Fifty-Fifth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 3-6, 2017

Polar Coding for Multiple Descriptions Using Monotone Chain Rules

Alankrita Bhatt, Nadim Ghaddar
Electrical and Computer Engineering
University of California San Diego
{a2bhatt,nghaddar}@ucsd.edu

Lele Wang
Electrical Engineering
Stanford University and Tel Aviv University
wanglele@stanford.edu

***Abstract*—In this paper, we present a polar coding scheme for the multiple description coding problem. The proposed scheme improves upon the existing joint polarization based scheme by Shi, Song, Tian, Chen, and Dumitrescu and achieves the entire El Gamal–Cover inner bound for this problem.**

***Keywords*—Polar codes, Multiple description coding, El Gamal–Cover inner bound, Monotone chain rules**

I. Introduction

We consider the multiple description coding (MDC) problem, which was formulated by A. Gersho and H. S. Witsenhausen and initally studied by [1]–[3]; see also [4, Chapter 13]. Given a discrete memoryless source, we wish to generate two descriptions such that each description by itself can be used to reconstruct the source with some desired distortion and the two descriptions together can be used to reconstruct the source with a lower distortion. This is motivated by applications such as multimedia communication over networks, where the network suffers from data loss. By compressing the multimedia, such as movies, with multiple descriptions, users receiving any one description of the movie wish to reconstruct it to some acceptable quality and users receiving both descriptions can reconstruct it to a higher quality. More formally, consider a discrete memoryless source $X^N \sim \prod_{i=1}^N p_X(x_i)$ with three distortion measures $d_j \colon \mathcal{X} \times \hat{\mathcal{X}}_j \to [0, \infty)$, $j = 0, 1, 2$, as depicted in Fig. 1. Each encoder generates a description of the source sequence so that decoder 1 (and resp. 2) that only receives one description can reconstruct the source with distortion D_1 (and resp. D_2) and decoder 0 that receives both descriptions can recover the source with distortion D_0. We wish to find the optimal tradeoff between the description rate pair (R_1, R_2) and the distortion triple (D_0, D_1, D_2).

A $(2^{NR_1}, 2^{NR_2}, N)$ *multiple description code* consists of

- two encoders, where encoder 1 assigns an index $y(x^N) \in \{1, 2, \ldots, 2^{\lfloor NR_1 \rfloor}\} := [2^{NR_1}]$ and encoder 2 assigns an index $z(x^N) \in [2^{NR_2}]$ to each sequence $x^N \in \mathcal{X}^N$, and
- three decoders, where decoder 1 assigns an estimate $\hat{x}_1^N$ to each index y, decoder 2 assigns an estimate $\hat{x}_2^N$ to each index z, and decoder 0 assigns an estimate $\hat{x}_0^N$ to each index pair (y, z).

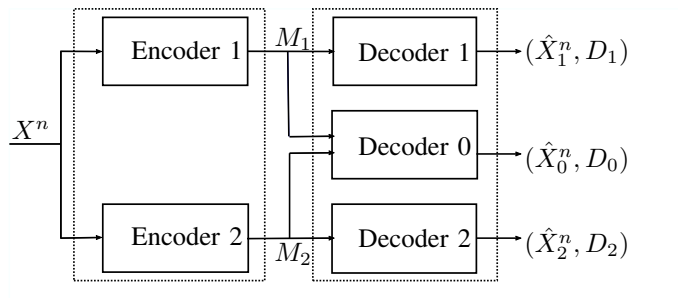

Fig. 1: The Multiple Description Coding problem

A rate–distortion quintuple $(R_1, R_2, D_0, D_1, D_2)$ is said to be *achievable* if there exists a sequence of $(2^{NR_1}, 2^{NR_2}, N)$ codes that satisfy

$$\limsup_{N \to \infty} \mathbb{E}[d_j(X^N, \hat{X}_j^N)] \le D_j, \quad j = 0, 1, 2.$$

where $d_j : X \times \hat{X}_j \to [0, d_{max}]$, $j = 0, 1, 2$, are bounded distortion measures. The optimal rate–distortion region $\mathcal{R}(D_0, D_1, D_2)$ is defined as the closure of the set of rate pairs (R_1, R_2) such that $(R_1, R_2, D_0, D_1, D_2)$ is achievable.

The optimal rate–distortion region for the MDC problem is not known in general. A number of random coding based achievability results have been proposed by El Gamal and Cover [5], Chen, Tian, Berger, and Hemami [6], Berger and Zhang [7], among others. A primitive component in these coding schemes is a joint typicality encoding that generates two descriptions from which we can obtain arbitrarily correlated reconstructions. While producing the best known achievability results, joint typicality encoding is nontrivial to implement in a time/space efficient manner, as it involves multiple codeword-sequence detection at the core of its operation. In this paper, we investigate how to implement joint typicality encoding at low-complexity using the recently invented polar codes [8]. To this end, we focus on the El Gamal–Cover (EGC) inner bound [5], an equivalent form of which is given as follows [9].

