A Note on the Probability of Rectangles for Correlated Binary Strings

Or Ordentlich, Yury Polyanskiy, and Ofer Shayevitz

***Abstract*—Consider two sequences of n independent and identically distributed fair coin tosses, $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$, which are ρ-correlated for each j, i.e. $\mathbb{P}[X_j = Y_j] = \frac{1+\rho}{2}$. We study the question of how large (small) the probability $\mathbb{P}[X \in A, Y \in B]$ can be among all sets $A, B \subset \{0,1\}^n$ of a given cardinality. For sets $|A|, |B| = \Theta(2^n)$ it is well known that the largest (smallest) probability is approximately attained by concentric (anti-concentric) Hamming balls, and this can be proved via the hypercontractive inequality (reverse hypercontractivity). Here we consider the case of $|A|, |B| = 2^{\Theta(n)}$. By applying a recent extension of the hypercontractive inequality of Polyanskiy-Samorodnitsky (J. Functional Analysis, 2019), we show that Hamming balls of the same size approximately maximize $\mathbb{P}[X \in A, Y \in B]$ in the regime of $\rho \to 1$. We also prove a similar tight lower bound, i.e. show that for $\rho \to 0$ the pair of opposite Hamming balls approximately minimizes the probability $\mathbb{P}[X \in A, Y \in B]$.**

I. Introduction

Let $X \sim \mathrm{Uniform}(\{0,1\}^n)$ and $Y \in \{0,1\}^n$ be a ρ-correlated copy of X, where $0 \le \rho < 1$, i.e.,

$$\begin{aligned}&\Pr(Y = y|X = x)\\ &= \prod_{i=1}^{n} \left(\frac{1-\rho}{2}\right)^{d(x_i,y_i)} \left(\frac{1+\rho}{2}\right)^{1-d(x_i,y_i)}\\ &= \left(\frac{1+\rho}{2}\right)^{n} \cdot \left(\frac{1-\rho}{1+\rho}\right)^{d(x,y)}, \end{aligned} \tag{1}$$

where $d(x_i, y_i) = \mathbb{1}_{\{x_i \neq y_i\}}$ and $d(x, y) = \sum_{i=1}^{n} d(x_i, y_i)$. For $A, B \subset \{0,1\}^n$, we denote $P_{XY}(A \times B) \triangleq \Pr(X \in A, Y \in B)$ – probability of a rectangle with sides A and B. In this paper we are interested in the following question: *Among all sets of a given size, how large/small can the probability of a rectangle be?* Previous works addressing similar questions relied on hypercontractive and reverse hypercontractive inequalities, as we describe below. Our main innovation is applying a new tool from [1] that is a refinement of the direct hypercontractive inequality to functions with sparse support.

A direct application of the hypercontractive inequality [2], [3], [4], [5], [6] (see Section III for more details) yields that for A and B of equal cardinalities, i.e. $|A| = |B| \triangleq \eta \cdot 2^n$, we have

$$P_{XY}(A \times B) \le \eta^{\frac{2}{1+\rho}}, \tag{2}$$

whereas the reverse hypercontractive inequality of [7] was applied in [8] to obtain

$$P_{XY}(A \times B) \ge \eta^{\frac{2}{1-\rho}}. \tag{3}$$

Both bounds become quite tight for the regime of $\eta = \Theta(1)$, i.e. for very large sets of cardinalities $|A| = |B| = \Theta(2^n)$. In particular, (2) is approximately attained by taking A and B as the zero-centered Hamming balls containing all vectors with Hamming weight smaller than $\frac{n}{2} - s\sqrt{n}$, for large s independent of n, whereas (3) is approximately attained by taking A as such zero-centered ball and B as the same ball shifted such that its center is the all-ones vector. A special case of the construction in [9] also gives more constructions of sets approximately attaining (2): namely, for any $k \in \mathbb{Z}_+$ and all sufficiently large $n \ge n_0(k)$ they constructed sets $A = B$ of cardinality 2^{n-k} such that

$$P_{XY}(A \times B) \ge \Omega_\rho(1/\sqrt{k}) 2^{-k \cdot \frac{2}{1+\rho}}, \tag{4}$$

thus showing that the estimate (2) is tight (up to a polylog factor $(\log \frac{1}{\eta}))^{-\frac{1}{2}}$).

In this paper we are interested in estimating the probability of rectangles for sets A, B of much smaller cardinalities (such as those frequently encountered in information and coding theories), namely $|A| = 2^{n\alpha}, |B| = 2^{n\beta}$ for $\alpha, \beta < 1$. Our original motivation stems from the bounds on the adder multiple access channel (MAC) zero-error capacity, obtained in [10]. Sets $A, B \subset \{0,1\}^n$ are called a zero-error code for the adder MAC, if $|A + B| = |A| \cdot |B|$, where $A + B \subset \{0,1,2\}^n$ is the

O. Ordentlich is with the Rachel and Selim Benin School of Computer Science and Engineering, Hebrew University of Jerusalem, Israel (email: or.ordentlich@mail.huji.ac.il). Y. Polyanskiy is with the Massachusetts Institute of Technology, MA, USA (email: yp@mit.edu). O. Shayevitz is with the Department of Electrical Engineering Systems, Tel Aviv University, Tel Aviv, Israel (e-mail: ofersha@eng.tau.ac.il).

The work of O.O. was supported by the ISF under Grant 1791/17. The work of Y.P. was supported (in part) by the National Science Foundation under Grant No CCF-17-17842, and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370. The work of O.S. was supported by the European Research Council, under grant agreement 639573.

Minkowski sum (over the reals) of the sets A and B. The problem of finding all pairs $(R_1, R_2) \in [0,1]^2$ for which there exist a zero-error code with sizes $|A| = 2^{nR_1}$, $|B| = 2^{nR_2}$ is a long standing open problem [11], [12], [13], [14], [15], [16], [17], [18]. One of the first results in the area, due to van Tilborg [12], states that if A, B form a zero error code, then

$$W_d(A,B) \triangleq \frac{1}{n}\log|\{(a,b) \in A \times B \ : \ d(a,b) = nd\}| \tag{5}$$

$$\le \frac{1}{n}\log\binom{n}{nd} + \min(d, 1-d), \tag{6}$$

for all $d \in \{0, \frac{1}{n}, \ldots, 1\}$. The basic idea in [10] was to use (6) for upper bounding

$$P_{XY}(A \times B) = 2^{-n}\left(\frac{1+\rho}{2}\right)^n \sum_{d=0}^{1} 2^{nW_d(A,B)} \left(\frac{1-\rho}{1+\rho}\right)^{nd} \tag{7}$$

for any zero-error code (A, B), and to contrast this with lower bounds on $P_{XY}(A \times B)$ for sets $|A| = 2^{nR_1}$, $|B| = 2^{nR_2}$ obtained in [8] (see Remark 4 below). A simple modification of this approach [10] yielded the best known outer bounds on (R_1, R_2) to date, and possibly, replacing the lower bound from [8] on $P_{XY}(A \times B)$ with a sharper one, could yield stronger bounds on (R_1, R_2). For instance, if our main conjecture, stated below, turns out to be true, repeating the arguments in [10] with the improved bounds will yield that as R_1 approaches 1 we must have that $R_2 < 0.4177$, improving upon $R_2 < 0.4228$ established in [10], which is the best known bound to date.

