

Polar Codes for Sources with Finite Reconstruction Alphabets

Aria G. Sahebi and S. Sandeep Pradhan

Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48109, USA.
Email: ariaghs@umich.edu, pradhanv@umich.edu

Abstract—Polar codes are a new class of codes introduced by Arikan capable of achieving the symmetric capacity of discrete memoryless channels. It was later shown that polar codes can achieve the symmetric rate-distortion bound for the lossy source coding problem when the size of the reconstruction alphabet is a prime. In this paper, we show that polar codes achieve the symmetric rate-distortion bound for finite-alphabet sources regardless of the size of the reconstruction alphabet.

I. INTRODUCTION

Polar codes were originally proposed by Arikan in [1] for discrete memoryless channels with a binary input alphabet. Polar codes over binary input channels are coset codes capable of achieving the symmetric capacity of channels. These codes are constructed based on the Kronecker power of the 2×2 matrix $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and are the first known class of capacity achieving codes with an explicit construction.

It was subsequently shown by Korada and Urbanke [2] that polar codes employed with a successive cancellation encoder are also optimal for the lossy source coding problem when the size of the reconstruction alphabet is two. Karzand and Telatar extended this result to the case where the size of the reconstruction alphabet is a prime [3].

In an earlier work [4], we have shown that polar codes can achieve the symmetric capacity of arbitrary discrete memoryless channels regardless of the size of the channel input alphabet. It is shown in [4] that in general, when the channel input alphabet is an Abelian group \mathbf{G} , channel polarization occurs in several levels. For each subgroup H of \mathbf{G} , there exists a corresponding asymptotic case in which the channel output can determine the coset of the subgroup to which the channel input belongs. Then a modification is made to the encoding and decoding rules to achieve the symmetric capacity of the channel. As mentioned in [4], polar codes employed with the modified encoding rule fall into a class of codes called nested group codes.

In this paper, motivated by the results of [2]–[4], we show that polar codes achieve the symmetric rate-distortion bound for the lossy source coding problem when the size of the reconstruction alphabet is finite. We show that similar to the channel coding problem, polar transformations applied

to (test) channels can converge to several asymptotic cases each corresponding to a subgroup H of the reconstruction alphabet \mathbf{G} . We employ a modified randomized rounding encoding rule to achieve the symmetric rate-distortion bound.

This paper is organized as follows: In Section II, some definitions and basic facts are stated which are used in the paper. In Section III we show that polar codes achieve the symmetric rate-distortion bound when the size of the reconstruction alphabet is of the form $q = p^r$ where p is a prime and r is a positive integer. This result is generalized in Section IV to the case where the reconstruction alphabet is some arbitrary Abelian group (of arbitrary size and arbitrary group operation). We conclude in Section V.

II. PRELIMINARIES

1) *Source and Channel Models*: The source is modeled as a discrete-time random process X with each sample taking values in a fixed finite set \mathcal{X} with probability distribution p_X . The reconstruction alphabet is denoted by \mathcal{U} and the quality of reconstruction is measured by a single-letter distortion function $d : \mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R}^+$. We denote the source by $(\mathcal{X}, \mathcal{U}, p_X, d)$.

A (test) channel is specified by a tuple $(\mathcal{U}, \mathcal{X}, W)$ where \mathcal{U} is the channel input alphabet, \mathcal{X} is the channel output alphabet and W is the channel transition kernel.

2) *Achievability and the Rate-Distortion Function*: A transmission system with parameters $(n, \Theta, \Delta, \tau)$ for compressing a given source $(\mathcal{X}, \mathcal{U}, p_X, d)$ consists of an encoding mapping and a decoding mapping

$$\begin{aligned} e : \mathcal{X}^n &\rightarrow \{1, 2, \dots, \Theta\}, \\ g : \{1, 2, \dots, \Theta\} &\rightarrow \mathcal{U}^n \end{aligned}$$

such that the following condition is met:

$$P(d(X^n, g(e(X^n))) > \Delta) \leq \tau$$

where X^n is the random vector of length n generated by the source. In this transmission system, n denotes the block length, $\log \Theta$ denotes the number of channel uses, Δ denotes the distortion level and τ denotes the probability of exceeding the distortion level Δ .

Given a source, a pair of non-negative real numbers (R, D) is said to be achievable if there exists for every $\epsilon > 0$, and for

all sufficiently large numbers n a transmission system with parameters $(n, \Theta, \Delta, \tau)$ for compressing the source such that

$$\frac{1}{n} \log \Theta \leq R + \epsilon, \quad \Delta \leq D + \epsilon, \quad \tau \leq \epsilon$$

The optimal rate distortion function $R^*(D)$ of the source is given by the infimum of the rates R such that (R, D) is achievable.

It is known that the optimal rate-distortion function is given by:

$$R(D) = \min_{\substack{p_{U|X} \\ \mathbb{E}_{p_X p_{U|X}} \{d(X, U)\} \leq D}} I(X; U)$$

where $p_{U|X}$ is the conditional probability of U given X and $p_X p_{U|X}$ is the joint distribution of X and U .

The *symmetric rate-distortion function* $\bar{R}(D)$ is defined as follows:

$$\bar{R}(D) = \min_{\substack{p_U \\ \mathbb{E}_{p_X p_{U|X}} \{d(X, U)\} \leq D \\ p_U = \frac{1}{q}}} I(X; U)$$

where p_U is the marginal distribution of U given by $p_U(u) = \sum_{x \in \mathcal{X}} p_X(x) p_{U|X}(u|x)$ and q is the size of the reconstruction alphabet \mathcal{U} .

