Rényi Bounds on Information Combining

Christoph Hirche

QMATH, Department of Mathematical Sciences, University of Copenhagen

Universitetsparken 5, 2100 Copenhagen, Denmark.

Email: christoph.hirche@gmail.com

Abstract

Bounds on information combining are entropic inequalities that determine how the information, or entropy, of a set of random variables can change when they are combined in certain prescribed ways. Such bounds play an important role in information theory, particularly in coding and Shannon theory. The arguably most elementary kind of information combining is the addition of two binary random variables, a CNOT gate, and the resulting quantities play an important role in belief propagation and polar coding.

In this work we will generalize the concept to Rényi entropies. We give optimal bounds on the conditional Rényi entropy after combination, based on a certain convexity or concavity property and discuss when this property indeed holds. Since there is no generally agreed upon definition of the conditional Rényi entropy, we consider four different versions from the literature.

Finally, we discuss the application of the bounds to the polarization of Rényi entropies under polar codes.

I. Introduction

Many tasks in information theory are concerned with the evolution of random variables and their corresponding entropies under certain "combining operations". A particularly relevant example of such an operation is the addition of two independent random variables (with values in some group). In this case, the entropy can be easily computed since we know that the addition of two random variables has a probability distribution which corresponds to the convolution of the probability distributions of the individual random variables. The picture changes when we have random variables with *side information*. Now we are interested in the entropy of the sum conditioned on all the available side information. Evaluating this is substantially more difficult. The field of *bounds on information combining* is concerned with finding optimal entropic bounds on the resulting conditional entropy.

A particular important case is that of two binary random variables with side information of arbitrary finite dimension. This is the setting we will consider in this work as it has many applications e.g. in coding theory for polar codes [1] and Shannon theory [2], [3]. An optimal lower bound for the resulting Shannon entropy was given by Wyner and Ziv in [4], the well known *Mrs. Gerber's Lemma*. Following this result, additional approaches to the problem have been found which also led to an *upper* bound on the conditional entropy of the combined random variables. One proof method and several additional applications can be found e.g. in [3] along with the optimal upper bound.

In this work, we discuss the extension of information combining bounds to Rényi entropies. While the unconditional Rényi entropy is clearly defined, one can find many different definitions of the conditional Rényi entropy in the literature. We focus on four of the most commonly found conditional Rényi entropies, which we detail in the next section.

In all cases, we find the optimal lower and upper bounds given a certain convexity or concavity property of an associated function. In many cases we will give the ranges of α for which the desired property holds. Furthermore, we identify some remarkable cases where equality holds. Similar to the Shannon entropy case, our bounds are optimal in the sense that they are attained by the binary symmetric channel and the binary erasure channel with equality.

Finally, we briefly discuss the application of our results to channel polarization. We show that our results can easily be combined with a simplified proof strategy for polarization from [1] to show polarization of H_α^J under the polar coding transformation, therefore reproducing a result previously found in [5].

II. Background and previous results

In this section we will give a brief overview over the known results for the Shannon entropy. The main quantities under investigation are

$$H(X_1 + X_2 | Y_1 Y_2) \tag{1}$$

and

$$H(X_2 | X_1 + X_2, Y_1 Y_2), \tag{2}$$

which are related by

$$H(X_1 + X_2 | Y_1 Y_2) + H(X_2 | X_1 + X_2, Y_1 Y_2) = H(X_1 | Y_1) + H(X_2 | Y_2). \tag{3}$$

In classical information theory, the optimal bounds are well known as follows:

$$h_2(h_2^{-1}(H_1) * h_2^{-1}(H_2)) \leq H(X_1 + X_2 | Y_1 Y_2) \leq \log 2 - \frac{(\log 2 - H_1)(\log 2 - H_2)}{\log 2}, \tag{4}$$

where the lower bound is called the Mrs. Gerber's Lemma, and

$$\frac{H_1 H_2}{\log 2} \leq H(X_2|X_1 + X_2, Y_1 Y_2) \leq H_1 + H_2 - h_2(h_2^{-1}(H_1) * h_2^{-1}(H_2)), \tag{5}$$

with $H_1 = H(X_1|Y_1)$ and $H_2 = H(X_2|Y_2)$. In many situations, it is more intuitive to look at the special case where the two underlying entropies are equal ($H = H_1 = H_2$). In that case we can state the following inequalities

$$0.799 \frac{H(\log 2 - H)}{\log 2} + H \leq h_2(h_2^{-1}(H_1) * h_2^{-1}(H_2)) \tag{6}$$

$$\leq H(X_1 + X_2|Y_1 Y_2) \leq \frac{H(\log 2 - H)}{\log 2} + H, \tag{7}$$

where the first is an additional convenient lower bound from [6] and the other two follow from Equation 4.

It follows that it is sufficient to prove the inequalities for either Equation (4) or (5). As we will see, most Rényi generalization of conditional entropy do not obeys Equation (3). In the following we will focus on bounding the term in Equation (1) and leave bounds on that in (2) for future research.

Furthermore, considering the random variables X_i and Y_i as input and output of a channel respectively, it is even known for which channels equality is achieved in the above equations (see e.g. [3]). For the lower bound in Equation (4) this is the binary symmetric channel (BSC) and for the upper bound it is the binary erasure channel (BEC). For clarity, it will later be useful to refer to the generalized bounds as BSC and BEC bounds according to the channel that achieves equality.

III. CONDITIONAL RÉNYI ENTROPIES

The Rényi entropy $H_\alpha(X)$ is defined as

$$H_\alpha(X) := \frac{1}{1-\alpha} \log \sum_x p(x)^\alpha. \tag{8}$$

In analogy to the case of the Shannon entropy, we have

$$H_\alpha(X_1 + X_2) = h_\alpha(h_\alpha^{-1}(H_\alpha(X_1) \star h_\alpha^{-1}(H_\alpha(X_2)))). \tag{9}$$

While this definition is well established, the picture becomes much less clear when we consider the conditional Rényi entropy. In fact, a multitude of different definitions exists [7], each of which has found a number of applications. In the following we will consider the following definitions:

$$H_\alpha^A(X|Y) = \frac{\alpha}{1-\alpha} \log \left(\sum_y p(y) \left(\sum_x p(x|y)^\alpha \right)^{\frac{1}{\alpha}} \right), \tag{10}$$

$$H_\alpha^H(X|Y) = \frac{1}{1-\alpha} \log \left(\sum_y \sum_x p(y) p(x|y)^\alpha \right), \tag{11}$$

$$H_\alpha^J(X|Y) = \frac{1}{1-\alpha} \left[\log \sum_{x,y} p(x,y)^\alpha - \log \sum_y p(y)^\alpha \right], \tag{12}$$

$$H_\alpha^C(X|Y) = \frac{1}{1-\alpha} \sum_y p(y) \log \sum_x p(x|y)^\alpha, \tag{13}$$

where the first was given by Arimoto [8], the second by Hayashi and Skoric *et al.* [9], [10], the third by Jizba and Arimitsu [11] and the fourth by Cachin [12].

