

# On the Polarization of Rényi Entropy

Mengfan Zheng  
Imperial College London,  
United Kingdom  
Email: m.zheng@imperial.ac.uk

Ling Liu  
Huawei Technologies Co. Ltd.  
Shenzhen, P. R. China  
Email: liuling\_88@pku.edu.cn

Cong Ling  
Imperial College London,  
United Kingdom  
Email: c.ling@imperial.ac.uk

**Abstract**—Existing polarization theories have mostly been concerned with Shannon’s information measures, such as Shannon entropy and mutual information, and some related measures such as the Bhattacharyya parameter. In this work, we extend polarization theories to a more general information measure, namely, the Rényi entropy. Our study shows that under conditional Rényi entropies of different orders, the same synthetic sub-channel may exhibit opposite extremal states. This result reveals more insights into the polarization phenomenon on the micro scale (probability pairs) rather than on the average scale.

## I. INTRODUCTION

The polarization technique (including channel polarization [1] and source polarization [2]) is one of the most significant breakthrough in information theory over the past decade. Arkan showed us that as the size of the polar transformation goes to infinity, the conditional entropies of the synthetic sub-channels (or random variable pairs) equal 0 or 1 almost everywhere (a.e.) [1]. Also, their varentropies (variance of the conditional entropy random variable) asymptotically decrease to zero [3]. These results imply that the sub-channels’ transition probability matrices tend to either deterministic (noiseless channels) or uniform with respect to any channel input (completely noisy channels). However, are they still close to uniform or deterministic distributions under stricter criteria? Polarization results using Shannon’s information measures fail to answer this question.

In this work, we study polarization using a more general information measure, i.e., the Rényi entropy [4]. The Rényi entropy is more sensitive to deviations from uniform or deterministic distributions, and has been used in many areas where Shannon entropy may not be a good metric. For example, the collision entropy and min-entropy, both of which are special cases of the Rényi entropy, are convenient metrics for privacy amplification in secret-key agreement [5]. The Rényi entropy of a random variable  $X$  is defined as follows.

**Definition 1** (Rényi Entropy [4]). *The Rényi entropy of a random variable  $X \in \mathcal{X}$  of order  $\alpha$  is defined as*

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P_X(x)^\alpha, \quad (1)$$

It can be shown that as  $\alpha \rightarrow 1$ , the Rényi entropy reduces to the Shannon entropy. Two other special cases of the Rényi entropy which will be discussed later include the *max-entropy*:

$$H_0(X) = \log |\mathcal{X}|, \quad (2)$$

which is the Rényi entropy of order 0, and the *min-entropy*:

$$H_\infty(X) = \min_i (-\log p_i) = -\log \max_i p_i, \quad (3)$$

which equals the limiting value of  $H_\alpha(X)$  as  $\alpha \rightarrow \infty$ .

Unlike the conditional Shannon entropy, there is no generally accepted definition of the conditional Rényi entropy yet. In this paper, we adopt the following definition of conditional Rényi entropy in the study of polarization.

**Definition 2** (conditional Rényi entropy [6], [7]). *The conditional Rényi entropy of order  $\alpha$  of  $X$  given  $Y$  is defined as*

$$H_\alpha(X|Y) = \frac{1}{1-\alpha} \log \frac{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} P_{X,Y}(x,y)^\alpha}{\sum_{y \in \mathcal{Y}} P_Y(y)^\alpha}. \quad (4)$$

Note that this type of conditional Rényi entropy satisfies the chain rule:

$$H_\alpha(X|Y) + H_\alpha(Y) = H_\alpha(X,Y). \quad (5)$$

This means that it reduces to the conditional Shannon entropy when  $\alpha = 1$ .

There have been very limited researches in regard to polarization of conditional Rényi entropies. In [8] it is shown that the following chain rule inequality holds for the polar transformation for  $\alpha \leq 1$ ,

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

whenever  $U_1, U_2$  are i.i.d. uniform on  $\mathbb{F}_2$ . The inequality holds with equality if and only if the channel  $W$  is perfect, or the channel  $W$  is completely noisy, or  $\alpha = 1$ . Note that  $H_\alpha^*(X|Y)$  in [8] is defined as  $H_\alpha^*(X|Y) = \frac{\alpha}{1-\alpha} \log \sum_{y \in \mathcal{Y}} P_Y(y) \left[ \sum_{x \in \mathcal{X}} P_{X|Y}(x|y)^\alpha \right]^{\frac{1}{\alpha}}$ .

