Entropy under additive Bernoulli and spherical noises

Entropy Under Additive Bernoulli and Spherical Noises

Or Ordentlich
Hebrew University of Jerusalem
or.ordentlich@mail.huji.ac.il

Yury Polyanskiy
MIT
yp@mit.edu

***Abstract*—Let Z^n be iid $\mathrm{Bernoulli}(\delta)$ and U^n be uniform on the set of all binary vectors of weight δn (Hamming sphere). As is well known, the entropies of Z^n and U^n are within $O(\log n)$. However, if X^n is another binary random variable independent of Z^n and U^n, we show that $H(X^n+U^n)$ and $H(X^n+Z^n)$ are within $O(\sqrt{n})$ and this estimate is tight. The bound is shown via coupling method. Tightness follows from the observation that the channels $x^n \mapsto x^n+U^n$ and $x^n \mapsto x^n+Z^n$ have similar capacities, but the former has zero dispersion. Finally, we show that despite the $\sqrt{n}$ slack in general, the Mrs. Gerber Lemma for $H(X^n+U^n)$ holds with only an $O(\log n)$ correction compared to its brethren for $H(X^n+Z^n)$.**

I. Introduction

This paper studies the difference between the output entropy of a binary symmetric channel with crossover probability $0<\delta<1/2$, and that of a binary additive channel with noise uniformly distributed over the sphere of vectors with Hamming weight δn. In particular, we are interested in the scaling of this difference with the blocklength n.

For $n\in\mathbb{N}$ and $0\le k\le n$ define the set

$$\mathcal{S}_{k,n}\triangleq\{x^n\in\{0,1\}^n\ :\ |x^n|=k\}, \tag{1}$$

where $|x^n|=\sum_{i=1}^n x_i$ is the Hamming weight of the vector x^n. For $0<\delta<1/2$, let $Z^n\sim\mathrm{Bernoulli}(\delta)^{\otimes n}$ and $U^n\sim\mathrm{Uniform}(S_{\delta n,n})$[1] be two n-dimensional random vectors. For a random vector X^n in $\{0,1\}^n$, statistically independent of Z^n and U^n, define the function[2]

$$\Psi_\delta(X^n)=H(X^n+Z^n)-H(X^n+U^n), \tag{2}$$

where $+$ corresponds to componentwise mod-2 addition. Our main goal is to characterize the asymptotic behavior in n of the quantities

$$\Psi^+_{n,\delta}\triangleq\sup_{X^n}\Psi_\delta(X^n)=\sup_{X^n}H(X^n+Z^n)-H(X^n+U^n),$$
$$\Psi^-_{n,\delta}\triangleq-\inf_{X^n}\Psi_\delta(X^n)=\sup_{X^n}H(X^n+U^n)-H(X^n+Z^n),$$
$$\Psi^*_{n,\delta}\triangleq\sup_{X^n}|\Psi_\delta(X^n)|=\sup_{X^n}|H(X^n+U^n)-H(X^n+Z^n)|,$$

where the supremum and infimum are w.r.t. all distributions on $\{0,1\}^n$. This problem is in line with some other work done on comparing "closeness" of the channels $x^n\mapsto x^n+Z^n$ and $x^n\mapsto x^n+U^n$ [1], [2].

Let $h(t)=-t\log(t)-(1-t)\log(1-t)$, $0\le t\le 1$ be the binary entropy function, and let $h^{-1}(\cdot)$ be its inverse restricted to the interval $[0,1/2]$. Among all random vectors $V^n\in\{0,1\}^n$ with $\mathbb{E}|V^n|\le n\delta$ the maximum of $H(V^n)$ is $nh(\delta)$ and this is attained by $V^n\sim\mathrm{Bernoulli}(\delta)^{\otimes n}$. In particular, taking X^n with a single mass point distribution, e.g., $X^n=0^n$, where 0^n is the all-zeros vector, we have that [3, Chapter 10, Lemma 7]

$$\begin{aligned}\Psi_\delta(0^n)&=H(Z^n)-H(U^n)\\&=nh(\delta)-\log|\mathcal{S}_{\delta n,n}|\\&\in\frac{1}{2}\log n+\left[\frac{1}{2}\log(2\pi\delta(1-\delta)),\frac{1}{2}\log(8\delta(1-\delta))\right].\end{aligned} \tag{3}$$

In light of this, one might be tempted to suspect that $\Psi_\delta(X^n)>0$ for all X^n, and that $\Psi^*_\delta=\theta(\log n)$. Nevertheless, we prove the following.

