

A Class of Transformations that Polarize Symmetric Binary-Input Memoryless Channels

Satish Babu Korada, Eren Şaşoğlu

October 16, 2008

Abstract

A generalization of Arikan's polar code construction using transformations of the form $G^{\otimes n}$ where G is an $\ell \times \ell$ matrix is considered. It is shown that a large class of such transformations polarize symmetric binary-input memoryless channels. Necessary and sufficient conditions are given for these transformations to ensure channel polarization.

1 Introduction

Polar codes, introduced by Arikan in [1], are the first provably capacity achieving codes for any symmetric binary-input discrete memoryless channel (B-DMC) with low encoding and decoding complexity. Polar code construction is based on the following observation: Let

$$G = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \quad (1)$$

Consider applying the transform $G^{\otimes n}$ (where “ $\otimes n$ ” denotes the n th Kronecker power) to a block of $N = 2^n$ bits and transmitting the output through independent copies of a B-DMC W (see Figure 1). As n grows large, the channels seen by individual bits (suitably defined in [1]) start *polarizing*: they approach either a noiseless channel or a pure-noise channel, where the fraction of channels becoming noiseless is close to the symmetric mutual information $I(W)$.

It was conjectured in [1] that polarization is a general phenomenon, and is not restricted to the particular transformation $G^{\otimes n}$ defined by (1). In this note we give a partial affirmation to this conjecture. In particular, we consider transformations of the form $G^{\otimes n}$ where G is an $\ell \times \ell$ matrix for $\ell \geq 3$ and provide necessary and sufficient conditions for such G 's to polarize symmetric B-DMCs.

2 Preliminaries

Let $W : \{0, 1\} \rightarrow \mathcal{Y}$ be a B-DMC. Let $I(W) \in [0, 1]$ denote the mutual information between the input and output of W with uniform distribution on the inputs. Also let $Z(W) \in [0, 1]$ denote the Bhattacharyya parameter of W , i.e., $Z(W) = \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)}$.

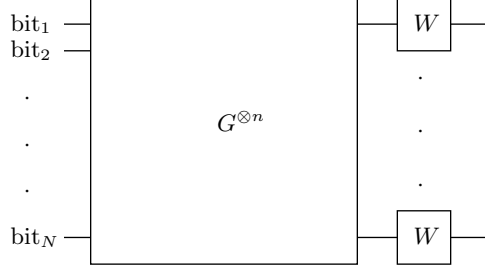


Figure 1:

Consider a random ℓ -vector U_1^ℓ that is uniformly distributed over $\{0, 1\}^\ell$. Let $X_1^\ell = U_1^\ell G$, where G is an $\ell \times \ell$ invertible $\{0, 1\}$ matrix and the multiplication is performed over $\text{GF}(2)$. Let u_1^ℓ and x_1^ℓ denote realizations of U_1^ℓ and X_1^ℓ respectively. Define $W^{(i)} : \{0, 1\} \rightarrow \{0, 1\}^{i-1} \times \mathcal{Y}^\ell$ as the channel with input u_i , output (y_1^ℓ, u_1^{i-1}) and transition probabilities given by

$$W^{(i)}(y_1^\ell, u_1^{i-1} | u_i) = \frac{1}{2^{\ell-1}} \sum_{u_{i+1}^\ell} W_\ell(y_1^\ell | u_1^\ell),$$

where

$$W_\ell(y_1^\ell | u_1^\ell) \triangleq \prod_{i=1}^{\ell} W(y_i | x_i) = \prod_{i=1}^{\ell} W(y_i | (u_1^\ell G)_i).$$

For $k \geq 1$ let W^k denote the product B-DMC with transition probabilities

$$W^k(y_1^k | x) = \prod_{j=1}^k W(y_j | x).$$

Also let $Z^{(i)}$ denote the Bhattacharyya parameter of $W^{(i)}$, i.e.,

$$Z^{(i)} = \sum_{y_1^\ell, u_1^{i-1}} \sqrt{W^{(i)}(y_1^\ell, u_1^{i-1} | 0) W^{(i)}(y_1^\ell, u_1^{i-1} | 1)}.$$

Finally, let $I^{(i)}$ denote the mutual information between the input and output of channel $W^{(i)}$. Since G is invertible, it is easy to check that

$$\sum_{i=1}^{\ell} I^{(i)} = \ell I(W).$$

3 Polarization

We will say that G is a *polarizing* matrix if

$$\exists i \text{ such that } W^{(i)} = W^k \text{ for some } k \geq 2. \quad (2)$$

