Controlled polarization for q-ary alphabets

Woomyoung Park* Alexander Barg‡

***Abstract*—We design polar codes for q-ary input, $q = 2^r$, that polarize to an arbitrary given subset of extremal configurations out of the original triangular array of $r+1$ such configurations. We also obtain a new proof of the known result about 2-level polarization of q-ary codes.**

I. Introduction

Polarization is a new paradigm in discrete-time random processes with applications in information theory. In the context of channel coding, polarization amounts to the statement that it is possible to isolate extremal dependencies by applying transformations to the input data. The original construction of polar codes by Arıkan [1] relies on a linear transformation of input bits that amounts to constructing special subcodes of Reed-Muller codes, and implies polarization of bit channels to virtual channels with capacity 0 and 1. Application of Arıkan's construction to nonbinary alphabets leads to multi-level polarization [2], [3], [4]. In particular, for alphabets of size $q = 2^r$, channels for individual data symbols polarize to $r+1$ levels, resulting in symbols that transmit almost perfectly $0, 1, \ldots, r+1$ bits. At the same time, using carefully designed nonlinear transformations it is possible to attain two-level polarization for nonbinary alphabets [5]. In this paper we advance the approach of [3], [5]. We design data transformations (polarization kernels) that, when applied to q-ary symbols, result in any pre-defined set of polarization levels (throughout we assume that $q = 2^r$). As an example, suppose that $q = 2^3$. In the construction of [3] the channels polarize to 4 levels that correspond to 0,1,2, and 3 bits in data symbols. Suppose that we would like to remove the extremal configuration that corresponds to channels with capacity 1 bit. Our construction answers this question as well as provides a way to polarize symbols to any given subset of capacities in $\{0, 1, \ldots, r+1\}$ that contains 0 and $r+1$, for any $r \geq 2$. As in [3], to one type of symbols there corresponds a unique extremal configuration, which makes it easier to construct codes. As a byproduct, we also obtain a new proof of the polarization result of [5].

The results of [3]. Let $W : \mathcal{X} \to \mathcal{Y}$ be a discrete memoryless channel. The construction of binary polar codes [1] relies on the transformation based on recursive application of the linear operation $x_1 = u_1 \oplus u_2, x_2 = u_2$ to the data symbols to produce channel symbols. The same approach was used in [3] for q-ary input alphabets, $q = 2^r$. For $v \in \mathcal{X}, v = (v_1, v_2, \ldots, v_r)$ we denote $\mathrm{wt}_{Hr}(v) = \max(i : v_i = 1)$ (the location of the rightmost nonzero bit in the binary expansion of v). Let $\mathcal{X}_i = \{x \in \mathcal{X} : \mathrm{wt}_{Hr}(x) = i\}$. Define $Z(W_{\{x_1,x_2\}}) = \sum_{y\in\mathcal{Y}} \sqrt{W(y|x_1)W(y|x_2)}$ and

$$Z_v(W) = \frac{1}{q} \sum_{x\in\mathcal{X}} Z(W_{\{x,x+v\}}), \qquad v \in \mathcal{X}\backslash\{0\}$$

$$Z_i(W) = \frac{1}{2^{i-1}} \sum_{v\in\mathcal{X}_i} Z_v(W).$$

The main result of [3], which is also the starting point of the present work, states that the virtual channels for the data symbols converge to one of $r+1$ extremal configurations characterized by the vector of values of $(Z_i, 1 \leq i \leq r)$. As the recursion depth $n \to \infty$, with probability one

$$(Z_{i,n}, 1 \leq i \leq r) \in \{(1^k 0^{r-k}), k = 0, 1, \ldots, r\}. \qquad (1)$$

II. Two-level polarization

In the nonbinary case [3], addition is performed modulo q, so the polarization kernel becomes nonlinear. A different non-linear transformation, suggested by Şaşoğlu in [5], achieves full (i.e., two-level) polarization for nonbinary DMCs with arbitrary-size input alphabets. His proof relies on establishing entropy polarization, i.e., tracking the behavior of conditional entropies of the transmitted symbols and proving polarization based on their convergence. Şaşoğlu [5] designed a set of kernels that force the entropies to approach 0 and 1, thereby polarizing the data symbols into fully noisy and almost noiseless. In this section, we give another proof of Şaşoğlu's result, establishing two-level polarization for $q = 2^r$ relying on Bhattacharyya distances rather than on entropies. Our proof draws on methods developed in [3], serving as a warm-up for the construction in the next section.

Suppose that $g : \mathcal{X}^2 \to \mathcal{X}$ is a map that combines two data symbols into one channel symbol. Let the combined channel W_2 and the channels W^- and W^+ be defined as follows:

$$W_2(y_1, y_2|u_1, u_2) = W(y_1|g(u_1, u_2))W(y_2|u_2)$$

$$W^-(y_1, y_2|u_1) = \sum_{u_2\in\mathcal{X}} \frac{1}{q} W_2(y_1, y_2|u_1, u_2)$$

$$W^+(y_1, y_2, u_1|u_2) = (1/q)W_2(y_1, y_2|u_1, u_2),$$

where $u_1, u_2 \in \mathcal{X}, y_1, y_2 \in \mathcal{Y}$.