In the remainder of this section we will discuss some useful properties of the above definitions, see also [13], [14], [7]. We have

$$\lim_{\alpha \to 1} H_\alpha(X) = H(X) \tag{14}$$

and we would expect for every conditional Rényi entropy H_α^* that the following holds

$$\lim_{\alpha \to 1} H_\alpha^*(X|Y) = H(X|Y). \tag{15}$$

Two further important properties are monotonicity

$$H_\alpha^*(X|YZ) \leq H_\alpha^*(X|Z) \tag{16}$$

and the (strong) chain rule

$$H_\alpha^*(X|YZ) = H_\alpha^*(XY|Z) - H_\alpha^*(Y|Z), \tag{17}$$

TABLE I
TABLE OF SEVERAL PROPERTIES FOR THE CONSIDERED RÉNYI ENTROPIES.

$\lim_{\alpha \to 1}$	$\lim_{\alpha \to \infty}$	Chain rule	Monotonicity	*
✓ [8]	✓ [7], [14]	(weak) [7]	✓ [8]	H_α^A
✓	✗	✗ [7]	✓ [14]	H_α^H
✓	✗	✓	✗ [7]	H_α^J
✓	✗	✗ [12]	✗ [7]	H_α^C

both of which hold for the conditional Shannon entropy, but not necessarily for all Rényi generalizations.

Also, note that while conditional Rényi entropies generally are not even subadditive, they are additive for independent pairs of random variables, i.e.

$$H_\alpha^*(X_1 X_2 | Y_1 Y_2) = H_\alpha^*(X_1 | Y_1) + H_\alpha^*(X_2 | Y_2), \tag{18}$$

when (X_1, Y_1) and (X_2, Y_2) are independent of each other.

Finally, a commonly used quantity is the so-called *min-entropy*, defined as

$$H_\infty(X) = -\log \max_x p(x) \tag{19}$$

and its conditional version

$$H_\infty(X|Y) = -\log \sum_y p(y) \max_x p(x|y). \tag{20}$$

Ideally, we would have

$$\lim_{\alpha \to \infty} H_\alpha^*(X|Y) = H_\infty(X|Y). \tag{21}$$

Which of the above properties hold for the several conditional Rényi entropies considered is summarized in Table I.

In the remainder of this section, we will show how to generalize information combining bounds to the Rényi entropies discussed above. We will devote one section to each of the entropies, starting with Arimotos conditional entropy.

IV. RÉNYI BOUNDS ON INFORMATION COMBINING

A. Bounds for $H_\alpha^A(X_1 + X_2 | Y_1 Y_2)$

A particular feature of the Shannon entropy that allowed to prove the classical bounds, is that a conditional Shannon entropy can be written as convex combination of unconditioned Shannon entropies. This important feature does not hold for Rényi entropies in general. Nevertheless, for H_α^A we have the following equality, providing us with a similar tool:

$$H_\alpha^A(X|Y) = \frac{\alpha}{1-\alpha} \log \left(\sum_y p(y) e^{\frac{1-\alpha}{\alpha} H_\alpha^A(X|Y=y)} \right). \tag{22}$$

This motivates us to define the following quantity:

$$K_\alpha^A(X|Y) = e^{\frac{1-\alpha}{\alpha} H_\alpha^A(X|Y)}, \tag{23}$$

which takes values on $[1, \delta_\alpha^A := 2^{\frac{1-\alpha}{\alpha}}]$. For a more convenient notation we will furthermore use, in analogy to the binary entropy, $k_\alpha^A(p)$ whenever X is a binary random variable with probability distribution $\{p, 1-p\}$ (and trivial conditioning system): $K_\alpha^A(X) = k_\alpha^A(p)$.

We will see that the crucial quantity here is the following

$$\Bbbk_\alpha^A(x, y) = k_\alpha^A \left(k_\alpha^{A-1}(x) \star k_\alpha^{A-1}(y) \right). \tag{24}$$

Following the proof technique of the standard bounds on information combining, we get the following results.

Theorem IV.1 (H_α^A BSC-bound)**.** *If, for a given α, the function $\Bbbk_\alpha^A(x, y)$ is convex in x for fixed y and vice versa, then one of the two following equations holds:*
If $\alpha > 1$, then

$$H_\alpha^A(X_1 + X_2 | Y_1 Y_2) \leq h_\alpha(h_\alpha^{-1}(H_\alpha^A(X_1 | Y_1)) \star h_\alpha^{-1}(H_\alpha^A(X_2 | Y_2))). \tag{25}$$

If $\alpha < 1$, then

$$H_\alpha^A(X_1 + X_2 | Y_1 Y_2) \geq h_\alpha(h_\alpha^{-1}(H_\alpha^A(X_1 | Y_1)) \star h_\alpha^{-1}(H_\alpha^A(X_2 | Y_2))). \tag{26}$$

If $\Bbbk_\alpha^A(x,y)$ is concave instead, the inequalities hold with $\leq$ and $\geq$ exchanged. These bounds are optimal, in the sense that equality is achieved by binary symmetric channels.

Proof. To prove the theorem, we will show Equation 25 in the case of $\Bbbk_\alpha^A(x,y)$ being convex. All the other combinations, for 25 and 26, follow directly. The prove strategy is similar to the proof in the Shannon entropy setting.

Consider the following chain of equations:

$$H_\alpha^A(X_1+X_2|Y_1Y_2) = \frac{\alpha}{1-\alpha}\log\left(\sum_{y_1,y_2}p(y_1)p(y_2)e^{\frac{1-\alpha}{\alpha}H_\alpha^A(X_1+X_2|Y_1=y_1,Y_2=y_2)}\right) \tag{27}$$

$$= \frac{\alpha}{1-\alpha}\log\left(\sum_{y_1,y_2}p(y_1)p(y_2)e^{\frac{1-\alpha}{\alpha}h_\alpha(h_\alpha^{-1}(H_\alpha^A(X_1|Y_1=y_1))\star h_\alpha^{-1}(H_\alpha^A(X_2|Y_2=y_2)))}\right) \tag{28}$$

$$= \frac{\alpha}{1-\alpha}\log\left(\sum_{y_1,y_2}p(y_1)p(y_2)k_\alpha^A\left(k_\alpha^{A^{-1}}(K_\alpha^A(X_1|Y_1=y_1))\star k_\alpha^{A^{-1}}(K_\alpha^A(X_2|Y_2=y_2))\right)\right) \tag{29}$$

$$\leq \frac{\alpha}{1-\alpha}\log\left(k_\alpha^A\left(k_\alpha^{A^{-1}}(\sum_{y_1}p(y_1)K_\alpha^A(X_1|Y_1=y_1))\star k_\alpha^{A^{-1}}(\sum_{y_2}p(y_2)K_\alpha^A(X_2|Y_2=y_2))\right)\right) \tag{30}$$

$$= h_\alpha(h_\alpha^{-1}(H_\alpha^A(X_1|Y_1))\star h_\alpha^{-1}(H_\alpha^A(X_2|Y_2))), \tag{31}$$

where all equalities follow simply by definition and rearranging. The inequality follows by using the convexity of $\Bbbk_\alpha^A(x,y)$ twice, once in the first argument and once in the second. $\qquad\square$

This results gives an extension of the lower bound for Shannon entropy information combining. In a similar fashion, we can also generalize the upper bound:

Theorem IV.2 (H_α^A BEC-bound). *If, for a given α, the function $\Bbbk_\alpha^A(x,y)$ is convex in x for fixed y and vice versa, then one of the following equations holds:*
If $\alpha > 1$, then

$$H_\alpha^A(X_1+X_2|Y_1Y_2) \geq \frac{\alpha}{1-\alpha}\log\frac{(\delta_\alpha^A-K_\alpha^A(X_1|Y_1))(\delta_\alpha^A-K_\alpha^A(X_2|Y_2))}{1-\delta_\alpha^A}+\delta_\alpha^A. \tag{32}$$

If $\alpha < 1$, then

$$H_\alpha^A(X_1+X_2|Y_1Y_2) \leq \frac{\alpha}{1-\alpha}\log\frac{(\delta_\alpha^A-K_\alpha^A(X_1|Y_1))(\delta_\alpha^A-K_\alpha^A(X_2|Y_2))}{1-\delta_\alpha^A}+\delta_\alpha^A. \tag{33}$$

If $\Bbbk_\alpha^A(x,y)$ is concave instead, the inequalities hold with $\leq$ and $\geq$ exchanged. These bounds are optimal, in the sense that equality is achieved by binary erasure channels.

Proof. We will show Equation 32 in the case of $\Bbbk_\alpha^A(x,y)$ being convex. All the other statements follow similarly.

The crucial part of this proof is to find an upper bound on $\Bbbk_\alpha^A(x,y)$. Since we assume that $\Bbbk_\alpha^A(x,y)$ is convex, an upper bound is given by the line connecting the endpoints of graph. Let's for now consider y to be fixed. It can easily be seen that $\Bbbk_\alpha^A(\delta_\alpha^A,y)=\delta_\alpha^A$ and $\Bbbk_\alpha^A(1,y)=y$. Therefore we can give the bound:

$$\Bbbk_\alpha^A(x,y) \leq \frac{y-\delta_\alpha^A}{1-\delta_\alpha^A}x+\frac{\delta_\alpha^A(1-y)}{1-\delta_\alpha^A} \tag{34}$$

$$= \frac{(\delta_\alpha^A-x)(\delta_\alpha^A-y)}{1-\delta_\alpha^A}+\delta_\alpha^A. \tag{35}$$

With this, the proof follows via the standard arguments:

$$H_\alpha^A(X_1+X_2|Y_1Y_2) = \frac{\alpha}{1-\alpha}\log\left(\sum_{y_1,y_2}p(y_1)p(y_2)e^{\frac{1-\alpha}{\alpha}H_\alpha^A(X_1+X_2|Y_1=y_1,Y_2=y_2)}\right) \tag{36}$$

$$= \frac{\alpha}{1-\alpha}\log\left(\sum_{y_1,y_2}p(y_1)p(y_2)e^{\frac{1-\alpha}{\alpha}h_\alpha(h_\alpha^{-1}(H_\alpha^A(X_1|Y_1=y_1))\star h_\alpha^{-1}(H_\alpha^A(X_2|Y_2=y_2)))}\right) \tag{37}$$

$$= \frac{\alpha}{1-\alpha}\log\left(\sum_{y_1,y_2}p(y_1)p(y_2)k_\alpha^A\left(k_\alpha^{A^{-1}}(K_\alpha^A(X_1|Y_1=y_1))\star k_\alpha^{A^{-1}}(K_\alpha^A(X_2|Y_2=y_2))\right)\right) \tag{38}$$

$$\leq \frac{\alpha}{1-\alpha}\log\left(\sum_{y_1,y_2}p(y_1)p(y_2)\frac{(\delta_\alpha^A-K_\alpha^A(X_1|Y_1=y_1))(\delta_\alpha^A-K_\alpha^A(X_2|Y_2=y_2))}{1-\delta_\alpha^A}+\delta_\alpha^A\right) \tag{39}$$
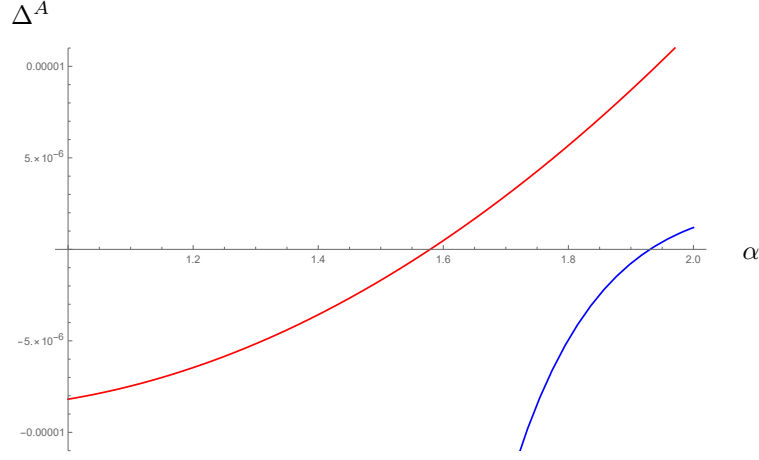
Fig. 1. The blue curve is $\Delta^A(BSC(10^{-3}), \alpha)$ and the red curve is $\Delta^A(BSC(0.46), \alpha)$. This shows that Δ^A can be positive or negative for the same value of α, which means ultimately that $\Bbbk^A$ cannot posses the desired convexity or concavity property.

$$= \frac{\alpha}{1-\alpha} \log \left(\frac{(\delta_\alpha^A - K_\alpha^A(X_1|Y_1))(\delta_\alpha^A - K_\alpha^A(X_2|Y_2))}{1 - \delta_\alpha^A} + \delta_\alpha^A \right) \tag{40}$$

□

Now, the crucial question is: When does the function $\Bbbk_\alpha^A(x, y)$ posses the desired convexity or concavity property?

One important result in this direction can be found in [15], which handles a different, but related, problem, and we rephrase it here for our convenience.

Lemma IV.3 (Convexity result from [15].). *For $\alpha \geq 2$, the function*

$$k_\alpha^A(k_\alpha^{A^{-1}}(x) \star k_\alpha^{A^{-1}}(y)) \tag{41}$$

is convex in x for fixed y and vice versa.

So, is the function always convex? It turns out, that this is not the case. In the following lemma we show by concrete examples that the function is neither convex nor concave for some values of α.

Lemma IV.4 (Counterexample for $\Bbbk_\alpha^A(x, y)$). *For $\alpha \in (1.58, 1.97)$, the function*

$$k_\alpha^A(k_\alpha^{A^{-1}}(x) \star k_\alpha^{A^{-1}}(y)) \tag{42}$$

is neither convex nor concave in x for fixed y and vice versa.

Proof. Remember that the convexity or concavity of $\Bbbk_\alpha^A(x, y)$ would directly imply the validity of the BSC and BEC bounds for all channels for that value of α. This means also that there would be a natural order between the two, since one is an upper bound and one a lower bound, again for all channels. In the following, we will show that for some values of α this order relation does not exist, i.e. for some channels the order is opposite to what it is for other channels. Therefore, for this values of α the function could not have been convex or concave.