It will be shown that channel transformations of the form $G^{\otimes n}$ polarize symmetric channels if and only if G is polarizing. This statement is made precise in the following theorem:

Theorem 1. *Let $G^{\otimes n}$ denote the n^{th} Kronecker power of G and consider the transformation $G^{\otimes n} : W \rightarrow (W^{(i)} : i = 1, \dots, \ell^n)$.*

i. If G is polarizing, then for any $\delta > 0$

$$\lim_{n \rightarrow \infty} \frac{\#\{i \in \{1, \dots, \ell^n\} : I(W^{(i)}) \in (\delta, 1 - \delta)\}}{\ell^n} = 0.$$

ii. If G is not polarizing, then

$$I(W^{(i)}) = I(W) \text{ for all } n \text{ and } i \in \{1, \dots, \ell^n\}.$$

Theorem 1 is a direct consequence of Lemmas 1 and 2 below.

Note that any invertible $\{0, 1\}$ matrix G can be written as a (real) sum $G = P + P'$, where P is a permutation matrix, and P' is a $\{0, 1\}$ matrix. This fact can be inferred from Hall's Theorem [3, Theorem 16.4.]. Therefore, for any such matrix G , there exists a column permutation that results in $G_{ii} = 1$ for all i . Since the transition probabilities defining $W^{(i)}$ are invariant (up to a permutation of the outputs y_1^ℓ) under column permutations on G , we only consider matrices with 1's on the diagonal.

The following lemma gives necessary and sufficient conditions for (2) to be satisfied:

Lemma 1. *For any B-DMC W ,*

i. If G is not upper triangular, then there exists an i for which $W^{(i)} = W^k$ for some $k \geq 2$.

ii. If G is upper triangular, then $W^{(i)} = W$ for all $1 \leq i \leq \ell$.

Proof. Let $G^{(\ell-i)}$ be the $(\ell - i) \times (\ell - i)$ matrix obtained from G by removing its last i rows and columns. Let the number of 1's in the last row of G be k . Clearly $W^{(\ell)} = W^k$. If $k \geq 2$ then G is not upper triangular and the first claim of the lemma holds. If $k = 1$, then $W^{(\ell)} = W$, and $(x_1, \dots, x_{\ell-1})$ is independent of u_ℓ . One can then write

$$\begin{aligned} W^{(\ell-i)}(u_1^{\ell-i-1}, y_1^\ell | u_{\ell-i}) &= \frac{1}{2^{\ell-1}} \sum_{u_{\ell-i+1}^\ell} W_\ell(y_1^\ell | u_1^\ell) \\ &= \frac{1}{2^{\ell-1}} \sum_{u_{\ell-i+1}^{\ell-1}, u_\ell} \Pr[Y_1^{\ell-1} = y_1^{\ell-1} | U_1^\ell = u_1^\ell] \Pr[Y_\ell = y_\ell | Y_1^{\ell-1} = y_1^{\ell-1}, U_1^\ell = u_1^\ell] \\ &\stackrel{(a)}{=} \frac{1}{2^{\ell-1}} \sum_{u_{\ell-i+1}^{\ell-1}, u_\ell} W_{\ell-1}(y_1^{\ell-1} | u_1^{\ell-1}) \Pr[Y_\ell = y_\ell | Y_1^{\ell-1} = y_1^{\ell-1}, U_1^\ell = u_1^\ell] \\ &= \frac{1}{2^{\ell-1}} \sum_{u_{\ell-i+1}^{\ell-1}} W_{\ell-1}(y_1^{\ell-1} | u_1^{\ell-1}) \sum_{u_\ell} \Pr[Y_\ell = y_\ell | Y_1^{\ell-1} = y_1^{\ell-1}, U_1^\ell = u_1^\ell] \\ &= \frac{1}{2^{\ell-1}} [W(y_\ell | 0) + W(y_\ell | 1)] \sum_{u_{\ell-i+1}^{\ell-1}} W_{\ell-1}(y_1^{\ell-1} | u_1^{\ell-1}) \end{aligned}$$

where (a) follows from the fact that $G_{lk} = 0$, for all $k < \ell$. Therefore y_ℓ is independent of the inputs to the channels $W^{(\ell-i)}$ for $i = 1, \dots, \ell - 1$. This is equivalent to saying that channels $W^{(1)}, \dots, W^{(\ell-1)}$ are defined by the matrix $G^{(\ell-1)}$. Applying the same argument to $G^{(\ell-1)}$ and repeating, we see that if G is upper triangular, then we have $W^{(i)} = W$ for all i . On the other hand, if G is not upper triangular, then there either exists an i for which $G^{(\ell-i)}$ has at least two 1s in the last row, which in turn implies $W^{(i)} = W^k$ for some $k \geq 2$. \square