Theorem 1: For any $0<\delta<1/2$ we have that

$$c_1(\delta)\sqrt{n}+o(\sqrt{n})\le\Psi^-_{n,\delta}\le\sqrt{2\pi}\cdot c_1(\delta)\sqrt{n} \tag{4}$$
$$\frac{1}{2}\log n+c_3(\delta)\le\Psi^+_{n,\delta}\le c_2(\delta)\sqrt{n} \tag{5}$$
$$c_1(\delta)\sqrt{n}+o(\sqrt{n})\le\Psi^*_{n,\delta}\le c_2(\delta)\sqrt{n} \tag{6}$$

where

$$c_1(\delta)=\log\left(\frac{1-\delta}{\delta}\right)\sqrt{\frac{\delta(1-\delta)}{2\pi}}, \tag{7}$$
$$c_2(\delta)=4\log\left(\frac{1}{\delta}\right)\sqrt{\frac{h(\delta)(1-\delta)}{\delta}}, \tag{8}$$
$$c_3(\delta)=\frac{1}{2}\log\left(2\pi\delta(1-\delta)\right). \tag{9}$$

Proof. The upper bound in (4) follows from Proposition 2, stated and proved in Section III, whereas the lower bound in (4) follows from Lemma 5, stated and proved in Section IV. The upper bound in (5) follows from Corollary 1, stated and proved in Section III, and the lower bound trivially follows from (3). The bounds in (6) are deduced from (4) and (5). ■

Finding the correct scaling of $\Psi^+_{n,\delta}$ remains an open problem. It can be shown that the channels $x^n\mapsto x^n+Z^n$ and

The work of OO was supported by ISF under Grant 1791/17. The work of YP was supported in part by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370, and by the NSF grant CCF-17-17842.

[1]We assume throughout that $\delta n\in\mathbb{N}$.

[2]Throughout this paper $H(\cdot)$, $I(\cdot;\cdot)$ and $D(\cdot\|\cdot)$ denote Shannon entropy, mutual information and KL divergence, respectively. All logarithms and exponents are natural.

$x^n \mapsto x^n + U^n$ are not comparable in the less-noisy order. We currently do not know whether the latter channel is more capable than the former.

In Section V we prove the following variant of Mrs. Gerber's Lemma (MGL) [4] for additive channels with noise uniformly distributed on a sphere.

Theorem 2 (Spherical MGL): Let $U^n \sim \mathrm{Uniform}(\mathcal{S}_{\delta n,n})$ and let X^n be a binary n dimensional random vector, statistically independent of U^n. Then,

$$H(X^n + U^n) \geq nh\left(\delta * h^{-1}\left(\frac{H(X^n)}{n}\right)\right) - 8\frac{1-\delta}{\delta}\log n.$$

Note that the bound $H(X^n + U^n) \geq nh\left(\delta * h^{-1}\left(\frac{H(X^n)}{n}\right)\right) - c_2(\delta)\sqrt{n}$ trivially follows from Wyner and Ziv's MGL [4] and Theorem 1. Theorem 2 tightens the $O(\sqrt{n})$ gap to an $O(\log n)$ gap.

II. Preliminaries

A. Useful Information Inequalities

The two lemmas below will be useful in the derivations that follow. The proof of Lemma 1 is omitted, and Lemma 2 is proved in Appendix A.

