Polarization improves E_0

Mine Alsan and Emre Telatar
Information Theory Laboratory
Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland
Email: {mine.alsan,emre.telatar}@epfl.ch

***Abstract*—We prove that channel combining and splitting via Arıkan's polarization transformation improves Gallager's reliability function E_0 for binary input channels. In this sense polarization 'creates' E_0. This observation gives yet another justification as to why the polar transform yields capacity achieving and low complexity codes: the improvement in E_0 translates to an improvement in complexity–error-probability trade-off. In analyzing polar codes, one examines auxiliary random processes that follow the evolution of information measures as an underlying communication channel undergoes a sequence of transformations. The conclusion of this paper shows that the E_0 process associated to such an analysis is a submartingale.**

***Index Terms*—Channel polarization, reliability function, reliability-complexity trade-off, Rényi's entropies**

I. Introduction

Arıkan's polar codes [1] are constructed by the repeated application of the polar transform. From two independent copies of a given binary input channel $W : \mathbb{F}_2 \to \mathcal{Y}$, this transform synthesizes two new binary input channels $W^- : \mathbb{F}_2 \to \mathcal{Y}^2$ and $W^+ : \mathbb{F}_2 \to \mathcal{Y}^2 \times \mathbb{F}_2$, with transition probabilities given by

$$\begin{aligned} W^-(y_1y_2|u_1) &= \sum_{u_2\in\mathbb{F}_2} \tfrac{1}{2} W(y_1|u_1\oplus u_2)W(y_2|u_2) \\ W^+(y_1y_2u_1|u_2) &= \tfrac{1}{2} W(y_1|u_1\oplus u_2)W(y_2|u_2). \end{aligned}$$

With $I(W)$ denoting the mutual information between the input and the output of a binary input channel W when the input is uniformly distributed, and $Z(W)$ denoting the Bhattacharyya parameter, it is observed in [1] that

$$I(W^-) + I(W^+) = 2I(W) \tag{1}$$
$$Z(W^-) + Z(W^+) \le 2Z(W) \tag{2}$$

that is, polar transform conserves the symmetric mutual information and improves the Bhattacharyya parameter. In a paper that predates polar coding [2] Arıkan had already shown that this transform improves the symmetric cutoff rate $R_0(W)$, that is[1],

$$R_0(W^-) + R_0(W^+) \ge 2R_0(W). \tag{3}$$

Since both the symmetric mutual information and the symmetric cutoff rate are quantities that can be obtained from Gallager's reliability function E_0, namely,

$$R_0(W) = E_0(1,W) \quad \text{and} \quad I(W) = \lim_{\rho\to 0} E_0(\rho,W)/\rho,$$

a natural question to ask is if the polar transform improves E_0, which would make (3) a special case (and also show that the left hand side of (1) is at least as large as the right hand side). In this paper we will show:

Theorem 1: For any binary input channel W and any $\rho \ge 0$,

$$E_0(\rho, W^-) + E_0(\rho, W^+) \ge 2E_0(\rho, W)$$

where $E_0(\rho, W)$ denotes "Gallager's E_0" [3, p. 138] evaluated for the uniform input distribution

$$E_0(\rho, W) = -\log \sum_{y\in\mathcal{Y}} \Big[\sum_{x\in\mathbb{F}_2} \tfrac{1}{2} W(y|x)^{\frac{1}{1+\rho}} \Big]^{1+\rho}. \tag{4}$$

The inequality in Theorem 1 holds with equality if and only if the channel W is perfect, or the channel W is completely noisy, or $\rho = 0$.

