair. So Theorem 7.4 mainly has three use cases: (a) To estimate the behavior of a minimalist kernel such as $\left[\begin{smallmatrix} 1 & 0 \\ c & 1 \end{smallmatrix}\right]$, (b) to estimate the behavior of a larger kernel with a bound on ρ, and (c) to estimate the behavior of random dynamic kerenling. That leads to the following corollary.

Corollary 7.5 (Log-log code for DMC)**.** *Given any q-ary DMC (presumably with virtual symbols) and any ergodic kernel $G \in \mathbb{F}_q^{\ell\times\ell}$, pruned polar coding achieves encoding and decoding complexity $O(\log(\log N))$ per channel usage, block error probability $1/N$, and code rate $1/N^\rho$ less than the channel . Here, ρ is a number guaranteed to be positive, lower bounded if you know more about the eigen/en23/een13/eplin behavior of G, and very close to $1/2$ as $\ell \to \infty$ if you allow dynamic kerneling.*

The same can be stated regarding lossless and lossy compression, and is omitted.

The next section discusses a continuous trade-off between P_{e} and complexity.

3. Error–Complexity Trade-off

In the proof of Theorem 7.4, the complexity $E[\mathsf{s}]$ and $\theta = \theta(n)$ are linked by the equation $\exp(-\ell^{\pi m}) = \theta$ that determines how many m's are small. The root thereof $m = m(\theta) = m(\theta(n))$ will be the complexity. We may as well alter $\theta(n)$ and obtain a different error–complexity pair. The only restriction is $\ell^{-2n} \ll \theta(n) \ll \exp(-\ell^{\pi n})$ to avoid the necessity to deal with special/edge cases.

Corollary 7.6 (Continuous error–complexity trade-off)**.** *Let $\theta(n)$ be an asymptote lying between ℓ^{-2n} and $\exp(-\ell^{\pi n})$. Then pruned polar coding achieves complexity $O(\log|\log\theta(n)|)$ per channel usage and block error probability $\ell^n\theta(n)$.*

In particular, if $\theta(n) := \exp(-\ell^{\pi n})$, then the complexity is $O(\log|\log\theta(n)|) = O(\log|\ell^{\pi m}|) = O(n) = O(\log N)$. This restores the case where one insists on retaining the $\exp(-N^\pi)$ error while pruning, which only gives you constant-scalar

TABLE 7.1. A comparison concerning the error–rate–complexity asymptotes of some well-known capacity-achieving codes.

Code	Error	Gap	Complexity	Channel
random	e^{-N^π}	$N^{-\rho}$	$\exp(N)$	DMC
concatenation	e^{-N^π}	$\to 0$	$\mathrm{poly}(N)$	DMC
RM	$\to 0$	$\to 0$	$O(N^2)$	BEC
LDPC	$\to 0$	$\to 0$	unclear	SBDMC
RA family	$\to 0$	$\to 0$	$O(1)$	BEC
MDP-polar	e^{-N^π}	$N^{-\rho}$	$O(\log N)$	DMC
loglog-polar	e^{-n^τ}	$N^{-\rho}$	$O(\log(\log N))$	DMC

improvement in complexity. In fact, we can almost prove this tight: The complexity, $\mathsf{E}[\mathsf{s}]$, is the (average) number of transformations a trajectory W_n undergoes. Since each transformation, at best, raises $Z_{\mathrm{mxd}}(\mathsf{W}_n)$ to the power of ℓ, you need $\log_\ell(\log_{Z_{\mathrm{mxd}}(W)}\theta)$ transformations to lower $Z_{\mathrm{mxd}}(\mathsf{W}_n)$ to θ. That implies that $\mathsf{E}[\mathsf{s}] \geqslant \log_\ell(\log_{Z_{\mathrm{mxd}}(W)}\theta)$.

On the other hand, if $\theta(n) \coloneqq \exp(-n^\tau)$ for some very large $\tau > 0$, then $\mathsf{E}[\mathsf{s}] = O(\log|\log\theta(n)|) = O(\log|-n^\tau|) = O(\tau\log(\log N))$. If τ is a constant despite of being very large, then the complexity is still $O(\log(\log N))$. This is the complexity paradigm code that was promised in the abstract. Table 7.1 compare this result and Corollary 6.16 to past works.

The same can be stated concerning lossless and lossy compression, and is omitted.

The next chapter generalizes the second-moment paradigm and complexity paradigm to some network coding scenarios.

CHAPTER 8

Distributed Lossless Compression

Network coding emerges as a generalization of one-to-one communication as there are, naturally, more than one party willing to participate. One of the easiest scenarios (in comparison with other network scenarios, not necessarily easy compared to one-to-one) is when there are more than two random sources to be compressed, each by a compressor that does not talk the other compressors, and a decompressor will gather all messages and reconstruct the original random sources. This scenario is referred to as *distributed compression*.

Distributed compression can be further divided into several sub-scenarios that are treated differently. Depending on whether a reconstruction needs to be faithful or can be fuzzy, there are distributed *lossless* compression and its *lossy* variants. Depending on whether a random source needs to be reconstructed or it provides side information to the other sources, the responsible compressor is called a *sender* or a *helper*. To summarize, there are three types of sources—those that need to be reconstructed as is, those whose reconstruction can be less accurate, and those that need no reconstruction at all. And a distributed compression problem consists of any combination of theses three sources.

Thanks to the infrastructures built in the past three chapters, if we manage to reduce a network coding problem to several one-to-one problems, then each of the one-to-one problems can be solved by polar coding that achieves capacity at a good pace and with low complexity. In this chapter, I will overview distributed lossless compression problems with two senders, one sender plus one helper, three senders, and finally many senders plus one helper. These are the problems whose rate region is known [**EGK11**], thus it makes sense to pursue the second-order behavior of the rate tuples.

1. Slepian–Wolf: The Two Sender Problem

A *Slepian–Wolf problem* is a distributed lossless compression problem with two senders, which is the first of the several cases we will consider. See Figure 8.1 for the specification. In this and the other scenarios, the "array access" operator $[\bullet]$ will be used to distinguish random sources. The pair $(R[1], R[2])$ will be called the *rate pair*. The N is still the block length. And P_{e} is the block error probability, the probability that either $\hat{X}[1]_1^N \neq X[1]_1^N$ or $\hat{X}[2]_1^N \neq X[2]_1^N$.

The *rate region* of a Slepian–Wolf problem is a region in $\mathbb{R}^2$ of all achievable rate pairs $(R[1], R[2])$. There are clearly three pessimistic criteria:

- About $H(X[1] \mid X[2])$ bits of information is only available at source[1], hence compressor[1] should at least send out this much information, $R[1] \geqslant H(X[1] \mid X[2])$.

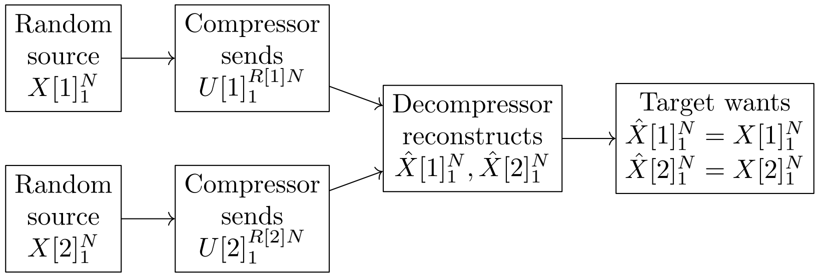

FIGURE 8.1. A Slepian–Wolf problem.

- About $H(X[2] \mid X[1])$ bits of information is only available at source[2], hence compressor[2] should at least send out this much information, $R[2] \geqslant H(X[2] \mid X[1])$.
- About $H(X[1]X[2])$ bits of information are generated in total, hence the two compressors should at least send out this many bits in total, $R[1] + R[2] \geqslant H(X[1]X[2])$.

As it turns out, these necessary criteria are sufficient.

Theorem 8.1 (Slepian–Wolf Theorem)**.** [**SW73**] *The rate region of a Slepian–Wolf problem consists of pairs* $(R[1], R[2]) \in \mathbb{R}^2$ *such that*

$$\begin{aligned} R[1] &\geqslant H(X[1] \mid X[2]), \\ R[2] &\geqslant H(X[2] \mid X[1]), \\ R[1] + R[2] &\geqslant H(X[1]X[2]), \end{aligned}$$

and is supported by the vertices

$$\big(H(X[1] \mid X[2]),\, H(X[2])\big) \quad \textit{and} \quad \big(H(X[1]),\, H(X[2] \mid X[1])\big).$$

See also Figure 8.2.

Notation: When the context is clear, $H(1), H(2), H(12), H(1|2), H(2|1)$ denote the corresponding (conditional) entropies with "$X[\bullet]$" wrapping around every Arabic number.