Lemma 1: For all $x \in (0, 1/2]$ and $\epsilon \in [-x, 1-x]$ we have

$$h(x+\epsilon) \geq h(x) + \log\left(\frac{1-x}{x}\right)\epsilon - \frac{4h(x)}{x^2}\epsilon^2.$$

Lemma 2: Let A, B, C be random variables, $f(\cdot,\cdot)$ some $\mathcal{X}$-valued function, and $\bar{B}$ a copy of B independent of (A, B, C). Then

$$H(f(A,B)|C) - H(f(A,\bar{B})|C) \leq \gamma_1 \cdot \sqrt{I(A,C;B)} \quad (10)$$

$$H(f(A,\bar{B})|C) - H(f(A,B)|C) \leq \gamma_2 \cdot \sqrt{I(A,C;B)} \quad (11)$$

where $\gamma_1 = \sqrt{2}\log|\mathcal{X}|$ and $\gamma_2 = -\frac{1}{\sqrt{2}}\log\min_x \Pr[f(A,\bar{B}) = x]$.

B. General Facts About Spherical Noise

Let $U^n \sim \mathrm{Uniform}(\mathcal{S}_{\delta n,n})$. Clearly, $\Pr(U_m = 1) = \delta$ for any $m \in [n]$. Let us define the random variables

$$A_m \triangleq \Pr(U_m = 1 | U_1^{m-1}),\ m = 2, \ldots, n. \quad (12)$$

A_m is a deterministic function of the random variable $W_{m-1} = |U_1^{m-1}|$. In particular,

$$\begin{aligned} A_m &= \frac{\delta n - W_{m-1}}{n-(m-1)} \\ &= \frac{\delta(n-(m-1)) + (\delta(m-1) - W_{m-1})}{n-(m-1)} \\ &= \delta - \frac{T_{m-1}}{n-(m-1)}, \end{aligned} \quad (13)$$

where

$$T_m \triangleq W_m - \delta m = W_m - \mathbb{E}(W_m). \quad (14)$$

By definition, $\mathbb{E}(T_m) = 0$. Moreover,

$$\begin{aligned} &\mathbb{E}(T_m^2) = \mathrm{Var}(W_m) \\ &= \sum_{1\leq i,j\leq m} \mathrm{Cov}(U_i, U_j) \quad (15) \\ &= m\delta(1-\delta) \\ &+ m(m-1)\left(\Pr(U_1=1, U_2=1) - \Pr(U_1=1)\Pr(U_2=1)\right) \\ &= m\delta(1-\delta) + m(m-1)\left(\delta\frac{n\delta-1}{n-1} - \delta^2\right) \\ &= m\delta(1-\delta) + m(m-1)\left(\delta\frac{\delta(n-1)-(1-\delta)}{n-1} - \delta^2\right) \\ &= m\delta(1-\delta)\left(1 - \frac{m-1}{n-1}\right) \\ &= m\delta(1-\delta)\left(\frac{n-m}{n-1}\right). \quad (16) \end{aligned}$$

Note that (16) implies that

$$\begin{aligned} \frac{\mathbb{E}(T_m^2)}{(n-m)^2} &= \delta(1-\delta)\frac{m}{(n-m)(n-1)} \\ &\leq \frac{\delta(1-\delta))}{n-m}, \quad \forall m = 1, \ldots, n-1. \end{aligned} \quad (17)$$

Lemma 3: Let $U^n \sim \mathrm{Uniform}(\mathcal{S}_{\delta n,n})$. For any $m = 2, \ldots, n$ it holds that

$$I(U_m; U_1^{m-1}) \leq \frac{c(\delta)}{n-(m-1)}. \quad (18)$$

where

$$c(\delta) \triangleq \frac{4h(\delta)(1-\delta)}{\delta}. \quad (19)$$

Proof. We have that

$$\begin{aligned} I(U_m; U_1^{m-1}) &= H(U_m) - H(U_m|U_1^{m-1}) \\ &= h(\delta) - \mathbb{E}h(A_m). \end{aligned}$$

