Polar Coding Theorems for Discrete Systems

THÈSE N^{O} 5219 (2011)

PRÉSENTÉE LE 2 NOVEMBRE 2011

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS

LABORATOIRE DE THÉORIE DE L'INFORMATION

PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Eren ŞAŞOĞLU

acceptée sur proposition du jury:

Prof. B. Rimoldi, président du jury
Prof. E. Telatar, directeur de thèse
Prof. E. Arikan, rapporteur
Prof. R. Urbanke, rapporteur
Prof. A. Vardy, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2011

Abstract

Polar coding is a recently invented technique for communication over binary-input memoryless channels. This technique allows one to transmit data at rates close to the symmetric-capacity of such channels with arbitrarily high reliability, using low-complexity encoding and decoding algorithms. As such, polar coding is the only explicit low-complexity method known to achieve the capacity of symmetric binary-input memoryless channels.

The principle underlying polar coding is *channel polarization*: recursively combining several copies of a mediocre binary-input channel to create noiseless and useless channels. The same principle can also be used to obtain optimal low-complexity compression schemes for memoryless binary sources.

In this dissertation, the generality of the polarization principle is investigated. It is first shown that polarization with recursive procedures is not limited to binary channels and sources. A family of low-complexity methods that polarize all discrete memoryless processes is introduced. In both data transmission and data compression, codes based on such methods achieve optimal rates, i.e., channel capacity and source entropy, respectively. The error probability behavior of such codes is as in the binary case.

Next, it is shown that a large class of recursive constructions polarize memoryless processes, establishing the original polar codes as an instance of a large class of codes based on polarization methods. A formula to compute the error probability dependence of generalized constructions on the coding length is derived. Evaluating this formula reveals that substantial error probability improvements over the original polar codes can be achieved at large coding lengths by using generalized constructions, particularly over channels and sources with non-binary alphabets.

Polarizing capabilities of recursive methods are shown to extend beyond memoryless processes: Any construction that polarizes memoryless processes will also polarize a large class of processes with memory.

The principles developed are applied to settings with multiple memoryless processes. It is shown that separately applying polarization constructions to two correlated processes polarizes both the processes themselves as well as the correlations between them. These observations lead to polar coding theorems for multiple-access channels and separate compression of correlated sources.

The proposed coding schemes achieve optimal sum rates in both problems.

Keywords: Polar codes, channel polarization, source polarization, capacity-achieving codes, optimal compression, multiple-access channels, distributed source coding, coding for ergodic channels and sources.

Résumé

Le codage polaire est une technique inventée récemment pour la communication sur des canaux sans mémoire avec entrées binaires. Cette technique permet de transmettre des données à des taux approchant la capacité symétrique de tels canaux avec une fiabilité arbitrairement grande, tout en utilisant des algorithmes de faible complexité pour l'encodage et le décodage. De fait, le codage polaire est la seule méthode explicite connue de faible complexité qui atteigne la capacité de canaux sans mémoire symétriques avec entrées binaires.

Le principe qui sous-tend le codage polaire est la *polarisation de canal*: en combinant de manière récursive plusieurs copies d'un canal à entrées binaires, on obtient des canaux qui sont soit sans bruit, soit inutiles. Le même principe s'applique pour l'obtention de schémas de compression de faible complexité pour des sources binaires sans mémoire.

Dans cette thèse, la généralité du principe de polarisation est étudiée. On montre tout d'abord que la procédure récursive de polarisation ne s'applique pas seulement aux canaux et sources binaires. Une famille de méthodes de faible complexité qui polarisent tous les processus sans mémoire discrets est introduite. Que ce soit dans le cas de la transmission ou de la compression de données, les codes basés sur de telles méthodes atteignent des taux optimaux, i.e., la capacité du canal ou l'entropie de la source, respectivement. Le comportement de la probabilité d'erreur de tels codes est semblable à celui obtenu dans le cas binaire.

On montre ensuite qu'une grande classe de constructions récursives polarisent des processus sans mémoire, démontrant ainsi que les codes polaires d'origine constituent un exemple particulier d'une grande classe de codes basés sur des méthodes de polarisation. Une formule est dérivée pour le calcul de la dépendance de la probabilité d'erreur de constructions généralisées en fonction de la longueur d'un code. En évaluant cette formule, on montre que par rapport aux codes polaires d'origine, des améliorations substantielles de la probabilité d'erreur peuvent être obtenues pour de longs codes en utilisant des constructions généralisées, et ceci plus particulièrement pour des canaux et des sources avec des alphabets non-binaires.

On montre également que la polarisation par des méthodes récursives s'étend au-delà des processus sans mémoire. En particulier, on montre que toute con-

struction qui polarise un processus sans mémoire polarise également une grande classe de processus avec mémoire.

Les principes développés dans cette thèse sont également appliqués à des processus multivariés sans mémoire. On démontre qu'appliquer séparément des techniques de polarisation à deux processus corrélés polarise non seulement les processus eux-mêmes, mais aussi les corrélations entre ces deux processus. Ces observations mènent à des théorèmes de codage polaire pour les canaux à accès multiples et la compression de sources corrélées. Les schémas de codage proposés atteignent des taux optimaux dans les deux cas.

Mots clés: Codes polaires, polarisation de canal, polarisation de source, codes atteignant la capacité, compression optimale, canaux à accès multiples, codage de source distribué, codage pour des canaux et des sources ergodiques.

Acknowledgements

It has been a great pleasure to know Professor Emre Telatar in person and to be his doctoral student. I am grateful for his guidance and support, without which this work could not have been possible. Over the years I have benefited immensely from Emre's exceptional generosity and wisdom on a vast range of subjects, technical and otherwise.

I thank my thesis committee members, Professors Erdal Arıkan, Rüdiger Urbanke, and Alexander Vardy for their helpful comments, and Professor Bixio Rimoldi for presiding over the thesis committee. Prof. Urbanke has been influential in my personal and professional development, for which I am grateful. I am indebted to Prof. Arıkan for his contributions to this work and earnest career advice. I am very fortunate to have come across his polar codes while looking for a thesis topic. I also thank Professor Gerhard Kramer for his guidance during my stay at Bell Labs and his interest in my career.

I thank Yvonne Huskie for her efficient administrative support. Along with her, Muriel Bardet and Françoise Behn were always willing to help and made life easy. I also thank Damir Laurenzi for managing the computer network.

I thank all past and present members of IPG for the innumerable (and often pointless) discussions we had. In particular, the companionship of Mohammad Karzand and Marc Vuffray made the office a fun place to be.

I thank Işık Karahanoğlu for her friendship and affection.

I thank my parents Esin and Şadan Şaşoğlu, and my brother Umut for their love and support.

Contents

1 Introduction

Figure 1.1 depicts the setting for the fundamental problem in communication theory. A sender has K bits of information to send, which, after appropriate processing, are transmitted through a noisy channel that accepts input symbols one at a time and produces a sequence of output symbols. The task of the communication engineer is to design an encoding/decoding scheme that ensures that the K bits are (i) transmitted in as few uses of the channel as possible, and (ii) correctly reproduced at the receiver with as high a probability as desired. In [1], Shannon showed that these seemingly conflicting requirements can be met simultaneously so long as K and N (number of channel uses) are large and K/N (called the rate of transmission) is below the *capacity* of the channel.

Figure 1.1

Shannon's proof of the channel coding theorem shows not only that reliable communication at rates below capacity is possible, but also that almost all encoding schemes, i.e., channel codes, with rates below channel capacity will perform well as long as optimal decoders are used at the receiver. Unfortunately, optimal decoding is in general prohibitively difficult—its complexity grows exponentially in the coding length—and how to construct practical coding schemes, and especially low-complexity decoders, is not immediately clear from Shannon's coding theorem alone.

Significant progress has been made in the past sixty years toward developing practical and capacity-achieving coding methods. The bulk of the research effort to this end can be broadly divided into two groups: algebraic coding and

iterative coding. Research in algebraic coding was grounded in the recognition that the words of a code must be as different from each other as possible in order to ensure their distinguishability at the receiver. Iterative codes (e.g., Turbo codes and LDPC codes) on the other hand are designed to work well with a low-complexity decoding algorithm. Despite remarkable advances in both fields, especially in iterative coding, finding codes that (i) operate at rates close to capacity, (ii) have low computational complexity, and (iii) have provable reliability guarantees was an elusive goal until recently.[1]

Polar codes, invented recently by Arıkan [3], have all of these desirable properties. In particular,

- they achieve the symmetric capacity of all binary-input memoryless channels. Consequently they are capacity-achieving for symmetric channels, which include several channel classes of practical relevance such as the binary-input additive white Gaussian noise channel, the binary symmetric channel, and the binary erasure channel.

- they are low-complexity codes, and therefore are practical: The time and space complexities of the encoding/decoding algorithms Arıkan proposes in [3] are $O(N \log N)$, where N is the blocklength.

- the block error probability of polar codes is roughly $O(2^{-\sqrt{N}})$ [4]. This performance guarantee is analytical, and is not only based on empirical evidence.

- for symmetric channels, polar code construction is deterministic. That is, the above statements are true not only for ensembles of codes, but for individual polar codes. Further, construction of polar codes can be accomplished with time complexity $O(N)$ and space complexity $O(\log N)$ [5].

The design philosophy of polar codes is fundamentally different from those of both algebraic codes and iterative codes (although the codes themselves are closely related to the algebraic Reed–Muller codes). It is interesting to note that the invention of these codes is in fact the culmination of Arıkan's efforts to improve the rates achievable by convolutional codes and *sequential decoding* [6], a decoding method developed in the late 1950s.

The technique underlying polar codes is 'channel polarization': creating extremal channels—those that are either noiseless or useless—from mediocre ones. Soon after the publication of [3], Arıkan showed that a similar technique can be used to construct optimal source codes [7]—he calls this technique 'source polarization'. It is clear in his work that a single *polarization* principle underlies both techniques; channel polarization and source polarization are specific applications of this principle.

[1] See [2] for a historical account of the development of coding theory in general.

1.1 Extremal Distributions and Polarization

Suppose we are interested in guessing (i.e., decoding) the value of a binary N-vector U_1^N after observing a related random vector Y_1^N. Here, U_1^N may represent a codeword chosen randomly from a channel code, and Y_1^N the output of a channel when U_1^N is the input. Alternatively, U_1^N may be viewed as the output of a random source, and Y_1^N as side information about U_1^N. In order to minimize the probability of decoding error, one chooses the value of U_1^N that maximizes[2]

$$p(u_1^N \mid y_1^N) = \prod_{i=1}^{N} p(u_i \mid y_1^N, u_1^{i-1}).$$

There are two extremal cases in terms of the probability of decoding error. First, if U_1^N is a function of Y_1^N—i.e., if the above probability is either 0 or 1—then its value can always be guessed correctly. Second, if U_1^N is independent of Y_1^N and uniformly distributed, then all guesses are equally good and will be correct with probability $1/2^N$. The first of these cases is trivial provided that the function computations can be done easily, and the second is hopeless.

A more interesting extremal case is one in which the conditional distribution of U_1^N is neither $\{0,1\}$-valued nor uniform, but it is *polarized* in the sense that all distributions in the product formula above are either $\{0,1\}$-valued or uniform. One can view this as a case where all randomness in U_1^N is concentrated in a subset of its components. Clearly, one cannot in general correctly decode such a random vector with high probability. On the other hand, decoding U_1^N again becomes trivial if one has prior knowledge of its random component. The polarized structure in the probability distribution even suggests that U_1^N can be decoded *successively*: Suppose, for the sake of argument, that the odd-numbered factors in the product formula above are $\{0,1\}$-valued distributions whereas the even-numbered factors are uniform. Then, if one has prior knowledge of the even indices of U_1^N, then the odd indices can be determined in increasing order as follows. The decoder first computes U_1 as a function of Y_1^N, then produces U_2 (which is already available to it) then uses its knowledge of U_1 and U_2 to compute U_3 as a function of (Y_1^N, U_1^2), etc.

A realistic model of the input/output process of a noisy channel or the output/side information process of a data source rarely fits this description. On the other hand, one may attempt to transform the process in question into one that does fit it. This is precisely the aim of Arıkan's polarization technique. In its original form, this technique consists in combining two identically distributed binary random variables so as to create two disparate random variables and repeating this operation several times to amplify the disparity, eventually approaching a polarized set of random variables. A review of this

[2]Throughout, probability distributions will be denoted by p as long as their arguments are lower case versions of the random variables they represent. For example we will write $p(x, y \mid z)$ for $p_{XY|Z}(x, y \mid z)$, denoting the joint distribution of X and Y conditioned on Z.

technique along with its applications to channel and source coding is given in Chapter 2.

The desirable properties of codes based on the polarization principle amply motivate an investigation of this principle's generality. This dissertation is the outcome of one such investigation. We begin in Chapter 3 by studying how discrete memoryless processes of arbitrary alphabet sizes, not just binary ones, can be polarized by recursive transforms. We show that this can be accomplished through a linear transform similar to Arıkan's when the alphabet size is prime. Interestingly, linear transforms lose their ability to polarize *all* stationary memoryless processes when the underlying alphabet size is not a prime number. There are, however, non-linear transforms that do polarize all stationary memoryless processes for all finite alphabet sizes. In Section 3.2 we provide sufficient conditions for a recursive transform to polarize all such processes, and give an example of a family of transforms that satisfy these conditions for all finite alphabet sizes. The complexity and the error probability behavior of codes obtained by such transforms are as in the binary case.

While the error probability guarantees of polar codes are unprecedented, it is of interest to know whether even stronger codes can be obtained by combining more than two random variables in each recursion of a polarizing construction. This study is undertaken in Chapter 4: We first show that a large class of recursive linear transforms that combine several random variables at a time polarize memoryless processes with prime alphabet sizes. We then characterize how a single recursion of a given polarizing transform affects error probability behavior, from which results on the large-blocklength behavior follow easily. The implications of this characterization are of a mixed nature: While in the binary case one cannot improve on the $O(2^{-\sqrt{N}})$ error probability decay by combining a small number of random variables at a time, strong improvements become possible as the alphabet size grows.

Results in Chapters 3 and 4 provide extensive evidence that polarization is a fairly general—in fact, almost inevitable—phenomenon. We further substantiate this claim in Chapter 5, where we show that recursive constructions also polarize processes with memory.

In Chapter 6, we make use of the polarization theorems of earlier chapters to study *joint* polarization of multiple processes. We show that recursive transforms, applied separately to multiple processes, not only polarize the individual processes, but the correlations between the processes are also polarized. These results immediately lead to polar coding theorems for multi-user settings such as the separate encoding of correlated sources and the multiple-access channel.

Polarization and Polar Codes

2

In this chapter, we will review the polarization method for binary memoryless processes and show how it can be used to obtain channel and source codes that achieve optimal rates. Owing to the recursive nature of these codes, the techniques for analyzing their performance (rate, error probability, complexity) are fairly simple. In the subsequent chapters we will frequently invoke the techniques discussed here. This chapter is based entirely on [3], [7], and [4].

Consider a pair of discrete random variables (X, Y) with $X \in \{0, 1\}$ and $Y \in \mathcal{Y}$. The alphabet $\mathcal{Y}$ and the joint distribution of (X, Y) may be arbitrary. Suppose we are given N independent copies $(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$ of (X, Y). We may view X_1^N as the output of a binary memoryless source, and Y_1^N as side information about X_1^N. Alternatively, one may interpret X_1^N as independent and identically distributed (i.i.d.) inputs to a binary-input memoryless channel, and Y_1^N as the corresponding output. We will initially focus on the first of these interpretations and discuss the second shortly.

Suppose that a receiver observes Y_1^N and is interested in decoding X_1^N. We know that in addition to Y_1^N, it is necessary and sufficient to provide the receiver with approximately $H(X_1^N \mid Y_1^N) = NH(X_1 \mid Y_1)$ bits of information[1] about X_1^N for it to decode with small error probability. As we mentioned in the introduction, there are two cases where decoding is a trivial task: First, if $H(X_1 \mid Y_1) = 0$, the receiver can decode X_1^N with no other information than Y_1^N and make no errors. Second, if $H(X_1 \mid Y_1) = 1$, any strategy short of providing X_1^N itself to the receiver—which would render the receiver's task trivial—will result in unreliable decoding.

Arıkan's polarization technique is a method that transforms the X_1^N sequence so as to reduce the decoder's task into a series of these two trivial

[1] Logarithms in this chapter are to the base 2, and thus entropies of binary random variables are $[0, 1]$-valued.

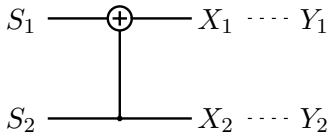

Figure 2.1: The first step of the recursive construction. The distribution on (S_1, S_2) is induced by the distribution on (X_1^2, Y_1^2).

tasks. While any good source or channel code can in fact be thought of in this way[2], Arıkan's technique yields low-complexity encoding and decoding algorithms due to its recursive nature.

2.1 A Basic Transform

In this section we review a single step of the polarization technique. Although the reader may find some of the details here trivial, we find it worthwhile to go through them since most polarization ideas are contained in the one-step construction.

Consider the case $N = 2$. Given (X_1, Y_1) and (X_2, Y_2), we define $S_1, S_2 \in \{0, 1\}$ through the mapping (see Figure 2.1)

$$S_1 = X_1 + X_2 \qquad \text{and} \qquad S_2 = X_2, \tag{2.1}$$

where '+' denotes modulo-2 addition. Notice that the correspondence between S_1, S_2 and X_1, X_2 is one-to-one, and therefore the independence of (X_1, Y_1) and (X_2, Y_2) implies

$$2H(X_1 \mid Y_1) = H(S_1^2 \mid Y_1^2) = H(S_1 \mid Y_1^2) + H(S_2 \mid Y_1^2 S_1).$$

It easily follows from (2.1) and the above equalities that

$$H(S_2 \mid Y_1^2 S_1) \leq H(X_1 \mid Y_1) \leq H(S_1 \mid Y_1^2). \tag{2.2}$$

Due to these entropy relations, one intuitively expects that observing $(Y_1^2 S_1)$ yields a more reliable estimate of S_2 (i.e., X_2) than observing Y_2 alone does. (It is in fact clear that the 'channel' $S_2 \to Y_1^2 S_1$ is *upgraded* with respect to the channel $X_2 \to Y_2$.) Similarly, observing Y_1^2 alone leads to a less reliable estimate of S_1. If we let $P_e(X_1 \mid Y_1)$ denote the average error probability of optimally decoding X_1 by observing Y_1, we indeed have

$$P_e(S_2 \mid Y_1^2 S_1) \leq P_e(X_1 \mid Y_1) \leq P_e(S_1 \mid Y_1^2). \tag{2.3}$$

The left-hand inequality above is obtained through the relations

$$P_e(S_2 \mid Y_1^2 S_1) \leq P_e(S_2 \mid Y_2) = P_e(X_1 \mid Y_1)$$

[2]A brief discussion on this is offered on pages 47–48.

and the right-hand inequality through

$$\begin{aligned} P_e(X_1 \mid Y_1) &= P_e(X_1 + X_2 \mid Y_1 X_2) \\ &= P_e(X_1 + X_2 \mid Y_1^2 X_2) \\ &\leq P_e(X_1 + X_2 \mid Y_1^2). \end{aligned}$$

The second equality above is due to the Markov chain $(X_1 + X_2)$—Y_1X_2—Y_2.

One can see the use of these relations in the following coding scheme: Upon observing X_1^2, the encoder computes S_1^2 and reveals S_1 to the receiver. The receiver then uses the optimal decision rule to decode S_2 from $(Y_1^2 S_1)$, and computes $(\hat{X}_1, \hat{X}_2) = (S_1 + \hat{S}_2, \hat{S}_2)$, where $\hat{S}_2$ is its estimate of S_2.

This is in fact the simplest instance of polar source coding, with code blocklength 2, rate 1/2, and average block error probability $P_e(S_2 \mid Y_1^2 S_1)$. Simple as it is, this scheme contains the essence of polarization and polar coding ideas: Out of two identical entropy terms $H(X_1 \mid Y_1)$ and $H(X_2 \mid Y_2)$, we have created two different entropies one of which is closer to 0 than the original and the other closer to 1, thereby approaching (albeit not very closely) the trivial cases we have mentioned above. By revealing to the decoder those random variables with high conditional entropies, we can decode those that have lower entropies with higher reliability.

2.2 An Improved Transform and Coding Scheme

Since the random variables S_1 and S_2 created by the above transform are $\{0, 1\}$-valued, one can apply the same transform to these in order to enhance the disparity between their entropies. In order to do so, let $N = 4$ and define, in addition to S_1, S_2 in (2.1),

$$T_1 = X_3 + X_4 \qquad \text{and} \qquad T_2 = X_4,$$

and also define $\tilde{Y}_1 = Y_1^2$ and $\tilde{Y}_2 = Y_3^4$ (see Figure 2.2). Observe that $(S_1, \tilde{Y}_1)$ and $(T_1, \tilde{Y}_2)$ are i.i.d., just as were (X_1, Y_1) and (X_2, Y_2). It then follows similarly to (2.2) that

$$H(T_1 \mid \tilde{Y}_1^2, S_1 + T_1) \leq H(S_1 \mid \tilde{Y}_1) \leq H(S_1 + T_1 \mid \tilde{Y}_1^2). \tag{2.4}$$

Similarly, defining $\bar{Y}_1 = (Y_1^2 S_1)$ and $\bar{Y}_2 = (Y_3^4 T_1)$ and noting that $(S_2, \bar{Y}_1)$ and $(T_2, \bar{Y}_2)$ are also i.i.d., we have

$$H(T_2 \mid \bar{Y}_1^2, S_2 + T_2) \leq H(S_2 \mid \bar{Y}_1) \leq H(S_2 + T_2 \mid \bar{Y}_1^2). \tag{2.5}$$

The relevance of the entropy terms above can be seen by an inspection of Figure 2.2. In particular, we have

$$\begin{aligned} 4H(X_1 \mid Y_1) &= 2H(S_1^2 \mid Y_1^2) \\ &= H(U_1^4 \mid Y_1^4) \\ &= H(U_1 \mid Y_1^4) + H(U_2 \mid Y_1^4 U_1) + H(U_3 \mid Y_1^4 U_1^2) + H(U_4 \mid Y_1^4 U_1^3). \end{aligned}$$

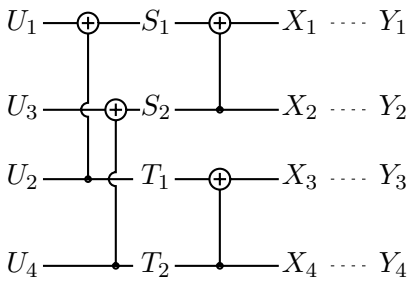

Figure 2.2

It is also easily seen that the last four entropy terms above are those appearing in (2.4) and (2.5):

$$\begin{aligned}
H(U_1 \mid Y_1^4) &= H(S_1 + T_1 \mid \tilde{Y}_1^2) \\
H(U_2 \mid Y_1^4 U_1) &= H(T_1 \mid \tilde{Y}_1^2, S_1 + T_1) \\
H(U_3 \mid Y_1^4 U_1^2) &= H(S_2 + T_2 \mid Y_1^4 S_1 T_1) = H(S_2 + T_2 \mid \bar{Y}_1^2) \\
H(U_4 \mid Y_1^4 U_1^3) &= H(T_2 \mid Y_1^4 S_1 T_1, S_2 + T_2) = H(T_2 \mid \bar{Y}_1^2, S_2 + T_2).
\end{aligned}$$

It follows from these relations, along with (2.4) and (2.5), that

$$\begin{aligned}
H(U_2 \mid Y_1^4 U_1) \leq H(S_1 \mid Y_1^2) \leq H(U_1 \mid Y_1^4) \\
H(U_4 \mid Y_1^4 U_1^3) \leq H(S_2 \mid Y_1^2 S_1) \leq H(U_3 \mid Y_1^4 U_1^2).
\end{aligned}$$

That is, from the two entropy terms $H(S_1 \mid Y_1^2)$ and $H(S_2 \mid Y_1^2 S_1)$ we obtain four new entropies that are separated from the original two as in the above inequalities. Since $H(S_1 \mid Y_1^2)$ and $H(S_2 \mid Y_1^2 S_1)$ were somewhat polarized towards 1 and 0, the above inequalities say that the polarization effect is enhanced by the second application of the transform.

Consider now the following source code of blocklength 4: We choose a set $\mathcal{A} \subset \{1, 2, 3, 4\}$ with $|\mathcal{A}| = 4 - k$. Upon observing $X_1^4 = x_1^4$, the encoder computes $U_1^4 = u_1^4$ and sends all $u_i, i \in \mathcal{A}^c$ to the decoder, therefore the rate of the code is $k/4$ bits/symbol. The decoder outputs its estimate $\hat{u}_1^4$ of u_1^4 *successively* as

$$\hat{u}_i = \begin{cases} u_i & \text{if } i \in \mathcal{A} \\ 0 & \text{if } i \in \mathcal{A}^c \text{ and } L(y_1^4, \hat{u}_1^{i-1}) > 1\,, \\ 1 & \text{otherwise} \end{cases} \tag{2.6}$$

where

$$L(y_1^4, u_1^{i-1}) = \frac{\Pr[U_i = 0 \mid Y_1^4 = y_1^4, U_1^{i-1} = \hat{u}_1^{i-1}]}{\Pr[U_i = 1 \mid Y_1^4 = y_1^4, U_1^{i-1} = \hat{u}_1^{i-1}]}. \tag{2.7}$$

A sensible choice of set $\mathcal{A}$ that will yield a small error probability under the above decoding scheme is

$$\mathcal{A} = \big\{i\colon P_e(U_i \mid Y_1^4 U_1^{i-1}) \text{ is among the } k \text{ smallest}\big\}.$$

This choice can be justified by the following result:

Proposition 2.1. *The average block error probability of the above coding scheme is at most*

$$\sum_{i\in\mathcal{A}} P_e(U_i \mid Y_1^4 U_1^{i-1}). \tag{2.8}$$

Proof. Consider a decoder with output $\tilde{u}_1^N$, whose decision rule for $\tilde{u}_i$ is obtained from (2.6) by replacing $L(y_1^4, \hat{u}_1^{i-1})$ with $L(y_1^4, u_1^{i-1})$. This is a *genie-aided* version of the original decoder: at each step of decoding, a genie provides the decoder with the correct value of the previously decoded bits. Clearly, the average error probability of the ith constituent of this decoder is $P_e(U_i \mid Y_1^4 U_1^{i-1})$, and therefore the block error probability is upper bounded by the expression in (2.8). In order to conclude the proof, we will show that the block error events for the original decoder described in (2.6)–(2.7) and its genie-aided version are identical. To see the latter claim, note that $\hat{u}_1 = \tilde{u}_1$ for each realization (y_1^4, u_1^4), as both decisions depend on $L(y_1^4)$ alone. Hence, if $\hat{u}_1 = \tilde{u}_1 = u_1$ (otherwise both decoders commit a block error in the first step), it then follows that $\hat{u}_2 = \tilde{u}_2$, as both decisions are based on $L(y_1^4, u_1)$. Continuing in this manner, we see that at each step, either both decoders have already committed an error, or their next decisions will be identical. This in turn implies that the block error events (but not necessarily the bit error events) under the original decoder and its genie-aided version are identical, yielding the claim. □

Proposition 2.1 highlights two simple but important aspects of the design and analysis of polar codes (of which the above code is an instance). First, the block error probability behavior of these codes can be deduced from the error behavior of the created 'channels' (e.g., channels $U_i \to Y_1^4 U_1^{i-1}$ above), which as we will see greatly simplifies error analysis. Second, minimizing the upper bound in (2.8) amounts to finding a good code, as it consists in determining the bit indices with the smallest probability of decoding error. This is one of the several appeals of polar codes: their design and construction on one hand and analysis on the other are closely linked and do not require separate techniques.

2.3 Recursive Construction: Polarization

We saw the first two steps of Arıkan's construction in the previous sections. The recursive nature of this construction is evident: The second step merely involves applying the transform in (2.1) to the random variables obtained in

the first. Similarly, in the general form of this construction, each recursion consists in applying (2.1) to the random variables obtained in the previous one. For this technique to create the desired effect of driving the entropies close to 0 and 1, it is therefore necessary that the basic transform in (2.1) lead to a strict separation of entropies, i.e., that the inequalities in (2.2) be strict, for otherwise the transform would have no effect. The following result guarantees that this requirement is always met, except in trivial cases.

Lemma 2.1. *Let $\alpha, \beta \in [0,1]$ and also let (X_1, Y_1) and (X_2, Y_2) be independent pairs of discrete random variables with $X_1, X_2 \in \{0,1\}$, $H(X_1 \mid Y_1) = \alpha$, and $H(X_2 \mid Y_2) = \beta$. Then, the entropy $H(X_1 + X_2 \mid Y_1^2)$*

(i) is minimized when $H(X_1 \mid Y_1 = y_1) = \alpha, H(X_2 \mid Y_2 = y_2) = \beta$ for all y_1, y_2 with $p(y_1), p(y_2) > 0$.

(ii) is maximized when $H(X_1 \mid Y_1 = y_1), H(X_2 \mid Y_2 = y_2) \in \{0,1\}$ for all y_1, y_2 with $p(y_1), p(y_2) > 0$.

It also follows from (i) that if $\alpha, \beta \in (\delta, 1-\delta)$ for some $\delta > 0$, then there exists $\epsilon(\delta) > 0$ such that

$$H(X_1 + X_2 \mid Y_1^2) - H(X_1 \mid Y_1) \geq \epsilon(\delta).$$

Proof. See Appendix 2.A. □

We can now describe the general form of the polarization construction: Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be an i.i.d. sequence as above. For $n = 0, 1, \ldots$, let $N = 2^n$ and define a sequence of transforms $G_n \colon \{0,1\}^N \to \{0,1\}^N$ recursively through

$$\begin{aligned} G_0(u) &= u, \\ G_n(u_1, u_2) &= \pi_n\big(G_{n-1}(u_1) + G_{n-1}(u_2), G_{n-1}(u_2)\big) \quad n = 1, 2, \ldots \end{aligned}$$

where $u = (u_1, u_2)$ and $\pi_n \colon \{0,1\}^N \to \{0,1\}^N$ permutes the components of its argument vector through

$$\begin{aligned} \pi_n(u)_{2i-1} &= u_i \\ \pi_n(u)_{2i} &= u_{i+N/2} \end{aligned} \quad , \qquad i = 1, \ldots, N/2.$$

It is easy to show [3] that G_n is one-to-one and that $G_n^{-1} = G_n$. Now define

$$U_1^N = G_n(X_1^N).$$

The general form of the transform G_n is shown in Figure 2.3. The inclusion of π_n in the definition of G_n is not necessary for the polarization technique to work, but it will greatly simplify the notation. Observe that G_1 and G_2 are equivalent to the transforms in the previous sections (Figures 2.1 and 2.2).