Theorem 1 will be obtained as a corollary to a slightly more general result pertaining to a more general polar transform that synthesizes two channels from two independent (but not necessarily identical) binary input channels $W_1 : \mathbb{F}_2 \to \mathcal{Y}_1$ and $W_2 : \mathbb{F}_2 \to \mathcal{Y}_2$. Given two such channels, define $W_{1,2}^- : \mathbb{F}_2 \to \mathcal{Y}_1 \times \mathcal{Y}_2$ and $W_{1,2}^+ : \mathbb{F}_2 \to \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathbb{F}_2$ by

$$W_{1,2}^-(y_1y_2|u_1) = \sum_{u_2\in\mathbb{F}_2} \tfrac{1}{2} W_1(y_1|u_1\oplus u_2)W_2(y_2|u_2), \tag{5}$$
$$W_{1,2}^+(y_1y_2u_1|u_2) = \tfrac{1}{2} W_1(y_1|u_1\oplus u_2)W_2(y_2|u_2). \tag{6}$$

In the next section we will prove

Theorem 2: For any two binary input channels W_1 and W_2 and any $\rho \ge 0$,

$$E_0(\rho, W_{1,2}^-) + E_0(\rho, W_{1,2}^+) \ge E_0(\rho, W_1) + E_0(\rho, W_2). \tag{7}$$

Theorem 1 trivially follows from Theorem 2 by setting $W_1 = W_2 = W$.

It may be of interest to note that in general the channel $W_{1,2}^-$ is not the same channel as $W_{2,1}^-$, so the order of W_1 and W_2 does matter. However the two channels have the same E_0, (and consequently the same symmetric cutoff rate, same symmetric mutual information, same Bhattacharyya parameter, etc.) The same remarks also apply to $W_{1,2}^+$ and $W_{2,1}^+$.

The apparent 'creation' of E_0 by the polar transform does not violate any 'conservation' theorem. While mutual

[1] This conclusion may be derived as a consequence of (2), via the relationship $R_0 = -\log\big((1+Z)/2\big)$, and the concavity and monotonicity of log.

information cannot be improved by processing of the input or output of the channel, E_0 is not a conserved quantity and may be created out of thin air by processing. Indeed, any good coding method implicitly relies on the possibility to create E_0 by processing.

A further result concerning the E_0 parameter of the synthesized channels is reported in [4]. It is shown therein that the binary erasure channel (BEC) and the binary symmetric channel (BSC) are extremal in the evolution of the E_0 parameter under the basic polarization transformations. In particular, for a given value of $E_0(\rho, W)$, while the BEC minimizes, and the BSC maximizes the value of $E_0(\rho, W^-)$ for any $\rho \geq 0$, and the value of $E_0(\rho, W^+)$ for any $\rho \in [1, 2]$; the minimizing and maximizing roles are reversed for the value of $E_0(\rho, W^+)$ when $\rho \in [0, 1] \cup [2, \infty)$.

The next section is devoted to proving Theorem 2. The subsequent section discusses some implications of the theorem on the chain rule for Rényi's entropies [5], and improving the reliability-complexity trade-off of random block codes by Arıkan's method of channel combining and splitting [2].

II. Results

Before proving Theorem 2, we first note that the quantity

$$\sum_y \Big[\tfrac{1}{2} W(y|0)^{\frac{1}{1+\rho}} + \tfrac{1}{2} W(y|1)^{\frac{1}{1+\rho}}\Big]^{1+\rho}$$

that appears in the definition of E_0 in (4) can be rewritten as

$$\sum_y W(y) \Big[\tfrac{1}{2}\big(1+\Delta(y)\big)^{\frac{1}{1+\rho}} + \tfrac{1}{2}\big(1-\Delta(y)\big)^{\frac{1}{1+\rho}}\Big]^{1+\rho}$$

with $W(y) := \frac{1}{2}W(y|0) + \frac{1}{2}W(y|1)$ and $\Delta(y) := \big[W(y|0) - W(y|1)\big]/\big[W(y|0)+W(y|1)\big]$. Observing that this last expression is in the form of an expectation, and that $|\Delta(y)| \leq 1$, we see that $E_0(\rho, W)$ can be expressed as[2]

$$E_0(\rho, W) = -\log \mathbb{E}\big[g(\rho, Z)\big],$$

where

$$g(\rho, z) = \Big(\tfrac{1}{2}(1+z)^{\frac{1}{1+\rho}} + \tfrac{1}{2}(1-z)^{\frac{1}{1+\rho}}\Big)^{1+\rho}, \qquad (8)$$

and Z a random variable taking values in the interval $[0, 1]$.

