Polar Codes for Multiple Descriptions

POLAR CODES FOR MULTIPLE DESCRIPTIONS

BY

QI SHI, B.Eng.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

Master of Applied Science (2013)	McMaster University
(Electrical & Computer Engineering)	Hamilton, Ontario, Canada

TITLE:	Polar Codes for Multiple Descriptions
AUTHOR:	Qi Shi B.Eng., (Electrical Engineering) Mcmaster University, Hamilton, Canada
SUPERVISOR:	Dr. Jun Chen
NUMBER OF PAGES:	x, 50

This thesis is dedicated to my family for supporting me all the time

Abstract

Two coding schemes based on polar codes are proposed for the multiple description (MD) problem. The first scheme is an adaptation of the one developed by Şaşoğlu *et al.* for the multiple access channel to the MD setting. Specifically, it is shown that the scheme is able to achieve a certain rate pair on the dominant line of the achievable rate region determined by El Gamal and Cover (EGC). Different from polar coding for the multiple access channel considered by Şaşoğlu *et al.*, the auxiliary random variables in the MD problem can be dependent. The second scheme is based on the idea of rate splitting. We show that it can achieve the entire EGC rate region. The effectiveness of the proposed polar coding schemes is verified by the experimental results.

Acknowledgements

First of all, I would like to thank my advisor Dr. Jun Chen. Thanks for providing me the opportunity to work with this wonderful group. His tutors enlighten me and encourage me to explorer. Through his guidance, this research becomes motivational, interesting, and inspiring.

I also thank Dr. D. Zhao, and Dr. S. Dumitrescu for their time and advice. My gratitude also goes to Dr. Chao Tian. Without his previous work on polar coding, this thesis would not have been written.

Thanks to Dr. Lin Song, for the time she dedicates on the discussion with me. She always points me the right direction and her advice means a lot to me.

I wish to thank my colleagues and friends Shuo Shao and Jihai Cao, for the two years of graduate school we spent together.

Special Thanks to my parents for encouraging me and giving me confidence.

Notation and abbreviations

EGC	El Gamal and Cover
MAC	multiple access channel
iid	identical independent distribution

$\mathcal{X}$	alphabet set
$\mathcal{X} \times \mathcal{Y}$	multiple alphabet set
$\lvert\mathcal{X}\rvert$	cardinality of the alphabet set
$O()$	big-O notation
$o()$	small-o notation
$H(.)$	entropy
$I(.;.)$	mutual information
$\mathbb{E}$	expectation
$p(.)$	distribution
x^n	a sequence containing n elements
$P^+(.), P^-(.)$	the distribution after applying polarization
Pr	probabilty
$\sum$	sum of a sequence

$\prod$	product of a sequence
$X - Y - Z$	Markov chain

Contents

List of Figures

Chapter 1

Introduction and Problem Statement

Polar codes, invented by Arikan (Arikan, 2009), are "the first provably capacity-achieving codes for any symmetric input discrete memoryless channel that have low encoding and decoding complexity" (Korada, 2008). Specifically, the encoding and decoding complexity of polar codes is of order $O(n \log n)$, where n is the code block length. Furthermore, Arikan and Telatar (Arikan and Telatar, 2009) upper-bounded the block error probability to order $o(\exp(-n^{1/2-\epsilon}))$. The most intriguing aspect of this new coding technique is that through recursive channel splitting and combining, n uses of the same memoryless channel are converted to successive uses of n different channels, and, asymptotically, each one of these converted channels is extremal in the sense that it is either a perfect channel or a pure noise channel. Subsequently, Korada and Urbanke proposed a lossy source coding scheme based on polar codes (Korada and Urbanke, 2010); see also (Karzand and Teletar, 2010; Korada, 2008). Şaşoğlu *et al.* (E. Sasoglu and Yeh, 2009) extended the polar coding technique to the two-user multiple access channel, where the channel polarizes into one of five extremal multiple access channels.

Motivated by the success in (E. Sasoglu and Yeh, 2009) for the multiple access channel, we apply the polar coding technique to the multiple description (MD) problem. In MD coding, a single source X is encoded into two descriptions of rate R_1 and rate R_2, respectively, such that the reconstruction distortion based on description i is D_i, $i = 1, 2$, and the reconstruction distortion based on the two descriptions is D_0. The goal is to find efficient coding schemes to achieve the optimal tradeoff between (R_1, R_2) and (D_0, D_1, D_2). Unfortunately, the optimal rate-distortion tradeoff in MD coding is unknown except for certain special cases. In this work we focus on the achievable rate pairs subject to distortion constraints (D_0, D_1, D_2) determined by El Gamal and Cover (Gamal and Cover, 1982) (sometimes referred to as the EGC rate region) and propose two MD coding schemes based on polar codes.

The first scheme exhibits a similar polarizing behavior as that in (E. Sasoglu and Yeh, 2009), yet the specific coding requirement in the MD problem introduces a crucial difference, and the alphabet sizes of the two descriptions play a significant role in our result. In particular, we show that this scheme can achieve certain rate pairs on the dominant line of the EGC rate region when the associated auxiliary random variables are marginally uniformly distributed in two alphabets of different sizes; moreover, the codes essentially polarize to four extreme cases, and the same error behavior of order $o(\exp(-n^{1/2-\epsilon}))$ as in the point-to-point case can be obtained. However, when the alphabet sizes of the two auxiliary random variables are the same, a complication arises with an additional extreme case, and though the codes also polarize, its error behavior becomes more difficult to characterize.

In contrast to the first proposed scheme which can only achieve certain rate pairs on the dominant line of the EGC rate region, our second scheme can achieve the entire rate region. Specifically, for the second scheme, we first use the rate splitting method (Y. Zhang, 2012) to convert the MD problem to a successive lossy source coding problem, then implement each coding step using polar codes.

The remainder of this thesis is organized as follows. We give a short review of the EGC rate region in Section 2.1. A joint polarization scheme together with its performance analysis can be found in Chapter 2. A different polar coding scheme based on the rate splitting method is discussed in Chapter 3. Chapter 4 contains the experimental results for the proposed schemes. We conclude the thesis in Chapter 5.

For any set $\mathcal{S}$, we use $|\mathcal{S}|$ to denote its cardinality. Throughout this thesis, he logarithm function is to base 2 unless stated otherwise.

Chapter 2

Polar Codes for Multiple Descriptions with Joint Polarization

2.1 The EGC Rate-Distortion Region

It was shown in (J. Wang and Permuter, 2009) that the achievable region given in (Gamal and Cover, 1982) has an alternative form, and thus the main result in (Gamal and Cover, 1982) can be given as follows.

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables drawn according to a probability mass function $p_X(x)$ defined over finite alphabet $\mathcal{X}$. Let $d_i : \mathcal{X} \times \hat{\mathcal{X}}_i \to [0, d_{\max}]$, $i = 0, 1, 2$, be bounded distortion measures, where $\hat{\mathcal{X}}_i$, $i = 0, 1, 2$, are the reconstruction alphabets.

Definition 1 *A rate pair* (R_1, R_2) *is said to be achievable subject to distortion constraints* (D_0, D_1, D_2) *if, for every* $\epsilon > 0$*, there exist encoding functions* $f_i^{(n)} : \mathcal{X}^n \to \mathcal{C}_i$*,* $i = 1, 2$*, and decoding functions* $g_0^{(n)} : \mathcal{C}_1 \times \mathcal{C}_2 \to \hat{\mathcal{X}}_0^n$ *and* $g_i^{(n)} : \mathcal{C}_i \to \hat{\mathcal{X}}_i^n$*,* $i = 1, 2$*, such that*

$$\frac{1}{n}\log|\mathcal{C}_i| \le R_i + \epsilon, \quad i = 1, 2,$$
$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}d_i(X_t, \hat{X}_{i,t}) \le D_i + \epsilon, \quad i = 0, 1, 2,$$

where $\hat{X}_0^n = g_0^{(n)}(f_1^{(n)}(X^n), f_1^{(n)}(X^n))$ *and* $\hat{X}_i^n = g_i^{(n)}(f_i^{(n)}(X^n))$*,* $i = 1, 2$*.*

The following result provides a sufficient condition on the achievability of rate pair (R_1, R_2) subject to distortion constraints (D_0, D_1, D_2). It is shown in J. Wang and Permuter (2009) that this sufficient condition is essentially equivalent to the one derived by El Gamal and Cover Gamal and Cover (1982).

Theorem 1 *A rate pair* (R_1, R_2) *is achievable subject to subject to distortion constraints* (D_0, D_1, D_2) *if*

$$R_1 \ge I(X;Y),$$
$$R_2 \ge I(X;Z),$$
$$R_1 + R_2 \ge I(X;Y,Z) + I(Y;Z),$$

for some probability mass function $p(x, y, z) \in \mathcal{P}$*, where* $\mathcal{P}$ *is the set of distributions* $p(x, y, z) = p_X(x)p(y, z|x)$ *defined over the alphabet* $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$*, for which there exist* *some deterministic mappings* ϕ_0, ϕ_1, *and* ϕ_2 *such that*

$$D_0 \geq \mathbb{E}[d_0(X, \phi_0(Y, Z))],$$
$$D_1 \geq \mathbb{E}[d_1(X, \phi_1(Y))],$$
$$D_2 \geq \mathbb{E}[d_2(X, \phi_2(Z))].$$

For a generic distribution in $\mathcal{P}$, we shall write it in capital letter as $P(x, y, z)$, and sometimes write it as P when no ambiguity arises. Following the notation in E. Sasoglu and Yeh (2009), define

$$I^{(0)}(P) = I(X; Y, Z) + I(Y; Z),$$
$$I^{(1)}(P) = I(X; Y), \quad \overline{I^{(1)}}(P) = I(Z; X, Y),$$
$$I^{(2)}(P) = I(X; Z), \quad \overline{I^{(2)}}(P) = I(Y; X, Z).$$

Further define $\mathcal{K}(P) = (I^{(0)}(P), I^{(1)}(P), I^{(2)}(P))$ and

$$\mathcal{J}(P) = \{(R_1, R_2) : R_1 \geq I^{(1)}(P),\ R_2 \geq I^{(2)}(P),\ R_1 + R_2 \geq I^{(0)}(P)\}.$$

Since $I^{(0)}(P) \geq I^{(1)}(P) + I^{(2)}(P)$, the set

$$\mathcal{F}(P) = \{(R_1, R_2) \in \mathcal{J}(P) : R_1 + R_2 = I^{(0)}(P)\}$$

is non-empty. We shall refer to $\mathcal{J}(P)$ as the EGC rate region and $\mathcal{F}(P)$ as its dominant line.

2.2 Source Combination and Splitting

In this section we mainly focus on a special class of distributions in $\mathcal{P}$ where Y is uniformly distributed in $\mathcal{Y}$, and Z is uniformly distributed in $\mathcal{Z}$, and furthermore $\mathcal{Y} = \{0, 1, \ldots, q_1 - 1\} \triangleq \mathbb{F}_{q_1}$ and $\mathcal{Z} = \{0, 1, \ldots, q_2 - 1\} \triangleq \mathbb{F}_{q_2}$ with q_1 and q_2 being two non-identical primes; the case $q_1 = q_2$ is discussed separately in Section 2.6. Note that every random variable can be approximated arbitrarily well by a random variable uniformly distributed over a sufficiently large alphabet through a deterministic mapping; as a consequence, there is no essential loss in focusing on this special class of distributions.

Let us fix a distribution $P(x, y, z)$. Consider random variables X_1 and X_2. We can view them jointly as a new vector source (X_1, X_2) and apply Theorem 1 to this source with (X_1, Y_1, Z_1) independent of (X_2, Y_2, Z_2), resulting in a rate region $\mathcal{J}(P^2)$ specified by the following quantities

$$2I^{(0)}(P) = I(X_1, X_2; Y_1, Y_2, Z_1, Z_2) + I(Y_1, Y_2; Z_1, Z_2),$$
$$2I^{(1)}(P) = I(X_1, X_2; Y_1, Y_2), \quad 2I^{(2)}(P) = I(X_1, X_2; Z_1, Z_2).$$

Similarly as in Arikan (2009); E. Sasoglu and Yeh (2009), define

$$Y_1 = U_1 + U_2, \quad Y_2 = U_2,$$
$$Z_1 = V_1 + V_2, \quad Z_2 = V_2,$$

where $+$ is modulo-q_1 (or modulo-q_2) addition. The random variables U_1 and U_2 are each uniformly distributed in the alphabet $\mathbb{F}_{q_1}$; the random variables V_1 and V_2 are each uniformly distributed in the alphabet $\mathbb{F}_{q_2}$. Since the mapping between (Y_1, Y_2) and (U_1, U_2) is bijective, it is easily seen that U_1 and U_2 are independent; for the same reason, V_1 and V_2 are also independent. However, (U_1, V_1) and (U_2, V_2) are not necessarily independent.