The main result in [3] and [7] is that as the construction size N grows, the entropies $H(U_i \mid Y_1^N U_1^{i-1})$ approach either 0 or 1:

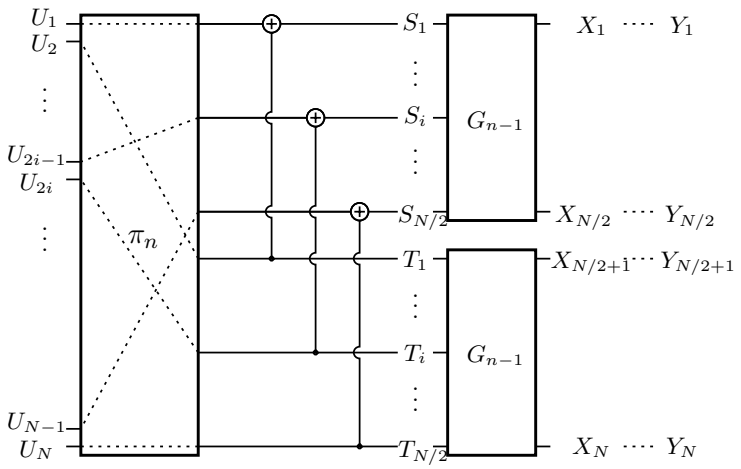

Figure 2.3

Theorem 2.1. *For all $\epsilon > 0$,*

$$\lim_{n\to\infty} \frac{1}{N} \left|\left\{i\colon H(U_i \mid Y_1^N U_1^{i-1}) > 1-\epsilon\right\}\right| = H(X \mid Y),$$
$$\lim_{n\to\infty} \frac{1}{N} \left|\left\{i\colon H(U_i \mid Y_1^N U_1^{i-1}) < \epsilon\right\}\right| = 1 - H(X \mid Y).$$

In order to simplify the notation in the proofs, we will often use the following definition.

Definition 2.1. *For i.i.d. (X_1, Y_1) and (X_2, Y_2) with $H := H(X_1 \mid Y_1)$, we define*

$$\begin{aligned} H^- &:= H(X_1 + X_2 \mid Y_1^2), \\ H^+ &:= H(X_2 \mid Y_1^2, X_1 + X_2). \end{aligned} \tag{2.9}$$

With the above definitions, we claim that

$$\begin{aligned} H(U_1 \mid Y_1^N) &= H^{-\cdots--} \\ H(U_2 \mid Y_1^N U_1) &= H^{-\cdots-+} \\ H(U_3 \mid Y_1^N U_1^2) &= H^{-\cdots+-} \\ &\vdots \\ H(U_{N-1} \mid Y_1^N U_1^{N-2}) &= H^{+\cdots+-} \\ H(U_N \mid Y_1^N U_1^{N-1}) &= H^{+\cdots++}, \end{aligned} \tag{2.10}$$

where the superscripts on the right-hand terms are of length n. These equivalences can be verified by an inspection of Figure 2.3. In particular, let us suppose that claim (2.10) holds for the entropy terms obtained after G_{n-1}, i.e., that for every $1 \le i \le N/2$ there is a distinct $\mathbf{s} \in \{-,+\}^{n-1}$ such that $H(S_i \mid$

$Y_1^{N/2}S_1^{i-1}) = H^{\mathbf{s}}$. Then, since the pairs $(S_i, Y_1^{N/2}S_1^{i-1})$ and $(T_i, Y_{N/2+1}^{N}T_1^{i-1})$ in the figure are i.i.d., it is easily seen that $H(U_{2i-1} \mid Y_1^N U_1^{2i-2}) = H(S_i \mid Y_1^{N/2}S_1^{i-1})^- = H^{\mathbf{s}-}$, and that $H(U_{2i} \mid Y_1^N U_1^{2i-1}) = H(S_i \mid Y_1^{N/2}S_1^{i-1})^+ = H^{\mathbf{s}+}$. It follows that for every $i \in \{1, \ldots, N\}$ there is a distinct $\mathbf{s} \in \{-,+\}^n$ such that $H(U_i \mid Y_1^N U_1^{i-1}) = H^{\mathbf{s}}$. It also follows from the definition of the permutation function π_n that these equivalences are as in (2.10). Since we have already seen in Section 2.1 that (2.10) holds for $n = 1$, it follows by induction that it holds for all n.

In order to prove Theorem 2.1 we define an i.i.d. process $B_1, B_2, \ldots$ where B_1 is uniformly distributed over $\{-,+\}$. We then define a $[0,1]$-valued random process $H_0, H_1, \ldots$ recursively as

$$\begin{aligned} H_0 &= H(X_1 \mid Y_1), \\ H_n &= H_{n-1}^{B_n}, \qquad n = 1, 2, \ldots \end{aligned} \tag{2.11}$$

As $B_1, \ldots, B_n$ is uniformly distributed over $\{-,+\}^n$, the equivalence of entropies in (2.10) imply that for all n,

$$\Pr[H_n \in \mathcal{I}] = \frac{1}{N}\left|\left\{i\colon H(U_i \mid Y_1^N U_1^{i-1}) \in \mathcal{I}\right\}\right|$$

for any $\mathcal{I} \subseteq [0,1]$. Therefore, Theorem 2.1 is implied by

Theorem 2.1*. *H_n converges almost surely to a $\{0,1\}$-valued random variable H_∞ with $\Pr[H_\infty = 1] = 1 - \Pr[H_\infty = 0] = H(X_1 \mid Y_1)$.*

Proof. Definitions (2.9) and (2.11) imply that $H_n^- + H_n^+ = 2H_n$. It follows that the process $H_1, H_2, \ldots$ is a bounded martingale and therefore converges almost surely to a random variable H_∞. As almost sure convergence implies convergence in $\mathcal{L}^1$, we have $E[|H_{n+1} - H_n|] = \frac{1}{2}E[H_n^- - H_n] + \frac{1}{2}E[H_n - H_n^+] = E[H_n^- - H_n] \to 0$. Also since Lemma 2.1 implies that $H_n^- - H_n > \delta(\epsilon)$ if $H_n \in (\epsilon, 1-\epsilon)$, it follows that $H_n \to \{0,1\}$ with probability 1, i.e., that H_∞ is $\{0,1\}$-valued. The claim on the distribution of H_∞ then follows from the relation $E[H_\infty] = E[H_0] = H(X_1 \mid Y_1)$. □

This is the main polarization theorem. It states that Arıkan's construction distills the randomness in an i.i.d. binary process into a sequence of uniform or constant binary random variables. Equivalently, this construction can be interpreted as one that creates a sequence of noiseless and useless channels $U_i \to Y_1^N U_1^{i-1}$ out of several copies of a memoryless channel $X_1 \to Y_1$.

Theorem 2.1 can be exploited to construct entropy-achieving polar source codes as follows: We fix $\delta, \epsilon > 0$, and find the set

$$\mathcal{A} := \{i : P_e(U_i \mid Y_1^N U_1^{i-1}) \leq \epsilon\}.$$

As $H(U_i \mid Y_1^N U_1^{i-1}) \to 0$ implies $P_e(U_i \mid Y_1^N U_1^{i-1}) \to 0$, it follows from Theorem 2.1 that $\mathcal{A}$ must be of size at least $(1 - H(X \mid Y) - \delta)N$ provided that

the blocklength N is sufficiently large. The encoder observes X_1^N, computes $U_1^N = G_n(X_1^N)$, and reveals $U_i, i \in \mathcal{A}^c$ to the receiver, i.e., the code is of rate $H(X \mid Y) + \delta$. Upon observing Y_1^N and $U_i, i \in \mathcal{A}^c$, the receiver decodes U_1^N successively as in (2.6) and (2.7). Similarly to the previous section, the block error probability of this code is at most

$$\sum_{i \in \mathcal{A}} P_e(U_i \mid Y_1^N U_1^{i-1}) \leq \epsilon N.$$

This bound on the error probability is not very useful, however, as we have chosen the threshold ϵ independently of N. Fortunately, the choice of set $\mathcal{A}$ in the above scheme can be modified slightly to include a blocklength-dependent ϵ, yielding codes with vanishing block error probability. More precisely, instead of $\mathcal{A}$ consider the set

$$\mathcal{A}'_\beta := \{i \colon P_e(U_i \mid Y_1^N U_1^{i-1}) \leq 2^{-N^\beta}\}$$

for some $\beta > 0$. Note that for large N we have $\mathcal{A}'_\beta \subset \mathcal{A}$. The next result states that as long as $\beta < 1/2$, the set difference $\mathcal{A} \backslash \mathcal{A}'_\beta$ is negligibly small, in the sense that $|\mathcal{A}'_\beta|/|\mathcal{A}| \to 1$. That is, at large blocklengths if the bit error probability $P_e(U_i \mid Y_1^N U_1^{i-1})$ is small, then it must indeed be exponentially small in the square root of the blocklength.

Theorem 2.2. *For all $\beta < 1/2$ and $\delta > 0$, there exists $N_o = N_o(\beta, \delta)$ such that*

$$|\mathcal{A}'_\beta| > (1 - H(X \mid Y) - \delta)N$$

for all $N \geq N_o$.

Corollary 2.1. *For all $\beta < 1/2$ and rates strictly above $H(X \mid Y)$, the average block error probability of the above source coding scheme is $o(2^{-N^\beta})$.*

In order to prove Theorem 2.2 one needs to compute the $P_e(U_i \mid Y_1^N U_1^{i-1})$ terms during the polarization process. The difficulty in doing so is that the joint distributions of $(U_i, Y_1^N U_1^{i-1})$ become increasingly complex as the blocklength grows, and consequently the exact computation of error probabilities becomes intractible. One may hope instead to find useful bounds on the error probabilities that are also independent of the details of the joint distributions. For this purpose, consider a $[0, 1]$-valued parameter $Z(X \mid Y)$ defined as

$$Z(X \mid Y) = 2 \sum_{y \in \mathcal{Y}} \sqrt{p_{XY}(0, y) p_{XY}(1, y)}.$$

Arıkan calls $Z(X \mid Y)$ the *source Bhattacharyya parameter* [7]. It is well-known that the Bhattacharyya parameter upper bounds the error probability of the optimal decision rule, and therefore may be used as a measure of reliability:

Proposition 2.2. $P_e(X \mid Y) \leq Z(X \mid Y)$.

Proof.

$$\begin{aligned}
P_e(X \mid Y) &\leq p_X(0) \sum_y p(y \mid 0) \mathbb{1}_{[p(0|y) \leq p(1|y)]} + p_X(1) \sum_y p(y \mid 1) \mathbb{1}_{[p(1|y) \leq p(0|y)]} \\
&\leq p_X(0) \sum_y \frac{p(0 \mid y)p(y)}{p_X(0)} \frac{\sqrt{p(1 \mid y)}}{\sqrt{p(0 \mid y)}} \\
&\quad + p_X(1) \sum_y \frac{p(1 \mid y)p(y)}{p_X(1)} \frac{\sqrt{p(0 \mid y)}}{\sqrt{p(1 \mid y)}} \\
&= 2 \sum_y \sqrt{p(0, y)p(1, y)} \\
&= Z(X \mid Y).
\end{aligned}$$

□

As a measure of reliability, it would be natural for $Z(X \mid Y)$ to satisfy

$$\begin{aligned}
Z(X \mid Y) \approx 1 &\iff H(X \mid Y) \approx 1, \\
Z(X \mid Y) \approx 0 &\iff H(X \mid Y) \approx 0.
\end{aligned}$$

The following relations show that this is indeed the case:

Proposition 2.3 ([7])**.**

$$\begin{aligned}
Z(X \mid Y)^2 &\leq H(X \mid Y) \\
H(X \mid Y) &\leq \log(1 + Z(X \mid Y)).
\end{aligned}$$

One may also expect to observe a disparity between the Bhattacharyya parameters after one step of the polarization transform, similar to the disparity between the entropies (2.2) and the error probabilities (2.3). We indeed have

$$Z(U_2 \mid Y_1^2 U_1) \leq Z(X_1 \mid Y_1) \leq Z(U_1 \mid Y_1^2).$$

It can also be shown that these inequalities are strict unless $Z(X_1 \mid Y_1)$ is either 0 or 1. Clearly, the exact values of these parameters depend on the details of the joint distribution of (X_1, Y_1). Nevertheless, there are bounds on these that are distribution-independent and are also sufficiently good for proving Theorem 2.2:

Lemma 2.2. *For all (X_1, Y_1), we have*

$$Z(U_1 \mid Y_1^2) \leq 2Z(X_1 \mid Y_1), \tag{2.12}$$

$$Z(U_2 \mid Y_1^2 U_1) = Z(X_1 \mid Y_1)^2. \tag{2.13}$$

Proof. First note that $p(u_1, u_2, y_1, y_2) = p_{XY}(u_1 + u_2, y_1)p_{XY}(u_2, y_2)$. The first bound can be seen through the following inequalities:

$$\begin{aligned}
Z(U_1 \mid Y_1^2) &= 2\sum_{y_1^2}\Bigg[\sum_{u_2} p_{XY}(u_2, y_1)p_{XY}(u_2, y_2) \\
&\qquad\qquad \cdot \sum_{v_2} p_{XY}(1 + v_2, y_1)p_{XY}(v_2, y_2)\Bigg]^{1/2} \\
&\le 2\sum_{y_1^2, u_2, v_2}\Big[p_{XY}(u_2, y_1)p_{XY}(1 + v_2, y_1)p_{XY}(u_2, y_2)p_{XY}(v_2, y_2)\Big]^{1/2} \\
&= 2\sum_{u_2, v_2}\sum_{y_1}\big[p_{XY}(u_2, y_1)p_{XY}(1 + v_2, y_1)\big]^{1/2} \\
&\qquad\qquad \cdot \sum_{y_2}\big[p_{XY}(u_2, y_2)p_{XY}(v_2, y_2)\big]^{1/2}
\end{aligned}$$

The term inside the outermost summation is equal to $p(u_2)Z(X_1 \mid Y_1)/2$ for all u_2, v_2. This yields the first claim. To obtain the second claim we write

$$\begin{aligned}
Z(U_2 \mid Y_1^2 U_1) &= 2\sum_{y_1^2, u_1}\big[p_{XY}(u_1, y_1)p_{XY}(0, y_2)p_{XY}(u_1 + 1, y_1)p_{XY}(1, y_2)\big]^{1/2} \\
&= 2\sum_{u_1}\sum_{y_1}\big[p_{XY}(u_1, y_1)p_{XY}(u_1 + 1, y_1)\big]^{1/2} \\
&\qquad\qquad \cdot \sum_{y_2}\big[p_{XY}(0, y_2)p_{XY}(1, y_2)\big]^{1/2} \\
&= 4\Bigg[\sum_{y}\big[p_{XY}(0, y)p_{XY}(1, y)\big]^{1/2}\Bigg]^2 \\
&= Z(X_1 \mid Y_1)^2.
\end{aligned}$$

□

In order to prove Theorem 2.2, we will define, similarly to the proof of Theorem 2.1, a random process that mirrors the behavior of the Bhattacharyya parameters obtained during the polarization construction. For this purpose, we first let $Z := Z(X_1 \mid Y_1)$ and define

$$\begin{aligned}
Z^- &:= Z(U_1 \mid Y_1^2), \\
Z^+ &:= Z(U_2 \mid Y_1^2 U_1).
\end{aligned}$$

We will see that bounds (2.12) and (2.13) on Z^- and Z^+ suffice to prove Theorem 2.2. To get an initial idea about the reason for this, let us neglect, for a moment, the factor 2 in the bound (2.12) on Z^-. It is now easy to see that on a 'polarization path' consisting of n consecutive '+' and '−' operations, the resulting $Z(U_i \mid Y_1^N U_1^{i-1})$ will be upper bounded by $Z(X \mid Y)^{2^{n_p}}$, where n_p is the number of the occurrences of '+'. Since on a typical path the plus and the minus operations occur with roughly the same frequency, i.e., $n_p \approx n/2$,

it follows that most Bhattacharyya parameters will be of the form $Z(U_i \mid Y_1^N U_1^{i-1}) \approx Z(X \mid Y)^{2^{n/2}} = Z(X \mid Y)^{\sqrt{N}}$, as claimed in Theorem 2.2.

The reason for us to resort to Bhattacharyya parameters instead of working directly with error probabilities is the lack of useful bounds on the latter. More precisely, although we have

$$P_e(U_2 \mid Y_1^2 U_1) \leq P_e(X_1 \mid Y_1) \leq P_e(U_1 \mid Y_1^2)$$

after the first step of polarization, how close these error terms are to each other depends strongly on the distribution of (X_1, Y_1). In particular, it can easily be verified that if X_1 is uniformly distributed and Y_1 is the output of an arbitrary binary symmetric channel whose input is X_1, then the left-hand bound above is satisfied with equality. In other words, the tightest upper bound on $P_e(U_2 \mid Y_1^2 U_1)$ in terms of $P_e(X_1 \mid Y_1)$ only (i.e., independent of the particular distribution of X_1 and Y_1) is

$$P_e(U_2 \mid Y_1^2 U_1) \leq P_e(X_1 \mid Y_1).$$

Comparing this with (2.13) reveals the advantage of the latter.

We will prove Theorem 2.2 as a corollary to Lemma 2.2 and the following result.

Lemma 2.3. *Let $B_1, B_2, \ldots$ be an i.i.d. binary process where B_1 is uniformly distributed over $\{-, +\}$. Also let $Z_0, Z_1, \ldots$ be a $[0, 1]$-valued random process where Z_0 is constant and*

$$Z_{n+1} \leq \begin{cases} K Z_n & \text{if } B_n = - \\ K Z_n^2 & \text{if } B_n = + \end{cases}$$

for some finite $K > 0$. Suppose also that Z_n converges almost surely to a $\{0, 1\}$-valued random variable Z_∞ with $\Pr[Z_\infty = 0] = z$. Then, for any $\beta < 1/2$,

$$\lim_{n \to \infty} \Pr[Z_n \leq 2^{-2^{n\beta}}] = z.$$

We defer the proof of Lemma 2.3 until Chapter 4, where we prove a more general result. We are now ready to prove Theorem 2.2:

Proof of Theorem 2.2. We will show that for all $\delta > 0$ and sufficiently large N, the size of the set

$$\mathcal{A}''_\beta := \left\{ i \colon Z(U_i \mid Y_1^N U_1^{i-1}) \leq 2^{-N^\beta} \right\}$$

is at least $(1 - H(X \mid Y) - \delta)N$, which will yield the lemma since the Bhattacharyya parameter upper bounds the average error probability. For this

purpose, observe that the Bhattacharyya parameters obtained along the polarization construction satisfy the equalities

$$\begin{aligned} Z(U_1 \mid Y_1^N) &= Z^{-\cdots--} \\ Z(U_2 \mid Y_1^N U_1) &= Z^{-\cdots-+} \\ Z(U_3 \mid Y_1^N U_1^2) &= Z^{-\cdots+-} \\ &\vdots \\ Z(U_{N-1} \mid Y_1^N U_1^{N-2}) &= Z^{+\cdots+-} \\ Z(U_N \mid Y_1^N U_1^{N-1}) &= Z^{+\cdots++}, \end{aligned} \tag{2.14}$$

for any N. As in the proof of Theorem 2.1, define an i.i.d. process $B_1, B_2, \ldots$ with $\Pr[B_1 = -] = \Pr[B_1 = +] = 1/2$, and a $[0,1]$-valued process $Z_0, Z_1, \ldots$ with

$$\begin{aligned} Z_0 &= Z(X \mid Y) \\ Z_n &= Z_{n-1}^{B_n}, \qquad n = 1, 2, \ldots \end{aligned}$$

Observe that $B_1, B_2, \ldots$ induces a uniform distribution on Z_n over the set $\left\{Z^{-\cdots--}, \ldots, Z^{+\cdots++}\right\}$, and that Proposition 2.3 implies the almost sure convergence of Z_n to the set $\{0,1\}$ with $\Pr[\lim_{n\to\infty} Z_n = 0] = 1 - H(X \mid Y)$. The claim then follows follows from Lemma 2.3. □

It is evident that the bounds in Lemma 2.2 are the only properties of the polarization construction that have a bearing upon the above proof. This brings out another technical appeal of polar codes: their large blocklength behavior can be inferred directly from the effect of the underlying one-step transformation on the Bhattacharyya parameters. This proves especially useful when one considers polar codes based on combining more than two random variables at a time. The recursive nature of such constructions ensure that the error probability behavior of the resulting codes can be analyzed with relative ease. We will discuss these constructions and their analysis in Chapter 4.

2.4 Polar Channel Coding

In the previous section, we saw an entropy-achieving source coding scheme whose average error probability decays roughly exponentially in the square root of the blocklength. We will now see that the techniques we have reviewed can be used, almost verbatim, to obtain capacity-achieving codes for binary-input symmetric memoryless channels. Consider a binary-input discrete memoryless channel $W \colon \{0,1\} \to \mathcal{Y}$. Let $X_1, \ldots, X_N$ be a sequence of i.i.d. inputs to N uses of W, and let $Y_1, \ldots, Y_N$ be the corresponding output (see Figure 2.4). Since the channel is memoryless and the inputs are i.i.d., the sequence $(X_1, Y_1), \ldots, (X_N, Y_N)$ is also i.i.d. This is exactly the same situation as in the previous sections, and one can imagine the following transmission scheme,

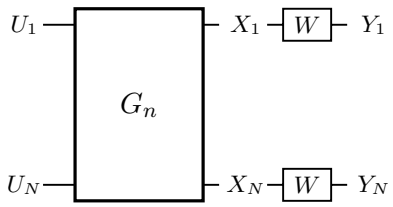

Figure 2.4

which mimics the techniques we have seen: To send the message corresponding to X_1^N, the encoder first computes $U_1^N = G_n(X_1^N)$ and reveals the bits with $P_e(U_i \mid Y_1^N U_1^{i-1}) \geq 2^{-N^\beta}$ to the decoder, and sends X_1^N through the channel. Upon receiving the channel output Y_1^N, the receiver decodes the unknown part of U_1^N successively as in (2.6) and (2.7). It follows from Theorem 2.2 that the average block error probability of this coding scheme is $O(2^{-N^\beta})$. Note that while all length-N binary sequences are potential codewords in this scheme, a codeword chosen in an i.i.d. fashion will belong to the 'typical set' of size $\approx 2^{NH(X)}$ with high probability. Further, since approximately $NH(X \mid Y)$ bits of information are revealed to the receiver in advance, the effective rate of this code is approximately $I(X;Y)$. Hence, by assigning the appropriate distribution to X_1, the capacity of the channel can be achieved.

The above coding argument is identical to the one in Section 2.3 but, while it is mathematically correct, it is inadequate from a channel coding perspective: First, observe that in the channel coding problem, the distribution on the channel inputs X_1^N is induced by the encoder's choice of the distribution on U_1^N. This is in contrast with the source coding case, where the distribution of X_1^N is intrinsic to the source, and the distribution of U_1^N is induced by the transformation G_n. The difficulty is that in order to generate i.i.d. inputs X_1^N to the channel, the encoder would have to choose U_1^N from a non-uniform distribution, conflicting with the common assumption that the sender's messages are uniformly distributed. Second, in the source coding problem the values of the bits to be revealed to the receiver depend on the realization of the source X_1^N. In channel coding, however, these values need to be revealed to the receiver prior to communication, and therefore cannot depend on the particular message to be sent as proposed in the above scheme.

The first of these issues is of a somewhat technical nature, and can be dealt with most easily by insisting on uniformly distributed channel inputs X_1^N since this would impose a uniform distribution on U_1^N. One can also circumvent the second issue by choosing the bits to be revealed in advance, and taking averages over the values of these bits. To make these arguments precise, let us consider the following coding scheme:

Code construction: Given a blocklength $N = 2^n$, fix $0 < \beta' < \beta < 1/2$ and find the set

$$\mathcal{A}_\beta := \{i \colon P_e(U_i \mid Y_1^N U_1^{i-1}) \leq 2^{-N^\beta}\}.$$

Choose $U_i, i \in \mathcal{A}_\beta^c$ independently and uniformly at random, and reveal their values to the receiver. The rate of the code will be $|\mathcal{A}_\beta|/N$.

Encoding: Given a uniformly distributed message $M \in \{0,1\}^{|\mathcal{A}_\beta|}$ to be transmitted, set $U_{\mathcal{A}_\beta} = M$. Transmit $X_1^N = G_n(U_1^N)$ over the channel.

Decoding: Upon receiving Y_1^N, the receiver decodes U_1^N successively as in (2.6) and (2.7).

Rate and error probability: As X_1^N is i.i.d. uniform, we have $H(X) = 1$, and therefore it follows from Theorem 2.2 that if N is sufficiently large, the rate of the code is

$$|\mathcal{A}_\beta|/N > 1 - H(X \mid Y) - \delta = I(X;Y) - \delta.$$

Note that $I(X;Y)$ here is the *symmetric capacity* of the channel $X \to Y$, the maximum rate achievable by binary codebooks with an equal fraction of zeros and ones. Note also that this is the true capacity for symmetric channels. It similarly follows from Theorem 2.2 and Proposition 2.1 that the block error probability of the above scheme, averaged over all messages and values of $U_i, i \in \mathcal{A}^c$, is $o(2^{-N^{\beta'}})$. Therefore there exists at least one set of values of bits $U_i, i \in \mathcal{A}^c$ (so-called the *frozen bits*) for which the average block error probability of the resulting code is at most $o(2^{-N^{\beta'}})$.

2.5 Complexity

An important practical issue that we did not discuss in this review is computational complexity. It is clear from the coding schemes we have seen that there are three problems of complexity that need to be addressed: (i) complexity of encoding, i.e., computing the function G_n, (ii) complexity of decoding, i.e., computing the probabilities appearing in equation (2.6), and (iii) complexity of construction, i.e., determining the set of bit indices with small error probabilities. Thanks to the recursive nature of the construction, all three tasks can be broken down to similar tasks of smaller sizes. An $O(N \log N)$ (both time and space complexities on a single-processor machine that performs infinite-precision arithmetic in unit time) encoding and decoding algorithm that exploits this structure was proposed in [3]. Later, Tal and Vardy [5] proposed an algorithm to determine the reliable bit indices, with time complexity $O(N)$ and space complexity $O(\log N)$. We refer the reader to these references for the details.

In the next chapter, we will study polarization for memoryless processes with arbitrary discrete alphabets. We will see that all such processes can be polarized by a recursive application of an appropriately chosen transform.

2.A Proof of Lemma 2.1

Let R_1 and R_2 be $[0, 1/2]$-valued random variables defined through

$$R_1 = \min\{p_{X_1|Y_1}(0 \mid y_1), p_{X_1|Y_1}(1 \mid y_1)\} \text{ whenever } Y_1 = y_1,$$
$$R_2 = \min\{p_{X_2|Y_2}(0 \mid y_2), p_{X_2|Y_2}(1 \mid y_2)\} \text{ whenever } Y_2 = y_2.$$

For $a, b \in [0, 1]$ define

$$a * b = a(1 - b) + (1 - a)b.$$

Also let $h\colon [0, 1/2] \to [0, 1]$ denote the binary entropy function. With these definitions, we have

$$H(X_1 + X_2 \mid Y_1^2) = E[h(R_1 * R_2)].$$

Both claims of the lemma follow from the convexity of the function $h\big(a*h^{-1}(t)\big)$ in $t \in [0, 1/2]$, which was established in [8]. In particular, we have

$$\begin{aligned} H(X_1 + X_2 \mid Y_1^2) &= E\left[h(R_1 * R_2)\right] \\ &= E\big[E\left[h(R_1 * R_2)\right] \mid R_1\big] \\ &= E\left[E\left[h\big(R_1 * h^{-1}(h(R_2))\big)\right] \mid R_1\right] \\ &\geq E\left[h\big(R_1 * h^{-1}(E\left[h(R_2)\right])\big)\right] \\ &= E\left[h\big(R_1 * h^{-1}(\beta)\big)\right]. \end{aligned}$$

Applying the convexity of $h(a * h^{-1}(t))$ a second time we obtain

$$\begin{aligned} H(X_1 + X_2 \mid Y_1^2) &\geq E\left[h\big(R_1 * h^{-1}(\beta)\big)\right] \\ &= E\left[h\big(h^{-1}(h(R_1)) * h^{-1}(\beta)\big)\right] \\ &\geq h\big(h^{-1}(E\left[h(R_1)\right]) * h^{-1}(\beta)\big) \\ &= h\big(h^{-1}(\alpha) * h^{-1}(\beta)\big). \end{aligned}$$

It is easy to see that the last term is the equal to $H(X_1 + X_2 \mid Y_1^2)$ when (X_1, Y_1) and (X_2, Y_2) are distributed as in (i), yielding the claim. To see the second claim, note that the convexity of $h\big(a * h^{-1}(t)\big)$ implies

$$\begin{aligned} h(a * h^{-1}(t)) &\leq th\big(a * h^{-1}(1)\big) + (1 - t)h(a * h^{-1}(0)) \\ &= t + (1 - t)h(a). \end{aligned}$$

It then follows that

$$\begin{aligned} H(X_1 + X_2 \mid Y_1^2) &= E\left[h\big(R_1 * R_2\big)\right] \\ &= E\left[h\big(R_1 * h^{-1}(h(R_2))\big)\right] \\ &\leq E\left[h(R_1) + h(R_2) - h(R_1)h(R_2)\right] \\ &= E[h(R_1)] + E[h(R_2)] - E[h(R_1)]E[h(R_2)]. \end{aligned}$$

where the last equality follows from the independence between R_1 and R_2. A simple calculation shows that the last term is equal to $H(X_1 + X_2 \mid Y_1^2)$ when (X_1, Y_1) and (X_2, Y_2) are distributed as in (ii), completing the proof.

Memoryless Processes with Arbitrary Discrete Alphabets

3

We saw in Chapter 2 that Arıkan's recursive method creates random variables with extremal entropies out of a binary memoryless process with moderate entropy. The cause of this polarization effect is simple: If a memoryless process $(X_1, Y_1), (X_2, Y_2), \ldots$ with binary X_1 has moderate entropy $H = H(X_1 \mid Y_1) \in (\epsilon, 1-\epsilon)$, then the entropies $H^- = H(U_1 \mid Y_1^2)$ and $H^+ = H(U_2 \mid Y_1^2 U_1)$ of

$$U_1 = X_1 + X_2 \quad \text{and} \quad U_2 = X_2 \tag{3.1}$$

are strictly away from each other (Lemma 2.1), i.e.,

$$H^+ + \delta(\epsilon) \leq H \leq H^- - \delta(\epsilon) \quad \text{for some } \delta(\epsilon) > 0. \tag{3.2}$$

This is illustrated in Figure 3.1. If H^- and H^+ are also moderate, applying (3.1) a second time will cause further separation in the resulting entropies. Continuing in this fashion, we see that if the 'entropy paths' we create converge at all—they indeed do—they can converge only to zero or to one, yielding polarization. It is then clear that for polarization to take place, the only requirement for a recursive transform and the underlying process is that the resulting entropies satisfy (3.2) at each step. This raises the question with which much of this thesis is concerned: What classes of processes can be polarized recursively, and what types of transforms polarize these processes?

By the end of this monograph, it will become clear that polarization is a fairly general phenomenon, taking place for a large class of processes, and under a large class of constructions. We will begin demonstrating this generality by showing how to polarize non-binary memoryless processes. Our motivation for this study is simple: Several source and channel coding problems of practical interest are in a non-binary setting. Perhaps the most prominent example is the additive white Gaussian channel, where the coding gains achieved by using non-binary inputs can be significant.