Our interest is in the greatest and smallest exponential decay rate of $P_{XY}(A \times B)$ among all possible sets A, B of sizes $2^{n\alpha}$ and $2^{n\beta}$, respectively. To that end, for fixed $0 < \alpha, \beta < 1$ we define

$$\overline{E}(\alpha,\beta,\rho) \triangleq -\limsup_{n\to\infty} \max_{\{A\},\{B\}} \frac{1}{n}\log P_{XY}(A \times B)\,, \tag{8}$$

$$\underline{E}(\alpha,\beta,\rho) \triangleq -\liminf_{n\to\infty} \min_{\{A\},\{B\}} \frac{1}{n}\log P_{XY}(A \times B)\,, \tag{9}$$

where $\max_{\{A\},\{B\}}$ and $\min_{\{A\},\{B\}}$ denote optimizations over the sequences of sets $A_n \subset \{0,1\}^n$, $B_n \subset \{0,1\}^n$, $n \in \mathbb{Z}_+$ such that

$$|A_n| = 2^{n\alpha + o(n)}, \qquad |B_n| = 2^{n\beta + o(n)}\,.$$

Our **main conjecture** is that *both* $\overline{E}(\alpha,\beta,\rho)$ *and* $\underline{E}(\alpha,\beta,\rho)$ *are optimized by concentric (resp., anti-concentric) Hamming balls*. In this work we show partial progress towards establishing this conjecture. Our conjecture is in line with the well-known facts that among all pairs of sets $A, B \subset \{0,1\}^n$ of given sizes, the maximal distance $d_{\max}(A,B) = \max_{a\in A, b\in B} d(a,b)$ is minimized by concentric Hamming (quasi) balls [19], [20], whereas the minimum distance $d_{\min}(A,B) = \min_{a\in A, b\in B} d(a,b)$ is maximized by anti-concentric Hamming (quasi) balls [21].

Notation: Logarithms are taken to base 2 throughout, unless stated otherwise. We denote the Shannon entropy of a random variable V by $H(V)$. For a binary random variable $V \sim \mathrm{Ber}(p)$ we denote the entropy by $h(p) \triangleq -p\log p - (1-p)\log(1-p)$ and its inverse restricted to $[0, 1/2]$ by $h^{-1}(\cdot)$. For $0 \le p, q \le 1$ we denote $p * q \triangleq p(1-q) + q(1-p)$.

Our main results characterize $\overline{E}(\alpha,\alpha,\rho)$ in the low noise (large ρ) regime, and $\underline{E}(\alpha,\beta,\rho)$ in the high noise (small ρ) regime, as follows.

Theorem 1: As $\rho \to 1$ we have

$$\overline{E}(\alpha,\alpha,\rho) = (1-\alpha) + \frac{\frac{1}{2} - \sqrt{h^{-1}(\alpha)\,(1-h^{-1}(\alpha))}}{\ln 2}(1-\rho) + o(1-\rho). \tag{10}$$

Theorem 1 will follow from combining Proposition 1 and Proposition 3, proved in Section II and Section III, respectively.

Theorem 2: As $\rho \to 0$ we have

$$\underline{E}(\alpha,\beta,\rho) = (1-\alpha) + (1-\beta) + \rho\log e\left(1 - 2h^{-1}(\alpha) * h^{-1}(\beta)\right) + o(\rho). \tag{11}$$

Theorem 2 will follow from combining Proposition 2 and Proposition 6, proved in Section II and Section IV, respectively.

In both cases, the optimal exponents are obtained (up to $o(\rho)$ and $o(1-\rho)$ terms) by taking A and B to be Hamming spheres. In Section II we compute $P_{XY}(A \times B)$ for Hamming spheres, and prove the corresponding upper bound for $\overline{E}(\alpha,\alpha,\rho)$ obtained by concentric spheres, and the lower bound on $\underline{E}(\alpha,\beta,\rho)$, obtained by spheres with opposite centers. In Section III we prove the lower bound on $\overline{E}(\alpha,\alpha,\rho)$. What is interesting is that while (2) is shown via the classical hypercontractivity inequality [2], [3], [4], [5], [6], our result is shown by applying a recent improvement [1] of this inequality for functions of small support (cf. Section III). In Section IV we prove the upper bound on $\underline{E}(\alpha,\beta,\rho)$ by bounding the maximal average Hamming distance between members of A and B, subject to the cardinality constraint – another combinatorial optimization problem of possible interest.

Remark 1: After this work had been completed, we have learned from Naomi Kirshner and Alex Samorodnitsky about their concurrent work [22] in which, among other things, they were able to prove that $\overline{E}(\alpha,\alpha,\rho)$ is attained by concentric spheres for all $0<\rho<1$. Their result subsumes our Theorem 1 and relies on a different strengthening of a hypercontractive inequality.[1] The problems of characterizing $\overline{E}(\alpha,\beta,\rho)$ for $\alpha\neq\beta$ and that of $\underline{E}(\alpha,\beta,\rho)$ remain open.

II. Bounds via Spheres

For $x=(x_1,\ldots,x_n)\in\{0,1\}^n$ denote the Hamming weight of x and the Hamming sphere centered at zero as

$$|x| \triangleq |\{j : x_j = 1\}| \tag{12}$$
$$\mathbb{S}_j \triangleq \{x : |x| = j\}\,. \tag{13}$$

For the size of Hamming spheres we have [23, Exc. 5.8]

$$|\mathbb{S}_{\lfloor dn\rfloor}| = \binom{n}{\lfloor dn\rfloor} = 2^{nh(d)-\frac{1}{2}\log n+O(1)}\,, \qquad n\to\infty \tag{14}$$

where the estimate is a consequence of Stirling's formula, $O(1)$ is uniform in δ on compact subsets of $(0,1)$.

Existential results (an upper bound on $\overline{E}$ and a lower bound on $\underline{E}$) follow from taking A and B as Hamming spheres $\mathbb{S}_i$, $\mathbb{S}_j$ for a suitable i,j. Here we compute the probability of such spherical rectangles.

For any two sets $A,B\subset\{0,1\}^n$, we have

$$\begin{aligned} P_{XY}(A\times B) &= \sum_{x\in A, y\in B} 2^{-n\left(2-\log(1+\rho)-\frac{d(x,y)}{n}\log\left(\frac{1-\rho}{1+\rho}\right)\right)} \\ &= 2^{-n(1+o(1))E(A,B,\rho)}, \end{aligned} \tag{15}$$

where

$$\begin{aligned} E(A,B,\rho) \triangleq \min_{0\le d\le 1}\Bigg(&2-\log(1+\rho) \\ &- W_d(A,B) - d\log\left(\frac{1-\rho}{1+\rho}\right)\Bigg), \end{aligned} \tag{16}$$

and $W_d(A,B)$ is as defined in (5). Note that if $nd\notin\mathbb{N}$ we have that $W_d(A,B)=-\infty$, and therefore the minimization in (16) can indeed be performed on $[0,1]$ and need not be restricted to $d\in\{0,\frac{1}{n},\ldots,1\}$.