The problem here is, How fast can $R \coloneqq (R[1], R[2])$ approach the boundary that passes $(H(1|2), +\infty)$–$(H(1|2), H(2))$–$(H(1), H(2|1))$–$(+\infty, H(2|1))$? Random coding allowed, this was discussed in [**TK12**] for the CLT regime. More elaborately, fixing a P_e, the (Euclidean) distance from R to the boundary scales as $O(1/\sqrt{N})$. Although there seems to be no references for the LDP and MDP regimes, we may presume that it obeys the same law as in the one-to-one case—namely,

$$\frac{-\ln P_\mathrm{e}}{\mathrm{dist}(R, \mathrm{boundary})^2} \approx N.$$

And we thus pose to ourselves a challenge about constructing polar codes (or any low-complexity codes) with $P_\mathrm{e} \approx \exp(-N^\pi)$ and dist $\approx N^{-\rho}$ whenever $\pi + 2\rho < 1$.

An obvious strategy is to execute time-sharing, which is based on this simple idea: If we know how to achieve the point $(H(1), H(2|1))$ using a coding scheme and the point $(H(1|2), H(2))$ using another coding scheme, then we can alternate

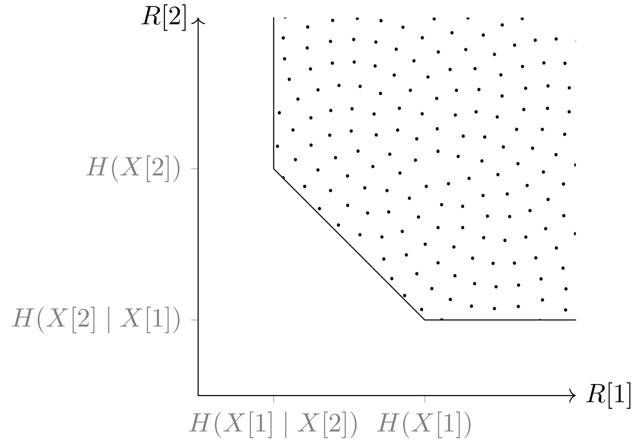

FIGURE 8.2. Slepian–Wolf problem's rate region.

between two said schemes to approach any point lying on the *sum-rate segment* $(H(1), H(2|1))$–$(H(1|2), H(2))$. And this time-sharing scheme, indeed, reaches the second moment goal in part.

Theorem 8.2 (Timed Slepian–Wolf)**.** *Let B be any point on the boundary of the rate region. Fix exponents $\pi + 2\rho < 1$. Then combining polar coding and time-sharing yields $P_{\mathrm{e}} < \exp(-N^{\pi/(1+\rho/2)})$ and $\operatorname{dist}(R, B) < N^{-\rho/(1+\rho/2)}$ at the cost of $O(\log N)$ complexity per source observation. (Notice the penalty $(1 + \rho/2)$.)*

PROOF. Let us first eliminate some trivial cases. Case one: If B is on the vertical ray $(H(1|2), +\infty)$–$(H(1|2), H(2))$, then the problem reduces to achieving the nontrivial vertex $(H(1|2), H(2))$. For this vertex, ask compressor[2] to compress $X[2]$ and ask compressor[1] to compress $X[1]$ with side information $X[2]$, both using polar coding. Note that compressor[1] need not, and cannot, access the side information $X[2]$.

Case two: If B is on the horizontal ray $(H(1), H(2|1))$–$(+\infty, H(2|1))$, then the problem reduces to achieving $(H(1), H(2|1))$. For this vertex, ask compressor[1] to compress $X[1]$ and ask compressor[2] to compress $X[2]$ with side information $X[1]$. For this case and case one, the problem reduces to one-to-one lossless compression and can be solved by polar coding within the specified gap to boundary/block error probability/complexity.

Case three: Assume that B is lying on the sum-rate segment $(H(1|2), H(2))$–$(H(1), H(2|1))$ and is a rational combination of the two end points. That is, there exist positive integers s, t such that $(s + t)B = s(H(1|2), H(2)) + t(H(1), H(2|1))$. Then this is what we do: For every $s + t$ code blocks, apply the coding scheme in case one for s blocks and then apply the coding scheme in case two for t blocks. Note that s and t are fixed constants, so the penalties imposed on N, P_{e}, $\operatorname{dist}(R, B)$, and the complexity are all constant scalars.

Now we deal with the nontrivial case. Case four: Assume B is on the sum-rate segment $(H(1|2), H(2))$–$(H(1), H(2|1))$ and is an irrational combination of the two ends. Then this is what we do: Prepare the case-one scheme and case-two scheme with a large block length M. By the polar coding infrastructure in the past few chapters, the gap to entropy is $M^{-\rho}$ and the error is $\exp(-M^{\pi})$.

Here comes the punchline for case four: Pick large integers $s,t>0$ such that $s+t$ is about the size of $M^{\rho/2}$ and B is very close to (measured in the Euclidean distance)

$$\frac{s}{s+t}(H(1|2),H(2))+\frac{t}{s+t}(H(1),H(2|1)). \tag{8.1}$$

In other words, we use the denominator $s+t\approx M^{\rho/2}$ to approximate the irrational coefficients in the combination. Now, for every $s+t$ code blocks, we apply the case-one scheme for s blocks and then apply the case-two scheme for t blocks. Unlike the rational case, where the penalty is constant scalars, the penalty here scales as M grows. Hence, in particular, the de facto block length is $N=(s+t)M\approx M^{1+\rho/2}$. The error in terms of the de facto block length is $\exp(-M^{\pi})=\exp(-N^{\pi/(1+\rho/2)})$, which suggests that the "de facto pi" is $\pi/(1+\rho/2)$. And the gap to boundary caused by the imperfect coding is $M^{-\rho}=N^{-\rho/(1+\rho/2)}$, which suggests a "de facto rho" of $\rho/(1+\rho/2)$.

But coding is not the only cause of the gap. Approximation (8.1) is very far away from B. In fact, by the Thue–Siegel–Roth theorem and its converse, the difference between an irrational number and its rational approximation is roughly the inverse square of the denominator, unless the irrational number lies in a measure-zero set. As a consequence, we almost alwayse have

$$\operatorname{dist}(B,\text{approximation (8.1)})=\Theta((s+t)^{-2})=\Theta(M^{-\rho})=\Theta(N^{-\rho/(1+\rho/2)}).$$

This gap is comparable to the coding gap, so the overall gap is still $O(N^{-\rho/(1+\rho/2)})$. That finishes the proof. $\square$

Remark: the de facto pi and rho satisfy $2\cdot\pi/(1+\rho/2)+5\cdot\rho/(1+\rho/2)=(2\pi+4\rho+\rho)/(1+\rho/2)<(2+\rho)/(1+\rho/2)=2$. Thus the region of de facto pi–rho pairs is strictly smaller than $\pi+2\rho<1$.

Bibliographical remark: It is once suggested that a Slepian–Wolf problem can be solved by *one* polar code via a technique called *monotonic chain rule* [**Bil12**]. However, the CLT aspect of the monotonic chain rule is as capable as time-sharing; in fact, its CLT behavior is worse than my estimate here because the denominator therein can only be a power of ℓ. It would not help us cancel the $(1+\rho/2)$ penalty.

In the next section, I borrow a technique that avoids approximating an irrational number using rational numbers. Intuitively speaking, this technique tunes the distribution of random variables (note that probabilities are real numbers, mostly irrational) to attain any necessary irrational number.

2. Slepian–Wolf via Source-Splitting

Source-splitting, in one sentence, divides the randomness carried by $X[2]$ into two random variables $X[2]\langle 1\rangle$ and $X[2]\langle 2\rangle$ and then uses them to sandwich $X[1]$. By choosing a proper configuration of $X[2]\langle 1\rangle$ and $X[2]\langle 2\rangle$, we can attain any irrational combination on the sum-rate segment without referring to the time-sharing technique.

In more detail, there will be $X[2]\langle 1\rangle$ and $X[2]\langle 2\rangle$ and a global "knob variable" Q satisfying the axioms:

- Q is independent of $X[1]X[2]$, i.e., the knob is a purely artificial variable;
- $H(X[2]\langle 1\rangle X[2]\langle 2\rangle \mid X[2]Q)=0$, i.e., the knob fully controls how $X[2]$ is split;

- $H(X[2] \mid X[2]\langle 1\rangle X[2]\langle 2\rangle) = 0$, i.e., piecing together the fragments of $X[2]$ yields the complete $X[2]$; and
- $H(X[2]\langle 1\rangle \mid Q)$, when Q is tuned properly, varies from 0 to $H(X[2])$, continuously and inclusively.
- (Alternative to the fourth) $H(X[2]\langle 2\rangle \mid Q)$ varies from 0 to $H(X[2])$, continuously and inclusively.