Proof of Theorem 2: By the observation just above and Lemmas 2 and 3 proved in [4], we know that

$$\begin{aligned} E_0(\rho, W_1) &= -\log \mathbb{E}\big[g(\rho, Z_1)\big], \\ E_0(\rho, W_2) &= -\log \mathbb{E}\big[g(\rho, Z_2)\big], \\ E_0(\rho, W_{1,2}^-) &= -\log \mathbb{E}\big[g(\rho, Z_1 Z_2)\big], \\ E_0(\rho, W_{1,2}^+) &= -\log \mathbb{E}\big[h(\rho, Z_1, Z_2)\big], \end{aligned}$$

where Z_1, Z_2 are *independent* random variables taking values in the interval $[0, 1]$, g is as in (8), and

$$\begin{aligned} h(\rho, z_1, z_2) = \tfrac{1}{2}(1+z_1 z_2) g\Big(\rho, \frac{z_1+z_2}{1+z_1 z_2}\Big) \\ + \tfrac{1}{2}(1-z_1 z_2) g\Big(\rho, \frac{z_1-z_2}{1-z_1 z_2}\Big). \qquad (9) \end{aligned}$$

[2]This representation was observed in 2008 in an unpublished manuscript by Arıkan and Telatar.

By these identities, showing (7) is equivalent to showing

$$\mathbb{E}\big[g(\rho, Z_1)\big]\mathbb{E}\big[g(\rho, Z_2)\big] \geq \mathbb{E}\big[g(\rho, Z_1 Z_2)\big]\mathbb{E}\big[h(\rho, Z_1, Z_2)\big].$$

The proof is carried in two steps. We first claim that the following inequality is satisfied:

$$g(\rho, z_1) g(\rho, z_2) \geq g(\rho, z_1 z_2) h(\rho, z_1, z_2) \qquad (10)$$

for any $z_1, z_2 \in [0, 1]$ and $\rho \geq 0$.

From (10) and noting the independence of Z_1 and Z_2 we see that

$$\begin{aligned} \mathbb{E}\big[g(\rho, Z_1)\big]\mathbb{E}\big[g(\rho, Z_2)\big] &= \mathbb{E}\big[g(\rho, Z_1) g(\rho, Z_2)\big] \\ &\geq \mathbb{E}\big[g(\rho, Z_1 Z_2) h(\rho, Z_1, Z_2)\big]. \end{aligned}$$

By Lemma 2 in the Appendix, the function $g(\rho, z_1 z_2)$ is non-increasing in z_1 and z_2 separately for any $\rho \geq 0$. Similarly, by Lemma 3 in the Appendix the function $h(\rho, z_1, z_2)$ is also non-increasing in z_1 and z_2 separately for any $\rho \geq 0$. These monotonicity properties are useful, as they imply, via Lemma 4 in the Appendix, that the random variables $g(\rho, Z_1 Z_2)$ and $h(\rho, Z_1, Z_2)$ are positively correlated. As a result

$$\begin{aligned} \mathbb{E}\big[g(\rho, Z_1)\big]\mathbb{E}\big[g(\rho, Z_2)\big] &\geq \mathbb{E}\big[g(\rho, Z_1 Z_2) h(\rho, Z_1, Z_2)\big] \\ &\geq \mathbb{E}\big[g(\rho, Z_1 Z_2)\big]\mathbb{E}\big[h(\rho, Z_1, Z_2)\big], \end{aligned}$$

concluding the proof of the relation in (7).

