Adder MAC and estimates for Rényi entropy

Ganesh Ajjanagadde, Yury Polyanskiy
Department of EECS, MIT, Cambridge, MA, 02139
email: {gajjanag,yp}@mit.edu

***Abstract*—This paper discusses a possible program for improving the outer (converse) bounds on the finite-blocklength performance of multiple-access codes. The program is based on a certain conjecture involving Rényi entropy of a sum of two independent binary vectors. Some partial results towards showing the conjecture are presented. The problem of bounding the joint Rényi entropy in terms of the marginal entropies is addressed.**

I. Introduction

Consider the following noiseless multiple-access channel, or *adder MAC*:

$$Y = A + B, \qquad A, B \in \{0,1\},\ Y \in \{0,1,2\}\,.$$

The capacity region of all MACs was found classically by Ahslwede and Liao [1], [2]. For the adder MAC we have

$$R_1 \le \log 2,\ R_2 \le \log 2,\ R_1 + R_2 \le \frac{3}{2}\log 2\,.$$

Let us state explicitly what this means. Given a pair of subsets $\mathcal{C}_1, \mathcal{C}_2 \subset \{0,1\}^n$ the probability of error ϵ is defined as

$$\epsilon(\mathcal{C}_1, \mathcal{C}_2) \triangleq 1 - \frac{|\mathcal{C}_1 + \mathcal{C}_2|}{|\mathcal{C}_1||\mathcal{C}_2|}$$

and measures the average multiplicity of elements of the sumset $\mathcal{C}_1 + \mathcal{C}_2$. Given n and ϵ we define

$$M^*(n,\epsilon) \triangleq \max\{|\mathcal{C}_1| \cdot |\mathcal{C}_2| : \epsilon(\mathcal{C}_1, \mathcal{C}_2) \le \epsilon\}\,.$$

The results of Ahslwede and Liao state:

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log M^*(n,\epsilon) = \frac{3}{2}\log 2\,.$$

Later improvements of Dueck and Ahlswede [3], [4] show:

$$\log M^*(n,\epsilon) \le \frac{3n}{2}\log 2 + K_\epsilon\sqrt{n}\log n\,, \tag{1}$$

where $K_\epsilon > 0$ is a constant. Using random coding and estimates similar to those in [5] it is easy to show

$$\log M^*(n,\epsilon) \ge \frac{3n}{2}\log 2 - \frac{\sqrt{n}}{2}Q^{-1}(\epsilon) + O(\log n)\,, \tag{2}$$

where $Q^{-1}(\epsilon)$ is the standard normal quantile function.

The state-of-the-art summarized by (1)-(2) is quite unsatisfactory. First, the sign of the second-order term is not clear. Second, even for $\epsilon = 0$ the best known upper-bound, cf. [6], is a rather simple

$$\log M^*(n,0) \le \frac{3n}{2}\log 2\,,$$

obtained by maximizing the entropy $H(X_1 + X_2)$ over $X_1 \perp\!\!\!\perp X_2$. Third, and perhaps most importantly, if it happens that the random-coding estimate (2) provides a correct second-order term, it would be a very strong indicator that random-like signaling is not only sufficient but also necessary for optimal communication. Indeed, note that appearance of the $\sqrt{n}$-type second-order terms, cf. [5], [7], has so far been purely due to the i.i.d.-randomness of the channel noise. The adder MAC does not have any channel noise, so the $\sqrt{n}$-type second-order can only arise from randomness of the multi-user interference.

The present paper outlines our program aimed at improving upper bound (1). Although not successful yet, we think some of our partial results are nevertheless of interest.

II. Main Results

The main idea is to follow the general method proposed in [8] (and even earlier in the quantum community): In order to prove the strong converse it is sufficient to maximize the Rényi mutual information K_λ, defined in [9]. In the context of the adder MAC, we do not need the more complicated definition of K_λ as it coincides with the Rényi entropy:

$$H_\alpha(X) \triangleq \frac{1}{1-\alpha}\log\sum_x [P_X(x)]^\alpha\,.$$

Our main conjecture is the following:

Conjecture 1. *For any $A^n \perp\!\!\!\perp B^n$ taking values in $\{0,1\}^n$*

$$H_\alpha(A^n + B^n) \le nH_\alpha(Y^*) \quad \forall\alpha \in [0,1) \tag{3}$$

where $P_{Y^*} = [\frac{1}{4}, \frac{1}{2}, \frac{1}{4}]$.

By [8, (32),(60)], Conj. 1 implies

$$\log M^*(n,\epsilon) \le \frac{3n}{2}\log 2 + O(\sqrt{n})$$

an improvement of (1).

Note that Y^* can only be generated by channel inputs A and B that are i.i.d Bernoulli($\frac{1}{2}$). We first offer a proposition establishing the truth of Conj. 1 for $n = 1$:

Proposition 1. *For all $A \perp\!\!\!\perp B \in \{0,1\}$ and $\alpha \in [0,1]$,*

$$H_\alpha(A + B) \le H_\alpha(Y^*). \tag{4}$$

We also prove Conj. 1 for $n = 2, \alpha \le 0.5$:

Proposition 2. *For all $A^2 \perp\!\!\!\perp B^2 \in \{0,1\}^2$ and $\alpha \in \left[0, \frac{1}{2}\right]$,*

$$H_\alpha(A^2 + B^2) \le 2H_\alpha(Y^*). \tag{5}$$

We prove some results of secondary importance regarding the setup of Conj. 1, and adder MAC's in general. We show why it is essential to take $\alpha < 1$ for Conj. 1:

Proposition 3. *For all $\alpha > 1$, there exists a $A^2 \perp\!\!\!\perp B^2$ satisfying:*

$$H_\alpha(A^2 + B^2) > 2H_\alpha(Y^*).$$

One might believe that for general adder MAC's ($Y = A + B$), $H(Y)$ is maximized at P_A and P_B both being uniform. Here is a proposition showing this to be false:

Proposition 4. *Consider alphabet $X = (0,1,2,\ldots,m)$, and let P_A and P_B be two distributions on it. Consider the channel $Y = A + B$. Then, for $m \ge 2, \alpha \in (0,\infty)$, $H_\alpha(Y)$ is not maximized with P_A and P_B being uniform distributions.*

Conj. 1 is nontrivial since although

$$H(P_{XY}) \le H(P_X) + H(P_Y) \quad \text{(sub-additivity)},$$

$\forall \alpha \notin \{0,1\}$, cf. [10]:

$$H_\alpha(P_{XY}) \not\leq H_\alpha(P_X) + H_\alpha(P_Y). \tag{6}$$

This prevents the application of the standard single-letterization. In Section IV, we address this issue in greater depth, and prove Thm. 1 and Thm. 2 that try to address this issue. Unfortunately, neither of these resolve Conj. 1.