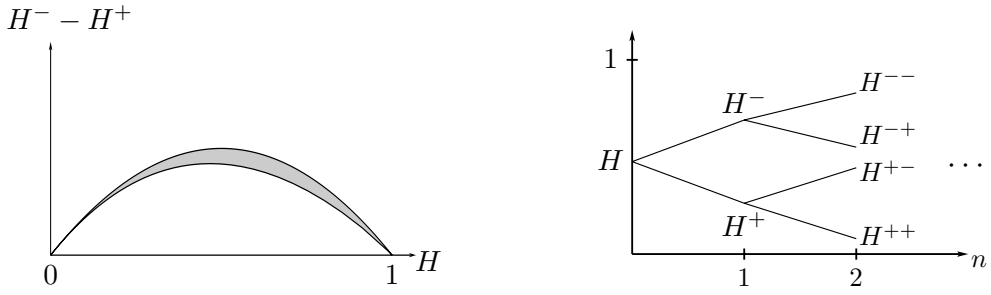

Figure 3.1: (left) In the binary case, allowed values of the difference $H^- - H^+$ versus H are inside the shaded region, and are away from zero except at $H = 0$ and $H = 1$. (right) The entropy paths created by the recursive construction keep bifurcating until they converge to zero or one.

As in the binary case, the memorylessness of the underlying processes will allow us to focus our attention on one-step transforms; once the properties of these are established, the large-blocklength behavior will readily follow. (We will have to partially forgo this convenience when we study polarization for processes with memory in Chapter 5.) We will first discuss processes with prime alphabet sizes. As we will see, such processes can be polarized by a simple extension of Arıkan's original method. We will then establish sufficient conditions for an Arıkan-like transform to polarize processes with arbitrary alphabets, and provide an example of a transform family that satisfies these conditions for all alphabet sizes. In all cases, the speed with which polarization takes place will be as in the binary case. We will leave out the translation of these results to low-complexity polar source and channel coding schemes, as we hope that these will be evident from the exposition in Chapter 2.

Suppose $(X_1, Y_1), (X_2, Y_2), \ldots$ is an i.i.d. process, where $X_1 \in \{0, \ldots, q-1\}$, and q is an arbitrary integer. As in the binary case, Y_1 takes values in a finite but arbitrary set $\mathcal{Y}$. We are interested in finding an invertible transform $G : X_1^2 \to U_1^2$ for which (3.2) holds for all joint distributions on (X_1, Y_1). Out of the many possibilities, perhaps the simplest guess is to use (3.1) by replacing the modulo-2 addition with a modulo-q addition. Before studying when this transform polarizes memoryless processes, it is useful to consider the following example, which shows when it does *not*:

Example 3.1. *Let X_1 be uniformly distributed over $\mathcal{X} = \{0, 1, 2, 3\}$ and let $Y_1 \in \{0, 1\}$ be such that $p_{Y|X}(0 \mid 0) = p_{Y|X}(0 \mid 2) = p_{Y|X}(1 \mid 1) = p_{Y|X}(1 \mid 3) = 1$. Then,*[1]

$$H(X_1 \mid Y_1) = 1/2.$$

[1] In this and the succeeding chapters, entropies will be computed with base-q logarithms, and therefore will be $[0, 1]$-valued. Also, addition of q-ary random variables will be modulo-q unless stated otherwise.

Also let $U_1 = X_1 + X_2$ *and* $X_2 = U_2$. *Then, the pairs* (X_1, Y_1), (U_1, Y_1^2), *and* $(U_2, Y_1^2 U_1)$ *are identically distributed (after appropriate grouping and labelling), and therefore*

$$H(U_2 \mid Y_1^2 U_1) = H(X_1 \mid Y_1) = H(U_1 \mid Y_1^2). \tag{3.3}$$

That is, the transformation has no effect on the resulting distributions. Clearly, this also implies that applying the same transform a second time (and further) will have no effect on the distributions or on the entropies.

At a first look, the anomaly in the above example may seem artificial: it is indeed easy to see that if we relabel the alphabet $\mathcal{X}$ by swapping 0 and 1, then the equalities in (3.3) become strict inequalities. Nevertheless, renaming the symbols alone may not be sufficient for polarization, as it may not guarantee that the resulting distributions will lead to a strict separation of entropies in the further steps of the construction.

The difficulty illustrated the above example is in fact common to all alphabets $\mathcal{X}$ of composite size. It is not peculiar to the particular transform in (3.1) either: Suppose that f is an operation for which the pair $(\mathcal{X}, f)$ is a group, and consider the mapping $(X_1, X_2) \to (U_1, U_2)$

$$U_1 = f(X_1, X_2), \qquad U_2 = X_2. \tag{3.4}$$

Then we have

Proposition 3.1. *If* $q = |\mathcal{X}|$ *is composite, then there exists an* $\epsilon > 0$ *and a distribution on* (X_1, Y_1) *for which* $H(X_1, Y_1) \in (\epsilon, 1 - \epsilon)$ *and*

$$H(U_2 \mid Y_1^2 U_1) = H(X_1 \mid Y_1) = H(U_1 \mid Y_1^2).$$

Proof. It is known [9, p. 28] that if q is composite, then the group $(\mathcal{X}, f)$ has a proper nontrivial subgroup. That is, there exists a set $S \subsetneq \mathcal{X}$ with $|S| > 1$ such that (S, f) is a group. Now let Y_1 be a constant random variable and X_1 be uniformly distributed over S. It is easy to verify that this choice of (X_1, Y_1) satisfies the claim. □

While the relations in (3.1) (and more generally (3.4)) fail to describe all one-to-one mappings on $\mathcal{X}^2$, we will focus our attention to transforms of this form. In view of Proposition 3.1, we will first restrict our attention to processes with prime $q = |\mathcal{X}|$. The reason for us to discuss the prime-q case before considering arbitrary alphabet sizes is twofold: First, we will see that proving polarization is relatively simple when the construction is based on (3.1). The observations we will make to this end will also be helpful in identifying the necessary properties of a transform to polarize processes over arbitrary alphabets. Second, constructions based on (3.1) are linear. As we will see in Chapter 4, generalizations of linear constructions are easy to analyze, and they can lead to higher rates of polarization.

3.1 Alphabets of Prime Size

Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be an i.i.d. process with prime $q = |\mathcal{X}|$. Define

$$U_1 = X_1 + X_2 \quad \text{and} \quad U_2 = X_2, \tag{3.5}$$

where the addition is modulo-q. Our first result states that the anomaly described in Example 3.1 and Proposition 3.1 vanish when q is prime.

Lemma 3.1. *For all $\delta > 0$, there exists $\epsilon(\delta) > 0$ such that if (X_1, Y_1) and (X_2, Y_2) are independent (but not necessarily identically distributed) pairs of random variables, then*

$$H(X_1 \mid Y_1), H(X_2 \mid Y_2) \in (\delta, 1 - \delta)$$

implies

$$H(X_1 + X_2 \mid Y_1^2) \geq \max\big\{H(X_1 \mid Y_1), H(X_2 \mid Y_2)\big\} + \epsilon(\delta),$$

provided that $q = |\mathcal{X}|$ is prime.

Before proving Lemma 3.1, let us describe the recursive construction and show that Lemma 3.1 implies polarization. These will be exactly as in the binary case: For $n = 0, 1, \ldots$, let $N = 2^n$ and define a sequence of transforms $G_n \colon \mathcal{X}^N \to \mathcal{X}^N$ recursively through

$$\begin{aligned} G_0(u) &= u \\ G_n(u) &= \pi_n\big(G_{n-1}(u_1) + G_{n-1}(u_2), G_{n-1}(u_2)\big) \quad n = 1, 2, \ldots \end{aligned}$$

where $u = (u_1, u_2)$ and $\pi_n \colon \{0, \ldots, q-1\}^N \to \{0, \ldots, q-1\}^N$ permutes the components of its argument vector through

$$\begin{aligned} \pi_n(u)_{2i-1} &= u_i \\ \pi_n(u)_{2i} &= u_{i+N/2} \end{aligned} \quad , \qquad i = 1, \ldots, N/2.$$

Now define

$$U_1^N = G_n(X_1^N).$$

As in the binary case, the transform G_n polarizes the underlying process.

Theorem 3.1. *For all $\epsilon > 0$,*

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{N} \left|\Big\{i \colon H(U_i \mid Y_1^N U_1^{i-1}) > 1 - \epsilon\Big\}\right| &= H(X_1 \mid Y_1), \\ \lim_{n\to\infty} \frac{1}{N} \left|\Big\{i \colon H(U_i \mid Y_1^N U_1^{i-1}) < \epsilon\Big\}\right| &= 1 - H(X_1 \mid Y_1). \end{aligned}$$

For the proof of the above theorem, we set the notation

$$H(X_1 \mid Y_1)^- := H(U_1 \mid Y_1^2), \qquad H(X_1 \mid Y_1)^+ := H(U_2 \mid Y_1^2 U_1),$$

similarly to the binary case. We also define a $\{-,+\}$-valued i.i.d. process $B_1, B_2, \ldots$ with $\Pr[B_1 = -] = 1/2$, and a $[0,1]$-valued process $H_0, H_1, \ldots$ through

$$\begin{aligned} H_0 &= H(X_1 \mid Y_1) \\ H_n &= H_{n-1}^{B_i}, \quad n = 1, 2, \ldots \end{aligned} \tag{3.6}$$

Proof. It follows from the equivalences in (2.10) that

$$\Pr[H_n \in \mathcal{I}] = \frac{1}{N} \left| \left\{ i \colon H(U_i \mid Y_1^N U_1^{i-1}) \in \mathcal{I} \right\} \right|$$

for all $\mathcal{I} \subseteq [0,1]$. It therefore suffices to show that for all $\epsilon > 0$

$$\begin{aligned} \lim_{n\to\infty} \Pr[H_n > 1 - \epsilon] &= H(X_1 \mid Y_1), \\ \lim_{n\to\infty} \Pr[H_n < \epsilon] &= 1 - H(X_1 \mid Y_1). \end{aligned}$$

We will show the stronger result that H_n converges almost surely (i.e., not only in probability) to a random variable H_∞ with $\Pr[H_\infty = 1] = 1 - \Pr[H_\infty = 0] = H(X_1 \mid Y_1)$. To that end, observe that $H_n^- + H_n^+ = 2H_n$, from which it follows that the process $H_0, H_1, \ldots$ is a bounded martingale and therefore converges almost surely to a random variable H_∞. As almost sure convergence implies convergence in $\mathcal{L}^1$, we have $E[|H_{n+1} - H_n|] = \frac{1}{2}E[H_n^- - H_n] + \frac{1}{2}E[H_n - H_n^+] = E[H_n^- - H_n] \to 0$. On the other hand, Lemma 3.1 implies that $H_n^- - H_n > \delta(\epsilon)$ if $H_n \in (\epsilon, 1-\epsilon)$, from which it follows that $H_n \to \{0,1\}$ with probability 1, i.e., that H_∞ is $\{0,1\}$-valued. The claim on the distribution of H_∞ follows from the relation $E[H_\infty] = E[H_0] = H(X_1 \mid Y_1)$. □

The first proof of polarization for the non-binary case consisted in showing that the source Bhattacharyya parameters (defined in the next section) polarize, and that this convergence implies the convergence of the entropies. This (somewhat convoluted) proof is included in Appendix 3.C for the interested reader. The present proof is direct and simple once Lemma 3.1 is obtained, as it is clearly a verbatim reproduction of Arıkan's original proof. Note, however, that Lemma 3.1 is weaker than Lemma 2.1, which identifies the distributions that are extremal in terms of how much they are polarized. Our preliminary studies suggest that such simple characterizations may not be possible in full generality in the q-ary case.

3.1.1 Proof of Lemma 3.1

We will first prove the unconditional version of Lemma 3.1, the proof for the conditional case will then follow easily. In particular, we will first show that if

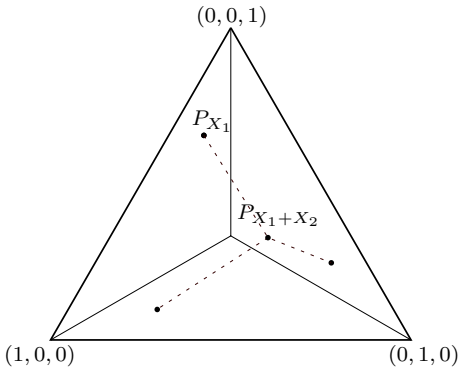

Figure 3.2: Cyclic convolution of two probability distributions over a ternary alphabet. The corners of the triangle represent the unit mass distributions and the center represents the uniform distribution.

X_1 and X_2 are independent random variables with moderate entropies, then the entropy of $X_1 + X_2$ is strictly larger than the entropy of either random variable (Lemma 3.4). To see why q has to be prime for this to hold, note that $p_{X_1+X_2}$ is obtained through a cyclic convolution, i.e., by taking a weighted sum of the cyclic shifts of p_{X_1}, where the weights are given by the coefficients of p_{X_2} (or vice versa, see Figure 3.2). These cyclic shifts are guaranteed to be away from each other only if q is prime and $H(X_1)$ is not too large, which in turn implies that $H(X_1 + X_2)$ is strictly larger than $H(X_1)$.

We now obtain a few simple lemmas in order to formalize these arguments. Some notation first: We let both $H(p)$ and $H(X)$ denote the entropy of a random variable $X \in \mathcal{X}$ with probability distribution p. We let p_i, $i \in \mathcal{X}$ denote the cyclic shifts of p, i.e.,

$$p_i(m) = p(m - i).$$

The cyclic convolution of probability distributions p and r will be denoted by $p * r$. That is,

$$p * r = \sum_{i \in \mathcal{X}} p(i) r_i = \sum_{i \in \mathcal{X}} r(i) p_i.$$

We also let $\mathrm{uni}(\mathcal{X})$ denote the uniform distribution over $\mathcal{X}$.

We first show that the $\mathcal{L}_1$ distance of a distribution from the uniform one is lower bounded by the corresponding Kullback–Leibler divergence. This result partially complements Pinsker's inequality.

Lemma 3.2. *Let p be a distribution over $\mathcal{X}$. Then,*

$$\|p - \mathrm{uni}(\mathcal{X})\|_1 \geq \frac{1}{q \log e} [1 - H(p)].$$

Proof.

$$\begin{aligned} 1 - H(p) &= \sum_{i \in \mathcal{X}} p(i) \log \frac{p(i)}{1/q} \\ &\leq \log e \sum_i p(i) \left[\frac{p(i) - 1/q}{1/q} \right] \\ &\leq q \log e \sum_i p(i) |p(i) - 1/q| \\ &\leq q \log e \| p - \mathrm{uni}(\mathcal{X}) \|_1, \end{aligned}$$

where we used the relation $\ln t \leq t - 1$ in the first inequality. □

Note that Lemma 3.2 holds for distributions over arbitrary finite sets. That $|\mathcal{X}|$ is a prime number has no bearing upon the above proof.

We next show that for prime q, if a distribution does not have too high an entropy, then its cyclic shifts will be away from each other:

Lemma 3.3. *Let p be a distribution over $\mathcal{X}$. Then,*

$$\| p_i - p_j \|_1 \geq \frac{1 - H(p)}{2q^2(q-1) \log e}.$$

for all $i, j \in \mathcal{X}$, $i \neq j$.

Proof. Given $i \neq j$, let $m = j - i$. We will show that there exists a $k \in \mathcal{X}$ satisfying

$$|p(k) - p(k+m)| \geq \frac{1 - H(p)}{2q^2(q-1) \log e},$$

which will yield the claim since $\| p_i - p_j \|_1 = \sum_{k \in \mathcal{X}} |p(k) - p(k+m)|$.

Suppose that $H(p) < 1$, as the claim is trivial otherwise. Let $p^{(\ell)}$ denote the ℓth largest element of p, and let $S = \{\ell : p^{(\ell)} \geq \frac{1}{q}\}$. Note that S is a proper subset of $\mathcal{X}$. We have

$$\begin{aligned} \sum_{\ell=1}^{|S|} [p^{(\ell)} - p^{(\ell+1)}] &= p^{(1)} - p^{(|S|+1)} \\ &\geq p^{(1)} - 1/q \\ &\geq \frac{1}{2(q-1)} \| p - \mathrm{uni}(\mathcal{X}) \|_1 \\ &\geq \frac{1 - H(p)}{2q(q-1) \log e}. \end{aligned}$$

In the above, the second inequality is obtained by observing that $p^{(1)} - 1/q$ is minimized when $p^{(1)} = \cdots = p^{(q-1)}$, and the third inequality follows from Lemma 3.2. Therefore, there exists at least one $\ell \in S$ such that

$$p^{(\ell)} - p^{(\ell+1)} \geq \frac{1 - H(p)}{2q^2(q-1) \log e}.$$

Given such an ℓ, let $A = \{1, \ldots, \ell\}$. Since q is prime, $\mathcal{X}$ can be written as

$$\mathcal{X} = \{k, k+m, k+m+m, \ldots, k\underbrace{+m+\ldots+m}_{q-1 \text{ times}}\}$$

for any $k \in \mathcal{X}$ and $m \in \mathcal{X}\backslash\{0\}$. Therefore, since A is a proper subset of $\mathcal{X}$, there exists a $k \in A$ such that $k + m \in A^c$, implying

$$p(k) - p(k+m) \geq \frac{1 - H(p)}{2q^2(q-1)\log e},$$

which yields the claim. □

We can now show that unless two independent random variables are both uniformly distributed or are both constants, their modulo-q addition strictly increases entropy:

Lemma 3.4. *Let $A, B \in \mathcal{X}$ be two independent random variables. For all $\delta > 0$, there exists $\epsilon_1(\delta) > 0$ such that*

$$\min\{H(A), 1 - H(B)\} \geq \delta$$

implies

$$H(A+B) \geq H(B) + \epsilon_1(\delta).$$

Proof. Let p and r denote the probability distributions of A and B, respectively, and let e_i denote the distribution with a unit mass on $i \in \mathcal{X}$. Since $H(p) \geq \delta > H(e_i) = 0$, it follows from the continuity of entropy that

$$\min_i \|p - e_i\|_1 \geq \mu(\delta) \tag{3.7}$$

for some $\mu(\delta) > 0$. On the other hand, since $H(r) \leq 1 - \delta$, we have by Lemma 3.3 that

$$\|r_i - r_j\|_1 \geq \frac{\delta}{2q^2(q-1)\log e} > 0 \tag{3.8}$$

for all pairs $i \neq j$. Relations (3.7), (3.8), and the strict concavity of entropy implies the existence of $\epsilon_1(\delta) > 0$ such that

$$\begin{aligned} H(p * r) &= H\left(\sum_i p(i) r_i\right) \\ &\geq \sum_i p(i) H(r_i) + \epsilon_1(\delta) \\ &= H(r) + \epsilon_1(\delta). \end{aligned}$$

□

Proof of Lemma 3.1. Let P_1 and P_2 be two random probability distributions on $\mathcal{X}$, with

$$P_1 = P_{X_1|Y_1}(\cdot \mid y_1) \text{ whenever } Y_1 = y_1,$$
$$P_2 = P_{X_2|Y_2}(\cdot \mid y_2) \text{ whenever } Y_2 = y_2.$$

It is then easy to see that

$$H(X_1 \mid Y_1) = \mathbb{E}[H(P_1)],$$
$$H(X_2 \mid Y_2) = \mathbb{E}[H(P_2)],$$
$$H(X_1 + X_2 \mid Y_1^2) = \mathbb{E}[H(P_1 * P_2)].$$

Suppose, without loss of generality, that $H(X_1 \mid Y_1) \leq H(X_2 \mid Y_2)$. We need to show that if $\mathbb{E}[H(P_1)], \mathbb{E}[H(P_2)] \in (\delta, 1-\delta)$ for some $\delta > 0$, then there exists an $\epsilon(\delta) > 0$ such that $\mathbb{E}[H(P_1 * P_2)] \geq \mathbb{E}[H(P_2)] + \epsilon(\delta)$. To that end, define the event

$$C = \{H(P_1) > \delta/2,\ H(P_2) < 1 - \delta/2\}.$$

Observe that

$$\begin{aligned}\delta &< \mathbb{E}[H(P_1)] \\ &\leq \big(1 - \Pr[H(P_1) > \delta/2]\big) \cdot \delta/2 + \Pr[H(P_1) > \delta/2],\end{aligned}$$

implying $\Pr[H(P_1) > \delta/2] > \frac{\delta}{2-\delta}$. It similarly follows that $\Pr[H(P_2) < 1 - \delta/2] > \frac{\delta}{2-\delta}$. Note further that since Y_1 and Y_2 are independent, so are $H(P_1)$ and $H(P_2)$. Thus, the event C has probability at least $\frac{\delta^2}{(2-\delta)^2} =: \epsilon_2(\delta)$. On the other hand, Lemma 3.4 implies that conditioned on C we have

$$H(P_1 * P_2) \geq H(P_2) + \epsilon_1(\delta/2) \tag{3.9}$$

for some $\epsilon_1(\delta/2) > 0$. Thus,

$$\begin{aligned}\mathbb{E}[H(P_1 * P_2)] &= \Pr[C] \cdot \mathbb{E}[H(P_1 * P_2) \mid C] + \Pr[C^c] \cdot \mathbb{E}[H(P_1 * P_2) \mid C^c] \\ &\geq \Pr[C] \cdot \mathbb{E}[H(P_2) + \epsilon_1(\delta/2) \mid C] \\ &\qquad + \Pr[C^c] \cdot \mathbb{E}[H(P_2) \mid C^c] \\ &\geq \mathbb{E}[H(P_2)] + \epsilon_1(\delta/2)\epsilon_2(\delta),\end{aligned}$$

where in the first inequality we used (3.9) and the relation $H(p * r) \geq H(p)$. Setting $\epsilon(\delta) := \epsilon_1(\delta/2)\epsilon_2(\delta)$ yields the result. □

3.1.2 Rate of Polarization

We have seen that a similar construction to Arıkan's polarizes q-ary memoryless processes for prime q. We will now show that polarization takes place

sufficiently fast—in fact as fast as in the binary case—so that source and channel codes based on such constructions have small error probability. We will do so following the approach in the binary case. For this purpose, we first need to define a reliability parameter, analogously to the Bhattacharyya parameter defined in Chapter 2, whose behavior through the polarization process is easy to track. For the q-ary case, a convenient choice turns out to be

$$Z(X \mid Y) := \frac{1}{q-1} \sum_{\substack{x,x' \in \mathcal{X}: \\ x \neq x'}} \sum_{y} \sqrt{p_{XY}(x,y) p_{XY}(x',y)}.$$

It is easy to see that this parameter takes values in $[0,1]$. As a measure of reliability, it is natural to expect that $Z(X \mid Y)$ upper bound the average error probability of the optimal decoder, and that

$$\begin{aligned} Z(X \mid Y) \approx 1 &\iff H(X \mid Y) \approx 1, \\ Z(X \mid Y) \approx 0 &\iff H(X \mid Y) \approx 0. \end{aligned}$$

The following propositions show that these requirements are indeed met:

Proposition 3.2. $P_e(X \mid Y) \leq (q-1) Z(X \mid Y)$.

Proof. Let $P_{e,x}$ denote the error probability of the optimal decision rule conditioned on $X = x$. We have

$$\begin{aligned} P_{e,x} &\leq \sum_{y} p(y \mid x) \mathbb{1}_{[x' \colon p_{X|Y}(x'|y) \geq p_{X|Y}(x|y)]} \\ &\leq \sum_{y} p(y \mid x) \sum_{x' \colon x' \neq x} \mathbb{1}_{[p_{X|Y}(x'|y) \geq p_{X|Y}(x|y)]} \\ &\leq \sum_{x' \colon x' \neq x} \sum_{y} \frac{p_{X|Y}(x \mid y) p(y)}{p(x)} \sqrt{\frac{p_{X|Y}(x' \mid y)}{p_{X|Y}(x \mid y)}} \\ &= \sum_{x' \colon x' \neq x} \sum_{y} \frac{1}{p(x)} \sqrt{p_{XY}(x',y) p_{XY}(x,y)}. \end{aligned}$$

Averaging the above relation over x yields the claim. □

Proposition 3.3.

$$Z(X \mid Y)^2 \leq H(X \mid Y) \tag{3.10}$$

$$H(X \mid Y) \leq \log(1 + (q-1) Z(X \mid Y)). \tag{3.11}$$

Proof. See Appendix 3.A. □

Since the polarization construction is recursive as in the binary case, the limiting behavior of the Z parameters along the polarization process is determined by their one-step behavior. In particular, the following bounds will suffice to conclude that polarization takes place fast:

Lemma 3.5. *Let $f\colon \mathcal{X}^2 \to \mathcal{X}$ be such that both functions $f(x_1,\cdot)\colon \mathcal{X} \to \mathcal{X}$ and $f(\cdot,x_2)\colon \mathcal{X} \to \mathcal{X}$ are invertible for all x_1 and x_2, respectively. Defining $V_1 := f(X_1,X_2)$ and $V_2 := X_2$ we have*

$$Z(V_1 \mid Y_1^2) \le (q^2-q+1)Z(X_1 \mid Y_1) \tag{3.12}$$
$$Z(V_2 \mid Y_1^2V_1) \le (q-1)Z(X_1 \mid Y_1)^2. \tag{3.13}$$

Clearly, bounds that are relevant to the present case are obtained by taking f to be the modulo-q addition. The reason for us to state these bounds in a slightly more general setting will be evident when we consider polarization for arbitrary alphabet sizes in the next section.

Proof. The assumptions on the function f imply that there exist q permutations $\pi_i\colon \mathcal{X} \to \mathcal{X}$, $i = 0,\dots,q-1$ with

$$\pi_i(x) \neq \pi_j(x) \text{ for all } i \neq j, x \in \mathcal{X}$$

such that $\pi_i(j) = f(j,i)$. We therefore have

$$p(v_1,v_2,y_1,y_2) = p_{XY}(\pi_{v_2}^{-1}(v_1),y_1)p_{XY}(v_2,y_2).$$

To obtain the first claim, we write

$$\begin{aligned}
Z(V_1 \mid Y_1^2) &= \frac{1}{q-1}\sum_{\substack{v_1,v_1':\\ v_1\neq v_1'}}\sum_{y_1^2}\left[p(v_1,y_1,y_2)p(v_1',y_1,y_2)\right]^{1/2} \\
&= \frac{1}{q-1}\sum_{\substack{v_1,v_1':\\ v_1\neq v_1'}}\sum_{y_1^2}\left[\sum_{v_2}p(v_1,v_2,y_1,y_2)\sum_{v_2'}p(v_1',v_2',y_1,y_2)\right]^{1/2} \\
&\le \frac{1}{q-1}\sum_{\substack{v_1,v_1':\\ v_1\neq v_1'}}\sum_{y_1^2}\sum_{v_2,v_2'}\left[p(v_1,v_2,y_1,y_2)p(v_1',v_2',y_1,y_2)\right]^{1/2} \\
&= \frac{1}{q-1}\sum_{v_2,v_2'}\sum_{y_2}\left[p_{XY}(v_2,y_2)p_{XY}(v_2',y_2)\right]^{1/2} \\
&\qquad\cdot\sum_{\substack{v_1,v_1':\\ v_1\neq v_1'}}\sum_{y_1}\left[p_{XY}(\pi_{v_2}^{-1}(v_1),y_1)p_{XY}(\pi_{v_2'}^{-1}(v_1'),y_1)\right]^{1/2}.
\end{aligned}$$

Splitting the summation over (v_2,v_2') into two parts $v_2 = v_2'$ and $v_2 \neq v_2'$, and considering the first part we have

$$\begin{aligned}
&\sum_{v_2=v_2'}\sum_{y_2}\left[p_{XY}(v_2,y_2)p_{XY}(v_2',y_2)\right]^{1/2} \\
&\qquad\cdot\frac{1}{q-1}\sum_{\substack{v_1,v_1':\\ v_1\neq v_1'}}\sum_{y_1}\left[p_{XY}(\pi_{v_2}^{-1}(v_1),y_1)p_{XY}(\pi_{v_2'}^{-1}(v_1'),y_1)\right]^{1/2}.
\end{aligned}$$

The sums on the second line above are equivalent to $Z(X_1 \mid Y_1)$ for all v_2 and y_2, and those on the first line add to 1. Therefore the above term is equal to $Z(X_1 \mid Y_1)$. On the other hand, when $v_2 \neq v_2'$ we have

$$\frac{1}{q-1} \sum_{\substack{v_2, v_2' : \\ v_2 \neq v_2'}} \sum_{y_2} \big[p_{XY}(v_2, y_2) p_{XY}(v_2', y_2) \big]^{1/2} \cdot \sum_{\substack{v_1, v_1' : \\ v_1 \neq v_1'}} \sum_{y_1} \big[p_{XY}(\pi_{v_2}^{-1}(v_1), y_1) p_{XY}(\pi_{v_2'}^{-1}(v_1'), y_1) \big]^{1/2}.$$

Here, the summation over y_1 is upper bounded by 1, and the upper sums are equal to $Z(X_1 \mid Y_1)$. Therefore the above term is upper bounded by $q(q-1)Z(X_1 \mid Y_1)$. Combining this with the first part yields (3.12). To obtain (3.13), we write

$$\begin{aligned} Z(V_2 \mid Y_1^2 V_1) &= \frac{1}{q-1} \sum_{\substack{v_2, v_2' : \\ v_2 \neq v_2'}} \sum_{y_1^2, v_1} \big[p_{XY}(\pi_{v_2}^{-1}(v_1), y_1) p_{XY}(v_2, y_2) \\ &\qquad \cdot p_{XY}(\pi_{v_2'}^{-1}(v_1), y_1) p_{XY}(v_2', y_2) \big]^{1/2} \\ &= \frac{1}{q-1} \sum_{\substack{v_2, v_2' : \\ v_2 \neq v_2'}} \sum_{y_2} \big[p_{XY}(v_2, y_2) p_{XY}(v_2', y_2) \big]^{1/2} \\ &\qquad \cdot \sum_{v_1} \sum_{y_1} \big[p_{XY}(\pi_{v_2}^{-1}(v_1), y_1) p_{XY}(\pi_{v_2'}^{-1}(v_1), y_1) \big]^{1/2}. \end{aligned}$$

For all $v_2 \neq v_2'$ and y_2, the lower sums on the second line are upper bounded by $(q-1)Z(X_1 \mid Y_1)$, and those on the first are equivalent to $Z(X_1 \mid Y_1)$. This yields the second claim. □

We are now ready to state and prove the main result on the rate of polarization:

Theorem 3.2. *For all $0 < \beta < 1/2$,*

$$\lim_{n \to \infty} \frac{1}{N} \Big| \Big\{ i \colon Z(U_i \mid Y_1^N U_1^{i-1}) \leq 2^{-N^\beta} \Big\} \Big| = 1 - H(X_1 \mid Y_1).$$

Proof. The proof is identical to that of Theorem 2.2: Set the shorthand notation

$$Z(X_1 \mid Y_1)^- := Z(U_1 \mid Y_1^2), \qquad Z(X_1 \mid Y_1)^+ := Z(U_2 \mid Y_1^2 U_1).$$

Define a $\{-,+\}$-valued i.i.d. process $B_1, B_2, \ldots$ with $\Pr[B_1 = -] = 1/2$ and a $[0,1]$-valued process $Z_0, Z_1, \ldots$ with

$$\begin{aligned} Z_0 &= Z(X_1 \mid Y_1) \\ Z_n &= Z_{n-1}^{B_n}, \qquad n = 1, 2, \ldots \end{aligned} \tag{3.14}$$

Then, the equivalences in (2.14) imply that

$$\Pr[Z_n \in \mathcal{I}] = \frac{1}{N}\Big|\Big\{i\colon Z(U_i \mid Y_1^N U_1^{i-1}) \in \mathcal{I}\Big\}\Big|$$

for all $\mathcal{I} \subseteq [0,1]$. Further, recall that the process $H_0, H_1, \ldots$ defined in (3.6) converges almost surely to the set $\{0,1\}$ (see proof of Theorem 3.1). It then follows from Proposition 3.3 that the process $Z_0, Z_1, \ldots$ also converges almost surely to the set $\{0,1\}$ with $\Pr[\lim_{n\to\infty} Z_n = 0] = 1 - H(X_1 \mid Y_1)$. The claim then follows from Lemma 2.3 by taking $\mathcal{I} = [0, 2^{-N^\beta}]$. □

3.2 Arbitrary Finite Alphabets

We have seen in the previous section that the mapping $(X_1, X_2) \to (X_1 + X_2, X_2)$ fails to polarize certain processes whenever $q = |\mathcal{X}|$ is a composite number (Example 3.1). We have also seen that the difficulty with such alphabets persists so long as '+' is replaced by any group operation over $\mathcal{X}$ (Proposition 3.1). We are now interested in finding transforms $(X_1, X_2) \to (U_1, U_2)$ that will polarize all i.i.d. processes over all finite alphabets. We will in particular study mappings of the form

$$\begin{aligned} U_1 &= f(X_1, X_2) \\ U_2 &= X_2, \end{aligned} \tag{3.15}$$

for some $f\colon \mathcal{X}^2 \to \mathcal{X}$. While not all one-to-one mappings $(X_1, X_2) \to (U_1, U_2)$ can be reduced to this form, we restrict our attention to these due to their relative simplicity.