Now, we prove the inequality claimed in (10). For that purpose, we first apply the change of variables

$$\begin{aligned} t &= \operatorname{arctanh} z_1, \quad w = \operatorname{arctanh} z_2, \\ k &= \operatorname{arctanh}(z_1 z_2), \quad s = \frac{1}{1+\rho}, \end{aligned}$$

where $s \in (0, 1]$ and $t, w, k \in [0, \infty)$. Using these, we obtain

$$g(\rho, z_1) = g\Big(\frac{1-s}{s}, \tanh(t)\Big) = \frac{\cosh(st)^{\frac{1}{s}}}{\cosh(t)}, \qquad (11)$$

$$g(\rho, z_2) = g\Big(\frac{1-s}{s}, \tanh(w)\Big) = \frac{\cosh(sw)^{\frac{1}{s}}}{\cosh(w)}, \qquad (12)$$

$$g(\rho, z_1 z_2) = g\Big(\frac{1-s}{s}, \tanh(k)\Big) = \frac{\cosh(sk)^{\frac{1}{s}}}{\cosh(k)}, \qquad (13)$$

and

$$\begin{aligned} h(\rho, z_1, z_2) &= h\Big(\frac{1-s}{s}, \tanh(t), \tanh(w)\Big) \\ &= \frac{\cosh(s(t+w))^{\frac{1}{s}} + \cosh(s(t-w))^{\frac{1}{s}}}{2\cosh(t)\cosh(w)}. \qquad (14) \end{aligned}$$

We further define

$$a = t + w, \quad b = t - w$$

so that $t = (a+b)/2$, $w = (a-b)/2$, and $a \geq |b|$. Then, the variable k is given by

$$k = \frac{1}{2}\log\left(\frac{\cosh(a)}{\cosh(b)}\right), \qquad (15)$$

and the expression in (13) becomes

$$g(\rho, z_1 z_2) = \frac{\left(\dfrac{\cosh(a)^s + \cosh(b)^s}{2}\right)^{\frac{1}{s}}}{\dfrac{\cosh(a) + \cosh(b)}{2}}. \tag{16}$$

With (11) and (12) at hand, a bit of algebra reveals that the left hand side of (10) is given by

$$\frac{\left(\frac{\cosh(sa)+\cosh(sb)}{2}\right)^{\frac{1}{s}}}{\cosh(t)\cosh(w)}. \tag{17}$$

Similarly, using equations (14) and (16), the right hand side of (10) is given by

$$\frac{\left(\frac{\cosh(a)^s+\cosh(b)^s}{2}\right)^{\frac{1}{s}}}{\frac{\cosh(a)+\cosh(b)}{2}} \times \frac{\cosh(sa)^{\frac{1}{s}} + \cosh(sb)^{\frac{1}{s}}}{2\cosh(t)\cosh(w)}. \tag{18}$$

Therefore, we see that the inequality (10) is equivalent to

$$\frac{\left(1 + \left(\frac{\cosh(sb)}{\cosh(sa)}\right)\right)^{\frac{1}{s}}}{1 + \frac{\cosh(sb)^{\frac{1}{s}}}{\cosh(sa)^{\frac{1}{s}}}} \geq \frac{\left(1 + \left(\frac{\cosh(b)}{\cosh(a)}\right)^s\right)^{\frac{1}{s}}}{1 + \frac{\cosh(b)}{\cosh(a)}}. \tag{19}$$

Let $u = \left(\frac{\cosh(sb)}{\cosh(sa)}\right)^{\frac{1}{s}}$ and $v = \frac{\cosh(b)}{\cosh(a)}$. Then, by Lemma 2 in the Appendix, whenever $a \geq |b|$, we have $1 \geq u \geq v \geq 0$ since

$$f_s(b) = \frac{\cosh(s|b|)^{\frac{1}{s}}}{\cosh(|b|)} \geq \frac{\cosh(sa)^{\frac{1}{s}}}{\cosh(a)} = f_s(a).$$