To this end, let's consider both bounds for a binary symmetric channel with crossover probability p, denoted $BSC(p)$. We have

$$H_\alpha^A(BSC(p)) = h_\alpha(p) \tag{43}$$

and therefore the BSC bound for two of these channels is given by $h_\alpha(p \star p)$. Evaluating the BEC bound for two $BSC(p)$ and defining the difference between both as

$$\Delta^A(BSC(p), \alpha) := h(p \star p) - \frac{\alpha}{1-\alpha} \log \frac{(\delta_\alpha^A - K_\alpha^A(BSC(p)))^2}{1 - \delta_\alpha^A} + \delta_\alpha^A \tag{44}$$

gives us an easily testable quantity. We simply verify numerically that

$$\Delta^A(BSC(10^{-6}), \alpha) < 0 \quad \text{for } \alpha \in (1, 1.97), \tag{45}$$

$$\Delta^A(BSC(0.49), \alpha) > 0 \quad \text{for } \alpha \in (1.58, 2). \tag{46}$$

By the above argument this leads to the desired result. The behavior is illustrated in Figure 1 for similar values of α. □

Remark IV.5. *The values in Lemma IV.4 are those that we could conclusively verify. Furthermore, additional numerics suggests that the same methods should provide counterexamples in the range $\alpha \in (1.5783, 2)$ when evaluated with sufficiently high precision.*

Additionally, the numerical evidence leads us to making the following conjecture for the Arimoto conditional entropy.

Conjecture IV.6. *There exists a value $1 < \hat{\alpha} < 1.5783$, such that*

$$k_\alpha^A(k_\alpha^{A^{-1}}(x) \star k_\alpha^{A^{-1}}(y)) \tag{47}$$

is convex for $0 < \alpha < 1$ and concave for $1 < \alpha \leq \hat{\alpha}$.

Finally, remember that

$$\lim_{\alpha \to \infty} H_\alpha^A(X|Y) = H_\infty(X|Y). \tag{48}$$

This value of α shows a particular behavior and deserves special attention.

Lemma IV.7. *We have that $\mathbb{k}_\infty^A(x, y)$ is independently linear in x and y and therefore*

$$H_\infty^A(X_1 + X_2|Y_1 Y_2) = h_\infty(h_\infty^{-1}(H_\infty^A(X_1|Y_1)) \star h_\infty^{-1}(H_\infty^A(X_2|Y_2))). \tag{49}$$

Proof. Remember from the definition, for binary X

$$k_\infty^A(p) = \max\{p, 1 - p\}. \tag{50}$$

Take $x \in [\frac{1}{2}, 1]$ and fix $c \geq \frac{1}{2}$. Now, $k_\infty^{A^{-1}}(x)$ can be either $x \geq \frac{1}{2}$ or $1 - x \leq \frac{1}{2}$; in the former case $x \star c \leq \frac{1}{2}$ and therefore $k_\infty^A(x \star c) = 1 - x \star c$, in the latter case $(1-x) \star c \geq \frac{1}{2}$ and $k_\infty^A((1-x) \star c) = (1-x) \star c = 1 - x \star c$. Therefore, the function is identical and linear in both cases. Similarly one can show that when $c \leq \frac{1}{2}$ then the function simply becomes $x \star c$ which is again linear. $\square$

The last lemma is remarkable as it reproduces an equality that we usually only have for unconditioned entropies.

B. Bounds for $H_\alpha^H(X_1 + X_2|Y_1 Y_2)$

Similar to H_α^A we get the following equality for H_α^H:

$$H_\alpha^H(X|Y) = \frac{1}{1 - \alpha} \log \left(\sum_y p(y) e^{(1-\alpha)H_\alpha^H(X|Y=y)} \right). \tag{51}$$

This motivates us to define the following quantity,

$$K_\alpha^H(X|Y) = e^{(1-\alpha)H_\alpha^H(X|Y)}, \tag{52}$$

which takes values on $[1, \delta_\alpha^H := 2^{1-\alpha}]$. For a more convenient notation we will furthermore use, in analogy to the binary entropy, $k_\alpha^H(p)$ whenever X is a binary random variable with probability distribution $\{p, 1 - p\}$ (and trivial conditioning system): $K_\alpha^H(X) = k_\alpha^H(p)$.

We will see that the crucial quantity here is the following

$$\mathbb{k}_\alpha^H(x, y) = k_\alpha^H \left(k_\alpha^{H^{-1}}(x) \star k_\alpha^{H^{-1}}(y) \right). \tag{53}$$

Following the proof technique of the standard bounds on information combining, we get the following results.

Theorem IV.8 (H_α^H BSC-bound). *If, for a given α, the function $\mathbb{k}_\alpha^H(x, y)$ is convex in x for fixed y and vice versa, then one of the two following equations holds:*
If $\alpha > 1$, then

$$H_\alpha^H(X_1 + X_2|Y_1 Y_2) \leq h_\alpha(h_\alpha^{-1}(H_\alpha^H(X_1|Y_1)) \star h_\alpha^{-1}(H_\alpha^H(X_2|Y_2))). \tag{54}$$

If $\alpha < 1$, then

$$H_\alpha^H(X_1 + X_2|Y_1 Y_2) \geq h_\alpha(h_\alpha^{-1}(H_\alpha^H(X_1|Y_1)) \star h_\alpha^{-1}(H_\alpha^H(X_2|Y_2))). \tag{55}$$

If $\mathbb{k}_\alpha^H(x, y)$ is concave instead, the inequalities hold with $\leq$ and $\geq$ exchanged. These bounds are optimal, in the sense that equality is achieved by binary symmetric channels.

Proof. The proof works analogous to that of Theorem IV.1. $\square$

Theorem IV.9 (H_α^H BEC-bound). *If, for a given α, the function $\Bbbk_\alpha^H(x,y)$ is convex in x for fixed y and vice versa, then one of the following equations holds:*
If $\alpha > 1$, then

$$H_\alpha^H(X_1 + X_2|Y_1 Y_2) \geq \frac{1}{1-\alpha} \log \frac{(\delta_\alpha^H - K_\alpha^H(X_1|Y_1))(\delta_\alpha^H - K_\alpha^H(X_2|Y_2))}{1 - \delta_\alpha^H} + \delta_\alpha^H. \tag{56}$$

If $\alpha < 1$, then

$$H_\alpha^H(X_1 + X_2|Y_1 Y_2) \leq \frac{1}{1-\alpha} \log \frac{(\delta_\alpha^H - K_\alpha^H(X_1|Y_1))(\delta_\alpha^H - K_\alpha^H(X_2|Y_2))}{1 - \delta_\alpha^H} + \delta_\alpha^H. \tag{57}$$

If $\Bbbk_\alpha^H(x,y)$ is concave instead, the inequalities hold with $\leq$ and $\geq$ exchanged. These bounds are optimal, in the sense that equality is achieved by binary erasure channels.

Proof. The proof works analogous to that of Theorem IV.2. $\qquad\square$

Let's now consider the case of $\Bbbk_\alpha^H$.

Lemma IV.10. *The function*

$$k_\alpha^H(k_\alpha^{H^{-1}}(x) \star k_\alpha^{H^{-1}}(y)) \tag{58}$$

is convex for $0 < \alpha < 1$ and $2 < \alpha \leq 3$ and concave for $1 < \alpha \leq 2$ and $\alpha \geq 3$.