In this paper, we also restrict ourselves to the binary case (i.e.,  $\mathcal{X} = \mathbb{F}_2$ ) as a starting point. As a result, logarithms in this paper will all be base-2. However, we do not assume  $X$  to be uniformly distributed. We prove that the order- $\alpha$  conditional Rényi entropies of the synthetic sub-channels also polarize to 0 and 1, but the fraction of 1 tends to  $H_\alpha(X|Y)$ . This is to say that for a given sub-channel, its conditional Rényi entropies of different orders may exhibit opposite extremal states. Intuitively, if a sub-channel is truly noiseless (or truly completely noisy), its conditional Rényi entropies of different orders should all be 0 (or 1). We show both analytically and numerically that this strange phenomenon is caused by a

vanishing deviation from truly deterministic or truly uniform distributions. The different extremal states that a sub-channel exhibits for various  $\alpha$  reflect the polarization level of the sub-channel at the micro scale, i.e., how close is its joint distribution to truly uniform or truly deterministic.

## II. POLARIZATION OF RÉNYI ENTROPY

### A. Polarization of a Basic Polar Transformation

Consider two binary-input discrete memoryless channels (B-DMC)  $P_{Y_1|X_1}$  and  $P_{Y_2|X_2}$  with the same input distribution  $P_X$  in the channel coding scenario, or consider  $(X_1, Y_1)$  and  $(X_2, Y_2)$  as two samples of a memoryless source  $(X, Y) \sim P_{X,Y}$  with  $X$  being the binary source to be compressed and  $Y$  being the side information about  $X$  in the source coding scenario. Note that in the former case,  $P_{Y_1|X_1}$  and  $P_{Y_2|X_2}$  can be different, which corresponds to the compound channel setting. Let

$$U_1 = X_1 \oplus X_2, \quad U_2 = X_2, \quad (6)$$

and denote

$$\begin{aligned} & P_A(u_1, u_2, y_1, y_2) \\ & \triangleq P_{U_1, U_2}(u_1, u_2) P_{Y_1|X_1}(y_1|u_1 \oplus u_2) P_{Y_2|X_2}(y_2|u_2) \end{aligned}$$

for short. From (6) we know that

$$\begin{aligned} P_{U_1, U_2}(0, 0) &= P_X(0)P_X(0), \quad P_{U_1, U_2}(0, 1) = P_X(1)P_X(1), \\ P_{U_1, U_2}(1, 0) &= P_X(1)P_X(0), \quad P_{U_1, U_2}(1, 1) = P_X(0)P_X(1). \end{aligned} \quad (7)$$

For the basic polar transformation, we have the following polarization result.

**Lemma 1.** *For  $\alpha \geq 0$ , we have*

$$H_\alpha(U_2|Y_1Y_2U_1) \leq \min\{H_\alpha(X_1|Y_1), H_\alpha(X_2|Y_2)\}. \quad (8)$$

$$H_\alpha(U_1|Y_1Y_2) \geq \max\{H_\alpha(X_1|Y_1), H_\alpha(X_2|Y_2)\}, \quad (9)$$

$$H_\alpha(U_1U_2|Y_1Y_2) = H_\alpha(U_1|Y_1Y_2) + H_\alpha(U_2|Y_1Y_2U_1). \quad (10)$$

Let us first recall the Minkowski inequality before proving Lemma 1.

**Proposition 1.** *The Minkowski inequality states that for  $1 \leq p \leq \infty$ ,*

$$\left( \sum_{k=1}^n |x_k + y_k|^p \right)^{\frac{1}{p}} \leq \left( \sum_{k=1}^n |x_k|^p \right)^{\frac{1}{p}} + \left( \sum_{k=1}^n |y_k|^p \right)^{\frac{1}{p}}, \quad (11)$$

with equality if and only if  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  and  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  are positively linearly dependent, i.e.,  $\mathbf{x} = \lambda \mathbf{y}$  for some  $\lambda \geq 0$  or  $\mathbf{y} = 0$ . For  $0 < p < 1$ , the inequality in (11) is reversed.

*Proof of Lemma 1.* (I) First, we consider (10). Denote

$$S_1 = \sum_{\{y_1, y_2\} \in \mathcal{Y}^2} \left( \sum_{\{u_1, u_2\} \in \mathcal{X}^2} P_A(u_1, u_2, y_1, y_2) \right)^\alpha \quad (12)$$

$$S_2 = \sum_{\{y_1, y_2\} \in \mathcal{Y}^2} \sum_{u_1 \in \mathcal{X}} \left( \sum_{u_2 \in \mathcal{X}} P_A(u_1, u_2, y_1, y_2) \right)^\alpha \quad (13)$$

$$S_3 = \sum_{\{y_1, y_2\} \in \mathcal{Y}^2} \sum_{\{u_1, u_2\} \in \mathcal{X}^2} \left( P_A(u_1, u_2, y_1, y_2) \right)^\alpha \quad (14)$$