III. Proofs

A. Blocklength $n = 1$

In order to prove Prop. 1, we establish a lemma which characterizes the possible single-letter channel output distributions:

Lemma 1. *Let P_Y be (r_0, r_1, r_2) on $(0,1,2)$ respectively. Then, P_Y is a channel output distribution iff $\sqrt{r_0} + \sqrt{r_2} \leq 1$.*

Proof. Let P_A be $(p, 1-p)$ and let P_B be $(q, 1-q)$ on $(0,1)$ respectively. Necessity follows via Cauchy-Schwarz:

$$\begin{aligned}(\sqrt{r_0} + \sqrt{r_2})^2 &= \left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right)^2 \\ &\leq (p + (1-p))(q + (1-q)) \\ &= 1.\end{aligned}$$

Now we show sufficiency. From the channel model, we see that p, q are roots of the quadratic equation:

$$f(x) = x^2 - (r_0 - r_2 + 1)x + r_0 = 0. \tag{7}$$

The roots of (7) are in $[0,1]$ iff its discriminant is nonnegative:

$$(r_0 - r_2 + 1)^2 \geq 4r_0 \tag{8}$$

$$\Leftrightarrow \sqrt{r_0} + \sqrt{r_2} \leq 1. \tag{9}$$

This equivalence follows from repeated squaring. □

We now return to the proof of Prop. 1. Here, and subsequently, we use the following definition of "trading mass":

Definition 1. *$q \prec p$ with respect to the partial order of majorization iff q may be derived from p by successive applications of linear transformations T ([11, 2.19 Lemma 2]) mapping vector p to p' in $\mathbb{R}^n$ as follows: $p = (p_0, p_1, \ldots, p_n)$, $p' = (p_0, \ldots, p_{i-1}, p_i - \epsilon, p_{i+1}, \ldots, p_{j-1}, p_j + \epsilon, p_{j+1}, \ldots, p_n)$, where without loss $p_i \geq p_j$, and we ensure $\epsilon \leq \frac{|p_i - p_j|}{2}$. We call such T "trading mass".*

Proof of Prop. 1. "Trading mass" between r_0, r_2 increases Rényi entropy until $\sqrt{r_0} + \sqrt{r_2} = 1$, implying:

$$p_A = (p, 1-p), p_B = (p, 1-p)$$

for some $p \in [0,1]$. It hence suffices to show that:

$$\left(p^2\right)^\alpha + \left((1-p)^2\right)^\alpha + (2p(1-p))^\alpha \tag{10}$$

is maximized at $p = 0.5$, and without loss assume $p \leq 0.5$ by symmetry. The proof of (10) proceeds in two steps. First, we show that we can reduce to $p \in [0.25, 0.5]$ via majorization. Then, we analyze this interval more carefully by bounding expressions occuring in $\frac{d}{dp}$. For the first step, observe that when $p < 0.25$, $(p^2 < 2p(1-p) < (1-p)^2)$ majorizes $(0.25 \leq 0.25 < 0.5)$. We now analyze $p \in [0.25, 0.5]$ more carefully. Expanding the derivative, we get

$$\frac{1}{2\alpha}\frac{d}{dp} = p^{2\alpha-1} - (1-p)^{2\alpha-1} + (1-2p)(2p(1-p))^{\alpha-1}. \tag{11}$$

First of all, if $\alpha \leq 0.5$, $p^{2\alpha-1} - (1-p)^{2\alpha-1} \geq 0$, yielding no stationary points apart from $p = 0.5$. Thus, we may now restrict to $\alpha > 0.5$. Here, consider $f(x) = x^{2\alpha-1}$ and apply Lagrange's mean value theorem to get for some $c \in (p, 1-p)$:

$$\begin{aligned}\frac{(1-p)^{2\alpha-1} - p^{2\alpha-1}}{(1-p) - p} &= (2\alpha - 1)c^{2\alpha-2} \\ &\leq (2\alpha-1)p^{2\alpha-2}.\end{aligned}$$

Thus, from (11) it suffices to show that:

$$\left(\frac{2(1-p)}{p}\right)^{\alpha-1} > 2\alpha - 1.$$

But we have:

$$\frac{2(1-p)}{p} \leq 6$$

reducing our task to:

$$6^{\alpha-1} \geq 2\alpha - 1. \tag{12}$$

But (12) is clear, since at $\alpha = 1$, we have equality, for $\alpha = 0.5$, we have strict inequality, and

$$\begin{aligned}&\frac{d}{d\alpha}(6^{\alpha-1} - (2\alpha - 1)) \\ &= 6^{\alpha-1}\ln(6) - 2 \\ &\leq \ln(6) - 2 \\ &< 0\end{aligned}$$

giving (12). Thus, we have shown that 0.5 is the only stationary point. Since the maximum is not attained on boundary, and we have a continuous function over a compact set, 0.5 indeed yields the maximum. This completes the proof. □

B. KL divergence to product measures

Conj. 1 is a statement about the set of all distributions that arise at the output of an adder MAC, namely:

$$\mathcal{P}_n \triangleq \{P_{Y^n} : Y^n = A^n + B^n, A^n \perp\!\!\!\perp B^n \in \{0,1\}^n\}.$$

Here we provide information about $\mathcal{P}_n$ in terms of its KL divergence to product distributions. These may be used in large deviations analysis of this channel via Sanov's [12] and related theorems. In the propositions of this subsection, we make use of Lemma 1 repeatedly.

Proposition 5. *Let $P_V = (v_0, v_1, v_2)$. Then,*

$$\min_{P \in \mathcal{P}_n} D(P||{P_V}^n) = nD({P_Y}^*||P_V),$$

where

$${P_Y}* = \begin{cases}(p^2, 2p(1-p), (1-p)^2) & \text{if } \sqrt{v_0} + \sqrt{v_2} \leq 1 \\ P_V & \text{otherwise}\end{cases}$$

with p satisfying:

$$\frac{2^{2p-1}{v_0}^p {v_2}^{p-1}(1-p)}{(1-v_0-v_2)^{2p-1}p} = 1. \tag{13}$$

Remark: If $v_0 = v_2 = v$, and $v \in \left[0.25, \frac{1}{3}\right]$, we obtain ${P_Y}* = (0.25, 0.5, 0.25)$, and

$$\min_{P \in \mathcal{P}_n} D(P||{P_V}^n) = n(-1.5 - 0.5\log(v(1-2v))).$$

Remark: Note that the solution to (13) may not be unique, in which case Prop. 5 does not specify p. Conditions can be derived for the uniqueness in special cases of interest, such as $v_0 = v_2 = v$.

Proof. It suffices to examine the single letter case, since:

$$D({P_{Y^n}}||{P_V}^n) \geq \sum_{i=1}^n D(P_{Y_i}||P_V). \tag{14}$$

For $\sqrt{v_0}+\sqrt{v_2} \leq 1$, we may apply Lemma 1 to conclude that $P_V \in \mathcal{P}_n$, giving the result.