For two natural numbers $j\ge i$ and $d\in[0,1]$ such that $j-i+nd$ is even, we have that

$$\begin{aligned} &W_d(\mathbb{S}_i,\mathbb{S}_j) \\ &= \frac{1}{n}\log\binom{n}{i}\binom{i}{\frac{1}{2}(j+i-nd)}\binom{n-i}{\frac{1}{2}(j-i+nd)} \end{aligned} \tag{17}$$

for $j-i\le nd\le j+i$, and $W_d(\mathbb{S}_i,\mathbb{S}_j)=-\infty$ otherwise. Let $0<\alpha\le\beta\le 1$ and $d\in[0,1]$ be such that $i=nh^{-1}(\alpha)$ and $j=nh^{-1}(\beta)$ are integers and $j-i+nd$ is an even integer. Approximating $\binom{n}{k}=2^{n(h(k/n)+o(1))}$ as in (14), we have (18), (19) and (20) at the top of the next page, and it therefore follows from (17) that

$$W_d(\mathbb{S}_i,\mathbb{S}_j) = w_d(\alpha,\beta)+o(1), \tag{21}$$

where

$$\begin{aligned} w_d(\alpha,\beta) \triangleq\ &\alpha + h^{-1}(\alpha)h\left(\frac{1}{2}+\frac{h^{-1}(\beta)-d}{2h^{-1}(\alpha)}\right) \\ &+ \left(1-h^{-1}(\alpha)\right)h\left(\frac{1}{2}+\frac{d-(1-h^{-1}(\beta))}{2(1-h^{-1}(\alpha))}\right) \end{aligned} \tag{22}$$

for $h^{-1}(\beta)-h^{-1}(\alpha)\le d\le h^{-1}(\beta)+h^{-1}(\alpha)$, and $w_d(\alpha,\beta)=-\infty$ otherwise. Since the values of $d\in[0,1]$ for which $j-i+nd$ is an even integer become arbitrarily dense as n grows, by continuity of $d\mapsto w_d(\alpha,\beta)$, we have that

$$\begin{aligned} E(\mathbb{S}_{nh^{-1}(\alpha)},\mathbb{S}_{nh^{-1}(\beta)},\rho) = \min_{0\le d\le 1}\Bigg(&2-\log(1+\rho) \\ &- w_d(\alpha,\beta) - d\log\left(\frac{1-\rho}{1+\rho}\right)\Bigg) + o(1). \end{aligned} \tag{23}$$

Proposition 1: For large ρ we have

$$\begin{aligned} &\overline{E}(\alpha,\alpha,\rho) \le (1-\alpha) \\ &+ \frac{\frac{1}{2}-\sqrt{h^{-1}(\alpha)\left(1-h^{-1}(\alpha)\right)}}{\ln 2}(1-\rho)+o(1-\rho). \end{aligned} \tag{24}$$

Proof. Let $0<\alpha\le 1$. We establish the claim by evaluating $P_{XY}(A\times B)$ for $A=B=\mathbb{S}_{nh^{-1}(\alpha)}$ and $\rho=1-\epsilon$. By (23), it holds that

$$\begin{aligned} &E\left(\mathbb{S}_{nh^{-1}(\alpha)},\mathbb{S}_{nh^{-1}(\alpha)},1-\epsilon\right) \\ &= \min_d\left(2-\log(2-\epsilon)-w_d(\alpha,\alpha)+d\log\left(\frac{2-\epsilon}{\epsilon}\right)\right) \\ &+o(1) \\ &= 1+\frac{\epsilon}{2}\log(e) - \max_d\left(w_d(\alpha,\alpha)-d\log\left(\frac{2}{\epsilon}\right)+d\frac{\epsilon}{2}\log(e)\right) \\ &+o(\epsilon)+o(1). \end{aligned} \tag{25}$$

Denoting $r=r_\alpha=h^{-1}(\alpha)$, we have that

$$w_d(\alpha,\alpha) = h(r)+r\cdot h\left(\frac{d/2}{r}\right)+(1-r)\cdot h\left(\frac{d/2}{1-r}\right). \tag{26}$$

[1] In the notation of Section III, our work leverages the inequality $\|T_\rho f\|_{q_0}\le\|f\|_q$ among all support-constrained functions f (with the best possible q), whereas the work [22] uses the inequality $\|T_\rho f\|_{q_0}\le e^{-n\lambda}\|f\|_{1+(q_0-1)\rho^2}$ with the largest possible λ, which depends on the support size of f.

$$\frac{1}{n}\log\binom{n}{i} = h(h^{-1}(\alpha)) + o(1), \tag{18}$$

$$\frac{1}{n}\log\binom{i}{\frac{1}{2}(j+i-nd)} = h^{-1}(\alpha)h\left(\frac{\frac{1}{2}h^{-1}(\alpha)+h^{-1}(\beta)-d}{h^{-1}(\alpha)}\right) + o(1), \tag{19}$$

$$\frac{1}{n}\log\binom{n-i}{\frac{1}{2}(j-i+nd)} = \left(1-h^{-1}(\alpha)\right)h\left(\frac{\frac{1}{2}h^{-1}(\beta)-h^{-1}(\alpha)+d}{1-h^{-1}(\alpha)}\right). \tag{20}$$

The function $d \mapsto w_d(\alpha,\alpha) - d\log\left(\frac{2}{\epsilon}\right) + d\frac{\epsilon}{2}\log(e)$ is concave and its derivative

$$\frac{1}{2}\log\left(\frac{1-\frac{d/2}{r}}{\frac{d/2}{r}}\right) + \frac{1}{2}\log\left(\frac{1-\frac{d/2}{1-r}}{\frac{d/2}{1-r}}\right) - \log\left(\frac{2}{\epsilon}\right) + \frac{\epsilon}{2}\log(e) \tag{27}$$

equals zero at $d^* = \epsilon\sqrt{r(1-r)} + o(\epsilon)$. Thus, the optimizing d in (25) is $d^* = \epsilon\sqrt{r(1-r)} + o(\epsilon)$, and therefore

$$\begin{aligned} &E\left(\mathbb{S}_{nh^{-1}(\alpha)}, \mathbb{S}_{nh^{-1}(\alpha)}, 1-\epsilon\right) = 1-h(r) \\ &+\frac{\epsilon}{2}\log(e) + \epsilon\sqrt{r(1-r)} + \epsilon\log\left(\frac{1}{\epsilon}\right)\sqrt{r(1-r)} \\ &-\left[r\cdot h\left(\sqrt{\frac{1-r}{r}}\frac{\epsilon}{2}\right) + (1-r)\cdot h\left(\sqrt{\frac{r}{1-r}}\frac{\epsilon}{2}\right)\right] \\ &+o(\epsilon)+o(1) \end{aligned} \tag{28}$$

We approximate the term in the square brackets in equations (29), (30) and (31) at the bottom of the page. Substituting (31) into (28), we obtain

$$\begin{aligned} &E\left(\mathbb{S}_{nh^{-1}(\alpha)}, \mathbb{S}_{nh^{-1}(\alpha)}, 1-\epsilon\right) = 1-h(r) \\ &+\left(\frac{1}{2}-\sqrt{r(1-r)}\right)\epsilon\log(e) + o(\epsilon) + o(1). \end{aligned} \tag{32}$$

The claim now follows by definition of $\overline{E}(\alpha,\alpha,\rho)$. ■

Proposition 2: For small ρ we have that

$$\begin{aligned} \underline{E}(\alpha,\beta,\rho) \geq (1-\alpha) + (1-\beta) \\ +\rho\log e\left(1-2h^{-1}(\alpha)*h^{-1}(\beta)\right) + o(\rho). \end{aligned} \tag{33}$$