Remark 1. *The above lemma says that all transformations that are not upper triangular are polarizing. Moreover, upper triangular transformations have no effect on the channel, i.e., each bit sees an independent copy of W after an upper triangular transformation.*

Corollary 1. *For any polarizing transformation G , there exists an $i \in \{1, \dots, \ell\}$ and $k \geq 2$ for which*

$$I^{(i)} = I(W^k) \tag{3}$$

$$Z^{(i)} = Z(W)^k. \tag{4}$$

Proof. The first claim is trivial. The second claim follows from the fact that the Bhattacharyya parameter of any product channel $\prod_j W_j$ is given by $\prod_j Z(W_j)$. \square

4 Convergence

Consider the recursive channel combination given in [1], using a polarizing transformation G . Following Arkan, associate to this construction a tree process $\{W_n; n \geq 0\}$ with

$$\begin{aligned} W_0 &= W \\ W_{n+1} &= W_n^{(B_{n+1})}, \end{aligned}$$

where $\{B_n; n \geq 1\}$ is a sequence of i.i.d. random variables defined on a probability space $(\Omega, \mathcal{F}, \mu)$, B_n being uniformly distributed over the set $\{1, \dots, \ell\}$. Define $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n = \sigma(B_1, \dots, B_n)$ for $n \geq 1$. Define the processes $\{I_n; n \geq 0\} = \{I(W_n); n \geq 0\}$ and $\{Z_n; n \geq 0\} = \{Z(W_n); n \geq 0\}$.

Observation 1. *$\{(I_n, \mathcal{F}_n)\}$ is a bounded martingale and therefore converges a.s. and in \mathcal{L}^1 to a random variable I_∞ .*

Lemma 2. *If W is symmetric and G is polarizing, then*

$$I_\infty = \begin{cases} 1 & \text{w.p. } I(W), \\ 0 & \text{w.p. } 1 - I(W). \end{cases}$$

Proof. By the convergence in \mathcal{L}^1 of I_n we have $\mathbb{E}[|I_{n+1} - I_n|] \xrightarrow{n \rightarrow \infty} 0$. Since G is a polarizing matrix, Lemma 1 implies

$$I_{n+1} = I(W_n^k) \text{ with probability at least } \frac{1}{\ell},$$

for some $k \geq 2$. This in turn implies

$$\mathbb{E}[|I_{n+1} - I_n|] \geq \frac{1}{\ell} \mathbb{E}[(I(W_n^k) - I(W_n))] \rightarrow 0. \quad (5)$$

It is shown in the Appendix that for any symmetric B-DMC W_n , if $I(W_n) \in (\delta, 1 - \delta)$ for some $\delta > 0$, then there exists an $\eta(\delta) > 0$ such that $I(W_n^k) - I(W_n) > \eta(\delta)$. We therefore conclude that convergence in (5) implies $I_\infty \in \{0, 1\}$ w.p. 1. The claim on the probability distribution of I_∞ follows from the fact that $\{I_n\}$ is a martingale, i.e., $\mathbb{E}[I_\infty] = \mathbb{E}[I_0] = I(W)$. \square

Corollary 2. *If W is symmetric and G is polarizing, then $\{Z_n\}$ converges a.s. to a random variable Z_∞ and*

$$Z_\infty = \begin{cases} 0 & \text{w.p. } I(W), \\ 1 & \text{w.p. } 1 - I(W). \end{cases}$$

Proof. The proof follows from the fact that $I_n \rightarrow I_\infty$ a.s. and the inequalities [1]

$$\begin{aligned} I(Q)^2 + Z(Q)^2 &\leq 1 \\ I(Q) + Z(Q) &\geq 1. \end{aligned}$$

for any B-DMC Q . \square

Theorem 2. *Given a symmetric B-DMC W , an $\ell \times \ell$ polarizing matrix G , and any $\beta < 1/\ell$,*

$$\lim_{n \rightarrow \infty} \Pr[Z_n \leq 2^{-2^{n\beta}}] = I(W).$$

Proof Idea. For any polarizing matrix it can be shown that $Z_{n+1} \leq \ell Z_n$ with probability 1 and that $Z_{n+1} \leq Z_n^2$ with probability at least $1/\ell$. The proof then follows by adapting the proof of [2, Theorem 3]. \square