Once we find an appropriate transform f, we will use it recursively as in the binary case. That is, we will define for all $n = 0, 1, \ldots$ and $N = 2^n$ a sequence of transforms $G_n\colon \{0, \ldots, q-1\}^N \to \{0, \ldots, q-1\}^N$ through

$$\begin{aligned} G_0(u) &= u \\ G_n(u) &= \pi_n\Big(f\big(G_{n-1}(u_1), G_{n-1}(u_2)\big), G_{n-1}(u_2)\Big) \quad n = 1, 2, \ldots \end{aligned} \tag{3.16}$$

where $u = (u_1, u_2)$, the action of f on its arguments is componentwise as in (3.15), and the permutation π_n is as in the previous sections. Let us now introduce the notion of a *polarizing* mapping:

Definition 3.1. *We call a mapping* $f\colon \mathcal{X}^2 \to \mathcal{X}$ polarizing *if*

(p.i) for all $x_2 \in \mathcal{X}$*, the mapping* $x_1 \to f(x_1, x_2)$ *is invertible,*

(p.ii) for all $x_1 \in \mathcal{X}$*, the mapping* $x_2 \to f(x_1, x_2)$ *is invertible,*[2] *and*

[2] In group theory, a pair $(\mathcal{X}, f)$ with f satisfying (p.i) and (p.ii) is known as a *quasigroup*.

(p.iii) for all $2 \le K \le q-1$ *and distinct* $a_0, \dots, a_{K-1} \in \mathcal{X}$*, the matrix*

$$B_{ij} = f(a_i, a_j), \quad i, j = 0, \dots, K-1$$

has at least $K+1$ *distinct entries.*

Example 3.2. *Consider a matrix* F *with* $F_{ij} = f(i,j), i, j = 0, \dots q-1$*. (That is,* F *is the* Cayley table *of* f*.) Then it is easy to see that, of the operations corresponding to*

$$F = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{bmatrix}, \qquad G = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 2 & 3 & 0 \\ 2 & 3 & 0 & 1 \\ 3 & 0 & 1 & 2 \end{bmatrix},$$

F is polarizing, whereas G is not, since $G_{00} = G_{22} = 0$ and $G_{02} = G_{20} = 2$, violating (p.iii). Note that F and G correspond to modulo-3 and modulo-4 addition, respectively (see also Example 3.1).

In the rest of this section, we will give meaning to Definition 3.1 by showing that the construction in (3.16) leads to polarization if f is a polarizing mapping: (p.i) guarantees that the one-step transform in (3.15) is one-to-one, and (p.iii) guarantees that anomalous distributions such as the one in Example 3.1 are also polarized; it turns out that this is indeed the only type of irregularity that needs handling. Condition (p.ii) is in fact not necessary for polarization to take place, and can be relaxed. We include it Definition 3.1 only because it helps simplify the proofs. This condition is also not a very restrictive one; there are several simple families of mappings that satisfy (p.i)–(p.iii) for all alphabet sizes. We give one example here:

Example 3.3. *The mapping* $f(x_1, x_2) = x_1 + \pi(x_2)$*, where* $\pi \colon \mathcal{X} \to \mathcal{X}$ *is the permutation*

$$\pi(x) = \begin{cases} \lfloor q/2 \rfloor & \text{if } x = 0 \\ x - 1 & \text{if } 1 \le x \le \lfloor q/2 \rfloor \\ x & \text{otherwise} \end{cases}$$

is polarizing for all $q = |\mathcal{X}|$*. A proof of this is given in Appendix 3.B. The Cayley table of* f *is given below for* $q = 6$*.*

$$\begin{bmatrix} 3 & 0 & 1 & 2 & 4 & 5 \\ 4 & 1 & 2 & 3 & 5 & 0 \\ 5 & 2 & 3 & 4 & 0 & 1 \\ 0 & 3 & 4 & 5 & 1 & 2 \\ 1 & 4 & 5 & 0 & 2 & 3 \\ 2 & 5 & 0 & 1 & 3 & 4 \end{bmatrix}$$

Before proceeding to the proof of polarization, let us introduce a definition in order to capture the anomaly described in Example 3.1: Given a distribution p over $\mathcal{X}$, let a_i, $i = 0, \ldots, q-1$ be any labelling of the elements of $\mathcal{X}$ for which $p(a_0) \geq p(a_1) \geq \ldots \geq p(a_{q-1})$. For all $\nu > 0$, let

$$K_\nu := \min\{i \leq q-2 \colon a_i - a_{i+1} > \nu\} \cup \{q-1\}$$

and define

$$M_{p,\nu} := \{a_0, \ldots, a_{K_\nu}\}.$$

The general form of the anomaly described in Proposition 3.1 can be stated as $M_{p_{X_1},\nu} = M_{p_{X_2},\nu}$ for random variables X_1 and X_2. The next lemma shows that a polarizing mapping will strictly increase entropy even under such irregularities:

Lemma 3.6. *For all $\epsilon, \nu > 0$, there exists $\delta(\epsilon, \nu) > 0$ such that if $X_1, X_2 \in \mathcal{X}$ are independent random variables with $H(X_1), H(X_2) \in (\epsilon, 1-\epsilon)$ and $M_{p_{X_1},\nu} = M_{p_{X_2},\nu} = M$ for some M with $1 \leq |M| \leq q-1$, and if f is a polarizing mapping, then*

$$H(f(X_1, X_2)) \geq H(X_i) + \delta(\epsilon, \nu), \qquad i = 1, 2.$$

Proof. We will prove the claim for $i = 2$, the proof for $i = 1$ follows similarly by the symmetry in the assumptions. It follows from (p.ii) that there exist q distinct permutations $\pi_i \colon \mathcal{X} \to \mathcal{X}$, $i = 0, \ldots, q-1$ such that $f(j, i) = \pi_i(j)$. Observe also that (p.i) implies

$$\pi_i(x) \neq \pi_j(x) \text{ for all } i \neq j, x \in \mathcal{X}. \tag{3.17}$$

Defining probability distributions r_i through $r_i(u) = p_{X_2}(\pi_i^{-1}(u))$, we have

$$p_{f(X_1,X_2)} = \sum_{i=0}^{q-1} p_{X_1}(i) r_i. \tag{3.18}$$

It suffices to show that there exist $a, b \in \mathcal{X}$ for which

(i) $p_{X_1}(a), p_{X_1}(b) \geq \eta(\epsilon, \nu)$ for some $\eta(\epsilon, \nu) > 0$, and

(ii) $\|r_a - r_b\|_1 \geq \nu$,

since the claim will then follow immediately from (3.18), the strict concavity of entropy, and that $H(r_i) = H(X_2)$ for all i.

First consider the case $M = \{a\}$ for some $a \in \mathcal{X}$, and observe that $H(X_1) > \epsilon$ implies $p_{X_1}(a) \geq p_{X_1}(b) \geq \eta(\epsilon)$ for some $b \neq a$ and $\eta(\epsilon) > 0$, satisfying (i). It also follows from (3.17) that $r_a(\pi_a(a)) - r_b(\pi_a(a)) = p_{X_1}(a) - p_{X_1}(c)$ for some $c \neq a$, implying (ii) since the latter difference is at least ν, and therefore yielding the claim.

Suppose now that $2 \leq |M| \leq q-1$. Define, for all $x \in \mathcal{X}$ and $T \subset \mathcal{X}$, the sets

$$S_{x,T} = \{i \colon \pi_x^{-1}(i) \in T\},$$

and observe that (p.iii) implies that

$$\forall T \subset \mathcal{X},\ 2 \le |T| \le q-1,\ \exists a, b \in T \text{ such that } S_{a,T} \neq S_{b,T}. \tag{3.19}$$

Now let $a, b \in M$ be such that $S_{a,M} \neq S_{b,M}$. It then follows from the definition of M that there exists $x \in \mathcal{X}$ for which $|r_a(x) - r_b(x)| \ge \nu$, satisfying (ii). That (i) is also satisfied can be seen by noting that $|M| \le q-1$ and $a, b \in M$ imply $p_{X_2}(a), p_{X_2}(b) \ge \nu$. This concludes the proof. □

We are now ready to prove the main result of this section, which will lead to a polarization theorem for arbitrary discrete alphabets.

Theorem 3.3. *For all $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that if $(X_1, Y_1), (X_2, Y_2)$ are i.i.d. random variable pairs with $H(X_1 \mid Y_1) \in (\epsilon, 1-\epsilon)$, and if $f\colon \mathcal{X}^2 \to \mathcal{X}$ is a polarizing mapping, then*

$$H(f(X_1, X_2) \mid Y_1^2) \ge H(X_1 \mid Y_1) + \delta(\epsilon).$$

Proof. Let H_1, H_2 and H_u be $[0,1]$-valued random variables with

$$\begin{aligned} H_1 &= H(X_1 \mid Y_1 = y_1) \\ H_2 &= H(X_2 \mid Y_2 = y_2) \\ H_u &= H(f(X_1, X_2) \mid Y_1 = y_1, Y_2 = y_2) \end{aligned}$$

whenever $(Y_1, Y_2) = (y_1, y_2)$. Clearly, H_1 and H_2 are i.i.d. with

$$E[H_1] = E[H_2] = H(X_1 \mid Y_1).$$

Suppose first that $\Pr[H_1 \le \epsilon/2], \Pr[H_1 \ge 1 - \epsilon/2] \ge \epsilon/2(2-\epsilon)$. Then, the event

$$A = \big\{y_1, y_2\colon H_1 \le \epsilon/2, H_2 \ge 1 - \epsilon/2\big\}$$

has probability at least $\big[\epsilon/2(2-\epsilon)\big]^2$. Further, as both functions $x_1 \to f(x_1, x_2)$ and $x_2 \to f(x_1, x_2)$ are invertible for all x_2 and x_1 respectively, we have $H_u \ge H_1, H_2$ for all $(Y_1, Y_2) = (y_1, y_2)$. Thus,

$$\begin{aligned} H(f(X_1, X_2) \mid Y_1 Y_2) &= E[H_u] \\ &= \Pr[A] \cdot E[H_u \mid A] + \Pr[A^c] \cdot E[H_u \mid A^c] \\ &\ge \Pr[A] \cdot E[H_2 \mid A] + \Pr[A^c] \cdot E[H_1 \mid A^c] \\ &\ge \Pr[A] \cdot E[H_1 + 1 - \epsilon \mid A] + \Pr[A^c] \cdot E[H_1 \mid A^c] \\ &\ge E[H_1] + \big[\tfrac{\epsilon}{2(2-\epsilon)}\big]^2 (1-\epsilon) \\ &= H(X_1 \mid Y_1) + \big[\tfrac{\epsilon}{2(2-\epsilon)}\big]^2 (1-\epsilon), \end{aligned}$$

yielding the claim.

Now suppose instead that $\Pr[H_1 \le \epsilon/2] < \frac{\epsilon}{2(2-\epsilon)}$. Then, since

$$\Pr[H_1 \ge 1 - \epsilon/2] \le \frac{E[H_1]}{1 - \epsilon/2} \le \frac{2 - 2\epsilon}{2 - \epsilon},$$

it follows that

$$\Pr[H_1 \in (\epsilon/2, 1-\epsilon/2)] \geq \frac{\epsilon}{2(2-\epsilon)}. \tag{3.20}$$

A similar argument shows that the above inequality also holds when $\Pr[H_1 \geq 1-\epsilon/2] < \frac{\epsilon}{2(2-\epsilon)}$. We will now show that the conditions of Lemma 3.6 hold with positive probability whenever we have (3.20). For that purpose, note that it follows from Lemma 3.2 that for all $\epsilon > 0$, there exists $\nu(\epsilon) > 0$ for which $H(V) \leq 1-\epsilon/2$ implies $|M_{p_V,\nu}| \leq q-1$. Given such a ν, let $S_1 \subset \mathcal{X}$ and $S_2 \subset \mathcal{X}$ be random sets with

$$\begin{aligned} S_1 &= M_{p_{X_1|Y_1=y_1},\nu} \qquad \text{whenever } Y_1 = y_1 \\ S_2 &= M_{p_{X_2|Y_2=y_2},\nu} \qquad \text{whenever } Y_2 = y_2. \end{aligned}$$

As S_1 and S_2 are independent and identically distributed, it follows from (3.20) and the above argument that there exists $S \subset \mathcal{X}$ with $1 \leq |S| \leq q-1$ such that the event

$$B = \{y_1, y_2 \colon S_1 = S_2 = S\}$$

has probability at least $[\epsilon/2^q(2-\epsilon)]^2$. It then follows from Lemma 3.6 that $H_u \geq H_1 + \delta(\epsilon, \nu(\epsilon))$ for some $\delta(\epsilon, \nu(\epsilon)) > 0$ whenever $y_1, y_2 \in B$. Therefore

$$\begin{aligned} E[H_u] &= \Pr[B] \cdot E[H_u \mid B] + \Pr[B^c] \cdot E[H_u \mid B^c] \\ &\geq \Pr[B] \cdot E[H_1 + \delta(\epsilon, \nu(\epsilon)) \mid B] + \Pr[B^c] \cdot E[H_1 \mid B^c] \\ &= E[H_1] + \left[\epsilon/2^q(2-\epsilon)\right]^2 \cdot \delta(\epsilon, \nu(\epsilon)), \end{aligned}$$

completing the proof. □

We can now state the polarization theorem for arbitrary finite alphabets. Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be a discrete, i.i.d. process with $|\mathcal{X}| < \infty$. Also let f be a polarizing mapping, and define

$$U_1^N = G_n(X_1^N),$$

where G_n is as in (3.16). We have

Theorem 3.4. *For all $\epsilon > 0$,*

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{N} \left|\left\{i \colon H(U_i \mid Y_1^N U_1^{i-1}) > 1-\epsilon\right\}\right| &= H(X_1 \mid Y_1), \\ \lim_{n\to\infty} \frac{1}{N} \left|\left\{i \colon H(U_i \mid Y_1^N U_1^{i-1}) < \epsilon\right\}\right| &= 1 - H(X_1 \mid Y_1). \end{aligned}$$

Proof. The proof follows from Theorem 3.3, and is identical to those of Theorems 2.1 and 3.1. □

The rate of polarization for the construction in (3.16) is also as in the binary case:

Theorem 3.5. *For all* $0 < \beta < 1/2$*,*

$$\lim_{n\to\infty} \left|\left\{i\colon Z(U_i \mid Y_1^N U_1^{i-1}) \leq 2^{-N^\beta}\right\}\right| = 1 - H(X_1 \mid Y_1).$$

Proof. The proof follows from Lemma 3.5 and is identical to that of Theorem 3.2. □

3.3 How to Achieve Capacity

Polarization results in this chapter immediately yield polar source coding methods that compress any discrete memoryless source to its entropy. Recall from the discussion in Section 2.4, however, that translating polarization results to channel coding schemes becomes trivial only for uniformly distributed channel inputs. Clearly, this statement is equally valid for channels with non-binary input alphabets. Therefore one can achieve the *symmetric capacity* of discrete memoryless channels with the methods discussed so far, as opposed to the true capacity. In channels where the gap between these two rates is significant, one can use the following generic method, discussed in [10, p. 208], to approach the true capacity: Given a channel $W\colon \mathcal{X} \to \mathcal{Y}$, one can construct a new channel $W'\colon \mathcal{X}' \to \mathcal{Y}$ with $|\mathcal{X}'| \geq |\mathcal{X}|$, where $W'(y \mid x') = W(y \mid f(x'))$ and $f\colon \mathcal{X}' \to \mathcal{X}$ is a deterministic map. Note that the mutual informations $I(X;Y)$ and $I(X';Y)$ developed across W and W' respectively are identical for any distribution on input X' to W' and the induced distribution on X. Observe further that if X' is uniformly distributed, then one can induce, using an appropriate mapping f, any distribution p_X on X with $p_X(x) = k_x/|\mathcal{X}'|$, where k_x's are integer-valued. Consequently, one can approach the true capacity of any discrete memoryless channel W by choosing f so as to approximate the capacity-achieving input distribution of this channel, and using a symmetric capacity-achieving polar code for the created channel W'.

3.4 Complexity

Non-binary codes based on the polarization transforms discussed in this chapter will have low-complexities like their binary counterparts. In particular, if one assumes that the computation of a one-step polarizing mapping takes one unit of time, then the time and space complexity of encoding these codes will be $O(N \log N)$ in the blocklength. Similarly, it readily follows from the results in [3] that successive cancellation decoding with such codes can be performed with $O(q^2 N \log N)$ time and $O(qN \log N)$ space complexities. Also by a straightforward extension of the algorithm proposed in [5], these codes can be constructed with $O(q^2 N)$ time and $O(q \log N)$ space complexities.

In the next chapter, we will continue studying the universality of polarization. In particular, we will show that memoryless processes can be polarized by generalizations of Arıkan's construction. As we will see, such generalizations

can produce substantial gains in error probability without too much added complexity.

3.A Proof of Proposition 3.3

Proof of (3.10)*:* The proof of this inequality was given in [7] for the binary case; the proof of the q-ary version is identical. We nevertheless include it here for completeness.

The Rényi entropy of order α of a random variable X is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_x p(x)^\alpha$$

for all $\alpha > 0, \alpha \neq 1$. (The logarithm is taken to the base q.) It is known that $H_\alpha(X)$ is decreasing in α and that $\lim_{\alpha \to 1} H_\alpha(X) = H(X)$. We thus have

$$\begin{aligned} H(X \mid Y = y) \leq H_{1/2}(X \mid Y = y) &= \log \Big[\sum_x \sqrt{p(x \mid y)} \Big]^2 \\ &= \log \big[1 + (q-1) Z(X \mid Y = y) \big], \end{aligned}$$

where we define $Z(X \mid Y = y) = \frac{1}{q-1} \sum_{x \neq x'} \sqrt{p(x \mid y) p(x' \mid y)}$. The desired inequality is obtained by averaging the above relation over y and using the concavity of $t \to \log(1 + (q-1)t)$.

Proof of (3.11)*:* We define two new random variables S and T with $p(x, y, s, t) = p(x) p(y \mid x) p(s, t \mid x)$, where

$$p(s, t \mid x) = \begin{cases} \frac{1}{2(q-1)} & \text{if } s = x, t \neq x \\ \frac{1}{2(q-1)} & \text{if } s \neq x, t = x \\ 0 & \text{otherwise} \end{cases} .$$

Note that the conditional probability $p(x, y \mid s, t)$ is defined only if $s \neq t$ and is non-zero only if $x = s$ or $x = t$. Therefore, if we define for $s \neq t$

$$Z_{s,t}(X \mid Y) = \sum_y \sqrt{p_{XY|ST}(s, y \mid s, t) p_{XY|ST}(t, y \mid s, t)},$$

we have from Proposition 2.3 that

$$H(X \mid Y, S = s, T = t) \geq [2 Z_{s,t}(X \mid Y)]^2.$$

The proof then follows from the relations

$$\begin{aligned}
H(X \mid Y) &\geq H(X \mid YST) \\
&\geq \sum_{\substack{s,t:\\ s\neq t}} p(s,t)[2Z_{s,t}(X \mid Y)]^2 \\
&= \sum_{\substack{s,t:\\ s\neq t}} p(s,t)\Bigg[2\sum_y \Big(\frac{p_{XY}(s,y)p_{ST|X}(s,t \mid s)}{p_{ST}(s,t)}\Big)^{1/2} \\
&\qquad\qquad \cdot \Big(\frac{p_{XY}(t,y)p_{ST|X}(s,t \mid t)}{p_{ST}(s,t)}\Big)^{1/2}\Bigg]^2 \\
&\geq \Bigg[\sum_{\substack{s,t:\\ s\neq t}} p(s,t)2\sum_y \Big(\frac{p_{XY}(s,y)p_{ST|X}(s,t \mid s)}{p_{ST}(s,t)}\Big)^{1/2} \\
&\qquad\qquad \cdot \Big(\frac{p_{XY}(t,y)p_{ST|X}(s,t \mid t)}{p_{ST}(s,t)}\Big)^{1/2}\Bigg]^2 \\
&= \Bigg[\sum_{\substack{s,t:\\ s\neq t}} \sum_y \frac{1}{q-1}\big[p_{XY}(s,y)p_{XY}(t,y)\big]^{1/2}\Bigg]^2 \\
&= Z(X \mid Y)^2.
\end{aligned}$$

In the above, the second inequality follows from the convexity of the function $x \to x^2$.

3.B A Family of Polarizing Transforms

Here we show that for all $q = |\mathcal{X}|$, the function $f\colon \mathcal{X}^2 \to \mathcal{X}$, $f(x_1, x_2) \to x_1 + \pi(x_2)$ with

$$\pi(x) = \begin{cases} \lfloor q/2 \rfloor & \text{if } x = 0 \\ x - 1 & \text{if } 1 \leq x \leq \lfloor q/2 \rfloor \\ x & \text{otherwise} \end{cases}$$

is polarizing (see Definition 3.1). That (p.i) and (p.ii) are satisfied readily follows from π being a permutation. It remains to show (p.iii), i.e., that for all $2 \leq K \leq q-1$ and $a_0 < a_1 < \ldots < a_{K-1}$ in $\mathcal{X}$, the matrix

$$B_{ij} = a_i + \pi(a_j), \quad i, j = 0, \ldots, K-1$$

has at least $K+1$ distinct entries. We will consider two cases:

$\underline{K \geq 3}$: We will show, by contradiction, that the sets $\{B_{i1}\}$ and $\{B_{i(K-1)}\}$ are not identical, which leads to the claim. For this purpose, note first that $1 \leq a_1 < a_{K-1}$. Also, since $B_{i1} = a_i + \pi(a_1)$ and $B_{i(K-1)} = a_i + \pi(a_{K-1})$, it

follows that if $\{B_{i1}\} = \{B_{i(K-1)}\}$, then there exists an $L \leq K$ and distinct $i_1, \ldots, i_L \in \{0, 2, 3 \ldots, K-1\}$ such that

$$\begin{aligned} B_{1(K-1)} &= B_{i_1 1} \\ B_{i_1(K-1)} &= B_{i_2 1} \\ &\vdots \\ B_{i_{L-1}(K-1)} &= B_{i_L 1} \\ B_{i_L(K-1)} &= B_{11}. \end{aligned}$$

This implies

$$\begin{aligned} \pi(a_{K-1}) - \pi(a_1) &= a_{i_1} - a_1 \\ &= a_{i_2} - a_{i_1} \\ &\vdots \\ &= a_1 - a_{i_L}. \end{aligned} \tag{3.21}$$

Since the terms on the right-hand side above sum to 0, we have $L[\pi(a_{K-1}) - \pi(a_0)] = 0$. As $a_{i_1}, \ldots, a_{i_L} \neq a_1$, this implies that L divides q, which in turn implies

$$\max_{i=0,\ldots,K-1} (a_i - a_{i-1}) \leq \lfloor q/2 \rfloor \tag{3.22}$$

(where $a_{-1} = a_{K-1}$) and thus

$$a_{K-1} - a_0 \geq \lfloor q/2 \rfloor.$$

We therefore have $1 \leq a_1 \leq \lfloor q/2 \rfloor < a_{K-1}$. It then follows from (3.21) that $a_{i_1} - a_1 = a_{K-1} - a_1 + 1$, i.e., $a_{i_1} = a_{K-1} + 1$, a contradiction.

$\underline{K = 2}$: Suppose, contrary to the claim, that $\{B_{00}, B_{10}\} = \{B_{01}, B_{11}\}$. This implies $B_{01} = B_{10}$, i.e.,

$$a_1 - a_0 = \pi(a_0) - \pi(a_1). \tag{3.23}$$

A similar reasoning to the one for the case $K \geq 3$ also yields (3.22). Since $K = 2$, it follows that $a_1 - a_0 = \lfloor q/2 \rfloor$. On the other hand, it follows from the definition of π that

$$a_1 - a_0 = \lfloor q/2 \rfloor \quad \text{implies} \quad \pi(a_0) - \pi(a_1) \neq \lfloor q/2 \rfloor,$$

contradicting (3.23). This completes the proof.

3.C An Alternative Proof of Polarization for Prime q

One can prove Theorem 3.1 by first showing that the Z parameters polarize through Arıkan's construction, which by Proposition 3.3 implies the polarization of entropies.

For this purpose, let us first define, for $d = 1, \ldots, q-1$, the parameters

$$Z_d(X \mid Y) := \sum_x \sum_y \sqrt{p(x,y)p(x+d,y)}.$$

It is easy to verify that $Z_d(X \mid Y)$ takes values in $[0,1]$. Clearly, $Z(X \mid Y)$ is the mean of Z_d's:

$$Z(X \mid Y) = \frac{1}{q-1} \sum_{d \neq 0} Z_d(X \mid Y).$$

We also define

$$Z_{\max}(X \mid Y) := \max_{d \neq 0} Z_d(X \mid Y).$$

We will show that the $Z_{\max}$'s created by Arıkan's construction converge to 0 or 1. In order to translate this to a polarization result for entropies, we need $Z_{\max}$ to satisfy

$$Z_{\max}(X \mid Y) \approx 1 \iff H(X \mid Y) \approx 1$$
$$Z_{\max}(X \mid Y) \approx 0 \iff H(X \mid Y) \approx 0.$$

The second of these relations is evident, since $Z(X \mid Y) \leq Z_{\max}(X \mid Y) \leq (q-1)Z(X \mid Y)$. The following lemma implies that the first relation also holds when q is prime:

Lemma 3.7. *For all prime q and $\delta > 0$, there exists $\eta(\delta, q) > 0$ such that $Z_{max}(X \mid Y) \geq 1 - \eta(\delta, q)$ implies $Z(X \mid Y) \geq 1 - \delta$.*

Proof. Let d be such that $Z_d(X \mid Y) = Z_{\max}(X \mid Y)$. Since q is prime, $\mathcal{X}$ can be written as

$$\mathcal{X} = \{a_i \colon a_i = x + id, i = 0, \ldots, q-1\}$$

for all $x \in \mathcal{X}$. Setting $\zeta_{x,x'} := \sum_y \sqrt{p(y \mid x)p(y \mid x')}$ we thus have

$$Z_d(X \mid Y) = \sum_{i=0}^{q-1} \sqrt{p_X(a_i)p_X(a_{i+1})} \cdot \zeta_{a_i,a_{i+1}}$$

It is easily verified that $Z_d(X \mid Y)$ is strictly concave in p_{XY}, attaining its maximum when p_X is the uniform distribution, and $\zeta_{a_i,a_{i+1}} = 1$ for all i. It then follows that there exists $\nu(\delta)$ such that $Z_d(X \mid Y) \geq 1 - \eta(\delta)$ implies

(i) $p_X(x) \geq 1/q - \nu(\delta)$ for all x,

(ii) $\zeta_{a_i,a_{i+1}} \geq 1 - \nu(\delta)$ for all i,

where $\nu \to 0$ as $\eta \to 0$. Now define

$$b_y = \sqrt{p(y \mid a_i)} - \sqrt{p(y \mid a_{i+1})},$$
$$c_y = \sqrt{p(y \mid a_{i+1})} - \sqrt{p(y \mid a_{i+2})}.$$

for all $y \in \mathcal{Y}$. The triangle inequality states that

$$\left(\sum_y (b_y + c_y)^2\right)^{1/2} \le \left(\sum_y b_y^2\right)^{1/2} + \left(\sum_y c_y^2\right)^{1/2},$$

or equivalently, that

$$\begin{aligned}\sqrt{1-\zeta_{a_i,a_{i+2}}} &\le \sqrt{1-\zeta_{a_i,a_{i+1}}} + \sqrt{1-\zeta_{a_{i+1},a_{i+2}}} \\ &\le 2\sqrt{\nu(\delta)}.\end{aligned}$$

Applying the above inequality repeatedly yields

$$\sqrt{1-\zeta_{x,x'}} \le (q-1)\sqrt{\nu(\delta)}$$

for all $x, x' \in \mathcal{X}$, which implies

$$\begin{aligned}Z(X \mid Y) &= \frac{1}{q-1} \sum_{x,x':x\neq x'} \sqrt{p(x)p(x')} \cdot \zeta_{x,x'} \\ &\ge [1 - q\nu(\delta)][1 - (q-1)^2\nu(\delta)],\end{aligned}$$

yielding the claim. □

Proposition 3.4. *If (X_1, Y_1) and (X_2, Y_2) are i.i.d., then*

$$\begin{aligned}Z_{max}(X_1 + X_2 \mid Y_1^2) &\le (q-1)(q^2-q+1) Z_{max}(X_1 \mid Y_1) \\ Z_{max}(X_2 \mid Y_1^2, X_1 + X_2) &= Z_{max}(X_1 \mid Y_1)^2.\end{aligned}$$

Proof. The first claim follows from (3.12):

$$\begin{aligned}Z_{\max}(X_1 + X_2 \mid Y_1^2) &\le (q-1) Z(X_1 + X_2 \mid Y_1^2) \\ &\le (q-1)(q^2-q+1) Z(X_1 \mid Y_1) \\ &\le (q-1)(q^2-q+1) Z_{\max}(X_1 \mid Y_1).\end{aligned}$$

To obtain the second claim we write

$$\begin{aligned}Z_d(X_2 \mid Y_1^2, X_1 + X_2) &= \sum_{x_2} \sum_{u,y_1,y_2} \big[p_{XY}(x_2, y_2) p_{XY}(x_2 + d, y_2)\big]^{1/2} \\ &\qquad \cdot \big[p_{XY}(u - x_2, y_1) p_{XY}(u - x_2 - d, y_1)\big]^{1/2} \\ &= \sum_{x_2,y_2} \big[p_{XY}(x_2, y_2) p_{XY}(x_2 + d, y_2)\big]^{1/2} \\ &\qquad \cdot \sum_{u,y_1} \big[p_{XY}(u - x_2, y_1) p_{XY}(u - x_2 - d, y_1)\big]^{1/2}\end{aligned}$$

Observing that both of the summations above are equal to $Z_d(X_1 \mid Y_1)$, we have $Z_d(X_2 \mid Y_1, Y_2, X_1 + X_2) = Z_d(X_1 \mid Y_1)^2$. This implies the claim since $t \to t^2$ is increasing for non-negative t. □

Lemma 3.8. *Suppose $B_1, B_2, \ldots$ are i.i.d., $\{-,+\}$-valued random variables with*

$$P(B_1 = -) = P(B_1 = +) = \tfrac{1}{2}$$

defined on a probability space $(\Omega, \mathcal{F}, P)$. Set $\mathcal{F}_0 = \{\phi, \Omega\}$ as the trivial σ-algebra and set $\mathcal{F}_n$, $n \geq 1$ to be the σ-field generated by $(B_1, \ldots, B_n)$.