Theorem 1 (El Gamal–Cover inner bound)**.** *A rate pair (R_1, R_2) is achievable for the multiple description problem*

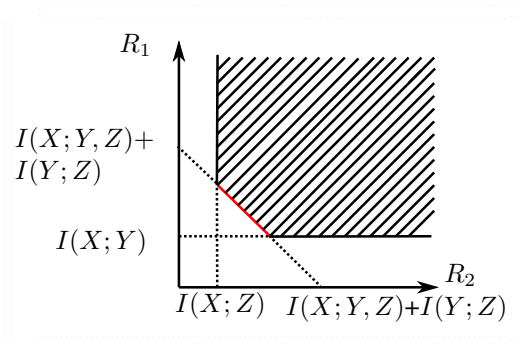

Fig. 2: El Gamal–Cover inner bound for a fixed pmf. The dominant face of this region is highlighted in red.

with distortion triple (D_0, D_1, D_2) *if*

$$\begin{aligned} R_1 &\geq I(X;Y), \\ R_2 &\geq I(X;Z), \\ R_1 + R_2 &\geq I(X;Y,Z) + I(Y;Z) \end{aligned} \tag{1}$$

for some conditional pmf $p(y,z|x)$ *and some deterministic mappings* $\phi_0 \colon \mathcal{Y} \times \mathcal{Z} \to \hat{\mathcal{X}}_0$, $\phi_1 \colon \mathcal{Y} \to \hat{\mathcal{X}}_1$, *and* $\phi_2 \colon \mathcal{Z} \to \hat{\mathcal{X}}_2$ *such that* $D_0 \geq \mathbb{E}[d_0(X, \phi_0(Y,Z))]$, $D_1 \geq \mathbb{E}[d_1(X, \phi_1(Y))]$, *and* $D_2 \geq \mathbb{E}[d_2(X, \phi_2(Z))]$.

Here Y and Z can be seen as the two descriptions representing the source X, and functions $\phi_j, j = 0,1,2$, are the reconstruction functions based on the available descriptions at each decoder. For a fixed $p(y,z|x)$ and functions $\phi_j, j = 0,1,2$, the subset of achievable rate pairs (R_1, R_2) that satisfy $R_1 + R_2 = I(X;Y,Z) + I(Y;Z)$ is called the *dominant face* of rate–distortion region, as illustrated in the red line of Fig. 2.

Invented by Arıkan in [8], polar codes are one of the most recent breakthroughs in coding theory. These codes achieve the symmetric capacity on any binary discrete memoryless channel and also have low encoding and decoding complexities. The basic idea behind polar coding is a phenomenon referred to as channel polarization: Given $N = 2^k$ independent copies of a channel, it is possible to synthesize N new channels, a fraction of which are near perfect (capacity 1) and the rest are near useless (capacity 0). Arıkan showed that the fraction of near perfect channels approaches the symmetric capacity of the original channel. Thus, one can transmit uncoded bits over the near perfect channels, and set the bits as constant over the near useless channels. In the past ten years, polar coding schemes have been extended to several communication settings, such as multiple access channels [10], [11], broadcast channels [12], [13], interference channels [14], [15], relay channels [16]–[18], wiretap channels [18]–[23], compound channels [24], [25], source coding [26], Slepian–Wolf coding [27], multiple description coding [28], among many others. In these problems, when the achievability result is built on some *single user point-to-point* random coding scheme, such as superposition in degraded broadcast channels and successive cancellation decoding in the multiple access channels, a polar coding scheme can be readily developed using *point-to-point polar codes*. For the MDC problem, however, it is unclear how its crucial random coding technique, namely joint typicality encoding, can be implemented using point-to-point polar codes.

Implementing joint typicality encoding using polar codes has been previously investigated by Shi, Song, Tian, Chen, and Dumitrescu [28] for the MDC problem, Goela, Abbe, and Gastpar [12] and Mondelli, Hassani, Sason, and Urbanke [13] in the context of Marton coding in broadcast channels, among others. In all these work, the polar coding schemes are designed for some specific rate points in the corresponding joint typicality encoding region. Specifically in [28], the target rate point is the one induced by the *joint polarization* technique, which was initially proposed by Şaşoğlu, Telatar and Yeh in the context of multiple access channels (MAC) [10]. This polar coding scheme is insufficient to achieve the entire EGC inner bound for the MDC problem. In [12], [13], the target rate points are the corner points of the Marton inner bound for broadcast channels. In all these work, it is unclear *without time-sharing* whether any rate point in the joint typicality encoding region can be achieved using polar codes.

In this paper, we explore polar coding schemes that implement joint typicality encoding in full generality. By exploring Arıkan's *monotone chain rule* polarization technique [27], which was initially proposed in the context of Slepian–Wolf coding, we show the proposed polar coding scheme achieves the entire EGC inner bound without any time-sharing. The technique can be adapted for Marton coding in broadcast channels as well.