Define g by the relation

$$g(u_1, u_2) = u_1 + \pi(u_2) \mod q \qquad (2)$$

* Samsung Electronics, Suwon, Gyeonggi-do, Korea. Research done while at Department of ECE and Institute for Systems Research, University of Maryland, College Park, MD. Email: wm6110.park@samsung.com. Research supported by NSF grant CCF0916919.

‡Department of ECE and Institute for Systems Research, University of Maryland, College Park, MD. E-mail: abarg@umd.edu. Research supported by NSF grants CCF1217245, CCF1217894, and NSA H98230-12-1-0260.

where $\pi : \mathcal{X} \to \mathcal{X}$ is a permutation with the following property: there is at least one number $x \in \mathcal{X}$ satisfying

$$\mathrm{wt}_{Hr}(\pi(x) - \pi(x + q/2)) = r. \tag{3}$$

As an example, one can take

$$\pi(u_2) = \begin{cases} q/2, & \text{if } u_2 = 0 \\ 1 - u_2, & \text{if } 1 \le u_2 \le q/2 \\ -u_2, & \text{if } u_2 > q/2 \end{cases} \tag{4}$$

where the operations are performed modulo q. Note that this map is the inverse of the map used in [5].

We will prove that the channel combining operation (2) supports convergence of the Bhattacharyya parameters.

Proposition 2.1: For $v \in \mathcal{X}$, the quantities $Z_v(W^-)$ and $Z_v(W^+)$ are related to $Z_v(W)$ by

$$Z_v(W^-) \ge Z_v(W), \tag{5}$$

$$Z_v(W^-) \le Z_v(W) + \frac{1}{q}\sum_x \frac{1}{q} \sum_{\substack{u_1,u_2 \\ g(x,u_1)=g(x+v,u_2)}} Z(W_{\{u_1,u_2\}}) + \frac{1}{q}\sum_x \frac{1}{q} \sum_{\substack{u_1,u_2 \\ u_1\neq u_2 \\ g(x,u_1)\neq g(x+v,u_2)}} Z(W_{\{g(x,u_1),g(x+v,u_2)\}})Z(W_{\{u_1,u_2\}}), \tag{6}$$

$$Z_v(W^+) = \frac{1}{q}\sum_x Z_{s(x,x+v)}(W)Z(W_{\{x,x+v\}}) \tag{7}$$

where $s(x_1, x_2) = \pi(x_1) - \pi(x_2) \mod q$ and $x_1, x_2 \in \mathcal{X}$.

Proof: First let us prove the upper bound for $Z_v(W^-)$. From the definitions of $Z_v(\cdot)$ and W^- we have

$$\begin{aligned}
Z_v(W^-) &= \frac{1}{q}\sum_x Z(W^-_{\{x,x+v\}}) \\
&= \frac{1}{q}\sum_x \sum_{y_1,y_2} \sqrt{W^-(y_1,y_2|x)W^-(y_1,y_2|x+v)} \\
&= \frac{1}{q}\sum_x \frac{1}{q}\sum_{y_1,y_2} \sqrt{\sum_{u_1,u_2} W(y_1|g(x,u_1))W(y_2|u_1)} \\
&\qquad \times \sqrt{W(y_1|g(x+v,u_2))W(y_2|u_2)} \\
&\le \frac{1}{q}\sum_x \frac{1}{q}\sum_{u_1,u_2}\sum_{y_1,y_2} \sqrt{W(y_1|g(x,u_1))W(y_2|u_1)} \\
&\qquad \times \sqrt{W(y_1|g(x+v,u_2))W(y_2|u_2)} \\
&= \frac{1}{q}\sum_x \frac{1}{q}\sum_u Z(W_{\{g(x,u),g(x+v,u)\}}) \\
&\qquad + \frac{1}{q}\sum_x \frac{1}{q} \sum_{\substack{u_1,u_2 \\ g(x,u_1)=g(x+v,u_2)}} Z(W_{\{u_1,u_2\}}) \\
&\quad + \frac{1}{q}\sum_x \frac{1}{q} \sum_{\substack{u_1,u_2 \\ u_1\neq u_2 \\ g(x,u_1)\neq g(x+v,u_2)}} Z(W_{\{g(x,u_1),g(x+v,u_2)\}})Z(W_{\{u_1,u_2\}}).
\end{aligned}$$