3) *Binary Polar Codes*: For any $N = 2^n$, a polar code of length N designed for the channel $(\mathbb{Z}_2, \mathcal{Y}, W)$ is a linear code characterized by a generator matrix G_N and a set of indices $A \subseteq \{1, \dots, N\}$ of *perfect channels*. The generator matrix for polar codes is defined as $G_N = B_N F^{\otimes n}$ where B_N is a permutation of rows, $F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and \otimes denotes the Kronecker product. The set A is a function of the channel. The decoding algorithm for polar codes is a specific form of successive cancellation [1].

For the source coding problem, when the size of the reconstruction alphabet is binary, the encoding rule is a form of soft decision successive encoding and the decoding rule is a matrix multiplication [2].

4) *Groups, Rings and Fields*: All groups referred to in this paper are *Abelian groups*. Given a group $(\mathbf{G}, +)$, a subset H of \mathbf{G} is called a *subgroup* of \mathbf{G} if it is closed under the group operation. In this case, $(H, +)$ is a group on its own right. This is denoted by $H \leq \mathbf{G}$. A *coset* C of a subgroup H is a shift of H by an arbitrary element $a \in \mathbf{G}$ (i.e. $C = a + H$ for some $a \in \mathbf{G}$). For any subgroup H of \mathbf{G} , its cosets partition the group \mathbf{G} . A *transversal* T of a subgroup H of \mathbf{G} is a subset of \mathbf{G} containing one and only one element from each coset (shift) of H .

We give some examples in the following: The simplest non-trivial example of groups is \mathbb{Z}_2 with addition mod-2 which is a *ring* and a *field* with multiplication mod-2. The group $\mathbb{Z}_2 \times \mathbb{Z}_2$ is also a ring and a field under component-wise mod-2 addition and a carefully defined multiplication. The group \mathbb{Z}_4 with mod-4 addition and multiplication is a ring but not a field since the element $2 \in \mathbb{Z}_4$ does not have a multiplicative inverse. The subset $\{0, 2\}$ is a subgroup of

\mathbb{Z}_4 since it is closed under mod-4 addition. $\{0\}$ and \mathbb{Z}_4 are the two other subgroups of \mathbb{Z}_4 . The group \mathbb{Z}_6 is neither a field nor a ring. Subgroups of \mathbb{Z}_6 are: $\{0\}$, $\{0, 3\}$, $\{0, 2, 4\}$ and \mathbb{Z}_6 .

5) *Polar Codes Over Abelian Groups*: For any discrete memoryless channel, there always exists an Abelian group of the same size as that of the channel input alphabet. In general, for an Abelian group, there may not exist a multiplication operation. Since polar encoders are characterized by a matrix multiplication, before using these codes for channels of arbitrary input alphabet sizes, a generator matrix for codes over Abelian groups needs to be properly defined. In Appendix A of [5] a convention is introduced to generate codes over groups using $\{0, 1\}$ -valued generator matrices.

6) *Channel Parameters*: For a test channel $(\mathcal{U}, \mathcal{X}, W)$ assume \mathcal{U} is equipped with the structure of a group $(\mathbf{G}, +)$. The symmetric capacity of the test channel is defined as $I^0(W) = I(U; X)$ where the channel input U is uniformly distributed over \mathcal{U} and X is the output of the channel; i.e. for $q = |\mathcal{U}|$,

$$I^0(W) = \sum_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} \frac{1}{q} W(x|u) \log \frac{W(x|u)}{\sum_{\tilde{u} \in \mathcal{U}} \frac{1}{q} W(x|\tilde{u})}$$

The Bhattacharyya distance between two distinct input symbols u and \tilde{u} is defined as

$$Z(W_{\{u, \tilde{u}\}}) = \sum_{x \in \mathcal{X}} \sqrt{W(x|u)W(x|\tilde{u})}$$

and the average Bhattacharyya distance is defined as

$$Z(W) = \sum_{\substack{u, \tilde{u} \in \mathcal{U} \\ u \neq \tilde{u}}} \frac{1}{q(q-1)} Z(W_{\{u, \tilde{u}\}})$$

We use the following two quantities in the paper extensively:

$$D_d(W) = \frac{1}{2q} \sum_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} |W(x|u) - W(x|u+d)|$$

$$\tilde{D}_d(W) = \frac{1}{2q} \sum_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} (W(x|u) - W(x|u+d))^2$$

where d is some element of \mathbf{G} and $+$ is the group operation.

7) *Notation*: We denote by $O(\epsilon)$ any function of ϵ which is right-continuous around 0 and that $O(\epsilon) \rightarrow 0$ as $\epsilon \downarrow 0$. We denote by $a \approx_\epsilon b$ to mean $a = b + O(\epsilon)$.

For positive integers N and r , let $\{A_0, A_1, \dots, A_r\}$ be a partition of the index set $\{1, 2, \dots, N\}$. Given sets T_t for $t = 0, \dots, r$, the direct sum $\bigoplus_{t=0}^r T_t^{A_t}$ is defined as the set of all tuples $u_1^N = (u_1, \dots, u_N)$ such that $u_i \in T_t$ whenever $i \in A_t$.

III. POLAR CODES FOR SOURCES WITH RECONSTRUCTION ALPHABET \mathbb{Z}_{p^r}

In this section, we consider sources whose reconstruction alphabet size q is of the form $q = p^r$ for some prime number p and a positive integer r . In this case, the reconstruction alphabet can be considered as a ring with addition and multiplication modulo p^r . We prove the achievability of the symmetric rate-distortion bound for these sources using polar codes and later in Section IV we will generalize this result to sources with arbitrary finite reconstruction alphabets.