Then we have

$$\begin{aligned} & H_\alpha(U_1|Y_1Y_2) + H_\alpha(U_2|Y_1Y_2U_1) \\ &= \frac{1}{1-\alpha} \log \frac{S_2}{S_1} + \frac{1}{1-\alpha} \log \frac{S_3}{S_2} \\ &= \frac{1}{1-\alpha} \log \frac{S_3}{S_1} \\ &= H_\alpha(U_1U_2|Y_1Y_2). \end{aligned}$$

(II) Next, we consider (9).  $H_\alpha(X_1|Y_1)$  can be expressed as (15) (on the top of the next page), where

$$\begin{aligned} S_4 &= \sum_{\{y_1, y_2\} \in \mathcal{Y}^2} \left\{ \left[ P_{Y_1, X_1}(y_1, 0) + P_{Y_1, X_1}(y_1, 1) \right]^\alpha \right. \\ &\quad \left. \times \left[ P_{Y_2, X_2}(y_2, 0)^\alpha + P_{Y_2, X_2}(y_2, 1)^\alpha \right] \right\} \end{aligned}$$

From (13) and (7) we have

$$\begin{aligned} S_2 &= \sum_{\{y_1, y_2\} \in \mathcal{Y}^2} \left\{ \left[ P_{Y_1, X_1}(y_1, 0) P_{Y_2, X_2}(y_2, 0) \right. \right. \\ &\quad \left. \left. + P_{Y_1, X_1}(y_1, 1) P_{Y_2, X_2}(y_2, 1) \right]^\alpha \right. \\ &\quad \left. + \left[ P_{Y_1, X_1}(y_1, 1) P_{Y_2, X_2}(y_2, 0) \right. \right. \\ &\quad \left. \left. + P_{Y_1, X_1}(y_1, 0) P_{Y_2, X_2}(y_2, 1) \right]^\alpha \right\}. \end{aligned}$$

For  $0 < \alpha < 1$ , by the Minkowski inequality we have

$$\begin{aligned} & \left\{ \left[ P_{Y_1, X_1}(y_1, 0) P_{Y_2, X_2}(y_2, 0) + P_{Y_1, X_1}(y_1, 1) P_{Y_2, X_2}(y_2, 1) \right]^\alpha \right. \\ &\quad \left. + \left[ P_{Y_1, X_1}(y_1, 1) P_{Y_2, X_2}(y_2, 0) \right. \right. \\ &\quad \left. \left. + P_{Y_1, X_1}(y_1, 0) P_{Y_2, X_2}(y_2, 1) \right]^\alpha \right\}^{\frac{1}{\alpha}} \\ &\geq \left\{ \left[ P_{Y_1, X_1}(y_1, 0) P_{Y_2, X_2}(y_2, 0) \right]^\alpha \right. \\ &\quad \left. + \left[ P_{Y_1, X_1}(y_1, 0) P_{Y_2, X_2}(y_2, 1) \right]^\alpha \right\}^{\frac{1}{\alpha}} \\ &\quad + \left\{ \left[ P_{Y_1, X_1}(y_1, 1) P_{Y_2, X_2}(y_2, 0) \right]^\alpha \right. \\ &\quad \left. + \left[ P_{Y_1, X_1}(y_1, 1) P_{Y_2, X_2}(y_2, 1) \right]^\alpha \right\}^{\frac{1}{\alpha}} \\ &= \left[ P_{Y_1, X_1}(y_1, 0) + P_{Y_1, X_1}(y_1, 1) \right] \\ &\quad \times \left[ P_{Y_2, X_2}(y_2, 0)^\alpha + P_{Y_2, X_2}(y_2, 1)^\alpha \right]^{\frac{1}{\alpha}}. \end{aligned} \quad (16)$$

$$H_\alpha(X_1|Y_1) = \frac{1}{1-\alpha} \log \frac{\left(\sum_{y_1 \in \mathcal{Y}} \sum_{x_1 \in \mathcal{X}} P_{Y_1, X_1}(y_1, x_1)^\alpha\right) \left(\sum_{y_2 \in \mathcal{Y}} \sum_{x_2 \in \mathcal{X}} P_{Y_2, X_2}(y_2, x_2)^\alpha\right)}{\left(\sum_{y_1 \in \mathcal{Y}} \left[\sum_{x_1 \in \mathcal{X}} P_{Y_1, X_1}(y_1, x_1)\right]^\alpha\right) \left(\sum_{y_2 \in \mathcal{Y}} \sum_{x_2 \in \mathcal{X}} P_{Y_2, X_2}(y_2, x_2)^\alpha\right)} = \frac{1}{1-\alpha} \log \frac{S_3}{S_4} \quad (15)$$