By the axioms, especially the fourth one,

- $H(X[2]\langle 1\rangle \mid X[1]X[2]\langle 2\rangle Q) + H(X[1] \mid X[2]\langle 2\rangle Q) + H(X[2]\langle 2\rangle \mid Q) = H(X[1]X[2] \mid Q)$; and
- $H(X[1] \mid X[2]\langle 2\rangle Q)$ varies from $H(X[1])$ to $H(X[1] \mid X[2])$, continuously and inclusively.

Now we let compressor[1] compress $X[1] \mid X[2]\langle 2\rangle Q$. Or equivalently, let it compress $X[1]$ given side information $X[2]\langle 2\rangle Q$. We also let compressor[2] compress $X[2]\langle 1\rangle \mid X[1]X[2]\langle 2\rangle Q$ and $X[2]\langle 2\rangle \mid Q$. Or equivalently, let it compress $X[2]\langle 1\rangle$ given side information $X[1]X[2]\langle 2\rangle Q$, and then compress $X[2]\langle 2\rangle$ given side information Q. By that $B[1] \coloneqq H(X[1] \mid X[2]\langle 2\rangle Q)$ varies from $H(X[1])$ to $H(X[1] \mid X[2])$ and that $B[2] \coloneqq$ (the sum of the other two conditional entropies) is $H(X[1]X[2]) - B[1]$, we conclude that $B \coloneqq (B[1], B[2])$, as a function in the distribution of Q, can exhaust all points on the sum-rate segment.

The formal statements are as follows.

Definition 8.3. Let random variable $Q \in \{1, 2\}$ be independent of $X[1]X[2]$. Let

$$X[2]\langle 1\rangle \coloneqq \begin{cases} X[2] & \text{if } Q = 1, \\ \spadesuit & \text{if } Q = 2, \end{cases}$$

where $\spadesuit$ is a placeholder symbol that replaces $X[2]$ when Q decides that it should hide $X[2]$. Similarly,

$$X[2]\langle 2\rangle \coloneqq \begin{cases} \spadesuit & \text{if } Q = 1, \\ X[2] & \text{if } Q = 2. \end{cases}$$

To abuse symbols, we can also say $X[2]\langle Q\rangle = X[2]$ and $X[2]\langle 3 - Q\rangle = \spadesuit$. This is called source-splitting or *coded time-sharing*.

Theorem 8.4 (Qed Slepian–Wolf)**.** *Let B be any point on the boundary of the rate region. Fix exponents $\pi + 2\rho < 1$. Then combining polar coding and source-splitting (coded time-sharing) yields $P_{\mathrm{e}} < \exp(-N^{\pi})$ and $\operatorname{dist}(R, B) < N^{-\rho}$ at the cost of $O(\log N)$ complexity per source observation. (Notice the absence of penalty.)*

PROOF. Given the definition of $X[2]\langle 1\rangle$ and $X[2]\langle 2\rangle$, we give

- compressor[2] the task of compressing $X[2]\langle 1\rangle$ given $X[1]X[2]\langle 2\rangle Q$,
- compressor[1] the task of compressing $X[1]$ given $X[2]\langle 2\rangle Q$, and
- compressor[2] the task of compressing $X[2]\langle 2\rangle$ given Q.

By the polar coding infrastructure developed before, these tasks can be done with the specified error and gap to entropy. The only problem is, Does this coding scheme approach the correct entropy pair?

To find out, let

$$B[1] \coloneqq H(X[1] \mid X[2]\langle 2\rangle Q),$$
$$B[2] \coloneqq H(X[2]\langle 1\rangle \mid X[1]X[2]\langle 2\rangle Q) + H(X[2]\langle 2\rangle \mid Q).$$

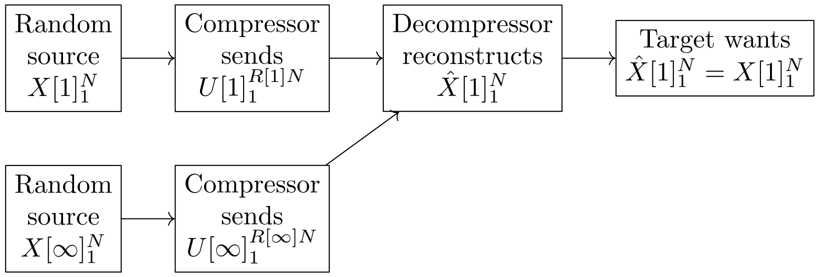

FIGURE 8.3. Lossless compression with one helper.

Then $B[1] + B[2] = H(X[2]\langle 1\rangle X[1]X[2]\langle 2\rangle \mid Q) = H(12)$, so $(B[1], B[2])$ is always on the sum-rate segment. When $E[Q] = 1$, that is, when Q is always 1, we see that $X[2]\langle 2\rangle$ is just a useless constant. In this case, $B[1] = H(X[1] \mid Q) = H(1)$ and thus $B[2] = H(12) - H(1) = H(2|1)$. When $E[Q] = 2$, that is, when Q is always 2, we see that $X[2]\langle 2\rangle = X[2]$. In this case, $B[1] = H(X[1] \mid X[2]Q) = H(1|2)$ and thus $B[2] = H(12) - B[2] = H(2)$.

All in all, when $E[Q]$ varies continuously from 1 to 2, the pair $(B[1], B[2])$ varies continuously from $(H(1), H(2|1))$ to $(H(1|2), H(2))$, and there must be a moment where $B = (B[1], B[2])$. Unless B is on the vertical or horizontal ray, in which case the theorem is trivial. That ends the proof. $\square$

That is how to generalize the second-moment paradigm to Slepian–Wolf problems. Similar statements can be made for $\exp(-n^\tau)$ error and $\log(\log N)$ complexity. The proof will be essentially the same and is omitted.

In the next section, we talk about a variant of Slepian–Wolf where the second source is not of interest, but compressing it helps the decompressor reconstruct the first.

3. Compression with Helper

A *lossless compression problem with a helper* is a distributed lossless compression problem with one sender and one helper. See Figure 8.3 for the specification. The pair $R \coloneqq (R[1], R[\infty])$ is still called the rate pair and the region of possible rate pairs is still called the rate region. The only difference is, this time, the block error probability P_e is the probability that $\hat{X}[1]_1^N \neq X[1]_1^N$.

Similar to the Slepian–Wolf case, the rate region of the one-helper problem can be characterized by an easy observation that "it should satisfy this" and a proof that confirms the observation.

The easy observation is that, if we manage to find a random variable $U[\infty]$ that represents $X[\infty]$ pretty well, then compressor[1] needs only to compress $X[1]$ given $U[\infty]$ while compressor[∞] needs to assure that the decompressor receives $U[\infty]$. For the latter, compressor[∞] sends $I(X[\infty]\,;U[\infty])$ bits per source observation.

Theorem 8.5 (Lossless compression with one helper)**.** [**AK75**] *The rate region for lossless source coding of $X[1]$ with a helper source $X[\infty]$ consists of all pairs $(R[1], R[\infty]) \in \mathbb{R}^2$ such that*

$$R[1] \geqslant H(X[1] \mid U[\infty]),$$

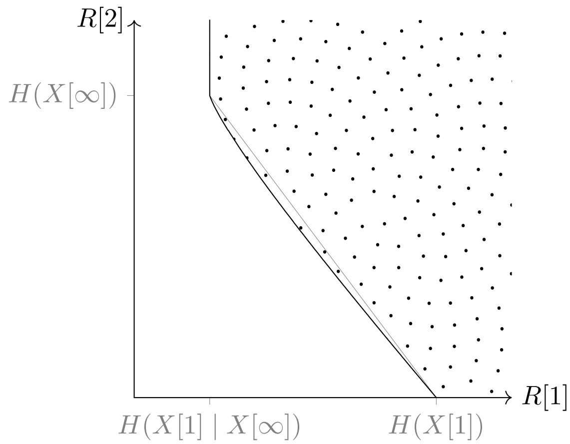

FIGURE 8.4. An example rate region of lossless compression problem with one helper. The gray segment is to contract the curved (unioned) boundary. Here, both $X[1]$ and $X[\infty]$ are uniform binary sources with correlation $I(X[1]\,;X[\infty]) = 3/4$.

$$R[\infty] \geqslant I(X[\infty]\,;U[\infty])$$

unioned over all random variables $U[\infty]$ that depend on $X[\infty]$ but not on $X[1]$. Moreover, it is sufficient to consider the size of the alphabet of $U[\infty]$ that is one plus the size of the alphabet of $X[\infty]$. See also Figure 8.4.

Remark: A subtle detail in the theorem statement is that the rate region is the union of all $(R[1], R[\infty])$, not the *convex hull* of the union of all $(R[1], R[\infty])$. In other words, the rate region for this one-sender one-helper scenario does not need time-sharing to become convex—every point on the boundary is achievable by some clever choice of $U[\infty]$. In fact, the variable $U[\infty]$ can itself be the knob that controls, continuously, the time-sharing coefficient if there are really two schemes to be combined.