Now assume $\sqrt{v_0}+\sqrt{v_2} > 1$. Let $P_A = (p, 1-p)$ and $P_B = (1-q, q)$. The divergence expression is:

$$g(p,q) = p(1-q)\log\left[\frac{p(1-q)}{v_0}\right] + q(1-p)\log\left[\frac{q(1-p)}{v_2}\right] + (1-p-q+2pq)\log\left[\frac{1-p-q+2pq}{1-v_0-v_2}\right]. \quad (15)$$

Setting $\frac{\partial g}{\partial p} = \frac{\partial g}{\partial q} = 0$, we get

$$(-1+2q)\log\left[\frac{1-p-q+2pq}{1-v_0-v_2}\right] + (1-q)\log\left[\frac{p(1-q)}{v_0}\right] - q\log\left[\frac{(1-p)q}{v_2}\right] = 0. \quad (16)$$

and

$$(-1+2p)\log\left[\frac{1-p-q+2pq}{1-2v}\right] + (1-p)\log\left[\frac{q(1-p)}{v_2}\right] - p\log\left[\frac{(1-q)p}{v_0}\right] = 0. \quad (17)$$

Adding (16) and (17), we get

$$(-2+2p+2q)\log\left[\frac{1-p-q+2pq}{1-v_0-v_2}\right] + (1-p-q)\log\left[\frac{pq(1-p)(1-q)}{v_0v_2}\right] = 0. \quad (18)$$

Suppose $p+q \neq 1$. Then we have

$$\left(\frac{1-p-q+2pq}{1-v_0-v_2}\right)^2 = \frac{pq(1-p)(1-q)}{v_0v_2}. \quad (19)$$

which we claim leads to a contradiction. We know that

$$\sqrt{v_0}+\sqrt{v_2} \geq 1 \Rightarrow 4v_0v_2 > (1-v_0-v_2)^2. \quad (20)$$

Also, we have:

$$[(1-p-q)+(2pq)]^2 \geq 4pq(pq+1-p-q) = 4pq(1-p)(1-q). \quad (21)$$

The two inequalities (20) and (21) contradict (19) as desired. Thus, $p+q=1$, and therefore (16) and (17) become after simplification the required (13). □

Remark: Similarly, one may consider a divergence minimization problem with the arguments of divergence exchanged. The single-letterization step, (14), is no longer valid. Nevertheless, for $n=1$ it is possible to compute the exact value:

$$\min_{P\in\mathcal{P}_1} D(P_V||P) = D(P_Y{}^*||P_V),$$

where

$$P_Y{}^* = \begin{cases} (p^2, 2p(1-p), (1-p)^2) & \text{if } \sqrt{v_0}+\sqrt{v_2} \leq 1 \\ P_V & \text{otherwise} \end{cases}$$

and

$$p = \frac{1+v_0-v_2}{2}.$$

C. Blocklength $n=2$

We now move to $n=2$, where we prove Prop. 2.

Proof of Prop. 2. Let $P_{A^2} = (p_0, p_1, p_2, p_3)$ and likewise $P_{B^2} = (q_0, q_1, q_2, q_3)$. The labelling is based on base two representation $(0\to 00, 1\to 01, 2\to 10, 3\to 11)$. Then, our conjecture is:

$$(p_0q_0)^\alpha + (p_1q_1)^\alpha + (p_2q_2)^\alpha + (p_3q_3)^\alpha + (p_0q_1+p_1q_0)^\alpha + (p_0q_2+p_2q_0)^\alpha + (p_1q_3+p_3q_1)^\alpha + (p_2q_3+p_3q_2)^\alpha + (p_0q_3+p_1q_2+p_2q_1+p_3q_0)^\alpha \leq \frac{1}{4^\alpha} + \frac{4}{8^\alpha} + \frac{4}{16^\alpha} \quad (22)$$

The idea is essentially "trading mass". We "equalize" (p_0, p_3) and (q_0, q_3) simultaneously. By "equalizing", we mean that we replace a pair (a,b) by $\left(\frac{a+b}{2}, \frac{a+b}{2}\right)$. We claim that this increases the left hand side of the inequality. Observe that $(p_0q_1+p_1q_0)+(p_1q_3+p_3q_1)$ is invariant under this operation. Likewise, $(p_0q_2+p_2q_0)+(p_2q_3+p_3q_2)$ is also invariant under this operation. Moreover, it is clear that the terms $p_0q_1+p_1q_0$ and $p_1q_3+p_3q_1$ have been "equalized". Similarly, $p_0q_2+p_2q_0$ and $p_2q_3+p_3q_2$ have been "equalized" as well. Thus, we have increased the sum of the corresponding four terms of (22). The terms p_1q_1 and p_2q_2 are unaffected by this operation. Thus, for the above claim, it suffices to check that the sum of the remaining three terms in (22) has not decreased. For this, observe that if (p_0, p_3) and (q_0, q_3) are "opposite sorted" (terminology that is used regarding rearrangements of sequences, see e.g [11, Section 10.1]), we have a successful majorization:

$\left(p_1q_2+p_2q_1+2\frac{p_0+p_3}{2}\frac{q_0+q_3}{2} \geq \frac{p_0+p_3}{2}\frac{q_0+q_3}{2} \geq \frac{p_0+p_3}{2}\frac{q_0+q_3}{2}\right)$
is majorized by
$(p_1q_2+p_2q_1+p_0q_3+p_3q_0 \geq p_0q_0?p_3q_3)$,
where '?' denotes an intermediate inequality that is not needed to have a definite direction for the majorization to hold.

In the case of similar sorting, by the rearrangement inequality [11, Thm. 368], it follows that the term $p_0q_3+p_1q_2+p_2q_1+p_3q_0$ has not decreased. Thus, it suffices to check that $(p_0q_0)^\alpha + (p_3q_3)^\alpha$ has increased. For this, we depend critically on $\alpha \leq 0.5$.

By normalizing, we may assume that $p_0+p_3 = q_0+q_3 = 2$ without loss of generality, since the desired inequality is homogenous. Thus, it suffices to check that for any variables w, x, y, z such that $w+z = x+y = 2$, and $\alpha \leq 0.5$, we have:

$$(wx)^\alpha + (yz)^\alpha \leq \sqrt{(w^{2\alpha}+z^{2\alpha})(x^{2\alpha}+y^{2\alpha})} \leq \sqrt{(w+z)(x+y)} \leq 2,$$

using $\alpha \leq 0.5$ as desired.

Thus, we have the claim that the "0-3" equalization can't decrease the left hand side when $\alpha \leq 0.5$. By symmetry, we may follow the "0-3" equalization by a "1-2" equalization to further not decrease the left hand side. Thus, for (22) (under $\alpha \leq 0.5$), it suffices to prove (22) for for all choices of $0 \leq p, q \leq 0.5$, where A^2 takes the p.m.f $(p, 0.5-p, 0.5-p, p)$, and B^2 takes the p.m.f $(0.5-q, q, q, 0.5-q)$.

Let $\vec{y}$ denote the channel output distribution for the above A^2 and B^2. Then,

$\vec{y} = (2p(0.5-q)+2q(0.5-p), pq+(0.5-p)(0.5-q), pq+(0.5-p)(0.5-q), pq+(0.5-p)(0.5-q), pq+(0.5-p)(0.5-q), p(0.5-q), p(0.5-q), q(0.5-p), q(0.5-p))$.