Proof. The proof is part of Appendix A. $\qquad\square$

Although technically included in the previous lemma, a particular case deserves some special attention.

Lemma IV.11. *The function*

$$k_\alpha^H(k_\alpha^{H^{-1}}(x) \star k_\alpha^{H^{-1}}(y)) \tag{59}$$

is linear in x and y for $\alpha = 2$ and $\alpha = 3$.

Proof. Checked by explicit calculation, see Appendix A. $\qquad\square$

This lemma is remarkable as it tells us that the BSC-bound and the BEC-bound both hold with equality. We have

$$H_2^H(X_1 + X_2|Y_1 Y_2) = h_2(h_2^{-1}(H_2^H(X_1|Y_1)) \star h_2^{-1}(H_2^H(X_2|Y_2))) \tag{60}$$

$$H_3^H(X_1 + X_2|Y_1 Y_2) = h_3(h_3^{-1}(H_3^H(X_1|Y_1)) \star h_3^{-1}(H_3^H(X_2|Y_2))). \tag{61}$$

These equations are remarkable as they give an equality in the conditional case, something we usually only get for unconditioned entropies.

To illustrate the results, we define in analogy to the technique used in Lemma IV.4 the quantity

$$\Delta^H(BSC(p), \alpha) := h(p \star p) - \frac{1}{1-\alpha} \log \frac{(\delta_\alpha^H - K_\alpha^H(BSC(p)))^2}{1 - \delta_\alpha^H} + \delta_\alpha^H, \tag{62}$$

which gives the difference between the BSC and the BEC bound evaluated on a binary symmetric channel with crossover probability p and we plot it in Figure 2.

C. Bounds for $H_\alpha^J(X_1 + X_2|Y_1 Y_2)$

From the definition, we have

$$H_\alpha^J(X|Y) = \frac{1}{1-\alpha} \left[\log \sum_y p(y)^\alpha \sum_x p(x|y)^\alpha - \log \sum_y p(y)^\alpha \right] \tag{63}$$

$$= \frac{1}{1-\alpha} \left[\log \sum_y p(y)^\alpha e^{(1-\alpha)H_\alpha(X|Y=y)} - \log \sum_y p(y)^\alpha \right] \tag{64}$$

This suggests that, again, the quantity

$$K_\alpha^H(X) = e^{(1-\alpha)H_\alpha(X)}. \tag{65}$$

is of crucial interest here. However, due to the exponent of $p(y)$ we can not directly invoke convexity, but have to play one extra trick compared to the proof of the previous bounds. For the conditional version we will also need

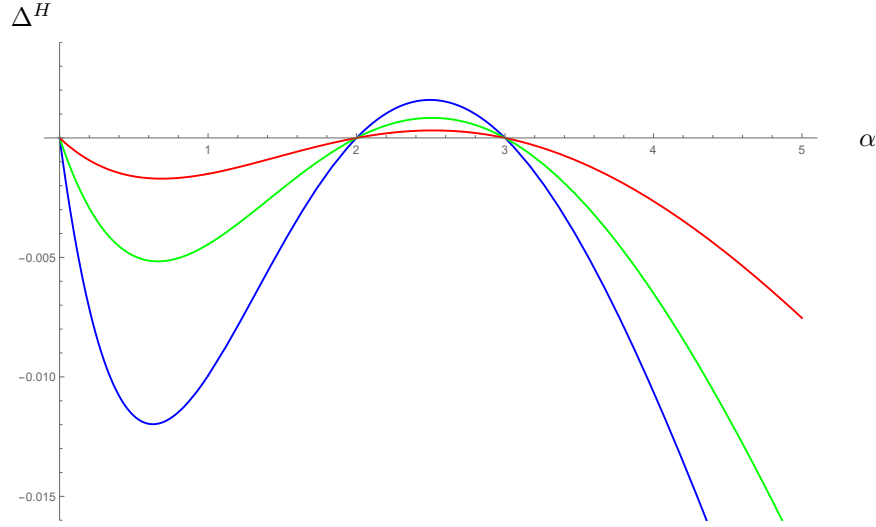$$K_\alpha^J(X|Y) = e^{(1-\alpha)H_\alpha^J(X|Y)}. \tag{66}$$

Fig. 2. The different curves show $\Delta^H(BSC(p), \alpha)$ for different values of p. Blue is $p = 0.25$, green is $p = 0.3$ and red is $p = 0.35$. One can clearly observe how the BSC and the BEC bound switch their respective roles as upper and lower bound and that both are actually identical for $\alpha = 2$ and $\alpha = 3$.

Theorem IV.12 (H_α^J BSC-bound)**.** *If, for a given α, the function $\Bbbk_\alpha^H(x, y)$ is convex in x for fixed y and vice versa, then one of the two following equations holds:*
If $\alpha > 1$, then

$$H_\alpha^J(X_1 + X_2|Y_1 Y_2) \leq h_\alpha(h_\alpha^{-1}(H_\alpha^J(X_1|Y_1)) \star h_\alpha^{-1}(H_\alpha^J(X_2|Y_2))). \tag{67}$$

If $\alpha < 1$, then

$$H_\alpha^J(X_1 + X_2|Y_1 Y_2) \geq h_\alpha(h_\alpha^{-1}(H_\alpha^J(X_1|Y_1)) \star h_\alpha^{-1}(H_\alpha^J(X_2|Y_2))). \tag{68}$$

If $\Bbbk_\alpha^H(x, y)$ is concave instead, the inequalities hold with $\leq$ and $\geq$ exchanged. These bounds are optimal, in the sense that equality is achieved by binary symmetric channels.

Proof. To prove the theorem, we will show Equation 67 in the case of $\Bbbk_\alpha^A(x, y)$ being convex. All the other combinations, for 67 and 68, follow directly. The prove strategy is similar to the proof in the Shannon entropy setting.

To make use of the convexity property, define the following probability distribution (also sometimes called the *tilted probability distribution*):

$$\tilde{p}(y) := \frac{p(y)^\alpha}{\sum_y p(y)^\alpha}. \tag{69}$$

Consider the following chain of equations:

$$H_\alpha^J(X_1 + X_2|Y_1 Y_2) = \frac{1}{1-\alpha}\left[\log \sum_{y_1, y_2} p(y_1)^\alpha p(y_2)^\alpha e^{(1-\alpha)H_\alpha(X_1+X_2|Y_1=y_1, Y_2=y_2)} - \log \sum_{y_1, y_2} p(y_1)^\alpha p(y_2)^\alpha\right] \tag{70}$$

$$= \frac{1}{1-\alpha}\left[\log \sum_{y_1, y_2} \tilde{p}(y_1)\tilde{p}(y_2) e^{(1-\alpha)H_\alpha(X_1+X_2|Y_1=y_1, Y_2=y_2)}\right] \tag{71}$$

$$= \frac{\alpha}{1-\alpha}\log\left(\sum_{y_1, y_2} \tilde{p}(y_1)\tilde{p}(y_2) k_\alpha^H\left(k_\alpha^{H-1}(K_\alpha^H(X_1|Y_1=y_1)) \star k_\alpha^{H-1}(K_\alpha^H(X_2|Y_2=y_2))\right)\right) \tag{72}$$

$$\leq \frac{\alpha}{1-\alpha}\log\left(k_\alpha^H\left(k_\alpha^{H-1}\left(\sum_{y_1} \tilde{p}(y_1) K_\alpha^H(X_1|Y_1=y_1)\right) \star k_\alpha^{H-1}\left(\sum_{y_2} \tilde{p}(y_2) K_\alpha^H(X_2|Y_2=y_2)\right)\right)\right) \tag{73}$$

$$= h_\alpha(h_\alpha^{-1}(H_\alpha^J(X_1|Y_1)) \star h_\alpha^{-1}(H_\alpha^J(X_2|Y_2))), \tag{74}$$

where all equalities follow simply by definition and rearranging. The inequality follows by using the convexity of $\Bbbk_\alpha^H(x, y)$ twice, once in the first argument and once in the second. $\square$