The rest of this paper is organized as follows. In Section II-A, we review the joint polarization technique introduced in [10] for MAC. In Section II-B, we explain how joint polarization can be generalized and improved by Arıkan's monotone chain rule polarization technique [27]. In Section III, we describe our proposed polar coding scheme for the MDC problem, outlining the differences with previous work of [28] on this problem. Namely, our scheme is shown to achieve the entire El Gamal–Cover inner bound. Finally, we conclude in Section IV.

II. Preliminaries

A. Joint Polarization for MAC

In this section, we review the joint polarization technique introduced in [10]. Consider a two-user MAC ($\mathcal{X}_1 \times \mathcal{X}_2$, $p(y|x_1,x_2)$, $\mathcal{Y}$), where two senders wish to communicate two messages M_1 and M_2 to a receiver through sending two codewords $X_1^N(M_1)$ and $X_2^N(M_2)$ over $N = 2^k$ uses of the channel. For a fix input pmf $p(x_1)p(x_2)$, the achievable rate region is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I(X_1;Y|X_2), \\ R_2 &\leq I(X_2;Y|X_1), \\ R_1 + R_2 &\leq I(X_1,X_2;Y). \end{aligned}$$

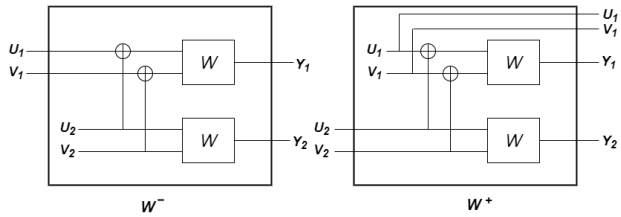

Fig. 3: Channel splitting operation for two uses of a two-user MAC under the technique of joint polarization

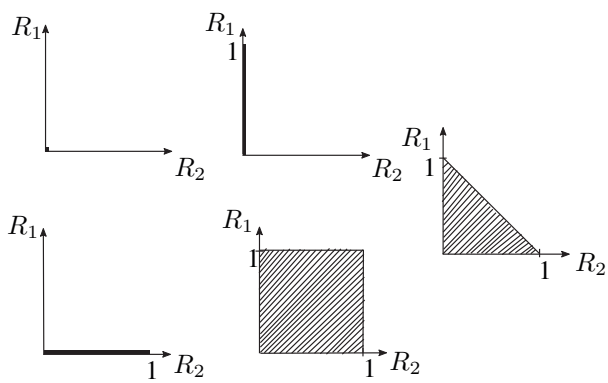

Fig. 4: Five extremal channels for MAC.

Similar to the MDC region, the rate pairs (R_1, R_2) that satisfy $R_1+R_2 = I(X_1, X_2; Y)$ in the above region is called the *dominant face* of the MAC region.

In [10], a technique, termed *joint polarization*, is proposed. Define the polar transform: let $U^N = X_1^N G_N$ and $V^N = X_2^N G_N$ for $G_N = B_N F_N$, where B_N is the bit-reversal matrix defined in [8] and $F_N = F_1^{\otimes k}$ is the k-th power Kronecker product of the matrix

$$F_1 \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Similar to the single-user case, two independent uses of W are transformed into two MACs W^- and W^+, as depicted in Fig. 3. Consecutively, by applying k levels of this transformation, $N = 2^k$ different MAC channels are created. For $i \in [N]$, consider the mutual information triple

$$\big(I(U_i; Y^N|U^{i-1}, V^i),$$
$$I(V_i; Y^N|U^i, V^{i-1}),$$
$$I(U_i, V_i; Y^N|U^{i-1}, V^{i-1})\big).$$

It is shown in [10] that, as k goes to infinity, the mutual information triple approaches one of the five points in following set with high probability

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

In other words, five extremal channels are approached for a two-user MAC, as compared to two in the single-user case (either noise-free or pure noise channels). The point (0,0,0) correspond to the case when the output provides no information about any of the two inputs. The points (1,0,1) and (0,1,1) correspond to the cases when the output provides full information about one of the inputs but provides no information about the other input. The point (1,1,2) corresponds to the case when the output provides full information about both inputs. Finally, the point (1,1,1) is a pure *contention* channel: if any of the two users communicates at zero rate, then the output will provide full information about the other user's input.

Coding over these extremal channels is simple: either send an information bit or freeze the bit to some known value by the decoder, depending on the corresponding extremal channel. For the (1,1,1) channel, simply assign an information bit to one of the users arbitrarily while freezing the other user's input. It is shown in [10] that the constructed polar code can achieve some rate point on the dominant face of the MAC region.

Note however that the above joint polarization technique does not achieve any rate point in the MAC region. This is because joint polarization only considers a single order in expanding (U^N, V^N) in the mutual information term, namely

$$I(U^N, V^N; Y^N) = \sum_{i=1}^{N} I(U_i, V_i; Y^N|U^{i-1}, V^{i-1}).$$

Thus, the symbols are decoded successively in the order $(U_1, V_1), (U_2, V_2), \cdots, (U_N, V_N)$. By exploiting different decoding orders, one can achieve different points on the dominant face of the MAC region. This motivates the next subsection.