Note that $g(x+v,u) - g(x,u) = x + v + \pi(u) - x - \pi(u) = v$ for any $x \in \mathcal{X}$. Since $x \mapsto g(x,u)$ is a permutation for a fixed u, $\frac{1}{q}\sum_x Z(W_{\{g(x,u),g(x+v,u)\}}) = Z_v(W)$ for any $u \in \mathcal{X}$. Thus the first term becomes $Z_v(W)$, which gives the estimate in (6). The lower bound on $Z_v(W^-)$ is shown as follows:

$$\begin{aligned}
Z_v(W^-) &= \frac{1}{q}\sum_x \frac{1}{q}\sum_{y_1,y_2} \sqrt{\sum_{u_1,u_2} W(y_1|g(x,u_1))} \\
&\qquad \times \sqrt{W(y_2|u_1)W(y_1|g(x+v,u_2))W(y_2|u_2)} \\
&\overset{(*)}{\ge} \frac{1}{q}\sum_x \frac{1}{q}\sum_u \sum_{y_1,y_2} \sqrt{W(y_1|g(x,u))W(y_2|u)} \\
&\qquad \times \sqrt{W(y_1|g(x+v,u))W(y_2|u)} \\
&= \frac{1}{q}\sum_x \frac{1}{q}\sum_u \sum_{y_1} \sqrt{W(y_1|g(x,u))W(y_1|g(x+v,u))} \\
&= \frac{1}{q}\sum_x \frac{1}{q}\sum_u Z(W_{\{g(x,u),g(x+v,u)\}}) \\
&= Z_v(W).
\end{aligned}$$

To prove (7), let us consider the Bhattacharyya distance of W^+ between $x, x' \in \mathcal{X}$, $x \neq x'$.

$$\begin{aligned}
Z(W^+_{\{x,x'\}}) &= \sum_{y_1,y_2,u_1} \sqrt{W^+(y_1,y_2,u_1|x)W^+(y_1,y_2,u_1|x')} \\
&= \frac{1}{q}\sum_{y_1,y_2,u_1} \sqrt{W(y_1|g(u_1,x)))W(y_2|x)} \\
&\qquad \times \sqrt{W(y_1|g(u_1,x'))W(y_2|x')} \\
&= \frac{1}{q}\sum_{u_1} Z(W_{\{g(u_1,x),g(u_1,x')\}})Z(W_{\{x,x'\}}) \\
&= Z_{s'}(W)Z(W_{\{x,x'\}})
\end{aligned}$$

where $s' = g(u_1,x) - g(u_1,x') = u_1 + \pi(x) - u_1 - \pi(x') = \pi(x) - \pi(x') = s(x,x') \mod q$ which is the same for all $u_1 \in \mathcal{X}$. The last equality holds because the set $\{g(u_1,x) : u_1 \in \mathcal{X}\} = \mathcal{X}$ for any $x \in \mathcal{X}$. From the definition of $Z_v(W)$ we now obtain (7). ∎

Our next goal is to prove that the map g given by (2) and (3) supports two-level polarization for channels with input alphabet of size q. Let $Z_{v,n}$ be the random Bhattacharyya distance after n steps of the random walk associated with the code construction. We label the paths of the random walk as $(b_1, b_2, \ldots, b_n, \ldots)$, where $b_n \in \{+,-\}$ for all n.

Lemma 2.2: Suppose that q-ary polar codes are designed using the kernel (2). For $v \in \mathcal{X} \setminus \{0\}$, $Z_{v,n}$ converges a.s. to a $(0,1)$-valued Bernoulli random variable Z_∞ that does not depend on v.

Proof: We begin with proving convergence of $Z_{\max,n} := \max_{v\neq 0} Z_{v,n}$. Let $v_1 = \operatorname{argmax}_{v\in\mathcal{X}} Z_{v,n}$ and $v_2 = \operatorname{argmax}_{v\in\mathcal{X}} Z_{v,n+1}$. Consider the case of moving along the $+$ path from level n to level $n+1$ in the tree. In this case we have

$$\begin{aligned}
Z_{\max,n+1} = Z_{v_2,n+1} &= \frac{1}{q}\sum_x Z_{s(x,x+v_2),n} Z(W_{\{x,x+v_2\},n}) \\
&\le Z_{v_1,n}\frac{1}{q}\sum_x Z(W_{\{x,x+v_2\},n})
\end{aligned}$$

$$= Z_{v_1,n} Z_{v_2,n}$$
$$\leq \big(Z_{\max,n}\big)^2.$$

Here the first inequality follows because, by definition of v_1, the quantity $Z_{s(x,x+v_2),n}$ for all $x \in \mathcal{X}$ can be bounded above by $Z_{v_1,n}$. Suppose that $v_1 = v_2$. Then the last inequality holds with equality. At the same time, if $v_1 \neq v_2$, then $Z_{v_2,n} \leq Z_{v_1,n} = Z_{\max,n}$.