By (13) and Lemma 1,

$$\begin{aligned} \mathbb{E}h(A_m) &= \mathbb{E}h\left(\delta - \frac{T_{m-1}}{n-(m-1)}\right) \\ &\geq h(\delta) + \log\left(\frac{1-\delta}{\delta}\right)\mathbb{E}\left(\frac{T_{m-1}}{n-(m-1)}\right) \\ &- \frac{4h(\delta)}{\delta^2}\mathbb{E}\left(\frac{T_{m-1}}{n-(m-1)}\right)^2 \\ &\geq h(\delta) - \frac{4h(\delta)}{\delta^2}\frac{\delta(1-\delta))}{n-(m-1)}, \end{aligned} \quad (20)$$

where we have used the fact that $\mathbb{E}(T_{m-1}) = 0$, and (17) in the last inequality. ■

III. Upper bounds on Entropy Difference

Before proving our $O(\sqrt{n})$ upper bound $|H(X^n + Z^n) - H(X^n + U^n)|$, we show how to obtain an easier, though slightly weaker, upper bound of $O(\sqrt{n}\log n)$. This bound is based on the coupling technique proposed in [5, Section 4]. In particular, [5, Proposition 8], specialized to the n-dimensional binary space, gives the following.

Proposition 1 ([5, Proposition 8]): Let A^n and B^n be random vectors on $\{0,1\}^n$, and let

$$\bar{d}(A^n, B^n) = \frac{1}{n} \inf \mathbb{E}|A^n + B^n|, \tag{21}$$

where the infimum is over all joint distributions $P_{A^n B^n}$ with marginals P_{A^n} and P_{B^n}. Then

$$|H(A^n) - H(B^n)| \le nh\left(\bar{d}(A^n, B^n)\right). \tag{22}$$

Since we are interested in $|H(X^n+Z^n) - H(X^n+U^n)|$, it suffices to find a good coupling for U^n and Z^n. To this end, we generate the random vectors U^n, Z^n as follows:

- Let Π be a uniform random permutation on $[n] = \{1, \dots, n\}$.
- Let $W \sim \text{Binomial}(n, \delta)$.
- Set $U_{\Pi(i)} = 1$ for $i = 1, \dots, \delta n$ and $U_{\Pi(i)} = 0$ for $i = \delta n + 1, \dots, n$.
- Set $Z_{\Pi(i)} = 1$ for $i = 1, \dots, W$ and $Z_{\Pi(i)} = 0$ for $i = W + 1, \dots, n$.

Clearly U^n and Z^n have the correct marginal distributions. Moreover, $|Z^n + U^n| = |W - \delta n| = |W - \mathbb{E}(W)|$. Thus,

$$\begin{aligned}\mathbb{E}|U^n + Z^n| &= \mathbb{E}\sqrt{(W - \mathbb{E}(W))^2} \\ &\le \sqrt{\text{Var}(W)} \\ &= \sqrt{n\delta(1-\delta)},\end{aligned}$$

where the inequality follows by the concavity of $x \mapsto \sqrt{x}$. We have therefore obtained that

$$\bar{d}(X^n + Z^n, X^n + U^n) = \bar{d}(Z^n, U^n) \le \sqrt{\frac{\delta(1-\delta)}{n}}. \tag{23}$$

Now applying Proposition 1, we see that for any random vector X^n on $\{0,1\}^n$ it holds that

$$\begin{aligned}|H(X^n + Z^n) - H(X^n + U^n)| &\le nh\left(\sqrt{\frac{\delta(1-\delta)}{n}}\right) \\ &\le \sqrt{n\delta(1-\delta)} \cdot \log\left(\frac{n}{\delta(1-\delta)}\right). \end{aligned} \tag{24}$$

Our goal is to improve the $O(\sqrt{n}\log n)$ bound to $O(\sqrt{n})$. First, one-sided improvement is easy:

Proposition 2: Let $U^n \sim \text{Uniform}(\mathcal{S}_{\delta n, n})$, $Z^n \sim \text{Bernoulli}(\delta)^{\otimes n}$ and $1 \le m \le n$. For any random vector $X^m \perp\!\!\!\perp (Z^n, U^n)$, supported on $\{0,1\}^m$ we have that

$$H(X^m + U^m) - H(X^m + Z^m) \le c_1(\delta)\sqrt{\frac{m^2}{n}}. \tag{25}$$

where $c_1(\delta)$ is as defined in (7).