Suppose further that two stochastic processes $\{I_n : n \geq 0\}$ and $\{T_n : n \geq 0\}$ are defined on this probability space with the following properties:

(i.1) I_n takes values in the interval $[0,1]$ and is measurable with respect to $\mathcal{F}_n$. That is, I_0 is a constant, and I_n is a function of $B_1, \ldots, B_n$.

(i.2) $\{(I_n, \mathcal{F}_n) : n \geq 0\}$ is a martingale.

(t.1) T_n takes values in the interval $[0,1]$ and is measurable with respect to $\mathcal{F}_n$.

(t.2) $T_{n+1} = T_n^2$ when $B_{n+1} = +$.

(i&t.1) For any $\epsilon > 0$ there exists $\delta > 0$ such that $I_n \in (\epsilon, 1-\epsilon)$ implies $T_n \in (\delta, 1-\delta)$.

Then, $I_\infty := \lim_{n\to\infty} I_n$ exists with probability 1, I_∞ takes values in $\{0,1\}$, and $P(I_\infty = 1) = I_0$.

Proof. The almost sure convergence of I_n to a limit follows from $\{I_n\}$ being a bounded martingale. Once it is known that I_∞ is $\{0,1\}$-valued it will then follow from the martingale property that $P(I_\infty = 1) = E[I_\infty] = I_0$. It thus remains to prove that I_∞ is $\{0,1\}$-valued. This in turn is equivalent to showing that for any $\eta > 0$,

$$P\big(I_\infty \in (\eta, 1-\eta)\big) = 0.$$

Since for any $0 < \epsilon < \eta$, the event $\big\{I_\infty \in (\eta, 1-\eta)\big\}$ is included in the event

$$J_\epsilon := \big\{\omega\colon \text{there exists } m \text{ such that for all } n \geq m,\ I_n \in (\epsilon, 1-\epsilon)\big\},$$

and since by property (i&t.1) there exists $\delta > 0$ such that $J_\epsilon \subset K_\delta$ where

$$K_\delta := \big\{\omega\colon \text{there exists } m \text{ such that for all } n \geq m,\ T_n \in (\delta, 1-\delta)\big\},$$

it suffices to prove that $P(K_\delta) = 0$ for any $\delta > 0$. This is trivially true for $\delta \geq 1/2$. Therefore, it suffices to show the claim for $0 < \delta < 1/2$. Given such a δ, find a positive integer k for which $(1-\delta)^{2^k} < \delta$. This choice of k guarantees that if a number $x \in [0, 1-\delta]$ is squared k times in a row, the result lies in $[0, \delta)$.

For $n \geq 1$ define E_n as the event that $B_n = B_{n+1} = \cdots = B_{n+k-1} = +$, i.e., E_n is the event that there are k consecutive $+$'s in the sequence $\{B_i : i \geq 1\}$ starting at index n. Note that $P(E_n) = 2^{-k} > 0$, and that $\{E_{mk} : m \geq 1\}$

is a collection of independent events. The Borel–Cantelli lemma thus lets us conclude that the event

$$\begin{aligned} E &= \{E_n \text{ occurs infinitely often}\} \\ &= \{\omega\colon \text{for every } m \text{ there exists } n \geq m \text{ such that } \omega \in E_n\} \end{aligned}$$

has probability 1, and thus $P(K_\delta) = P(K_\delta \cap E)$. We will now show that $K_\delta \cap E$ is empty, from which it will follow that $P(K_\delta) = 0$. To that end, suppose $\omega \in K_\delta \cap E$. Since $\omega \in K_\delta$, there exists m such that $T_n(\omega) \in (\delta, 1-\delta)$ whenever $n \geq m$. But since $\omega \in E$ there exists $n_0 \geq m$ such that $B_{n_0+1} = \cdots = B_{n_0+k-1} = +$, and thus $T_{n_0+k}(\omega) = T_{n_0}(\omega)^{2^k} \leq (1-\delta)^{2^k} < \delta$ which contradicts with $T_{n_0+k}(\omega) \in (\delta, 1-\delta)$. □

Proof of Theorem 3.1. Let $B_1, B_2, \ldots$ be an i.i.d. binary process with $\Pr[B_1 = +] = 1/2$. Define $H_0, H_1, \ldots$ and $Z_0, Z_1, \ldots$ as in (3.6) and (3.14), respectively. We will show that the conditions of Lemma 3.8 are satisfied if I_n and T_n are replaced with H_n and Z_n, respectively: That (i.1), (i.2) and (t.1) are satisfied is clear by the definitions of H_n and Z_n, (t.2) is established in Proposition 3.4, and (i&t.1) follows from Proposition 3.3 and Lemma 3.7. The claim is then a corollary to Lemma 3.8. □

4 Generalized Constructions

In the preceding chapters, polarization was achieved using a fixed recipe: Choose a transform that acts on two random variables, and use it recursively. For prime alphabet sizes, an appropriate choice of mapping was $(X_1, X_2) \to (X_1 + X_2, X_2)$, or equivalently

$$[U_1\ U_2] = [X_1\ X_2] \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Some thought reveals that an n-fold application of this mapping to a block of $N = 2^n$ symbols X_1^N is equivalent to [3]

$$U_1^N = X_1^N \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n} B_n,$$

where '$\otimes n$' is the nth Kronecker power of a matrix, and B_n is an $N \times N$ permutation matrix known as the *bit-reversal* operator. (Recall that the inclusion of the permutation matrix B_n is out of notational convenience only.) In this chapter, we will study generalizations of this method.

Finding transformations that polarize memoryless processes becomes an easy task if one completely disregards complexity issues. In fact, almost all invertible binary matrices polarize such processes. This is most easily seen in the following case. Consider an i.i.d. process $(X_1, Y_1), (X_2, Y_2), \ldots$ where X_1 is uniformly distributed on $\{0, 1\}$, and Y_1 is the output of a symmetric binary-input memoryless channel with input X_1. One can think of X_1^N as codewords obtained through

$$X_1^N = U_1^N G_N$$

where U_1^N is uniformly distributed over $\{0, 1\}^N$, and G_N is an invertible $\{0, 1\}$-matrix. Suppose that G_N is chosen through the following procedure: The

bottom $R = 1 - H(X_1 \mid Y_1) - \epsilon$ fraction of the rows are chosen independently and uniformly at random from $\{0,1\}^N$. These rows will be linearly independent with high probability. The remaining $1-R$ fraction of the rows are then chosen in any manner that ensures the invertibility of G_N. We know from [10, Section 6.2] that with high probability, the code generated by the bottom R fraction of the rows will have exponentially small error probability (in the blocklength) over the channel $X_1 \to Y_1$. This means, by virtue of Fano's inequality, that $H(U_{N(1-R)+1}^N \mid Y_1^N U_1^{N(1-R)})$ can be made arbitrarily small as N grows without bound, i.e.,

$$H(U_i \mid Y_1^N U_1^{i-1}) \to 0 \qquad \text{for all } i > N(1-R).$$

It also follows from the above relation and $H(U_1^N \mid Y_1^N) \geq NH(X_1 \mid Y_1)$ that almost all of the conditional entropies $H(U_i \mid Y_1^N U_1^{i-1})$ that are not close to zero must be close to one. That is, a typical random matrix generated in this fashion will polarize the underlying process. On the other hand, such matrices will typically have no useful structure, and thus one may not be able to find low-complexity algorithms to decode the generated codes. The decoding complexity of such codes will typically be exponential in the blocklength.

The above argument can be stated more generally. Observe that in a channel code with messages U_1^{NR}, codewords X_1^N, channel outputs Y_1^N and small block error probability, the entropy

$$H(U_1^{NR} \mid Y_1^N) = \sum_{i=1}^{NR} H(U_i \mid Y_1^N U_1^{i-1})$$

is also small. That is, almost all terms on the right-hand side of the above are close to 0. Hence, any good code can be thought of as one which polarizes the resulting process of channel inputs and outputs. A similar statement also holds for good source codes. Polarization, if defined as the creation of extremal entropies from mediocre ones, is then not peculiar to polar codes, but is common to all good codes. The main virtue of polar codes is not that they polarize processes, but that they do so in a recursive fashion. It is this recursive structure that enables their good performance under low-complexity successive cancellation decoding.

4.1 Recursive Transforms

In view of the above discussion, it is reasonable to restrict the search for methods of polarization to recursive ones. We will focus on the easiest way of obtaining such transforms: replacing the matrix $\left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$ in the original construction with another square matrix, possibly of a larger size. More precisely, we will assume that the process $(X_1, Y_1), (X_2, Y_2), \ldots$ is i.i.d. and X_1 takes values over a finite field $\mathbb{F}_q$ of prime size, and we will study transforms of the form

$$U_1^N = X_1^N G^{\otimes n} B_n, \tag{4.1}$$

where $N = \ell^n$, matrix multiplication is over $\mathbb{F}_q$, and G is an $\ell \times \ell$ $\mathbb{F}_q$-matrix with $\ell \geq 2$. The $N \times N$ permutation matrix B_n is defined analogously to the bit-reversal operation in the original construction: It corresponds to the permutation $f(i) = r_\ell(i-1)+1$, $i = 1, \ldots, N$, where $r_\ell(i) = j$ for i and j with ℓ-ary expansions $b_n \ldots b_1$ and $b_1 \ldots b_n$, respectively.

In addition to their low encoding and decoding complexity, codes based on recursive transforms are also amenable to error analysis. As in Arıkan's original construction, the large blocklength behavior of recursive transforms is dictated by certain properties of the basic transform G, and therefore several useful conclusions can be drawn simply by establishing these properties. We will in particular study the following questions: (i) What choices of G yield polarizing transforms? (ii) What is the error probability behavior of such codes? We will see that the answers to both questions are fairly simple.

4.2 Polarizing Matrices

We will say that a matrix G is a *polarizing matrix* if it is invertible and a recursive application of it as in (4.1) yields

$$\lim_{n\to\infty} \frac{1}{N} \left| \left\{ i \colon H(U_i \mid Y_1^N U_1^{i-1}) > 1 - \epsilon \right\} \right| = H(X_1 \mid Y_1)$$
$$\lim_{n\to\infty} \frac{1}{N} \left| \left\{ i \colon H(U_i \mid Y_1^N U_1^{i-1}) < \epsilon \right\} \right| = 1 - H(X_1 \mid Y_1)$$

for all $\epsilon > 0$ and all i.i.d. processes $(X_1, Y_1), (X_2, Y_2), \ldots$, exactly as in the original construction. Note that the invertibility of a matrix implies that the set of its non-zero entries includes a permutation. We will therefore assume throughout, and without loss of generality, that all of the diagonal entries of G are non-zero (for otherwise it can be reduced to this form by permuting its columns). Recall that a necessary condition for polarization is that the 'entropy paths' generated along the recursion always fork until they converge to 0 or 1 (see Figure 3.1), i.e., that at least one of the created entropies at each step be different from the others. This requirement is met by a large class of matrices:

Lemma 4.1. *Let $S_1^\ell = X_1^\ell G$ for some invertible matrix G.*

(i) If G is upper-triangular, then $H(S_i \mid Y_1^\ell S_1^{i-1}) = H(X_1 \mid Y_1)$ for all $i = 1, \ldots, \ell$.

(ii) If G is not upper-triangular, then for every $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ and $i \in \{1, \ldots, \ell\}$ such that

$$H(X_1 \mid Y_1) \in (\epsilon, 1-\epsilon)$$

implies

$$H(S_i \mid Y_1^\ell S_1^{i-1}) - H(X_1 \mid Y_1) > \delta(\epsilon).$$

Proof. Let g_{ij} denote the (i,j)th entry of G. If G is upper-triangular, $H(S_i \mid Y_1^\ell S_1^{i-1})$ can be written as

$$H(S_i \mid Y_1^\ell S_1^{i-1}) = H\Big(\sum_{j=1}^{i} g_{ji}X_j \,\Big|\, Y_1^\ell, g_{11}X_1, g_{12}X_1 + g_{22}X_2, \ldots, \sum_{j=1}^{i-1} g_{ji}X_j\Big).$$

Since G is invertible, its first $i-1$ columns are linearly independent, and therefore the above can be rewritten as

$$H(S_i \mid Y_1^\ell S_1^{i-1}) = H\Big(\sum_{j=1}^{i} g_{ji}X_j \,\Big|\, Y_1^\ell, X_1^{i-1}\Big) = H(X_i \mid Y_i),$$

proving (i). If on the other hand G is not upper-triangular, then let $i \in \{1,\ldots,\ell\}$ be the smallest index for which the ith column of G has at least two non-zero entries g_{ki} and g_{li} below and including the diagonal. (Such an i always exists.) Since $(X_1,Y_1),\ldots,(X_\ell,Y_\ell)$ are independent, and since summing independent random variables increases entropy, we have

$$\begin{aligned} H\big(S_i \mid Y_1^\ell S_1^{i-1}\big) &= H\Big(\sum_{j=1}^{\ell} g_{ji}X_j \,\Big|\, Y_1^\ell S_1^{i-1}\Big) \\ &\geq H\big(g_{ki}X_k + g_{li}X_l \mid Y_1^\ell S_1^{i-1}\big) \\ &= H\big(g_{ki}X_k + g_{li}X_l \mid Y_kY_l\big), \end{aligned}$$

where the second equality is due to the definition of i. Observe now that the last entropy term can be written as $H(\tilde{X}_k + \tilde{X}_l \mid Y_k, Y_l)$, where $\tilde{X}_k$ and $\tilde{X}_l$ are appropriately permuted versions of X_k and X_l, respectively. The claim then follows from Lemma 3.1. □

The following polarization result can be proven as a corollary to the above lemma, using the standard martingale argument (see proofs of Theorem 2.1, 3.1, or 3.4).

Theorem 4.1. *For all prime q, an invertible $\mathbb{F}_q$-matrix is polarizing if and only if it is not upper-triangular.*

The above theorem says that the class of polarizing matrices is large. One may therefore hope to find, in this large class, matrices that yield better codes than the original polar codes in terms of their error probabilities. We study this problem next.

4.3 Rate of Polarization

Recall that for constructions based on combining two random variables at a time, convergence of the Bhattacharyya parameters was exponential roughly

in the square root of the blocklength, i.e., we had

$$\lim_{n\to\infty} \frac{1}{N}\left|\left\{i\colon Z(U_i \mid Y_1^N U_1^{i-1}) < 2^{-N^\beta}\right\}\right| = 1 - H(X_1 \mid Y_1)$$

for all $\beta < 1/2$. Let us recall the reason behind this behavior: Throughout the recursion, a Bhattacharyya parameter is (roughly) squared in approximately half of the recursions, and is unaffected (i.e., raised to power 1) in the remaining recursions. Since each recursion also doubles the blocklength, a simple calculation shows that the *exponent* of a typical Bhattacharyya parameter Z is roughly $\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 1 = \frac{1}{2}$, i.e., $Z \approx 2^{-N^{1/2}}$. (Note that these statements still need proof, as they neglect the multiplicative constants appearing in the bounds on the Bhattacharyya parameters. See the discussion on page 15.) It is also intuitively evident that the same argument can be made for any recursive construction: If an $\ell \times \ell$ matrix G creates ℓ Bhattacharyya parameters that are roughly equal to $Z(X_1 \mid Y_1)^{a_1}, \ldots, Z(X_1 \mid Y_1)^{a_\ell}$, then after many recursions the exponent of a typical Bhattacharyya parameter would be given by $\mathtt{E} = \frac{1}{\ell}\log_\ell a_1 + \ldots + \frac{1}{\ell}\log_\ell a_\ell$, i.e., $Z \approx 2^{-N^{\mathtt{E}}}$. That is, the large scale behavior of the Bhattacharyya parameters—and therefore of the error probability—is determined by their one-step evolution. It thus suffices to study how the underlying matrix G transforms the Bhattacharyya parameters in a single recursion. It turns out that this transformation is determined largely by the *partial distances* of G^{-1}:

Definition 4.1. *Let G be an $\ell \times \ell$ matrix with rows $g_1, \ldots, g_\ell \in \mathbb{F}_q^\ell$. The partial distances $D_1, \ldots, D_\ell$ of G are defined as*

$$D_i = \mathtt{d}_H\left(\langle g_i\rangle, \langle g_{i+1}, \ldots, g_\ell\rangle\right),$$

where $\langle a\rangle$ denotes the vector space spanned by a, and

$$\mathtt{d}_H(\langle a\rangle, \langle b\rangle) := \min_{\substack{x\in\langle a\rangle, y\in\langle b\rangle \\ x\neq 0}} \mathtt{d}_H(x, y)$$

where $\mathtt{d}_H(x, y)$ denotes the Hamming distance between vectors x and y.

Proposition 4.1. *Let $S_1^\ell = X_1^\ell G$, and let $D_1, \ldots, D_\ell$ be the partial distances of G^{-1}. We have*

$$Z(S_i \mid Y_1^\ell S_1^{i-1}) \le q^{3\ell} Z(X_1 \mid Y_1)^{D_i}, \qquad i = 1, \ldots, \ell. \tag{4.2}$$

Proof. Note first that

$$p_{S_1^i Y_1^\ell}(s_1^i, y_1^\ell) = \sum_{s_{i+1}^\ell} p_{S_1^\ell Y_1^\ell}(s_1^\ell, y_1^\ell) = \sum_{s_{i+1}^\ell} \prod_{i=1}^{\ell} p_{XY}\left([s_1^\ell G^{-1}]_i, y_i\right).$$

We have

$$\begin{aligned}
&Z(S_i \,|Y_1^\ell S_1^{i-1}) \\
&\quad= \frac{1}{q-1}\sum_{s\neq s'}\sum_{y_1^\ell, s_1^{i-1}} \Big[p_{S_1^i Y_1^\ell}\big((s_1^{i-1}, s), y_1^\ell\big) p_{S_1^i Y_1^\ell}\big((s_1^{i-1}, s'), y_1^\ell\big)\Big]^{1/2} \\
&\quad= \frac{1}{q-1}\sum_{s\neq s'}\sum_{y_1^\ell, s_1^{i-1}} \Bigg[\sum_{v_{i+1}^\ell}\prod_i p_{XY}\big([(s_1^{i-1}, s, v_{i+1}^\ell)G^{-1}]_i, y_i\big) \\
&\qquad\qquad \cdot \sum_{w_{i+1}^\ell}\prod_i p_{XY}\big([(s_1^{i-1}, s', w_{i+1}^\ell)G^{-1}]_i, y_i\big)\Bigg]^{1/2} \\
&\quad\le \frac{1}{q-1}\sum_{s\neq s'}\sum_{y_1^\ell, s_1^{i-1}}\sum_{v_{i+1}^\ell, w_{i+1}^\ell} \Bigg[\prod_i p_{XY}\big([(s_1^{i-1}, s, v_{i+1}^\ell)G^{-1}]_i, y_i\big) \\
&\qquad\qquad \cdot p_{XY}\big([(s_1^{i-1}, s', w_{i+1}^\ell)G^{-1}]_i, y_i\big)\Bigg]^{1/2}.
\end{aligned} \tag{4.3}$$

Observe that for all s_1^{i-1}, v_{i+1}^ℓ, and w_{i+1}^ℓ we have

$$\mathtt{d}_H\left((s_1^{i-1}, s, v_{i+1}^\ell)G^{-1}, (s_1^{i-1}, s', w_{i+1}^\ell)G^{-1}\right) \ge D_i,$$

and therefore

$$\sum_{y_1^\ell}\Bigg[\prod_i p_{XY}\big([(s_1^{i-1}, s, v_{i+1}^\ell)G^{-1}]_i, y_i\big) \cdot p_{XY}\big([(s_1^{i-1}, s', w_{i+1}^\ell)G^{-1}]_i, y_i\big)\Bigg]^{1/2} \le \big[(q-1)Z(X_1 \mid Y_1)\big]^{D_i}.$$

Combining this relation with (4.3) yields the claim. □

We can now characterize the error probability behavior of general recursive polar codes. For this purpose, we first define the *exponent* $\mathtt{E}(G)$ of a matrix G, through the partial distances $D_1, \ldots, D_\ell$ of G^{-1}:

$$\mathtt{E}(G) := \frac{1}{\ell}\sum_{i=1}^{\ell}\log_\ell D_i. \tag{4.4}$$

Theorem 4.2. *Let G be an $\ell \times \ell$ polarizing matrix and U_1^N be defined as in (4.1). Then,*

$$\lim_{n\to\infty}\frac{1}{N}\left|\left\{i\colon Z(U_i \mid Y_1^N U_1^{i-1}) < 2^{-N^\beta}\right\}\right| = 1 - H(X_1 \mid Y_1)$$

for all $\beta < \mathtt{E}(G)$.

We defer the proof of Theorem 4.2 to Section 4.4.

4.3.1 Bounds on the Rate of Polarization

The importance of Proposition 4.1 and Theorem 4.2 is in identifying through $\mathtt{E}(G)$ the exponential dependence between the error probability and the blocklength. This significantly simplifies the search for good recursive constructions since $\mathtt{E}(G)$ is an easy-to-calculate algebraic quantity. One can also use the existing results on the minimum distance of codes to find useful bounds on the best possible $\mathtt{E}(G)$ for a given size, i.e., on

$$\mathtt{E}_\ell := \max_{G \in \mathbb{F}_q^{\ell \times \ell}} \mathtt{E}(G).$$

It is useful to note that recursive constructions may not be of much practical value for large values of ℓ: It can indeed be verified easily that the decoding complexity of codes based on a general $\ell \times \ell$ recursion is $O(q^\ell N \log N)$. We can therefore restrict our attention to small ℓ, for which one can either exactly compute or bound $\mathtt{E}_\ell$. Conveniently, even the simplest bounding techniques provide useful information at small sizes. The following upper and lower bounds on the partial distances—based on sphere packing and Gilbert–Varshamov type constructions, respectively—were given in [11] for the binary case:

Proposition 4.2.

$$\frac{1}{\ell} \sum_{i=1}^{\ell} \log_\ell \tilde{D}_i \le \mathtt{E}_\ell \le \frac{1}{\ell} \sum_{i=1}^{\ell} \log_\ell \hat{D}_i,$$

where

$$\hat{D}_i = \max \left\{ D \colon \sum_{j=0}^{\lfloor \frac{D-1}{2} \rfloor} \binom{\ell}{j} \le q^{i-1} \right\} \quad \textit{and} \quad \tilde{D}_i = \max \left\{ D \colon \sum_{j=0}^{D-1} \binom{\ell}{j} < q^i \right\}.$$

An improved version of these bounds, along with the exponents of a BCH code-based construction (both given in [11]) are plotted for $q = 2$ in Figure 4.1. These results are of a somewhat negative nature, as they show that the original exponent $1/2$ of Arıkan's construction cannot be improved with at small recursion sizes. It was in fact shown in [11] that $\mathtt{E}_\ell \le 1/2$ for $\ell < 15$, and that $\mathtt{E}_{16} \approx 0.51$. Nevertheless, it follows from the above bounds that one can attain 'almost exponential' error probability decay with the blocklength if the size of the recursion is sufficiently large:

Proposition 4.3 ([11])**.** *For all prime q,* $\lim_{\ell \to \infty} \mathtt{E}_\ell = 1$.

The case for generalized constructions is stronger in non-binary settings. The reason is that for a fixed matrix size, larger alphabet sizes allow for better separation (in the Hamming distance) between the rows of a matrix, yielding better exponents at any fixed ℓ. A simple evidence of this is given in the following result.

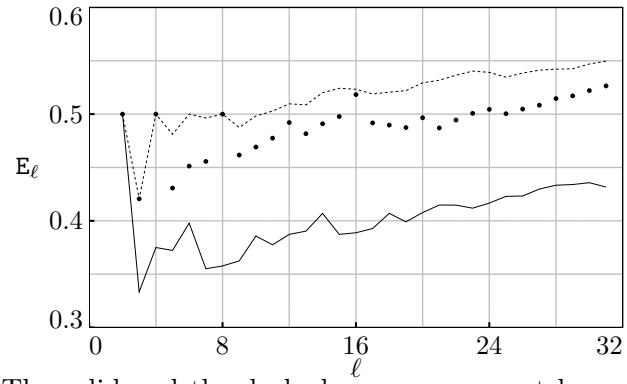

Figure 4.1: The solid and the dashed curves represent lower and upper bounds on $\mathtt{E}_\ell$ (in the binary case), respectively. The dots show the exponents of a BCH code-based construction (see [11]).

Theorem 4.3. *For $\ell \le q$, $\mathtt{E}_\ell = \frac{1}{\ell}\log_\ell(\ell!)$.*

Proof. Observe first that $D_i \le i$ for any invertible matrix. To see this, note that the invertibility of a matrix G with rows $g_1, \dots, g_\ell$ implies that $g_{i+1}, \dots, g_\ell$ have $\ell - i$ linearly independent columns, and thus span $\mathbb{F}_q^{\ell-i}$ at the locations corresponding to these columns. Therefore, g_i can at most be at a distance i from $\langle g_{i+1}, \dots, g_\ell\rangle$.

To prove the claim, we only need to find a matrix with $D_i = i$. To that end, let ω be an arbitrary element of $\mathbb{F}_q$ other than the identity, and let G be the matrix with rows

$$g_i = \left[1, \omega^i, \omega^{2i}, \dots, \omega^{(\ell-1)i}\right]$$

That is, G is the generator matrix of a Reed–Solomon code of rate 1. It is known that the minimum distance of the code $\langle g_i, \dots, g_\ell\rangle$ is i [12, Ch. 10.2], and therefore

$$D_i = \mathtt{d}_H\big(\langle g_i\rangle, \langle g_{i+1}, \dots, g_\ell\rangle\big) \ge i. \qquad \square$$

The above theorem implies that for $q \ge 5$, we have $\mathtt{E}_2 = 0.5$, $\mathtt{E}_3 \approx 0.54$, $\mathtt{E}_4 \approx 0.57$, and $\mathtt{E}_5 \approx 0.59$. Compare these with the upper bounds given in Figure 4.1 for the binary case.

4.4 Proof of Theorem 4.2

We will not provide the proof in full, since it is an almost identical reproduction of the proof of Theorem 2.2 once we obtain the following result.

Lemma 4.2. *Let $B_1, B_2, \dots$ be an i.i.d. process where B_1 is uniformly distributed over $\{1, 2, \dots, \ell\}$. Also let $Z_0, Z_1, \dots$ be a $[0, 1]$-valued random process where Z_0 is constant and*

$$Z_{n+1} \le K Z_n^{D_i} \qquad \text{whenever } B_n = i$$

for some $K > 0$ and $2 \leq D_1 \leq \ell$ and $1 \leq D_2, \ldots, D_\ell \leq \ell$. Suppose also that Z_n converges almost surely to a $\{0,1\}$-valued random variable Z_∞ with $\Pr[Z_\infty = 0] = z$. Then, for any $\beta < E$ where

$$E = \frac{1}{\ell} \sum_i \log_\ell D_i$$

we have

$$\lim_{n \to \infty} \Pr[Z_n \leq 2^{-\ell^{\beta n}}] = z.$$

Remark 4.1. *Note that the definition of the process $Z_0, Z_1, \ldots$ reflects the transformation of Bhattacharyya parameters in a single recursion (4.2): All partial distances $D_1, \ldots, D_\ell$ of a polarizing matrix are ≥ 1 (since the matrix is invertible), with at least one partial distance ≥ 2 (since the matrix is not upper-triangular).*

This result was originally proven for $\ell = 2$ by Arıkan and Telatar in [4]. We will provide the general proof in full for completeness, although it is a straightforward extension of the bounding technique given in [4]. As the technique is slightly intricate, it is useful to briefly explain the ideas contained in it: Note first that for $K \leq 1$ the result is a simple corollary to the weak law of large numbers: In a sufficiently long sequence $B_1, \ldots, B_n$, each exponent D_i appears nearly n/ℓ times with high probability, and thus a typical Z_n is less than

$$Z_0^{\prod_i D_i^{n/\ell}} = (1/Z_0)^{-\ell^{nE}}.$$

It can easily be seen that this method does not yield a useful bound when $K > 1$. The proof given below is instead based on the following observations: Whenever Z_n converges to zero, there must be a finite point n_0 for which the sequence Z_n, $n > n_0$ stays below a given positive threshold ϵ (Lemma 4.4). This threshold can be chosen sufficiently small so that if $Z_n \leq \epsilon$, then KZ_n^d is approximately the same as Z_n^d if $d > 1$, i.e., multiplying Z_n with K has negligible effect compared with exponentiating it. Once this is established, one can again appeal to the law large numbers as in the case $K \leq 1$ to obtain the result.

Lemma 4.3. *Let $a_0, a_1, \ldots$ be a sequence of numbers satisfying*

$$a_{i+1} = b_{i+1} a_i + K, \qquad i = 0, 1, \ldots$$

where $K > 0$ and $b_i \geq 1$ for all i. Then,

$$a_n \leq (a_0 + Kn) \prod_{i=1}^{n} b_i.$$

Proof. A straightforward computation shows that

$$a_n = a_0 \prod_{i=1}^{n} b_i + K \sum_{i=1}^{n} \prod_{j>i} b_j$$

from which the claim follows trivially. □

Lemma 4.4. *For every $\epsilon > 0$, there exists an $m(\epsilon)$ such that*

$$\Pr[Z_n \leq 1/K^{\ell+1} \text{ for all } n \geq m(\epsilon)] > z - \epsilon.$$

Proof. Let $\Omega = \{\omega \colon Z_n(\omega) \to 0\}$, and note that $\Pr[\Omega] = z$. Also observe that since Z_n is non-negative, Ω can be written as

$$\begin{aligned}\Omega &= \Big\{\omega \colon \text{for all } k \geq 1 \text{ there exists } n_0(\omega) \\ &\qquad \text{such that } Z_n(\omega) < 1/k \text{ for all } n \geq n_0(\omega)\Big\} \\ &= \bigcap_{k \geq 1} \bigcup_{n_0 \geq 0} A_{n_0,k},\end{aligned}$$

where $A_{n_0,k} = \{\omega \colon Z_n(\omega) < 1/k \text{ for all } n \geq n_0\}$. (Note that n_0 in the definition of $A_{n_0,k}$ is independent of ω.) Since the sets $A_{n_0,k}$ are increasing in n_0, for all $\epsilon > 0$ there exists an $m(\epsilon)$ for which $\Pr[A_{m(\epsilon),k}] > \Pr[\cup_{n_0 \geq 0} A_{n_0,k}] - \epsilon$, and thus taking $k = K^{\ell+1}$ we have

$$\Pr[A_{m(\epsilon),K^{\ell+1}}] > \Pr[\cup_{n_0 \geq 0} A_{n_0,K^{\ell+1}}] - \epsilon \geq \Pr[\Omega] - \epsilon,$$

yielding the claim. □

Lemma 4.5. *For all $\epsilon > 0$, there exists an $n(\epsilon)$ such that*

$$\Pr[\log_K Z_n < -n/4\ell] > z - \epsilon$$

for all $n \geq n(\epsilon)$.