Note that

$$\begin{aligned} 2I^{(1)}(P) &= I(U_1, U_2; X_1, X_2) \\ &= I(U_1; X_1, X_2) + I(U_2; X_1, X_2, U_1) \qquad (2.1) \\ &\leq I(U_1; X_1, X_2) + I(U_2; X_1, X_2, U_1, V_1). \qquad (2.2) \end{aligned}$$

Similarly

$$\begin{aligned} 2I^{(2)}(P) &= I(V_1, V_2; X_1, X_2) \\ &= I(V_1; X_1, X_2) + I(V_2; X_1, X_2, V_1) \qquad (2.3) \\ &\leq I(V_1; X_1, X_2) + I(V_2; X_1, X_2, U_1, V_1). \qquad (2.4) \end{aligned}$$

Moreover, we have

$$\begin{aligned} 2\overline{I^{(1)}}(P) &= I(V_1, V_2; X_1, X_2, U_1, U_2) \\ &= I(V_1; X_1, X_2, U_1, U_2) + I(V_2; X_1, X_2, U_1, U_2 | V_1) \\ &= I(V_1; X_1, X_2, U_1, U_2) + I(V_2; X_1, X_2, U_1, U_2, V_1) \\ &\geq I(V_1; X_1, X_2, U_1) + I(V_2; X_1, X_2, U_1, U_2, V_1), \qquad (2.5) \end{aligned}$$

and similarly

$$2\overline{I^{(2)}}(P) \geq I(U_1; X_1, X_2, V_1) + I(U_2; X_1, X_2, U_1, V_1, V_2). \tag{2.6}$$

Finally, it can be verified that

$$\begin{aligned}
2I^{(0)}(P) &= I(Y_1, Y_2, Z_1, Z_2; X_1, X_2) + I(Y_1, Y_2; Z_1, Z_2) \\
&= I(U_1, U_2, V_1, V_2; X_1, X_2) + I(U_1, U_2; V_1, V_2) \\
&= I(U_1, V_1; X_1, X_2) + I(U_2, V_2; X_1, X_2|U_1, V_1) + I(U_1, U_2; V_1, V_2) \\
&= I(U_1, V_1; X_1, X_2) + I(U_2, V_2; X_1, X_2, U_1, V_1) + I(U_1, U_2; V_1, V_2) - I(U_1, V_1; U_2, V_2) \\
&= I(U_1, V_1; X_1, X_2) + I(U_2, V_2; X_1, X_2, U_1, V_1) + I(U_1; V_1) + I(U_2; V_2).
\end{aligned} \tag{2.7}$$

Note that (2.7) is true because

$$\begin{aligned}
&I(U_1, U_2; V_1, V_2) - I(U_1, V_1; U_2, V_2) - [I(U_1; V_1) + I(U_2; V_2)] \\
&= H(U_1, U_2) + H(V_1, V_2) - H(U_1, U_2, V_1, V_2) - H(U_1, V_1) - H(U_2, V_2) + H(U_1, U_2, V_1, V_2) \\
&\quad - H(U_1) - H(V_1) + H(U_1, V_1) - H(U_2) - H(V_2) + H(U_2, V_2) \\
&= H(U_1, U_2) + H(V_1, V_2) - H(U_1) - H(V_1) - H(U_2) - H(V_2) \\
&= -I(U_1; U_2) - I(V_1; V_2) \\
&= 0,
\end{aligned}$$

where the last step follows from the independence between U_1 and U_2, as well as the independence between V_1 and V_2.

The triple

$$\Big(I(U_1, V_1; X_1, X_2) + I(U_1; V_1), I(U_1; X_1, X_2), I(V_1; X_1, X_2)\Big)$$

is in fact $\mathcal{K}(p((x_1, x_2), u_1, v_1))$, and the triple

$$\Big(I(U_2, V_2; X_1, X_2, U_1, V_1) + I(U_2; V_2), I(U_2; X_1, X_2, U_1, V_1), I(V_2; X_1, X_2, U_1, V_1)\Big)$$

is in fact $\mathcal{K}(p((x_1, x_2, u_1, v_1), u_2, v_2))$. If we view $\mathcal{K}(P)$ as encoding source X with random variables Y and Z, then $\mathcal{K}(p((x_1, x_2), u_1, v_1))$ can be viewed as encoding (X_1, X_2) with random variables U_1 and V_1, and $\mathcal{K}(p((x_1, x_2, u_1, v_1), u_2, v_2))$ as encoding (X_1, X_2, U_1, V_1) using random variables U_2 and V_2.

Definition 2 *Given a source X and the joint distribution $P(x, y, z)$ in the alphabet $\mathcal{X} \times \mathbb{F}_{q_1} \times \mathbb{F}_{q_2}$, define two new sources in the alphabet $\mathcal{X}^2$ and $\mathcal{X}^2 \times \mathbb{F}_{q_1} \times \mathbb{F}_{q_2}$, respectively, and the joint distributions together with the auxiliary encoding random variables (U_1, V_1) and (U_2, V_2), respectively, in the corresponding alphabets,*

$$P^-((x_1, x_2), u_1, v_1) = \sum_{u_2 \in \mathcal{Y}} \sum_{v_2 \in \mathcal{Z}} P(x_2, u_2, v_2) P(x_1, u_1 + u_2, v_1 + v_2) \tag{2.8}$$

$$P^+((x_1, x_2, u_1, v_1), u_2, v_2) = P(x_2, u_2, v_2) P(x_1, u_1 + u_2, v_1 + v_2), \tag{2.9}$$

where we have written (x_1, x_2) as a group and view it as a single random vector, and similarly for (x_1, x_2, u_1, v_1).

2.3 Polarization

Note that

$$\begin{aligned} I^{(1)}(P^+) &= I(U_2; X_1, X_2, U_1, V_1) \\ &\geq I(U_2; X_1, X_2, U_1) \\ &\geq I(U_2; X_2) \\ &= I(Y_2; X_2) \\ &= I^{(1)}(P), \end{aligned}$$

and thus from (2.1), we have

$$\begin{aligned} I^{(1)}(P^-) &= I(U_1; X_1, X_2) \\ &\leq I(Y_2; X_2) \\ &= I^{(1)}(P). \end{aligned}$$

Similarly, it can be shown that $I^{(2)}(P^-) \leq I^{(2)}(P) \leq I^{(2)}(P^+)$. Since

$$I(U_2, V_2; X_1, X_2, U_1, V_1) \geq I(U_2, V_2; X_1, X_2)$$

and

$$I(U_2; V_2) = I(Y_2; Z_2),$$

it follows that

$$\begin{aligned} I^{(0)}(P^+) &= I(U_2, V_2; X_1, X_2, U_1, V_1) + I(U_2; V_2) \\ &\geq I^{(0)}(P), \end{aligned}$$

which, together with (2.7), implies $I^{(0)}(P^-) \leq I^{(0)}(P) \leq I^{(0)}(P^+)$. Therefore, we have $\mathcal{K}(P^-) \preceq \mathcal{K}(P) \preceq \mathcal{K}(P^+)$, where $\preceq$ means that $\leq$ holds for each component of the vector.

We can now apply the same process to P^- and P^+ using Definition 2, resulting in four new sources $P^{--} = (P^-)^-$, $P^{-+} = (P^-)^+$, $P^{+-} = (P^+)^-$, $P^{++} = (P^+)^+$. Repeating this process k times, we derive a new set of 2^k sources.

The following Theorem formally establishes the polarization behavior, in a manner similar to those seen in the multiple access channel E. Sasoglu and Yeh (2009). For simplicity, we write $\log(q_1)$ as Δ_1 and $\log(q_2)$ as Δ_2.

Theorem 2 *Let $P(x, y, z)$ be a joint source-codebook distribution defined over $\mathcal{X} \times \mathbb{F}_{q_1} \times \mathbb{F}_{q_2}$ with q_1 and q_2 being two non-identical primes. Define*

$$M \triangleq \{(0, 0, 0), (\Delta_1, \Delta_1, 0), (\Delta_2, 0, \Delta_2), (\Delta_1 + \Delta_2, \Delta_1, \Delta_2)\}.$$

For $p \in \mathbb{R}^3$, let $||p - M|| \triangleq \min_{x \in M} ||p - x||$ denote the distance from a point p to M, where $|| \cdot ||$ is the Euclidean norm. Then for any $\delta > 0$,

$$\lim_{k\to\infty} \frac{1}{2^k} \left|\{s \in \{-, +\}^k : ||\mathcal{K}(P^s) - M|| \geq \delta\}\right| = 0. \tag{2.10}$$

Theorem 2 essentially states that for most of the new sources, the rate region

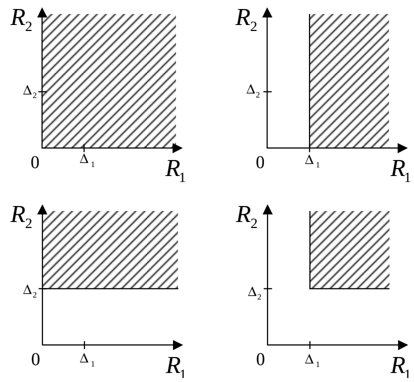

Figure 2.1: The four extremal rate regions.

$\mathcal{J}(P^s)$ approaches one of the four extremal ones, as depicted in Fig. 2.1.

The rate region associated with $(0, 0, 0)$ corresponds to the case that both descriptions reveal nothing about the source, thus are useless; the rate regions associated with $(\Delta_1, \Delta_1, 0)$ and $(\Delta_2, 0, \Delta_2)$ correspond to the cases where one description is useless, while the other provides full information regarding the source; the last case $(\Delta_1 + \Delta_2, \Delta_1, \Delta_2)$ is when both descriptions provide full information.

From a coding perspective, the case $(0, 0, 0)$ corresponds to the case that both coded symbols can be set arbitrarily, as long as they are revealed to both the encoder and the decoder; the cases $(\Delta_1, \Delta_1, 0)$ and $(\Delta_2, 0, \Delta_2)$ correspond to the case that the coded symbol in one of the descriptions (the one with non-zero rate) is fully determined by the source vector, while the other can be set arbitrarily; the case $(\Delta_1 + \Delta_2, \Delta_1, \Delta_2)$ corresponds tot the case that both symbols are fully determined. This will become clearer in Section 2.4 when we describe the codes in more details.

The proof of Theorem 2 closely follows that for the multiple access channel given in E. Sasoglu and Yeh (2009), and thus we only discuss some of the steps which are different from E. Sasoglu and Yeh (2009).