Theorem IV.13 (H_α^J BEC-bound). *If, for a given α, the function $\Bbbk_\alpha^H(x,y)$ is convex in x for fixed y and vice versa, then one of the following equations holds:*
If $\alpha > 1$, then

$$H_\alpha^J(X_1 + X_2|Y_1Y_2) \geq \frac{1}{1-\alpha} \log \frac{(\delta_\alpha^H - K_\alpha^J(X_1|Y_1))(\delta_\alpha^H - K_\alpha^J(X_2|Y_2))}{1 - \delta_\alpha^H} + \delta_\alpha^H. \tag{75}$$

If $\alpha < 1$, then

$$H_\alpha^J(X_1 + X_2|Y_1Y_2) \leq \frac{1}{1-\alpha} \log \frac{(\delta_\alpha^H - K_\alpha^J(X_1|Y_1))(\delta_\alpha^H - K_\alpha^J(X_2|Y_2))}{1 - \delta_\alpha^H} + \delta_\alpha^H. \tag{76}$$

If $\Bbbk_\alpha^H(x,y)$ is concave instead, the inequalities hold with $\leq$ and $\geq$ exchanged. These bounds are optimal, in the sense that equality is achieved by binary erasure channels.

Proof. The proof works analogous to that of Theorem IV.2 using the tilted probability distribution from the proof of Theorem IV.12. $\square$

Now the exact behavior of the bounds follows, as it did for $H_\alpha^H(X_1 + X_2|Y_1Y_2)$, from Lemma IV.10.

D. Bounds for $H_\alpha^C(X_1 + X_2|Y_1Y_2)$

Finally, we will investigate the properties of H_α^C. On first glance, this might appear simpler then the other quantities since the sum over y is outside the logarithm, which means that

$$H_\alpha^C(X|Y) = \sum_y p(y) H_\alpha(X|Y = y) \tag{77}$$

and therefore the quantity to investigate will be

$$\hbar_\alpha^A(x,y) = h_\alpha\left(h_\alpha^{-1}(x) \star h_\alpha^{-1}(y)\right). \tag{78}$$

However, it will turn out that proving convexity or concavity of this quantity is actually more complicated. Nevertheless, we will provide some partial results later in this section. First, let us state the BSC and BEC bound for this conditional entropy.

Theorem IV.14 (H_α^C BSC-bound). *If, for a given α, the function $\hbar_\alpha(x,y)$ is convex in x for fixed y and vice versa, then one of the two following equations holds:*
If $\alpha > 1$, then

$$H_\alpha^C(X_1 + X_2|Y_1Y_2) \leq h_\alpha(h_\alpha^{-1}(H_\alpha^C(X_1|Y_1)) \star h_\alpha^{-1}(H_\alpha^C(X_2|Y_2))). \tag{79}$$

If $\alpha < 1$, then

$$H_\alpha^C(X_1 + X_2|Y_1Y_2) \geq h_\alpha(h_\alpha^{-1}(H_\alpha^C(X_1|Y_1)) \star h_\alpha^{-1}(H_\alpha^C(X_2|Y_2))). \tag{80}$$

If $\hbar_\alpha(x,y)$ is concave instead, the inequalities hold with $\leq$ and $\geq$ exchanged. These bounds are optimal, in the sense that equality is achieved by binary symmetric channels.

Proof. The proof works analogous to that of Theorem IV.1. $\square$

Theorem IV.15 (H_α^C BEC-bound). *If, for a given α, the function $\hbar_\alpha(x,y)$ is convex in x for fixed y and vice versa, then one of the following equations holds:*
If $\alpha > 1$, then

$$H_\alpha^C(X_1 + X_2|Y_1Y_2) \geq \log 2 - \frac{(\log 2 - H_\alpha^C(X_1|Y_1))(\log 2 - H_\alpha^C(X_2|Y_2))}{\log 2}. \tag{81}$$

If $\alpha < 1$, then

$$H_\alpha^C(X_1 + X_2|Y_1Y_2) \leq \log 2 - \frac{(\log 2 - H_\alpha^C(X_1|Y_1))(\log 2 - H_\alpha^C(X_2|Y_2))}{\log 2}. \tag{82}$$

If $\hbar_\alpha(x,y)$ is concave instead, the inequalities hold with $\leq$ and $\geq$ exchanged. These bounds are optimal, in the sense that equality is achieved by binary erasure channels.

Proof. The proof works analogous to that of Theorem IV.2. $\square$

Lemma IV.16 (Counterexample for $\hbar_\alpha(x,y)$). *For $\alpha \in (1.387, 1.95)$, the function*

$$h_\alpha(h_\alpha^{-1}(x) \star h_\alpha^{-1}(y)) \tag{83}$$

is neither convex nor concave in x for fixed y and vice versa.

Proof. We will use the same strategy as for the counterexamples found in Lemma IV.4

For a binary symmetric channel with crossover probability p, denoted $BSC(p)$, we have

$$H_\alpha^C(BSC(p)) = h_\alpha(p) \tag{84}$$

and therefore the BSC bound for two of these channels is given by $h_\alpha(p \star p)$. Evaluating the BEC bound for two $BSC(p)$ and defining the difference between both as

$$\Delta^C(BSC(p), \alpha) := h(p \star p) - \log 2 + \frac{(\log 2 - h_\alpha(p))^2}{\log 2} \tag{85}$$

gives us an easily testable quantity. We simply verify numerically that

$$\Delta^C(BSC(10^{-7}), \alpha) < 0 \quad \text{for } \alpha \in (1, 1.95), \tag{86}$$

$$\Delta^C(BSC(0.49), \alpha) > 0 \quad \text{for } \alpha \in (1.387, 2). \tag{87}$$

By the above argument this leads to the desired result. $\square$

Remark IV.17. *Similar to remark IV.5, also here, numerics suggests that the same methods could provide counterexamples in the range $\alpha \in (1.3863, 2)$ when evaluated with sufficiently high computer precision.*

V. Polarization under H_α^J

Polar codes have gained a lot of attention recently as the first efficiently implementable codes that provable reach capacity. The underlying technique that enables Polar codes is that of polarization, allowing to transmit information with a rate equal to the capacity and vanishing error in the theoretical limit of infinite block length. In the original work by Arikan polarization was proven for the Shannon entropy using a Martingale approach. Later, Alsan and Teletar [1] gave a conceptually simpler proof using information combining bounds, in particular the Mrs. Gerbers Lemma. Very recently, polarization was shown to also happen when considering the H_α^J Rényi entropy, again via a Martingale approach [5]. As the main result of this section we show that our Rényi information combining bounds from the last section can be used to give a simpler and more intuitive proof following the technique established in [1].