Per the remark, we now have a very straightforward scheme to achieve the second moment behavior for this problem—for any B on the boundary, pick a $U[\infty]$ that achieves this B, use lossless compression to approach $H(X[1] \mid U[\infty])$, and use lossy compression to approach $I(X[\infty]\,;U[\infty])$.

Theorem 8.6 (Polar coding with one helper)**.** *Let B be any point on the boundary of the rate region. Fix exponents $\pi + 2\rho < 1$. Then polar coding alone yields $P_\mathrm{e} < \exp(-N^\pi)$ and $\operatorname{dist}(R, B) < N^{-\rho}$ at the cost of $O(\log N)$ complexity per source observation.*

PROOF. Let $U[\infty]$ be the auxiliary variable such that

$$B = \big(\, H(X[1] \mid U[\infty]),\, I(X[\infty]\,;[\infty]) \,\big).$$

Tell compressor[1] to compress $X[1]$ losslessly with $U[\infty]$ as the side information. Tell compressor$[\infty]$ to do lossy compression with $U[\infty] \to X[\infty]$ being the test channel. Both of them are covered by the infrastructure and can be done with the specified error and gap to entropy. $\square$

Remark: The same can be stated with error $\exp(n^\tau)$ and complexity $\log(\log N)$. The proof is exactly the same and thus omitted.

It is pointless to have two helpers and no senders. So two senders and one-sender–one-helper are all we need to consider for two random sources. In the upcoming sections, we will see scenarios with more than two sources. The very next scenario we will go over is when there are three senders.

4. Three-Sender Slepian–Wolf

Starting from three senders, a rate region of a distributed compression problem will be a subset in a higher-dimensional Euclidean space. Most importantly, the sum-rate segment will become a sum-rate polygon or even a sum-rate polyhedron. They are also called the *dominant face*, where dominance refers to the fact that it is the set of minimal points under coordinate-wise comparison, and face refers to that it has co-dimension 1 in the ambient space.

To achieve the sum-rate polyhedron, we can always apply time-sharing and accept the penalty $(1+2\rho/3)$. (Note that it is even harder to approximate multiple irrational numbers using a common denominator, so $\rho/2$ will become $2\rho/3$, $3\rho/4$, etc. as the dimension increases.) We can also generalize source-splitting to multiple senders. In the latter case, the problem boils down to why source-splitting exhausts all points in the sum-rate polyhedron; and this is nontrivial.

To demonstrate the non-triviality of source-splitting, consider three senders. See Figure 8.5 for the specification. An easy observation is that, for any sender, its corresponding rate is at least the entropy of its source conditioned on the other two sources. That is,

$$R[1] \geqslant H(1|23) \quad R[2] \geqslant H(2|13) \quad R[3] \geqslant H(3|12). \tag{8.2}$$

Similarly, any two senders should send out the entropy of their sources conditioned on the remaining one source. That is,

$$R[1]+R[2] \geqslant H(12|3) \quad R[1]+R[3] \geqslant H(13|2) \quad R[2]+R[3] \geqslant H(23|1). \tag{8.3}$$

And lastly, the sum-rate is no less than the overall entropy:

$$R[1]+R[2]+R[3] \geqslant H(123). \tag{8.4}$$

Theses inequalities turn out the be the only inequalities that a feasible rate tuple needs to satisfy.

With only inequalities (8.2), the rate region looks like a cube (which actually extends to infinity). Now inequalities (8.3) will chamfer the three edges that are closest to the axes. And finally, inequality (8.4) will truncate the corner that is closest to the origin. See Figure 8.6 for an illustration of the result. The *rate hexagon* is the intersection of the rate region with inequality (8.4) replaced by equality.

Now consider the following source-splitting setup.

- $X[1]$ will not be split, but will be denoted by $X[1]\langle 1\rangle$ for notational compatibility;
- $X[2]$ will be split into $X[2]\langle 1\rangle$ and $X[2]\langle 2\rangle$; and
- $X[3]$ will be split into $X[3]\langle 1\rangle$, $X[3]\langle 2\rangle$, $X[3]\langle 3\rangle$, and $X[3]\langle 4\rangle$.

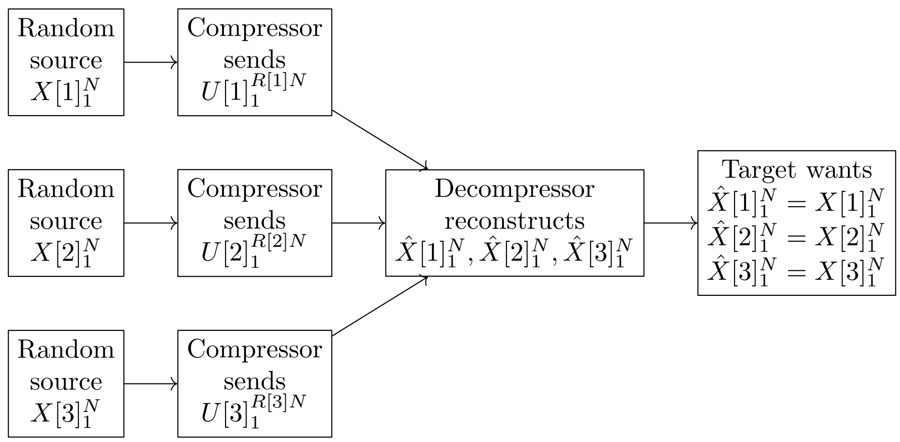

FIGURE 8.5. Distributed lossless compression with three senders.

And now we use the fragments of $X[3]$ to sandwich "the sandwich made out of $X[1]$ and $X[2]$". More precisely, we order them as

$$X[3]\langle 1\rangle, X[2]\langle 1\rangle, X[3]\langle 2\rangle, X[1]\langle 1\rangle, X[3]\langle 3\rangle, X[2]\langle 2\rangle, X[3]\langle 4\rangle. \tag{8.5}$$

In general, the fragment $X[m]\langle l\rangle$ will be placed at $(2l-1)/2^m$ on the number line, and then we read off the fragments from left to right.

Let Q be a random variable that outputs a permutation of $\{1, 2, 3\}$. That is to say, $Q \in S_3 \coloneqq \{123, 132, 312, 321, 231, 213\}$. Depending on Q, we want to assign each fragment a true value or a placeholder symbol. For a fixed m, all $X[m]\langle l\rangle$ will be $\spadesuit$ except that one will be $X[m]$. The fragments that get the true values are such that $X[Q(1)]$ will appear first on the number line, followed by $X[Q(2)]$, and finishing with $X[Q(3)]$. For example, if $Q = 231$, then $X[2]\langle 1\rangle$ gets the true value of $X[2]$ and $X[3][2]$ gets the true value of $X[3]$. Now sequence (8.5) becomes

$$\spadesuit,\ X[2]\langle 1\rangle = X[2],\ X[3]\langle 2\rangle = X[3],\ X[1]\langle 1\rangle = X[1],\ \spadesuit,\ \spadesuit,\ \spadesuit.$$

See Table 8.1 for the other Q's. The assignment is not necessarily unique (e.g., when $Q = 213$ or $Q = 312$); we will get back to this soon.

With the fragments defined, I will specify the coding scheme: For each $X[m]\langle l\rangle$, it will be compressed by compressor$[m]$ given all fragments to the right and Q. More precisely,

- compressor[3] will compress $X[3]\langle 1\rangle$ given $X[2]\langle 1\rangle X[3]\langle 2\rangle X[1]\langle 1\rangle X[3]\langle 3\rangle X[2]\langle 2\rangle X[3]\langle 4\rangle Q$,
- compressor[2] will compress $X[2]\langle 1\rangle$ given $X[3]\langle 2\rangle X[1]\langle 1\rangle X[3]\langle 3\rangle X[2]\langle 2\rangle X[3]\langle 4\rangle Q$,
- compressor[3] will compress $X[3]\langle 2\rangle$ given $X[1]\langle 1\rangle X[3]\langle 3\rangle X[2]\langle 2\rangle X[3]\langle 4\rangle Q$,
- compressor[1] will compress $X[1]\langle 1\rangle$ given $X[3]\langle 3\rangle X[2]\langle 2\rangle X[3]\langle 4\rangle Q$,
- compressor[3] will compress $X[3]\langle 3\rangle$ given $X[2]\langle 2\rangle X[3]\langle 4\rangle Q$,
- compressor[2] will compress $X[2]\langle 2\rangle$ given $X[3]\langle 4\rangle Q$, and finally
- compressor[3] will compress $X[3]\langle 4\rangle$ given Q.