Suppose $2p(0.5-q)+2q(0.5-p) \leq 0.25$. Then, the equation $4r(0.5-r) = 2p(0.5-q)+2q(0.5-p)$ has a solution in $0 \leq r \leq 0.5$. In this case, consider $A^{*2} = (r, 0.5-r, 0.5-r, r)$ and $B^{*2} = (0.5-r, r, r, 0.5-r)$ respectively. Then, the channel output is:

$\vec{y^*} = (4r(0.5-r), r^2+(0.5-r)^2, r^2+(0.5-r)^2, r^2+(0.5-r)^2, r^2+(0.5-r)^2, r^2+(0.5-r)^2, r(0.5-r), r(0.5-r), r(0.5-r), r(0.5-r))$.

This effectively "matches" the desired $\vec{y}$. More precisely, it is clear that the first term is the same in both vectors; $r(0.5-r)$ is the average of the sum of the last 4 terms of $\vec{y}$; and hence in fact the first five terms are the same in both vectors.

Putting these claims together, we have that $\vec{y^*}$ is majorized by $\vec{y}$ in this case.

Now suppose $2p(0.5-q)+2q(0.5-p) > 0.25$. Then, the equation $2r^2 + 2(0.5-r)^2 = 2p(0.5-q) + 2q(0.5-p)$ has a solution in $0 \le r \le 0.5$. In this case, consider $A^{**2} = (r, 0.5-r, 0.5-r, r)$ and $B^{**2} = (r, 0.5-r, 0.5-r, r)$ respectively. Then, the channel output is:

$\vec{y^{**}} = (2r^2+2(0.5-r)^2, 2r(0.5-r), 2r(0.5-r), 2r(0.5-r), 2r(0.5-r), r^2, r^2, (0.5-r)^2, (0.5-r)^2)$.

We claim that this effectively "matches" the desired $\vec{y}$. The first term is the same in both vectors by the choice of r. Also, the sum of the last four terms is the same in both vectors. Thus, the first five terms are the same in both vectors.

We now claim that $|r^2 - (0.5-r)^2| \le |p(0.5-q) - q(0.5-p)|$.

This claim shows that $(r^2, (0.5-r)^2)$ can be obtained by "trading mass" between $(p(0.5-q), q(0.5-p))$.

We now prove the claim.

For ease of dealing with absolute values, we assume without loss of generality that $r \ge 0.25$ and $p \ge q$.

Using the fact that r and $0.5-r$ are the two roots of a quadratic $x^2 + (0.5-x)^2 = p(0.5-q) + q(0.5-p)$, and simplifying, we see that this is equivalent to:

$\frac{\sqrt{4p+4q-16pq-1}}{2} \le p-q$.

Squaring both sides, it suffices to show that:

$4p^2 + 4q^2 - 4p - 4q + 8pq + 1 \ge 0$, or equivalently,

$(2p+2q-1)^2 \ge 0$, which is clearly true.

Collecting all these claims, we see that $\vec{y^{**}}$ is majorized by $\vec{y}$ in this case.

Altogether, we have now reduced the task to proving two single variable inequalities, one corresponding to $\vec{y^*}$, and the other corresponding to $\vec{y^{**}}$ parametrized by the variable r.

These single variable inequalities are easy to establish by derivative tests, similar to the proof of Prop. 1. Thus, we have resolved the conjecture when $n = 2, \alpha \le 0.5$. □

D. Counterexamples for $\alpha > 1$ and non-binary cases

We prove Prop. 3 which shows why $\alpha \le 1$ is essential to Conj. 1:

Proof of Prop. 3. Take

$$P_{A^2} = (0.5-p, p, p, 0.5-p), P_{B^2} = (p, 0.5-p, 0.5-p, p)$$

for $0 \le p \le 0.5$. We examine

$$2^{(1-\alpha)H_\alpha} = 4(p(0.5-p))^\alpha + 4(p^2+(0.5-p)^2)^\alpha + (4p(0.5-p))^\alpha.$$

Then from a derivative test (with respect to p) it is clear that $p = 0.25$ is a stationary point. However, $\frac{d^2}{dp^2} = -2^{5-4\alpha}(2^\alpha - 2)^2\alpha < 0$ for $\alpha \ne 1$. This shows that for $\alpha > 1$, the Rényi entropy is actually at a local minimum at $p = 0.25$, completing the proof of this proposition. □

Note that the construction for $n = 2$ can be combined with an independent collection of product measures (over $n-2$ channel uses) to generate counterexamples for all $n \ge 2, \alpha > 1$.

We now prove Prop. 4 which demonstrates that for general adder MAC's, capacity achieving input distributions are not necessarily uniform:

Proof of Prop. 4. Fix P_B to be uniform. Then, we claim that P_A is not uniform at optimality. Suppose not, then P_Y is a triangle shaped distribution (convolution of two rectangles). For $P_A = (p_0, p_1, \ldots, p_m)$, "trade mass" slightly between p_0 and p_1 to make $p_0 > p_1$ while keeping $p_0 + p_1$ fixed. Then, only two elements of P_Y are affected, namely the two that are proportional to p_0 and $1-p_0$. But by this "mass trade", we have moved $(p_0, 1-p_0)$ closer to $(0.5, 0.5)$, so the entropy must have increased. Note that since the above proof only relies on majorization, this works for any H_α with $\alpha \in (0, \infty)$. □

The above proof suggests that when there is a large amount of "additive energy", uniform input distributions do not maximize output entropy.

IV. General conjecture and additivity of Rényi entropy

In Section II, it was remarked that sub-additivity fails for Rényi entropy (6). In fact, more is true. It turns out that one can fix $H_\alpha(P_X)$ and $H_\alpha(P_Y)$ and make $H_\alpha(P_{XY}) \nearrow \infty$ $(\forall \alpha \in (0,1))$ [13]. Also, $\forall \alpha \in (1, \infty)$, there does not exist a non-trivial homogeneous inequality involving Rényi entropies of order α [14]. Nevertheless, some estimates resembling sub-additivity can be made, such as the following theorem:

Theorem 1.

$$H_\alpha(P_{XY}) \le H_\alpha(P_X) + H_{\frac{1}{\alpha}}(P_{Y_\alpha}) \, (\forall \alpha \ge 0), \tag{23}$$

where

$$\mathbb{P}[Y_\alpha = y] = \sum_x P_{XY}(x,y)^\alpha \exp[(\alpha-1)H_\alpha(X,Y)].$$

More generally, for any $0 < u < v < \infty$, we have:

$$\frac{1-u}{u} H_u(X,Y) + \frac{v-1}{v} H_v(X,Y) \le \left(\frac{1}{u} - \frac{1}{v}\right)\left(H_{\frac{u}{v}}(X_v) + H_{\frac{v}{u}}(Y_u)\right). \tag{24}$$

Proof. The summation form of Minkowski's inequality for mixed l_u, l_v norms gives for $0 < u < v < \infty$:

$$\left[\sum_y \left(\sum_x p_{ij}^u\right)^{\frac{v}{u}}\right]^{\frac{1}{v}} \le \left[\sum_x \left(\sum_y p_{ij}^v\right)^{\frac{u}{v}}\right]^{\frac{1}{u}}. \tag{25}$$

Using $\sum_i {p_i}^\alpha = \exp((1-\alpha)H_\alpha(P))$ and simplifying (25), we get (24). □

Note that the right hand sides of (23) and (24) are *not* functions of marginals of P_X, P_Y and thus cannot be used in an induction to prove Conj. 1. Below, we discuss an alternative bounding technique in terms of tilted marginals. The idea is that with sufficient tilt towards the uniform, the Rényi entropy can be boosted significantly.