5 Discussion

Using Arıkan's rule for choosing the information set, polar codes of blocklength $N = \ell^n$ can be constructed starting with any polarizing $\ell \times \ell$ matrix G . The encoding and successive cancellation decoding complexities of such codes are $O(N \log N)$. Using similar arguments, it is easy to show that polar codes of blocklength $N = \prod_{i=1}^n \ell_i$ can be constructed from generator matrices of the form $\otimes_i G_i$, where each G_i is a polarizing matrix of size $\ell_i \times \ell_i$. The encoding and successive cancellation decoding complexities of these codes are also $O(N \log N)$.

Appendix

In this section we prove the following:

Lemma 3. *Let W be a symmetric B-DMC and let W^k denote the product channel*

$$W^k(y_1^k | x) = \prod_{i=1}^k W(y_i | x).$$

If $I(W) \in (\delta, 1 - \delta)$ for some $\delta > 0$, then there exists an $\eta(\delta) > 0$ such that $I(W^k) - I(W) > \eta(\delta)$.

We will use the following theorem in proving Lemma 3:

Theorem 3 ([4]). *Let W_1, \dots, W_k be k symmetric B-DMCs with capacities I_1, \dots, I_k respectively. Let $W^{(k)}$ denote the channel with transition probabilities*

$$W^{(k)}(y_1^k | x) = \prod_{i=1}^k W_i(y_i | x).$$

Also let $W_{BSC}^{(k)}$ denote the channel with transition probabilities

$$W_{BSC}^{(k)}(y_1^k | x) = \prod_{i=1}^k W_{BSC(\epsilon_i)}(y_i | x),$$

where $BSC(\epsilon_i)$ denotes the BSC with crossover probability $\epsilon_i \in [0, \frac{1}{2}]$, $\epsilon_i \triangleq h^{-1}(1 - I_i)$, where h denotes the binary entropy function. Then, $I(W^{(k)}) \geq I(W_{BSC}^{(k)})$.

Remark 2. *Consider the transmission of a single bit X using k independent symmetric B-DMCs W_1, \dots, W_k with capacities I_1, \dots, I_k . Theorem 3 states that over the class of all symmetric channels with given mutual informations, the mutual information between the input and the output vector is minimized when each of the individual channels is a BSC.*

Proof of Lemma 3. Let $\epsilon \in [0, \frac{1}{2}]$ be the crossover probability of a BSC with capacity $I(W)$, i.e., $\epsilon = h^{-1}(1 - I(W))$. Note that for $k \geq 2$,

$$I(W^k) \geq I(W^2) \geq I(W).$$

By Theorem 3, we have $I(W^2) \geq I(W_{BSC(\epsilon)}^2)$. A simple computation shows that

$$I(W_{BSC(\epsilon)}^2) = 1 + h(2\epsilon\bar{\epsilon}) - 2h(\epsilon).$$

We can then write

$$\begin{aligned} I(W^k) - I(W) &\geq I(W_{BSC(\epsilon)}^2) - I(W) \\ &= I(W_{BSC(\epsilon)}^2) - I(W_{BSC(\epsilon)}) \\ &= h(2\epsilon\bar{\epsilon}) - h(\epsilon) \\ &> \eta(\delta) \end{aligned} \tag{6}$$

for some $\eta(\delta) > 0$. In the above, (6) follows from the fact that $I(W) \in (\delta, 1 - \delta)$ implies $\epsilon \in (\phi(\delta), \frac{1}{2} - \phi(\delta))$ where $\phi(\delta) > 0$. \square

References

- [1] E. Arıkan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” submitted to *IEEE Trans. Inform. Theory*. Available at: <http://arxiv.org/pdf/0807.3917>.
- [2] E. Arıkan and E. Telatar, “On the rate of channel polarization”. Available at <http://arxiv.org/pdf/0807.3806>.
- [3] J.A. Bondy and U.S.R. Murty, “Graph Theory,” Springer, 2008.
- [4] I. Sutskever, S. Shamai and J. Ziv, “Extremes of information combining,” *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1313–1325, Apr. 2005.