As a result, we have reduced the inequality (10) to the following form:

$$F_s(u) \geq F_s(v) \quad \text{when} \quad 1 \geq u \geq v \geq 0$$

where

$$F_s(u) = \frac{(1+u^s)^{1/s}}{1+u}.$$

But, we know this is true by Lemma 1 in the Appendix. Hence, inequality (10) holds as claimed. ■

III. Discussion

A. Submartingale Property of E_0

Applying the polar transform to the channel W one obtains the channels W^- and W^+. If one applies the polar transform to these new channels one would obtain the channels $W^{--} := (W^-)^-$, $W^{-+} := (W^-)^+$ and $W^{+-} := (W^+)^-$, $W^{++} := (W^+)^+$. Repeated application, will yield at stage n, a set of 2^n channels

$$\{W^s : s \in \{+,-\}^n\}.$$

In analyzing the properties of this channel it is useful to introduce an auxiliary stochastic process, $W_0, W_1, \ldots$, defined by $W_0 := W$, and for $n \geq 0$

$$W_{n+1} := \begin{cases} W_n^- & \text{with probability 1/2} \\ W_n^+ & \text{with probability 1/2} \end{cases}$$

with the successive choices taken independently. In this way, W_n is uniformly distributed over the set of 2^n channels above. Theorem 1 is equivalent to the statement that the stochastic process $\{E_0(\rho, W_n) : n \geq 0\}$ is a submartingale.

B. Improving the Reliability-Complexity Trade-off

Beside its usefulness as an argument in channel polarization related proofs, an interesting interpretation of the inequality in Theorem 2 is given in [2]. Arıkan introduces the concept of the reliability-complexity exponent under maximum likelihood decoding of a code drawn from a random code ensemble. He formalizes the problem as a trade-off, and suggests the general method of channel combining and splitting can be used to improve this trade-off. Here we explore this idea.

For a given rate R and B-DMC W, consider the particular ρ value, say ρ^*, which maximizes the random coding exponent

$$E(R, W) = \max_{\rho} \left[E_0(\rho, W) - \rho R\right].$$

For that particular ρ^*, we have

$$\begin{aligned} 2E(R, W) &= 2E_0(\rho^*, W) - 2\rho^* R \\ &\leq E_0(\rho^*, W^+) + E_0(\rho^*, W^-) - \rho^* 2R. \end{aligned}$$

Now, if the rate R is split into two parts R^+ and R^- proportional to $E_0(\rho^*, W^+)$ and $E_0(\rho^*, W^-)$, respectively, i.e. they satisfy $R = R^+ + R^-$ and

$$\frac{R^+}{E_0(\rho^*, W^+)} = \frac{R^-}{E_0(\rho^*, W^-)},$$

then the reliability-complexity trade-off function $E(R, W)/R$ of random codes will satisfy both

$$\frac{E(R, W)}{R} \leq \frac{E_0(\rho^*, W^-) - \rho^* R^-}{R^-} \leq \frac{E(R^-, W^-)}{R^-}$$

and

$$\frac{E(R, W)}{R} \leq \frac{E(R^+, W^+)}{R^+}.$$

Therefore, both of the synthesized channels will have a better reliability-complexity exponent function than the original channel. In that respect, the inequality in Theorem 2 implies the particular polar transform combined with a successive cancellation decoder does improve the reliability-complexity exponent of random codes.