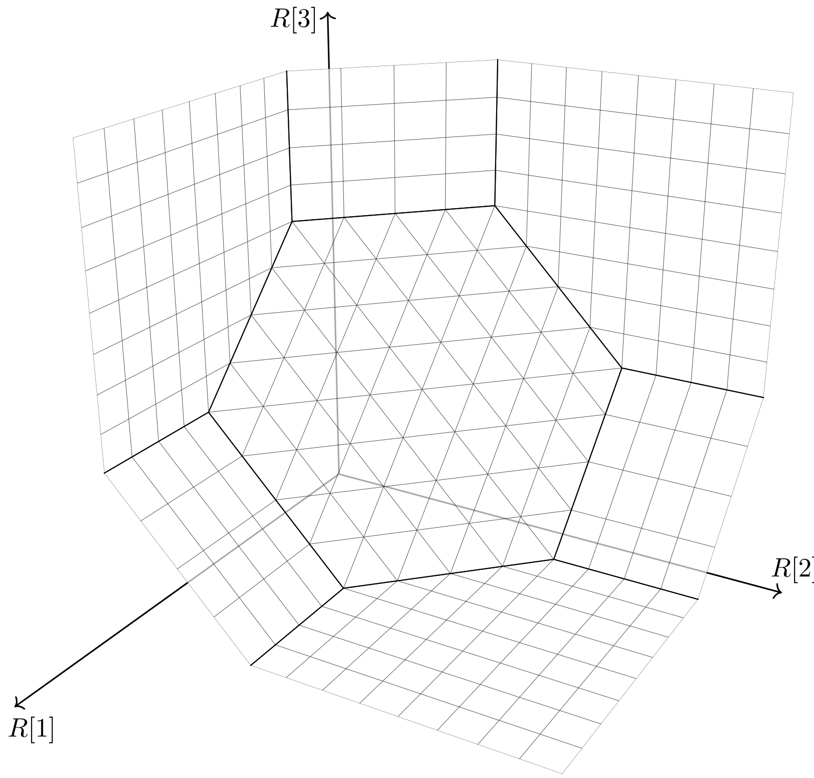

FIGURE 8.6. The rate region for if there are three senders. The camera is inside the rate region, looking at the origin. A pentagon with square mesh is when one of inequalities (8.2) is forced to be an equality. A rectangle with rectangular mesh is when one of inequalities (8.3) is forced to be an equality. The hexagon is when the sum-rate equals the total entropy.

Let $B[m]$ be the sum of the conditional entropies of the duties of compressor$[m]$. It is clear that the sum of these conditional entropies is $H(X[1]X[2]X[3])$ by the chain rule; so $B \coloneqq (B[1], B[2], B[3])$ lies on the sum-rate hexagon. The claim is that, by equipping Q with the correct distribution, B achieves any point on the sum-rate hexagon.

Theorem 8.7 (Onto the hexagon)**.** $B \coloneqq (B[1], B[2], B[3])$ *exhausts all points on the sum-rate hexagon as Q varies over all distributions on S_3.*

PROOF. The sum-rate hexagon has vertices

$$
\begin{array}{cccc}
 & (H(1|3), H(2|13), H(3)), & (H(1|23), H(2|3), H(3)), & \\
(H(1), H(2|31), H(3|1)), & & & (H(1|32), H(2), H(3|2)), \\
 & (H(1), H(2|1), H(3|21)), & (H(1|2), H(2), H(3|12)). &
\end{array}
$$

Each permutation on $\{1, 2, 3\}$ corresponds to a vertex by specifying which sources are compressed in full and which are conditioned on the others. For instance, the permutation 123 corresponds to the vertex on the top right, and the permutation

TABLE 8.1. Splitting three sources per the permutation Q. Note that there are eight rows because two permutations (312 and 213) assume two solutions.

Q	$X[3]\langle 1\rangle$	$X[2]\langle 1\rangle$	$X[3]\langle 2\rangle$	$X[1]\langle 1\rangle$	$X[3]\langle 3\rangle$	$X[2]\langle 2\rangle$	$X[3]\langle 4\rangle$
123	$\spadesuit$	$\spadesuit$	$\spadesuit$	$X[1]$	$\spadesuit$	$X[2]$	$X[3]$
132	$\spadesuit$	$\spadesuit$	$\spadesuit$	$X[1]$	$X[3]$	$X[2]$	$\spadesuit$
312	$\spadesuit$	$\spadesuit$	$X[3]$	$X[1]$	$\spadesuit$	$X[2]$	$\spadesuit$
312	$X[3]$	$\spadesuit$	$\spadesuit$	$X[1]$	$\spadesuit$	$X[2]$	$\spadesuit$
321	$X[3]$	$X[2]$	$\spadesuit$	$X[1]$	$\spadesuit$	$\spadesuit$	$\spadesuit$
231	$\spadesuit$	$X[2]$	$X[3]$	$X[1]$	$\spadesuit$	$\spadesuit$	$\spadesuit$
213	$\spadesuit$	$X[2]$	$\spadesuit$	$X[1]$	$X[3]$	$\spadesuit$	$\spadesuit$
213	$\spadesuit$	$X[2]$	$\spadesuit$	$X[1]$	$\spadesuit$	$\spadesuit$	$X[3]$

321 corresponds to the vertex on the bottom left. See Figure 8.7 for more on this correspondence. (Note that $(H(3|21), H(2|1), H(1))$ does not make sense, because compressor[1] cannot access $X[3]$.)

Our strategy is as follows: We first show that every vertex is achievable. We then show that every edge is achievable. We lastly show that the entire hexagon is achievable.

The first goal is straightforward. If we want to achieve, for instance, the vertex $(H(1), H(2|31), H(3|1))$ corresponding to the permutation 231, then let $Q = 231$ with probability 1. This means that $X[2]\langle 1\rangle = X[2]$ and $X[3]\langle 2\rangle = X[3]$ constantly, and the other fragments are all $\spadesuit$ constantly. As a result, compressor[2] will have to compress $X[2] \mid X[3]X[1]$, compressor[3] will have to compress $X[3] \mid X[1]$, and compressor[1] will have to compress $X[1]$. Now $B \coloneqq (B[1], B[2], B[3])$ becomes $(H(1), H(2|31), H(3|1))$, as desired. For any other vertex, the argument is similar and thus omitted.

To achieve the second goal, take the edge 312–321 as an example. Now we let $Q = 312$ with probability $1 - t$ and let $Q = 321$ with probability t. As t goes from 0 to 1, the knob Q varies from constantly 312 to constantly 321. This means that B moves from the vertex 312 to the vertex 321. Along this process, $X[3]$ is always compressed conditioned on the other two, so $B[3]$ is always $H(3|12) = H(3|21)$. This means that $B(t)$, as a function in t, maps surjectively onto the edge 312–321 of the hexagon. For any other edge, the argument is similar and thus omitted.

It remains to show that B, as a function in the distribution of Q, maps surjectively onto the hexagon. To this end, consider the following "Tour de France"

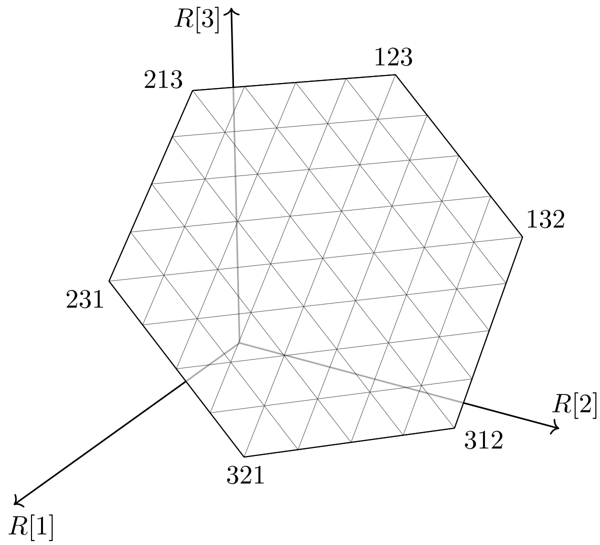

FIGURE 8.7. The sum-rate hexagon with vertices indexed by permutations.

definition of Q_t

$$Q_t = \begin{cases} \text{constantly } 123 & \text{when } t = 0, \\ \text{constantly } 132 & \text{when } t = 1, \\ \text{constantly } 312 & \text{when } t = 2, \\ \text{constantly } 321 & \text{when } t = 3, \\ \text{constantly } 231 & \text{when } t = 4, \\ \text{constantly } 213 & \text{when } t = 5, \\ \text{constantly } 123 & \text{when } t = 6, \end{cases}$$

and filling in the non-integer t by interpolation

$$Q_t = \begin{cases} Q_{\lceil t \rceil} & \text{with probability } \lceil t \rceil - t, \\ Q_{\lfloor t \rfloor} & \text{with probability } t - \lfloor t \rfloor. \end{cases}$$

By the previous paragraph, $B(t)$ will travel through each edge of the hexagon (although we have no idea the velocity it travels) in the order given in Table 8.1.