First, we define what we mean by exponential tilting towards the uniform distribution:

Definition 2. *Let $P = (p_1, p_2, \ldots, p_N)$ denote a probability vector on N atoms. Define $P^\beta = \frac{1}{Z}\left(p_1^\beta, p_2^\beta, \ldots, p_N^\beta\right)$, where*

$$Z = \sum_{i=1}^N p_i^\beta$$

is a normalization constant. We call P^β the "β-tilt of P".

Note that this is a special form of the standard exponential tilting encountered in information geometry, large deviations, etc.

Definition 3. *Define the set of allowable tilts*

$$T_{\alpha,n} = \{\beta : \forall X^n \; H_\alpha(X^n) \leq \sum_{i=1}^n H_\alpha({P_{X_i}}^\beta)\}. \tag{26}$$

Note that $0 \in T_{\alpha,n}$ since $P^{\beta=0}$ is a uniform distribution.

We now prove a lemma which relates the Rényi entropy of the "β-tilt of P" to Rényi entropies of two different orders, namely $\alpha\beta$ and β:

Lemma 2. *For all $\alpha \in [0,\infty)\backslash\{1\}, \beta \in [0,\infty)$, we have:*

$$H_\alpha(P^\beta) = \frac{1-\alpha\beta}{1-\alpha} H_{\alpha\beta}(P) - \frac{\alpha(1-\beta)}{1-\alpha} H_\beta(P). \tag{27}$$

Proof.

$$\begin{aligned} H_\alpha(P^\beta) &= \frac{1}{1-\alpha} \log\left(\frac{1}{Z^\alpha}\left(\sum_{i=1}^n {p_i}^{\beta\alpha}\right)\right) \\ &= -\frac{\alpha}{1-\alpha}\log(Z) + \frac{1-\alpha\beta}{1-\alpha} H_{\alpha\beta}(P) \\ &= -\frac{\alpha(1-\beta)}{1-\alpha} H_\beta(P) + \frac{1-\alpha\beta}{1-\alpha} H_{\alpha\beta}(P). \end{aligned}$$

□

Thus our approach also probes inequalities involving Rényi entropies of two different orders, namely $\alpha\beta$ and β.

The sets $T_{\alpha,n}$ turn out to have one of the two forms: $T_{\alpha,n} = [0,t)$, or $T_{\alpha,n} = [0,t]$. This is intuitively clear, since tilting towards the uniform should increase entropy monotonically. The formal proof follows from the following lemma:

Lemma 3. *For all $\alpha \geq 0$, $H_\alpha(P^\beta)$ is non-increasing with β. Moreover, it is strictly decreasing unless P is uniform.*

Proof.

$$\begin{aligned} &\frac{d}{d\beta} H_\alpha(P^\beta) \leq 0 \\ \Leftrightarrow \frac{\alpha}{1-\alpha}&\left(\frac{\sum_{i=1}^n p_i^{\alpha\beta}\log(p_i)}{\sum_{i=1}^n p_i^{\alpha\beta}} - \frac{\sum_{i=1}^n p_i^\beta \log(p_i)}{\sum_{i=1}^n p_i^\beta}\right) \leq 0. \end{aligned}$$

Thus, it suffices to prove that

$$f(\beta) = \frac{\sum_{i=1}^n p_i^\beta \log(p_i)}{\sum_{i=1}^n p_i^\beta}$$

is non-decreasing in β, and strictly increasing unless P is uniform. For this, note that numerator of $\frac{d}{d\beta} f(\beta)$ is:

$$\left(\sum_{i=1}^n p_i^\beta (\log(p_i))^2\right)\left(\sum_{i=1}^n p_i^\beta\right) - \left(\sum_{i=1}^n p_i^\beta \log(p_i)\right)^2 \geq 0$$

by the Cauchy-Schwarz inequality, and equality holds iff P is uniform, as desired. □

Combining Lemma 3 with the fact that Rényi entropy is not sub-additive for $\alpha \notin \{0,1\}$, we now have the fact that $\{0\} \subseteq T_{\alpha,n} \subseteq [0,1)$ for such α.

The next simple result is the following:

$$\forall \alpha > 1 : \frac{1}{\alpha} \in T_{\alpha,n}\,. \tag{28}$$

Indeed, consider the chain

$$\begin{aligned} \sum_{i=1}^n H_\alpha({P_{X_i}}^{\frac{1}{\alpha}}) &= \sum_{i=1}^n H_{\frac{1}{\alpha}}(P_{X_i}) \\ &\geq \sum_{i=1}^n H(P_{X_i}) \\ &\geq H(P_{X^n}) \\ &\geq H_\alpha(P_{X^n}) \end{aligned} \tag{29}$$

by Lemma 2 together with the monotonicity of Rényi entropy with respect to its order α [15, Prop. 5.3.1].

The main result of this section (Thm. 2) states that asymptotically as $n \to \infty$, the estimate (28) is the best possible for $\alpha > 1$. It also demonstrates that for $\alpha \in (0,1)$, as $n \to \infty$, no nontrivial tilt is allowed.

Theorem 2.

$$\forall \alpha \in (0,1) \quad \sup(T_{\alpha,n}) \in \left[0, \frac{1}{n-(n-1)\alpha}\right]. \tag{30}$$

$$\forall \alpha \in (1,\infty) \quad \sup(T_{\alpha,n}) \in \left[\frac{1}{\alpha}, \frac{1}{n} + \frac{n-1}{n\alpha}\right]. \tag{31}$$

We also have the special cases:

$$T_{0,n} = [0,\infty). \tag{32}$$

$$T_{1,n} = [0,1]. \tag{33}$$

$$T_{\infty,n} = \{0\}. \tag{34}$$

Corollary 1. *As $n \to \infty$,*

$$T_{\alpha,n} \to \{0\} \quad \forall \alpha \in (0,1). \tag{35}$$

$$\sup(T_{\alpha,n}) \to \frac{1}{\alpha} \quad \forall \alpha \in (1,\infty). \tag{36}$$

In order to prove this result, we first explore the idea of what we call "maximum Rényi entropy couplings". First, we define what we mean by couplings:

Definition 4. *$\mathcal{C}(P_{X_1}, P_{X_2}, \ldots, P_{X_n})$ is called the set of couplings with marginals $P_{X_1}, P_{X_2}, \ldots, P_{X_n}$. It consists of all joint distributions P_{X^n} whose marginals are $P_{X_1}, P_{X_2}, \ldots, P_{X_n}$.*

From Definition 4, we define the maximum Rényi entropy coupling:

Definition 5. *Fix $P_{X_1}, P_{X_2}, \ldots, P_{X_n}$, and also fix α. Then,*

$$P^*_{X^n} = \operatorname*{argmax}_{P_{X^n} \in \mathcal{C}(P_{X_1}, P_{X_2}, \ldots, P_{X_n})} H_\alpha(P_{X^n})$$

is called the maximum Rényi entropy coupling of $P_{X_1}, P_{X_2}, \ldots, P_{X_n}$ and order α.