B. MAC Polarization Using Monotone Chain Rules

Now we give an overview of the scheme proposed by Arıkan in [27] which achieves any point on the dominant face of the MAC region. The scheme is based on exploiting all possible expansions of the joint mutual information between the channel inputs and outputs. Let $U^N = X_1^N G_N$ and $V^N = X_2^N G_N$ be the random variables induced by the polar transform of X_1^N and X_2^N respectively, where X_1 and X_2 are the inputs to a two-user MAC, as defined in Section II-A.

Consider now the expansions of $I(U^N, V^N; Y^N)$ that are of the form

$$\sum_{i=1}^{2N} I(S_i; Y^N|S^{i-1}),$$

where $S^{2N} = (S_1, \ldots, S_{2N})$ is a *monotone* permutation of $U^N V^N$, i.e. a permutation where the relative order of the elements of both U^N and V^N is preserved. For example, $(U_1, U_2, U_3, U_4, V_1, V_2, V_3, V_4)$ and $(U_1, V_1, U_2, V_2, U_3, V_3, U_4, V_4)$ are monotone permutations of U^4V^4, but $(U_1, U_2, U_3, U_4, V_1, V_4, V_3, V_2)$ is not because the order on (V_1, V_2, V_3, V_4) is not preserved. S^{2N} is assumed to be known by both transmitters and the receiver. Also,

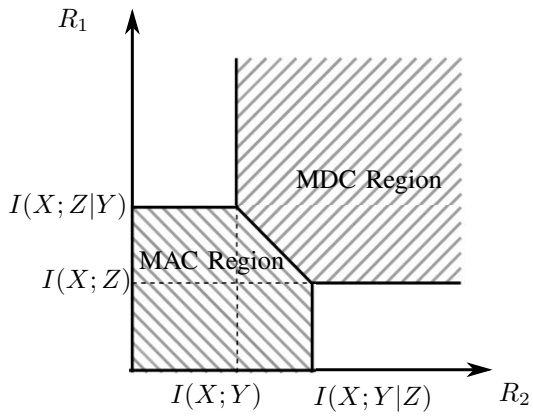

Fig. 5: MDC-MAC duality

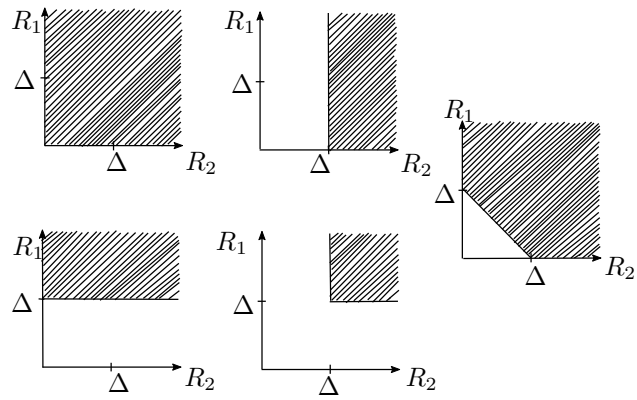

Fig. 6: Five extremal channels for the MDC problem.

let $\mathcal{S}_U$ and $\mathcal{S}_V$ denote the set of indices of S^{2N} such that $S_i = U_k$ and $S_i = V_k$ respectively and define the rates

$$R_1 = \frac{1}{N} \sum_{i \in \mathcal{S}_U} I(S_i; Y^N | S^{i-1}),$$
$$R_2 = \frac{1}{N} \sum_{i \in \mathcal{S}_V} I(S_i; Y^N | S^{i-1}).$$

The main contribution of [27] is that the pair (R_1, R_2) can approach any rate pair on the dominant face of the capacity region by selecting a valid permutation S^{2N} and that the mutual informations $I(S_i; Y^N | S^{i-1})$ become polarized with increasing N (i.e. asymptotically approach either 0 or 1). Also, it is shown that permutations of the form $S^{2N} = (U^i, V^N, U_{i+1}^N)$ are sufficient to guarantee this result. Namely, the following theorem holds:

Theorem 2. *([27]) For each $\epsilon > 0$ and $\beta < 1/2$, and rate pair (R_x, R_y) on the dominant face of the MAC capacity region, there exists an N and a permutation $S^{2N} = (U^i, V^N, U_{i+1}^N)$ for some i where*

(i) $|R_1 - R_x| < \epsilon$ and $|R_2 - R_y| < \epsilon$

(ii)

$$\frac{N - |\mathcal{F}_1|}{N} > R_1 - \epsilon \quad \text{and} \quad \frac{N - |\mathcal{F}_2|}{N} > R_2 - \epsilon,$$

where

$$\mathcal{F}_1 = \{1 \le i \le 2N : i \in \mathcal{S}_U, I(S_i; Y^N | S^{i-1}) < 2^{-N^\beta}\},$$
$$\mathcal{F}_2 = \{1 \le i \le 2N : i \in \mathcal{S}_V, I(S_i; Y^N | S^{i-1}) < 2^{-N^\beta}\}.$$