Proof. Given $\epsilon > 0$, choose m and $A_{m,K^{\ell+1}}$ as in the proof Lemma 4.4. Observe that inside the set $A_{m,K^{\ell+1}}$ we have, conditioned on $B_n = i$,

$$\begin{aligned}Z_{n+1} &\leq K Z_n^{D_i} \\ &\leq K^{1-(D_i-1)(\ell+1)} Z_n \\ &\leq \begin{cases} K^{-\ell} Z_n & \text{if } B_n = 1 \\ K Z_n & \text{if } B_n = 2, \ldots, \ell \end{cases},\end{aligned}$$

or equivalently

$$\begin{aligned}\log_K Z_{n+1} &\leq \log_K Z_n - \ell \quad \text{if } B_n = 1 \\ \log_K Z_{n+1} &\leq \log_K Z_n + 1 \quad \text{if } B_n = 2, \ldots, \ell.\end{aligned}$$

This implies that inside the set $A_{m,K^{\ell+1}}$

$$\log_K Z_n \le \log_K Z_m + (n-m)(1-\alpha(\ell+1))$$

where α is the fraction of 1's in the sequence $B_m, \ldots, B_n$. Let $T^n_{m,\alpha}$ denote the event that the sequence $B_m, \ldots, B_n$ contains at least an α fraction of each letter $k \in \{1, \ldots, \ell\}$. Now choose $n_0 \ge 2m$ such that $\Pr[T^n_{m,\alpha}] > 1-\epsilon$ for all $n \ge n_0$ with $\alpha = (2\ell+1)/[(2\ell+2)\ell]$. Note that such an n_0 exists since $\alpha < 1/\ell$. Then we have inside the set $A_{m,K^{\ell+1}} \cap T^n_{m,\alpha}$

$$\begin{aligned}\log_K Z_n &\le \log_K Z_m - \tfrac{n}{2}(1-\alpha(\ell+1)) \\ &\le -n/4\ell.\end{aligned}$$

Observing that $\Pr[A_{m,K^{\ell+1}} \cap T^n_{m,\alpha}] \ge z - 2\epsilon$ yields the claim. □

Proof of Lemma 4.2. We only need to prove the claim for $K > 1$. Given $\epsilon > 0$, choose $\alpha < 1/\ell$ and $\gamma < 1$ such that $\alpha\gamma\ell > 1-\epsilon$. Also let n be sufficiently large so that $n_1 := \log_\ell(2nK)8K/E\alpha$ $n_2 := n_1/8\ell K$ satisfy

(i) $n_1 > \max(n_0, 8\ell)$, where n_0 is as in Lemma 4.5,

(ii) $\Pr[T^{n_1+n_2}_{n_1,\alpha}] > 1-\epsilon$, where $T^{n_1+n_2}_{n_1,\alpha}$ is defined as in the proof of Lemma 4.5,

(iii) $\Pr[T^n_{n_1+n_2,\alpha}] > 1-\epsilon$, and

(iv) $n - (n_1+n_2) \ge \gamma n$.

Conditions (i)–(iii) imply that the probability of the set

$$A = \{\log_K Z_{n_1} \le -n_1/4\ell\} \cap T^{n_2}_{n_1,\alpha} \cap T^n_{n_1+n_2,\alpha}$$

is at least $z - 3\epsilon$. Observe also that the process $L_n = \log_K Z_n$ satisfies

$$L_{n+1} \le D_i L_n + K \quad \text{if } B_n = i.$$

Since inside the set A we have $B_n = i$ for at least an α fraction of B_n, it follows from Lemma 4.3 that

$$\begin{aligned}L_{n_1+n_2} &\le (-n_1/4\ell + n_2 K) \prod_{m=n_1}^{n_1+n_2} D_{B_m} \\ &\le -\prod_{m=n_1}^{n_1+n_2} D_{B_m} \\ &\le -\prod_{i=1}^{\ell} D_i^{\alpha n_2} \\ &= -\ell^{E\ell\alpha n_2}.\end{aligned}$$

Similarly bounding L_n we obtain

$$\begin{aligned}
L_n &\leq (L_{n_2} + [n - n_1 - n_2]K) \prod_{m=n_1+n_2}^{n} D_{B_m} \\
&\leq (-\ell^{E\ell\alpha n_2} + nK) \prod_{m=n_1+n_2}^{n} D_{B_m} \\
&\leq (-\ell^{E\alpha n_1/8K} + nK) \prod_{m=n_1+n_2}^{n} D_{B_m} \\
&\leq (-\ell^{E\alpha n_1/8K}/2) \prod_{m=n_1+n_2}^{n} D_{B_m} \\
&\leq -\prod_{m=n_1+n_2}^{n} D_{B_m} \\
&\leq -\prod_{i=1}^{\ell} D_i^{\alpha(n-n_1-n_2)} \\
&= -\ell^{E\ell\alpha(n-n_1-n_2)} \\
&\leq -\ell^{E\ell\alpha\gamma n} \\
&\leq -\ell^{En(1-\epsilon)}
\end{aligned}$$

which implies that with probability at least $z - 3\epsilon$

$$Z_n \leq K^{-\ell^{(1-\epsilon)En}} = 2^{-\ell^{[(1-\epsilon)E - \log_\ell(\log_2 K)/n]n}},$$

yielding the claim. □

5 Processes with Memory

We have seen in Chapters 3 and 4 that memoryless processes with finite alphabets can be polarized by recursive transforms, generalizing Arıkan's results on binary channel and source polarization to all stationary memoryless processes. In this chapter, we will see that the boundaries of this generality extend beyond memoryless processes. We will show in particular that any recursive transform that polarizes memoryless processes can also be used, as is, for polarizing a large class of processes with memory.

In order to keep the notation simple, we will restrict our attention to transforms that combine two random variables at a time, although the results in this chapter apply also to more general transforms. Recall once again that all that is required of a transform $(X_1, X_2) \to (f(X_1, X_2), X_2)$ to polarize a memoryless process is the strict separation of the created entropies, i.e., that the function f be such that

$$H(X_1 \mid Y_1) \leq H(f(X_1, X_2) \mid Y_1^2) \tag{5.1}$$

holds strictly for all i.i.d. (X_1, Y_1) and (X_2, Y_2) with moderate conditional entropy $H(X_1 \mid Y_1)$ (see Lemma 3.1). The nature of the recursive construction then ensures that the random variables combined at each step are i.i.d. and thus satisfy (5.1) with strict inequality unless they already have extremal entropies.

Unfortunately, the above argument does not hold in the presence of memory in the underlying process, as the strictness of the inequality in (5.1) relies strongly on the independence assumption. Not only (5.1) may hold with equality for such a process, but it may also not hold at all. On the other hand, for polarization to take place it suffices that the inequality be strict *eventually* at every step of the construction and for *almost all* random variables, ensuring the bifurcation of the entropy paths. We will prove polarization by showing that this requirement is fulfilled by a large class of processes.

5.1 Problem Statement and Main Result

Suppose $(X_1, Y_1), (X_2, Y_2), \dots$ is a stationary and ergodic (i.e., positive recurrent and aperiodic), Markov process of order $\kappa < \infty$, taking values in a finite set $\mathcal{X} \times \mathcal{Y}$. We will let $\mathcal{H}_{X|Y}$ denote the entropy rate of $X_1, X_2, \dots$ conditioned on $Y_1, Y_2, \dots$, i.e.,

$$\mathcal{H}_{X|Y} = \lim_{N\to\infty} \frac{1}{N} H(X_1^N \mid Y_1^N) = \lim_{N\to\infty} H(X_N | X_1^{N-1} Y_1^N).$$

Let $f\colon \mathcal{X}^2 \to \mathcal{X}$ be a polarizing mapping (see Definition 3.1). For all n and $N = 2^n$ let $G_n\colon \mathcal{X}^N \to \mathcal{X}^N$ be the recursive mapping defined via f (i.e., equations (3.16)), and let

$$U_1^N = G_n(X_1^N).$$

As the invertibility of G_n implies

$$\sum_{i=1}^{N} H(U_i \mid Y_1^N U_1^{i-1}) = H(X_1^N \mid Y_1^N),$$

we have

$$\lim_{n\to\infty} \frac{1}{N} \sum_{i=1}^{N} H(U_i \mid Y_1^N U_1^{i-1}) = \mathcal{H}_{X|Y}. \tag{5.2}$$

We will show that if G_n is a polarizing transform for memoryless processes, then it also polarizes Markov processes of arbitrary finite order:

Theorem 5.1. *For all $\epsilon > 0$,*

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{N} \left| \left\{ i\colon H(U_i \mid Y_1^N U_1^{i-1}) > 1 - \epsilon \right\} \right| &= \mathcal{H}_{X|Y}, \\ \lim_{n\to\infty} \frac{1}{N} \left| \left\{ i\colon H(U_i \mid Y_1^N U_1^{i-1}) < \epsilon \right\} \right| &= 1 - \mathcal{H}_{X|Y}. \end{aligned} \tag{5.3}$$

The remainder of this chapter is devoted to the proof of Theorem 5.1. We will prove the result for prime $q = |\mathcal{X}|$ and take f to be the modulo-q addition. The proof in the composite case follows similar arguments but is more tedious.

The techniques we will use for proving Theorem 5.1 are similar to those in the memoryless case. As we have discussed above, however, the memory in the underlying processes introduces several technical difficulties to be handled. It is therefore useful to construct the proof through a number of intermediate lemmas that resolve each of these difficulties. Let us first outline the basic ideas: For notational convenience, we will define

$$V_1^N = G_n(X_{N+1}^{2N}).$$

Our aim is to show that for large N, the inequality

$$H(U_i \mid Y_1^N U_1^{i-1}) \le H(U_i + V_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}) \tag{5.4}$$

holds strictly for almost all i's for which $H(U_i \mid Y_1^N U_1^{i-1})$ is moderate. For this purpose, first observe that since the process $(X_1, Y_1), (X_2, Y_2), \ldots$ is stationary, there is vanishing amount of per-letter dependence between non-overlapping blocks of it. That is, for large N,

$$\frac{1}{N} I(U_1^N; V_1^N \mid Y_1^{2N}) = \frac{1}{N} I(X_1^N; X_{N+1}^{2N} \mid Y_1^{2N}) \approx 0,$$

from which we will conclude that for almost all i, U_i and V_i are almost independent given their past, i.e.,

$$I(U_i; V_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}) \approx 0.$$

Thus, the conditional distribution of $U_i + V_i$ will be approximately equal to the convolution of the two distributions. This alone does not suffice to yield (5.4), however: Although we have seen in Lemma 3.4 that modulo-q addition of independent, moderate-entropy random variables strictly increases entropy, a conditional version of this statement is not true in general. This can be seen in the following example.

Example 5.1. *Let $X_1, X_2 \in \{0, \ldots, q-1\}$ and $Y \in \{0, 1\}$ be random variables with*

$$p(x_1, x_2 \mid y) = \begin{cases} 1/q^2 & \text{if } y = 0 \\ 1 & \text{if } y = 1 \text{ and } x_1 = x_2 = 0 \\ 0 & \text{otherwise} \end{cases}.$$

Note that X_1 and X_2 are i.i.d. conditioned on Y. It is also easy to see that $H(X_1 + X_2 \mid Y) = H(X_1 \mid Y) = p_Y(0)$.

The above example illustrates the only case where a modulo-q addition of conditionally independent random variables does not increase entropy: the case in which X_1 and X_2 are simultaneously constant or simultaneously uniform for all realizations of Y. We now proceed to the proof of Theorem 5.1, by first showing that a conditional version of Lemma 3.4 holds excluding this anomalous case. We will later see that the nature of ergodic Markov processes precludes this anomaly in a polarization construction.

Given $X_1, X_2 \in \mathcal{X}$ and Y, let H_i denote the random variable that takes the value $H(X_i \mid Y = y)$ whenever $Y = y$. Given $0 < \delta < \frac{1}{2}$, define two random variables $S_1, S_2 \in \{0, 1, 2\}$ through

$$S_i = \begin{cases} 0 & \text{if } H_i \in [0, \delta) \\ 1 & \text{if } H_i \in [\delta, 1-\delta] \\ 2 & \text{if } H_i \in (1-\delta, 1] \end{cases}, \qquad i = 1, 2.$$

Note that the irregularity described in Example 5.1 corresponds to the case where $S_1 = S_2 \in \{0, 2\}$ with probability 1.

Lemma 5.1. *If*

(i) $I(X_1; X_2 \mid Y) \leq \epsilon_2$, *and*

(ii) $\Pr[S_1 = S_2 \in \{0, 2\}] < 1 - \eta$ *for some* $\eta > 0$,

then there exist $\mu(\delta, \eta) > 0$ *and* $\nu(\epsilon_2)$ *such that*

$$H(X_1 + X_2 \mid Y) \geq \min_{i \in \{1,2\}} H(X_i \mid Y) + \mu(\delta, \eta) - \nu(\epsilon_2),$$

where $\nu(\epsilon_2) \to 0$ *as* $\epsilon_2 \to 0$.

Proof. We have

$$I(X_1; X_2 \mid Y) = \sum_{y \in \mathcal{Y}} P_Y(y) D\big(P_{X_1 X_2 | Y=y} \| P_{X_1|Y=y} \cdot P_{X_2|Y=y}\big) \leq \epsilon_2.$$

Therefore, the set

$$C = \big\{y \colon D\big(P_{X_1 X_2 | Y=y} \| P_{X_1|Y=y} \cdot P_{X_2|Y=y}\big) \leq \sqrt{\epsilon_2}\big\}$$

has probability at least $1 - \sqrt{\epsilon_2}$. Also, Pinsker's inequality implies that

$$\| P_{X_1 X_2 | Y=y} - P_{X_1|Y=y} \cdot P_{X_2|Y=y} \| \leq 2\epsilon_2^{1/4} \tag{5.5}$$

for all $y \in C$. Let $\tilde{X}_1$ and $\tilde{X}_2$ be random variables with

$$P_{\tilde{X}_1, \tilde{X}_2 | Y}(x_1, x_2 \mid y) = P_{X_1|Y}(x_1 \mid y) P_{X_2|Y}(x_2 \mid y).$$

Since $H(X_1 + X_2 \mid Y = y)$ is continuous in $P_{X_1 X_2 | Y=y}$, (5.5) implies that for all $y \in C$ we have

$$H(X_1 + X_2 \mid Y = y) \geq H(\tilde{X}_1 + \tilde{X}_2 \mid Y = y) - \epsilon(\epsilon_2),$$

where $\epsilon(\epsilon_2) \to 0$ as $\epsilon_2 \to 0$. We can then write

$$\begin{aligned} H(X_1 + X_2 \mid Y) &= \sum_{y \in Y} p(y) H(X_1 + X_2 \mid Y = y) \\ &\geq \sum_{y \in C} p(y) H(X_1 + X_2 \mid Y = y) \\ &\geq \sum_{y \in C} p(y) H(\tilde{X}_1 + \tilde{X}_2 \mid Y) - \epsilon(\epsilon_2). \end{aligned} \tag{5.6}$$

Next, note that the event $\big\{y \colon S_1 = S_2 \in \{0, 2\}\big\}^c$ is identical to

$$\Big\{y \colon \min\{H_1(y), 1 - H_2(y)\} > \delta \text{ or } \min\{H_2(y), 1 - H_1(y)\} > \delta\Big\}.$$

We can assume, without loss of generality, that condition (ii) of the lemma then implies that the event

$$D = \big\{y\colon \min\{H_2(y), 1 - H_1(y)\} > \delta\big\}$$

has probability at least $\eta/2$, and therefore

$$\Pr[C \cap D] \geq \eta/2 - \sqrt{\epsilon_2}.$$

It then follows from Lemma 3.4 that

$$H(\tilde{X}_1 + \tilde{X}_2 \mid Y = y) \geq H(X_1 \mid Y = y) + \epsilon_1(\delta)$$

for all $y \in C \cap D$, where $\epsilon_1(\delta) > 0$. Since we also have $H(A + B) \geq \max\{H(A), H(B)\}$ for independent random variables A and B, We can continue (5.6) as

$$\begin{aligned}
H(X_1 + X_2 \mid Y) &\geq \sum_{y \in C} p(y) H(\tilde{X}_1 + \tilde{X}_2 \mid Y = y) - \epsilon(\epsilon_2) \\
&\geq \sum_{y \in C \cap D^c} p(y) H(X_1 \mid Y = y) \\
&\quad + \sum_{y \in C \cap D} p(y)[H(X_1 \mid Y = y) + \epsilon_1(\delta)] - \epsilon(\epsilon_2) \\
&= \sum_{y \in C} p(y) H(X_1 \mid Y = y) \\
&\quad + \Pr[Y \in C \cap D]\epsilon_1(\delta) - \epsilon(\epsilon_2) \\
&\geq H(X_1 \mid Y) - \Pr[Y \notin C] + [\eta/2 - \sqrt{\epsilon_2}]\epsilon_1(\delta) - \epsilon(\epsilon_2) \\
&\geq H(X_1 \mid Y) + \tfrac{\eta}{2}\epsilon_1(\delta) - 2\sqrt{\epsilon_2} - \epsilon(\epsilon_2).
\end{aligned}$$

Defining $\mu(\delta, \eta) := \frac{\eta}{2}\epsilon_1(\delta)$ and noting that $2\sqrt{\epsilon_2} + \epsilon(\epsilon_2) \to 0$ as $\epsilon_2 \to 0$ yields the claim. □

We next show (Lemma 5.4) that there is sufficient independence between the pasts of U_i and V_i, i.e., between $(Y_1^N U_1^{i-1})$ and $(Y_{N+1}^{2N} V_1^{i-1})$ to satisfy condition (ii) of Lemma 5.1: For this purpose, we will let $H_{u,i}$ denote the random variable that takes the value $H(U_i \mid Y_1^N = y_1^N, U_1^{i-1} = u^{i-1})$ whenever $(Y_1^N U_1^{i-1}) = (y_1^N u^{i-1})$, similarly to H_i above. Also analogously to the above, we define a sequence of random variables $S_{u,i}$ through

$$S_{u,i} = \begin{cases} 0 & \text{if } H_{u,i} \in [0, \delta/2) \\ 1 & \text{if } H_{u,i} \in [\delta/2, 1 - \delta/2] \\ 2 & \text{if } H_{u,i} \in (1 - \delta/2, 1] \end{cases}, \quad i = 1, \ldots, N. \tag{5.7}$$

Similarly define random variables $H_{v,i}$ and $S_{v,i}$ by replacing the U's with V's above.

Lemma 5.2. *Let $(X_1, Y_1), (X_2, Y_2) \ldots$ be as before, and let $S = f(X_1^N, Y_1^N)$ and $T = f(X_{N+1}^{2N}, Y_{N+1}^{2N})$ for some (possibly probabilistic) mapping f. Then,*

$$H(S \mid T) + H(T \mid S) \geq H(S) - I(X_1^{\kappa+1} Y_1^{\kappa+1}; X_{N-\kappa}^N Y_{N-\kappa}^N).$$

We will use the following inequality in the proof of Lemma 5.2.

Lemma 5.3. *For random variables A, B, and C, we have*

$$H(A \mid B) + H(A \mid C) \geq H(A) - I(B; C).$$

Proof.

$$\begin{aligned} H(A) - H(A \mid B) &= I(A; B) \\ &\leq I(AC; B) \\ &\leq H(AC) - H(C \mid B) \\ &= H(A \mid C) + I(B; C). \end{aligned}$$

□

Proof. The conditions of the lemma, in addition to the stationarity of $\{X_i, Y_i\}$ imply that $(S, X_1^{\kappa+1}, Y_1^{\kappa+1})$ and $(T, X_{N+1}^{N+\kappa+1}, Y_{N+1}^{N+\kappa+1})$ are identically distributed, and that S—$X_{N-\kappa}^N Y_{N-\kappa}^N$ —$X_{N+1}^{N+\kappa+1} Y_{N+1}^{N+\kappa+1}$—$T$ is a Markov chain. We therefore have

$$\begin{aligned} H(S \mid T) + H(T \mid S) &\geq H(S \mid X_{N-\kappa}^N Y_{N-\kappa}^N) + H(T \mid X_{N+1}^{N+\kappa+1} Y_{N+1}^{N+\kappa+1}) \\ &= H(S \mid X_{N-\kappa}^N Y_{N-\kappa}^N) + H(S \mid X_1^{\kappa+1} Y_1^{\kappa+1}) \\ &\geq H(S) - I(X_1^{\kappa+1} Y_1^{\kappa+1}; X_{N-\kappa}^N Y_{N-\kappa}^N), \end{aligned}$$

where the second inequality follows from Lemma 5.3. □

Lemma 5.4. *For any $\delta > 0$, there exists $N_0(\delta)$ and $\eta(\delta) > 0$ such that whenever $N > N_0(\delta)$, $H(U_i \mid Y_1^N U_1^{i-1}) \in (\delta, 1 - \delta)$ implies*

$$\Pr\left[S_{u,i} = S_{v,i} \in \{0, 2\}\right] < 1 - \eta(\delta).$$

Proof. Note first that when $H(U_i \mid Y_1^N U_1^{i-1}) \in (\delta, 1 - \delta)$, there exists $\epsilon(\delta) > 0$ such that if $H(S_{u,i}) < \epsilon$, then $\Pr[S_{u,i} = 1] > \epsilon$. The latter inequality would then yield the claim. If, on the other hand $H(S_{u,i}) > \epsilon$, then we have from Lemma 5.2 that

$$\begin{aligned} \max\{H(S_{u,i} \mid S_{v,i}), &H(S_{v,i} \mid S_{u,i})\} \\ &\geq \tfrac{1}{2}[H(S_{u,i}) - I(X_1^{\kappa+1} Y_1^{\kappa+1}; X_{N-\kappa}^N Y_{N-\kappa}^N)] \\ &\geq \tfrac{1}{2}[\epsilon - I(X_1^{\kappa+1} Y_1^{\kappa+1}; X_{N-\kappa}^N Y_{N-\kappa}^N)] \\ &\geq \tfrac{1}{4}\epsilon, \end{aligned}$$

where the last inequality is obtained by choosing N sufficiently large, and by noting that aperiodic and positive recurrent Markov processes are mixing, from which it follows that $I(X_1^{\kappa+1} Y_1^{\kappa+1}; X_{N-\kappa}^N Y_{N-\kappa}^N) \to 0$. The lemma is then an easy corollary to the last inequality. □

5.2 Proof of Theorem 5.1

Following the proof of polarization in the memoryless case, we first define a $\{-,+\}$-valued i.i.d. process $B_1, B_2, \ldots$ with $\Pr[B_1 = -] = 1/2$, and associate to it a $[0,1]$-valued process $H_0, H_1, \ldots$ through

$$\begin{aligned} H_0 &= H(X_1 \mid Y_1) \\ H_n &= H_{n-1}^{B_n}, \qquad n = 1, 2, \ldots \end{aligned}$$

where we define for every N

$$\begin{aligned} H(U_i \mid Y_1^N U_1^{i-1})^- &:= H(U_i + V_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}) \\ H(U_i \mid Y_1^N U_1^{i-1})^+ &:= H(V_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}, U_i + V_i) \end{aligned}$$

It suffices to show that H_n converges almost surely to a $\{0,1\}$-valued random variable. To that end, we first write

$$\begin{aligned} H(U_i \mid Y_1^N U_1^{i-1})^- + H(U_i \mid Y_1^N U_1^{i-1})^+ & \\ &= H(U_i V_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}) \\ &\le H(U_i \mid Y_1^N U_1^{i-1}) + H(V_i \mid Y_{N+1}^{2N} V_1^{i-1}) \\ &= 2H(U_i \mid Y_1^N U_1^{i-1}) \end{aligned} \tag{5.8}$$

In the above, the last equality is due to the stationarity assumption. Since H_n takes values in $[0,1]$, it follows from (5.8) that the process $\{H_n\}$ is a bounded supermartingale, and therefore converges almost surely to a random variable H_∞. Since almost sure convergence implies convergence in probability, we conclude that the limit

$$\lim_{n\to\infty} \Pr[H_n(\delta, 1-\delta)]$$

exists. We will obtain the claim if we can show that this limit is equal to zero for all $\delta > 0$. This is equivalent to showing that for all $\delta, \epsilon > 0$, there exists n_0 such that

$$\Pr[H_n \in (\delta, 1-\delta)] < \epsilon \tag{5.9}$$

for all $n > n_0$. We do this next: Note first that

$$\begin{aligned} \tfrac{1}{2} E\left[|H_n^- - H_n|\right] &\le \tfrac{1}{2} E\left[|H_n^- - H_n|\right] + \tfrac{1}{2} E\left[|H_n^+ - H_n|\right] \\ &= E\left[|H_{n+1} - H_n|\right] \to 0, \end{aligned}$$

where the convergence to zero is due to the almost sure convergence of H_n. It then follows that for all $\zeta > 0$, there exists $n_1(\zeta)$ such that

$$\Pr\left[|H_n^- - H_n| \le \zeta\right] \ge 1 - \frac{\epsilon}{4} \quad \text{for all } n \ge n_1(\zeta). \tag{5.10}$$

Now take $\eta = \eta(\delta/2)$ as in Lemma 5.4 and $\mu(\delta/2, \eta)$ as in Lemma 5.1, and let $\zeta = \mu(\delta/2, \eta)$. Then, (5.10) implies that the set

$$\mathcal{M}_n := \left\{ i \colon \left| H(U_i \mid Y_1^N U_1^{i-1})^- - H(U_i \mid Y_1^N U_1^{i-1}) \right| < \mu(\delta/2, \eta) \right\}$$

satisfies

$$\frac{|\mathcal{M}_n|}{N} \geq 1 - \frac{\epsilon}{4} \tag{5.11}$$

for all $n \geq n_1(\mu(\delta/2, \eta))$. We will prove (5.9) by contradiction. To that end, define the set

$$\mathcal{L}_n := \left\{ i \colon H(U_i \mid Y_1^N U_1^{i-1}) \in (\delta, 1-\delta) \right\}$$

and suppose, contrary to (5.9), that there exists $n > n_1(\mu(\delta/2, \eta))$ for which

$$\frac{|\mathcal{L}_n|}{N} \geq \epsilon. \tag{5.12}$$

Define $\lambda := \kappa \log(|\mathcal{X}||\mathcal{Y}|)$ and the sets

$$\begin{aligned}
\mathcal{K}_n &:= \left\{ i \colon I(U_i; V_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}) \leq \sqrt{\lambda/N} \right\}, \\
\mathcal{J}_{n,1} &:= \left\{ i \colon I(U_i; Y_{N+1}^{2N} V_1^{i-1} \mid Y_1^N U_1^{i-1}) \leq \sqrt{\lambda/N} \right\}, \\
\mathcal{J}_{n,2} &:= \left\{ i \colon I(V_i; Y_1^N U_1^{i-1} \mid Y_{N+1}^{2N} V_1^{i-1}) \leq \sqrt{\lambda/N} \right\}, \\
\mathcal{J}_n &:= \mathcal{J}_{n,1} \cap \mathcal{J}_{n,2}.
\end{aligned}$$

Note that for all $N = 2^n$ we have

$$\begin{aligned}
\frac{\lambda}{N} &\geq \frac{1}{N} I(X_1^N Y_1^N; X_{N+1}^{2N} Y_{N+1}^{2N}) \\
&\geq \frac{1}{N} I(X_1^N; X_{N+1}^{2N} Y_{N+1}^{2N} \mid Y_1^N) \\
&= \frac{1}{N} I(U_1^N; Y_{N+1}^{2N} V_1^N \mid Y_1^N) \\
&= \frac{1}{N} \sum_{i=1}^{N} I(U_i; Y_{N+1}^{2N} V_1^N \mid Y_1^N U_1^{i-1}) \\
&\geq \frac{1}{N} \sum_{i=1}^{N} I(U_i; Y_{N+1}^{2N} V_1^i \mid Y_1^N U_1^{i-1}) \\
&\geq \frac{1}{N} \sum_{i=1}^{N} \big[I(U_i; Y_{N+1}^{2N} V_1^{i-1} \mid Y_1^N U_1^{i-1}) \\
&\qquad\qquad + I(U_i; V_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}) \big].
\end{aligned}$$

This in particular implies that

$$\frac{|\mathcal{J}_{n,1} \cap \mathcal{K}_n|}{N} \geq 1 - \sqrt{\lambda/N}.$$

By swapping the U's with the V's above, one also obtains $\frac{|\mathcal{J}_{n,2}|}{N} \geq 1 - \sqrt{\lambda/N}$. Hence,

$$\frac{|\mathcal{J}_n \cap \mathcal{K}_n|}{N} = \frac{|\mathcal{J}_{n,1} \cap \mathcal{J}_{n,2} \cap \mathcal{K}_n|}{N} \geq 1 - 2\sqrt{\lambda/N}. \tag{5.13}$$

Take $n > \max\{n_0(\delta), n_1(\mu(\delta/2, \eta))\}$ (where $n_0(\delta)$ is as in Lemma 5.4) such that

$$\sqrt{\lambda/N} < \frac{\delta}{2}, \tag{5.14}$$

$$2\sqrt{\lambda/N} < \frac{\epsilon}{2}, \tag{5.15}$$

$$\nu\left(\sqrt{\lambda/N}\right) + \sqrt{\lambda/N} \leq \frac{\mu(\delta/2, \eta)}{2}.$$

Observe that for such n and for all $i \in \mathcal{J}_n \cap \mathcal{L}_n$, relation (5.14) implies

$$H(U_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}), H(V_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}) \in (\delta/2, 1 - \delta/2),$$

Now let $\tilde{H}_{u,i}$ be a random variable that takes the value

$$H(U_i \mid Y_1^{2N} = y_1^{2N} U_1^{i-1} V_1^{i-1} = u^{i-1} v^{i-1})$$

whenever $(Y_1^{2N} U_1^{i-1} V_1^{i-1}) = (y_1^{2N} u^{i-1} v^{i-1})$ and define

$$\tilde{S}_{u,i} = \begin{cases} 0 & \text{if } \tilde{H}_{u,i} \in [0, \delta/2) \\ 1 & \text{if } \tilde{H}_{u,i} \in [\delta/2, 1 - \delta/2] \\ 2 & \text{if } \tilde{H}_{u,i} \in (1 - \delta/2, 1] \end{cases}.$$

Also define $\tilde{H}_{v,i}$ and $\tilde{S}_{v,i}$ analogously. It can easily be shown that for $i \in \mathcal{J}_n$, the joint distribution of the pair $(\tilde{S}_{u,i}, \tilde{S}_{v,i})$ approaches that of $(S_{u,i}, S_{v,i})$ (defined in (5.7)) as n grows. It then follows from Lemma 5.4 that for sufficiently large n we have

$$\Pr\left[\tilde{S}_{u,i} = \tilde{S}_{v,i} \in \{0, 2\}\right] < 1 - \eta/2.$$

For such n, and for all $i \in \mathcal{J}_n \cap \mathcal{K}_n \cap \mathcal{L}_n$, it is easily seen that $(Y_1^{2N} U_1^i V_1^i)$, along with $\tilde{S}_{u,i}$ and $\tilde{S}_{v,i}$ satisfy the conditions of Lemma 5.1 with

$$\begin{aligned} X_1 &= U_i, & X_2 &= V_i, & Y &= (Y_1^{2N} U_1^{i-1} V_1^{i-1}), \\ S_1 &= \tilde{S}_{u,i}, & S_2 &= \tilde{S}_{v,i}, & \epsilon_2 &= \sqrt{\lambda/N}, \quad \eta = \eta/2. \end{aligned}$$

Consider now $i \in \mathcal{J}_n \cap \mathcal{K}_n \cap \mathcal{L}_n$. We have

$$\begin{aligned} H(U_i \mid Y_1^N U_1^{i-1})^- - H(U_i \mid Y_1^N U_1^{i-1}) & \\ &= H(U_i + V_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}) - H(U_i \mid U_1^{i-1} Y_1^N) \\ &\geq H(U_i + V_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}) \\ &\qquad - H(U_i \mid Y_1^{2N} U_1^{i-1} V_1^{i-1}) - \sqrt{\lambda/N} \\ &\geq \mu(\delta/2, \eta) - \nu\left(\sqrt{\lambda/N}\right) - \sqrt{\lambda/N} \\ &\geq \frac{\mu(\delta/2, \eta)}{2}, \end{aligned}$$

from which we obtain

$$\mathcal{J}_n \cap \mathcal{K}_n \cap \mathcal{L}_n \cap \mathcal{M}_n = \emptyset.$$

This, in addition to (5.12), (5.13) and (5.15), implies

$$\begin{aligned}\frac{\epsilon}{2} &\leq \frac{|\mathcal{J}_n \cap \mathcal{K}_n \cap \mathcal{L}_n|}{N} \\ &= \frac{|\mathcal{J}_n \cap \mathcal{K}_n \cap \mathcal{L}_n \cap \mathcal{M}_n^c|}{N} \\ &\leq \frac{|\mathcal{M}_n^c|}{N},\end{aligned}$$

which contradicts (5.11), yielding the claim.