Note that for $\alpha = 1$, the maximum Rényi entropy coupling reduces to the familiar product distribution

$$P_{X_1} \otimes P_{X_2} \otimes \cdots \otimes P_{X_n}$$

due to the sub-additivity of Shannon entropy, where $\otimes$ denotes the product measure. Moreover, for any order α, computation of the maximum Rényi entropy coupling is a convex optimization problem due to the concavity of Rényi entropy with respect to the underlying distribution. Concavity of Rényi entropy also yields the statistically pleasing property of exchangeability (see e.g [16] for a discussion of exchangeability) as follows. Consider a discrete vector valued random variable $(X_1, X_2, \ldots, X_n)$. Suppose one has

some prior information that certain indices are indistinguishable, i.e $P_{X_{i_1}} = P_{X_{i_2}} = \ldots = P_{X_{i_m}}$ for some indices $(i_1, i_2, \ldots, i_m)$. The principle of maximum entropy [17] suggests that in absence of further information, one should use a maximum entropy prior subject to the given constraints. However, the classical method of maximizing the Shannon entropy results in forcing independence in the prior. Statistically speaking, often one merely needs the idea of exchangeability [18]. This means that in the example above, the maximum entropy prior should be invariant under permutations of the indices $(i_1, i_2, \ldots, i_M)$. However, this is always satisfied by the Rényi entropy as well. Evaluation of the use of maximum Rényi entropy priors in statistical settings is thus an intriguing question that we do not consider here.

Our motivation in defining this notion of coupling is because it represents the tightest gap in (26) for fixed right hand side marginals. To give a flavor of subsequent analysis, we first prove Thm. 2 for the only nontrivial special case, namely $\alpha = \infty$.

For this, we first prove an easy lemma:

Lemma 4.

$$H_\infty(P_{XY}) \leq \max(H_\infty(P_X) + H_0(P_Y), H_\infty(P_Y) + H_0(P_X)).$$

Proof. Recall that $H_\infty(P) = -\log(\max(p_i))$, and $H_0(P) = \log(\text{supp}(P))$. But we know that for all i, there must exist at least one $P_{XY}(i,j) \geq \frac{P_X(i)}{\text{supp}(P_Y)}$. Likewise with j. Taking logarithms, one gets the result. □

Now consider the case when $P_X = P_Y = (p_1, p_2, \ldots, p_N)$, where without loss $p_1 \geq p_2 \geq \ldots \geq p_N$. Now, if $\frac{p_1}{N} \leq p_N$, we can create a maximum Rényi entropy coupling ($\alpha = \infty$), i.e one that achieves equality in the bound of Lemma 4. This may be done by taking $P_{XY}(1,j) = P_{XY}(i,1) = \frac{p_1}{N}$ for all $1 \leq i, j \leq N$. The remaining entries of $P_{XY}(i,j)$ with $2 \leq i, j \leq N$ may then be filled as a diagonal matrix, with entries $P_{XY}(i,i) = p_i - \frac{p_1}{N}$. The validity of this joint distribution follows from $\frac{p_1}{N} \leq p_N$. Thus, in order to show that $T_{\infty,2} = \{0\}$, it suffices to find for any $\beta > 0$ $(p_1, p_2, \ldots, p_N)$ with $\frac{p_1}{N} \leq p_N$, and:

$$(1-2\beta)H_\infty(P) + \log(N) > 2(1-\beta)H_\beta(P). \tag{37}$$

In order to do this we prove our first "Rényi entropy transition lemma":

Lemma 5. *Let $P = (p_1, p_1, \ldots, p_1, \frac{p_1}{N}, \frac{p_1}{N}, \ldots, \frac{p_1}{N})$ where p_1 occurs $M = N^{1-\beta}$ times, and normalization is ensured by taking $p_1 = \frac{N}{MN+N-M}$. Then,*

$$H_\alpha(P) = (1-\beta)\log(N) + O(1) \quad \forall \alpha \in [\beta, \infty]\,.$$

Proof. Essentially, the proof uses a discrete version of Laplace's method (see e.g [19, Thm. 4.3.1]). More rigorously,

$$\begin{aligned} H_\infty(P) &= \log\left(\frac{MN+N-M}{N}\right) \\ &= \log(M + 1 - N^{-\beta}) \\ &= (1-\beta)\log(N) + O(1) \end{aligned}$$

at $\alpha = \infty$. For $\alpha = \beta$,

$$\begin{aligned} H_\beta(P) &= \frac{1}{1-\beta}\left({p_1}^\beta\left(M + \frac{N-M}{N^\beta}\right)\right) \\ &= \frac{\beta}{1-\beta}\log\left(\frac{N}{MN+N-M}\right) + \\ &\frac{1}{1-\beta}\log(M + N^{1-\beta} - N^{1-2\beta}) \\ &= -\beta\log(N) + \log(N) + O(1). \end{aligned}$$

For $\alpha \geq \beta$, the entropy is sandwiched between the two, giving the result. □

Remark: While deriving this result, M was taken to be N^γ for some $0 < \gamma < 1$. We then optimized over γ to get a tight bound. This taking of $M = N^\gamma$ will recur in subsequent proofs.

(37) is now verified:

$$\begin{aligned} &(1-2\beta)H_\infty(P) + \log(N) - 2(1-\beta)H_\beta(P) \\ &= ((1-2\beta)(1-\beta) + 1 - 2(1-\beta)^2)\log(N) + O(1) \\ &= \beta\log(N) + O(1) \\ &> 0. \end{aligned}$$

Thus, we have proved that $T_{\infty,2} = \{0\}$, implying $T_{\infty,n} = \{0\}$ for all $n \geq 2$.

The remainder of the proof of Thm. 2 falls into two cases, namely $0 < \alpha < 1$ and $1 < \alpha < \infty$. For $\alpha = \infty$, our proof used a maximum Rényi entropy coupling to give a tight bound. In a similar manner, our subsequent proof probes "good couplings" that are possibly suboptimal in order to generate useful bounds. It turns out that for $0 < \alpha < 1$, a slightly generalized version of the coupling given in [13] gives the result. We take it up first.