Thus,  $S_2 \geq S_4$ , which means

$$\begin{aligned} H_\alpha(U_2|Y_1 Y_2 U_1) &= \frac{1}{1-\alpha} \log \frac{S_3}{S_2} \\ &\leq \frac{1}{1-\alpha} \log \frac{S_3}{S_4} = H_\alpha(X_1|Y_1). \end{aligned} \quad (17)$$

For  $\alpha > 1$ , the inequality in (16) is reversed. However, since  $1 - \alpha < 0$ , the result remains the same. For  $\alpha = 1$ , the Renyi entropy reduces to the Shannon entropy, and the polarization result is identical to the result here [1]. For  $\alpha = 0$ , it is obvious from (2) and (4) that (10) holds, while (8) and (9) hold with equality.

Similarly,  $H_\alpha(U_2|Y_1 Y_2 U_1) \leq H_\alpha(X_2|Y_2)$  can be proved.

(III) Finally, equality (10) and inequality (9) immediately imply (8).  $\square$

### B. Recursive Polar Transformation

Now consider extending the basic transformation recursively to higher orders. For  $N = 2^n$  with  $n$  being an arbitrary integer, the recursive transformation can be expressed as  $U^{1:N} = X^{1:N} \mathbf{G}_N$  [1], where  $\mathbf{G}_N = \mathbf{B}_N \mathbf{F}^{\otimes n}$  with  $\mathbf{B}_N$  being the bit-reversal matrix and  $\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ . Denote  $H_N(i) = H_\alpha(U^i|Y^{1:N} U^{1:i-1})$ . We have the following theorem.

**Theorem 1.** *For any B-DMC  $P_{Y|X}$  (or any discrete memoryless source  $(X, Y) \sim P_{X,Y}$  over  $\mathcal{X} \times \mathcal{Y}$  with  $\mathcal{X} = \{0, 1\}$  and  $\mathcal{Y}$  an arbitrary countable set) and any  $\alpha \geq 0$ , as  $N \rightarrow \infty$  through the power of 2, the fraction of indices  $i \in [N] \triangleq \{1, 2, \dots, N\}$  with  $H_N(i) \in (1 - \delta, 1]$  goes to  $H_\alpha(X|Y)$ , and the fraction with  $H_N(i) \in [0, \delta)$  goes to  $1 - H_\alpha(X|Y)$ .*

*Proof.* We follow Arkan's martingale approach [1], [9] to complete the proof. First, we introduce the same infinite binary tree as in the proof of [1, Theorem 1], with a root node at level 0 and  $2^n$  nodes at level  $n$ . Then define a random walk  $\{B_n; n \geq 0\}$  in this tree as follows. The random walk starts at the root node with  $B_0 = (0, 1)$ , and moves to one of the two child nodes in the next level with equal probability at each integer time. If  $B_n = (n, i)$ ,  $B_{n+1}$  equals  $(n+1, 2i-1)$  or  $(n+1, 2i)$  with probability  $1/2$  each. Denote  $H(0, 1) = H_\alpha(X|Y)$  and  $H(n, i) = H_\alpha(U^i|Y^{1:2^n} U^{1:i-1})$  for  $n \geq 1$ ,  $i \in [2^n]$ . Define a random process  $\{H_n; n \geq 0\}$  with  $H_n = H(B_n)$ . It can be shown that the process  $\{H_n; n \geq 0\}$  is a martingale due to the chain rule equality of (10), i.e.,

$$E[H_{n+1}|B_0, B_1, \dots, B_n] = H_n. \quad (18)$$

Since  $\{H_n; n \geq 0\}$  is a uniformly integrable martingale, it converges a.e. to an RV  $H_\infty$  such that  $E[|H_n - H_\infty|] = 0$ . Then we have

$$E[|H_n - H_{n+1}|] \rightarrow 0. \quad (19)$$

Note that

$$\begin{aligned} &E[|H_n - H_{n+1}|] \\ &= \frac{1}{2} \left( E[|H_\alpha(U^i|Y^{1:2^n}, U^{1:i-1}) - H_\alpha(U^{2i-1}|Y^{1:2^{n+1}}, U^{1:2i-2})|] \right. \\ &\quad \left. + E[|H_\alpha(U^i|Y^{1:2^n}, U^{1:i-1}) - H_\alpha(U^{2i}|Y^{1:2^{n+1}}, U^{1:2i-1})|] \right). \end{aligned} \quad (20)$$

In Lemma 1, by letting  $X_1 = X_2 = U^i$ ,  $Y_1 = Y_2 = (Y^{1:2^n}, U^{1:i-1})$ ,  $U_1 = U^{2i-1}$ , and  $U_2 = U^{2i}$ , we can see that (19) and (20) force (16) to hold with equality for  $i \in [2^n]$  a.e. as  $n \rightarrow \infty$ . From Proposition 1 we know that (16) holds with equality only in the following two cases:

- (Case 1)  $P_{Y_1, X_1}(y_1, 0) = 0$  or  $P_{Y_1, X_1}(y_1, 1) = 0$  for all effective  $y_1 \in \mathcal{Y}$ .
- (Case 2)  $P_{Y_2, X_2}(y_2, 0) = P_{Y_2, X_2}(y_2, 1)$  for all effective  $y_2 \in \mathcal{Y}$ .