Lemma 1 (Polarization Lemma Sasoglu (2010)) *Let p be a prime number. For any $\epsilon > 0$, there is a $\delta > 0$ such that if*

1. $Q : \mathbb{F}_q \to \mathcal{B}$ *is a q-ary input channel with arbitrary output alphabet* $\mathcal{B}$,
2. A_1, A_2, B_1, B_2 *are random variables jointly distributed as*

$$p_{A_1,A_2,B_1,B_2}(a_1, a_2, b_1, b_2) = \frac{1}{q^2} Q(b_1|a_1 + a_2)Q(b_2|a_2),$$

3. $I(A_2; B_1, B_2, A_1) - I(A_2; B_2) < \delta$, *where the mutual information is of logarithm base-q,*

then

$$I(A_2; B_2) \notin (\epsilon, 1 - \epsilon).$$

Corollary 1 *For any $\epsilon > 0$ there exists a $\delta > 0$ such that if $P(x, y, z)$ is a joint distribution with $I^{(1)}(P^+) - I^{(1)}(P) \leq \delta$, then $I^{(1)}(P) \notin (\epsilon, \Delta_1 - \epsilon)$; similarly if $I^{(2)}(P^+) - I^{(2)}(P) \leq \delta$, then $I^{(2)}(P) \notin (\epsilon, \Delta_2 - \epsilon)$.*

Proof 1 *By symmetry, we only prove the first case with $j = 1$ where j is the index of $I^{(j)}$. Apply Lemma 1 with $A_i = U_i$ and $B_i = X_i$ for the first statement, then the result follows.*

Corollary 2 *For any $\epsilon > 0$ there exists a $\delta > 0$ such that if $P(x, y, z)$ is a joint distribution with $\overline{I^{(1)}}(P^+) - \overline{I^{(1)}}(P) \leq \delta$, then $\overline{I^{(1)}}(P) \notin (\epsilon, \Delta_2 - \epsilon)$; similarly if $\overline{I^{(2)}}(P^+) - \overline{I^{(2)}}(P) \leq \delta$, then $\overline{I^{(2)}}(P) \notin (\epsilon, \Delta_1 - \epsilon)$.*

Proof 2 *Again by symmetry, only $j = 1$ needs to be proved. Observe that*

$$\begin{aligned}&\overline{I^{(1)}}(P^+) - \overline{I^{(1)}}(P)\\&= I(V_2; X_1, X_2, U_1, U_2, V_2) - I(V_2; X_2, U_2)\\&\geq I(V_2; X_1, X_2, U_1, U_2) - I(V_2; X_2, U_2),\end{aligned}$$

thus the assumption implies

$$I(V_2; X_1, X_2, U_1, U_2) - I(V_2; X_2, U_2) \leq \delta.$$

Now applying Lemma 1 with $A_i = V_i$ and $B_i = (X_i, U_i)$ gives the result.

Similarly to Arikan and Telatar (2009); E. Sasoglu and Yeh (2009), let $B_1, B_2, \ldots,$ be an i.i.d. sequence of random variables taking values in the set $\{-, +\}$, with $\Pr(B_i = -) = \Pr(B_i = +) = 0.5$. Define a random process $\{P_k : k \geq 0\}$ by

$$P_0 = P, \quad P_k = P_{k-1}^{B_k}, \, k \geq 1. \tag{2.11}$$

Now define the random process $\{I_k^{(1)} : k \geq 0\}$, $\{I_k^{(2)} : k \geq 0\}$, $\overline{\{I_k^{(1)}} : k \geq 0\}$, $\overline{\{I_k^{(2)}} : k \geq 0\}$, and $\{I_k^{(0)} : k \geq 0\}$ as

$$\begin{aligned}I_k^{(j)} &= I^{(j)}(P_k), \quad j = 0, 1, 2,\\\overline{I_k^{(j)}} &= \overline{I^{(j)}}(P_k), \quad j = 1, 2.\end{aligned}$$

We have the following result.

Theorem 3 *When q_1 and q_2 are non-identical primes, the process $(I_k^{(1)}, I_k^{(2)}, I_k^{(0)})$ converges almost surely, and the limit belongs to the set M in Theorem 2 with probability 1.*

Proof 3 *The almost sure convergence follows from the simple fact that $\{I_k^{(1)} : k \geq 0\}$ and $\{I_k^{(2)} : k \geq 0\}$ are bounded submartigales by (2.2) and (2.4), $\{\overline{I_k^{(1)}} : k \geq 0\}$ and $\{\overline{I_k^{(2)}} : k \geq 0\}$ are bounded supermartigales by (2.5) and (2.6), and $\{I_k^{(0)} : k \geq 0\}$ is a bounded martigale by (2.7).*

By applying Corollary 1, it is not too difficult to see that

$$\lim_{k\to\infty} I_k^{(j)} \in \{0, \Delta_j\}, \quad j = 1, 2.$$

Similarly by applying Corollary 3, we have

$$\lim_{k\to\infty} \overline{I_k^{(1)}} \in \{0, \Delta_2\},$$
$$\lim_{k\to\infty} \overline{I_k^{(2)}} \in \{0, \Delta_1\}.$$

For $I_k^{(0)}$, we only need to observe that

$$I_k^{(1)} + \overline{I_k^{(1)}} = I_k^{(2)} + \overline{I_k^{(2)}} = I_k^{(0)} \tag{2.12}$$

for any $k \geq 0$. However when $I_k^{(1)} = I_k^{(2)} = 0$, it must be true that $I_k^{(0)} = 0$, because otherwise $\overline{I_k^{(1)}} = \overline{I_k^{(2)}} \neq 0$, which is impossible. This completes the proof.

Theorem 2 now follows immediately. The difference between the proof of this theorem and that given in E. Sasoglu and Yeh (2009) for the multiple access channel is the dependence on the alphabet sizes q_1 and q_2, and that instead of considering

$(I_k^{(1)}, I_k^{(2)}, I_k^{(0)})$ as in E. Sasoglu and Yeh (2009), we in fact need to consider the process $(I_k^{(1)}, I_k^{(2)}, \overline{I_k^{(1)}}, \overline{I_k^{(2)}}, I_k^{(0)})$.

2.4 Polar Coding

Now we proceed to describe the polar coding algorithm. Let $n = 2^k$ be the code length. Let B_n denote the $n \times n$ "bit reversal" permutation matrix in Arikan (2009), and let $G_n = G_1^{\otimes k}$ be the k-th power Kronecker product of the matrix

$$G_1 \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Write

$$Y^n = U^n B_n G_n, \quad Z^n = V^n B_n G_n. \tag{2.13}$$

Define $P_{(i)}$ to be the joint distribution $P((X^{i-1}, U^{i-1}, V^{i-1}), U_i, V_i)$. Following the analysis in Arikan (2009) (see also E. Sasoglu and Yeh (2009)), it is seen that

$$P_{(1)} = P^{--\ldots-}, \quad P_{(2)} = P^{--\ldots+}, \quad \ldots P_{(n)} = P^{++\ldots+}.$$

Theorem 2 assures that for almost all $P_{(i)}$'s, the triple $\mathcal{K}(P_{(i)})$ is close to one of the extremals when n is sufficiently large. We shall partition the indices of U^n into a frozen set or an information set, and similarly for V^n. More precisely, fix some small $\epsilon > 0$ and let $\mathcal{F}_1$ and $\mathcal{F}_2$ be subsets of $\{1, 2, \ldots, n\}$ defined as follows:

1. If $||\mathcal{K}(P_{(i)}) - (0, 0, 0)|| < \epsilon$, then $i \in \mathcal{F}_1$ and $i \in \mathcal{F}_2$;

2. If $||\mathcal{K}(P_{(i)}) - (\Delta_1, \Delta_1, 0)|| < \epsilon$, then $i \notin \mathcal{F}_1$ and $i \in \mathcal{F}_2$;

3. If $||\mathcal{K}(P_{(i)}) - (\Delta_2, 0, \Delta_2)|| < \epsilon$, then $i \in \mathcal{F}_1$ and $i \notin \mathcal{F}_2$;

4. If $||\mathcal{K}(P_{(i)}) - (\Delta_1 + \Delta_2, \Delta_1, \Delta_2)|| < \epsilon$, then $i \notin \mathcal{F}_1$ and $i \notin \mathcal{F}_2$;

5. If none of the above condition is satisfied, or more than one of the conditions are satisfied, then $i \notin \mathcal{F}_1$ and $i \notin \mathcal{F}_2$.

Let the joint distribution among x^n, u^n, v^n be specified by the product distribution $\prod_{i=1}^n p(x, y, z)$ and the transforms in (2.13), and denoted as $p_{X^n,U^n,V^n}(x^n, u^n, v^n)$; the marginals of $p_{X^n,U^n,V^n}(x^n, u^n, v^n)$ are written in a similar way; when clear from the context, we shall omit the subscripts. The randomized encoding function $f^{(n)} : \mathcal{X}^n \to \mathbb{F}_{q_1}^n \times \mathbb{F}_{q_2}^n$ is given as follows:

1. For each $i \in \mathcal{F}_1$, randomly set the value of u_i according to the uniform distribution over $\mathbb{F}_{q_1}$; similarly, for each $i \in \mathcal{F}_2$, randomly set the value of v_i according to the uniform distribution over $\mathbb{F}_{q_2}$; $u_{\mathcal{F}_1} \triangleq (u_i)_{i \in \mathcal{F}_1}$ and $v_{\mathcal{F}_2} \triangleq (v_i)_{i \in \mathcal{F}_2}$ are referred to as the frozen symbols of u^n and v^n, respectively, and are revealed to both the encoder and the decoder;

2. For $i = 1, 2, \ldots, n$, if $i \notin \mathcal{F}_1$, then u_i takes value $a \in \mathbb{F}_{q_1}$ with probability

$$\Pr(u_i = a) = \frac{p_{X^n,U^i,V^{i-1}}(x^n, (u^{i-1}, a), v^{i-1})}{p_{X^n,U^{i-1},V^{i-1}}(x^n, u^{i-1}, v^{i-1})}; \tag{2.14}$$

3. For $i = 1, 2, \ldots, n$, if $i \in \mathcal{F}_1$ and $i \notin \mathcal{F}_2$, then v_i takes value $a \in \mathbb{F}_{q_2}$ with probability

$$\Pr(v_i = a) = \frac{p_{X^n,U^{i-1},V^i}(x^n, u^{i-1}, (v^{i-1}, a))}{p_{X^n,U^{i-1},V^{i-1}}(x^n, u^{i-1}, v^{i-1})}; \tag{2.15}$$

else if $i \notin \mathcal{F}_1$ and $i \notin \mathcal{F}_2$, then v_i takes value $a \in \mathbb{F}_{q_2}$ with probability

$$\Pr(v_i = a) = \frac{p_{X^n,U^i,V^i}(x^n, u^i, (v^{i-1}, a))}{p_{X^n,U^i,V^{i-1}}(x^n, u^i, v^{i-1})}. \tag{2.16}$$

As we previously noticed, the symbols that are not frozen are (almost) completely determined by the source vectors, and thus we could also pick the most likely symbol (ML encoding) instead of using a randomized approach. However, we follow the randomized approach in Korada and Urbanke (2010) in order to simplify the analysis.

With only the first description (i.e., given u^n), the decoder forms $y^n = u^n B_n G_n$; it further applies ϕ_1 in Theorem 1 to each symbol of y^n individually and then concatenates the outputs as the reconstruction. Similarly, with only the second description (i.e., given v^n), the decoder forms $z^n = v^n B_n G_n$; it further applies ϕ_2 in Theorem 1 to z^n and then concatenates the outputs as the reconstruction. When both descriptions are available (i.e., given u^n and v^n), the decoder can apply ϕ_0 in Theorem 1 to (y_i, z_i), $i = 1, 2, \cdots, n$, and then concatenate the outputs as the reconstruction.

2.5 Performance Analysis

Now we are in a position to analyze the performance of the proposed polar coding scheme. Recall that the probability distribution $p(x^n, u^n, v^n)$ is induced by the product distribution $\prod_{i=1}^n p(x, y, z)$ and the transforms in (2.13). Clearly if our encoding procedure replicates this probability distribution, then at the decoder, we can recover y^n and z^n which are distributed jointly with x^n as $\prod_{i=1}^n p(x, y, z)$, and thus meet the distortion constraints (D_0, D_1, D_2) as given in Theorem 1; moreover, Theorem 2 and (2.7) ensure that some rate pairs on the dominant line of the EGC rate region are achieved asymptotically. However, the encoding procedure does not completely replicate $p(x^n, u^n, v^n)$, but only closely approximates it. As such, our goal is to show this approximation does not cause significant performance degradation in terms of the achieved distortions, i.e., the excess distortion can be bounded.

The following lemma Karzand and Teletar (2010) is needed in the proof, which is a consequence of Pinsker's inequality.