Now let me borrow some algebraic topology nonsense: The space of the distributions on S_3 is contractible to the uniform distribution. Thus the Tour de France Q_t is a cycle (mapped to 0 by the co-differential operator ∂) that happens to be the boundary (the image under ∂) of some disk D. This disk D has its boundary mapped to the boundary of the hexagon with winding number 1 (degree 1), so D will map surjectively onto the hexagon. That completes the proof. □

The proof is inspired by [**GRUW01**], where multiple access channels are considered. Once we know how to use Q to attain any point on the sum-rate hexagon, use the infrastructure to code.

Corollary 8.8 (Polar coding for three senders)**.** *For three-sender distributed lossless compression, polar coding coupled with source-splitting attains every point on the boundary of the rate region with $N^{-\rho}$ gap.*

The next section combines everything we learned so far in this chapter to attack the problem with more than three senders and one helper.

5. Many Senders with One Helper

Below [**EGK11**, Theorem 10.4], the authors commented that the optimal rate region is unknown when there are two helpers. That is to say, the most general case with known rate region is when there are multiple senders and one helper. We stick to the cases with known rate region because we want to state theorems about the pace of convergence; only when the aimed limit is optimal is this pace meaningful.

For a many-sender one helper problem, the description of the rate region is a combination of inequalities of the form $R[1]+R[3]+R[5] \geqslant H(135|246)$ and $R[\infty] \geqslant$ the capacity of a proper test channel.

Theorem 8.9 (Distributed lossless compression with a helper)**.** *The optimal rate region for lossless source coding of $X[1], \dots, X[M]$ with helper source $X[\infty]$ is described by*

$$\sum_{m \in \mathcal{S}} R[m] \leqslant H(X[\mathcal{S}] \mid U[\infty], X[\mathcal{S}^{\complement}]) \tag{8.6}$$

for all subsets $\mathcal{S} \subseteq \{1, \dots, M\}$ and

$$R[\infty] \geqslant I(X[\infty]\,; U[\infty])$$

unioned over all random variables $U[\infty]$. Here, $X[\mathcal{S}]$ is the tuple $(X[m] : m \in \mathcal{S})$ and $X[\mathcal{S}^{\complement}]$ is what is left $(X[m] : m \notin \mathcal{S})$.

Note that the theorem implicitly uses $U[\infty]$ as a coded time-sharing knob so there is not need to take the convex hull of the union. Now fix a point B on the boundary of the rate region. Fix a $U[\infty]$ that achieve this point in the rate region. Then inequalities (8.6) is a family of inequalities parametrized by the subset $\mathcal{S}$ of $\{1, \dots, m\}$. The right-hand side of the inequalities,

$$H(X[\mathcal{S}] \mid U[\infty], X[\mathcal{S}^{\complement}]) = H\big((X[m] : m \in \mathcal{S}) \mid U[\infty], (X[m] : m \notin \mathcal{S})\big),$$

is a supermodular function in $\mathcal{S}$. To verify this, it suffices to check the three variable case.

Lemma 8.10 (Supermodularity)**.** *For any random variables X, Y, Z,*

$$H(XY \mid Z) + H(YZ \mid X) \leqslant H(Y \mid XZ) + H(XYZ).$$

Proof. Subtract $2H(XYZ)$ from both sides; the desired inequality is equivalent to $H(X) + H(Z) \geqslant H(XZ)$. □

A supermodular function comes with a contra-polymatroid defined by the sum of coordinates over a subset $\mathcal{S}$ being greater than or equal to the function evaluation at $\mathcal{S}$. In other words, the rate region with a fixed $U[\infty]$ is a contra-polymatroid. This is the dual case of a polymatroid defined by a submodular function. The latter is seen when one considers the capacity region of a multiple access channel [**GRUW01**].

By the duality between (submodular function, polymatroid) and (supermodular function, contra-polymatroid), the proof given in [**GRUW01**] applies here. The proof therein says that the *rate*-splitting technique exhausts all points on the sum-rate polyhedron of a multiple access channel. And here, we conclude that the source-splitting technique exhausts all points on the sum-rate polyhedron of a distributed lossless compression.

Corollary 8.11 (Polar coding for many-sender one-helper)**.** *Polar coding coupled with source-splitting attains every point on the boundary of the rate region with* $N^{-\rho}$ *gap.*

Sketch of the proof. Given the polar coding infrastructure, it suffices to construct a scheme to split sources and show that it exhausts all point on the sum-rate polyhedron (aka. the dominant face).

The splitting scheme will look like the following: For each m, the mth source $X[m]$ will be split into 2^{m-1} fragments and the latter are named $X[m]\langle 1\rangle, \dots,$ $X[m]\langle 2^{m-1}\rangle$. Each fragment $X[m]\langle l\rangle$ will be placed at $(2l-1)/2^m$ on the number line. The mth compressor will compress $X[m]\langle l\rangle$ given everything to its right. The mth duty entropy, $B[m]$, will be the sum of $H(X[m]\langle l\rangle \mid \text{fragments to its right})$ over all l, which will also be the limit of $R[m]$ as the block length goes to infinity.

Now apply induction to show that every d-dimensional facet of the sum-rate polyhedron is achievable by source-splitting.

- Show that every vertex of the sum-rate polyhedron corresponds to a permutation of $\{1, \dots, M\}$; and show that by reordering the sources in all possible ways, the duty tuple $(B[1], \dots, B[M])$ attains all vertices.
- Show that every edge of the sum-rate polyhedron corresponds to a smooth transition between two permutations that differ by a swap.
- Show that every face (2-dimensional facet) is mapped surjectively because there is a Tour de France Q_t whose image goes around the boundary once while the domain is contractible.
- Show the similar argument that if a map from a topological ball maps the boundary to the boundary of a facet of the polygon, plus the induced map on the top homology groups is 1 (multiplying by one), then the map is surjective.

For more details, see [**GRUW01**]. □

A similar statement can be made with $\log(\log N)$ complexity. The proof is essentially the same except that when we invoke the infrastructure, the log-log code is used instead of the second-moment code.

I will end the dissertation with a remark on the dual of distributed compression.

6. On Multiple Access Channels

Multiple access channels are noisy channels that take multiple inputs—each from a different encoder—and output to a unified decoder. The assumption that the encoders cannot talk to each other and that the decoder has all information in hand to make decisions make this problem a proper dual of distributed compression. See Figure 8.8 for an example specification of a multiple access channel with three senders and compare it with Figure 8.5.

The capacity region of a multiple access channel is defined similarly to the rate region of distributed compression. For instance for three senders, the capacity region is the set of $(R[1], R[2], R[3])$ such that a reliable communication can be carried. Then the pessimistic criteria are, for instance

- sender[1] can send out at most $I(X[1]\,;Y \mid X[2]X[3])$ bits reliably,
- senders [1] and [2] together can send out at most $I(X[1]X[2]\,;Y \mid X[3])$ bits reliably, and

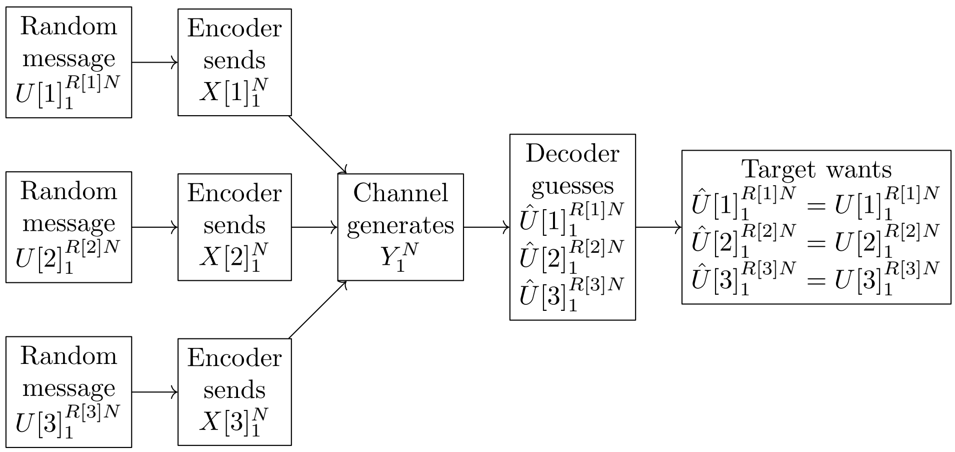

FIGURE 8.8. Multiple access channel with three senders.

- all three senders together can send out at most $I(X[1]X[2]X[3] \,;\, Y)$ bits reliably.

Hence the following characterization of the capacity region. See also Figure 8.9.

Theorem 8.12 (Rate region of multiple access channel)**.** *For any multiple access channel with M senders, the capacity region is the set of points $(R[1], \ldots, R[M])$ such that, for all subsets $\mathcal{S} \subseteq \{1, \ldots, M\}$,*

$$\sum_{m \in \mathcal{S}} R[m] \leqslant I(X[\mathcal{S}] \,;\, Y \mid S[\mathcal{S}^\complement], Q),$$

unioned over all possible distributions of the inputs $X[1], \ldots, X[M]$ and the knob variable Q.