Proof. We establish the claim by evaluating $P_{XY}(A\times B)$ for $A = \mathbb{S}_{nh^{-1}(\alpha)}$ and $B = 1^n + \mathbb{S}_{nh^{-1}(\beta)}$, i.e., a zero-centered Hamming sphere and a Hamming sphere centered around the all-ones vector 1^n. First, note that for any $A, B \subset \{0,1\}^n$ it holds that

$$W_d(A, 1^n + B) = W_{1-d}(A, B). \tag{34}$$

Thus, applying (23), we see that for $0 < \alpha \leq \beta \leq 1$ it

---

$$\begin{aligned} r\cdot h\left(\sqrt{\frac{1-r}{r}}\frac{\epsilon}{2}\right) &= -\sqrt{r(1-r)}\frac{\epsilon}{2}\log\left(\sqrt{\frac{1-r}{r}}\frac{\epsilon}{2}\right) - r\left(1-\sqrt{\frac{1-r}{r}}\frac{\epsilon}{2}\right)\log\left(1-\sqrt{\frac{1-r}{r}}\frac{\epsilon}{2}\right) \\ &= -\sqrt{r(1-r)}\frac{\epsilon}{2}\log\left(\sqrt{\frac{1-r}{r}}\frac{\epsilon}{2}\right) + \frac{\epsilon}{2}\sqrt{r(1-r)}\log(e) + o(\epsilon),, \end{aligned} \tag{29}$$

$$\begin{aligned} (1-r)\cdot h\left(\sqrt{\frac{r}{1-r}}\frac{\epsilon}{2}\right) &= -\sqrt{r(1-r)}\frac{\epsilon}{2}\log\left(\sqrt{\frac{r}{1-r}}\frac{\epsilon}{2}\right) - (1-r)\left(1-\sqrt{\frac{r}{1-r}}\frac{\epsilon}{2}\right)\log\left(1-\sqrt{\frac{r}{1-r}}\frac{\epsilon}{2}\right) \\ &= -\sqrt{r(1-r)}\frac{\epsilon}{2}\log\left(\sqrt{\frac{r}{1-r}}\frac{\epsilon}{2}\right) + \frac{\epsilon}{2}\sqrt{r(1-r)}\log(e) + o(\epsilon), \end{aligned} \tag{30}$$

$$\begin{aligned} r\cdot h\left(\sqrt{\frac{1-r}{r}}\frac{\epsilon}{2}\right) + (1-r)\cdot h\left(\sqrt{\frac{r}{1-r}}\frac{\epsilon}{2}\right) &= -\sqrt{r(1-r)}\epsilon\log\left(\frac{\epsilon}{2}\right) + \epsilon\sqrt{r(1-r)}\log(e) + o(\epsilon) \\ &= \sqrt{r(1-r)}\epsilon\log\left(\frac{1}{\epsilon}\right) + \epsilon\sqrt{r(1-r)} + \epsilon\sqrt{r(1-r)}\log(e) + o(\epsilon). \end{aligned} \tag{31}$$

holds that

$$\begin{aligned}&E\left(\mathbb{S}_{nh^{-1}(\alpha)},1^n+\mathbb{S}_{nh^{-1}(\beta)},\rho\right)\\&=\min_{0\le d\le 1}\left(2-\log(1+\rho)-w_d(\alpha,\beta)\right.\\&\qquad\left.-(1-d)\log\left(\frac{1-\rho}{1+\rho}\right)\right)+o(1)\\&=2-\log(1-\rho)-\max_{0\le d\le 1}\left(w_d(\alpha,\beta)-d\log\left(\frac{1-\rho}{1+\rho}\right)\right)\\&+o(1)\end{aligned} \tag{35}$$

Let us consider the case of $\rho \ll 1$. In this case, we have that $\log(1+\rho)=\rho\log e+o(\rho)$, so that (35) reads

$$\begin{aligned}&E\left(\mathbb{S}_{nh^{-1}(\alpha)},1^n+\mathbb{S}_{nh^{-1}(\beta)},\rho\right)=2+\rho\log e\\&-\max_d\left(w_d(\alpha,\beta)+2d\rho\log e\right)+o(\rho)+o(1).\end{aligned} \tag{36}$$

The function $d\mapsto w_d(\alpha,\beta)\triangleq g(d)$ is strictly concave, and it is straightforward to verify that $g'(h^{-1}(\alpha)*h^{-1}(\beta))=0$ and that $g(h^{-1}(\alpha)*h^{-1}(\beta))=\alpha+\beta$. Denoting $c=2g''(h^{-1}(\alpha)*h^{-1}(\beta))<0$ and setting $\delta=d-h^{-1}(\alpha)*h^{-1}(\beta)$, we therefore have

$$g(d)=\alpha+\beta+c\delta^2+o(\delta^2). \tag{37}$$

Consequently,

$$\begin{aligned}&w_d(\alpha,\beta)+2d\rho\log e=g(d)+2d\rho\log e\\&=\alpha+\beta+c\delta^2+2(h^{-1}(\alpha)*h^{-1}(\beta)+\delta)\rho\log e+o(\delta^2)\\&=\alpha+\beta+\rho\log e\cdot 2h^{-1}(\alpha)*h^{-1}(\beta)\\&+\delta\left(2\rho\log e+c\delta+o(\delta)\right)\\&\le\alpha+\beta+\rho\log e\cdot 2h^{-1}(\alpha)*h^{-1}(\beta)+o(\rho),\end{aligned} \tag{38}$$

where the last inequality follows since $c<0$. Substituting (38) into (36) we obtain

$$\begin{aligned}&E\left(\mathbb{S}_{nh^{-1}(\alpha)},1^n+\mathbb{S}_{nh^{-1}(\beta),\beta}\right)\ge(1-\alpha)+(1-\beta)\\&+\rho\log e\left(1-2h^{-1}(\alpha)*h^{-1}(\beta)\right)+o(\rho)+o(1).\end{aligned} \tag{39}$$

The claim now follows by definition of $\underline{E}(\alpha,\beta,\rho)$. ■

III. Lower Bound on $\overline{E}(\alpha,\alpha,\rho)$

For a function $f:\{0,1\}^n\to\mathbb{R}^+$ and $p\ge 1$ we define $\|f\|_p=\mathbb{E}^{1/p}[|f(X)|^p]$. For a set $A\subset\{0,1\}^n$ denote

$$\mathbb{1}_A(x)\triangleq\begin{cases}0, & x\notin A\\1, & x\in A\end{cases}$$

We have that

$$\begin{aligned}P_{XY}(A\times B)&=\mathbb{E}\left[\mathbb{1}_A(X)\mathbb{1}_B(Y)\right]\\&=\mathbb{E}\left[\mathbb{1}_B(Y)\mathbb{E}\left[\mathbb{1}_A(X)|Y\right]\right]\\&=\mathbb{E}\left[\mathbb{1}_B(Y)(T_\rho\mathbb{1}_A)(Y)\right],\end{aligned} \tag{40}$$

where

$$(T_\rho f)(y)\triangleq\mathbb{E}[f(X)|Y=y]. \tag{41}$$

Denoting the inner-product $(f,g)=\mathbb{E}[f(Y)g(Y)]$ and noticing that T_ρ is self-adjoint and satisfies the semi-group property $T_{\rho_1}T_{\rho_2}=T_{\rho_1\rho_2}$ (for $0<\rho_1,\rho_2<1$), we obtain

$$\begin{aligned}P_{XY}(A\times B)&=(\mathbb{1}_B,T_\rho\mathbb{1}_A)\\&=(T_{\rho_1}\mathbb{1}_B,T_{\rho_2}\mathbb{1}_A)\qquad\forall\rho_1\rho_2=\rho &(42)\\&\le\|T_{\rho_1}\mathbb{1}_B\|_2\|T_{\rho_2}\mathbb{1}_A\|_2\,, &(43)\end{aligned}$$

where the last step is Cauchy-Schwarz inequality.