C. Chain Rule for Rényi's Entropies

Rényi's entropy of order α of a discrete random variable $X \sim P(x)$ is defined in [5] as

$$H_\alpha(X) = \frac{\alpha}{1-\alpha} \log \left(\sum_x P(x)^\alpha\right)^{\frac{1}{\alpha}}.$$

The Rényi's conditional entropy of order α of a discrete random variable X given Y with joint distribution $P(x, y)$

is defined in [6] as

$$H_\alpha(X \mid Y) = \frac{\alpha}{1-\alpha} \log \sum_y \left(\sum_x P(x,y)^\alpha \right)^{\frac{1}{\alpha}}$$

$$= H_\alpha(X) + \frac{\alpha}{1-\alpha} \log \sum_y \left(\sum_x Q(x) P(y \mid x)^\alpha \right)^{\frac{1}{\alpha}}$$

where $Q(x) = \dfrac{P(x)^\alpha}{\sum_x P(x)^\alpha}$ is a distribution. If P is the uniform input distribution on the set $\mathcal{X}$ of the values of X, then Q is also the uniform distribution on $\mathcal{X}$, and letting $\alpha = \frac{1}{1+\rho}$, we get

$$H_{\frac{1}{1+\rho}}(X) = \frac{1}{\rho} \log \left(\sum_x P(x)^{\frac{1}{1+\rho}} \right)^{\frac{1}{1+\rho}} = \log |\mathcal{X}|, \tag{20}$$

$$H_{\frac{1}{1+\rho}}(X \mid Y) = H_{\frac{1}{1+\rho}}(X) + \frac{1}{\rho} \log \sum_y \left(\sum_x P(x) P(y \mid x)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{21}$$

From the definition of $E_0(\rho, W)$ in (4), we deduce

$$\frac{E_0(\rho, W)}{\rho} = H_{\frac{1}{1+\rho}}(X) - H_{\frac{1}{1+\rho}}(X \mid Y). \tag{22}$$

The quantity in the right hand side of (22) is called the mutual information of order $\frac{1}{1+\rho}$ in [6].

Using the definitions, $E_0(\rho, W_{1,2}^-)$ and $E_0(\rho, W_{1,2}^+)$ can be expressed as follows

$$\frac{E_0(\rho, W_{1,2}^-)}{\rho} = \log 2 - H_{\frac{1}{1+\rho}}(U_1 \mid Y_1 Y_2),$$

$$\frac{E_0(\rho, W_{1,2}^+)}{\rho} = \log 2 - H_{\frac{1}{1+\rho}}(U_2 \mid Y_1 Y_2 U_1),$$

where Y_1 is the output of the channel W_1 with input $X_1 = U_1 \oplus U_2$, and Y_2 is the output of the channel W_2 with input $X_2 = U_2$. In addition,

$$\frac{E_0(\rho, W_1)}{\rho} + \frac{E_0(\rho, W_2)}{\rho} = 2 \log 2 - H_{\frac{1}{1+\rho}}(X_1 X_2 \mid Y_1 Y_2).$$

Since the mapping between (X_1, X_2) and (U_1, U_2) is one-to-one, we have

$$H_{\frac{1}{1+\rho}}(X_1 X_2 \mid Y_1 Y_2) = H_{\frac{1}{1+\rho}}(U_1 U_2 \mid Y_1 Y_2).$$

The relationship derived between $E_0(\rho, W_1)$, $E_0(\rho, W_2)$, $E_0(\rho, W_{1,2}^-)$, and $E_0(\rho,_{1,2} W^+)$ in Theorem 2 implies a certain "chain rule inequality" holds for the polarization transformation, i.e.,

$$H_{\frac{1}{1+\rho}}(U_1 U_2 \mid Y_1 Y_2) \geq H_{\frac{1}{1+\rho}}(U_1 \mid Y_1 Y_2) + H_{\frac{1}{1+\rho}}(U_2 \mid Y_1 Y_2 U_1)$$

for $\rho \geq 0$ whenever U_1, U_2 are i.i.d., uniform on $\mathbb{F}_2$.

Equivalently, we can conclude that whenever (i) (X_1, Y_1) and (X_2, Y_2) are independent, and (ii) X_1 and X_2 are uniformly distributed on $\mathbb{F}_2$, then, with $U_1 = X_1 \oplus X_2$ and $U_2 = X_2$,

$$H_\alpha(U_1 U_2 | Y_1 Y_2) \geq H_\alpha(U_1 | Y_1 Y_2) + H_\alpha(U_2 | Y_1 Y_2 U_1)$$

for any $\alpha \leq 1$.