Lemma 2 *Let W denote the transition probability of a discrete channel, and $I(W)$ denote the mutual information between the input X and the output Y when X is uniformly distributed in the alphabet $\mathcal{X}$, then*

$$\sum_{x\in\mathcal{X}} \mathbb{E}\left|\frac{1}{|\mathcal{X}|} - p(x|Y)\right| \leq \sqrt{(2\log e)I(W)}. \tag{2.17}$$

Let $\hat{p}(x^n, u^n, v^n)$ be a probability distribution defined as follows

$$\hat{p}(u_i, v_i|x^n, u^{i-1}, v^{i-1}) = \begin{cases} \frac{1}{q_1 q_2} & i \in \mathcal{F}_1 \text{ and } i \in \mathcal{F}_2 \\ \frac{1}{q_2} p(u_i|x^n, u^{i-1}, v^{i-1}) & i \notin \mathcal{F}_1 \text{ and } i \in \mathcal{F}_2 \\ \frac{1}{q_1} p(v_i|x^n, u^{i-1}, v^{i-1}) & i \in \mathcal{F}_1 \text{ and } i \notin \mathcal{F}_2 \\ p(u_i, v_i|x^n, u^{i-1}, v^{i-1}) & i \notin \mathcal{F}_1 \text{ and } i \notin \mathcal{F}_2 \end{cases}. \tag{2.18}$$

For the coding procedure described in Section 2.4, we can write the resultant reconstruction distortion based on the first description as

$$\begin{aligned}\hat{D}_1 = &\sum_{u_{\mathcal{F}_1}, v_{\mathcal{F}_2}} \frac{1}{q_1^{|\mathcal{F}_1|} q_2^{|\mathcal{F}_2|}} \sum_{x^n} p(x^n) \\ &\sum_{\substack{u_{\mathcal{F}_1^c} \\ v_{\mathcal{F}_2^c}}} \prod_{i \notin \mathcal{F}_1} p(u_i|x^n, u^{i-1}, v^{i-1}) \cdot \prod_{\substack{i \in \mathcal{F}_1 \\ i \notin \mathcal{F}_2}} p(v_i|x^n, u^{i-1}, v^{i-1}) \cdot \prod_{\substack{i \notin \mathcal{F}_1 \\ i \notin \mathcal{F}_2}} p(v_i|x^n, u^{i}, v^{i-1}) \\ &\qquad \cdot d_1^{(n)}(x^n, \phi_1^{(n)}(u^n B_n G_n)) \\ &= \mathbb{E}_{\hat{p}} d_1^{(n)}(X^n, \phi_1^{(n)}(U^n B_n G_n)),\end{aligned}$$

where $d_1^{(n)}$ and $\phi_1^{(n)}$ are, respectively, the n-letter extensions of d_1 and ϕ_1 in Theorem 1. Similarly, the reconstruction distortion based on the second description and the reconstruction distortion based on both descriptions are given by

$$\begin{aligned}\hat{D}_2 &= \mathbb{E}_{\hat{p}} d_2^{(n)}(X^n, \phi_2^{(n)}(V^n B_n G_n)) \\ \hat{D}_0 &= \mathbb{E}_{\hat{p}} d_0^{(n)}(X^n, \phi_0^{(n)}(U^n B_n G_n, V^n B_n G_n)),\end{aligned}$$

where $d_0^{(n)}$, $d_2^{(n)}$, $\phi_0^{(n)}$, and $\phi_2^{(n)}$ are, respectively, the n-letter extensions of d_0, d_2, ϕ_0, and ϕ_2 in Theorem 1. It is clear that

$$\begin{aligned}D_1 &\geq \mathbb{E}_p d_1(X, \phi_1(Y)) = \mathbb{E}_p d_i^{(n)}(X^n, \phi_1^{(n)}(U^n B_n G_n)) \triangleq D_1^*, \\ D_2 &\geq \mathbb{E}_p d_2(X, \phi_2(Z)) = \mathbb{E}_p d_i^{(n)}(X^n, \phi_2^{(n)}(V^n B_n G_n)) \triangleq D_2^*, \\ D_0 &\geq \mathbb{E}_p d_0(X, \phi_0(Y, Z)) = \mathbb{E}_p d_i^{(n)}(X^n, \phi_0^{(n)}(U^n B_n G_n, V^n B_n G_n)) \triangleq D_0^*.\end{aligned}$$

Thus we only need to compare D_i^* and $\hat{D}_i$ for $i = 0, 1, 2$. For this purpose we can write

$$
\begin{aligned}
|\hat{D}_1 - D_1^*| &= |\mathbb{E}_{\hat{p}} d_1^{(n)}(X^n, \phi_1^{(n)}(U^n B_n G_n)) - \mathbb{E}_p d_1^{(n)}(X^n, \phi_1^{(n)}(U^n B_n G_n))| \\
&\leq d_{\max} \sum_{x^n, u^n, v^n} |\hat{p}(x^n, u^n, v^n) - p(x^n, u^n, v^n)|. \qquad (2.19)
\end{aligned}
$$

Similarly, we have

$$
|\hat{D}_2 - D_2^*| \leq d_{\max} \sum_{x^n, u^n, v^n} |\hat{p}(x^n, u^n, v^n) - p(x^n, u^n, v^n)|, \qquad (2.20)
$$

$$
|\hat{D}_0 - D_0^*| \leq d_{\max} \sum_{x^n, u^n, v^n} |\hat{p}(x^n, u^n, v^n) - p(x^n, u^n, v^n)|. \qquad (2.21)
$$

We can further write

$$
\begin{aligned}
&\sum_{x^n, u^n, v^n} p(x^n) |\hat{p}(u^n, v^n | x^n) - p(u^n, v^n | x^n)| \\
&= \sum_{x^n, u^n, v^n} p(x^n) \left| \prod_{i=1}^{n} p(u_i, v_i | x^n, u^{i-1}, v^{i-1}) - \prod_{i=1}^{n} \hat{p}(u_i, v_i | x^n, u^{i-1}, v^{i-1}) \right| \\
&= \sum_{x^n, u^n, v^n} p(x^n) \left| \sum_{i=1}^{n} (p(u_i, v_i | x^n, u^{i-1}, v^{i-1}) - \hat{p}(u_i, v_i | x^n, u^{i-1}, v^{i-1})) \right. \\
&\quad \left. \cdot \left(\prod_{j=1}^{i-1} p(u_j, v_j | x^n, u^{j-1}, v^{j-1}) \prod_{j=i+1}^{n} \hat{p}(u_j, v_j | x^n, u^{j-1}, v^{j-1}) \right) \right| \qquad (2.22) \\
&\leq \sum_{i=1}^{n} \sum_{x^n, u^n, v^n} p(x^n) \left| (p(u_i, v_i | x^n, u^{i-1}, v^{i-1}) - \hat{p}(u_i, v_i | x^n, u^{i-1}, v^{i-1})) \right. \\
&\quad \left. \cdot \left(\prod_{j=1}^{i-1} p(u_j, v_j | x^n, u^{j-1}, v^{j-1}) \prod_{j=i+1}^{n} \hat{p}(u_j, v_j | x^n, u^{j-1}, v^{j-1}) \right) \right| \\
&= \sum_{i=1}^{n} \sum_{x^n, u^n, v^n} p(x^n, u^{i-1}, v^{i-1}) |(p(u_i, v_i | x^n, u^{i-1}, v^{i-1}) - \hat{p}(u_i, v_i | x^n, u^{i-1}, v^{i-1}))|,
\end{aligned}
$$

where in (2.22) the following telescoping expansion Korada and Urbanke (2010) is applied

$$\prod_{i=1}^{n} A_i - \prod_{i=1}^{n} B_i = \sum_{i=1}^{n}(A_i - B_i)\prod_{j=1}^{i-1} A_j \prod_{j=i+1}^{n} B_j.$$

Now one can readily obtain the following upper bounds on the excess distortions

$$|\hat{D}_j - D_j^*| \leq d_{\max} \sum_{i=1}^{n} E_i, \quad j = 0, 1, 2,$$

where

$$E_i = \sum_{u_i=0}^{q_i-1} \sum_{v_i=0}^{q_2-1} \mathbb{E}_p |p(u_i, v_i|X^n, U^{i-1}, V^{i-1}) - \hat{p}(u_i, v_i|X^n, U^{i-1}, V^{i-1})|.$$

We need to bound E_i for each of the five cases as given in the definition of $\mathcal{F}_1$ and $\mathcal{F}_2$ (see Section 2.4).

- Case 1: $i \in \mathcal{F}_1$ and $i \in \mathcal{F}_2$. It can be shown that

$$\begin{aligned}
E_i &= \sum_{u_i=0}^{q_1-1}\sum_{v_i=0}^{q_2-1}\mathbb{E}_p\left|\frac{1}{q_1q_2} - p(u_i, v_i|X^n, U^{i-1}, V^{i-1})\right| \\
&= \sum_{x^n,u^i,v^i}\left|\frac{1}{q_1q_2}p(x^n, u^{i-1}, v^{i-1}) - p(x^n, u^i, v^i)\right| \\
&= \sum_{x^n,u^i,v^i}\left|\frac{1}{q_1q_2}p(x^n, u^{i-1}, v^{i-1}) - \frac{1}{q_2}p(x^n, u^i, v^{i-1}) + \frac{1}{q_2}p(x^n, u^i, v^{i-1}) - p(x^n, u^i, v^i)\right| \\
&\leq \sum_{x^n,u^i,v^i}\left|\frac{1}{q_1q_2}p(x^n, u^{i-1}, v^{i-1}) - \frac{1}{q_2}p(x^n, u^i, v^{i-1})\right| \qquad (2.23) \\
&+ \sum_{x^n,u^i,v^i}\left|\frac{1}{q_2}p(x^n, u^i, v^{i-1}) - p(x^n, u^i, v^i)\right| \\
&= \sum_{v_i=0}^{q_2-1}\frac{1}{q_2}\sum_{x^n,u^i,v^{i-1}}\left|\frac{1}{q_1}p(x^n, u^{i-1}, v^{i-1}) - p(x^n, u^i, v^{i-1})\right| \qquad (2.24) \\
&+ \sum_{x^n,u^i,v^i}\left|\frac{1}{q_2}p(x^n, u^i, v^{i-1}) - p(x^n, u^i, v^i)\right| \\
&= \sqrt{(2\log e)I(U_i; X^nU^{i-1}V^{i-1})} + \sqrt{(2\log e)I(V_i; X^nU^iV^{i-1})} \qquad (2.25) \\
&= \sqrt{(2\log e)I^{(1)}(P_{(i)})} + \sqrt{(2\log e)\overline{I^{(1)}}(P_{(i)})} \\
&\leq \sqrt{(2\log e)\overline{I^{(2)}}(P_{(i)})} + \sqrt{(2\log e)\overline{I^{(1)}}(P_{(i)})},
\end{aligned}$$

where (2.25) follows from Lemma 2.

- Case 2: $i \notin \mathcal{F}_1$ and $i \in \mathcal{F}_2$. We have

$$E_i = \sum_{v_i=0}^{q_2-1}\mathbb{E}_p\left|\frac{1}{q_2} - p(v_i|X^n, U^i, V^{i-1})\right| \leq \sqrt{(2\log e)\overline{I^{(1)}}(P_{(i)})}$$

- Case 3: $i \in \mathcal{F}_1$ and $i \notin \mathcal{F}_2$. We have

$$E_i = \sum_{u_i=0}^{q_1-1} \mathbb{E}_p \left| \frac{1}{q_1} - p(u_i|X^n, U^{i-1}, V^i) \right| \leq \sqrt{(2\log e)\overline{I^{(2)}}(P_{(i)})}$$

- Case 4: $i \notin \mathcal{F}_1$ and $i \notin \mathcal{F}_2$. For this case, we have $E_i = 0$.

- Case 5 is similar to case 4 and we have $E_i = 0$. Moreover, this case occurs with asymptotically zero probability.

The following two lemmas Arikan and Telatar (2009) and Karzand and Teletar (2010) provide a way to bound cases 1, 2, and 3; of course cases 4 and 5 are trivial.

Lemma 3 *If a stochastic process Z_n has the property that*

$$Z_{n+1} = \begin{cases} \leq 2Z_n & \textit{with probability } \frac{1}{2} \\ \leq Z_n^2 & \textit{with probability } \frac{1}{2} \end{cases}$$

then for any $\beta < \frac{1}{2}$, $\lim_{n\to\infty} Pr(Z_n \leq 2^{-2^{n\beta}}) = Pr(Z_\infty = 0)$.