The next step is to use the hypercontractivity inequality to upper bound $\|T_\rho f\|_p$. Denote the support size of f by $\|f\|_0$. Since $\|f\|_0\ll 2^n$, we will use an improved hypercontractivity inequality from [1], that takes $\|f\|_0$ into account. The following result is a key ingredient:

Theorem 3 (Theorem 7 in [1]): Fix $1<p_0<\infty$ and $0\le\lambda_0\le(1-p_0^{-1})\ln 2$. For any $f:\{0,1\}^n\to\mathbb{R}_+$ with $\|f\|_{p_0}\ge e^{n\lambda_0}\|f\|_1$ we have

$$\|T_{e^{-t}}f\|_{p(t)}\le\|f\|_{p_0}\,,\qquad p(t)=1+e^{u(t)}\,, \tag{44}$$

where $u(t)$ is the unique solution on $[0,\infty)$ of the following ODE with initial condition $u(0)=\ln(p_0-1)$

$$u'(t)=C\left(\lambda_0(1+e^{-u(t)})\right) \tag{45a}$$

$$C(\ln 2(1-h(y)))=\frac{2-4\sqrt{y(1-y)}}{\ln 2(1-h(y))}. \tag{45b}$$

Furthermore, the function $C:[0,\ln 2]\to[2,2/\ln 2]$ is a smooth, convex and strictly increasing bijection.

From this result we derive the following implication for indicator functions.

Theorem 4: Fix $0<\alpha<1$ and $1<q_0<\infty$. Then there exists a function $q=q(t)$ defined on an interval $t\in[0,\epsilon)$ for some $\epsilon>0$ such that for all sets $A\subset\{0,1\}^n$ with $|A|\le 2^{n\alpha}$ we have

$$\|T_{e^{-t}}1_A\|_{q_0}\le\|1_A\|_{q(t)}\qquad\forall t\in[0,\epsilon)\,. \tag{46}$$

The function $q(t)$ satisfies

$$q(t)=q_0-(q_0-1)C((1-\alpha)\ln 2)t+O(t^2)\qquad\text{as } t\to 0\,. \tag{47}$$

Remark 2: Note that the standard hypercontractivity estimate [2], [3], [4], [5] yields the same result without restriction on the size of the set A but with a strictly worse (larger) function $q(t)=(q_0-1)e^{-2t}+1$. See [1, Remark 3].

Proof. Denote by $u_f(a,b,t)$ the solution of the ordinary differential equation (ODE)

$$\frac{d}{dt}u(t)=C(b(1+e^{-u(t)}))\,,$$

with $u(0) = a$. Here $C(\cdot)$ is a function defined in (60), $a \in \mathbb{R}$ and $0 < b < (1+e^{-a})^{-1}\ln 2$. For a fixed a, b the standard results on ODEs imply that this solution exists and is unique in some neighborhood $-\epsilon < t < \epsilon$ of zero. Furthermore, for any a_0, b_0 satisfying $0 < b_0 < (1+e^{-a_0})^{-1}\ln 2$ there exists an $\epsilon_1 > 0$ such that the map

$$(a, b, t) \mapsto u_f(a, b, t)$$

is smooth for $|a - a_0| < \epsilon_1, |b - b_0| < \epsilon_1, |t| < \epsilon_1$ (for both of these results, cf. [24, Chapter 2, Section 7, Corollary 6]. We set $a_0 = \ln(q_0 - 1)$ and $b_0 = (1-\alpha)(1-q_0^{-1})\ln 2$. We will call triplets (a, b, t) in the above neighborhood of $(a_0, b_0, 0)$ *admissible*.

From (44) we have for any admissible (a, b, s) with $s \geq 0$ and any A with $|A| \leq 2^{n\alpha}$:

$$\|T_{e^{-s}} 1_A\|_{1+e^{u_f(a,b,s)}} \leq \|1_A\|_{1+e^a}\,, \tag{48}$$

provided that $b(1+e^{-a}) \leq (1-\alpha)\ln 2$ (this is just the condition $\|f\|_{p_0} \geq e^{n\lambda_0}\|f\|_1$ of Theorem 3).

Our aim is to set $s = t$ in (48) and show that there exists a choice of $a = a(t)$ and $b = b(t)$ and $\epsilon < \epsilon_1$ such that the following conditions are satisfied: (C1) $a(0) = a_0$, $b(0) = b_0$ and both functions are smooth on $|t| < \epsilon$; (C2) for any $|t| < \epsilon$ the triplet $(a(t), b(t), t)$ is admissible; (C3) for each $|t| < \epsilon$

$$\begin{cases} b(t)(1+e^{-a(t)}) & = (1-\alpha)\ln 2 \\ u_f(a(t), b(t), t) & = \ln(q_0 - 1) \end{cases} \tag{49}$$

It is clear that if indeed such a choice of $a(t), b(t)$ were found we get from (48) with $s = t$ the statement of the Theorem with $q(t) = 1 + e^{a(t)}$.

We claim that it is sufficient to show that the system of equations

$$\begin{cases} f(a, b) & = 0, \\ u_f(a, b, t) & = \ln(q_0 - 1) \end{cases} \tag{50}$$

where $f(a, b) \triangleq b - (1-\alpha)(1 - (1+e^a)^{-1})\ln 2$, is uniquely solvable (for a, b) in the interval $-\epsilon < t < \epsilon$ and that solution $a(t), b(t)$ is smooth. Indeed, since the triplet $(a_0, b_0, 0)$ is a solution, we get (C1). Smoothness of $a(t), b(t)$ implies (C2). And, finally, (C3) is automatic. Smooth solvability, in turn, follows from the fact that the map

$$(a, b, t) \mapsto (f(a, b), u_f(a, b, t), t) \tag{51}$$

has non-trivial Jacobian at $(a_0, b_0, 0)$. Indeed, denoting $\partial_x = \frac{\partial}{\partial x}$ the Jacobian is given by

$$\mathrm{Jac}(a, b, t) = (\partial_a f)(\partial_b u_f) - (\partial_b f)(\partial_a u_f)\,.$$

To evaluate this we note an identity $u_f(a, b, 0) = a$ and thus

$$\left.\frac{\partial}{\partial a}\right|_{t=0} u_f(a, b, t) = 1, \tag{52}$$

$$\left.\frac{\partial}{\partial b}\right|_{t=0} u_f(a, b, t) = 0, \tag{53}$$

$$\left.\frac{\partial}{\partial t}\right|_{t=0} u_f(a, b, t) = C(b(1+e^{-a})). \tag{54}$$

Therefore, at $(a = a_0, b = b_0, t = 0)$ the Jacobian evaluates to

$$\mathrm{Jac}(a_0, b_0, 0) = -1 \neq 0\,.$$

Since the Jacobian is non-zero in some neighborhood of $(a_0, b_0, 0)$, the map (51) can be locally inverted, and we take for $a(t), b(t)$ the pre-image of $(0, 0, t)$ under (51).

Finally, we need to show that $q(t) = 1 + e^{a(t)}$ satisfies the expansion (47). To that end, we differentiate over t the identity

$$u_f(a(t), b(t), t) = \ln(q_0 - 1) \tag{55}$$

to get

$$\dot{a}(t)\partial_a u_f(a(t), b(t), t) + \dot{b}(t)\partial_b u_f(a(t), b(t), t) + \partial_t u_f(a(t), b(t), t) = 0 \tag{56}$$

where $\dot{a}(t) \triangleq \frac{da(t)}{dt}$ and $\dot{b}(t) \triangleq \frac{db(t)}{dt}$. At $t = 0$ this is evaluated via (52)-(54) to give

$$\dot{a}(0) + C((1-\alpha)\ln 2) = 0\,. \tag{57}$$

This clearly implies that $q(t) = 1 + e^{a(t)}$ satisfies (47). ■

The following application of the previous result establishes the hard direction of Theorem 1.