Appendix

Lemma 1: For $s \in [0,1]$, define the function $F_s : [0,1] \to [1, 2^{\frac{1-s}{s}}]$ as

$$F_s(x) = \frac{(1+x^s)^{\frac{1}{s}}}{1+x}. \tag{23}$$

Then, F_s is a non-decreasing function.

Proof: Taking the derivative of $F_s(x)$ with respect to x, we have

$$\frac{\partial}{\partial x} F_s(x) = \frac{(1+x^s)^{\frac{1}{s}-1}(x^s - x)}{x(1+x)^2} \geq 0$$

since $(x^s - x) \geq 0$ for $x, s \in [0,1]$. ■

Lemma 2: For $s \in [0,1]$, define the function $f_s : [0, \infty) \to [2^{\frac{s-1}{s}}, 1]$ as

$$f_s(k) = \frac{\cosh(ks)^{\frac{1}{s}}}{\cosh(k)}. \tag{24}$$

Then, f_s is a non-increasing function. Moreover, this implies the function $g(\rho, z)$ defined in (8) is non-increasing in the variable $z \in [0,1]$ for any fixed $\rho \geq 0$.

Proof: We can equivalently show that $\log(f_s(k))$ is non-increasing in k. Taking the first derivative gives

$$\frac{\partial}{\partial k} \left(\frac{1}{s} \log(\cosh(ks)) - \log(\cosh(k)) \right) = \tanh(sk) - \tanh(k) \leq 0$$

as $\tanh(\cdot)$ is increasing in its argument.
To prove the second monotonicity relation, we let $k = \operatorname{arctanh} z$ and $s = \frac{1}{1+\rho}$. Then,

$$g(\rho, z) = f_{\frac{1}{1+\rho}}(\operatorname{arctanh} z).$$

Since arctanh is a monotone increasing function, it follows that the function $g(\rho, z)$ is non-increasing in z. ■

Lemma 3: The function $h(\rho, z_1, z_2) : [0, \infty) \times [0,1] \times [0,1] \to [2^{-\rho}, 1]$ defined in (9), is non-increasing in the variables z_1 and z_2 separately for any $\rho \geq 0$.

Proof: By the symmetry of h with respect to z_1 and z_2, it suffices to show the claim for z_1 alone. In the expression below, we will suppress ρ in all function arguments, and denote $g'(u) = \frac{\partial}{\partial u} g(\rho, u)$. Taking the derivative of h with respect to z_1, we get

$$\begin{aligned}
&\frac{\partial}{\partial z_1} h(z_1, z_2) \\
&\quad = \frac{1}{2} z_2\, g\Big(\frac{z_1+z_2}{1+z_1z_2}\Big) + \frac{1-{z_2}^2}{2(1+z_1z_2)}\, g'\Big(\frac{z_1+z_2}{1+z_1z_2}\Big) \\
&\qquad - \frac{1}{2} z_2\, g\Big(\frac{z_1-z_2}{1-z_1z_2}\Big) + \frac{1-{z_2}^2}{2(1-z_1z_2)}\, g'\Big(\frac{z_1-z_2}{1-z_1z_2}\Big) \\
&\quad = \frac{1}{2} z_2 \Big[g\left(\frac{z_1+z_2}{1+z_1z_2}\right) - g\left(\frac{z_1-z_2}{1-z_1z_2}\right)\Big] \\
&\qquad + \frac{1-{z_2}^2}{2(1+z_1z_2)}\, g'\Big(\frac{z_1+z_2}{1+z_1z_2}\Big) \\
&\qquad + \frac{1-{z_2}^2}{2(1-z_1z_2)}\, g'\Big(\frac{z_1-z_2}{1-z_1z_2}\Big).
\end{aligned}$$