Lemma 4 *For any q, there is a constant $\kappa(q)$ such that for any q-ary input channel W*

$$I(W^+) \leq 2I(W), \quad I(W^-) \leq \kappa(q)I(W)^2.$$

Thus indeed in the case where q_1 and q_2 are two non-identical primes, for any $\delta > 0$ and $0.5 > \beta > 0$, when n is sufficiently large, the proposed scheme achieves rate-distortion tuple $(R_1+\delta, R_2+\delta, D_0^*+\epsilon, D_1^*+\epsilon, D_2^*+\epsilon)$, where ϵ is of order $O(2^{-n^\beta})$ and (R_1, R_2) is a rate pair on the dominant line of the EGC rate region.

2.6 The Difficulty for the Case When $q_1 = q_2$

For the case where q_1 and q_2 are two identical primes, the sources P^s also polarize, for which we have the following theorem.

Theorem 4 *Let $P(x, y, z)$ be a joint source-codebook distribution where q_1 and q_2 are two identical primes, and write $\Delta = \log(q_1) = \log(q_2)$. Let*

$$M \triangleq \{(0, 0, 0), (\Delta, \Delta, 0), (\Delta, 0, \Delta), (2\Delta, \Delta, \Delta), (\Delta, 0, 0)\}.$$

Then for any $\delta > 0$,

$$\lim_{k\to\infty} \frac{1}{2^k} \#\{s \in \{-, +\}^k : ||\mathcal{K}(P^s) - M|| \geq \delta\} = 0. \tag{2.26}$$

The proof of Theorem 4 is almost the same as that of Theorem 2 and thus is omitted. Note that we have an additional extremal case $(\Delta, 0, 0)$ when $q_1 = q_2$, corresponding to the scenario where two descriptions jointly provide full information whereas each individual description by itself is useless. From a coding perspective, this case means any one symbol in the two descriptions can be set arbitrarily, and the other is then fully determined.

Although the sources also polarize when $q_1 = q_2$, to show that minimum sum-rate pair can be achieved, we also need to bound the excess distortions caused by the mismatch of the distributions as in Chapter 2.5. This however appears non-trivial precisely due to the additional extremal case. We suspect that in fact this additional extremal case appears only with asymptotically zero probability, and thus does not contribute meaningfully to the overall distortions.

Chapter 3

Polar Codes for Multiple Descriptions with Rate Splitting

The joint polarization scheme discussed in the previous Chapter can achieve a certain rate pair on the dominant line of the EGC rate region. However, this rate pair is determined by the coding scheme instead of being a design choice. Motivated by this observation, in this Chapter we propose a different polar coding scheme based on the rate splitting method, which can achieve the entire EGC rate region.

3.1 Rate Splitting

We shall only consider rate pairs on the dominant line of the EGC rate region since every rate pair $(R'_1, R'_2) \in \mathcal{J}(P)$ is dominated, in a componentwise manner, by some rate pair $(R_1, R_2) \in \mathcal{F}(P)$. Note that the dominant line $\mathcal{F}(P)$ has two endpoints: $V_1 = (I^{(1)}(P), \overline{I^{(1)}}(P))$ (i.e., $(I(X;Y), I(Z;X,Y))$) and $V_2 = (\overline{I^{(2)}}(P), I^{(2)}(P))$ (i.e.,

$(I(Y;X,Z), I(X;Z)))$. It is clear that these two endpoints are achievable via successive encoding. Specifically, for V_1, we can first encode X into Y at rate $I^{(1)}(P)$, then view (X,Y) as a joint source and further encode it into Z at rate $\overline{I^{(1)}}(P)$. The successive encoding scheme for V_2 is similar with the roles of Y and Z switched. Apparently, one can achieve other rate pairs on the dominant line of the EGC rate region by timesharing the encoding schemes for the two endpoints.

It turns out that a general rate pair on the dominant line can also be achieved via successive encoding without timesharing. Indeed, it is shown in (Y. Zhang, 2012) that, for every $(R_1, R_2) \in \mathcal{F}(P)$, there exist random variables $\hat{Y}$ and $\tilde{Y}$ (jointly distributed with (X,Y,Z)) as well as a deterministic function ψ such that:

- $\hat{Y}$ and $\tilde{Y}$ are independent;
- $Y = \psi(\hat{Y}, \tilde{Y})$;
- $(X,Z) \leftrightarrow Y \leftrightarrow (\hat{Y}, \tilde{Y})$ form a Markov chain;
- $R_1 = I(X;\hat{Y}) + I(X,\hat{Y},Z;\tilde{Y})$ and $R_2 = I(X,\hat{Y};Z)$.

Now one can readily see that (R_1, R_2) can be achieved by the following successive encoding scheme: first encode X into $\hat{Y}$ at rate $I(X;\hat{Y})$, then view $(X,\hat{Y})$ as a joint source and encode it into Z at rate $I(X,\hat{Y};Z)$, and finally view $(X,\hat{Y},Z)$ as a joint source and encode it into $\tilde{Y}$. Here we essentially split the first description, i.e., Y, into two sub-descriptions $\hat{Y}$ and $\tilde{Y}$ of rate $R_{1,1}$ and rate $R_{1,2}$, respectively, where $R_{1,1} = I(X;\hat{Y})$ and $R_{1,2} = I(X,\hat{Y},Z;\tilde{Y})$; moreover, given $\hat{Y}$ and $\tilde{Y}$, one can recover Y by using the deterministic function ψ. Note that three successive encoding steps are needed for a general point on the dominant line of the EGC rate region whereas only two encoding steps are needed for the endpoints.

Let (R_1, R_2) be a rate pair on the domain face of EGC region, then we must have

$$R_1 + R_2 = I(X;Y,Z) + I(Y;Z),$$
$$R_1 \in [I(X;Y), I(X,Z;Y)],$$
$$R_2 \in [I(X;Z), I(X,Y;Z)].$$

$A_1 = (I(X;Y), I(X,Y;Z))$ and $A_2 = (I(X,Z;Y), I(X;Z))$ are the two corner points of the domain face. It is obvious the corner points can be achieved by successive encoding scheme. Let us consider A_1 as an example. The code book of Y contains $2^{nI(X;Y)}$ sequences of y^n which are generated according to $p(y)$. The code book of Z contains $2^{nI(X,Z;Y)}$ sequences of z^n which are generated according to $p(z)$. When the encoder receives the source sequence x^n, it finds a y^n sequence that is joint typical with the source sequence and sends the corresponding index as the firs description. The encoder then finds a z^n sequence that is joint typical with (x^n, y^n) and sends the corresponding index as the second description.

For other points on the domain face, we introduce an auxiliary random variable $\hat{Y}$ such that $\hat{Y} - Y - (X, Z)$. We can find a class of transition probability $p(\hat{y}|y)$ such that $I(Y;\hat{Y})$ varies from 0 to $H(Y)$. Hence for $\hat{Y}$, the point on the domain face with

$$R_1 = I(X;\hat{Y}) + I(X,Z;Y|\hat{Y}),$$
$$R_2 = I(X,\hat{Y};Z),$$

is achievable. From the above expressions, we can see that by introducing the auxiliary variable $\hat{Y}$ the encoding rate of the first description is split into two parts. Intuitively, we have three code books for $\hat{Y}$, Z and Y receptively. The code book for $\hat{Y}$ contains

$2^{I(X;\hat{Y})}$ sequences which are generated according to $p(\hat{y})$. The code book for Z contains $2^{I(X,\hat{Y};Z)}$ sequences which are generated according to $p(z)$. The code books for Y are conditional code books. For each $\hat{y}^n$ sequences, generate $2^{I(X,Z;Y|\hat{Y})}$ sequence of y^n according to $p(\hat{y}|y)$ and form the sub-code book of Y. When the encoder receives the source sequence x^n, it finds a $\hat{y}^n$ sequence that is joint typical with x^n. Let m_{11} denote the corresponding index. The encoder then finds a z^n sequence that is joint typical with $(x^n, \hat{y}^n(m_{11}))$. Let m_2 denote the corresponding index. Finally, the encoder finds a y^n sequence in m_{11}th sub-code book of Y that is joint typical with $(x^n, \hat{y}^n(m_{11}), z^n(m_2))$. Let m_{12} denote the index in the sub-code book. Hence, the first description is formed by (m_{11}, m_{12}) and the second description is m_2. It can be noticed that by introducing $\hat{Z}$ instead of $\hat{Y}$ such that $\hat{Z} - Z - (X, Y)$ we can split rate R_2. Additional, by applying functional representation lemma we can replace the conditional code books of Y by a code book of new auxiliary variable $\tilde{Y}$. By functional representation lemma, there exists a variable $\tilde{Y}$ and function $f : \tilde{\mathcal{Y}} \times \hat{\mathcal{Y}} \to \mathcal{Y}$ such that $\tilde{Y}$ and $\hat{Y}$ are independent, $Y = f(\tilde{Y}, \hat{Y})$ and $\tilde{Y} - (\hat{Y}, Y) - (X, Z)$ form a Markov chain. Hence, the rate pair

$$
\begin{aligned}
R_1 &= I(X;\hat{Y}) + I(X, Z;\tilde{Y}), \\
R_2 &= I(X, \hat{Y}; Z).
\end{aligned}
$$

is achievable by successive encoding with the encoding order $\hat{Y}$, Z, $\tilde{Y}$. At the receiver end, y^n is reconstructed by $y_i = f(\hat{y}_i, \tilde{y}_i)$.

3.2 Polarization and Polar Codes

Considering a distribution $P_s(x, \hat{y}, z, \tilde{y})$ in the alphabet $\mathcal{X} \times \mathbb{F}_{\hat{q}_1} \times \mathbb{F}_{q_2} \times \mathbb{F}_{\tilde{q}_1}$ as described in section 3.1 with uniform marginals $P_s(\hat{y})$, $P_s(z)$ and $P_s(\tilde{y})$, define

$$I(W_1) = I(X; \hat{Y}),$$
$$I(W_2) = I(X, \hat{Y}; Z),$$
$$I(W_3) = I(X, \hat{Y}, Z; \tilde{Y}).$$

where $W_1(x, \hat{y}) = P_s(x, \hat{y}), W_2((x, \hat{y}), z) = P_s(x, \hat{y}, z)$ and $W_3((x, \hat{y}, z), \tilde{y}) = P_s(x, \hat{y}, z, \tilde{y})$.

Let $(X_1, \hat{Y}_1, Z_1, \tilde{Y}_1)$ and $(X_2, \hat{Y}_2, Z_2, \tilde{Y}_2)$ be two independent copies, and

$$\hat{Y}_1 = \hat{U}_1 + \hat{U}_2, \quad \hat{Y}_2 = \hat{U}_2,$$
$$Z_1 = V_1 + V_2, \quad Z_2 = V_2,$$
$$\tilde{Y}_1 = \tilde{U}_1 + \tilde{U}_2, \quad \tilde{Y}_2 = \tilde{U}_2,$$

where + is modulo addition, then we have the following equalities

$$
\begin{aligned}
2I(W_1) &= I(X_1, X_2; \hat{Y}_1, \hat{Y}_2) = I(X_1, X_2; \hat{U}_1, \hat{U}_2) \\
&= I(X_1, X_2; \hat{U}_1) + I(X_1, X_2; \hat{U}_2 | \hat{U}_1) \\
&= I(X_1, X_2; \hat{U}_1) + I(X_1, X_2, \hat{U}_1; \hat{U}_2), \\
2I(W_2) &= I(X_1, X_2, \hat{Y}_1, \hat{Y}_2; Z_1, Z_2) = I(X_1, X_2, \hat{Y}_1, \hat{Y}_2; V_1, V_2) \\
&= I(X_1, X_2, \hat{Y}_1, \hat{Y}_2; V_1) + I(X_1, X_2, \hat{Y}_1, \hat{Y}_2; V_2 | V_1) \\
&= I(X_1, X_2, \hat{Y}_1, \hat{Y}_2; V_1) + I(X_1, X_2, \hat{Y}_1, \hat{Y}_2, V_1; V_2), \\
&= I(X_1, X_2, \hat{U}_1, \hat{U}_2; V_1) + I(X_1, X_2, \hat{U}_1, \hat{U}_2, V_1; V_2), \\
2I(W_3) &= I(X_1, X_2, \hat{Y}_1, \hat{Y}_2, Z_1, Z_1; \tilde{Y}_1, \tilde{Y}_2) = I(X_1, X_2, \hat{Y}_1, \hat{Y}_2, Z_1, Z_1; \tilde{U}_1, \tilde{U}_2) \\
&= I(X_1, X_2, \hat{Y}_1, \hat{Y}_2, Z_1, Z_1; \tilde{U}_1) + I(X_1, X_2, \hat{Y}_1, \hat{Y}_2, Z_1, Z_1; \tilde{U}_2 | \tilde{U}_1) \\
&= I(X_1, X_2, \hat{Y}_1, \hat{Y}_2, Z_1, Z_1; \tilde{U}_1) + I(X_1, X_2, \hat{Y}_1, \hat{Y}_2, Z_1, Z_1, \tilde{U}_1; \tilde{U}_2) \\
&= I(X_1, X_2, \hat{U}_1, \hat{U}_2, V_1, V_1; \tilde{U}_1) + I(X_1, X_2, \hat{U}_1, \hat{U}_2, V_1, V_1, \tilde{U}_1; \tilde{U}_2).
\end{aligned}
$$

The random variables $\hat{U}_1$ and $\hat{U}_2$ are uniformly distributed in alphabet $\mathbb{F}_{\hat{q}_1}$; V_1 and V_2 are uniformly distributed in alphabet $\mathbb{F}_{q_2}$; $\tilde{U}_1$ and $\tilde{U}_2$ are uniformly distributed in alphabet $\mathbb{F}_{\tilde{q}_1}$. Here, we do not need to impose that $\hat{q}_1$, q_2 and $\tilde{q}_1$ are different.