As commented before, [**GRUW01**] showed that one can split a multiple access channel into several one-to-one DMCs. Once that is done, we can apply the infrastructure.

Corollary 8.13 (Polar coding for multiple access channel)**.** *Polar coding coupled with rate-splitting attains every point on the boundary of the capacity region with $N^{-\rho}$ gap.*

PROOF. [**GRUW01**] reduces this problem into at most $2M$ (twice the number of senders) DMCs. For each DMC, apply the polar coding infrastructure. □

A similar statement can be made with $\log(\log N)$ complexity. The proof is essentially the same except that when we invoke the infrastructure, the log-log code is used instead of the second-moment code.

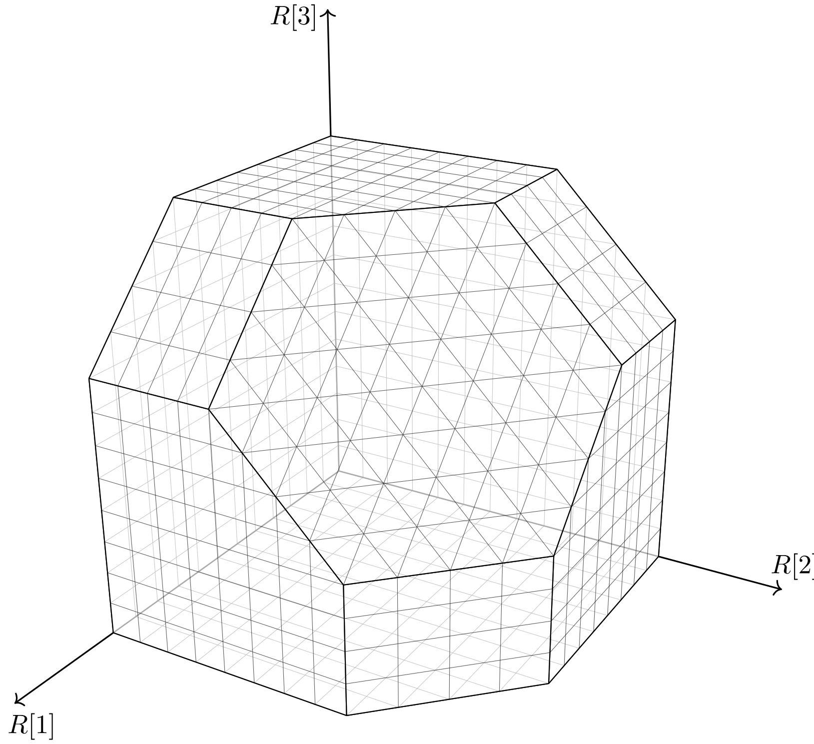

FIGURE 8.9. The capacity region of a three-sender multiple access channel.

Bibliography

[AK75] R. Ahlswede and J. Korner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, 21(6):629–637, November 1975. doi:10.1109/TIT.1975.1055469.

[Ari06] E. Arikan. Channel combining and splitting for cutoff rate improvement. *IEEE Transactions on Information Theory*, 52(2):628–639, Feb 2006. doi:10.1109/TIT.2005.862081.

[Ari09] E. Arikan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, July 2009. doi:10.1109/TIT.2009.2021379.

[Ari10] E. Arikan. Source polarization. In *2010 IEEE International Symposium on Information Theory*, pages 899–903, June 2010. doi:10.1109/ISIT.2010.5513567.

[AT09] E. Arikan and E. Telatar. On the rate of channel polarization. In *2009 IEEE International Symposium on Information Theory*, pages 1493–1495, June 2009. doi:10.1109/ISIT.2009.5205856.

[AW14] Y. Altuğ and A. B. Wagner. Moderate deviations in channel coding. *IEEE Transactions on Information Theory*, 60(8):4417–4426, Aug 2014. doi:10.1109/TIT.2014.2323418.

[AYK11] A. Alamdar-Yazdi and F. R. Kschischang. A simplified successive-cancellation decoder for polar codes. *IEEE Communications Letters*, 15(12):1378–1380, December 2011. doi:10.1109/LCOMM.2011.101811.111480.

[BF02] A. Barg and G. D. Forney. Random codes: minimum distances and error exponents. *IEEE Transactions on Information Theory*, 48(9):2568–2573, Sep. 2002. doi:10.1109/TIT.2002.800480.

[BGN+18] Jarosław Błasiok, Venkatesan Guruswami, Preetum Nakkiran, Atri Rudra, and Madhu Sudan. General strong polarization. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 485–492, New York, NY, USA, 2018. Association for Computing Machinery. doi:10.1145/3188745.3188816.

[BGS18] Jaroslaw Blasiok, Venkatesan Guruswami, and Madhu Sudan. Polar Codes with Exponentially Small Error at Finite Block Length. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, volume 116 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 34:1–34:17, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. URL: http://drops.dagstuhl.de/opus/volltexte/2018/9438, doi:10.4230/LIPIcs.APPROX-RANDOM.2018.34.

[Bil12] E. A. Bilkent. Polar coding for the slepian-wolf problem based on monotone chain rules. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 566–570, July 2012. doi:10.1109/ISIT.2012.6284254.

[Bla87] Richard E. Blahut. *Principles and Practice of Information Theory*. Addison-Wesley Longman Publishing Co., Inc., USA, 1987.

[CK10] H. S. Cronie and S. B. Korada. Lossless source coding with polar codes. In *2010 IEEE International Symposium on Information Theory*, pages 904–908, June 2010. doi:10.1109/ISIT.2010.5513561.

[CS07] C. Chang and A. Sahai. Universal quadratic lower bounds on source coding error exponents. In *2007 41st Annual Conference on Information Sciences and Systems*, pages 714–719, March 2007. doi:10.1109/CISS.2007.4298398.

[Dur19] Rick Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 5 edition, 2019.

[DZ10] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition. doi:10.1007/978-3-642-03311-7.

[EGK11] Abbas El Gamal and Young-Han Kim. *Network information theory*. Cambridge University Press, Cambridge, 2011. doi:10.1017/CBO9781139030687.

[EKMF+17] M. El-Khamy, H. Mahdavifar, G. Feygin, J. Lee, and I. Kang. Relaxed polar codes. *IEEE Transactions on Information Theory*, 63(4):1986–2000, April 2017. doi:10.1109/TIT.2017.2657691.

[FHMV17] Arman Fazeli, S. Hamed Hassani, Marco Mondelli, and Alexander Vardy. Binary linear codes with optimal scaling and quasi-linear complexity. *CoRR*, abs/1711.01339, 2017. URL: http://arxiv.org/abs/1711.01339, arXiv:1711.01339.

[FHMV18] A. Fazeli, H. Hassani, M. Mondelli, and A. Vardy. Binary linear codes with optimal scaling: Polar codes with large kernels. In *2018 IEEE Information Theory Workshop (ITW)*, pages 1–5, Nov 2018. doi:10.1109/ITW.2018.8613428.

[FHMV20] A. Fazeli, H. Hassani, M. Mondelli, and A. Vardy. Binary linear codes with optimal scaling: Polar codes with large kernels. *IEEE Transactions on Information Theory*, pages 1–1, 2020. doi:10.1109/TIT.2020.3038806.

[FM94] M. Feder and N. Merhav. Relations between entropy and error probability. *IEEE Transactions on Information Theory*, 40(1):259–266, Jan 1994. doi:10.1109/18.272494.

[FT17] Silas L. Fong and Vincent Y. F. Tan. Scaling exponent and moderate deviations asymptotics of polar codes for the awgn channel. *Entropy*, 19(7), 2017. URL: http://www.mdpi.com/1099-4300/19/7/364, doi:10.3390/e19070364.

[FV14] A. Fazeli and A. Vardy. On the scaling exponent of binary polarization kernels. In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 797–804, Sept 2014. doi:10.1109/ALLERTON.2014.7028536.

[Gal68] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., New York, NY, USA, 1968.

[GB14] D. Goldin and D. Burshtein. Improved bounds on the finite length scaling of polar codes. *IEEE Transactions on Information Theory*, 60(11):6966–6978, Nov 2014. doi:10.1109/TIT.2014.2359197.

[GHU12] A. Goli, S. H. Hassani, and R. Urbanke. Universal bounds on the scaling behavior of polar codes. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 1957–1961, July 2012. doi:10.1109/ISIT.2012.6283641.

[GRUW01] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting. Rate-splitting multiple access for discrete memoryless channels. *IEEE Transactions on Information Theory*, 47(3):873–890, March 2001. doi:10.1109/18.915637.

[GRY19] Venkatesan Guruswami, Andrii Riazanov, and Min Ye. Arıkan meets shannon: Polar codes with near-optimal convergence to channel capacity. *CoRR*, abs/1911.03858, 2019. URL: http://arxiv.org/abs/1911.03858, arXiv:1911.03858.