III. Polar Codes for the MDC problem

A. Previous work

Polar codes have been considered in [28] for the multiple description coding problem. Based on a joint polarization technique similar to [10], a polar coding scheme is shown to achieve one point on the dominant face of the EGC inner bound. The key point in this result is an MDC-MAC duality, that we outline in this section. Consider again the MDC problem described in section I, with random variables (X, Y, Z) distributed over $\mathbb{F}_q$ and having a joint probability mass function $p(x, y, z)$. In what follows, we assume that Y and Z are uniformly distributed, and this assumption holds because any random variable can be always approximated by a uniform random variable over a sufficiently large alphabet size and through a deterministic mapping function. Given $p(x, y, z)$, the conditional distribution $p(x|y, z)$ can be viewed as a MAC with inputs (Y, Z) and output X. Recall that for fixed MAC distribution $p(x|y, z)$ where $p_{Y,Z}(y, z) = p_Y(y)p_Z(z)$, the rate region is the set of non-negative rate pairs (R_1, R_2) such that

$$R_1 \le I(Y; X|Z),$$
$$R_2 \le I(Z; X|Y),$$
$$R_1 + R_2 \le I(X; Y, Z).$$

In comparison with the MDC rate region in (1), it can be seen that the two regions will have the same *sum-rate* if Y and Z are independent. Fig. 5 shows the two regions in the case that Y and Z are independent. This is not necessarily true for the MDC problem in general, where Y and Z are the two descriptions of the source. Nevertheless, the independence of the two descriptions can be achieved via a "dithering step" ([28]): Let Z' be a random variable uniformly distributed over $\mathbb{F}_q$ and independent of (X, Y, Z). Define $\widetilde{X} = (X, Z')$, $\widetilde{Y} = Y$ and $\widetilde{Z} = Z \oplus Z'$. Then clearly $\widetilde{Y}$ and $\widetilde{Z}$ are independent and the following holds:

$$\begin{aligned} I(\widetilde{X}; \widetilde{Y}) &= I(X; Y), \\ I(\widetilde{X}, \widetilde{Z}; \widetilde{Y}) &= I(X, Z; Y), \\ I(\widetilde{X}; \widetilde{Z}) &= I(X; Z), \\ I(\widetilde{X}, \widetilde{Y}; \widetilde{Z}) &= I(X, Y; Z). \end{aligned} \tag{2}$$

Also, we have that

$$\begin{aligned} I(\widetilde{X};\widetilde{Y},\widetilde{Z}) + I(\widetilde{Y};\widetilde{Z}) &= I(\widetilde{X};\widetilde{Y},\widetilde{Z}) \\ &= I(\widetilde{X};\widetilde{Y}) + I(\widetilde{X};\widetilde{Z}|\widetilde{Y}) + I(\widetilde{Y};\widetilde{Z}) \\ &= I(\widetilde{X};\widetilde{Y}) + I(\widetilde{X},\widetilde{Y};\widetilde{Z}) \\ &= I(X;Y) + I(X,Y;Z) \\ &= I(X;Y,Z) + I(Y;Z). \end{aligned}$$

Hence, the EGC rate region does not change under this transformation.

Following this, it can be shown that, similar to a MAC, the sum-rate on the dominant face of the EGC region is preserved under the polar transformation. Also, we have that

$$\begin{aligned} I(\widetilde{X};\widetilde{Y},\widetilde{Z}) &= I(X;Y,Z) + I(Y;Z) \\ &= I(X;Y) + I(X,Y;Z) \\ &= I(\widetilde{X};\widetilde{Y}) + I(X,Y;Z) \end{aligned} \tag{3}$$

and, similarly,

$$I(\widetilde{X};\widetilde{Y},\widetilde{Z}) = I(\widetilde{X};\widetilde{Z}) + I(X,Z;Y). \tag{4}$$

Hence, it follows from the results of [10] for the polarization of $I(X,Z;Y)$ and $I(X,Y;Z)$ that $I(\widetilde{X};\widetilde{Y})$ and $I(\widetilde{X};\widetilde{Z})$ also polarize. Hence, five extremal regions are approached asymptotically for the MDC problem as well. On each of these regions, encoding follows naturally from the MAC case. Fig. 6 shows the extremal regions for the MDC problem, where $\Delta = \log_2 q$.

It is also shown in [28] that the induced distribution from the polar transformation satisfies the distortion constraints, and consequently one rate pair on the dominant face of the EGC region can be achieved asymptotically using this scheme.

In what follows, we argue that the whole dominant face of the EGC region can be achieved, if we consider different monotone chain rule expansions of the mutual information (i.e. similar to the approach of [27]), while still satisfying the distortion constraints imposed by the problem.

B. Proposed Scheme

We now describe our proposed polar coding scheme for the MDC problem. Let (X_i, Y_i, Z_i) be N i.i.d. copies of $(X,Y,Z), i = 1,\ldots,N$ distributed according to $p_{X,Y,Z}(x,y,z)$ where $N = 2^k$ is the code block length. As indicated before, we assume that random variables Y and Z are uniformly distributed over $\mathbb{F}_q$.