Definition 3 *Given a joint distribution $W_1(x, \hat{y})$ in the alphabet $\mathcal{X} \times \mathbb{F}_{\hat{q}_1}$, define*

$$
\begin{aligned}
W_1^-((x_1, x_2), \hat{u}_1) &= \sum_{\hat{u}_2} W_1(x_2, \hat{u}_2) W_1(x_1, \hat{u}_1 + \hat{u}_2), \\
W_1^+((x_1, x_2, \hat{u}_1), \hat{u}_2) &= W_1(x_2, \hat{u}_2) W_1(x_1, \hat{u}_1 + \hat{u}_2).
\end{aligned}
$$

Similarly, given $W_2((x,\hat{y}),z)$ in the alphabet $\mathcal{X}\times\mathbb{F}_{\hat{q}_1}\times\mathbb{F}_{q_2}$, define

$$W_2^-((x_1,x_2,\hat{u}_1,\hat{u}_2),v_1)=\sum_{v_2}W_2(x_2,\hat{u}_2,v_2)W_2(x_1,\hat{u}_1+\hat{u}_2,v_1+v_2),$$

$$W_2^+((x_1,x_2,\hat{u}_1,\hat{u}_2,v_1),v_2)=W_2(x_2,\hat{u}_2,v_2)W_2(x_1,\hat{u}_1+\hat{u}_2,v_1+v_2).$$

Given $W_3((x,\hat{y},z),\tilde{y})$ in the alphabet $\mathcal{X}\times\mathbb{F}_{\hat{q}_1}\times\mathbb{F}_{q_2}\times\mathbb{F}_{\tilde{q}_1}$, define

$$W_3^-((x_1,x_2,\hat{u}_1,\hat{u}_2,v_1,v_2),\tilde{u}_1)=\sum_{v_2}W_3(x_2,\hat{u}_2,v_2,\tilde{u}_2)W_3(x_1,\hat{u}_1+\hat{u}_2,v_1+v_2,\tilde{u}_1+\tilde{u}_2)$$

$$W_3^+((x_1,x_2,\hat{u}_1,\hat{u}_2,v_1,v_2,\tilde{u}_1),\tilde{u}_2)=W_3(x_2,\hat{u}_2,v_2,\tilde{u}_2)W_3(x_1,\hat{u}_1+\hat{u}_2,v_1+v_2,\tilde{u}_1+\tilde{u}_2).$$

Note that $I(W_i^-)\leq I(W_i)\leq I(W_i^+)$, $I(W_i^-)+I(W_i^+)=2I(W_i)$ for $i=1,2,3$. The next corollary followed from lemma 1 and the mappings $(\hat{U}_1,\hat{U}_2)\rightarrow(\hat{Y}_1,\hat{Y}_2)$, $(V_1,V_2)\rightarrow(Z_1,Z_2)$ and $(\tilde{U}_1,\tilde{U}_2)\rightarrow(\tilde{Y}_1,\tilde{Y}_2)$ are bijection.

Corollary 3 *For any $\epsilon>0$ there exists a $\delta>0$ such that if $W_1(x,\hat{y})$ is a joint distribution with $I(W_1^+)-I(W_1)\leq\delta$, then $I(W_1)\notin(\epsilon,\Delta_1'-\epsilon)$; similarly if $W_2(x,\hat{y},z)$ is a joint distribution with $I(W_2^+)-I(W_2)\leq\delta$, then $I(W_2)\notin(\epsilon,\Delta_2'-\epsilon)$; $W_3(x,\hat{y},z,\tilde{y})$ is a joint distribution with $I(W_3^+)-I(W_3)\leq\delta$, then $I(W_3)\notin(\epsilon,\Delta_1''-\epsilon)$ where $\Delta_1'=\log\hat{q}_1$ and $\Delta_1''=\log\tilde{q}_1$.*

Similarly to previous section, let $B_1,B_2,\ldots,$ be an i.i.d. sequence of random variables taking values in the set $\{-,+\}$, with $\Pr(B_i=-)=\Pr(B_i=+)=0.5$. Define random process $\{W_{ik}:k\geq 0\}$, $i=1,2,3$ by

$$W_{i0}=W_i,\quad W_{ik}=W_{i,k-1}^{B_k},\ k\geq 1. \tag{3.1}$$

Now define the random process $\{I_{ik} : k \geq 0\}$, $i = 1, 2, 3$ as

$$I_{ik} = I(P_{ik}), \quad i = 1, 2, 3.$$

The next theorem followed from Corollary 3 and $\{I_{ik} : k \geq 0\}$, $i = 1, 2, 3$ are bounded martingales.

Theorem 5 $\lim_{k\to\infty} I_{1k} \in \{0, \Delta_1'\}$, $\lim_{k\to\infty} I_{2k} \in \{0, \Delta_2\}$ *and* $\lim_{k\to\infty} I_{3k} \in \{0, \Delta_1''\}$.

We can also use the randomized polar coding algorithm to facilitate analysis. Write

$$\hat{Y}^n = \hat{U}^n B_n G_n, \qquad \tilde{Y}^n = \tilde{U}^n B_n G_n.$$

Let $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ be subsets of $\{1, 2, \cdots, n\}$ and fix some $\epsilon > 0$.

1. If $|I(X^n, \hat{U}^{i-1}; \hat{U}_i) - 0| < \epsilon$, then $i \in \mathcal{F}_1$, otherwise $i \notin \mathcal{F}_1$;

2. If $|I(X^n, \hat{U}^n, V^{i-1}; V_i) - 0| < \epsilon$, then $i \in \mathcal{F}_2$, otherwise $i \notin \mathcal{F}_2$;

3. If $|I(X^n, \hat{U}^n, V^n, \tilde{U}_{i-1}; \tilde{U}_i) - 0| < \epsilon$, then $i \in \mathcal{F}_3$, otherwise $i \notin \mathcal{F}_3$;

Fix vector $\hat{u}'^{|\mathcal{F}_1|} \in \mathbb{F}_{\hat{q}_1}^{|\mathcal{F}_1|}$, $v'^{|\mathcal{F}_2|} \in \mathbb{F}_{q_2}^{|\mathcal{F}_2|}$ and $\tilde{u}'^{|\mathcal{F}_3|} \in \mathbb{F}_{\tilde{q}_1}^{|\mathcal{F}_3|}$. The encoding function is given as follows.

1. Set the frozen symbols in $\hat{u}^n$ to be $\hat{u}'^{|\mathcal{F}_1|}$, the frozen symbols in v^n to be $v'^{|\mathcal{F}_2|}$ and the frozen symbols in $\tilde{u}^n$ to be $\tilde{u}'^{|\mathcal{F}_1|}$.

2. For $i = 1, 2, \ldots, n$, if $i \notin \mathcal{F}_1$, then $\hat{u}_i$ takes value $a \in \mathbb{F}_{\hat{q}_1}$ with probability

$$\Pr(\hat{u}_i = a) = \frac{P_{X^n, \hat{U}^i}(x^n, \hat{u}^{i-1}, a)}{P_{X^n, \hat{U}^{i-1}}(x^n, \hat{u}^{i-1})};$$

3. For $i = 1, 2, \ldots, n$, if $i \notin \mathcal{F}_2$, then v_i takes value $a \in \mathbb{F}_{q_2}$ with probability

$$\Pr(v_i = a) = \frac{P_{X^n,\hat{U}^n,V^i}(x^n, \hat{u}^n, v^{i-1}, a)}{P_{X^n,\hat{U}^n,V^{i-1}}(x^n, \hat{u}^n, v^{i-1})};$$

4. For $i = 1, 2, \ldots, n$, if $i \notin \mathcal{F}_3$, then $\tilde{u}_i$ takes value $a \in \mathbb{F}_{\tilde{q}_1}$ with probability

$$\Pr(\tilde{u}_i = a) = \frac{P_{X^n,\hat{U}^n,V^i,\tilde{U}^n}(x^n, \hat{u}^n, v^{i-1}, a)}{P_{X^n,\hat{U}^n,V^{i-1}}(x^n, \hat{u}^n, v^{i-1})}.$$

3.3 Performance Under Randomized Frozen Symbols

For simplicity, we do not distinguish $\hat{q}_1$, q_2 and $\tilde{q}_1$ in this section, since there is no essential difference in the analysis. Let $\hat{p}(x^n, u^i, v^i)$ be a probability distribution defined as follows

$$\hat{p}(\hat{u}_i|x^n, \hat{u}^{i-1}) = \begin{cases} 1/q & i \in \mathcal{F}_1 \\ p(\hat{u}_i|x^n, \hat{u}^{i-1}) & i \notin \mathcal{F}_1 \end{cases}$$

$$\hat{p}(v_i|x^n, \hat{u}^n, v^{i-1}) = \begin{cases} 1/q & i \in \mathcal{F}_2 \\ p(v_i|x^n, \hat{u}^n, v^{i-1}) & i \notin \mathcal{F}_2 \end{cases}$$

$$\hat{p}(\tilde{u}_i|x^n, \hat{u}^n, v^n, \tilde{u}^{i-1}) = \begin{cases} 1/q & i \in \mathcal{F}_3 \\ p(\tilde{u}_i|x^n, \hat{u}^n, v^n, \tilde{u}^{i-1}) & i \notin \mathcal{F}_3 \end{cases}$$

Follow the coding procedure, we can write the resultant distortion D_1 as

$$\hat{D}_1 = \sum_{\hat{u}'^{\mathcal{F}_1}} \frac{1}{q^{|\mathcal{F}_1|}} \sum_{v'^{\mathcal{F}_2}} \frac{1}{q^{|\mathcal{F}_2|}} \sum_{\tilde{u}'^{\mathcal{F}_3}} \frac{1}{q^{|\mathcal{F}_3|}} \sum_{x^n} p(x^n) \sum_{\hat{u}^{n-|\mathcal{F}_1|}} \prod_{i \notin \mathcal{F}_1} p(\hat{u}_i|x^n, \hat{u}^{i-1}) \sum_{v^{n-|\mathcal{F}_2|}} \prod_{i \notin \mathcal{F}_2} p(v_i|x^n, \hat{u}^i, v^{i-1})$$
$$\sum_{\tilde{u}^{n-|\mathcal{F}_3|}} \prod_{i \notin \mathcal{F}_3} p(\tilde{u}_i|x^n, \hat{u}^i, v^i, \tilde{u}^{i-1}) d_1(x^n, \phi_1(f(\hat{u}^n B_n G_n, \tilde{u}^n B_n G_n)))$$
$$= \mathbb{E}_{\hat{p}} d_1(x^n, \phi_1(f(\hat{u}^n B_n G_n, \tilde{u}^n B_n G_n)))$$