The effective elements of  $\mathcal{Y}$  (denoted by  $\mathcal{Y}_e \subset \mathcal{Y}$ ) mean that

$$\frac{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}_e} P_{X,Y}(x,y)^\alpha}{\sum_{y \in \mathcal{Y}_e} P_Y(y)^\alpha} \rightarrow \frac{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} P_{X,Y}(x,y)^\alpha}{\sum_{y \in \mathcal{Y}} P_Y(y)^\alpha}$$

as  $n \rightarrow \infty$ . A more detailed discussion about it will be given in the next subsection. Since  $P_{Y_1, X_1}$  and  $P_{Y_2, X_2}$  are identical in the recursive transformation, the joint distributions of  $P_{Y^i, U^i}$  ( $i \in [2^n]$ ) tend to either Case 1 or Case 2 a.e. as  $n \rightarrow \infty$ , where  $\mathbf{Y}^i = (Y^{1:2^n}, U^{1:i-1})$ . It is easy to verify that  $H_\alpha(U^i|\mathbf{Y}^i)$  equals 0 in Case 1 and 1 in Case 2. This shows that  $H_n$  converges a.e. to 0 and 1 as  $n \rightarrow \infty$ .

The convergence result together with the chain rule equality of (10) imply that the fraction of  $\{i : H(n, i) \in (1 - \delta, 1]\}$  goes to  $H_\alpha(X|Y)$  as  $n \rightarrow \infty$ .  $\square$

Now we can answer the question raised at the beginning of this paper. From the definition of the max-entropy  $H_0(X)$  and (5) we know that

$$H_0(X|Y) = \log |\mathcal{X} \times \mathcal{Y}| - \log |\mathcal{Y}| = 1, \quad (21)$$

provided that the probabilities  $P_{Y^i, U^i}(\mathbf{y}^i, u^i)$  are nonzero. Since  $P_{Y^i, U^i}(\mathbf{y}^i, u^i)$  never really become 0 by the polar transformation, we know that the fraction of sub-channels with  $H_0(U^i|\mathbf{Y}^i) \rightarrow 0$  is always 0, which means that the synthetic

sub-channels never really polarize to truly deterministic state. On the contrary, the min-conditional-entropy,

$$\begin{aligned} H_\infty(X|Y) &= H_\infty(XY) - H_\infty(Y) \\ &= \log \frac{\max_{y \in \mathcal{Y}} p_y}{\max_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} p_{x,y}}, \end{aligned} \quad (22)$$

is not a constant in general. As will be explained in Section III, only truly uniform distribution yields  $H_\infty(X|Y) = 1$ . Therefore, as  $n \rightarrow \infty$ , a fraction  $H_\infty(X|Y)$  of the sub-channels will become truly completely noisy. This result can be concluded by the following corollary.

**Corollary 1.** *As  $n \rightarrow \infty$ , the fraction of truly completely noisy sub-channels tends to  $H_\infty(X|Y)$ , while the fraction of truly noiseless sub-channels is 0.*

### C. An Example of Effective Elements

The reason we introduce the term of effective elements is that, without the "effective" in the definitions of Case 1 and Case 2, the conditional Rényi entropies in Case 1 and Case 2 of different orders should all be 0 and 1, respectively. In this subsection, we further discuss this issue. First, let us present an example to show that for a given  $\alpha = \alpha_0$  and  $\alpha' = \alpha_0 + 1$ ,  $H_\alpha(X|Y)$  can be arbitrarily close to 1 while  $H_{\alpha'}(X|Y)$  can be arbitrarily close to 0.