A. \mathbb{Z}_{p^r} Rings

Let $\mathbf{G} = \mathbb{Z}_{p^r} = \{0, 1, 2, \dots, p^r - 1\}$ with addition and multiplication modulo p^r be the input alphabet of the channel, where p is a prime and r is an integer. For $t = 0, 1, \dots, r$, define the subgroups H_t of \mathbf{G} as the set:

$$H_t = p^t \mathbf{G} = \{0, p^t, 2p^t, \dots, (p^{r-t} - 1)p^t\}$$

and for $t = 0, 1, \dots, r$, define the subsets K_t of \mathbf{G} as $K_t = H_t \setminus H_{t+1}$; i.e. K_t is defined as the set of elements of \mathbf{G} which are a multiple of p^t but are not a multiple of p^{t+1} . Note that K_0 is the set of all invertible elements of \mathbf{G} and $K_r = \{0\}$. One can sort the sets $K_0 > K_1 > \dots > K_r$ in a decreasing order of “invertibility” of its elements. Let T_t be a transversal of H_t in \mathbf{G} ; i.e. a subset of \mathbf{G} containing one and only one element from each coset of H_t in \mathbf{G} . One valid choice for T_t is $\{0, 1, \dots, p^t - 1\}$. Note that given H_t and T_t , each element g of \mathbf{G} can be represented uniquely as a sum $g = \hat{g} + \tilde{g}$ where $\hat{g} \in T_t$ and $\tilde{g} \in H_t$.

B. Recursive Channel Transformation

1) *The Basic Channel Transforms:* For a test channel $(\mathcal{U} = \mathbf{G}, \mathcal{X}, W)$ where $|\mathbf{G}| = q$, the channel transformations are given by:

$$W^-(x_1, x_2 | u_1) = \sum_{u'_2 \in \mathbf{G}} \frac{1}{q} W(x_1 | u_1 + u'_2) W(x_2 | u'_2) \quad (1)$$

$$W^+(x_1, x_2, u_1 | u_2) = \frac{1}{q} W(x_1 | u_1 + u_2) W(x_2 | u_2) \quad (2)$$

for $x_1, x_2 \in \mathcal{X}$ and $u_1, u_2 \in \mathbf{G}$. Repeating these operations n times recursively, we obtain $N = 2^n$ channels $W_N^{(1)}, \dots, W_N^{(N)}$. For $i = 1, \dots, N$, these channels are given by:

$$W_N^{(i)}(x_1^N, u_1^{i-1} | u_i) = \sum_{u_{i+1}^N \in \mathbf{G}^{N-i}} \frac{1}{q^{N-i}} W^N(x_1^N | u_1^N G_N) \quad (3)$$

where G_N is the generator matrix for polar codes.

Let V_1^N be a random vector uniformly distributed over \mathbf{G}^N and assume the random vectors V_1^N , U_1^N and X_1^N are distributed over $\mathbf{G}^N \times \mathbf{G}^N \times \mathcal{X}^N$ according to

$$p_{V_1^N U_1^N X_1^N}(v_1^N, u_1^N, x_1^N) = \frac{1}{q^N} \mathbb{1}_{\{u_1^N = v_1^N G_N\}} \prod_{i=1}^N W(x_i | u_i)$$

The conditional probability distribution induced from the above equation is consistent with (3). We use this probability distribution extensively throughout the paper.

C. Encoding and Decoding

Let n be a positive integer and let G_N be the generator matrix for polar codes where $N = 2^n$. Let $\{A_t | 0 \leq t \leq r\}$ be a partition of the index set $\{1, 2, \dots, N\}$. Let b_1^N be an arbitrary element from the set $\bigoplus_{t=0}^r H_t^{A_t}$. Assume that the partition $\{A_t | 0 \leq t \leq r\}$ and the vector b_1^N are known to both the encoder and the decoder. The encoder maps a source sequence x_1^N to a vector $v_1^N \in \mathbf{G}^N$ by the following rule: For $i = 1, \dots, N$, if $i \in A_t$, let v_i be a random element g from the set $b_i + T_t$ picked with probability

$$P(v_i = g) = \frac{P_{V_i | V_1^{i-1}, X_1^N}(g | v_1^{i-1}, x_1^N)}{P_{V_i | V_1^{i-1}, X_1^N}(b_i + T_t | v_1^{i-1}, x_1^N)}$$

This encoding rule is a generalization of the randomized rounding encoding rule used for the binary case in [2].

Given a sequence $v_1^N \in \mathbf{G}^N$, the decoder decodes it to $v_1^N G_N$.

For the sake of analysis we assume that the vector b_1^N is uniformly randomly distributed over the set $\bigoplus_{t=0}^r H_t^{A_t}$ (Although it's common information between the encoder and the decoder). The average distortion is given by $\frac{1}{N} \mathbb{E}\{d(X_1^N, U_1^N)\}$ and the rate of the code is given by

$$R = \frac{1}{N} \sum_{t=0}^r |A_t| \log |T_t| = \sum_{t=0}^r \frac{|A_t|}{N} t \log p$$

D. Test Channel Polarization

The following result has been proved in [5]: For all $\epsilon > 0$, there exists a number $N = N(\epsilon) = 2^{n(\epsilon)}$ and a partition $\{A_0^\epsilon, A_1^\epsilon, \dots, A_r^\epsilon\}$ of $\{1, \dots, N\}$ such that for $t = 0, \dots, r$ and $i \in A_t^\epsilon$, $Z_d(W_N^{(i)}) < O(\epsilon)$ if $d \in H_s$ for $0 \leq s < t$ and $Z_d(W_N^{(i)}) > 1 - O(\epsilon)$ if $d \in H_s$ for $t \leq s < r$. For $t = 0, \dots, r$ and $i \in A_t^\epsilon$, we have $I(W_N^{(i)}) = t \log(p) + O(\epsilon)$ and $Z^t(W_N^{(i)}) = O(\epsilon)$ where

$$Z^t(W) = \frac{1}{|H_t|} \sum_{d \in H_t} Z_d(W)$$

Moreover, as $\epsilon \rightarrow 0$, $\frac{|A_t^\epsilon|}{N} \rightarrow p_t$ for some probabilities p_0, \dots, p_r .