Similarly, we have

$$\hat{D}_2 = \mathbb{E}_{\hat{p}} d_2(x^n, \phi_2(v^n B_n G_n))$$
$$\hat{D}_0 = \mathbb{E}_{\hat{p}} d_0(x^n, \phi_0(f(\hat{u}^n B_n G_n, \tilde{u}^n B_n G_n), v^n B_n G_n)).$$

It is clear that

$$D_1 = \mathbb{E}_{p_s} d_1(x^n, \phi_1(f(\hat{u}^n B_n G_n, \tilde{u}^n B_n G_n)))$$
$$D_2 = \mathbb{E}_{p_s} d_2(x^n, \phi_2(v^n B_n G_n))$$
$$D_0 = \mathbb{E}_{p_s} d_0(x^n, \phi_0(f(\hat{u}^n B_n G_n, \tilde{u}^n B_n G_n), v^n B_n G_n))$$

Now let us compare D_i and $\hat{D}_i$ for $i = 0, 1, 2$.

$$|\hat{D}_1 - D_1| = \mathbb{E}_{\hat{p}} d_1(x^n, \phi_1(f(\hat{u}^n B_n G_n, \tilde{u}^n B_n G_n))) - \mathbb{E}_{p_s} d_1(x^n, \phi_1(f(\hat{u}^n B_n G_n, \tilde{u}^n B_n G_n)))$$
$$\leq d_{\max} \sum_{x^n, \hat{u}^n, v^n, \tilde{u}^n} |\hat{p}(x^n, \hat{u}^n, v^n, \tilde{u}^n) - p_s(x^n, \hat{u}^n, v^n, \tilde{u}^n)|$$

Similarly, we have

$$|\hat{D}_2 - D_2| \leq d_{\max} \sum_{x^n,\hat{u}^n,v^n,\tilde{u}^n} |\hat{p}(x^n,\hat{u}^n,v^n,\tilde{u}^n) - p_s(x^n,\hat{u}^n,v^n,\tilde{u}^n)|$$

$$|\hat{D}_0 - D_0| \leq d_{\max} \sum_{x^n,\hat{u}^n,v^n,\tilde{u}^n} |\hat{p}(x^n,\hat{u}^n,v^n,\tilde{u}^n) - p_s(x^n,\hat{u}^n,v^n,\tilde{u}^n)|.$$

In the following, we omit the subscript 's' for simplicity.

$$\begin{aligned}
&\hat{p}(x^n,\hat{u}^n,v^n,\tilde{u}^n) - p(x^n,\hat{u}^n,v^n,\tilde{u}^n) \\
=& \sum_{x^n,\hat{u}^n,v^n,\tilde{u}^n} p(x^n)|\hat{p}(\hat{u}^n,v^n,\tilde{u}^n|x^n) - p(\hat{u}^n,v^n,\tilde{u}^n|x^n)| \\
=& \sum_{x^n,\hat{u}^n,v^n,\tilde{u}^n} p(x^n)\Big|\prod_{i=1}^{n}\hat{p}(\hat{u}_i|x^n,\hat{u}^{i-1})\prod_{i=1}^{n}\hat{p}(v_i|x^n,\hat{u}^n,v^{i-1})\prod_{i=1}^{n}\hat{p}(\tilde{u}_i|x^n,\hat{u}^n,v^n,\tilde{u}^{i-1}) \\
&\quad - \prod_{i=1}^{n}p(\hat{u}_i|x^n,\hat{u}^{i-1})\prod_{i=1}^{n}p(v_i|x^n,\hat{u}^n,v^{i-1})\prod_{i=1}^{n}p(\tilde{u}_i|x^n,\hat{u}^n,v^n,\tilde{u}^{i-1})\Big|
\end{aligned}$$

Let

$$\begin{aligned}
A_i =& \hat{p}(\hat{u}_i|x^n,\hat{u}^{i-1}) \quad i=1,2,\ldots n, \\
A_{i+n} =& \hat{p}(\hat{v}_i|x^n,\hat{u}^n,v^{i-1}) \quad i=1,2,\ldots n, \\
A_{i+2n} =& \hat{p}(\tilde{u}_i|x^n,\hat{u}^n,v^n,\tilde{u}^{i-1}) \quad i=1,2,\ldots n,
\end{aligned}$$

and

$$\begin{aligned}
B_i =& p(\hat{u}_i|x^n,\hat{u}^{i-1}) \quad i=1,2,\ldots n, \\
B_{i+n} =& p(\hat{v}_i|x^n,\hat{u}^n,v^{i-1}) \quad i=1,2,\ldots n, \\
B_{i+2n} =& p(\tilde{u}_i|x^n,\hat{u}^n,v^n,\tilde{u}^{i-1}) \quad i=1,2,\ldots n,
\end{aligned}$$

We have

$$\sum_{x^n,\hat{u}^n,v^n,\tilde{u}^n} |\hat{p}(x^n,\hat{u}^n,v^n,\tilde{u}^n) - p(x^n,\hat{u}^n,v^n,\tilde{u}^n)| = \sum_{x^n,u^n,v^n} p(x^n)|\sum_{i=1}^{3n}(A_i - B_i)\prod_{j=1}^{i-1} A_j \prod_{j=i+1}^{3n} B_j|$$

Define

$$E_i = \sum_{x^n,\hat{u}^n,v^n,\tilde{u}^n} p(x^n)|(A_i - B_i)\prod_{j=1}^{i-1} A_i \prod_{j=I+1}^{3n} B_i|$$

then the excess distortion is upper-bounded as

$$|\hat{D}_1 - D_1| \leq d_{\max}\sum_{i=1}^{3n} E_i.$$

We need to bound E_i for each case.

- Case 1: $1 \leq i \leq n$ and $i \in \mathcal{F}_1$,

$$E_i \leq \sum_{\hat{u}_i=0}^{q-1} \mathbb{E}_p|\frac{1}{q} - p(\hat{u}_i|x^n,\hat{u}^{i-1}))| \leq \sqrt{2(\log e)I(\hat{U}_i;X^N,\hat{U}^{i-1})}$$

- Case 2: $1 \leq i \leq n$ and $i \in \mathcal{F}_2$,

$$E_{i+n} \leq \sum_{v_i=0}^{q-1} \mathbb{E}_p|\frac{1}{q} - p(v_i|x^n,\hat{u}^n,v^{i-1}))| \leq \sqrt{2(\log e)I(V_i;X^N,\hat{U}^n,V^{i-1})}$$

- Case 3: $1 \leq i \leq n$ and $i \in \mathcal{F}_3$,

$$E_{i+2n} \leq \sum_{\tilde{u}_i=0}^{q-1} \mathbb{E}_p|\frac{1}{q} - p(\tilde{u}_i|x^n,u^n,v^n,\tilde{u}^{i-1}))| \leq \sqrt{2(\log e)I(\tilde{U}_i;X^N,\hat{U}^n,V^n,\tilde{U}^{i-1})}$$

- Case 4: $1 \leq i \leq n$, $i \notin \mathcal{F}_1$, $i \notin \mathcal{F}_1$ and $i \notin \mathcal{F}_3$. For this case, $E_i = 0$.

Lemma 3 and Lemma 4 can now be invoked to bound the cases 1, 2 and 3 and case 4 is trivial. Thus indeed for any $\delta > 0$ and $0.5 > \beta > 0$, when n is sufficiently large, the proposed scheme has rate $(R_1+\delta, R_2+\delta)$, and the distortion is $(D_0+\epsilon, D_1+\epsilon, D_2+\epsilon)$, where ϵ is of order $O(2^{-n^\beta})$ and (R_1, R_2) achieves rate pair described in section 3.1, and (D_0, D_1, D_2) is the corresponding distortion triple.

Chapter 4

Experimental Results

In this Chapter we implement the proposed polar coding schemes and present the simulation results for binary sources with the Hamming distortion measure (i.e., $\mathcal{X} = \mathcal{X}_0 = \mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ and $d_0 = d_1 = d_2 = d_H$, where $d_H(x, \hat{x}) = 0$ if $x = \hat{x}$ and $d_H(x, \hat{x}) = 1$ otherwise). The construction of polar codes is based on degraded merge given in I. Tal and Vardy (2012). The length n of the sequences is selected to be 2^7, 2^9, and 2^{11}. The empirical distortions are averaged over 100 runs. For the ease of implementation, we set the frozen symbols to zero and use ML encoding instead of randomized encoding. It is worth mentioning that in the experiments we fix the threshold ϵ (which is used to determine the frozen sets); as a consequence, the empirical rates and distortions can change with the block length. But it will be seen from the experimental results that the empirical rates and distortions approach to the theoretical limits as the block length increases.

4.1 Experimental Results for the Joint Polarization Scheme

For the joint polarization scheme, we consider four different cases. Note that, given the joint distribution $p(x, y, z)$, the theoretical limits of the sum rate $R_1^* + R_2^*$ and distortions (D_0^*, D_1^*, D_2^*) in Theorem 1 are, respectively, given by

$$R_1^* + R_2^* = I(X; Y, Z) + I(Y; Z), \tag{4.1}$$

$$D_0^* = Ed_H(X, \phi_0(Y, Z)), \tag{4.2}$$

$$D_1^* = Ed_H(X, \phi_1(Y)), \tag{4.3}$$

$$D_2^* = Ed_H(X, \phi_2(Z)). \tag{4.4}$$

As previously pointed out, though the joint polarization scheme can achieve a certain rate pair (R_1^*, R_2^*) on the dominant line of the EGC rate region (and consequently one can fix $R_1^* + R_2^*$ in advance), the exact values of R_1^* and R_2^* are determined by the algorithm instead of being a design choice.

1. Two independent binary descriptions.

   The source distribution is given by $p_X(0) = \frac{3}{4}$ and $p_X(1) = \frac{1}{4}$. The conditional distribution $p(y, z|x)$ and the joint distribution $p(x, y, z)$ are given in Table 4.1. It can be verified that $p(y, z) = p(y)p(z)$ for $y \in \mathbb{F}_2$ and $z \in \mathbb{F}_2$. Moreover, for

$y \in \mathbb{F}_2$ and $z \in \mathbb{F}_2$, we define

$$\phi_0(y, z) = y \cdot z, \tag{4.5}$$

$$\phi_1(y) = y, \tag{4.6}$$

$$\phi_2(z) = z. \tag{4.7}$$

According to (4.1)-(4.4), we have $R_1^* + R_2^* = 0.8113$ and $(D_0^*, D_1^*, D_2^*) = (0, 0.2500, 0.2500)$. In the experiments we set the threshold $\epsilon = 0.1$. The empirical rates $(\hat{R}_1, \hat{R}_2)$ (as well as $\hat{R}_1 + \hat{R}_2$) and distortions $(\hat{D}_0, \hat{D}_1, \hat{D}_2)$ obtained through simulation are shown in Table 4.2.

Table 4.1: $p(y, z|x)$ and $p(x, y, z)$

x, y, z	$0,0,0$	$0,0,1$	$0,1,0$	$0,1,1$	$1,0,0$	$1,0,1$	$1,1,0$	$1,1,1$
$p(y, z\|x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0	0	0	0	1
$p(x, y, z)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	0	0	$\frac{1}{4}$

Table 4.2: Experimental Results Based on the Joint Distribution $p(x, y, z)$ in Table 4.1

n	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_1 + \hat{R}_2$	$\hat{D}_0$	$\hat{D}_1$	$\hat{D}_2$
2^7	0.5234	0.5234	1.0468	0.005	0.2652	0.2649
2^9	0.4746	0.4746	0.9492	0.0025	0.2470	0.2532
2^{11}	0.4512	0.4512	0.9024	6.64×10^{-4}	0.2481	0.2499

2. Two dependent binary descriptions.

The source distribution is given by $p_X(0) = \frac{2}{3}$ and $p_X(1) = \frac{1}{3}$. The conditional distribution $p(y, z|x)$ and the joint distribution $p(x, y, z)$ are given in Table 4.3. We adopt the same definition of ϕ_0, ϕ_1, and ϕ_2 as in (4.5)-(4.7). In this case we have $R_1^* + R_2^* = 1.0000$ and $(D_0^*, D_1^*, D_2^*) = (0, 0.1666, 0.1666)$. In the experiments we set the threshold $\epsilon = 0.1$. The empirical rates $(\hat{R}_1, \hat{R}_2)$ (as well as $\hat{R}_1 + \hat{R}_2$) and distortions $(\hat{D}_0, \hat{D}_1, \hat{D}_2)$ obtained through simulation are shown in Table 4.4.