Let $I(W)$ be the symmetric capacity of a channel as expressed by its (Shannon) mutual information. In its operational interpretation as the channel capacity the mutual information is uniquely defined and it is known to exhibit polarization under Arikans channel transformation. However, it appears natural to ask whether polarization also occurs when considering a different quantity.

Let W^+ and W^- denote the channels resulting from transforming two copies of the channel W. By carefully going through the simpler polarization proof in [1] one notices that essentially only two properties of the symmetric capacity are needed:

$$I(W^+) + I(W^-) = 2I(W) \tag{88}$$

$$\Delta(W) := \frac{1}{2}[I(W^+) - I(W^-)] \geq \kappa(a, b), \tag{89}$$

where for the inequality $I(W) \in [a, b]$ and $\kappa(a, b) > 0$ whenever $0 < a < b < 1$.

Now lets consider a alternative quantity $I^*(W)$, one can easily generalize the main result in [1] to be formulated in the following way.

Theorem V.1. *Let $I^*(\cdot)$ be a channel mutual information that fulfills the conditions in Equations 88 and 89. For any binary input channel W, and any $0 < a < b < 1$, we have*

$$\lim_{n \to \infty} \frac{1}{2^n} \#\{s^n \in \{+, -\}^n : I^*(W^{s^n}) \in [0, a)\} = 1 - I^*(W), \tag{90}$$

$$\lim_{n \to \infty} \frac{1}{2^n} \#\{s^n \in \{+, -\}^n : I^*(W^{s^n}) \in [a, b]\} = 0, \tag{91}$$

$$\lim_{n \to \infty} \frac{1}{2^n} \#\{s^n \in \{+, -\}^n : I^*(W^{s^n}) \in (b, 1]\} = I^*(W), \tag{92}$$

Proof. As described before, the proof follows immediately by following the proof of [1, Theorem 1]. $\square$

We are just left with showing that the above can be applied to a mutual information based on H_α^J. Let us take X and Y as the random variables describing the input and the output of the channel W, respectively, and define

$$I_\alpha^J(W) := H_\alpha(X) - H_\alpha^J(X|Y) \tag{93}$$

$$= \frac{1}{1 - \alpha} \left[\log \sum_{x,y} p(x, y)^\alpha - \log \sum_x p(x)^\alpha - \log \sum_y p(y)^\alpha \right].$$

Now, its easy to see that the definition obeys the chain rule the same way H_α^J does and therefore a standard proof applies to show that $I_\alpha^J(W)$ obeys the condition in Equation 88. Additionally, it is rather easy to see that our results in the previous section ensure that $I_\alpha^J(W)$ also obeys the condition in Equation 89 (compare e.g. [1, Lemma 1]. Therefore it follows directly that $I_\alpha^J(W)$ exhibits polarization and, since $I_\alpha^J(W) = \log 2 - H_\alpha^J(X|Y)$ for binary symmetric channels, so does H_α^J, which is the main result of [5].

We finish the section with two remark. First, in [1] the authors also consider the case where the channel transformation is applied to non identical channels. We remark that following the same ideas as above the result can easily be generalized to hold for generalized $I^*(W)$ that fulfill the corresponding generalizations of Equations 88 and 89. This holds in particular for $I_\alpha^J(W)$ and therefore also non-stationary channels polarize when considering this measure.

Second, since in the Shannon information combining bounds the equality in the lower bound is achieved by the binary symmetric channel, the BSC seems to be the channel that is the most difficult to polarize. However, when considering H_α^J, for some values of α the lower bound is satisfied by the binary erasure channel, suggesting that the roles of the channels are inverted when it comes to polarization.

VI. Conclusions

We have shown several extensions of information combining bounds to Rényi entropies. In general the results require a convexity or concavity property, we have however been able to show such a property for several settings. Additionally, we discussed a simple application of our results to the polarization of Rényi entropies.

Due to the number of possible definitions of conditional Rényi entropies we decided to focus on four of them. Primarily, they were chosen since all of them have found several applications in information theory and are therefore of practical relevance. Additionally, they all share the property that taking the limit $\alpha \to 1$ gives the Shannon entropy case. The latter can be useful as it gives an alternative approach towards proofing this special case. While this doesn't seem of particular interest in the traditional setting, recently the extension of information combining bounds to the realm of quantum information theory has been discussed [16], where proving optimal bounds remains an open problem, even for the von Neumann entropy which is the natural generalization of the Shannon entropy to the quantum setting. Proving a conjectured lower bound would have immediate applications investigating Polar codes for classical-quantum channels [16], [17], [18]. A potential strategy towards solving the quantum case, could be to look at quantum Rényi entropies and take the limit to the von Neumann entropy. In particular, H_α^A and H_α^H are special cases of the most commonly used quantum Rényi entropies [19]. Our results can therefore be seen as partial progress towards a new approach to solve the quantum case.

Finally, this work leaves some natural open problems. The most obvious one being to determine whether the desired convexity or concavity properties hold in those ranges of α where it is not yet known. Additionally, it remains open whether one can find different bounds that hold also where the convexity or concavity does not hold. Lastly, one might also consider entropies other than the Rényi entropies. Particularly interesting ones would again be those that contain the Shannon entropy as a special case.

Acknowledgments

References

[1] M. Alsan and E. Teletar, "A simple proof of polarization and polarization for non-stationary channels," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 301 – 305, July 2014.

[2] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. New York, U. S. A.: Cambridge University Press, January 2012.

[3] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York, NY, USA: Cambridge University Press, 2008.

[4] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications–i," *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 769 – 772, November 1973.

[5] M. Zheng, L. Liu, and C. Ling, "On the polarization of rényi entropy," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 2094–2098.

[6] V. Guruswami and P. Xia, "Polar codes: Speed of polarization and polynomial gap to capacity," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 3–16, January 2015, arXiv:1304.4321.

[7] S. Fehr and S. Berens, "On the conditional renyi entropy," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6801–6810, Nov 2014.

[8] S. Arimoto, "Information measures and capacity of order a for discrete memoryless channels," *Topics in Information Theory (Colloquia Mathematica Societatis Janos Bolyai)*, vol. 16, pp. 41–52, 1977.

[9] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3989–4001, 2011.

[10] B. Škorić, C. Obi, E. Verbitskiy, and B. Schoenmakers, "Sharp lower bounds on the extractable randomness from non-uniform sources," *Information and Computation*, vol. 209, no. 8, pp. 1184–1196, 2011.

[11] 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.

[12] C. Cachin, "Entropy measures and unconditional security in cryptography," Ph.D. dissertation, ETH Zurich, 1997.

[13] A. Teixeira, A. Matos, and L. Antunes, "Conditional rényi entropies," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4273–4277, 2012.

[14] M. Iwamoto and J. Shikata, "Information theoretic security for encryption based on conditional rényi entropies," in *International Conference on Information Theoretic Security*. Springer, 2013, pp. 103–121.