Let  $|\mathcal{Y}| = 2^N \triangleq M$ . Consider a joint distribution  $P_{X,Y}$  such that a fraction  $\frac{1}{L}$  of probability pairs  $(P_{X,Y}(0, y), P_{X,Y}(1, y))$  are completely deterministic with accumulated probability of  $\frac{1}{N}$ . Without loss of generality, we assume that  $P_{X,Y}(0, y_i) = \frac{1}{NM}$  and  $P_{X,Y}(1, y_i) = 0$  for  $i \in \mathcal{A}$ , where  $\mathcal{A} \subset [2^N]$  is the set of deterministic pairs. The rest  $\frac{L-1}{L}$  fraction of probability pairs are completely uniform, i.e.,  $P_{X,Y}(0, y_i) = P_{X,Y}(1, y_i) = \frac{(N-1)L}{2NM(L-1)}$  for  $i \in \mathcal{A}^C$ . Then

$$\begin{aligned} H_\alpha(X|Y) &= \frac{1}{1-\alpha} \log \frac{\sum_{i \in \mathcal{A}} (\frac{L}{NM})^\alpha + 2^{1-\alpha} \sum_{i \in \mathcal{A}^C} (\frac{(N-1)L}{NM(L-1)})^\alpha}{\sum_{i \in \mathcal{A}} (\frac{L}{NM})^\alpha + \sum_{i \in \mathcal{A}^C} (\frac{(N-1)L}{NM(L-1)})^\alpha} \\ &= \frac{1}{1-\alpha} \log \frac{\frac{M}{L} (\frac{L}{NM})^\alpha + 2^{1-\alpha} \frac{M(L-1)}{L} (\frac{(N-1)L}{NM(L-1)})^\alpha}{\frac{M}{L} (\frac{L}{NM})^\alpha + \frac{M(L-1)}{L} (\frac{(N-1)L}{NM(L-1)})^\alpha} \\ &= \frac{1}{1-\alpha} \log \frac{(L-1)^{\alpha-1} + 2^{1-\alpha} (N-1)^\alpha}{(L-1)^{\alpha-1} + (N-1)^\alpha}. \end{aligned} \quad (23)$$

For a considered  $\alpha = \alpha_0 > 1$ , let

$$L-1 = 2^{-1} (N-1)^{\frac{\alpha_0-0.5/\alpha_0}{\alpha_0-1}}. \quad (24)$$

Then we have

$$H_{\alpha_0}(X|Y) = \frac{1}{1-\alpha_0} \log \frac{2^{1-\alpha_0} [(N-1)^{\alpha_0-0.5/\alpha_0} + (N-1)^{\alpha_0}]}{2^{1-\alpha_0} (N-1)^{\alpha_0-0.5/\alpha_0} + (N-1)^{\alpha_0}}.$$

It is clear that as  $N \rightarrow \infty$ ,  $\frac{(N-1)^{\alpha_0}}{(N-1)^{\alpha_0-0.5/\alpha_0}} = (N-1)^{0.5/\alpha_0} \rightarrow \infty$ , thus  $H_{\alpha_0}(X|Y) \rightarrow 1$ .

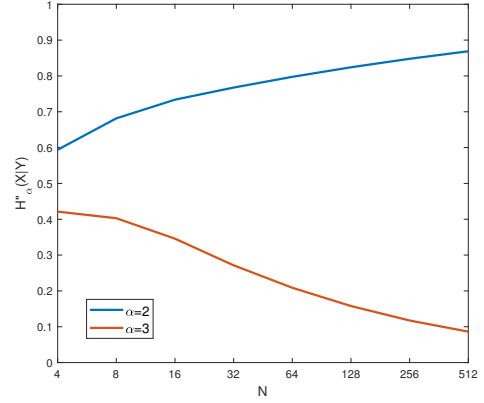


Fig. 1. conditional Rényi entropies of different  $\alpha$ .

Now let  $\alpha' = \alpha_0 + 1$ . From (23) we have

$$\begin{aligned} H_{\alpha'}(X|Y) &= \frac{1}{-\alpha_0} \log \frac{2^{-\alpha_0} (N-1)^{\frac{\alpha_0^2-0.5}{\alpha_0-1}} + 2^{-\alpha_0} (N-1)^{\alpha_0+1}}{2^{-\alpha_0} (N-1)^{\frac{\alpha_0^2-0.5}{\alpha_0-1}} + (N-1)^{\alpha_0+1}}. \end{aligned}$$

In this case, as  $N \rightarrow \infty$ ,  $\frac{(N-1)^{\alpha_0+1}}{(N-1)^{\frac{\alpha_0^2-0.5}{\alpha_0-1}}} = (N-1)^{\frac{-0.5}{\alpha_0-1}} \rightarrow 0$ ,

thus  $H_{\alpha'}(X|Y) \rightarrow 0$ .

From (24) we can see that  $\frac{1}{L} \rightarrow 0$  as  $N \rightarrow \infty$  if  $\alpha_0 > 1$ . Also,  $\frac{1}{N} \rightarrow 0$  as  $N \rightarrow \infty$ . Therefore, we have shown a case when  $H_\alpha(X|Y)$  and  $H_{\alpha+1}(X|Y)$  can be completely opposite asymptotically. Fig. 1 shows the example of  $\alpha_0 = 2$ .