Proposition 3: Fix $\rho \in (0, 1)$. Then for any sets A, B with $|A| \leq 2^{n\alpha}$, $|B| \leq 2^{n\alpha}$ we have

$$P_{XY}(A \times B) \leq 2^{-n\psi(\alpha,\rho)}\,, \tag{58}$$

where as $\rho \to 1$ we have

$$\psi(\alpha, \rho) = (1-\alpha) + \frac{1}{\ln 2}(1/2 - \sqrt{h^{-1}(\alpha)(1-h^{-1}(\alpha))})(1-\rho) + o(1-\rho)\,. \tag{59}$$

Remark 3: For bounding $\overline{E}(\alpha, \beta, \rho)$ with $\alpha \neq \beta$ this method does not give a bound matching that attained by Hamming spheres. The main reason is that if we take A, B as concentric (but grossly unequal) Hamming balls the Cauchy-Schwarz inequality (43) is applied to functions $T_{\rho_1}\mathbb{1}_A$, $T_{\rho_2}\mathbb{1}_B$ which have effectively disjoint supports for $\rho \to 1$.

Proof. Let $\rho = e^{-2t}$ for some fixed t. Suppose the sets A, B both have sizes at most $2^{n\alpha}$. Then from Theorem 4 we obtain

$$\|T_{e^{-t}}\mathbb{1}_A\|_2 \le \|\mathbb{1}_A\|_{p(t)} \quad (60a)$$
$$\|T_{e^{-t}}\mathbb{1}_B\|_2 \le \|\mathbb{1}_B\|_{p(t)} \quad (60b)$$
$$p(t) = 2 - (2-1)C((1-\alpha)\ln 2)t + o(t). \quad (60c)$$

Since $\|\mathbb{1}_A\|_q = 2^{-n(1-\alpha)/q}$ we get from (43) the following:

$$\frac{1}{n}\log P_{X,Y}(A\times B) \le -\frac{2}{p(t)}(1-\alpha) \quad (61)$$
$$= -(1-\alpha)\left(1+\frac{t}{2}C((1-\alpha)\ln 2) + o(t)\right) \quad (62)$$
$$= -(1-\alpha) - \frac{t(1-\alpha)}{2}\frac{2-4\sqrt{h^{-1}(\alpha)(1-h^{-1}(\alpha))}}{(1-\alpha)\ln 2} + o(t). \quad (63)$$

The statement now follows since $t = \frac{1-\rho}{2} + o(1-\rho)$. ■

IV. Upper Bound on $\underline{E}(\alpha,\beta,\rho)$

Note that

$$P_{XY}(A\times B) = \sum_{a\in A, b\in B}\Pr(X=a, Y=b)$$
$$= |A|\cdot|B| \cdot \frac{1}{|A|\cdot|B|}\sum_{a\in A,b\in B} 2^{-n}\left(\frac{1+\rho}{2}\right)^n\cdot\left(\frac{1-\rho}{1+\rho}\right)^{d(a,b)}$$
$$\ge |A|\cdot|B| \cdot 2^{-n}\left(\frac{1+\rho}{2}\right)^n\cdot\left(\frac{1-\rho}{1+\rho}\right)^{\frac{1}{|A|\cdot|B|}\sum_{a\in A,b\in B} d(a,b)} \quad (64)$$
$$= 2^{-n\left(2-\frac{\log(|A|\cdot|B|)}{n} - \log(1+\rho) - \frac{\log\frac{1-\rho}{1+\rho}}{|A|\cdot|B|}\sum_{a\in A,b\in B}\frac{d(a,b)}{n}\right)}, \quad (65)$$

where we have used Jensen's inequality in (64). As $\frac{1-\rho}{1+\rho} < 1$, we need to upper bound $\frac{1}{|A|\cdot|B|}\sum_{a\in A,b\in B} d(a,b)$ in terms of $|A|$ and $|B|$ in order to further lower bound (65). Consequently, we define

$$\bar{d}(n,\alpha,\beta) = \frac{1}{n}\max_{A,B:|A|=2^{n\alpha},|B|=2^{n\beta}}\frac{1}{|A|\cdot|B|}\sum_{a\in A,b\in B} d(a,b) \quad (66)$$

$$\underline{d}(n,\alpha,\beta) = \frac{1}{n}\min_{A,B:|A|=2^{n\alpha},|B|=2^{n\beta}}\frac{1}{|A|\cdot|B|}\sum_{a\in A,b\in B} d(a,b). \quad (67)$$

With these definitions we relax (65) to

$$P_{XY}(A\times B) \ge 2^{-n\left((1-\alpha)+(1-\beta)-\log(1+\rho)-\bar{d}(n,\alpha,\beta)\log\frac{1-\rho}{1+\rho}\right)}. \quad (68)$$

It is obvious that $\bar{d}(n,\alpha,\beta) = 1-\underline{d}(n,\alpha,\beta)$, since if the sets (A,B) achieve the minimal average distance, the sets $(A, B' = 1^n + B)$ must achieve the maximal average distance. A quantity similar to $\underline{d}(n,\alpha,\beta)$, where the optimization in (66) is performed over all families A of size $2^{n\alpha}$ while $B = A$ was defined in [20, p.10 eq. 1], and its asymptotic (in n) value, was characterized in [25]. Below we prove a lower bound on $\underline{d}(n,\alpha,\beta)$. The technique is quite similar to that of [25], and requires the following simple proposition.

Proposition 4: The function $\varphi(x,y) = h^{-1}(x) * h^{-1}(y)$ is jointly convex in $(x,y)\in[0,1]^2$.

The function $\varphi(x,y)$ is plotted in Figure 1. To prove Proposition 4, we will rely on the following simpler statement, which is essentially proved in [25]. For completeness we provide the proof in the appendix.

Proposition 5: The function $x \mapsto h^{-1}(x)\left(1-h^{-1}(x)\right)$ is convex in $[0,1]$.

Proof of Proposition 4. Let (X,Y) be two (possibly dependent) random variables on $[0,1]^2$. We use the identity $a*b = \frac{1}{2}(1-(1-2a)(1-2b))$ to write

$$\mathbb{E}[\varphi(X,Y)] = \frac{1}{2}\left(1-\mathbb{E}\left[\left(1-2h^{-1}(X)\right)\left(1-2h^{-1}(Y)\right)\right]\right)$$
$$\ge \frac{1}{2}\left(1-\sqrt{\mathbb{E}\left[\left(1-2h^{-1}(X)\right)^2\right]} \quad (69)$$
$$\sqrt{\mathbb{E}\left[\left(1-2h^{-1}(Y)\right)^2\right]}\right) \quad (70)$$
$$\ge \frac{1}{2}\left(1-\sqrt{\left(1-2h^{-1}(\mathbb{E}[X])\right)^2}\sqrt{\left(1-2h^{-1}(\mathbb{E}[Y])\right)^2}\right) \quad (71)$$
$$= \varphi(\mathbb{E}[X],\mathbb{E}[Y]), \quad (72)$$

where (70) follows from the Cauchy-Schwarz inequality, and (71) from Jensen's inequality and the fact that $t\mapsto (1-2h^{-1}(t))^2 = 1-4h^{-1}(t)\left(1-h^{-1}(t)\right)$ is concave due to Proposition 5. ■

Lemma 1: For any two independent n-dimensional random binary vectors V and W

$$h^{-1}\left(\frac{H(V)}{n}\right) * h^{-1}\left(\frac{H(W)}{n}\right) \le \frac{\mathbb{E}d(V,W)}{n} \le 1-h^{-1}\left(\frac{H(V)}{n}\right) * h^{-1}\left(\frac{H(W)}{n}\right). \quad (73)$$

Proof. Let V and W be two independent random vectors with $H(V) = n\alpha$ and $H(W) = n\beta$. Further, let $a_i \triangleq$