Let P_{X_i} for $1 \leq i \leq n$ be identically distributed as $(p_1, p_2, \ldots)$, where we assume without loss that p_i are monotonically nonincreasing. In particular, we have $p_n \leq \frac{1}{n}$ for all n. Then the following coupling is valid:

$$P_{X^n}(i_1, i_2, \ldots, i_n) = \frac{p_N}{N^{n-1+\delta}} + \epsilon_{i_1,i_2,\ldots,i_n}$$

for $1 \leq i_1, i_2, \ldots, i_n \leq N$.

$\epsilon_{i_1,i_2,\ldots,i_n} \geq 0$ are chosen to make P_{X^n} a valid coupling, i.e satisfy the marginal constraints. $\delta > 0$ denotes a small number. But then

$$\sum_{i_1,i_2,\ldots,i_n} (P_{X^n}(i_1, i_2, \ldots, i_n))^\alpha \geq N^{n-(n-1+\delta)\alpha}{p_N}^\alpha. \tag{38}$$

Now let p_N decay as $N^{-\gamma}$ for some $\gamma > 1$. In order to ensure that the marginal entropies of orders $\alpha\beta$ and β remain bounded, we need the following constraints:

$$\gamma > \frac{1}{\alpha\beta} \quad \text{and} \quad \gamma > \frac{1}{\beta}\,. \tag{39}$$

Since $\alpha < 1$, the second constraint is redundant. Now, we will pick γ so that the entropy of this coupling goes to ∞. By the bound (38), a sufficient condition is:

$$n - (n - 1 + \delta + \gamma)\alpha > 0 \tag{40}$$

In the limit as $\delta \searrow 0$, constraints (39) and (40) can be met simultaneously if $\beta > \frac{1}{n-(n-1)\alpha}$. This completes the proof of the $\alpha < 1$ case.

We now turn to the upper bound on β for the $\alpha > 1$ case. Here, we first compute analytically the maximum Rényi entropy coupling for $\alpha = 2$. To the best of our knowledge, this is the only $1 < \alpha < \infty$

which yields the maximum Rényi entropy coupling in closed form. The amenability of $\alpha = 2$ for closed form analysis comes from the fact that the Rényi entropy consists of a monotone transformation of a quadratic form, and also because we optimize over affine constraints with a simple structure. We then use this coupling to generate a bound for all $\alpha > 1$. Note that intuitively this bound should be tightest for $\alpha = 2$, and that its quality will be worse for α far away from 2. Nevertheless, this bound suffices for our purposes, since even this bound is asymptotically tight in the sense of Corollary 1.

First, we collect a direct application of the Karush-Kuhn-Tucker conditions (see e.g [20, Prop. 5.49]) on the optimal coupling for general α in the following lemma:

Lemma 6. *Let $\alpha \notin \{0, 1, \infty\}$. Suppose P_{X_i} for $1 \leq i \leq n$ are a set of marginal distributions with full support. Then, suppose there exists a P_{X^n} with full support satisfying the following constraints:*

1) *It is a member of $\mathcal{C}(P_{X_1}, P_{X_2}, \ldots, P_{X_n})$.*
2) *$\exp\left[(P_{X^n})^{\alpha-1}\right]$ is a product distribution. Here, the tilting notation is as before, and*

$$\exp(P) = \frac{1}{Z}(\exp(p_1), \exp(p_2), \ldots, \exp(p_N))$$

where Z is a usual normalization constant.

Then, such a P is the optimal Rényi entropy coupling of order α over $\mathcal{C}(P_{X_1}, P_{X_2}, \ldots, P_{X_n})$.

We do not prove this result here, simply because we do not need it for any subsequent proofs. In fact, as one can see, the general case with loss of support is omitted from the statement. Lemma 6 is provided in order to motivate our choice of coupling that we use for our bound (for $\alpha > 1$). Nevertheless, we give a corollary of Lemma 6 for the analytically solvable $\alpha = 2$ case. This is because it yields the coupling we are going to use, and also because it has a certain elegance.

Lemma 7. *Suppose $\alpha = 2$ in above Lemma 6. Also, suppose support constraints of Lemma 6 are met. Then, the optimal coupling P^* is given by:*

$$\begin{aligned} P^* = &U_1 \otimes P_{X_2} \otimes P_{X_3} \otimes \cdots \otimes P_{X_n} \\ &+ P_{X_1} \otimes U_2 \otimes P_{X_3} \otimes \cdots \otimes P_{X_n} \\ &+ \\ &\vdots \\ &+ P_{X_1} \otimes P_{X_2} \otimes \cdots \otimes P_{X_{n-1}} \otimes U_n \\ &- (n-1) U_1 \otimes U_2 \otimes \cdots \otimes U_n. \end{aligned} \tag{41}$$

Here, U_i denote uniform random variables over the respective alphabets of the P_{X_i}.

Note that even with a mildly decaying tail, Lemma 7 forces many of the indices to have zero probability. This statement is for $\alpha = 2$, but we believe this loss of support is a general feature for $\alpha > 1$ as well. Thus, Lemma 7 resolves some of the mystery of the loss of support in maximum Rényi entropy distributions as illustrated by the following numerical example. Numerically, for $n = 3, \alpha = 4$, and channel being the binary adder MAC, it appears that

$$\begin{aligned} P_{A^*} &\approx (0, 0.1666, 0.1666, 0.1666, 0.1666, 0.1666, 0.1666, 0) \\ P_{B^*} &\approx (0.256, 0.081, 0.081, 0.081, 0.081, 0.081, 0.081, 0.256) \end{aligned} \tag{42}$$

is at least a local optimum.

This goes against the usual behavior one expects from a Shannon-like measure, where it is advantageous to exploit all available degrees of freedom. Our intuition that there is a sharp transition in behavior from $\alpha < 1$ to $\alpha > 1$ is illustrated neatly by the upcoming "Renyi entropy transition lemma" 8.

The astute reader may have noticed that the coupling we picked in Lemma 5 was designed to be as "extreme" as possible, i.e it met the necessary condition $\frac{p_1}{N} \leq p_N$ with equality. Furthermore, it yielded a distribution with two distinct probabilities that asymptotically had constant entropy held at $(1 - \beta)\log(N)$ for all orders $\alpha \in [\beta, \infty]$. Observe that Lemma 7 yields a similar such condition, namely:

$$p_N \geq \frac{d-1}{dN}.$$

The similarity of these two conditions lies in the fact that both enforce $\frac{p_1}{p_N} = O(N)$. This is not surprising, since we are dealing with $\alpha > 1$ in either case. Nevertheless, the extremization here yields a more interesting lemma, the second of our "Rényi entropy transition lemmas".

Lemma 8. *Let $P = (p_1, p_1, \ldots, p_1, \frac{n-1}{nN}, \frac{n-1}{nN}, \ldots, \frac{n-1}{nN})$ where n is held fixed, p_1 occurs $M = N^\gamma$ times, and normalization is ensured by taking $p_1 = \frac{1}{nM} + \frac{n-1}{nN}$. Then,*

$$\begin{aligned} H_\alpha(P) &= \gamma \log(N) + O(1) \quad \alpha \in (1, \infty]. \\ H_\alpha(P) &= \log(N) + O(1) \quad \alpha \in [0, 1). \end{aligned}$$

The proof of this Lemma 8 is very similar to that of Lemma 5.