5.2.1 Channels with Memory

Unlike in the memoryless case, it may not be immediately obvious how Theorem 5.1 translates to a channel polarization result; memory in discrete channels is typically modeled through a channel state sequence, which is absent from the above model. More precisely, a *finite state channel* is defined through the set of joint probability distributions on its inputs $x_1^N \in \mathcal{X}^N$, outputs $y_1^N \in \mathcal{Y}^N$, and states $s_0^N \in \mathcal{S}^N$ with $|\mathcal{S}| < \infty$,

$$p(y_1^N, x_1^N, s_1^N \mid s_0) = p(x_1^N) \prod_{i=1}^{N} p(y_i, s_i \mid x_i, s_{i-1}),$$

where s_0 is the initial channel state. In certain channel models (e.g., if the channel is *indecomposable* [10]) assigning a finite order ergodic Markov distribution on the input sequence $X_1, X_2, \ldots$ induces a similar distribution on the sequence $(X_1, Y_1, S_1), (X_2, Y_2, S_2), \ldots$. It is then easy—after minor modifications to the proof of Theorem 5.1—to obtain

Theorem 5.2. *Let $U_1^N = G_n(X_1^N)$. Then for all $\delta > 0$,*

$$\lim_{n\to\infty} \frac{1}{N} \left| \left\{ i \colon I(U_i; Y_1^N \mid U_1^{i-1}) \in (\delta, 1-\delta) \right\} \right| = 0.$$

5.3 Discussion

Results in this chapter complement previous polarization theorems (Theorems 2.1 and 3.4), showing that Arıkan-like constructions are fairly universal in distilling randomness, that is, in transforming a stochastic process into a set of uniform random variables and constants. Although the main result is stated only for finite-memory processes, we believe that this restriction is an artifact of our proof technique, and is not necessary for polarization to take place. In fact, one can check that most of the crucial steps in the proofs remain

valid without the finite-memory assumption. We conjecture that Arıkan's construction polarizes all *mixing* processes with finite alphabets.

The practical importance of the result presented here is perhaps less obvious than its theoretical relevance. The main practical appeal of polar codes for memoryless channels and sources is due to (i) their low encoding/decoding complexity, (ii) the ease with which they can be constructed, and (iii) their 'exponentially' decaying error probability. All of these desirable properties owe much to the 'tree structure' in the probability distributions induced by the polarization transform, which breaks down when the underlying process has memory. A problem of interest in this direction is to determine whether previous results on error probability and complexity can be generalized to such processes.

Joint Polarization of Multiple Processes

6

We have by now established that a large class of discrete stochastic processes can be polarized by a large class of recursive procedures. In memoryless cases, these procedures yield low-complexity point-to-point channel codes as well as source codes that achieve optimal rates, i.e., symmetric capacity and source entropy, respectively. Our aim in this chapter is to apply the principles developed so far in order to obtain *joint polarization* results for multiple sequences. In particular, we will consider i.i.d. processes of the form $(W_1, X_1, Y_1), (W_2, X_2, Y_2), \ldots$ where $W_1 \in \mathcal{W}$, $X_1 \in \mathcal{X}$, and $Y_1 \in \mathcal{Y}$ for finite sets $\mathcal{W}$, $\mathcal{X}$ and $\mathcal{Y}$. The joint distribution of (W_1, X_1, Y_1) will be arbitrary.

Polarizing such a process may be understood in several ways. One may for instance ask whether a block (W_1^N, X_1^N) can be transformed such that the result $(U_1^N, V_1^N) \in \mathcal{W}^N \times \mathcal{X}^N$ is polarized in the sense that

$$H(U_i V_i \mid Y_1^N U_1^{i-1} V_1^{i-1}) \approx 0 \text{ or } \approx 1 \text{ for almost all } i\text{'s}. \tag{6.1}$$

If no constraints are imposed on this transformation, then it is indeed easy to attain polarization: In light of the results in Chapter 3, this can be done simply by viewing (W_1, X_1) as a single $\mathcal{W} \times \mathcal{X}$-valued random variable, and using a polarizing transform for the alphabet $\mathcal{W} \times \mathcal{X}$. Naturally, then, such a definition of joint polarization is not very interesting.

In order to obtain a more useful definition, let us first place the underlying process $(W_1, X_1, Y_1), (W_2, X_2, Y_2), \ldots$ in an operational context. As in single source/channel polarization, two simple interpretations are possible:

Separate encoding of correlated sources: In this setting, W_1^N and X_1^N can be viewed as the outputs of two correlated i.i.d. sources, which are observed by separate source encoders. The sequence Y_1^N can be thought of as

side information about the source output, available to the decoder. The output sequences are encoded separately by their respective encoders, and are subsequently estimated by the decoder. It was shown by Slepian and Wolf [13] that the set of all achievable rate pairs (R_W, R_X) in this setup is characterized by the bounds

$$\begin{aligned} R_W &\geq H(W_1 \mid Y_1 X_1) \\ R_X &\geq H(X_1 \mid Y_1 W_1) \\ R_W + R_X &\geq H(W_1 X_1 \mid Y_1). \end{aligned}$$

Optimal points in this region can be achieved by employing a single-source polar code at each encoder. To see how this can be accomplished, consider first a corner point of the above region, with $R_W = H(W_1 \mid Y_1)$ and $R_X = H(X_1 \mid Y_1 W_1)$, and the following scheme:
Encoding: The encoders for W and X each choose a polarizing transform for alphabet sizes $|\mathcal{W}|$ and $|\mathcal{X}|$ respectively and compute the sets

$$\mathcal{A}_W = \{i \colon Z(U_i \mid Y_1^N U_1^{i-1}) \approx 0\}$$

and

$$\mathcal{A}_X = \{i \colon Z(V_i \mid Y_1^N W_1^N V_1^{i-1}) \approx 0\}.$$

Here U_1^N (respectively, V_1^N) is the result of the polarizing transform for W (respectively, X). Upon observing their corresponding source outputs W_1^N and X_1^N, both encoders apply their transforms to obtain U_1^N and V_1^N, and send $U_{\mathcal{A}_W^c}$, and $V_{\mathcal{A}_X^c}$ to the decoder.
Decoding: The decoder first estimates W_1^N from $U_{\mathcal{A}_W^c}$ and Y_1^N using the successive cancellation (SC) decoder for the sequence $(W_1, Y_1), (W_2, Y_2), \ldots$. (That is, it ignores its knowledge of $V_{\mathcal{A}_X^c}$.) It then assumes that its estimate $\hat{W}_1^N$ is correct and therefore that $\hat{W}_1^N$ is identically distributed as W_1^N, and uses the SC decoder for the sequence $(X_1, (Y_1 W_1)), (X_2, (Y_2 W_2)), \ldots$ to estimate X_1^N from $V_{\mathcal{A}_X^c}$ and $(Y_1^N \hat{W}_1^N)$.
Rate: It follows from single-source polarization theorems that $|\mathcal{A}_W^c| \approx NH(W_1 \mid Y_1)$ and $|\mathcal{A}_X^c| \approx NH(X_1 \mid Y_1 W_1)$, i.e., that the above scheme operates approximately at a corner point of the achievable region.
Error probability: A decoding error occurs if at least one of the two constituent SC decoders errs. The probability of this event can be upper bounded by the sum of the error probabilities of each decoder. (The proof of this fact is identical to that of Proposition 2.1.) It follows from previous results that each of these average block error probabilities, and thus also their sum, is approximately $2^{-\sqrt{N}}$.

Multiple-access channel: Recall that the capacity region of a multiple-access channel is the convex hull of

$$\bigcup_{W,X} \mathcal{R}_{W,X}$$

where

$$\mathcal{R}_{W,X} = \Big\{(R_1, R_2)\colon R_W \le I(W;YX) \\ R_X \le I(X;YW) \\ R_W + R_X \le I(WX;Y)\Big\}.$$

Here W and X are independently distributed inputs to the channel, and Y is the output. The sequence $(W_1, X_1, Y_1), (W_2, X_2, Y_2), \ldots$ naturally fits in such a setting. This is best seen by considering the case in which W_1 and X_1 are uniformly and independently distributed inputs to the channel, and Y_1 is the output. The region corresponding to this case is described by the rate bounds

$$\begin{aligned} R_W &\le 1 - H(W_1 \mid Y_1X_1) \\ R_X &\le 1 - H(X_1 \mid Y_1W_1) \\ R_W + R_X &\le 2 - H(W_1X_1 \mid Y_1). \end{aligned} \tag{6.2}$$

Corner points of this region can be achieved by the following coding scheme, which is similar to the one for the source coding case:

Code construction: The encoders for W and X each choose a polarizing transform G_W and G_X for alphabet sizes $|\mathcal{W}|$ and $|\mathcal{X}|$ respectively, and compute the sets

$$\mathcal{A}_W = \{i\colon Z(U_i \mid Y_1^N U_1^{i-1}) \approx 0\}$$

and

$$\mathcal{A}_X = \{i\colon Z(V_i \mid Y_1^N W_1^N V_1^{i-1}) \approx 0\}.$$

where $U_1^N = G_W(W_1^N)$ and $V_1^N = G_X(X_1^N)$ are the respective outputs of these transforms. The senders choose U_i, $i \in \mathcal{A}_W^c$ and V_i, $i \in \mathcal{A}_X^c$ independently and uniformly at random and reveal their values to the receiver.

Encoding: Given uniformly distributed messages $M_W \in \mathcal{W}^{|\mathcal{A}_W|}$ and $M_X \in \mathcal{X}^{|\mathcal{A}_X|}$, the receivers respectively set $U_{\mathcal{A}_W} = M_W$ and $V_{\mathcal{A}_X} = M_X$ and transmit $G_W^{-1}(U_1^N)$ and $G_X^{-1}(V_1^N)$ over the channel.

Decoding: The decoder first decodes $U_{\mathcal{A}_W}$ from $U_{\mathcal{A}_W^c}$ and Y_1^N using the SC decoder for the sequence $(W_1, Y_1), (W_2, Y_2), \ldots$ and produces $\hat{M}_W = G_W(\hat{W}_1^N)$ as its estimate of the message M_W. It then assumes that this estimate is correct, and uses the SC decoder for the sequence $(X_1, (Y_1W_1)), (X_2, (Y_2W_2)), \ldots$ to decode $V_{\mathcal{A}_X}$ from $V_{\mathcal{A}_X^c}$ and $(Y_1^N\hat{W}_1^N)$, and produces $\hat{M}_X = G_X(\hat{X}_1^N)$ as its estimate of M_X.

Rate: It follows from previous results that $|\mathcal{A}_W| \approx N(1 - H(W_1 \mid Y_1))$ and $|\mathcal{A}_X| \approx N(1 - H(X_1 \mid Y_1W_1))$, i.e., that the above scheme operates near a corner point of the region given in (6.2).

Error probability: The block error probability is as in the source coding case, i.e., $\approx 2^{-\sqrt{N}}$ averaged over all message pairs and all pairs of frozen vectors U_i, $i \in \mathcal{A}_X^c$ and V_i, $i \in \mathcal{A}_X^c$. It thus follows that there exists at least one frozen vector pair for which the average block error probability is $\approx 2^{-\sqrt{N}}$.

The above coding schemes are obtained by reducing the corresponding multi-user problem into two single-user problems, for which devising polar coding schemes is easy. Arbitrary points in the achievable rate region in each problem can be achieved via the 'rate splitting' technique of [14]. (In the multiple-access problem, one can also use the technique discussed in Section 3.3 to achieve rate regions with non-uniform inputs.) Clearly, these schemes can be generalized to settings with more than two users. They also yield an alternative polar coding method for single-sources and point-to-point channels when the source/channel-input alphabet size is a composite number. To see this, consider the sequence $(X_1, Y_1), (X_2, Y_2), \ldots$ with $X_1 \in \mathcal{X}$ and $|\mathcal{X}| = q_1 \cdot q_2 \ldots \cdot q_k$. To polarize $X_1, X_2, \ldots$, one may—instead of applying a polarizing transform for the alphabet $\mathcal{X}$ directly—view X_1 as a collection of random variables $(X^{(1)}, \ldots, X^{(k)})$ taking values in $\mathcal{X}^{(1)} \times \ldots \times \mathcal{X}^{(k)}$, with $|\mathcal{X}^{(i)}| = q_i$. This decomposition can be made in an arbitrary manner. Considering the expansion

$$\begin{aligned} H(X_1 \mid Y_1) &= H(X_1^{(1)}, \ldots, X_1^{(k)} \mid Y_1) \\ &= H(X_1^{(1)} \mid Y_1) + \ldots + H(X_1^{(k)} \mid Y_1, X_1^{(1)}, \ldots, X_1^{(k-1)}), \end{aligned}$$

one easily sees that long blocks of each component $X^{(i)}$ can be polarized separately as above, and can then be decoded in the order $X^{(1)}$, $X^{(2)}$,...,$X^{(k)}$, using the appropriate SC decoder in each step. Such a scheme also achieves optimal rates in both channel and source coding, with error probabilities comparable to those of direct polarization schemes.

Our aim here is not just to find polar coding schemes for multi-user settings. Instead, we would also like to know whether one can polarize multiple processes jointly in the sense that (a) polarization is achieved by applying a separate transform to the underlying sequences, and that (b) the resulting random variables ((U_i, V_i) above) are extremal conditioned on their past (U_1^{i-1}, V_1^{i-1}), in the sense that they consist only of deterministic and/or uniformly random parts. Observe that our first definition of joint polarization in (6.1) meets requirement (b) but not (a), since a polarizing transform for a single sequence may not necessarily be decomposed into two separate transforms on the constituent sequences. On the other hand, the second polarization method we discussed does meet (a), as it achieves polarization through separately applying a transform to each sequence. However, it is not clear at this point that it meets requirement (b), since the joint distributions $p_{U_i V_i | Y_1^N U_1^{i-1} V_1^{i-1}}$ one obtains by this method may not be extremal. (We will see that they indeed are.)

This aim can be motivated analogously to single source/channel polarization: In the single-user case, an extremal channel is one whose input is either determined by or independent of its output. In a multi-user setting, a channel may be called extremal if this property holds for all of its inputs: Some are determined by the output, others are independent of it. In the two-user case, this is equivalent to saying that an extremal channel (or equivalently, an extremal

joint source) is one for which the entropies $H(W_1 \mid Y_1X_1)$ and $H(X_1 \mid Y_1W_1)$ are $\{0,1\}$-valued, and $H(W_1X_1 \mid Y_1)$ is $\{0,1,2\}$-valued. It can easily be seen that there are five possible extremal channels/sources with these properties, the rate regions (6.2) associated with such channels are depicted in Figure 6.1. It is also easily seen that reliable communication over extremal channels is trivial, as in the single-user case. Our aim is to polarize several copies of a mediocre multiple-access channel (respectively, joint source) to a set of extremal ones, thereby simplifying the transmission (respectively, compression) task.

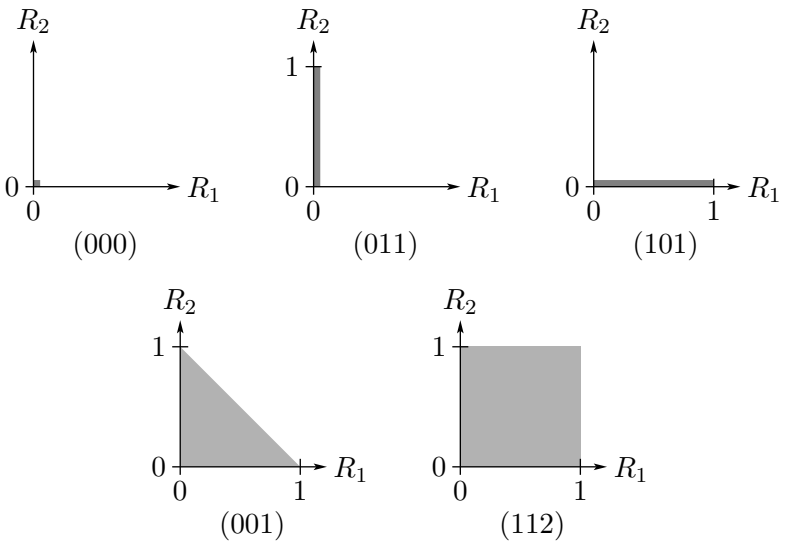

Figure 6.1: Rate regions of the extremal multiple-access channels (Achievable source coding rate regions for extremal sources are analogous to these.) (000) is a channel whose inputs are independent from its output, (011) and (101) are channels in which one input is determined by the output and the other is independent from it, (001) is one in which either of the inputs, but not both, can be determined from the output, and (112) is a noiseless multiple-access channel whose inputs are functions of the output.

6.1 Joint Polarization

Consider an i.i.d. process $(W_1, X_1, Y_1), (W_2, X_2, Y_2), \ldots$ as above. For notational convenience, we will assume in this section that $\mathcal{W} = \mathcal{X}$ and later discuss how the results here apply to processes with different alphabet sizes. We will be interested in determining how the entropies

$$\begin{aligned} H[1] &:= H(W_1 \mid Y_1X_1) \\ H[2] &:= H(X_1 \mid Y_1W_1) \\ H[12] &:= H(W_1X_1 \mid Y_1), \end{aligned}$$

which define the achievable rate regions evolve in the course of a joint polarization process. For this purpose, we first choose a polarizing mapping, which we will denote by the generic symbol '+', and apply it separately to (W_1, W_2) and (X_1, X_2) to obtain

$$\begin{aligned} U_1 &= W_1 + W_2 & V_1 &= X_1 + X_2 \\ U_2 &= W_2 & V_2 &= X_2. \end{aligned}$$

We also set the following shorthand notation for the resulting entropy terms of interest

$$\begin{aligned} H^b[1] &:= H(U_1 \mid Y_1^2 V_1) & H^g[1] &:= H(U_2 \mid Y_1^2 U_1 V_1 V_2) \\ H^b[2] &:= H(V_1 \mid Y_1^2 U_1) & H^g[2] &:= H(V_2 \mid Y_1^2 U_1 V_1 U_2) \\ H^b[12] &:= H(U_1 V_1 \mid Y_1^2) & H^g[12] &:= H(U_2 V_2 \mid Y_1^2 U_1 V_1) \end{aligned}$$

If one applies this transform to both sequences recursively in the usual manner, one obtains after n recursions $U_1^N = G_N(W_1^N)$ and $V_1^N = G_N(X_1^N)$, where again $N = 2^n$ and G_N represents n recursions of the polarizing transform. Our aim is to show that the resulting random variable triples $(U_i, V_i, (Y_1^N U_1^{i-1} V_1^{i-1}))$ are polarized in the sense that for all $\epsilon > 0$, we have

$$\begin{aligned} H^{(i)}[1] &:= H(U_i \mid Y_1^N U_1^{i-1} V_1^{i-1} V_i) \notin (\epsilon, 1-\epsilon) \\ H^{(i)}[2] &:= H(V_i \mid Y_1^N U_1^{i-1} V_1^{i-1} U_i) \notin (\epsilon, 1-\epsilon) \\ H^{(i)}[12] &:= H(U_i V_i \mid Y_1^N U_1^{i-1} V_1^{i-1}) \notin (\epsilon, 1-\epsilon) \cup (1+\epsilon, 2-\epsilon), \end{aligned} \tag{6.3}$$

for almost all $i \in \{1, \ldots, N\}$, provided that N is sufficiently large. This is equivalent to saying that the entropy triples $\big(H^{(i)}[1], H^{(i)}[2], H^{(i)}[12]\big)$ for almost all i's is close to one of the five extremal values

$$(0,0,0), \quad (0,1,1), \quad (1,0,1), \quad (0,0,1), \quad (1,1,2).$$

As in the previous chapters, the main ingredient of the proof of this polarization statement is a result on the single-step evolution of entropies $H[1]$, $H[2]$, and $H[12]$:

Lemma 6.1. *For every $\epsilon > 0$, there exists $\delta > 0$ such that*

$$H^b[12] - H[12] \leq \delta$$

implies

(i) $H^b[1] - H[1] \leq \delta$ *and* $H^b[2] - H[2] \leq \delta$,

(ii) $H[1], H[2] \notin (\epsilon, 1-\epsilon)$,

(iii) $H[12] \notin (2\epsilon, 1-\epsilon) \cup (1+\epsilon, 2-2\epsilon)$.

Proof. We have

$$\begin{aligned}\delta &\geq H^b[12] - H[12] \\ &= H(W_1 + W_2, X_1 + X_2 \mid Y_1^2) - H(W_1X_1 \mid Y_1) \\ &= H(W_1 + W_2 \mid Y_1^2) - H(W_1 \mid Y_1) \\ &\quad + H(X_1 + X_2 \mid Y_1^2, W_1 + W_2) - H(X_1 \mid Y_1W_1)\end{aligned} \tag{6.4}$$

Note that both entropy differences in (6.4) are non-negative, and thus are at most δ, implying $H^b[2] - H[2] \leq \delta$. Swapping the W's and the X's in the above relations also yields $H^b[1] - H[1] \leq \delta$, proving (i). One can continue (6.4) as

$$\begin{aligned}\delta \geq\ & H(W_1 + W_2 \mid Y_1^2) - H(W_1 \mid Y_1) \\ & + H(X_1 + X_2 \mid Y_1^2W_1^2) - H(X_1 \mid Y_1W_1).\end{aligned} \tag{6.5}$$

For sufficiently small δ, it follows from (6.5) and Theorem 3.3 that $H(W_1 \mid Y_1) \notin (\epsilon, 1-\epsilon)$, and $H(X_1 \mid Y_1W_1) = H[2] \notin (\epsilon, 1-\epsilon)$. Further, since

$$H(W_1X_1 \mid Y_1) = H(W_1 \mid Y_1) + H(X_1 \mid Y_1W_1),$$

it follows that $H(W_1X_1 \mid Y_1) = H[12] \notin (2\epsilon, 1-\epsilon) \cup (1+\epsilon, 2-2\epsilon)$, yielding (iii). By swapping the X's with the W's in the above chain of inequalities one also obtains $H(X_1 \mid Y_1) \notin (\epsilon, 1-\epsilon)$ and $H(W_1 \mid Y_1X_1) = H[1] \notin (\epsilon, 1-\epsilon)$, completing the proof. □

This lemma suffices to show the main polarization result of this chapter, which was also obtained independently by Arıkan (in an unpublished version of [7]).

Theorem 6.1. *Let* $M := \{(0,0,0), (0,1,1), (1,0,1), (0,0,1), (1,1,2)\}$*, and*

$$d(a, M) := \max_{b \in M} \|a - b\|, \qquad a \in \mathbb{R}^3.$$

For all $\epsilon > 0$*, we have*

$$\lim_{n\to\infty} \frac{1}{N} \left| \left\{ i \colon d\big((H^{(i)}[1], H^{(i)}[2], H^{(i)}[12]), M\big) \geq \epsilon \right\} \right| = 0.$$

Proof. The proof is similar to those of previous polarization theorems: Let $B_1, B_2, \ldots$ be an i.i.d. process with $\Pr[B_1 = b] = \Pr[B_1 = g] = 1/2$. Define a process $(H_0[1], H_0[2], H_0[12]), (H_1[1], H_1[2], H_1[12]), \ldots$ with

$$\begin{aligned}H_0[k] &= H[k], \\ H_n[k] &= H_{n-1}^{B_n}[k], \quad n = 1, 2, \ldots\end{aligned}$$

for $k = 1, 2, 12$. Observe that

$$\begin{aligned}H^b[12] + H^g[12] &= H(U_1V_1 \mid Y_1^2) + H(U_2V_2 \mid Y_1^2U_1V_1) \\ &= H(W_1^2X_1^2 \mid Y_1^2) \\ &= 2H[12],\end{aligned}$$

therefore the process $H_0[12], H_1[12], \ldots$ is a bounded martingale and converges almost surely to a $[0,2]$-valued random variable $H_\infty[12]$. It then follows from (i) in Lemma 6.1 that processes $H_0[1], H_1[1], \ldots$ and $H_0[2], H_1[2], \ldots$ also converge almost surely to $[0,1]$-valued random variables $H_\infty[1]$ and $H_\infty[2]$, respectively. It further follows from (ii) in Lemma 6.1 that $H_\infty[1]$ and $H_\infty[2]$ are $\{0,1\}$-valued, and from (iii) that $H_\infty[12]$ is $\{0,1,2\}$-valued, i.e., that the process $(H_0[1], H_0[2], H_0[12]), (H_1[1], H_1[2], H_1[12]), \ldots$ converges almost surely to a random vector taking values in the set M. The claim then follows from the equivalence between the probability distribution of $(H_n[1], H_n[2], H_n[12])$ and the distribution of $(H^{(i)}[1], H^{(i)}[2], H^{(i)}[12])$, $i = 1, \ldots, N$. □

6.1.1 Rate Region

We have seen that separately applying a polarizing transformation to two i.i.d. processes polarizes them jointly, i.e., the resulting joint distributions approach one of five extremal distributions as the construction size grows. We now consider the rate region obtained by this procedure. We will discuss the multiple-access channel interpretation of the result.

Let $\mathcal{R}$ denote the rate region defined by the bounds in (6.2). Also let $\mathcal{R}^b$ and $\mathcal{R}^g$ denote the rate regions obtained after the first polarization step, i.e., those with entropies $(H[1], H[2], H[12])$ in (6.2) replaced respectively by $(H^b[1], H^b[2], H^b[12])$ and $(H^g[1], H^g[2], H^g[12])$. One can similarly define the regions $\mathcal{R}^{\mathbf{s}}$, $\mathbf{s} \in \{b,g\}^n$ obtained after n polarization steps. Note that

$$\begin{aligned} 2H[1] &= H(W_1^2 \mid Y_1^2 X_1^2) \\ &= H(U_1^2 \mid Y_1^2 V_1^2) \\ &\leq H(U_1 \mid Y_1^2 V_1) + H(U_2 \mid Y_1^2 U_1 V_1 V_2) \\ &= H^b[1] + H^g[1]. \end{aligned}$$

It similarly follows that

$$\begin{aligned} 2H[2] &\leq H^b[2] + H^g[2] \\ 2H[12] &= H^b[12] + H^g[12], \end{aligned} \tag{6.6}$$

and therefore the set

$$\tfrac{1}{2}\mathcal{R}^b + \tfrac{1}{2}\mathcal{R}^g = \left\{ \tfrac{1}{2}a + \tfrac{1}{2}b \colon a \in \mathcal{R}^b, b \in \mathcal{R}^g \right\}$$

is a subset of $\mathcal{R}$. It is easy to find examples where this inclusion is strict. Nevertheless, due to equality in (6.6) and the polymatroidal nature of $\mathcal{R}$, $\frac{1}{2}\mathcal{R}^b + \frac{1}{2}\mathcal{R}^g$ and $\mathcal{R}$ share points on their dominant faces (see Figure 6.2). Polarizing the resulting regions $\mathcal{R}^b$ and $\mathcal{R}^g$ further will similarly lead to a loss of overall rate region, i.e., for all n

$$\frac{1}{N} \sum_{\mathbf{s} \in \{b,g\}^n} \mathcal{R}^{\mathbf{s}} \subset \mathcal{R}$$

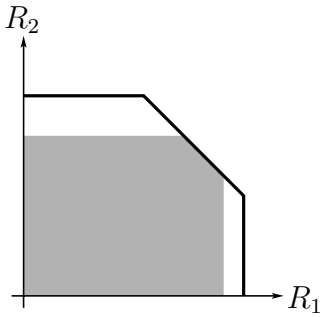

Figure 6.2: The average of the rate regions after n polarization steps (the shaded region) is a subset of the original region, but contains points on the dominant face of the latter.

although the regions on either side of the last relation will share at least one point on their dominant faces. Note that the situation here is in contrast with point-to-point channel polarization, where no rate penalty is incurred by the construction.