The polar coding scheme is defined by two parameters (S^{2N}, β), where S^{2N} is a monotone chain rule for $U^N V^N$, as defined in [27], and β is a threshold parameter where $0 < \beta < 1/2$. We also consider two subsets $\mathcal{F}_1$ and $\mathcal{F}_2$ of $\{1,2,\ldots,2N\}$, defined as follows:

$$\begin{aligned} \mathcal{F}_1 &= \{1 \le i \le 2N : i \in \mathcal{S}_U, I(S_i; X^N|S^{i-1}) < 2^{-N^\beta}\}, \\ \mathcal{F}_2 &= \{1 \le i \le 2N : i \in \mathcal{S}_V, I(S_i; X^N|S^{i-1}) < 2^{-N^\beta}\}. \end{aligned} \tag{5}$$

The induced distribution of the variables (X^N, S^{2N}) is given by

$$\begin{aligned} &p_{X^N,S^{2N}}(x^N, s^{2N}) \\ &\quad = p_{X^N,U^N,V^N}(x^N, u^N, v^N) \\ &\quad = \sum_{y^N}\sum_{z^N}\prod_{i=1}^{N} p_{X,Y,Z}(x_i,y_i,z_i)p(u^N|y^N)p(v^N|z^N), \end{aligned}$$

where

$$\begin{aligned} p(u^N|y^N) &= 1_{\{u^N = y^N G_N\}}, \\ p(v^N|z^N) &= 1_{\{v^N = z^N G_N\}}. \end{aligned}$$

1) Encoding: For each $i \in \mathcal{F}_1$ or $i \in \mathcal{F}_2$, generate s_i at random over $\mathbb{F}_q$. The frozen symbols are generated once and informed to both the encoder and the decoder, and are fixed throughout the communication. If $i \notin \mathcal{F}_1$ and $i \notin \mathcal{F}_2$, then s_i will take value $a \in \mathbb{F}_q$ with probability given by $\frac{P_{X^N,S^i}(x^N,(s^{i-1},a))}{P_{X^N,S^{i-1}}(x^N,s^{i-1})}$. So description 1 will be $s_{\mathcal{F}_1^C} = \{s_i : i \notin \mathcal{F}_1, i \in \mathcal{S}_U\}$ and description 2 will be $s_{\mathcal{F}_2^C} = \{s_i : i \notin \mathcal{F}_2, i \in \mathcal{S}_V\}$.

2) Decoding: Decoder 1 will form u^N using the first description and the known frozen symbols of u. Then it will generate $y^N = u^N G_N$ and apply ϕ_1 to each symbol of y^N and the output will be the reconstruction $\hat{X}_1^N$. Similarly decoder 2 will form v^N using the first description and the known frozen symbols of v. Then it will generate $z^N = v^N G_N$, and apply ϕ_2 to each symbol of z^N as reconstruction $\hat{X}_2^N$. Decoder 0 forms u^N and v^N, generates y^N and z^N and applies ϕ_0 to $(y_i, z_i), i = 1,2,\cdots,N$ and gets reconstruction $\hat{X}_0^N$.

3) Distortion Analysis: Now we show that this scheme satisfies the distortion constraints. Note that the encoding procedure described above only approximates $p(x^N, s^{2N})$. We want to show that the excess distortion due to this approximation is bounded. First define the distribution $\hat{p}$ to be the distribution induced by our encoding procedure. It follows that

$$\hat{p}(s_i|s^{i-1}, x^N) = \begin{cases} \frac{1}{q} & \text{if } i \in \mathcal{F}_1 \text{ or } i \in \mathcal{F}_2 \\ p(s_i|s^{i-1}, x^N) & \text{otherwise.} \end{cases}$$

Let $\hat{D}_j, j = 0,1,2$ be the expected distortions under distribution $\hat{p}$, while $D_j^*, j = 0,1,2$ be the expected distortions under p. The following theorem states that the difference between the two distortions can be made arbitrarily small.

Theorem 3. *For any $\epsilon > 0$, there exists a large enough blocklength N such that $|\hat{D}_j - D_j^*| < \epsilon, j = 0,1,2$.*

Proof: First note that

$$\begin{aligned} |\hat{D}_1 - D_1^*| &= \big|\mathbb{E}_{\hat{p}}[d_1^{(N)}(X^N, \phi_1^{(N)}(U^N G_N))] \\ &\qquad - \mathbb{E}_p[d_1^{(N)}(X^N, \phi_1^{(N)}(U^N G_N))]\big| \\ &\le d_{\max} \sum_{x^N, s^{2N}} \big|\hat{p}(x^N, s^{2N}) - p(x^N, s^{2N})\big|. \end{aligned}$$

Similarly,

$$|\hat{D}_2 - D_2^*| \le d_{\max} \sum_{x^N, s^{2N}} \big|\hat{p}(x^N, s^{2N}) - p(x^N, s^{2N})\big|,$$

$$|\hat{D}_0 - D_0^*| \le d_{\max} \sum_{x^N, s^{2N}} \big|\hat{p}(x^N, s^{2N}) - p(x^N, s^{2N})\big|.$$