Table 4.3: $p(y, z|x)$ and $p(x, y, z)$

x, y, z	$0,0,0$	$0,0,1$	$0,1,0$	$0,1,1$	$1,0,0$	$1,0,1$	$1,1,0$	$1,1,1$
$p(y, z\|x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	0	0	1
$p(x, y, z)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	0	0	0	0	$\frac{1}{3}$

Table 4.4: Experimental Results Based on the Joint Distribution $p(x, y, z)$ in Table 4.3

n	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_1 + \hat{R}_2$	$\hat{D}_0$	$\hat{D}_1$	$\hat{D}_2$
2^7	0.6406	0.6406	1.2812	8.60×10^{-4}	0.1221	0.1243
2^9	0.5801	0.5801	1.1602	1.75×10^{-4}	0.1401	0.1389
2^{11}	0.5439	0.5439	1.0878	6.33×10^{-5}	0.1524	0.1525

Although we are unable to handle analytically the case where the alphabet sizes of the two descriptions are identical, the experimental results indicate that the empirical rates and distortions indeed approach the theoretical limits as the block length increases.

3. Two independent descriptions: one binary and one ternary.

   The source distribution is given by $p_X(0) = \frac{2}{3}$ and $p_X(1) = \frac{1}{3}$. The conditional distribution $p(y, z|x)$ and the joint distribution $p(x, y, z)$ are given in Table 4.5. It can be verified that $p(y, z) = p(y)p(z)$ for $y \in \mathbb{F}_3$ and $z \in \mathbb{F}_2$. Moreover, for

$y \in \mathbb{F}_3$ and $z \in \mathbb{F}_2$, we define

$$\phi_0(y,z) = \begin{cases} 0, & (y,z) = (0,0),(0,1),(1,0),(2,0) \\ 1, & (y,z) = (1,1),(2,1) \end{cases}, \tag{4.8}$$

$$\phi_1(y) = \begin{cases} 0, & y = 0 \\ 1, & y = 1,2 \end{cases}, \tag{4.9}$$

$$\phi_2(z) = z. \tag{4.10}$$

In this case we have $R_1^* + R_2^* = 0.9183$ and $(D_0^*, D_1^*, D_2^*) = (0, 0.333, 0.166)$. In the experiments we set the threshold $\epsilon = 0.3$. The empirical rates $(\hat{R}_1, \hat{R}_2)$ (as well as $\hat{R}_1 + \hat{R}_2$) and distortions $(\hat{D}_0, \hat{D}_1, \hat{D}_2)$ obtained through simulation are shown in Table 4.6.

Table 4.5: $p(y,z|x)$ and $p(x,y,z)$

x,y,z	$0,0,0$	$0,0,1$	$0,1,0$	$0,1,1$	$0,2,0$	$0,2,1$	$1,0,0$	$1,0,1$	$1,1,0$	$1,1,1$	$1,2,0$	$1,2,1$
$p(y,z\|x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$	0	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
$p(x,y,z)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$	0	0	0	0	$\frac{1}{6}$	0	$\frac{1}{6}$

Table 4.6: Experimental Results Based on the Joint Distribution $p(x,y,z)$ in Table 4.5

n	$\hat{R}_1$	$\hat{R}_1$	$\hat{R}_1 + \hat{R}_2$	$\hat{D}_0$	$\hat{D}_1$	$\hat{D}_2$
2^7	0.7306	0.5234	1.2540	0.0039	0.3332	0.1782
2^9	0.6160	0.4980	1.1140	0.0077	0.3330	0.1792
2^{11}	0.5502	0.4844	1.0346	0.0067	0.3286	0.1830

4. Two dependent descriptions: one binary and one ternary.

   The source distribution is given by $p_X(0) = \frac{7}{12}$ and $p_X(1) = \frac{5}{12}$. The conditional distribution $p(y,z|x)$ and the joint distribution $p(x,y,z)$ are given in Table 4.7.

We adopt the same definition of ϕ_0, ϕ_1, and ϕ_2 as in (4.8)-(4.10). In this case we have $R_1^* + R_2^* = 1.0939$ and $(D_0^*, D_1^*, D_2^*) = (0, 0.2500, 0.0833)$. In the experiments we set the threshold $\epsilon = 0.3$. The empirical rates $(\hat{R}_1, \hat{R}_2)$ (as well as $\hat{R}_1 + \hat{R}_2$) and distortions $(\hat{D}_0, \hat{D}_1, \hat{D}_2)$ obtained through simulation are shown in Table 4.8.

Table 4.7: $p(y, z|x)$ and $p(x, y, z)$

x, y, z	$0,0,0$	$0,0,1$	$0,1,0$	$0,1,1$	$0,2,0$	$0,2,1$	$1,0,0$	$1,0,1$	$1,1,0$	$1,1,1$	$1,2,0$	$1,2,1$
$p(y, z\|x)$	$\frac{3}{7}$	$\frac{1}{7}$	$\frac{3}{14}$	0	$\frac{3}{14}$	0	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
$p(x, y, z)$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{8}$	0	$\frac{1}{8}$	0	0	0	0	$\frac{5}{24}$	0	$\frac{5}{24}$

Table 4.8: Experimental Results Based on the Joint Distribution $p(x, y, z)$ in Table 4.7

n	$\hat{R}_1$	$\hat{R}_1$	$\hat{R}_1 + \hat{R}_2$	$\hat{D}_0$	$\hat{D}_1$	$\hat{D}_2$
2^7	0.7430	0.7035	1.4465	0.0038	0.2574	0.0945
2^9	0.6067	0.6758	1.2825	0.0034	0.2761	0.0941
2^{11}	0.5350	0.6535	1.1885	0.0038	0.2828	0.0973

4.2 Experimental Results for the Rate Splitting Scheme

The source distribution is given by $p_X(0) = \frac{2}{3}$ and $p_X(1) = \frac{1}{3}$. The conditional distribution $p(y, z|x)$ and the joint distribution $p(x, y, z)$ are given in Table 4.3. We adopt the same definition of ϕ_0, ϕ_1, and ϕ_2 as in (4.5)-(4.7). In this case we have $R_1^* + R_2^* = 1.0000$ and $(D_0^*, D_1^*, D_2^*) = (0, 0.1666, 0.1666)$. In the experiments we set the threshold $\epsilon = 0.1$.

We first consider the following corner point of the EGC rate region

$$R_1^* = I(X;Y) = 0.4591,$$

$$R_2^* = I(X,Y;Z) = 0.5409.$$

Note that no rate splitting is needed for corner points. The empirical rates $(\hat{R}_1, \hat{R}_2)$ (as well as $\hat{R}_1 + \hat{R}_2$) and distortions $(\hat{D}_0, \hat{D}_1, \hat{D}_2)$ obtained through simulation are shown in Table 4.9.

Table 4.9: Experimental Results Based on the Joint Distribution $p(x,y,z)$ in Table 4.3: A Corner Point

n	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_1 + \hat{R}_2$	$\hat{D}_0$	$\hat{D}_1$	$\hat{D}_2$
2^7	0.5859	0.6563	1.2422	7.03×10^{-4}	0.1354	0.1240
2^9	0.5469	0.6211	1.1680	2.14×10^{-4}	0.1432	0.1311
2^{11}	0.5155	0.5917	1.1072	1.16×10^{-4}	0.1518	0.1430

Now we consider the following rate pair on the dominant line of the EGC rate region

$$R_1^* = 0.524,$$

$$R_2^* = 0.476.$$

Note that rate splitting is need for this rate pair since it is not a corner point. We construct two random variables $\hat{Y}$ and $\tilde{Y}$ such that

- $(X,Z) \leftrightarrow Y \leftrightarrow (\hat{Y}, \tilde{Y})$ form a Markov chain,
- $\hat{Y}$ and $\tilde{Y}$ are independent with $p_{\hat{Y}}(0) = \frac{1}{2}$, $p_{\hat{Y}}(1) = \frac{1}{2}$, $p_{\tilde{Y}}(0) = p_{\tilde{Y}}(2) = \frac{1}{5}$, and $p_{\tilde{Y}}(1) = \frac{3}{5}$,

- $Y = \psi(\hat{Y}, \tilde{Y})$, where

$$\psi(\hat{y}, \tilde{y}) = \begin{cases} 0, & (\hat{y}, \tilde{y}) = (0,0), (0,1), (1,0) \\ 1, & (\hat{y}, \tilde{y}) = (0,2), (1,1), (1,2) \end{cases}.$$

It can be verified that

$$R^*_{1,1} \triangleq I(X; \hat{Y}) = 0.1366,$$

$$R^*_{1,2} \triangleq I(X, \hat{Y}, Z; \tilde{Y}) = 0.3874,$$

$$R^*_{1,1} + R^*_{1,2} = R^*_1,$$

$$I(X, \hat{Y}; Z) = R^*_2.$$

The empirical rates $(\hat{R}_{1,1}, \hat{R}_{1,2}, \hat{R}_2)$ (as well as $\hat{R}_1 + \hat{R}_2$ with $\hat{R}_1 \triangleq \hat{R}_{1,1} + \hat{R}_{1,2}$) and distortions $(\hat{D}_0, \hat{D}_1, \hat{D}_2)$ are shown in Table 4.10.

Table 4.10: Experimental Results Based on the Joint Distribution $p(x, y, z)$ in Table 4.3: A General Point

n	$\hat{R}_{11}$	$\hat{R}_{12}$	$\hat{R}_2$	$\hat{R}_1 + \hat{R}_2$	$\hat{D}_0$	$\hat{D}_1$	$\hat{D}_2$
2^7	0.2188	0.4354	0.6016	1.2558	0.0055	0.1382	0.1600
2^9	0.1895	0.3720	0.5605	1.1220	0.0043	0.1465	0.1822
2^{11}	0.1652	0.3118	0.5259	1.0029	0.0116	0.1558	0.2123

Chapter 5

Conclusion

We have considered polar coding for the multiple description problem and proposed two different polar coding schemes: the first one is based on joint polarization while the second one is based on the rate splitting method. It is shown that the first scheme is able to asymptotically achieve a certain rate pair on the dominant line of the EGC rate region when the alphabet sizes of the two auxiliary random variables are non-identical primes. For the second scheme, we show that it can achieve the entire EGC rate region. Simulation results for the proposed schemes are also provided.

Bibliography

Arikan, E. (2009). Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, **55**(7), 3051–3073.

Arikan, E. and Telatar, E. (2009). On the rate of channel polarization. In *IEEE International Symposium on Information Theory*, pages 1493–1495, Seoul, Korea.

E. Sasoglu, E. T. and Yeh, E. (2009). Polar codes for the two-user multiple-access channel. In *IEEE International Symposium on Information Theory*, pages 1493–1495, Seoul, Korea.

Gamal, A. E. and Cover, T. M. (1982). Achievable rates for multiple descriptions. *IEEE Transactions on Information Theory*, **28**, 851–857.

I. Tal, A. S. and Vardy, A. (2012). Constructing Polar Codes for Non-Binary Alphabets and MACs. In *IEEE International Symposium on Information Theory*, pages 2142–2146, Cambridage, MA.

J. Wang, J. Chen, L. Z. P. C. and Permuter, H. (2009). A random variable substitution lemma with applications to multiple description coding. preprint, arXiv:0909.3135.

Karzand, M. and Teletar, E. (2010). Polar codes for q-ary source coding. In *IEEE International Symposium on Information Theory*, pages 909–912, Austin, TX.

Korada, S. (2008). *Polar codes for channel and source coding*. Ph.D. thesis, EPFL, Lausanne, Switzerland.

Korada, S. and Urbanke, R. (2010). Polar codes are optimal for lossy source coding. *IEEE Transactions on Information Theory*, **56**, 1751–1768.

Sasoglu, E. (2010). An entropy inequality for q-ary random variables and its application to channel polarization. In *IEEE International Symposium on Information Theory*, pages 1360–1363, Austin, TX.

Y. Zhang, S. Dumitrescu, J. C. Z. S. (2012). LDGM-Based Multiple Description Coding for Finite Alphabet Sources. *IEEE Transaction on Communications*, **60**(12).