Now consider the case of moving along the $-$ path. For any $v \in \mathcal{X} \setminus \{0\}$, the quantity $Z_{v,n} = \frac{1}{q}\sum_x Z(W_{\{x,x+v\},n}) \leq Z_{\max,n}$ and $Z(W_{\{x,x'\},n}) \leq qZ_{\max,n}$ for all $x, x' \in \mathcal{X}$, $x \neq x'$. Therefore the second term in the upper bound on $Z_{\max,n+1}$ in (6) is bounded above by $qZ_{\max,n}$ and the third term is bounded above by $(q^2 - 2q)Z_{\max,n}$. Now invoke Lemma 4 in [3], in which condition (ii) is replaced with $P(U_{n+1} \leq U_n^2 | \mathcal{F}_n) \geq 1/2$. This change does not affect the proof, so we conclude that $Z_{\max,n}$ converges a.s. to a Bernoulli random variable $Z_{\max,\infty}$.

Having established that $Z_{\max,n}$ converges to 0 or to 1, let us consider both cases. If it is the former, then $Z_{v,n}$ converges to 0 for all v by the definition. If $Z_{\max,n}$ converges to 1, at least one of $Z_{v,n}$ converges to 1, and from the monotone behavior of the Z_v's ([3, Lemma 5]), $Z_{v,n}$ with $\mathrm{wt}_{H_r}(v) = 1$ must converge to 1. Since there exists at least one $x \in \mathcal{X}$ that satisfies condition (3), without loss of generality we can assume that $\mathrm{wt}_{H_r}(s(0, q/2)) = r$. The quantity $Z_{1,n+1}$ with $b_{n+1} = +$ is

$$\begin{aligned}
Z_{1,n+1} &= \frac{2}{q}\big(Z(W_{\{0,q/2\},n+1}) + Z(W_{\{1,q/2+1\},n+1}) + \\
&\qquad \cdots + Z(W_{\{q/2-1,q-1\},n+1})\big) \\
&= \frac{2}{q}\big(Z_{s(0,q/2),n} Z(W_{\{0,q/2\},n}) + Z_{s(1,q/2+1),n} Z(W_{\{1,q/2+1\},n}) \\
&\qquad + \cdots + Z_{s(q/2-1,q-1),n} Z(W_{\{q/2-1,q-1\},n})\big) \\
&\leq \frac{2}{q}\big(Z_{s(0,q/2),n} + \frac{q}{2} - 1\big) \\
&= \frac{2}{q}\big(Z_{s(0,q/2),n} - 1\big) + 1
\end{aligned}$$

where the inequality follows from the fact that $Z(W_{\{x,x+q/2\},n}) \leq 1$ and $Z_{s(i,q/2+i),n} \leq 1, i = 1, \ldots, q/2 - 1$. Let $\Omega_a^{(1)} = \{\omega : Z_{1,n} \to a\}, a = 0, 1$. Since the above inequality holds sample path-wise and $\frac{2}{q}\big(Z_{s(0,q/2),n} - 1\big) + 1 \leq 1$, the sequence of random variables $\frac{2}{q}\big(Z_{s(0,q/2),n} - 1\big) + 1$ converges to 1, and thus $Z_{s(0,q/2),n} \to 1$ on the event $\Omega_1^{(1)}$. Finally, notice that $\mathrm{wt}_{H_r}(s(0, q/2)) = r$, and so from the monotone behavior of the Z_v's we conclude that $Z_{v,n}$ converges to 1 on $\Omega_1^{(1)}$ for all v. This completes the proof. ∎

The final step is to use Lemma 1 in [3] which establishes relations between $I(W)$ and $Z_i(W)$. It implies that if $Z_i(W) = 0$ for $i = 1, \ldots, r$, then $I(W) = r$, and if $Z_i(W) = 1$ for all i, then $I(W) = 0$. Thus, with probability one the virtual channels for the data symbols converge to either perfect channels or purely noisy channels.

III. Controlling Polarization: Any Number of Levels

In this section, we design mappings $(u_1, u_2) \mapsto (x_1, x_2)$ that polarize a DMC with the q-ary input to q-ary channels with capacity k bits, where $k \in T$ and T is a subset of $\{0, 1, \ldots, r\}$ of cardinality $3 \leq |T| \leq r$ that includes 0 and r.

Let k be the running index of the channels in (1) and suppose that capacity of channel k is k. In the previous section, using the map (2), we have removed all the levels from $k = 1$ to $k = r - 1$. The technical tool for accomplishing this is to force the random variables $Z_{i,n}, i = 1, 2, \ldots, r$ to converge to identical copies of a Bernoulli random variable; see also the proof of Lemma 2.2. Generalizing this idea, let us find a mapping that leaves all the extremal configurations except those in rows $r - j$ to $r - i$,, $i \leq j$. In this case, the set of extremal configurations becomes

$$\{1^k 0^{r-k}, k = 0, 1, \ldots, i-1, j+1, \ldots, r\}.$$

To obtain this set, we will make the random variables $Z_{i,n}, Z_{i+1,n}, \ldots, Z_{j+1,n}$ converge to the same $(0, 1)$-valued random variable, thereby collapsing the corresponding levels. For a number n, denote the set $\{0, 1, \ldots, n-1\}$ by $[n]$. Let $q_1 = 2^{j-i+2}$. In order to remove $j - i + 1$ consecutive levels, we define a permutation π by first defining $\pi_{q_1} : [q_1] \mapsto [q_1]$ so that it satisfies condition (3). The resulting permutation π is given by