Now we have

$$\begin{aligned}
&\sum_{x^N, s^{2N}} \big|\hat{p}(x^N, s^{2N}) - p(x^N, s^{2N})\big| \\
&= \sum_{x^N, s^{2N}} p(x^N)\big|\hat{p}(s^{2N}|x^N) - p(s^{2N}|x^N)\big| \\
&= \sum_{x^N, s^{2N}} p(x^N)\Big|\prod_{i=1}^{2N} \hat{p}(s_i|x^N, s^{i-1}) - \prod_{i=1}^{2N} p(s_i|x^N, s^{i-1})\Big| \\
&\overset{(a)}{=} \sum_{x^N, s^{2N}} p(x^N)\Bigg|\sum_{i=1}^{2N}(p(s_i|x^N, s^{i-1}) - \hat{p}(s_i|x^N, s^{i-1})) \\
&\qquad \cdot\Big(\prod_{j=1}^{i-1} p(s_j|x^N, s^{j-1}) \prod_{j=i+1}^{N} \hat{p}(s_j|x^N, s^{j-1})\Big)\Bigg| \\
&\le \sum_{i=1}^{2N} \sum_{x^N, s^{2N}} p(x^N)\Bigg|(p(s_i|x^N, s^{i-1}) - \hat{p}(s_i|x^N, s^{i-1})) \\
&\qquad \cdot\Big(\prod_{j=1}^{i-1} p(s_j|x^N, s^{j-1}) \prod_{j=i+1}^{N} \hat{p}(s_j|x^N, s^{j-1})\Big)\Bigg| \\
&= \sum_{i=1}^{2N} \sum_{x^N, s^i} p(s^{i-1}, x^N)\big|(p(s_i|x^N, s^{i-1}) - \hat{p}(s_i|x^N, s^{i-1}))\big|,
\end{aligned}$$

where (a) follows since

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

Therefore we have for $j = 0, 1, 2$,

$$|\hat{D}_j - D_j^*| \le d_{\max} \sum_{i=1}^{2N} E_i,$$

where $E_i = \sum_{s_i=0}^{q-1} \mathbb{E}_p[\big|p(s_i|X^N, S^{i-1}) - \hat{p}(s_i|X^N, S^{i-1})\big|]$. We have two cases:

Case 1: $i \in \mathcal{F}_1$ or $i \in \mathcal{F}_2$

$$\begin{aligned}
E_i &= \sum_{s_i=0}^{q-1} \mathbb{E}_p\Big[\big|p(s_i|X^N, S^{i-1}) - \frac{1}{q}\big|\Big] \\
&\overset{(a)}{\le} \sqrt{(2\log^{-1} e) I(X^N, S^{i-1}; S_i)} \\
&\le \sqrt{(2\log^{-1} e)\delta},
\end{aligned}$$

where (a) follows from Pinsker's inequality and δ is a threshold parameter such that $\delta = \mathcal{O}(2^{-N^\beta})$.

Case 2: $i \notin \mathcal{F}_1$ and $i \notin \mathcal{F}_2$. In this case, clearly $E_i = 0$. Combining these two cases, we get that for $j = 0, 1, 2$,

$$\begin{aligned}
|\hat{D}_j - D_j^*| &\le d_{\max}(2N)\sqrt{(2\log^{-1} e)\delta} \\
&= O(2^{-N^{\beta'}}),
\end{aligned}$$

for any $\beta' \in (0, \beta)$. ■

4) Achieving the entire EGC rate region: We will now show that our scheme can approach any point on the dominant face of the EGC region arbitrarily closely. First, define the rate pairs:

$$R_1 = \frac{1}{N} \sum_{i \in \mathcal{S}_U} I(S_i; X^N | S^{i-1}),$$
$$R_2 = \frac{1}{N} \sum_{i \in \mathcal{S}_V} I(S_i; X^N | S^{i-1}).$$

In an approach similar to Theorem 2 in [27], we can show that the terms $I(S_i; X^N|S^{i-1})$ asymptotically approach 0 or 1, and that

$$\frac{N - |\mathcal{F}_1|}{N} \to R_1 \quad and \quad \frac{N - |\mathcal{F}_2|}{N} \to R_2$$

where $\mathcal{F}_1$ and $\mathcal{F}_2$ are as defined in (5). Also, adapting an analysis similar to [27], we can directly see that if (U^N, V^N) is a pair obtained from (Y^N, Z^N) via the polar transformation defined in Section II-A, then any monotone chain rule expansion on U^N, V^N should satisfy

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

In general, corner points of this region are not necessarily the corner points of the EGC rate region. However, following the "dithering step" defined in section III-A, we know that:

$$I(\widetilde{X}; \widetilde{Y}, \widetilde{Z}) = I(X; Y, Z) + I(Y; Z).$$

Therefore, obtaining (U^N, V^N) from $\widetilde{Y}^N, \widetilde{Z}^N$ via the polar transformation, it follows that any monotone chain rule expansion on U^N, V^N should satisfy

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

with equality for the first inequality if $S^{2N} = (U^N, V^N)$ and equality for the second inequality if $S^{2N} = (V^N, U^N)$. Thus, the corner points can be achieved, and achieving any other point on the dominant face follows directly from Theorem 2. ■

IV. Conclusions

In this work, we presented a polar coding scheme that achieves the entire El Gamal–Cover inner bound for the multiple description coding problem. As discussed, choosing different decoding orders will achieve different points on the dominant face. Also, the independence of the two descriptions is crucial and we ensure that through a dithering argument.