[GRY20] Venkatesan Guruswami, Andrii Riazanov, and Min Ye. Arikan meets shannon: Polar codes with near-optimal convergence to channel capacity. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, pages 552–564, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3357713.3384323.

[GX13] V. Guruswami and P. Xia. Polar codes: Speed of polarization and polynomial gap to capacity. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 310–319, Oct 2013. doi:10.1109/FOCS.2013.41.

[GX15] V. Guruswami and P. Xia. Polar codes: Speed of polarization and polynomial gap to capacity. *IEEE Transactions on Information Theory*, 61(1):3–16, Jan 2015. doi:10.1109/TIT.2014.2371819.

[HAU14] S. H. Hassani, K. Alishahi, and R. L. Urbanke. Finite-length scaling for polar codes. *IEEE Transactions on Information Theory*, 60(10):5875–5898, Oct 2014. doi:10.1109/TIT.2014.2341919.

[Hay06] M. Hayashi. General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel. *IEEE Transactions on Information Theory*, 52(4):1562–1575, April 2006. doi:10.1109/TIT.2006.871040.

[HMF+20] Seyyed Ali Hashemi, Marco Mondelli, Arman Fazeli, Alexander Vardy, John Cioffi, and Andrea Goldsmith. Parallelism versus latency in simplified successive-cancellation decoding of polar codes, 2020. arXiv:2012.13378.

[HMTU13] S. H. Hassani, R. Mori, T. Tanaka, and R. L. Urbanke. Rate-dependent analysis of the asymptotic behavior of channel polarization. *IEEE Transactions on Information Theory*, 59(4):2267–2276, April 2013. doi:10.1109/TIT.2012.2228295.

[HY13] J. Honda and H. Yamamoto. Polar coding without alphabet extension for asymmetric models. *IEEE Transactions on Information Theory*, 59(12):7829–7838, Dec 2013. doi:10.1109/TIT.2013.2282305.

[JA18] T. S. Jayram and E. Arıkan. A note on some inequalities used in channel polarization and polar coding. *IEEE Transactions on Information Theory*, 64(8):5767–5768, Aug 2018. doi:10.1109/TIT.2017.2717598.

[KMTU10] S. B. Korada, A. Montanari, E. Telatar, and R. Urbanke. An empirical scaling law for polar codes. In *2010 IEEE International Symposium on Information Theory*, pages 884–888, June 2010. doi:10.1109/ISIT.2010.5513579.

[KSU10] S. B. Korada, E. Sasoglu, and R. Urbanke. Polar codes: Characterization of exponent, bounds, and constructions. *IEEE Transactions on Information Theory*, 56(12):6253–6264, Dec 2010. doi:10.1109/TIT.2010.2080990.

[KU10] S. B. Korada and R. L. Urbanke. Polar codes are optimal for lossy source coding. *IEEE Transactions on Information Theory*, 56(4):1751–1768, April 2010. doi:10.1109/TIT.2010.2040961.

[MHCG20] M. Mondelli, S. A. Hashemi, J. Cioffi, and A. Goldsmith. Sublinear latency for simplified successive cancellation decoding of polar codes. *IEEE Transactions on Wireless Communications*, pages 1–1, 2020. doi:10.1109/TWC.2020.3022922.

[MHU16] M. Mondelli, S. H. Hassani, and R. L. Urbanke. Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors. *IEEE Transactions on Information Theory*, 62(12):6698–6712, Dec 2016. doi:10.1109/TIT.2016.2616117.

[MT10] R. Mori and T. Tanaka. Channel polarization on q-ary discrete memoryless channels by arbitrary kernels. In *2010 IEEE International Symposium on Information Theory*, pages 894–898, June 2010. doi:10.1109/ISIT.2010.5513568.

[MT14] R. Mori and T. Tanaka. Source and channel polarization over finite fields and reed-solomon matrices. *IEEE Transactions on Information Theory*, 60(5):2720–2736, May 2014. doi:10.1109/TIT.2014.2312181.

[Mur21] Jun Muramatsu. Binary polar codes based on bit error probability, 2021. arXiv:2101.08943.

[Nas18a] Rajai Nasser. Continuity of channel parameters and operations under various dmc topologies. *Entropy*, 20(5), 2018. URL: https://www.mdpi.com/1099-4300/20/5/330, doi:10.3390/e20050330.

[Nas18b] Rajai Nasser. Topological structures on dmc spaces. *Entropy*, 20(5), 2018. URL: https://www.mdpi.com/1099-4300/20/5/343, doi:10.3390/e20050343.

[PU16] H. D. Pfister and R. Urbanke. Near-optimal finite-length scaling for polar codes over large alphabets. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 215–219, July 2016. doi:10.1109/ISIT.2016.7541292.

[PV10] Y. Polyanskiy and S. Verdú. Channel dispersion and moderate deviations limits for memoryless channels. In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1334–1339, Sept 2010. doi:10.1109/ALLERTON.2010.5707068.

[RU08] Tom Richardson and Rüdiger Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008. doi:10.1017/CBO9780511791338.

[Sas11] Eren Sasoglu. *Polar Coding Theorems for Discrete Systems*. PhD thesis, École polytechnique fédérale de Lausanne, Lausanne, 2011. URL: http://infoscience.epfl.ch/record/168993, doi:10.5075/epfl-thesis-5219.

[SRDR12] D. Sutter, J. M. Renes, F. Dupuis, and R. Renner. Achieving the capacity of any dmc using only polar codes. In *2012 IEEE Information Theory Workshop*, pages 114–118, Sep. 2012. doi:10.1109/ITW.2012.6404638.

[ŞTA09] E. Şaşoğlu, E. Telatar, and E. Arikan. Polarization for arbitrary discrete memoryless channels. In *2009 IEEE Information Theory Workshop*, pages 144–148, Oct 2009. doi:10.1109/ITW.2009.5351487.

[Str62] V. Strassen. Asymptotische abschätzungen in shannons informationstheorie. In *Transactions of the Third Prague Conference on Information Theory*, pages 689–723. Publishing House of the Czechoslovak Academy of Sciences, 1962. URL: https://www.math.cornell.edu/~pmlut/strassen.pdf.

[SW73] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, July 1973. doi:10.1109/TIT.1973.1055037.

[Ter99] Audrey Terras. *Fourier analysis on finite groups and applications*, volume 43 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1999. doi:10.1017/CBO9780511626265.

[TK12] V. Y. F. Tan and O. Kosut. The dispersion of slepian-wolf coding. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 915–919, July 2012. doi:10.1109/ISIT.2012.6284695.

[WD18a] Hsin-Po Wang and Iwan M. Duursma. Log-logarithmic time pruned polar coding on binary erasure channels. *CoRR*, abs/1812.08106, 2018. URL: http://arxiv.org/abs/1812.08106, arXiv:1812.08106.

[WD18b] Hsin-Po Wang and Iwan M. Duursma. Polar code moderate deviation: Recovering the scaling exponent. *CoRR*, abs/1806.02405, 2018. URL: http://arxiv.org/abs/1806.02405, arXiv:1806.02405.

[WD18c] Hsin-Po Wang and Iwan M. Duursma. Polar-like codes and asymptotic tradeoff among block length, code rate, and error probability. *CoRR*, abs/1812.08112, 2018. URL: http://arxiv.org/abs/1812.08112, arXiv:1812.08112.

[WD19a] Hsin-Po Wang and Iwan M. Duursma. Log-logarithmic time pruned polar coding. *CoRR*, abs/1905.13340, 2019. URL: http://arxiv.org/abs/1905.13340, arXiv:1905.13340.

[WD19b] Hsin-Po Wang and Iwan M. Duursma. Polar codes' simplicity, random codes' durability. *CoRR*, abs/1912.08995, 2019. URL: http://arxiv.org/abs/1912.08995, arXiv:1912.08995.

[WD21a] H. P. Wang and I. M. Duursma. Log-logarithmic time pruned polar coding. *IEEE Transactions on Information Theory*, 67(3):1509–1521, March 2021. doi:10.1109/TIT.2020.3041523.

[WD21b] H. P. Wang and I. M. Duursma. Polar codes' simplicity, random codes' durability. *IEEE Transactions on Information Theory*, 67(3):1478–1508, March 2021. doi:10.1109/TIT.2020.3041570.

[Wyn75] A. D. Wyner. The wire-tap channel. *Bell System Technical Journal*, 54(8):1355–1387, 1975. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1975.tb02040.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.1538-7305.1975.tb02040.x, doi:10.1002/j.1538-7305.1975.tb02040.x.

[YB15] M. Ye and A. Barg. Polar codes using dynamic kernels. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 231–235, June 2015. doi:10.1109/ISIT.2015.7282451.

[YFV19] H. Yao, A. Fazeli, and A. Vardy. Explicit polar codes with small scaling exponent. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 1757–1761, July 2019. doi:10.1109/ISIT.2019.8849741.