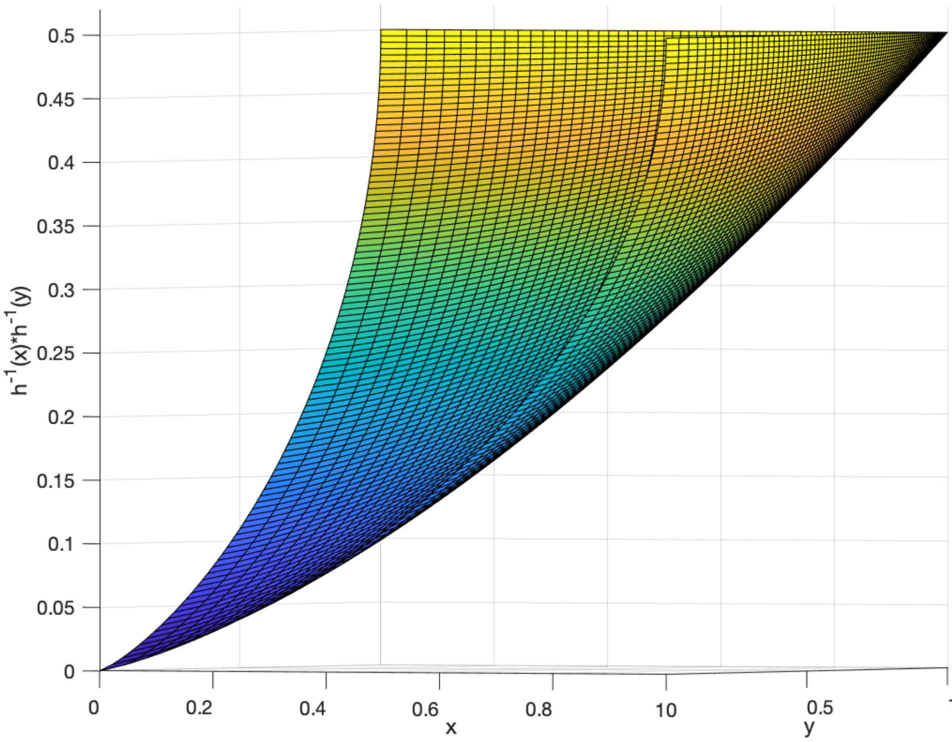

Fig. 1. Illustration of the function $h^{-1}(x) * h^{-1}(y)$.

$\Pr(V_i = 1)$, $b_i \triangleq \Pr(W_i = 1)$, be the induced marginal distributions for each coordinate. Our goal is to minimize and maximize $\sum_{i=1}^n a_i * b_i$ under the entropy constraints $H(V) = n\alpha, H(W) = n\beta$. We may and will assume without loss of generality that $a_i, b_i \leq 1/2$ for all i. We have

$$\inf_{\substack{V,W:\\ H(V)=n\alpha\\ H(W)=n\beta}} \sum_{i=1}^n a_i * b_i \geq \inf_{\substack{V,W:\\ H(V)\geq n\alpha\\ H(W)\geq n\beta}} \sum_{i=1}^n a_i * b_i$$

$$= \inf_{\substack{\{a_i\},\{b_i\}:\\ \sum_{i=1}^n h(a_i)\geq n\alpha\\ \sum_{i=1}^n h(b_i)\geq n\beta}} \sum_{i=1}^n a_i * b_i \tag{74}$$

$$= \inf_{\substack{\{\alpha_i\},\{\beta_i\}:\\ \frac{1}{n}\sum_{i=1}^n \alpha_i\geq \alpha\\ \frac{1}{n}\sum_{i=1}^n \beta_i\geq \beta}} \sum_{i=1}^n h^{-1}(\alpha_i) * h^{-1}(\beta_i) \tag{75}$$

where (74) follows since the cost function $\sum_{i=1}^n a_i * b_i$ depends only on the marginal distributions, and for every feasible distribution V, W the product of the marginalized distributions is also feasible. Our lower bound now immediately follows from Proposition 4. For the upper bound, note that if V and W minimize $\mathbb{E}d(V, W)$ under the entropy constraints, V and $W' = W + 1^n$ maximizes the expected distance under the same entropy constraints. ■

Taking $V \sim \text{Uniform}(A)$ and $W \sim \text{Uniform}(B)$, we immediately get the following.

Corollary 1:

$$\bar{d}(n, \alpha, \beta) \leq n\left(1 - h^{-1}(\alpha) * h^{-1}(\beta)\right), \tag{76}$$

$$\underline{d}(n, \alpha, \beta) \geq n h^{-1}(\alpha) * h^{-1}(\beta). \tag{77}$$

Combining (68) and Corollary 1, gives

$$\begin{aligned}\underline{E}(\alpha, \beta, \rho) &\leq (1-\alpha) + (1-\beta) - \log(1+\rho)\\ &\quad - \left(1 - h^{-1}(\alpha) * h^{-1}(\beta)\right)\log\frac{1-\rho}{1+\rho}\\ &= (1-\alpha) + (1-\beta) - \log(1-\rho)\\ &\quad + \left(h^{-1}(\alpha) * h^{-1}(\beta)\right)\log\frac{1-\rho}{1+\rho}.\end{aligned} \tag{78}$$

We have therefore obtained the following.

Proposition 6: We have

$$\begin{aligned}\underline{E}(\alpha, \beta, \rho) &\leq (1-\alpha) + (1-\beta)\\ &\quad + \rho\log(e)\left(1 - 2h^{-1}(\alpha) * h^{-1}(\beta)\right) + o(\rho).\end{aligned} \tag{79}$$

Remark 4: In [8] the bound

$$\underline{E}(\alpha,\beta,\rho) \le \frac{(1-\alpha)+(1-\beta)+2\rho\sqrt{(1-\alpha)(1-\beta)}}{1-\rho^2} \tag{80}$$

was proved, using reverse hypercontractivity. It is easy to verify that for $\alpha=\beta$ the bound (78) is strictly better than (80) for all $\alpha < 1-\frac{1-\rho}{2\rho}\log\left(\frac{1}{1-\rho}\right)$. Moreover, for any $0<\alpha,\beta<1$ the bound (78) is better than (80) for ρ large enough. The reverse hypercontractivity bound states that for $p<1$ we have $\|T_\rho f\|_{q(\rho,p)} \ge \|f\|_p$ where $q(\rho,p) = 1-\frac{1-p}{\rho^2} < p$ for $\rho<1$. The weakness of this bound in our setup is that the function $q(\rho,p)$ does not depend on the support of f, which is exponentially small. It is quite plausible that deriving support dependent reverse hypercontractivity bounds, analogous to the support dependent hypercontractivity bounds of [1], would result in tighter upper bounds on $\underline{E}(\alpha,\beta,\rho)$ in the high-correlation regime.

Acknowledgement

The authors would like to thank the anonymous reviewers and the associate editor for their excellent suggestions, and in particular for a simplification of the proof of Proposition 4.