The last two terms that contain $g'(\cdot)$ are negative by Lemma 2, so it suffices to show that

$$g\left(\frac{z_1+z_2}{1+z_1z_2}\right) \leq g\left(\frac{z_1-z_2}{1-z_1z_2}\right).$$

To that end, observe that, for any $z_1, z_2 \in [0,1]$ we have

$$\frac{z_1+z_2}{1+z_1z_2} \geq \frac{|z_1-z_2|}{1-z_1z_2}$$

and by Lemma 2 and the symmetry of g around $z=0$, the required inequality follows. ■

Lemma 4: Suppose $f : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ and $g : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ are two functions that satisfy

$$\big[f(x,y) - f(x',y)\big]\big[g(x,y') - g(x',y')\big] \geq 0,$$

and

$$\big[f(x,y) - f(x,y')\big]\big[g(x,y) - g(x,y')\big] \geq 0$$

for every x, x', y, y'. Then, for any independent random variables X, Y the random variables $f(X,Y)$ and $g(X,Y)$ are positively correlated.

Note that if $\mathcal{X}$ and $\mathcal{Y}$ are ordered sets and f and g are monotone (in the same sense) in their arguments then they satisfy the requirements of the lemma.

Proof: Let (X', Y') be an independent copy of (X, Y). By the first premise of the lemma

$$\big[f(X,Y) - f(X',Y)\big]\big[g(X,Y') - g(X',Y')\big] \geq 0.$$

Taking expectations, we get

$$\begin{aligned}
&\mathbb{E}[f(X,Y)g(X,Y')] + \mathbb{E}[f(X',Y)g(X',Y')] \\
&\qquad \geq \mathbb{E}[f(X,Y)g(X',Y')] + \mathbb{E}[f(X',Y)g(X,Y')],
\end{aligned}$$

equivalently,

$$\mathbb{E}[f(X,Y)g(X,Y')] \geq \mathbb{E}[f(X,Y)]\mathbb{E}[g(X,Y)]. \tag{25}$$

By the second premise of the lemma

$$\big[f(X,Y) - f(X,Y')\big]\big[g(X,Y) - g(X,Y')\big] \geq 0.$$

Taking expectations, we get

$$\begin{aligned}
&\mathbb{E}[f(X,Y)g(X,Y)] + \mathbb{E}[f(X,Y')g(X,Y')] \\
&\qquad \geq \mathbb{E}[f(X,Y)g(X,Y')] + \mathbb{E}[f(X,Y')g(X,Y)]
\end{aligned}$$

which is equivalent to

$$\mathbb{E}[f(X,Y)g(X,Y)] \geq \mathbb{E}[f(X,Y)g(X,Y')] \tag{26}$$

Putting together (25) and (26) concludes the proof. ■

IV. Acknowledgments

This work was supported by Swiss National Science Foundation under grant number 200021-125347/1.

References

[1] E. Arıkan, “Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels”, *IEEE Trans. Inf. Theor.*, 55(7):3051–3073, 2009.

[2] E. Arıkan, *“Channel combining and splitting for cutoff rate improvement”*, *IEEE Trans. Inf. Theor.*, 52(2):628-639, 2006.

[3] R. G. Gallager, “Information Theory and Reliable Communication”, John Wiley & Sons, Inc., New York, NY, USA, 1968.

[4] M. Alsan, “Extremality Properties for the Basic Polarization Transformations”, *eprint arXiv:1301.5258*, January 2013.

[5] A. Rényi, “On Measures of Entropy and Information”, *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*,Vol. 1, pages 547–561, 1961.

[6] S. Arimoto, “Information measures and capacity of order α for discrete memoryless channels”, in Topics in information theory, I.Csiszar and P. Elias, editors, Amsterdam, The Netherlands, 1977. North-Holland Publishing Co.