[15] H. Hsu, S. Asoodeh, S. Salamatian, and F. P. Calmon, "Generalizing bottleneck problems," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 531–535.

[16] C. Hirche and D. Reeb, "Bounds on information combining with quantum side information," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4739–4757, 2018.

[17] M. M. Wilde and S. Guha, "Polar codes for classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 1175–1187, February 2013, arXiv:1109.2591.

[18] C. Hirche, "Polar codes in quantum information theory," 2014, Master's thesis, Hannover, arXiv:1501.03737.

[19] M. Tomamichel, *Quantum Information Processing with Finite Resources: Mathematical Foundations*, ser. SpringerBriefs in Mathematical Physics. Springer, 2015, vol. 5.

APPENDIX

A. Proof of Lemma IV.10 and Lemma IV.11

Our goal is to investigate properties of the function

$$\Bbbk_\alpha^H(x) = k_\alpha^H(k_\alpha^{H^{-1}}(x) \star c), \tag{94}$$

in particular concavity, convexity or linearity in x. The results in Lemma IV.10 and Lemma IV.11 then follow by the symmetry of $\Bbbk_\alpha^H(x, y)$ under exchanging x and y.

Recall that a function $f(x)$ is convex iff

$$f(y) \geq f(x) + f'(x)(y - x) \tag{95}$$

for all x, y. Concavity holds if above inequality holds as $\leq$ and the function is linear if it becomes an equality.

It can easily be seen that

$$\partial_x \Bbbk_\alpha^H(x) = \Bbbk_\alpha^H(x) \frac{1 - 2c}{x} \frac{h'_\alpha(k_\alpha^{H^{-1}}(x) \star c)}{h'_\alpha(k_\alpha^{H^{-1}}(x))} . \tag{96}$$

With some rewriting Equation (95) becomes

$$\frac{\Bbbk_\alpha^H(y)}{\Bbbk_\alpha^H(x)} - \frac{y}{x}(1 - 2c) \frac{h'_\alpha(k_\alpha^{H^{-1}}(x) \star c)}{h'_\alpha(k_\alpha^{H^{-1}}(x))} \geq 1 - (1 - 2c) \frac{h'_\alpha(k_\alpha^{H^{-1}}(x) \star c)}{h'_\alpha(k_\alpha^{H^{-1}}(x))} . \tag{97}$$

Since this has to hold for all x and y in the parameter range we can do the re-parametrization $x \to k_\alpha^H(x)$ and $y \to k_\alpha^H(y)$, leading to

$$\frac{k_\alpha^H(y \star c)}{k_\alpha^H(x \star c)} - \frac{k_\alpha^H(y)}{k_\alpha^H(x)}(1 - 2c) \frac{h'_\alpha(x \star c)}{h'_\alpha(x)} \geq 1 - (1 - 2c) \frac{h'_\alpha(x \star c)}{h'_\alpha(x)} . \tag{98}$$

Let's define the function

$$g_\alpha^x(y) := \frac{k_\alpha^H(y \star c)}{k_\alpha^H(x \star c)} - \frac{k_\alpha^H(y)}{k_\alpha^H(x)}(1 - 2c) \frac{h'_\alpha(x \star c)}{h'_\alpha(x)} . \tag{99}$$

Now it can be easily seen that the inequality reduces to

$$g_\alpha^x(y) \geq g_\alpha^x(x) . \tag{100}$$

The original convexity/concavity problem is therefore reduced to whether the function $g_\alpha^x(y)$ has a minimum or maximum at x.

To simplify the expression a bit more we consider the following simple calculation,

$$h'_\alpha(x) = \frac{\alpha}{1 - \alpha} \frac{x^{\alpha - 1} - (1 - x)^{\alpha - 1}}{k_\alpha^H(x)} . \tag{101}$$

We find,

$$g_\alpha^x(y) = \frac{1}{k_\alpha^H(x \star c)} \left[k_\alpha^H(y \star c) - k_\alpha^H(y)(1 - 2c) \frac{(x \star c)^{\alpha - 1} - (1 - x \star c)^{\alpha - 1}}{x^{\alpha - 1} - (1 - x)^{\alpha - 1}} \right] . \tag{102}$$

As a start, we consider two special cases that will turn out to have an interesting property: Consider $\alpha = 2$ and $\alpha = 3$. First observe that interestingly,

$$\frac{(x \star c)^{\alpha - 1} - (1 - x \star c)^{\alpha - 1}}{x^{\alpha - 1} - (1 - x)^{\alpha - 1}} \bigg|_{\alpha=2} = \frac{(x \star c)^{\alpha - 1} - (1 - x \star c)^{\alpha - 1}}{x^{\alpha - 1} - (1 - x)^{\alpha - 1}} \bigg|_{\alpha=3} = (1 - 2c) \tag{103}$$

and also

$$k_2^H(y \star c) = (1 - c)^2 + c^2 - 2y(1 - y)(1 - 2c)^2 \tag{104}$$

$$k_3^H(y \star c) = 1 - 3c + 3c^2 - 3y(1 - y)(1 - 2c)^2 . \tag{105}$$

Therefore,

$$g_{2,1}^x(y) = \frac{1}{k_2^H(x \star c)} \left[(1-c)^2 + c^2 - (1-2c)^2 \right], \tag{106}$$

$$g_{3,1}^x(y) = \frac{1}{k_3^H(x \star c)} \left[1 - 3c + 3c^2 - (1-2c)^2 \right] \tag{107}$$

which makes the function independent of y in both cases. It follows that for these values of α and β Equation (95) holds even with equality and therefore $\Bbbk$ is linear, which proves Lemma IV.11.

Let's further investigate Equation 102, for which taking the derivative in y gives:

$$(g_\alpha^x)'(y) = \frac{\alpha(1-2c)}{k_\alpha^H(x \star c)} \left[\left[y_c^{\alpha-1} - (1-y_c)^{\alpha-1} \right] - \frac{x_c^{\alpha-1} - (1-x_c)^{\alpha-1}}{x^{\alpha-1} - (1-x)^{\alpha-1}} \left[y^{\alpha-1} - (1-y)^{\alpha-1} \right] \right], \tag{108}$$

where we use $y_c := y \star c$ and $x_c := x \star c$.

Obviously we have $(g_\alpha^x)'(x) = 0$ as needed. Further define $f(x) := x^{\alpha-1} - (1-x)^{\alpha-1}$ which gives

$$(g_\alpha^x)'(y) = \frac{\alpha(1-2c)}{k_\alpha^H(x \star c)} \left[f(y_c) - \frac{f(x_c)}{f(x)} f(y) \right], \tag{109}$$

Note that $y_c \geq y$ for $0 \leq y \leq 0.5$ and $y_c \leq y$ for $0.5 \leq y \leq 1$. Since the prefactor is always positive and independent of y, Lemma IV.10 follows from the properties of the function $f(x)$ and in particular its first and second derivative (note that these have factors $(\alpha - 1)$ and $(\alpha - 2)(\alpha - 1)$ respectively and latter has exponent $\alpha - 3$, which leads to the change between convexity and concavity at $\alpha = 1, 2, 3$).