In the next section, we show that for any $\beta < \frac{1}{2}$ and for $t = 0, \dots, r$,

$$\lim_{n \rightarrow \infty} P\left((Z^t)^{(n)} > 1 - 2^{-2^{\beta n}}\right) \geq P\left((Z^t)^{(\infty)} = 1\right) \quad (4)$$

$$= 1 - \sum_{s=t}^r p_s$$

Remark III.1. This observation implies the following stronger result: For all $\epsilon > 0$, there exists a number $N = N(\epsilon) = 2^{n(\epsilon)}$ and a partition $\{A_0^\epsilon, A_1^\epsilon, \dots, A_r^\epsilon\}$ of $\{1, \dots, N\}$ such that for $t = 0, \dots, r$ and $i \in A_t^\epsilon$,

$I(W_N^{(i)}) = t \log(p) + O(\epsilon)$ and $Z^t(W_N^{(i)}) > 1 - 2^{-2^{\beta n(\epsilon)}}$. Moreover, as $\epsilon \rightarrow 0$, $\frac{|A_t|}{N} \rightarrow p_t$ for some probabilities p_0, \dots, p_r .

E. Rate of Polarization

In this section we derive a rate of polarization result for the source coding problem. In this proof, we do not assume q is a power of a prime and hence the rate of polarization result derived in this section is valid for the general case. It is shown in [6] (with a slight generalization) that if a random process Z_n satisfies the following two properties

$$Z_{n+1} \leq kZ_n \text{ w.p. } \frac{1}{2} \quad (5)$$

$$Z_{n+1} \leq Z_n^2 \text{ w.p. } \frac{1}{2} \quad (6)$$

for some constant k , then for any $\beta < \frac{1}{2}$, $\lim_{n \rightarrow \infty} P(Z_n < 2^{-2^{\beta n}}) = P(Z_\infty = 0)$. We prove that the random process \tilde{D}_d^n satisfied these properties. First note that by definition

$$\tilde{D}_d(W^+) = \frac{1}{2q} \sum_{u_2 \in \mathbf{G}} \sum_{\substack{x_1, x_2 \in \mathcal{X} \\ u_1 \in \mathbf{G}}} \left[\frac{1}{q} W(x_1|u_1 + u_2) W(x_2|u_2) - \frac{1}{q} W(x_1|u_1 + u_2 + d) W(x_2|u_2 + d) \right]^2$$

If we add and subtract the term $\frac{1}{q} W(x_1|u_1 + u_2) W(x_2|u_2 + d)$ to the term inside brackets and use the inequality $(a + b)^2 \leq 2(a^2 + b^2)$ with

$$\begin{aligned} a &= W(x_1|u_1 + u_2) W(x_2|u_2) - W(x_1|u_1 + u_2) W(x_2|u_2 + d) \\ b &= W(x_1|u_1 + u_2) W(x_2|u_2 + d) - W(x_1|u_1 + u_2 + d) W(x_2|u_2 + d) \end{aligned}$$

we obtain

$$\begin{aligned} \tilde{D}_d(W^+) &\leq \frac{1}{2q} \sum_{u_2 \in \mathbf{G}} \sum_{x_1, x_2 \in \mathcal{X}, u_1 \in \mathbf{G}} \frac{2}{q^2} \\ &\left[(W(x_1|u_1 + u_2) W(x_2|u_2) - W(x_1|u_1 + u_2) W(x_2|u_2 + d))^2 \right. \\ &\quad + (W(x_1|u_1 + u_2) W(x_2|u_2 + d) \\ &\quad \left. - W(x_1|u_1 + u_2 + d) W(x_2|u_2 + d))^2 \right] \quad (7) \end{aligned}$$

This summation can be expanded into two separate summations. For the first summation, we have

$$\begin{aligned} &\frac{1}{2q} \sum_{u_2 \in \mathbf{G}} \sum_{x_1, x_2 \in \mathcal{X}, u_1 \in \mathbf{G}} \frac{2}{q^2} \\ &(W(x_1|u_1 + u_2) W(x_2|u_2) - W(x_1|u_1 + u_2) W(x_2|u_2 + d))^2 \\ &\leq \frac{2}{q^2} \frac{1}{2q} \sum_{u_2 \in \mathbf{G}} \sum_{x_1, x_2 \in \mathcal{X}, u_1 \in \mathbf{G}} \\ &\quad W(x_1|u_1 + u_2)^2 (W(x_2|u_2) - W(x_2|u_2 + d))^2 \\ &\leq \frac{2q}{q^2} \frac{1}{2q} \sum_{u_2 \in \mathbf{G}} \sum_{x_2 \in \mathcal{X}} (W(x_2|u_2) - W(x_2|u_2 + d))^2 \\ &= \frac{2}{q} \tilde{D}_d(W) \quad (8) \end{aligned}$$

Similarly, for the second summation we can show that

$$\begin{aligned} &\frac{1}{2q} \sum_{u_2 \in \mathbf{G}} \sum_{\substack{x_1, x_2 \in \mathcal{X} \\ u_1 \in \mathbf{G}}} \frac{2}{q^2} (W(x_1|u_1 + u_2) W(x_2|u_2 + d) \\ &\quad - W(x_1|u_1 + u_2 + d) W(x_2|u_2 + d))^2 \\ &\leq \frac{2}{q} \tilde{D}_d(W) \quad (9) \end{aligned}$$

Therefore, it follows from (7), (8) and (9) that condition (5) is satisfied for $k = \frac{4}{q}$.