We can similarly design such an example for  $0 < \alpha < 1$ . This shows that the extreme cases defined in the proof of Theorem 1 are relative. Even if almost all probability pairs of a sub-channel are uniform (so that it may seem to be of Case 2), when powered by some  $\alpha$ , these probability pairs may have little impact on the value of  $H_\alpha(X|Y)$  (so that the sub-channel is actually of Case 1), just as our example has shown. Thus, although we proved Theorem 1 for any  $\alpha \geq 0$  in a unified form, the criterion for judging whether a sub-channel converges to Case 1 or Case 2 depends on  $\alpha$ .

### D. Numerical Results

In this subsection we present the polarization of a binary symmetric channel (BSC) with crossover probability 0.2 numerically. We calculate the joint distribution of each synthetic sub-channel and then compute its conditional Rényi entropies of order 0.1, 0.5, 1, 2, 10 and 100, respectively. The result is shown in Fig. 2. The channel indices are reordered so that the conditional Shannon entropies increase monotonically. The dash lines demonstrate the proportions of extremal sub-channels as  $n \rightarrow \infty$ . Although the considered block-length is not long, this figure clearly shows that the polarized sets vary with  $\alpha$ . We can also see that a relatively higher conditional Shannon entropy does not necessarily imply a relatively higher conditional Rényi entropy of a different order.

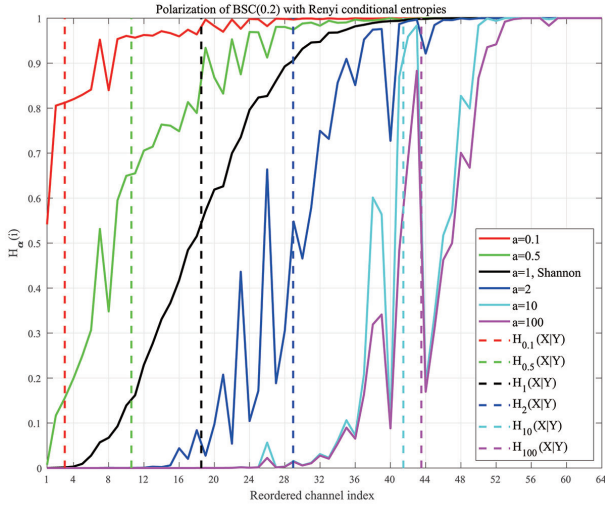


Fig. 2. Polarization of conditional Rényi entropies.

### III. A DISCUSSION ON RÉNYI ENTROPY

In this section, we further discuss how small the deviation from a uniform or deterministic distribution should be to make the conditional Rényi entropy achieve 1 or 0. Let  $\tilde{Q}_{X,Y}$  be a uniform distribution with respect to  $X$ , i.e.,  $\tilde{Q}_{X,Y}(0, y) = \tilde{Q}_{X,Y}(1, y) = \frac{1}{2}\tilde{Q}_Y(y)$  for any  $y \in \mathcal{Y}$ , and further assume that  $\tilde{Q}_Y(y) = P_Y(y)$ . Let  $\tilde{Q}_{X,Y}$  be a deterministic distribution with respect to  $X$ , i.e.,  $\tilde{Q}_{X,Y}(0, y) = 0, \tilde{Q}_{X,Y}(1, y) = \tilde{Q}_Y(y)$  or  $\tilde{Q}_{X,Y}(1, y) = 0, \tilde{Q}_{X,Y}(0, y) = \tilde{Q}_Y(y)$  for any  $y \in \mathcal{Y}$ , and further assume that  $\tilde{Q}_Y(y) = P_Y(y)$ . Then we have

$$\begin{aligned} H_\alpha(X|Y) &= \frac{1}{1-\alpha} \log \left\{ \frac{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} P_{X,Y}(x, y)^\alpha}{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} \tilde{Q}_{X,Y}(x, y)^\alpha} \right. \\ &\quad \times \left. \frac{2^{1-\alpha} \sum_{y \in \mathcal{Y}} \tilde{Q}_Y(y)^\alpha}{\sum_{y \in \mathcal{Y}} P_Y(y)^\alpha} \right\} \\ &= 1 + \frac{1}{1-\alpha} \log \frac{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} P_{X,Y}(x, y)^\alpha}{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} \tilde{Q}_{X,Y}(x, y)^\alpha}, \end{aligned}$$

and similarly

$$H_\alpha(X|Y) = \frac{1}{1-\alpha} \log \frac{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} P_{X,Y}(x, y)^\alpha}{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} \tilde{Q}_{X,Y}(x, y)^\alpha}.$$