$$\pi(u) = \begin{cases} 2^{r-j-1} \cdot \pi_{q_1}\big(\frac{u}{2^{r-j-1}}\big), & \text{if } 2^{r-j-1} | u, u < 2^{r-i+1} \\ \pi(u - u'') + u'', & \text{if } 2^{r-j-1} \nmid u, u < 2^{r-i+1} \\ \pi(u') + (u - u'), & \text{otherwise} \end{cases} \tag{8}$$

where $u' = u \mod 2^{r-i+1}$ and $u'' = u \mod 2^{r-j-1}$, and addition is modulo q.

The intuition behind this definition is as follows. Consider the set $B_l = \{g(u_1, u_2) : u_1, u_2 \in \mathcal{X}_0^l\}$, where $\mathcal{X}_0^l = \mathcal{X}_0 \cup \cdots \cup \mathcal{X}_l$, $l \geq 1$ and $\mathcal{X}_0 = \{0\}$. If the mapping g is defined using addition modulo q, then the size of the set B_l is 2^l, $l = 1, \ldots, r$, which is equal to the size of the set $\mathcal{X}_0^l$. At the same time, if we use the mapping g given by (2) and (3), then $|B_l|$ is strictly greater than 2^l when $1 \leq l \leq r - 1$. Based on this, we assume that the lth level is present among the extreme configurations if the size of the set B_l is equal to $|\mathcal{X}_0^l|$. Therefore, in order to remove levels i to j from the set of extremal configurations, we design g so that $|B_l| = |\mathcal{X}_0^l|$ if $l = 1, \ldots, i-1, j+1, \ldots, r$, while $|B_l| > |\mathcal{X}_0^l|$ for $l = i, \ldots, j$. To achieve this goal, we use the permutation $\pi_{q_1}(u)$ if $i \leq \mathrm{wt}_{H_r}(u) \leq j + 1$ and use modulo-q addition otherwise. The value $j+1$ is included for the following reason. For 2-level polarization, we make the random variables $Z_{k,n}$, $k = 1, 2, \ldots, r$ converge to the same $(0, 1)$-valued random variable by using the permutation $\pi(u)$ when $1 \leq \mathrm{wt}_{H_r}(u) \leq r$ to define a mapping g. Likewise, since the random variables $Z_{k,n}$, $k = i, \ldots, j+1$ must converge to the same Bernoulli random variable, we need a permutation $\pi(u)$ of size 2^{j-i+2} for vectors with ordered weight from i to $j + 1$.

The permutation π has the following property.

Proposition 3.1: Suppose that π is given by (8). Then for $u_1, u_2 \in \mathcal{X}$, the ordered weight $\mathrm{wt}_{Hr}(\pi(u_1) - \pi(u_2)) = t$ if and only if $\mathrm{wt}_{Hr}(u_1 - u_2) = t$, $t = 1, 2, \ldots, i-1, j+2, \ldots, r$.
Proof: See [2].

In the next theorem, we prove convergence of $Z_{v,n}$.

Theorem 3.2: Let the mapping $g : \mathcal{X}^2 \to \mathcal{X}$ be given by (8). For all $i = 1, 2, \ldots, r$

$$\lim_{n\to\infty} Z_{i,n} = Z_{i,\infty} \quad a.s.,$$

where $Z_{i,\infty} \in \{0, 1\}$. Moreover, with probability 1, the vector $(Z_{1,\infty}, Z_{2,\infty}, \ldots, Z_{r,\infty})$ is one of the vectors of the form $\{1^i 0^{r-i} : i = 0, 1, \ldots, i-1, j+1, \ldots, r\}$.

Proof: Let us consider the upper bound of $Z_v(W^-)$ first. Let $v \in \mathcal{X}_t$, $t = 1, 2, \ldots, r$. When $g(x, u_1) = g(x + v, u_2)$, $g(x, u_1) - g(x + v, u_2) = -v + \pi(u_1) - \pi(u_2) = 0$ and $\mathrm{wt}_{Hr}(\pi(u_1) - \pi(u_2)) = t$. By Proposition 3.1, $\mathrm{wt}_{Hr}(u_1 - u_2) = t$ if $t = 1, \ldots, i-1, j+2, \ldots, r$. Therefore, the second term on the r.h.s. of (6) in the case $\mathrm{wt}_{Hr}(v) = 1, \ldots, i-1, j+2, \ldots, r$ is bounded above by $qZ^{(t)}_{\max}(W)$ or $qZ^{[t,r]}_{\max}(W)$[1]. If $i \le t \le j+1$, the ordered weight of the vector $u_1 - u_2$ is also between i and $j+1$ so the second term in this case is upper bounded by $qZ^{[i,j+1]}_{\max}(W)$ or $qZ^{[i,r]}_{\max}(W)$.