Appendix

Let $\phi(x) = \left(1-2h^{-1}(x)\right)^2$. Since $h^{-1}(x)\left(1-h^{-1}(x)\right) = 1-\frac{\phi(x)}{4}$, it suffices to show that $x\mapsto\phi(x)$ is concave. We have

$$\begin{aligned}\phi'(x) &= -\frac{4}{\log\left(\frac{1-h^{-1}(x)}{h^{-1}(x)}\right)}\left(1-2h^{-1}(x)\right)\\ &= -\frac{4}{\log e}v(h^{-1}(x)),\end{aligned} \tag{81}$$

where

$$v(t) = \frac{1-2t}{\ln\left(\frac{1-t}{t}\right)}. \tag{82}$$

Showing that $x\mapsto\phi(x)$ is concave is equivalent to showing that $x\mapsto\phi'(x)$ is decreasing, which in turn is equivalent to showing that $t\mapsto v(t)$ is increasing in $(0,1/2)$, due to monotonicity of $x\mapsto h^{-1}(x)$. Thus, it remains to show that $v'(t)\ge 0$ for $t\in(0,1/2)$. Let $y=y_t=\frac{1-t}{t}\in(1,\infty)$. We have that $v'(t) = \frac{\frac{1-2t}{t(1-t)}-2\ln\left(\frac{1-t}{t}\right)}{\ln^2\left(\frac{1-t}{t}\right)}$ and since $\frac{1-2t}{t(1-t)} = \frac{y^2-1}{y}$, it suffices to show that $g(y)=\frac{y^2-1}{y}-2\ln(y)\ge 0$ for all $y>1$. Noting that $g(1)=0$ and $g'(y) = 1+\frac{1}{y^2}-\frac{2}{y} = \frac{(y-1)^2}{y^2}\ge 0$ for all $y\ge 1$, we see that indeed $g(y)\ge 0$ for all $y\ge 1$, which establishes our claim.

References

[1] Y. Polyanskiy and A. Samorodnitsky, "Improved log-Sobolev inequalities, hypercontractivity and uncertainty principle on the hypercube," *J. Functional Analysis*, 2019, to appear.

[2] E. Nelson, "A quartic interaction in two dimensions," in *Mathematical Theory of Elementary Particles*, R. Goodman and I. Segal, Eds. Cambridge, MA: M.I.T. Press, 1966.

[3] A. Bonami, "Čtude des coefficients de Fourier des fonctions de $l^p(g)$," *Annales de l'institut Fourier*, vol. 20, no. 2, pp. 335–402, 1970.

[4] W. Beckner, "Inequalities in Fourier analysis on $\mathbb{R}^n$," *Proceedings of the National Academy of Sciences*, vol. 72, no. 2, pp. 638–641, 1975.

[5] L. Gross, "Logarithmic Sobolev inequalities," *American Journal of Mathematics*, vol. 97, no. 4, pp. 1061–1083, 1975.

[6] R. O'Donnell, *Analysis of Boolean functions*. Cambridge University Press, 2014.

[7] C. Borell, "Positivity improving operators and hypercontractivity," *Math. Zeit.*, vol. 180, pp. 225–234, 1982.

[8] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov, "Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality," *Israel Journal of Mathematics*, vol. 154, no. 1, pp. 299–336, 2006.

[9] A. Bogdanov and E. Mossel, "On extracting common random bits from correlated sources," *IEEE Transactions on information theory*, vol. 57, no. 10, pp. 6351–6355, 2011.

[10] P. Austrin, P. Kaski, M. Koivisto, and J. Nederlof, "Sharper upper bounds for unbalanced uniquely decodable code pairs," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1368–1373, Feb 2018.

[11] B. Lindström, "Determination of two vectors from the sum," *Journal of Combinatorial Theory*, vol. 6, no. 4, pp. 402–407, 1969.

[12] H. van Tilborg, "An upper bound for codes in a two-access binary erasure channel (corresp.)," *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 112–116, 1978.

[13] T. Kasami and S. Lin, "Bounds on the achievable rates of block coding for a memoryless multiple-access channel," *IEEE Transactions on Information Theory*, vol. 24, no. 2, pp. 187–197, 1978.

[14] E. Weldon, "Coding for a multiple-access channel," *Information and Control*, vol. 36, no. 3, pp. 256–274, 1978.

[15] T. Kasami, S. Lin, V. Wei, and S. Yamamura, "Graph theoretic approaches to the code construction for the two-user multiple-access binary adder channel," *IEEE Transactions on Information Theory*, vol. 29, no. 1, pp. 114–130, 1983.

[16] R. Urbanke and Q. Li, "The zero-error capacity region of the 2-user synchronous BAC is strictly smaller than its shannon capacity region," in *Information Theory Workshop*, Jun 1998, p. 61.

[17] O. Ordentlich and O. Shayevitz, "An upper bound on the sizes of multiset-union-free families," *SIAM Journal on Discrete Mathematics*, vol. 30, no. 2, pp. 1032–1045, 2016.

[18] G. Ajjanagadde and Y. Polyanskiy, "Adder MAC and estimates for Rényi entropy," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2015, pp. 434–441.

[19] D. J. Kleitman, "On a combinatorial conjecture of Erdös," *Journal of Combinatorial Theory*, vol. 1, no. 2, pp. 209–214, 1966.

[20] R. Ahlswede and G. O. Katona, "Contributions to the geometry of Hamming spaces," *Discrete Mathematics*, vol. 17, no. 1, 1977.

[21] P. Frankl and Z. Füredi, "A short proof for a theorem of Harper about Hamming-spheres," *Discrete Mathematics*, vol. 34, no. 3, pp. 311 – 313, 1981.

[22] N. Kirshner and A. Samorodnitsky, "A moment ratio bound for polynomials and some extremal properties of Krawchouk polynomials and Hamming spheres," *arXiv preprint arXiv:1909.11929*, 2019.

[23] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[24] V. Arnold, *Ordinary Differential Equations*.
[25] R. Ahlswede and I. Althofer, “The asymptotic behavior of diameters in the average,” *Journal of Combinatorial Theory, Series B*, vol. 61, no. 2, pp. 167 – 177, 1994.

Or Ordentlich is a senior lecturer (assistant professor) in the School of Computer Science and Engineering at the Hebrew University of Jerusalem. He received the B.Sc. (cum laude), M.Sc. (summa cum laude), and Ph.D. degrees from Tel Aviv University, Israel, in 2010, in 2011, and 2016, respectively, all in electrical engineering. During the years 2015-2017 he was a postdoctoral fellow in the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology (MIT), and in the Department of Electrical and Computer Engineering at Boston University.

Yury Polyanskiy is an Associate Professor of Electrical Engineering and Computer Science and a member of IDSS and LIDS at MIT. Yury received M.S. degree in applied mathematics and physics from the Moscow Institute of Physics and Technology, Moscow, Russia in 2005 and Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ in 2010. His research interests span information theory, statistical learning, error-correcting codes, wireless communication and fault tolerance. Dr. Polyanskiy won the 2020 IEEE Information Theory Society James Massey Award, 2013 NSF CAREER award and 2011 IEEE Information Theory Society Paper Award.

Ofer Shayevitz received the B.Sc. degree from the Technion Institute of Technology, Haifa, Israel, in 1997 and the M.Sc. and Ph.D. degrees from the Tel-Aviv University, Tel Aviv, Israel, in 2004 and 2009, respectively, all in electrical engineering. He is currently an Associate Professor in the Department of EE - Systems at Tel Aviv University, and serves as the head of the Advanced Communication Center (ACC). Ofer’s research spans a wide cross-section of problems in information theory, statistical signal processing, and discrete mathematics. He is the recipient of the European Research Council (ERC) Starting Grant (2015–2020), and his group’s research activities have further been supported by Israel Science Foundation grants (2014–2022), and the Marie Curie Grant (2014–2018). Ofer is also actively involved in the Israeli hi-tech industry, and regularly consults to various startup companies. Before joining Tel Aviv University, Ofer was a postdoctoral fellow in the Information Theory and Applications (ITA) Center at the University of California, San Diego (2008 - 2011), and worked as a quantitative analyst with the D.E. Shaw group in New York (2011 - 2013). Prior to his graduate studies, he served as a digital communication engineer and team leader in the Israeli Defense Forces (1997 - 2003), and worked in statistical signal processing at CellGuide (2003 - 2004).