Next we show that $\tilde{D}_d(W^-) \leq (\tilde{D}_d(W))^2$. Note that

$$\begin{aligned} \tilde{D}_d(W^-) &= \frac{1}{2q} \sum_{v_1 \in \mathbf{G}} \sum_{x_1, x_2 \in \mathcal{X}} \left[\left(\frac{1}{q} \sum_{v_2 \in \mathbf{G}} W(x_1|v_1 + v_2) W(x_2|v_2) \right) \right. \\ &\quad \left. - \left(\frac{1}{q} \sum_{v_2 \in \mathbf{G}} W(x_1|v_1 + d + v_2) W(x_2|v_2) \right) \right]^2 \\ &= \frac{1}{2q} \sum_{v_1 \in \mathbf{G}} \sum_{x_1, x_2 \in \mathcal{X}} \frac{1}{q^2} \left[\sum_{v_2 \in \mathbf{G}} W(x_2|v_2) (W(x_1|v_1 + v_2) \right. \\ &\quad \left. - W(x_1|v_1 + d + v_2)) \right]^2 \end{aligned}$$

The squared term in the brackets can be expanded as

$$\begin{aligned} &\frac{1}{q^2} \sum_{v_2, v_2' \in \mathbf{G}} W(x_2|v_2) W(x_2|v_2') (W(x_1|v_1 + v_2) - \\ &\quad W(x_1|v_1 + d + v_2)) (W(x_1|v_1 + v_2') - W(x_1|v_1 + d + v_2')) \end{aligned}$$

Therefore, $\tilde{D}_d(W^-)$ can be written as a summation of four terms $\tilde{D}_d(W^-) = D_1^- + D_2^- + D_3^- + D_4^-$ where

$$\begin{aligned} D_1^- &= \frac{1}{2q} \sum_{v_1 \in \mathbf{G}} \sum_{x_1, x_2 \in \mathcal{X}} \frac{1}{q^2} \sum_{v_2, v_2' \in \mathbf{G}} W(x_2|v_2) \\ &\quad W(x_2|v_2') W(x_1|v_1 + v_2) W(x_1|v_1 + v_2') \\ D_2^- &= \frac{1}{2q} \sum_{v_1 \in \mathbf{G}} \sum_{x_1, x_2 \in \mathcal{X}} \frac{1}{q^2} \sum_{v_2, v_2' \in \mathbf{G}} W(x_2|v_2) \\ &\quad W(x_2|v_2') W(x_1|v_1 + v_2 + d) W(x_1|v_1 + v_2' + d) \\ D_3^- &= \frac{1}{2q} \sum_{v_1 \in \mathbf{G}} \sum_{x_1, x_2 \in \mathcal{X}} \frac{1}{q^2} \sum_{v_2, v_2' \in \mathbf{G}} W(x_2|v_2) \\ &\quad W(x_2|v_2') W(x_1|v_1 + v_2) W(x_1|v_1 + v_2' + d) \\ D_4^- &= \frac{1}{2q} \sum_{v_1 \in \mathbf{G}} \sum_{x_1, x_2 \in \mathcal{X}} \frac{1}{q^2} \sum_{v_2, v_2' \in \mathbf{G}} W(x_2|v_2) \\ &\quad W(x_2|v_2') W(x_1|v_1 + v_2 + d) W(x_1|v_1 + v_2' + d) \end{aligned}$$

For $d \in \mathbf{G}$ define

$$S_d(W) = \frac{1}{2q} \sum_{v \in \mathbf{G}} \sum_{x \in \mathcal{X}} W(x|v) W(x|v + d)$$

Note that $S_d(W) = S_{-d}(W)$. We have

$$\begin{aligned}
D_1^- &= \frac{1}{q^2} \sum_{x_2 \in \mathcal{X}} \sum_{v_2, v'_2 \in \mathbf{G}} W(x_2|v_2)W(x_2|v'_2) \\
&= \frac{1}{2q} \sum_{v_1 \in \mathbf{G}} \sum_{x_1 \in \mathcal{X}} W(x_1|v_1 + v_2)W(x_1|v_1 + v'_2) \\
&= \frac{1}{q^2} \sum_{x_2 \in \mathcal{X}} \sum_{v_2, v'_2 \in \mathbf{G}} W(x_2|v_2)W(x_2|v'_2)S_{v_2-v'_2}(W) \\
&= \frac{1}{q^2} \sum_{x_2 \in \mathcal{X}} \sum_{v_2 \in \mathbf{G}} \sum_{\substack{a \in \mathbf{G} \\ v'_2 = v_2 - a}} W(x_2|v_2)W(x_2|v_2 - a)S_a(W) \\
&= \frac{2}{q} \sum_{a \in \mathbf{G}} S_a(W) \frac{1}{2q} \sum_{v_2 \in \mathbf{G}} \sum_{x_2 \in \mathcal{X}} W(x_2|v_2)W(x_2|v_2 - a) \\
&= \frac{2}{q} \sum_{a \in \mathbf{G}} S_a(W)S_{-a}(W) \\
&= \frac{2}{q} \sum_{a \in \mathbf{G}} S_a(W)^2
\end{aligned}$$

With similar arguments we can show that

$$\begin{aligned}
D_2^- &= \frac{2}{q} \sum_{a \in \mathbf{G}} S_a(W)^2 \\
D_3^- &= \frac{2}{q} \sum_{a \in \mathbf{G}} S_a(W)S_{a-d}(W) \\
D_4^- &= \frac{2}{q} \sum_{a \in \mathbf{G}} S_a(W)S_{a-d}(W)
\end{aligned}$$

Therefore

$$\begin{aligned}
\tilde{D}_d(W^-) &= \frac{4}{q} \sum_{a \in \mathbf{G}} (S_a(W)^2 - S_a(W)S_{a-d}(W)) \\
&= \frac{2}{q} \sum_{a \in \mathbf{G}} (S_a(W) - S_{a-d}(W))^2
\end{aligned}$$