Consider the case $g(x, u_1) \neq g(x + v, u_2)$. If $\mathrm{wt}_{Hr}(u_1 - u_2) = s \ge t$, then

$$\begin{aligned} Z(W_{\{u_1,u_2\}})Z(W_{\{g(x,u_1),g(x+v,u_2)\}}) &\le qZ^{(s)}_{\max}(W) \\ &\le qZ^{[t,r]}_{\max}(W). \end{aligned}$$

On the other hand, if $\mathrm{wt}_{Hr}(u_1 - u_2) = s < t$, then term inside the last summation in (6) can also be bounded above by $qZ^{[t,r]}_{\max}(W)$ if $t \le i-1$ or $t \ge j+2$ and $qZ^{[i,r]}_{\max}(W)$ if $i \le t \le j+1$. Indeed, consider the difference $g(x, u_1) - g(x + v, u_2) = -v + \pi(u_1) - \pi(u_2)$. If $t \le i-1$ or $t \ge j+2$, then $\mathrm{wt}_{Hr}(\pi(u_1) - \pi(u_2)) = s < t$, so $\mathrm{wt}_{Hr}(g(x, u_1) - g(x + v, u_2)) = t$. This means that the term

$$\begin{aligned} Z(W_{\{u_1,u_2\}})Z(W_{\{g(x,u_1),g(x+v,u_2)\}}) &\le Z(W_{\{g(x,u_1),g(x+v,u_2)\}}) \\ &\le qZ^{(t)}_{\max}(W) \\ &\le qZ^{[t,r]}_{\max}(W), \end{aligned}$$

At the same time, if $i \le t \le j+1$, then the ordered weight of the vector $z = \pi(u_1) - \pi(u_2)$ satisfies $\mathrm{wt}_{Hr}(z) = s$ if $s < i$ or $i \le \mathrm{wt}_{Hr}(z) \le j+1$ if $s \ge i$. This implies that $i \le \mathrm{wt}_{Hr}(g(x, u_1) - g(x + v, u_2)) \le j+1$. Therefore,

$$\begin{aligned} Z(W_{\{u_1,u_2\}})Z(W_{\{g(x,u_1),g(x+v,u_2)\}}) &\le Z(W_{\{g(x,u_1),g(x+v,u_2)\}}) \\ &\le qZ^{[i,j+1]}_{\max}(W) \\ &\le qZ^{[i,r]}_{\max}(W). \end{aligned}$$

Summarizing, the upper bound on $Z_v(W^-)$ is

$$Z_v(W^-) \le Z_v(W) + qZ^{[t,r]}_{\max}(W) + q(q-2)Z^{[t,r]}_{\max}(W)$$

if $t = 1, \ldots, i-1, j+2, \ldots, r$ and

$$Z_v(W^-) \le Z_v(W) + qZ^{[i,r]}_{\max}(W) + q(q-2)Z^{[i,r]}_{\max}(W)$$

if $t = i, \ldots, j+1$. If $\mathrm{wt}_{Hr}(v) = t$,

$$Z_v(W) \le Z^{(t)}_{\max}(W) \le Z^{[t,r]}_{\max}(W) \le Z^{[s,r]}_{\max}(W)$$

for any $s \le t$. Therefore, the term $Z_v(W)$ in the first inequality can be replaced with $Z^{[t,r]}_{\max}(W)$ and the term in the second inequality with $Z^{[i,r]}_{\max}(W)$.

From Proposition 3.1, if $\mathrm{wt}_{Hr}(v) = t$, where $t = 1, \ldots, i-1, j+2, \ldots, r$, then $\mathrm{wt}_{Hr}(s(x, x+v)) = \mathrm{wt}_{Hr}(\pi(x) - \pi(x+v)) = t$ and $Z_{s(x,x+v)} \le Z^{(t)}_{\max}(W)$. Therefore we obtain

$$\begin{aligned} Z_v(W^+) &= \frac{1}{q}\sum_x Z_{s(x,x+v)}(W)Z(W_{\{x,x+v\}}) \\ &\le Z^{(t)}_{\max}(W)Z_v(W) \\ &\le \left(Z^{(t)}_{\max}(W)\right)^2 \le \left(Z^{[t,r]}_{\max}(W)\right)^2. \end{aligned}$$

In addition, if $t = i, \ldots, j+1$, the ordered weight of $s(x, x+v)$ is one of the values in the set $\{i, i+1, \ldots, j+1\}$. Thus,

$$\begin{aligned} Z_v(W^+) &= \frac{1}{q}\sum_x Z_{s(x,x+v)}Z(W_{\{x,x+v\}}) \\ &\le Z^{[i,j+1]}_{\max}(W)Z_v(W) \\ &\le \left(Z^{[i,j+1]}_{\max}(W)\right)^2 \le \left(Z^{[i,r]}_{\max}(W)\right)^2. \end{aligned}$$

This means that the random variable $Z^{[t,r]}_{\max,n}$ converges a.s. to a $(0, 1)$-valued random variable $Z^{[t,r]}_{\max,\infty}$ for $t = 1, 2, \ldots, i-1, i, j+2, \ldots, r$.