ACKNOWLEDGMENT

The authors would like to thank Prof. Young-Han Kim for helpful comments on the paper and for his guidance and support throughout the submission process. This work is supported partially by InterDigital, Inc.

REFERENCES

[1] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inf. Theory*, vol. 26, no. 5, pp. 518–521, 1980.

[2] J. K. Wolf, A. D. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell System Tech. J.*, vol. 59, no. 8, pp. 1417–1426, 1980.

[3] H. S. Witsenhausen and A. D. Wyner, "Source coding for multiple descriptions—II: A binary source," *Bell System Tech. J.*, vol. 60, no. 10, pp. 2281–2292, 1981.

[4] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

[5] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851–857, 1982.

[6] J. Chen, C. Tian, T. Berger, and S. S. Hemami, "Multiple description quantization via (g)ram-(s)chmidt orthogonalization," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5197–5217, Dec 2006.

[7] T. Berger and Z. Zhang, "Minimum breakdown degradation in binary source encoding," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, pp. 807–814, 1983.

[8] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3051–3073, July 2009.

[9] J. Wang, J. Chen, L. Zhao, P. Cuff, and H. Permuter, "On the role of the refinement layer in multiple description coding and scalable coding," *IEEE Trans. Inf. Theory*, vol. 57, pp. 1443–1456, March 2011.

[10] E. Şaşoğlu, E. Telatar, and E. Yeh, "Polar codes for the two-user multiple-access channel," in *IEEE Trans. Inf. Theory*, vol. 59, pp. 6583–6592, October 2013.

[11] E. Abbe and I. E. Telatar, "Polar codes for the m-user multiple access channel," in *IEEE Trans. Inf. Theory*, vol. 58, pp. 5437–5448, August 2012.

[12] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," in *Proc. IEEE Internat. Symp. Inf. Theory*, pp. 1127–1131, July 2013.

[13] M. Mondelli, S. H. Hassani, I. Sason, and R. L. Urbanke, "Achieving marton's region for broadcast channels using polar codes," in *IEEE Trans. Inf. Theory*, vol. 61, pp. 783–800, February 2015.

[14] L. Wang and E. Şaşoğlu, "Polar coding for interference networks," in *Proc. IEEE Internat. Symp. Inf. Theory*, pp. 311–315, June 2014.

[15] L. Wang, *Channel Coding Techniques for Network Communication*. Ph.D. thesis, University of California, San Diego, La Jolla, CA, 2015.

[16] L. Wang, "Polar coding for relay channels," in *Proc. IEEE Internat. Symp. Inf. Theory*, pp. 1532–1536, June 2015.

[17] M. Karzand, "Polar codes for degraded relay channels," in *Proc. Int. Zurich Seminar Commun.*, pp. 59–62, February 2012.

[18] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested polar codes for wiretap and relay channels," in *Communications Letters, IEEE*, vol. 14, pp. 752–754, August 2010.

[19] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6428–6443, Oct 2011.

[20] E. Şaşoğlu and A. Vardy, "A new polar coding scheme for strong security on wiretap channels," in *Proc. IEEE Internat. Symp. Inf. Theory*, pp. 1117–1121, July 2013.

[21] Y. P. Wei and S. Ulukus, "Polar coding for the general wiretap channel with extensions to multiuser scenarios," *IEEE Journal on Selected Areas in Communications*, vol. 34, pp. 278–291, Feb 2016.

[22] T. C. Gulcu and A. Barg, "Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component," in *Proc. IEEE Inf. Theory Workshop*, pp. 1–5, April 2015.

[23] R. A. Chou and M. R. Bloch, "Polar coding for the broadcast channel with confidential messages: A random binning analogy," *IEEE Trans. Inf. Theory*, vol. 62, pp. 2410–2429, May 2016.

[24] S. H. Hassani and R. Urbanke, "Universal polar codes," in *Proc. IEEE Internat. Symp. Inf. Theory*, pp. 1451–1455, June 2014.

[25] E. Şaşoğlu and L. Wang, "Universal polarization," *IEEE Trans. Inf. Theory*, vol. 62, pp. 2937–2946, June 2016.

[26] E. Arıkan, "Source polarization," in *Proc. IEEE Internat. Symp. Inf. Theory*, pp. 899–903, June 2010.

[27] E. Arıkan, "Polar coding for the slepian-wolf problem based on monotone chain rules," in *Proc. IEEE Internat. Symp. Inf. Theory*, pp. 566–570, July 2012.

[28] Q. Shi, L. Song, C. Tian, J. Chen, and S. Dumitrescu, "Polar codes for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 61, pp. 107–119, January 2015.