Note that

$$\begin{aligned}
\tilde{D}_d(W) &= \frac{1}{2q} \sum_{v \in \mathbf{G}} \sum_{x \in \mathcal{X}} (W(x|v) - W(x|v+d))^2 \\
&= 2S_0(W) - 2S_d(W)
\end{aligned}$$

Therefore

$$\left(\tilde{D}_d(W)\right)^2 = 4(S_0(W) - S_d(W))^2$$

To show that $\tilde{D}_d(W^-) \leq \left(\tilde{D}_d(W)\right)^2$ it suffices to show that $S_a(W) - S_{a-d}(W) \leq S_0(W) - S_d(W)$. We will make use of the rearrangement inequality:

Lemma III.1. *Let π be an arbitrary permutation of the set $\{1, \dots, n\}$. If $a_1 \leq \dots \leq a_n$ and $b_1 \leq \dots \leq b_n$ then*

$$\sum_{i=1}^n a_i b_i \geq \sum_{i=1}^n a_i b_{\pi(i)}$$

The rearrangement inequality implies that $S_0(W) - S_d(W) \geq S_a(W) - S_{a-d}(W)$. Therefore it follows that condition (6) is also satisfied and hence $\lim_{n \rightarrow \infty} P(\tilde{D}_d^n < 2^{-2^{\beta_n}}) = P(\tilde{D}_d^\infty = 0)$. It has been shown in Appendix D of [5] that $\tilde{D}_d(W) < \epsilon$ implies $Z_d(W) > 1 - \epsilon$. This completes the rate of polarization result.

F. Polar Codes Achieve the Rate-Distortion Bound

The average distortion for the encoding and decoding rules described in Section III-C is given by

$$\begin{aligned}
D_{avg} &= \sum_{x_1^N \in \mathcal{X}^N} p_X^N(x_1^N) \sum_{b_1^N \in \bigoplus_{t=0}^r H_t^{A_t}} \sum_{v_1^N \in b_1^N + \bigoplus_{t=0}^r T_t^{A_t}} \\
&\quad \left(\prod_{t=0}^r \prod_{i \in A_t} \frac{P_{V_i|V_1^{i-1}, X_1^N}(g|v_1^{i-1}, x_1^N)}{P_{V_i|V_1^{i-1}, X_1^N}(b_i + T_t|v_1^{i-1}, x_1^N)} \right) \\
&\quad \left(\prod_{t=0}^r \frac{1}{|H_t|^{A_t}} \right) d(x_1^N, v_1^N G_N) \\
&= \sum_{x_1^N \in \mathcal{X}^N} p_X^N(x_1^N) \sum_{b_1^N \in \bigoplus_{t=0}^r H_t^{A_t}} \sum_{v_1^N \in b_1^N + \bigoplus_{t=0}^r T_t^{A_t}} \\
&\quad \left(\prod_{t=0}^r \prod_{i \in A_t} \frac{P_{V_i|V_1^{i-1}, X_1^N}(g|v_1^{i-1}, x_1^N)}{P_{V_i|V_1^{i-1}, X_1^N}(b_i + T_t|v_1^{i-1}, x_1^N) \cdot |H_t|} \right) \\
&\quad d(x_1^N, v_1^N G_N)
\end{aligned}$$

This can be written as

$$D_{avg} = \mathbb{E}_Q\{d(X_1^N, V_1^N G_N)\}$$

where the distribution Q is defined by

$$Q(v_i|v_1^{i-1}, x_1^N) = \frac{P_{V_i|V_1^{i-1}, X_1^N}(v_i|v_1^{i-1}, x_1^N)}{P_{V_i|V_1^{i-1}, X_1^N}(b_i + T_t|v_1^{i-1}, x_1^N) \cdot |H_t|}$$

and

$$Q(x_1^N) = p_X^N(x_1^N)$$

and hence

$$\begin{aligned}
Q(v_1^N, x_1^N) &= \prod_{i=1}^N Q(v_i|v_1^{i-1}, x_1^N) \\
&= \prod_{t=0}^r \prod_{i \in A_t} \frac{P_{V_i|V_1^{i-1}, X_1^N}(v_i|v_1^{i-1}, x_1^N)}{P_{V_i|V_1^{i-1}, X_1^N}(b_i + T_t|v_1^{i-1}, x_1^N) \cdot |H_t|}
\end{aligned}$$

Recall that

$$P(v_1^N, x_1^N) = \prod_{i=1}^N P_{V_i|V_1^{i-1}, X_1^N}(v_i|v_1^{i-1}, x_1^N)$$

The total variation distance between the distributions P and Q is given by

$$\begin{aligned}
\|P - Q\|_{\text{t.v.}} &= \sum_{v_1^N \in \mathbf{G}^N, x_1^N \in \mathcal{X}^N} |Q(v_1^N, x_1^N) - P(v_1^N, x_1^N)| \\
&= \sum_{x_1^N \in \mathcal{X}^N} p_X^N(x_1^N) \sum_{v_1^N \in \mathbf{G}^N} |Q(v_1^N|x_1^N) - P(v_1^N|x_1^N)|
\end{aligned}$$