For the uniform distribution case, assume that  $P_{X,Y}(0, y_i) = \frac{1}{2}\tilde{Q}_Y(y_i) + \delta_i, P_{X,Y}(1, y_i) = \frac{1}{2}\tilde{Q}_Y(y_i) - \delta_i$ , where  $\delta_i \ll \tilde{Q}_Y(y_i)$ . Denote  $|\mathcal{Y}| = M$ . Define

$$\Delta = \frac{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} P_{X,Y}(x, y)^\alpha}{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} \tilde{Q}_{X,Y}(x, y)^\alpha} - 1 \quad (25)$$

$$\approx \frac{2\alpha(\alpha-1) \sum_{i=1}^M \delta_i^2 \tilde{Q}_Y^{\alpha-2}(y_i)}{\sum_{i=1}^M \tilde{Q}_Y(y_i)^\alpha}. \quad (26)$$

As a simple example, further assume that  $\tilde{Q}_Y$  is also uniform. Then to ensure that  $H_\alpha(X|Y) \rightarrow 1$ ,

$$\alpha(\alpha-1)M^2\delta_i^2 \rightarrow 0$$

is required. As  $\alpha \rightarrow \infty$ , only truly uniform distribution yields  $H_\alpha(X|Y) = 1$ .

For the deterministic distribution case, without loss of generality, assume  $\tilde{Q}_{X,Y}(0, y_i) = 0, \tilde{Q}_{X,Y}(1, y_i) = \tilde{Q}_Y(y_i)$ , and  $P_{X,Y}(0, y_i) = \delta_i, P_{X,Y}(1, y_i) = \tilde{Q}_Y(y_i) - \delta_i$  for any  $y_i \in \mathcal{Y}$ , where  $\delta_i \ll \tilde{Q}_Y(y_i)$ . Denote  $|\mathcal{Y}| = M$ . Define

$$\Delta = \frac{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} P_{X,Y}(x, y)^\alpha}{\sum_{\{x,y\} \in \mathcal{X} \times \mathcal{Y}} \tilde{Q}_{X,Y}(x, y)^\alpha} - 1 \quad (27)$$

$$\approx \frac{\sum_{i=1}^M [\delta_i^\alpha - \alpha\delta_i \tilde{Q}_Y^{\alpha-1}(y_i)]}{\sum_{i=1}^M \tilde{Q}_Y(y_i)^\alpha}. \quad (28)$$

For  $0 < \alpha < 1$ , the difference between  $\delta_i$  and  $\tilde{Q}_Y(y_i)$  shrinks by the power- $\alpha$  operation. As  $\alpha$  approaches 0,  $\sum_{i=1}^M \delta_i^\alpha$  gains more influence on  $\Delta$ . When  $\alpha$  is small enough, we will have

$$\Delta \approx \frac{\sum_{i=1}^M \delta_i^\alpha}{\sum_{i=1}^M \tilde{Q}_Y(y_i)^\alpha}. \quad (29)$$

In the extreme case when  $\alpha = 0$  and  $\delta_i > 0$  for all  $i \in [M]$ , we get  $\Delta = 1$ , and  $H_\alpha(X|Y)$  always equals 1.  $\Delta$  equals 0 if and only if  $\delta_i = 0$  for all  $i \in [M]$ .

### IV. CONCLUDING REMARKS

This work has revealed the polarization phenomenon of conditional Rényi entropies. To apply the results to specific problems, much work has yet to be done. For example, estimating conditional Rényi entropies accurately can be a hard problem when  $N$  grows large, and existing approximation methods for polar code constructions may not be directly applied. For another example, the polarization rate of conditional Rényi entropies has not been touched in this paper. We will leave them for future research.

### REFERENCES

- [1] E. Arkan, "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] —, "Source polarization," in *2010 IEEE International Symposium on Information Theory*, 2010, pp. 899–903.
- [3] —, "Varentropy decreases under the polar transform," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3390–3400, 2016.
- [4] A. Rényi, "On measures of entropy and information," Hungarian Academy of Sciences Budapest Hungary, Tech. Rep., 1961.
- [5] M. Bloch and J. Barros, *Physical-layer security*. Cambridge University Press, 2011.
- [6] P. Jizba and T. Arimitsu, "The world according to Rényi: thermodynamics of multifractal systems," *Annals of Physics*, vol. 312, no. 1, pp. 17–59, 2004.
- [7] L. Golshani, E. Pasha, and G. Yari, "Some properties of Rényi entropy and Rényi entropy rate," *Information Sciences*, vol. 179, no. 14, pp. 2426–2433, 2009.
- [8] M. Alsam and E. Telatar, "Polarization improves  $E_0$ ," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2714–2719, 2014.
- [9] E. Arkan, "A note on polarization martingales," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 1466–1468.