6.1.2 Processes with Different Alphabet Sizes

We have so far assumed that the processes we polarize jointly have identical alphabet sizes. However, this restriction is only for notational convenience, and is not necessary for polarization to take place. It can indeed be seen easily that the proofs given above are equally valid when the alphabet sizes of the processes differ, and the resulting random variables are still either uniformly random or deterministic. If one computes entropies with base-$|\mathcal{W}||\mathcal{X}|$ logarithms, then the extremal values for $(H[1], H[2], H[12])$ become

$$(0,0,0), \quad (0, \log|\mathcal{X}|, \log|\mathcal{X}|) \quad (\log|\mathcal{W}|, 0, \log|\mathcal{W}|), \quad (\log|\mathcal{W}|, \log|\mathcal{X}|, 1),$$

corresponding respectively to the previous cases $(000), (011), (101), (112)$. The case (001) is precluded from this setting. To see the reason for this, suppose that random variables (W, X, Y) with $|\mathcal{W}| < |\mathcal{X}|$ satisfy the conditions of the case (001): X is uniformly distributed conditioned on Y, but is a function of (W, Y), i.e., $H(X \mid Y) = \log|\mathcal{X}|$ and $H(X \mid YW) = 0$. This would imply $I(W; X \mid Y) = \log|\mathcal{X}|$, an impossibility since $I(W; X \mid Y) \leq \log|\mathcal{W}|$. Consequently, the rate region obtained by polarization is rectangular (i.e., it has a single point on the dominant face of the original region) when the alphabet sizes differ.

6.2 Rate of Polarization

Our purpose in this section is to give operational meaning to the rate region obtained after polarization. We will do so by describing a channel coding scheme

that achieves the corresponding rate region—the source coding counterpart is similar. We will restrict our attention to processes with prime alphabet sizes, and will assume that the polarizing mapping '+' for each alphabet is the corresponding modulo-addition operation.

Suppose $W_1, W_2, \dots$ and $X_1, X_2, \dots$ are i.i.d., uniformly distributed inputs to a multiple-access channel, and $Y_1, Y_2, \dots$ is the output. Let G_X and G_W be two polarizing transforms as above, and $U_1^N = G_W(W_1^N)$, $V_1^N = G_X(X_1^N)$ their outputs. Fix $\epsilon > 0$, and define the set

$$\mathcal{P}_\epsilon(a,b,c) := \left\{ i \colon \left\| (H^{(i)}[1], H^{(i)}[2], H^{(i)}[12]) - (a,b,c) \right\| < \epsilon \right\}$$

for $(a,b,c) \in \mathbb{R}^3$. Let $\mathcal{A}_W$, $\mathcal{A}_X \subset \{1,\dots,N\}$ denote sets of indices over which the users transmit their data, and choose these sets as follows:

(i.a) If $i \in \mathcal{P}_\epsilon(0,0,0)$, then set $i \in \mathcal{A}_W, i \in \mathcal{A}_X$,

(i.b) else if $i \in \mathcal{P}_\epsilon(0,1,1)$, then set $i \in \mathcal{A}_W, i \notin \mathcal{A}_X$,

(i.c) else if $i \in \mathcal{P}_\epsilon(1,0,1)$, then set $i \notin \mathcal{A}_W, i \in \mathcal{A}_X$,

(i.d) else if $i \in \mathcal{P}_\epsilon(0,0,1)$, then set either $i \in \mathcal{A}_W, i \notin \mathcal{A}_X$ or $i \notin \mathcal{A}_W, i \in \mathcal{A}_X$,

(ii) else, set $i \notin \mathcal{A}_W, i \notin \mathcal{A}_X$.

The senders set $U_i, i \in \mathcal{A}_W$ and $V_i, i \in \mathcal{A}_X$ to be the uniformly distributed data symbols. Symbols in $\mathcal{A}_W^c$ and $\mathcal{A}_X^c$ are frozen, i.e., they are chosen uniformly at random and revealed to the receiver. It follows from previous results that for all $\delta > 0$ there exists N_0 such that $|\mathcal{A}_W| + |\mathcal{A}_X| > N(2 - H(W_1X_1 \mid Y_1))$ for all $N \geq N_0$, i.e., that the operating point of this scheme is close to the dominant face of the original region. The whole dominant face of the region obtained by polarization can be spanned by varying the sizes of the data sets $\mathcal{A}_W$ and $\mathcal{A}_X$ through (i.d).

Decoding is performed successively as in the single-user case, in the order $(U_1,V_1),(U_2,V_2),\dots(U_N,V_N)$: In decoding (U_i,V_i) the receiver first sets the frozen symbol (if there is one), say U_i, to its known value, and decodes V_i using the optimal decision rule for the channel $V_i \to Y_1^N U_1^{i-1} V_1^{i-1} U_i$. If neither U_i nor V_i is frozen, then they are decoded in an arbitrary order, also using the optimal decision rules for the corresponding channels. Since these channels have the same recursive structure as in the single-user case, the complexity of the described decoding operation is $O(N \log N)$. The error probability of this scheme can similarly be bounded by those of the resulting channels:

$$\begin{aligned} P_e \leq & \sum_{i \in \mathcal{P}_\epsilon(0,0,0)} \left[Z(U_i \mid Y_1^N U_1^{i-1} V_1^{i-1}) + Z(V_i \mid Y_1^N U_1^{i-1} V_1^{i-1}) \right] \\ & + \sum_{i \in \mathcal{P}_\epsilon(0,1,1)} Z(U_i \mid Y_1^N U_1^{i-1} V_1^{i-1}) + \sum_{i \in \mathcal{P}_\epsilon(1,0,1)} Z(V_i \mid Y_1^N U_1^{i-1} V_1^{i-1}) \\ & + \sum_{i \in \mathcal{P}_\epsilon(0,0,1)} \max \left\{ Z(U_i \mid Y_1^N U_1^{i-1} V_1^{i-1} V_i), Z(V_i \mid Y_1^N U_1^{i-1} V_1^{i-1} U_i) \right\} \end{aligned}$$

Note that the Bhattacharyya parameters on the first two lines of the above sum are larger than those of the corresponding channels, since they each ignore the knowledge of one symbol (U_i or V_i) available at the output. We will see that this relaxation greatly simplifies error probability proofs. In particular, we will see that almost all Bhattacharyya parameters in the above sum are 'exponentially small', and therefore one can make the sum vanish by freezing a negligible fraction of the data symbols in both codes:

Lemma 6.2. *Define* $Z'(A \mid B) := Z(A \mid Y_1^N U_1^{i-1} V_1^{i-1} B)$. *There exists an* $\epsilon > 0$ *such that for all* $\beta < 1/2$,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{N} \left| \left\{ i \in \mathcal{P}_\epsilon(0,0,0) \colon Z'(U_i) + Z'(V_i) \geq 2^{-N^\beta} \right\} \right| = 0, \\
&\lim_{n\to\infty} \frac{1}{N} \left| \left\{ i \in \mathcal{P}_\epsilon(0,1,1) \colon Z'(U_i) \geq 2^{-N^\beta} \right\} \right| = 0, \\
&\lim_{n\to\infty} \frac{1}{N} \left| \left\{ i \in \mathcal{P}_\epsilon(1,0,1) \colon Z'(V_i) \geq 2^{-N^\beta} \right\} \right| = 0, \\
&\lim_{n\to\infty} \frac{1}{N} \left| \left\{ i \in \mathcal{P}_\epsilon(0,0,1) \colon \max\left\{ Z'(U_i \mid V_i), Z'(V_i \mid U_i) \right\} \geq 2^{-N^\beta} \right\} \right| = 0.
\end{aligned}$$

Proof. It is easy to see that

(i) $i \in \mathcal{P}_\epsilon(0,0,0)$ implies $Z'(U_i), Z'(V_i) \leq \delta(\epsilon)$,

(ii) $i \in \mathcal{P}_\epsilon(0,1,1)$ implies $Z'(U_i) \leq \delta(\epsilon)$,

(iii) $i \in \mathcal{P}_\epsilon(1,0,1)$ implies $Z'(V_i) \leq \delta(\epsilon)$,

(iv) $i \in \mathcal{P}_\epsilon(0,0,1)$ implies $Z'(U_i \mid V_i), Z'(V_i \mid U_i) \leq \delta(\epsilon)$,

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Therefore, the proof will be complete once we show that whenever the above Bhattacharyya parameters are close to 0, they are exponentially small in the square root of the blocklength. For this purpose, we will define stochastic processes that mirror the behavior of the Bhattacharyya parameters of interest, in the now-customary manner: We first define the Bhattacharyya parameters

$$\begin{aligned}
Z^b(W_1 \mid Y_1) &:= Z(W_1 + W_2 \mid Y_1^2) \\
Z^g(W_1 \mid Y_1) &:= Z(W_2 \mid Y_1^2, W_1 + W_2, X_1 + X_2),
\end{aligned}$$

obtained from $Z(W_1 \mid Y_1)$ after the first polarization step. Also define an i.i.d. process $B_1, B_2, \ldots$ with $\Pr[B_1 = g] = \Pr[B_1 = b] = 1/2$, and the processes

$$\begin{aligned}
Z_0 &= Z(W_1 \mid Y_1) \\
Z_n &= Z_{n-1}^{B_n}, \quad n = 1, 2, \ldots
\end{aligned}$$

It suffices to characterize the one-step evolution of the Bhattacharyya parameters, the rest of the proof being identical to previous ones (e.g., Theorem 3.2): Observe that

$$\begin{aligned} Z^b(W_1 \mid Y_1) &= Z(W_1 \mid Y_1)^- \\ Z^g(W_1 \mid Y_1) &\leq Z(W_2 \mid Y_1^2, W_1 + W_2) = Z(W_1 \mid Y_1)^+, \end{aligned}$$

where Z^- and Z^+ are defined as in the single-user case. Consequently, whenever Z_n converges to 0, it does so at least as fast as in single-user polarization. That is, whenever $Z'(U_i)$ is close to 0, it is almost surely exponentially small in the square root of the blocklength. By symmetry, a similar statement also holds for $Z'(V_i)$. This yields the first three claims.

The last claim is trivial when $|\mathcal{W}| \neq |\mathcal{X}|$, since we then have

$$\lim_{n\to\infty} \tfrac{1}{N} |\mathcal{P}_\epsilon(0, 0, 1)| = 0.$$

(See Section 6.1.2.) For the case $|\mathcal{W}| = |\mathcal{X}|$, we will prove that the claimed rate of convergence holds for the Bhattacharyya parameter $Z'(U_i + \alpha V_i)$, for some $\alpha \in \mathcal{W} \in \backslash\{0\}$ from which the result will follow since

$$Z'(U_i \mid V_i) = Z'(U_i + \alpha V_i \mid V_i) \leq Z'(U_i + \alpha V_i).$$

Consider the one-step evolution of the entropy $H(W_1 + \alpha X_1 \mid Y_1)$. We have

$$\begin{aligned} H^b(W_1 + \alpha X_1 \mid Y_1) &:= H\big((W_1 + \alpha X_1) + (X_2 + \alpha W_2) \mid Y_1^2\big) \\ &= H(W_1 + \alpha X_1 \mid Y_1)^- \end{aligned}$$

and

$$\begin{aligned} H^g(W_1 + \alpha X_1 \mid Y_1) &:= H(W_2 + \alpha X_2 \mid Y_1^2, W_1 + W_2, X_1 + X_2) \\ &\leq H\big(W_1 + \alpha X_1 \mid Y_1^2, (W_1 + W_2) + \alpha(X_1 + X_2)\big) \\ &= H\big(W_1 + \alpha X_1 \mid Y_1^2, (W_1 + \alpha X_1) + (W_2 + \alpha X_2)\big) \\ &= H(W_1 + \alpha X_1 \mid Y_1)^+. \end{aligned}$$

If one defines an entropy process $H_0, H_1, \ldots$ that tracks the evolution of $H(W_1 + \alpha X_1 \mid Y_1)$ in the course of the polarization procedure, then it can be shown using the above relations that $H_0, H_1, \ldots$ is a supermartingale and converges almost surely to a $\{0, 1\}$-valued random variable. Moreover, it is easily seen that the above chain of relations also holds with entropies replaced by the Bhattacharyya parameters, and thus we have

$$\begin{aligned} Z^b(W_1 + \alpha X_1 \mid Y_1) &= Z(W_1 + \alpha X_1 \mid Y_1)^- \\ Z^g(W_1 + \alpha X_1 \mid Y_1) &\leq Z(W_1 + \alpha X_1 \mid Y_1)^+. \end{aligned}$$

Defining once again a Bhattacharyya process $Z_0, Z_1, \ldots$ in the usual manner, it follows that whenever Z_n converges to 0, it does so at least as fast as in the

single-user case. It further follows from Lemma 6.3 in Appendix 6.A that for sufficiently large N,

$$i \in \mathcal{P}_\epsilon(0,0,1) \text{ implies } Z'(U_i + \alpha\, V_i) \le \delta(\epsilon) \text{ for some } \alpha \in \mathcal{W}\backslash\{0\},$$

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. We therefore have,

$$\lim_{n\to\infty} \frac{1}{N}\Big\{ i \in \mathcal{P}_\epsilon(0,0,1)\colon Z'(U_i + \alpha V_i) \ge 2^{-N^\beta} \Big\} = 0$$

for sufficiently small $\epsilon > 0$ and all $\beta < 1/2$, completing the proof. □

Corollary 6.1. *The average block error probability of the coding scheme described above is $o(2^{-N^\beta})$ for all $\beta < 1/2$.*

6.A Appendix

Lemma 6.3. *Let W, X, Y be random variables with $W, X \in \mathcal{W} = \mathbb{F}_q$. There exists $\delta > 0$ such that*

(i) $H(W \mid Y) > 1-\delta$, $H(X \mid Y) > 1-\delta$, $H(W \mid YX) < \delta$, $H(X \mid YW) < \delta$ and

(ii) $H(W + \alpha X \mid Y) \notin (\delta, 1-\delta)$ for all $\alpha \in \mathcal{W}\backslash\{0\}$,

imply

$$H(W + \alpha' X \mid Y) < \delta$$

for some $\alpha' \in \mathcal{W}$.

Proof. Let π be a permutation on $\mathcal{W}$, and let

$$p_\pi(w,x) = \begin{cases} \frac{1}{q} & \text{if } w = \pi(x) \\ 0 & \text{otherwise} \end{cases}.$$

Note that $H(W) = H(X) = 1$ and $H(W \mid X) = H(X \mid W) = 0$ whenever the joint distribution of (W, X) is p_π. We claim that for every π, there exists an $\alpha_\pi \in \mathcal{W}\backslash\{0\}$ such that

$$H(W + \alpha_\pi X) < 1 - c(q),$$

where $c(q) > 0$ depends only on q. To see this, given a permutation π, let

$$\alpha_\pi := \pi(0) - \pi(1) \tag{6.7}$$

Clearly, $\alpha_\pi \neq 0$. It is also easy to check that with these definitions we have

$$\Pr[W + \alpha_\pi X = \pi(0)] \ge \Pr[(W,X) = (\pi(0),0)] + \Pr[(W,X) = (\pi(1),1)] = \tfrac{2}{q},$$

which yields the claim. It also follows from the continuity of entropy in the L_1 metric that

$$\|p_{WX} - p_\pi\| \le o(\delta) \quad \text{implies} \quad H(W + \alpha_\pi X) \le 1 - c(q) + o(\delta).$$

We claim that the conditions of the lemma imply that with high probability (on Y) the distance

$$\|p_{WX|Y=y} - p_\pi\| \text{ is small for some } \pi. \tag{6.8}$$

Note first that

$$\begin{aligned}\delta > 1 - H(W \mid Y) &= \sum_y p(y)[1 - H(W \mid Y = y)] \\ &= \sum_y p(y) D\big(p_{W|Y=y} \| \mathrm{uni}(\mathcal{W})\big) \\ &\ge \sum_y p(y) \tfrac{1}{2} \|p_{W|Y=y} - \mathrm{uni}(\mathcal{W})\|^2,\end{aligned}$$

where the last relation is a consequence of Pinsker's inequality. It then follows that the set

$$G = \big\{y : \|p_{W|Y=y} - \mathrm{uni}(\mathcal{W})\| < \delta^{1/4}\big\}$$

has probability at least $1 - 2\delta^{1/4}$. Further, as

$$\delta > H(X \mid WY) = \sum_y p_Y(y) H(X \mid W, Y = y),$$

the set $B = \{y\colon H(X \mid W, Y = y) \le \sqrt{\delta}\}$ has probability at least $1 - \sqrt{\delta}$. Hence, set $S = G \cap B$ has probability at least $1 - 2\delta^{1/4} - \sqrt{\delta}$. Note that for all $y \in S$ we have for any w, $|\frac{1}{q} - p_{W|Y=y}(w)| < o(\delta)$, and $p_{X|WY}(x \mid w, y) \notin (o(\delta), 1 - o(\delta))$, and thus

$$\min_\pi \|p_{WX|Y=y} - p_\pi\| < o(\delta),$$

yielding the claim in (6.8). In particular, this implies that there exist π' and $S' \subset S$ with $p_Y(S') \ge p_Y(S)/q!$ such that

$$\|p_{WX|Y=y} - p_{\pi'}\| < o(\delta)$$

for all $y \in S'$. Choosing $\alpha' = \alpha_{\pi'}$ as in (6.7), we obtain

$$\begin{aligned}H(W + \alpha' X \mid Y) &\le p_Y(S')(1 - c(q) + o(\delta)) + p_Y(S'^c) \\ &= 1 - c_2 + o(\delta)\end{aligned}$$

where $c_2 > 0$ depends only on q. Since $H(W + \alpha' X \mid Y) \notin (\delta, 1 - \delta)$ by assumption, and we see that if δ is sufficiently small, then $H(W + \alpha' X \mid Y) \le \delta$. □

7 Conclusion

We conclude with a summary of our results and complementary remarks:

In Chapter 3, we showed that discrete memoryless processes with prime alphabet sizes can be polarized by a recursive linear transform similar to the original one for binary processes. We saw that linear transforms fail to polarize all memoryless processes with composite alphabet sizes. We then demonstrated a family of non-linear transforms that polarize stationary memoryless processes with arbitrary discrete alphabets. The crucial property of all basic polarizing transforms is their ability to create a high- and a low-entropy random variable out of two moderate-entropy ones, irrespective of the distribution of the latter. We also derived 'exponential' error probability bounds for channel codes (respectively, source codes) based on the proposed transforms, establishing their capacity-achieving (respectively, entropy-achieving) properties. Let us note that since the results here hold for codes on all discrete alphabets, one can approach the capacity of any memoryless channel with continuous inputs by approximating its capacity-achieving input distribution through the method discussed in Section 3.3.

In Chapter 4 we first showed that processes with prime alphabet sizes can be polarized by any linear transform whose matrix representation is not upper-triangular. This also implies that given any invertible and non-trivial transform, one can find a decoding order (i.e., a permutation of the columns of the transform) under which the resulting random variables are polarized. We observed that the exponential error probability behavior of recursive polar codes is closely related to the distance properties of a single recursion. We derived a simple formula that characterizes this behavior. Although we only provided upper bounds on the error probability in terms of this formula, one can in fact show that the minimum distance behavior of polar codes is given by the same formula, and conclude that successive cancellation decoding of

polar codes achieves optimal performance in the exponential sense. We also saw that the error probability improvements afforded by general constructions over Arıkan's original construction is significant especially for larger alphabet sizes. One should note, however, that our results on the error probability are asymptotic, as are the results in [4], and are not very informative about the performance of short polar codes. Two problems of interest in this direction are to determine whether generalized transforms yield stronger codes at practically relevant lengths, and to determine whether reliability gains can be attained by using non-binary polar codes over binary channels. To that end, one can use a generalized version of the algorithm given in [5] to evaluate the performance of various polar code constructions on various channels, although it is also of interest to develop a theory of polar code design for practically relevant blocklengths.

In Chapter 5, we extended the polarization results of the previous chapters to processes with memory, showing that such processes can be polarized by recursive transforms that polarize memoryless processes. These results are perhaps most relevant in the context of robustness of polar codes against memory in the channel or source, as memorylessness assumptions are used heavily in the proofs of polarization in Chapters 2–4 and it is not immediately clear whether these proofs hold without it. Crucial to our proof in Chapter 5 is the observation that a polarization transformation creates 'almost memoryless' distributions after sufficiently many recursions. One should note that the results here do not immediately lead to practical coding theorems by themselves, and need to be complemented by error probability bounds and low-complexity decoding algorithms. These results should therefore be seen as a first step toward showing the robustness of polar coding against memory. A natural next step in this direction is to investigate how a memoryless process's set of 'good indices' varies when a small amount of memory structure is imposed on the process, and also to determine the behavior of the Bhattacharyya parameters under such variations.

Robustness against uncertainty in the channel is also often studied as a *compound channel* problem, where the task is to design a code that will perform well over all memoryless channels in a given class. Polar coding for compound channels was considered in [15] by Hassani *et. al.*, where it was shown that over a compound channel that includes the binary symmetric and binary erasure channels with equal capacities, polar codes achieve strictly smaller rates than the compound channel capacity under SC decoding. In Appendix 7, it is shown that this gap to capacity is indeed due to the suboptimality of the SC decoder, and can be closed by employing optimal decoders at the receiver. An open problem of interest is to determine whether polar codes achieve compound channel capacity under low-complexity decoding algorithms.

In Chapter 6 we considered polarization for multi-user coding settings. We first showed that all optimal rates for multiple-access channels and the distributed source coding problems can be achieved using polar codes at each user. We then showed that applying polarizing transforms to multiple pro-

cesses separately not only polarizes the processes, but the correlations are also polarized. We saw that coding schemes exploiting this joint polarization phenomenon achieve some, but not always all, optimal points in the rate regions of the mentioned problems, with error probabilities comparable to those of single-user polar coding schemes. One should note that the unachievability of certain rate points by this scheme is not due to the way that the processes are polarized—they are indeed polarized using the same transform as in the first method discussed above—but rather to the proposed decoding order, which does not fully exploit the resulting probability structure. This rate loss is a good example that illustrates the strong dependence of polarization on how the probability structure in a process is decomposed through the choice of the decoding algorithm.

Although we have demonstrated that polarization is a fairly general phenomenon, the extent of the practical and the theoretical implications of this generality is largely unknown. We leave this problem for future study.

Appendix

Here we show that many good codes for a given binary symmetric channel also perform well over symmetric binary-input channels with higher capacities. In order to do so, we first prove that the binary symmetric channel is the *least capable* among all symmetric channels with a given capacity. Recall that a channel $W\colon \mathcal{X} \to \mathcal{Y}$ is said to be *more capable* [16, p. 116] than $V\colon \mathcal{X} \to \mathcal{Z}$ if

$$I(X;Y) \geq I(X;Z)$$

for all joint distributions $p_{XYZ}(x,y,z) = p(x)W(y \mid x)V(z \mid x)$.

Lemma 7.1. *The binary symmetric channel with capacity C is the least capable among all symmetric binary-input channels with capacity at least C.*

Proof. Let $h\colon [0,1/2] \to [0,1]$ denote the binary entropy function. Recall that any symmetric binary-input channel can be written as one with input $X \in \{0,1\}$, output $(T,Y) \in [0,1/2] \times \{0,1\}$, and

$$p(x,t,y) = p(x)p(t)p(y \mid x,t)$$

with

$$p(y \mid x,t) = \begin{cases} t & \text{if } y \neq x \\ 1-t & \text{if } y = x \end{cases}. \tag{7.1}$$

That is, any symmetric channel is a combination of binary symmetric channels.

It suffices to prove the claim for symmetric channels with capacity C, since channels with higher capacities are upgraded with respect to these. To that end, let ϵ be the crossover probability of a binary symmetric channel with

capacity C, i.e. $C = 1-h(\epsilon)$. Now note that with input distribution $p_X(0) = q$, the mutual information developed across a binary symmetric channel with capacity C is

$$h(q * \epsilon) - h(\epsilon).$$

where $a * b := a(1-b) + (1-a)b$. On the other hand, the mutual information developed across a symmetric channel W with capacity C under the same input distribution is

$$\begin{aligned} I(X;YT) &= H(YT) - H(YT \mid X) \\ &= H(T) + H(Y \mid T) - H(T \mid X) - H(Y \mid TX) \\ &= H(Y \mid T) - H(Y \mid TX) \\ &= H(Y \mid T) - H(Y + X \mid T) \\ &= E[h(q * T)] - h(\epsilon), \end{aligned}$$

where the third equality follows from the independence of T and X, and the last equality follows from (7.1) and the relation $H(Y + X \mid T) = E[h(T)] = 1 - C = h(\epsilon)$. It is known that the function $h(a * h^{-1}(t))$ is convex in t [8], and thus we can continue the above chain of equalities as

$$\begin{aligned} I(X;YT) &= E[h(q * T)] - h(\epsilon) \\ &= E\left[h\big(q * h^{-1}(h(T))\big)\right] - h(\epsilon) \\ &\geq h(q * h^{-1}(E[h(T)])) - h(\epsilon) \\ &= h(q * \epsilon) - h(\epsilon) \end{aligned}$$

completing the proof. □

We next show that the performance of a code over a channel W cannot be much worse than its performance over a less capable channel V. This and the above result will imply that a sequence of codes with sublinear error probability decay (in the blocklength) over a binary symmetric channel will have vanishing error probability over any symmetric channel with a higher capacity. It will also follow that the error probability of polar codes designed for a binary symmetric channel is roughly $O(2^{-\sqrt{N}})$ when used over a symmetric channel with higher capacity, provided that the code is decoded optimally.

Proposition 7.1. *Let $P_{e,W}$ denote the average error probability of a code $\mathcal{C}$ of length N over a binary-input channel W, under optimal decoding. If W is more capable than V, then*

$$P_{e,W} \leq N P_{e,V} + h(P_{e,V}).$$

Proof. Let X_1^N denote a randomly chosen codeword from $\mathcal{C}$. Let Y_1^N and Z_1^N denote the outputs of channels W and V, respectively, with input X_1^N. Fano's inequality states that

$$H(X_1^N \mid Z_1^N) \leq N P_{e,V} + h(P_{e,V}).$$

Since W is more capable than V we also have [16, p. 116] $H(X_1^N \mid Y_1^N) \leq H(X_1^N \mid Z_1^N)$, from which it follows that

$$H(X_1^N \mid Y_1^N) \leq NP_{e,V} + h(P_{e,V}).$$

The claim is then a corollary to Lemma 7.2 below, which states that

$$P_{e,W} \leq 1 - e^{-H(X_1^N|Y_1^N)},$$

and to the relation $1 - e^{-H(X_1^N|Y_1^N)} \leq H(X_1^N \mid Y_1^N)$. □

The following lemma is an upper bound on the error probability of optimal decoding in terms of conditional entropy, and is a variant of the one in [10, Problem 4.7]:

Lemma 7.2 ([17])**.** *Let X be a discrete random variable and Y an arbitrary random variable. The average error probability of optimally decoding X upon observing Y satisfies*

$$P_e \leq 1 - e^{-H(X|Y)}.$$

Proof. Let

$$x_y := \arg\max_x p(x \mid y)$$

and let $P_e(y)$ denote the probability of a decoding error conditioned on $Y = y$. We have

$$\begin{aligned} P_e(y) &= 1 - p(x_y \mid y) \\ &= 1 - \prod_{x \in \mathcal{X}} p(x_y \mid y)^{p(x|y)} \\ &\leq 1 - \prod_{x \in \mathcal{X}} p(x \mid y)^{p(x|y)} \\ &= 1 - e^{-H(X|Y=y)} \end{aligned}$$

Then,

$$P_e = \int_y p(y) P_e(y) dy \leq 1 - \int_y p(y) e^{-H(X|Y=y)} dy \leq 1 - e^{-H(X|Y)},$$

where the last inequality is due to the convexity of the function $t \to e^{-t}$. □

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, July and October 1948.

[2] D. J. Costello Jr. and G. D. Forney Jr., "Channel coding: The road to channel capacity," *Proceedings of the IEEE*, vol. 95, no. 6, June 2007.

[3] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[4] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Seoul, South Korea, July 2009, pp. 1493–1495.

[5] I. Tal and A. Vardy, "How to construct polar codes," in *Proc. of the IEEE Inform. Theory Workshop*, Dublin, Ireland, Sept. 2010.

[6] E. Arıkan, "A survey of Reed-Muller codes from polar coding perspective," in *Proc. of the IEEE Inform. Theory Workshop*, Cairo, Egypt, Jan. 2010.

[7] ——, "Source polarization," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Austin, Texas, June 2010.

[8] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 769–772, Nov. 1973.

[9] A. Clark, *Elemens of Abstract Algebra*. New York: Dover, 1971.

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

[11] S. B. Korada, E. Şaşoğlu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *accepted for publication in IEEE Transactions on Information Theory*, 2009.

[12] F. J. MacWilliams and N. J. Sloane, *The Theory of Error-Correcting Codes*. North-Holland, 1977.

[13] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[14] A. Grant, B. Rimoldi, R. Urbanke, and P. Whiting, "Rate-splitting multiple access for discrete memoryless channel," *IEEE Transactions on Information Theory*, vol. IT–47, no. 3, pp. 873–890, Mar. 2001.

[15] S. H. Hassani, S. B. Korada, and R. Urbanke, "The compound capacity of polar codes," *in preparation*, 2009.

[16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

[17] E. Telatar, private communication.

Curriculum Vitae

Education

Ecole Polytéchnique Fédérale de Lausanne (EPFL), Lausanne

Ph.D., Communication Systems, November 2011

M.Sc., Communication Systems, April 2007

Boğaziçi University, Istanbul

B.Sc., Electrical and Electronics Engineering, July, 2005

Publications

Monograph

E. Şaşoğlu, *Polar Codes: A Tutorial,* in preparation for the Foundations and Trends in Communications and Information Theory.

Journal Papers

S. Korada, E. Şaşoğlu, R. Urbanke, "Polar Codes: Characterization of exponent, bounds, and constructions," *IEEE Transactions on Information Theory,* vol. 56, pp. 6253–6264, Dec. 2010.

E. Şaşoğlu, E. Telatar, E. Yeh, "Polar codes for the two-user multiple-access channel," submitted to *the IEEE Transactions on Information Theory,* June 2010.

Online Papers

E. Şaşoğlu, E. Telatar, E. Arıkan, "Polarization for arbitrary discrete memoryless channels," available on arXiv.org [IT.cs] 0908.0302.

Conference Papers

E. Şaşoğlu, "Successive cancellation for cyclic interference channels," *Proc. IEEE Inform. Theory Workshop,* May 2008.

S. Korada, E. Şaşoğlu, "A class of channels that polarize binary input memoryless channels," *Proc. IEEE Intern. Symp. Inform. Theory,* Seoul, June–July 2009.

S. Korada, E. Şaşoğlu, R. Urbanke, "Polar Codes: Characterization of exponent, bounds, and constructions," *Proc. IEEE Intern. Symp. Inform. Theory,* Seoul, June–July 2009. vol. 56, pp. 6253–6264, Dec. 2010.

E. Şaşoğlu, E. Telatar, E. Arıkan, "Polarization for arbitrary discrete memoryless channels," *Proc. IEEE Inform. Theory Workshop,* Taormina, October 2009.

E. Şaşoğlu, E. Telatar, E. Yeh, "Polar codes for the two-user multiple-access channel," *Proc. IEEE Inform. Theory Workshop,* Cairo, January 2010.

E. Şaşoğlu, "An entropy inequality for q-ary random variables and its application to channel polarization," *Proc. IEEE Intern. Symp. Inform. Theory,* Austin, June 2010.

E. Şaşoğlu, "Polarization in the Presence of Memory," *Proc. IEEE Intern. Symp. Inform. Theory,* St. Petersburg, July–August 2011.