We have

$$\begin{aligned}
& \sum_{v_1^N \in \mathbf{G}^N} |Q(v_1^N | x_1^N) - P(v_1^N | x_1^N)| \\
&= \sum_{v_1^N \in \mathbf{G}^N} \left| \left(\prod_{t=0}^r \prod_{i \in A_t} \frac{P(v_i | v_1^{i-1}, x_1^N)}{P(b_i + T_t | v_1^{i-1}, x_1^N) \cdot |H_t|} \right) \right. \\
&\quad \left. - \left(\prod_{t=0}^r \prod_{i \in A_t} P(v_i | v_1^{i-1}, x_1^N) \right) \right| \\
&\stackrel{(a)}{=} \sum_{v_1^N \in \mathbf{G}^N} \left| \sum_{t=0}^r \sum_{i \in A_t} \left[\left(\frac{P(v_i | v_1^{i-1}, x_1^N)}{P(b_i + T_t | v_1^{i-1}, x_1^N) \cdot |H_t|} - P(v_i | v_1^{i-1}, x_1^N) \right) \right. \right. \\
&\quad \left. \cdot \left(\prod_{t=0}^r \prod_{\substack{j=1 \\ j \in A_t}}^{i-1} P(v_j | v_1^{j-1}, x_1^N) \right) \right. \\
&\quad \left. \cdot \left(\prod_{t=0}^r \prod_{\substack{j=i+1 \\ j \in A_t}}^N \frac{P(v_j | v_1^{j-1}, x_1^N)}{P(b_j + T_t | v_1^{j-1}, x_1^N) \cdot |H_t|} \right) \right] \right|
\end{aligned}$$

where in (a) we used the telescopic inequality introduced in [2]. It is straightforward to show that

$$\|P - Q\|_{\text{t.v.}} \leq \sum_{t=0}^r \sum_{i \in A_t} \mathbb{E} \left\{ \left| \frac{1}{P(b_i + T_t | v_1^{i-1}, x_1^N) \cdot |H_t|} - 1 \right| \right\}$$

It has been shown in Appendix D of [5] that if $Z_d(W) > 1 - \epsilon$ then $D_d(W) \leq 2\epsilon - \epsilon^2$. Therefore if $Z_d(W_N^{(i)}) > 1 - \epsilon$ for all $d \in H$ we have

$$\begin{aligned}
D_d(W_N^{(i)}) &= \frac{1}{2q} \sum_{v_i \in \mathbf{G}} \sum_{v_1^{i-1} \in \mathbf{G}^{i-1}, x_1^N \in \mathcal{X}^N} \\
&\quad \left| W_N^{(i)}(v_1^{i-1}, x_1^N | v_i) - W_N^{(i)}(v_1^{i-1}, x_1^N | v_i + d) \right| \\
&\leq 2\epsilon - \epsilon^2
\end{aligned}$$

Therefore for all $v_i \in \mathbf{G}$

$$\begin{aligned}
& \sum_{\substack{v_1^{i-1} \in \mathbf{G}^{i-1} \\ x_1^N \in \mathcal{X}^N}} \left| W_N^{(i)}(v_1^{i-1}, x_1^N | v_i) - W_N^{(i)}(v_1^{i-1}, x_1^N | v_i + d) \right| \\
&\leq 2q(2\epsilon - \epsilon^2)
\end{aligned}$$

We have

$$\begin{aligned}
& \mathbb{E} \left\{ \left| \frac{1}{|H_t|} - P(T_t | V_1^{i-1}, X_1^N) \right| \right\} \\
&= \sum_{v_1^{i-1} \in \mathbf{G}^{i-1}, x_1^N \in \mathcal{X}^N} P(v_1^{i-1}, x_1^N) \left| \frac{1}{|H_t|} - P(T_t | v_1^{i-1}, x_1^N) \right| \\
&= \sum_{\substack{v_1^{i-1} \in \mathbf{G}^{i-1} \\ x_1^N \in \mathcal{X}^N}} \left| \frac{1}{|H_t|} P(v_1^{i-1}, x_1^N) - \sum_{g \in T_t} P(g, v_1^{i-1}, x_1^N) \right| \\
&= \sum_{v_1^{i-1} \in \mathbf{G}^{i-1}, x_1^N \in \mathcal{X}^N} \left| \sum_{g \in T_t} \left[\frac{1}{|\mathbf{G}|} W(v_1^{i-1}, x_1^N | g) \right. \right. \\
&\quad \left. \left. - \sum_{d \in H} \frac{1}{|H_t| \cdot |\mathbf{G}|} W(v_1^{i-1}, x_1^N | g + d) \right] \right| \\
&\leq \sum_{v_1^{i-1} \in \mathbf{G}^{i-1}, x_1^N \in \mathcal{X}^N} \frac{1}{|\mathbf{G}|} \sum_{g \in T_t} \left| [W(v_1^{i-1}, x_1^N | g) \right. \\
&\quad \left. - \sum_{d \in H} \frac{1}{|H_t|} W(v_1^{i-1}, x_1^N | g + d)] \right| \\
&\leq \frac{1}{|\mathbf{G}|} \sum_{g \in T_t} \frac{1}{|H_t|} \sum_{d \in H} \sum_{v_1^{i-1} \in \mathbf{G}^{i-1}, x_1^N \in \mathcal{X}^N} |W(v_1^{i-1}, x_1^N | g) - W(v_1^{i-1}, x_1^N | g + d)| \\
&\leq \frac{2q(2\epsilon - \epsilon^2)}{|H_t|}
\end{aligned}$$

Therefore for if for all $d \in H_t$, $Z_d > 1 - \epsilon$ then for $\delta = \frac{2q(2\epsilon - \epsilon^2)}{|H_t|}$ we have

$$\mathbb{E}_P \left\{ \left| \frac{1}{|H_t|} - P(b_i + T_t | V_1^{i-1}, X_1^N) \right| \right\} < \delta$$

A similar argument as in [2] implies that

$$\begin{aligned}
D_{\text{avg}} &= \frac{1}{N} \mathbb{E}_Q \{ d(X_1^N, V_1^N G_N) \} \\
&\leq \frac{1}{N} \mathbb{E}_P \{ d(X_1^N, V_1^N G_N) \} + \frac{1}{N} \sum_{t=0}^r |A_t| d_{\text{max}} \delta
\end{aligned}$$

where d_{max} is the maximum value of the distortion function. Note that

$$\mathbb{E}_P \{ d(X_1^N, V_1^N G_N) \} = ND$$

Therefore,

$$D_{\text{avg}} \leq D + \frac{1}{N} \sum_{t=0}^r |A_t| d_{\text{max}} \delta$$

Note that from the rate of polarization derived in Section III-E we can choose ϵ to be $\epsilon = 2^{-2^{\theta n}}$. This implies that as $n \rightarrow \infty$, $D + \frac{1}{N} \sum_{t=0}^r |A_t| d_{\text{max}} \delta \rightarrow D$. Also note that the rate of the code $R = \sum_{t=0}^r \frac{|A_t|}{N} t \log p$ converges to $\sum_{t=0}^r p_t t \log p$ and this last quantity is equal to the symmetric capacity of the test channel since the mutual information is a martingale. This means the rate $I(X; U)$ is achievable with distortion D .