Now we will prove that the random variables $Z^{(t)}_{\max,n}$, $t = 1, \ldots, r$ converge a.s. to $Z^{(t)}_{\max,\infty}$ which is $(0, 1)$-valued random variables. By following the induction argument in the proof of Lemma 6 in [3] we can show that $Z^{(t)}_{\max,\infty}$ takes values 0 or 1 with probability 1 and so does $Z_{v,\infty}$, $v \in \mathcal{X}_t$, $t = j+2, \ldots, r$. Assume that we have proved the needed convergence for $Z^{(t)}_{\max,n}$, $t = j+2, \ldots, r$ and let us prove the convergence for $Z^{(s)}_{\max,n}$, $s = i, \ldots, j+1$. Suppose that $Z^{[i,r]}_{\max,\infty} = 0$, then $Z^{(s)}_{\max,\infty} = 0$, $s = i, \ldots, r$. On the other hand if $Z^{[i,r]}_{\max,\infty} = 1$, by the monotonicity principle (Lemma 5 in [3]), $Z^{(i)}_{\max,\infty} = 1$ and $Z_{v,\infty} = 1$ for all $v \in \mathcal{X}_i$. Observe that π_{q_1} is the same permutation as in (2) with q replaced with q_1. Therefore, monotonicity implies that the variables $Z_{v,n}$,$\mathrm{wt}_{Hr}(v) = i, \ldots, j+1$ converge to the same Bernoulli random variable. We conclude that $Z_{i,\infty} = \cdots = Z_{j+1,\infty}$, while configurations from the $(r-j)$th to the $(r-i)$th level occur with probability zero. By applying the same induction argument, we can also obtain the needed convergence for $t = i-1, \ldots, 1$. ∎

Let us give an example of the above mapping.

Example 1: Let $q = 2^3$. In the original construction of [3], the channels polarize to 4 levels, corresponding to 0,1,2, and 3 bits in the data symbols. Suppose we would like to remove the extremal configuration that corresponds to channels with capacity 1. Use the mapping $g(u_1, u_2) = u_1 + \pi(u_2)$ where π is given by (8), where we take $q = 8$, $i = j = 2$, and π_4 a permutation given by (4) with $q = 4$. The matrix G, $G_{i,j} = g(i, j)$, $i, j \in \mathcal{X}$ is given below. From the

[1] here we use the notation $Z^{[i,j]}_{\max} = \max(Z^k_{\max}, i \le k \le j)$.

relations between $Z_v(W^{\pm})$ and $Z_v(W)$ and Theorem 3.2, the random variable $Z_{i,n}$ converges to $Z_{i,\infty}$, $i = 1, 2, 3$, where the vector $(Z_{1,\infty}, Z_{2,\infty}, Z_{3,\infty})$ takes one of the following values: $(0, 0, 0)$, $(1, 0, 0)$, and $(1, 1, 1)$. This gives the desired polarization.

$$G = \begin{pmatrix} 2 & 0 & 3 & 1 & 6 & 4 & 7 & 5 \\ 3 & 1 & 4 & 2 & 7 & 5 & 0 & 6 \\ 4 & 2 & 5 & 3 & 0 & 6 & 1 & 7 \\ 5 & 3 & 6 & 4 & 1 & 7 & 2 & 0 \\ 6 & 4 & 7 & 5 & 2 & 0 & 3 & 1 \\ 7 & 5 & 0 & 6 & 3 & 1 & 4 & 2 \\ 0 & 6 & 1 & 7 & 4 & 2 & 5 & 3 \\ 1 & 7 & 2 & 0 & 5 & 3 & 6 & 4 \end{pmatrix}.$$

Generalizing the above idea, we can find a mapping that removes some of the intermediate extremal configurations that are not necessarily located on adjacent levels. For instance, suppose that we would like to remove levels from $(r - j_2)$ to $(r - i_2)$ and from $(r - j_1)$ to $(r - i_1)$, $1 \le i_1 \le j_1$, $i_2 \le j_2 \le r-1$, and $j_1+1 < i_2$. In this case, random variables $(Z_{i_1,n}, \ldots, Z_{j_1+1,n})$ converge to the same random variable, and so do random variables $(Z_{i_2,n}, \ldots, Z_{j_2+1,n})$. Thus, we need a permutation π_{q_1} with $q_1 = 2^{j_1-i_1+2}$ for removing $(j_1 - i_1+1)$ consecutive levels starting with the $(r-j_1)$th one, and a permutation π_{q_2} with $q_2 = 2^{j_2-i_2+2}$ that removes levels from $(r - j_2)$ to $(r - i_2)$. The permutation π is given by:

$$\pi(u_1) = \begin{cases} q_1^{(j)} \cdot \pi_{q_1}\big(\frac{u_1-u_1^{(j)}}{q_1^{(j)}}\big) + \pi^{(2)}(u_1 - u_1^{(j)}), & \text{if } u < q_1^{(i)} \\ \pi(u_1^{(i)}) + (u_1 - u_1^{(i)}), & \text{if } q_1^{(i)} \le u_1 \le q_1^{(j)} \end{cases}$$

where $q_1 = 2^{j_1-i_1+2}$, $q_1^{(i)} = 2^{r-i_1+1}$, $q_1^{(j)} = 2^{r-j_1-1}$, $u_1^{(j)} = u_1 \mod q_1^{(j)}$, and $u_1^{(i)} = u_1 \mod q_1^{(i)}$. Further, $\pi^{(2)}$ is a permutation from $[q_1^{(j)}]$ to $[q_1^{(j)}]$ with

$$\pi^{(2)}(u_2) = \begin{cases} q_2^{(j)} \cdot \pi_{q_2}\big(\frac{u_2-u_2^{(j)}}{q_2^{(j)}}\big) + (u_2 - u_2^{(j)}), & \text{if } u_2 < q_2^{(i)} \\ \pi(u_2^{(i)}) + (u_2 - u_2^{(i)}), & \text{if } q_2^{(i)} \le u_2 < q_2^{(j)} \end{cases}$$

where $q_2 = 2^{j_2-i_2+2}$, $q_2^{(i)} = 2^{r-i_2+1}$, $q_2^{(j)} = 2^{r-j_2-1}$, $u_2^{(j)} = u_2 \mod q_2^{(j)}$, and $u_2^{(i)} = u_2 \mod q_2^{(i)}$. Finally, π_{q_1} and π_{q_2} are permutations that satisfy condition (3) with q replaced with q_1 and q_2, respectively.

Let us consider the following example.

Example 2: Let $q = 16$ and suppose that we would like to remove the 1st and the 3rd levels from the set of 5 extremal configurations (1). Construct the following mapping: $g(u_1, u_2) = u_1 + \pi(u_2)$. where π is given above with $q = 16$, $i_1 = j_1 = 1$, and $i_2 = j_2 = 3$. The first row of the Cayley table G is [10 8 11 9 2 0 3 1 14 12 15 13 6 4 7 5].

In its most general form, the above construction results in polarization kernels that remove from $\{1, 2, \ldots, r\}$ any given set of disjoint segments of the form $(i_m, i_m + 1, \ldots, j_m)$, where $m \ge 1$ is some positive number, thereby reducing the set of extremal configurations (1) to any given subset. Space constraints do not permit us to include the description of this construction in the present paper (see [2, Theorem 4.8]).

IV. Concluding remarks: Video coding

We constructed polarization kernels that yield q-ary codes, $q = 2^r$, that polarize to any given subset of the set of $r + 1$ extremal configurations, resulting in symbols that carry precisely the desired number of perfect bits over the channel W. We note that in certain contexts, transmission can benefit from relying on symbols that carry different amounts of information over the channel. One such application relates to video coding. For instance, the video stream in MPEG-2 consists of six layers that encode information of different types [6]. In particular, the picture layer begins with an informational frame (I-frame) followed by predictive frames (P-frames) and bidirectional predictive frames (B-frames). An I-frame contains the reference picture and does not require additional information to reconstruct the image. Because of this, data in the I-frames has to be encoded for high reliability, and thus its compression ratio is low. At the same time, both P- and B-frames carry data only for blocks that are changed from one reference picture to the next one. The picture is decoded based on the prior I-frame as well as the P-frame. To reconstruct the image for P-frames, we start with the full image of the neighboring frame and obtain a new image by applying the new motion information. Although this information may be decoded with errors, we can still obtain good quality of the picture as long as the distortion level of the recovered image is low or if the user does not notice the distortion caused by errors. Therefore, usually data in P- and B-frames require less reliability than those in the I-frames, and so this portion of the data can be compressed at a higher rate. This simplified description (which becomes even more complicated in MPEG-4) already shows the benefit of data symbols of different types in the encoding because they enable one to avoid additional overhead in the encoder and the decoder that would arise from using two-level polar codes (binary or q-ary).

References

[1] E. Arıkan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[2] W. Park, "Applications of ordered weights in information transmission," Ph.D. dissertation, University of Maryland, November 2012. Online: drum.lib.umd.edu/handle/1903/1354.

[3] W. Park and A. Barg, "Polar codes for q-ary channels, $q = 2^r$," *IEEE Trans. Inform. Theory*, vol. 59, no. 2, pp. 955-969, 2013.

[4] A. G. Sahebi and S. S. Pradhan, "Multilevel polarization of polar codes over arbitrary discrete memoryless channels," arXiv:1107.1535.

[5] E. Şaşoğlu, "Polar coding theorems for discrete systems," Ph.D. dissertation, École polytechnique fédérale de Lausanne, November 2011.

[6] D. Le Gall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM*, vol. 34, no. 4, pp. 47–58, 1991.