Proof. The case of $\alpha = 0$ is clear since the distribution has full support. The case of $\alpha = \infty$ is also simple:

$$\begin{aligned} H_\infty(P) &= -\log(p_1) \\ &= -\log\left(\frac{1}{nM} + \frac{n-1}{nN}\right) \\ &= \log(M) + O(1) \\ &= \gamma \log(N) + O(1). \end{aligned}$$

For other $\alpha \neq 1$,

$$\begin{aligned} H_\alpha(P) = \frac{1}{1-\alpha} \log \big(&M^{1-\alpha}(1 + (n-1)N^{\gamma-1})^\alpha \\ &+ (n-1)^\alpha N^{1-\alpha} - (n-1)^\alpha N^{\gamma-\alpha} \big). \end{aligned}$$

Here, the dominant term is $M^{1-\alpha}$ if $\alpha > 1$ and $N^{1-\alpha}$ if $\alpha < 1$. Taking the logarithm, we thus get the desired result. □

A natural question is what role does n play in this Lemma 8? It turns out that n controls the Shannon entropy rate ($\alpha = 1$), which represents the boundary between the two regimes. Note that this result is true in spite of the fact that for a fixed distribution, Rényi entropy is continuous in its order α. Perhaps this result is a small instance of a larger phenomenon, roughly saying that the Rényi entropies can be controlled independently in the $\alpha < 1$ and $\alpha > 1$ regimes.

We now use Lemma 8 in our proof of Thm. 2 as follows. Consider P_{X_i} identically distributed with distribution P of Lemma 8 for $1 \leq i \leq n$. Let P_{X^n} be then distributed according to the coupling (41) of Lemma 7. A "type" consists of all terms where p_1 is chosen k out of the n possible times, and $\frac{n-1}{nN}$ is chosen in the remaining $n - k$ out of the n possible times. Thus, there are $n + 1$ types. The "size" of a type is its cardinality multiplied by the individual probability in that

type raised to the power α. Formally, by expanding and simplifying the expression for the Rényi entropy, we have:

$$H_\alpha(P_{X^n}) = \frac{1}{1-\alpha}\log\left(\sum_{k=0}^{n}(N-M)^{n-k}M^{k-\alpha}\binom{n}{k}N^{-\alpha(n-1)}\left(\frac{k}{n}\right)^{\alpha}\right). \tag{43}$$

From (43), $k=1$ is the dominant type, implying

$$\begin{aligned} H_\alpha(P_{X^n}) &= \log\left((N-M)^{n-1}M^{1-\alpha}N^{-\alpha(n-1)}\right) + O(1) \\ &= \log\left(N^{(1-\alpha)(n-1)+(1-\alpha)\gamma}\right) + O(1) \\ &= (n-1+\gamma)\log(N) + O(1). \end{aligned} \tag{44}$$

Now, since $\beta > \frac{1}{\alpha}$ ((29)), we may invoke Lemma 8 to get:

$$\begin{aligned} \sum_{i=1}^{n} H_\alpha(P_{X_i}^{\beta}) &= n\left(\frac{1-\alpha\beta}{1-\alpha}H_{\alpha\beta}(P) - \frac{\alpha(1-\beta)}{1-\alpha}H_\beta(P)\right) \\ &= n\left(\frac{\gamma-\alpha+(1-\gamma)\alpha\beta}{1-\alpha}\right)\log(N) + O(1). \end{aligned} \tag{45}$$

Comparing (44) and (45) we get (for a $\beta \in T_{\alpha,n}$)

$$\begin{aligned} n-1+\gamma &\le n\left(\frac{\gamma-\alpha+(1-\gamma)\alpha\beta}{1-\alpha}\right) \\ \Leftrightarrow \gamma\left(\frac{\alpha+(n-1)-n\alpha\beta}{\alpha-1}\right) &\le \frac{\alpha+(n-1)-n\alpha\beta}{\alpha-1} \\ \Leftrightarrow \alpha+(n-1) &\ge n\alpha\beta \\ \Leftrightarrow \beta &\le \frac{1}{n} + \frac{n-1}{n\alpha}. \end{aligned} \tag{46}$$

This completes the proof of Thm. 2.

Acknowledgements

This material is based upon work supported by the National Science Foundation CAREER award under grant agreement CCF-12-53205 and by the SuperUROP program of the Dept. of EECS at MIT.

References

[1] R. Ahlswede, "Multi-way communication channels," in *Proc. 1971 IEEE Int. Symp. Inf. Theory (ISIT)*, Tsahkadsor, Armenia, USSR, Sep. 1971, pp. 23–52.

[2] H. Liao, "Multiple access channels," Ph.D. dissertation, Dept. of Elect. Eng., U. of Hawaii, Honolulu, HI, 1972.

[3] G. Dueck, "The strong converse to the coding theorem for the multiple–access channel," *J. Comb. Inform. Syst. Sci*, vol. 6, no. 3, pp. 187–196, 1981.

[4] R. Ahlswede, "An elementary proof of the strong converse theorem for the multiple-access channel," *J. Combinatorics, Information and System Sciences*, vol. 7, no. 3, 1982.

[5] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[6] O. Ordentlich and O. Shayevitz, "A VC-dimension-based outer bound on the zero-error capacity of the binary adder channel," *arXiv preprint arXiv:1412.8670*, 2014.

[7] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," in *Trans. 3d Prague Conf. Inf. Theory*, Prague, 1962, pp. 689–723.

[8] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and rényi divergence," in *Proc. 2010 48th Allerton Conference*, Allerton Retreat Center, Monticello, IL, USA, Sep. 2010, pp. 1327–1333.

[9] I. Csiszár, "Generalized cutoff rates and Renyi's information measures," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 26 –34, Jan. 1995.

[10] J. Aczél and Z. Daróczy, "On measures of information and their characterizations," *New York*, 1975.

[11] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*. Cambridge university press, 1952.

[12] I. Sanov, *On the probability of large deviations of random variables*. United States Air Force, Office of Scientific Research, 1958.

[13] M. Kovacevic, I. Stanojevic, and V. Senk, "On the entropy of couplings," *CoRR*, vol. abs/1303.3235, 2013. [Online]. Available: http://arxiv.org/abs/1303.3235

[14] N. Linden, M. Mosonyi, and A. Winter, "The structure of Renyi entropic inequalities," *Royal Society of London Proceedings Series A*, vol. 469, p. 20737, Aug. 2013.

[15] C. Beck and F. Schögl, *Thermodynamics of chaotic systems: an introduction*. Cambridge University Press, 1995, no. 4.

[16] D. J. Aldous, *Exchangeability and related topics*. Springer, 1985.

[17] E. T. Jaynes, "Information theory and statistical mechanics," *Physical review*, vol. 106, no. 4, p. 620, 1957.

[18] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[19] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Springer Science & Business Media, 2009, vol. 38.

[20] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.