IV. ARBITRARY RECONSTRUCTION ALPHABETS

For an arbitrary Abelian group \mathbf{G} , The following polarization result has been provided in [5]: For all $\epsilon > 0$, there exists a number $N = N(\epsilon) = 2^{n(\epsilon)}$ and a partition $\{A_H^\epsilon | H \leq \mathbf{G}\}$ of $\{1, \dots, N\}$ such that for $H \leq \mathbf{G}$ and $i \in A_H^\epsilon$, $Z_d(W_N^{(i)}) < O(\epsilon)$ if $d \in H$ and $Z_d(W_N^{(i)}) > 1 - O(\epsilon)$ if $d \notin H$. For $H \leq \mathbf{G}$ and $i \in A_H^\epsilon$, we have $I(W_N^{(i)}) = \log \frac{|\mathbf{G}|}{|H|} + O(\epsilon)$ and $Z^H(W_N^{(i)}) = O(\epsilon)$ where

$$Z^H(W) = \frac{1}{|H|} \sum_{d \in H} Z_d(W)$$

Moreover, as $\epsilon \rightarrow 0$, $\frac{|A_H^\epsilon|}{N} \rightarrow p_H$ for some probabilities $p_H, H \leq \mathbf{G}$.

As mentioned earlier the rate of polarization result derived in Section III-E is valid for the general case. Therefore it follows that for any $\beta < \frac{1}{2}$ and for $H \leq \mathbf{G}$,

$$\lim_{n \rightarrow \infty} P\left((Z^H)^{(n)} > 1 - 2^{-2^{\beta n}}\right) \geq P\left((Z^H)^{(\infty)} = 1\right) \quad (10)$$

$$= 1 - \sum_{S \leq H} p_S$$

Remark IV.1. *This observation implies the following stronger result: For all $\epsilon > 0$, there exists a number $N = N(\epsilon) = 2^{n(\epsilon)}$ and a partition $\{A_H^\epsilon | H \leq \mathbf{G}\}$ of $\{1, \dots, N\}$ such that for $H \leq \mathbf{G}$ and $i \in A_H^\epsilon$, $I(W_N^{(i)}) = \log \frac{|\mathbf{G}|}{|H|} + O(\epsilon)$ and $Z^H(W_N^{(i)}) > 1 - 2^{-2^{\beta n(\epsilon)}}$. Moreover, as $\epsilon \rightarrow 0$, $\frac{|A_H^\epsilon|}{N} \rightarrow p_H$ for some probabilities $p_H, H \leq \mathbf{G}$.*

The encoding and decoding rules for the general case is as follows: Let n be a positive integer and let G_N be the generator matrix for polar codes where $N = 2^n$. Let $\{A_H | H \leq \mathbf{G}\}$ be a partition of the index set $\{1, 2, \dots, N\}$. Let b_1^N be an arbitrary element from the set $\bigoplus_{H \leq \mathbf{G}} H^{A_H}$. Assume that the partition $\{A_H | H \leq \mathbf{G}\}$ and the vector b_1^N are known to both the encoder and the decoder. The encoder maps a source sequence x_1^N to a vector $v_1^N \in \mathbf{G}^N$ by the following rule:

For a subgroup H of \mathbf{G} , let T_H be a transversal of H in \mathbf{G} . For $i = 1, \dots, N$, if $i \in A_H$, let v_i be a random element g from the set $b_i + T_H$ picked with probability

$$P(v_i = g) = \frac{P_{V_i | V_1^{i-1}, X_1^N}(g | v_1^{i-1}, x_1^N)}{P_{V_i | V_1^{i-1}, X_1^N}(b_i + T_H | v_1^{i-1}, x_1^N)}$$

Given a sequence $v_1^N \in \mathbf{G}^N$, the decoder decodes it to $v_1^N G_N$.

It follows from the analysis of the \mathbb{Z}_{p^r} case in a straightforward fashion that this encoding/decoding scheme achieves the symmetric rate-distortion bound when the group \mathbf{G} is an arbitrary Abelian group.

V. CONCLUSION

We have shown that polar codes achieve the symmetric rate-distortion bound for finite-alphabet sources regardless of the size of the reconstruction alphabet.

REFERENCES

- [1] E. Arikan, "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.
- [2] S. B. Korada and R. Urbanke, "Polar Codes are Optimal for Lossy Source Coding," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1751–1768, Apr. 2010.
- [3] M. Karzand and E. Telatar, "Polar Codes for Q-ary Source Coding," *Proceedings of IEEE International Symposium on Information Theory*, 2010, austin, TX.
- [4] A. G. Sahebi and S. S. Pradhan, "Multilevel Polarization of Polar Codes Over Arbitrary Discrete Memoryless Channels," *Proc. 49th Allerton Conference on Communication, Control and Computing*, Sept. 2011.
- [5] —, "Multilevel Polarization of Polar Codes over Arbitrary Discrete Memoryless Channels," July 2011, online: <http://arxiv.org/abs/1107.1535>.
- [6] E. Arikan and E. Telatar, "On the rate of channel polarization," *Proceedings of IEEE International Symposium on Information Theory*, 2009, Seoul, Korea.