Proof. Following the idea of [5, (14)] we get

$$H(X^m + U^m) - H(X^m + Z^m) \le \left[\log \frac{Q(X^m + Z^m)}{Q(X^m + U^m)}\right],$$

where we denoted $Q = P_{X^m + Z^m}$. As noticed in [6, (58)] this distribution is "smooth", in the sense that $x^m \mapsto \log \frac{1}{Q(x^m)}$ is a Lipschitz function (with respect to the Hamming distance) with Lipschitz constant bounded by $\log \frac{1-\delta}{\delta}$. Consequently,

$$\left[\log \frac{Q(X^m + Z^m)}{Q(X^m + U^m)}\right] \le \log \frac{1-\delta}{\delta} \mathbb{E}|U^m + Z^m|.$$

The proposition follows by using the coupling constructed in (23), which is symmetric, and therefore satisfies

$$\mathbb{E}|X^m + U^m| = \frac{m}{n}\mathbb{E}|X^n + U^n|. \tag{26}$$

■

To make improvement in the other direction, we will make a crucial observation that while $P_{X^n + U^n}$ is not "smooth" in the strong sense used above, it is still rather nice. For example, the atoms of $P_{X_m + U_m | X^{m-1} + U^{m-1} = y^{m-1}}$ are almost in the interval $(\delta, 1-\delta)$ since $P_{X_m + U_m | X^{m-1} + U^{m-1}} \approx P_{X_m + Z_m | X^{m-1} + U^{m-1}}$ (as a consequence of (18)). We proceed to the details.

Lemma 4: Let $U^n \sim \text{Uniform}(\mathcal{S}_{\delta n, n})$, $Z^n \sim \text{Bernoulli}(\delta)^{\otimes n}$ and $1 \le m \le n$. For any random vector $X^m \perp\!\!\!\perp (Z^n, U^n)$, supported on $\{0,1\}^m$ we have that

$$H(X^m + Z^m) - H(X^m + U^m) \le \tilde{c}_2(\delta) a_m, \tag{27}$$

where $a_1 = 0$, $a_m = \sum_{2 \le k \le m} \frac{1}{\sqrt{n-(k-1)}}$, and $\tilde{c}_2(\delta) = \sqrt{\frac{c(\delta)}{2}} \log \frac{1}{\delta}$, and $c(\delta)$ is defined in Lemma 3.

Substituting $m = n$ above, we obtain the following bound as an immediate corollary.

Corollary 1: Let $U^n \sim \text{Uniform}(\mathcal{S}_{\delta n, n})$, $Z^n \sim \text{Bernoulli}(\delta)^{\otimes n}$ and $1 \le m \le n$. For any random vector $X^n \perp\!\!\!\perp (Z^n, U^n)$, supported on $\{0,1\}^n$ we have that

$$H(X^n + Z^n) - H(X^n + U^n) \le c_2(\delta)\sqrt{n}, \tag{28}$$

where $c_2(\delta)$ is as defined in (8).

Proof of Lemma 4. For $m = 1$ the claim trivially holds as $Z_i \overset{d}{=} U_i$ for all $i \in [n]$. Proceed by induction and suppose we established the case for $m-1$ and arbitrary X^{m-1}. Since $H(X_m + U_m) = H(X_m + Z_m)$, we have from the chain rule

$$\begin{aligned}&H(X^m + Z^m) - H(X^m + U^m) \\ &= H(X^{m-1} + Z^{m-1} | X_m + Z_m) \\ &\quad - H(X^{m-1} + U^{m-1} | X_m + U_m)\end{aligned} \tag{29}$$

Since $Z_m \perp\!\!\!\perp (Z^{m-1}, U^{m-1})$ we have from the induction hypothesis

$$\begin{aligned}&H(X^{m-1} + Z^{m-1} | X_m + Z_m) \\ &\quad - H(X^{m-1} + U^{m-1} | X_m + Z_m) \le \tilde{c}_2(\delta) a_{m-1}.\end{aligned} \tag{30}$$

Thus, subtracting and adding $H(X^{m-1} + U^{m-1} | X_m + Z_m)$ from the RHS of (29), and defining $V^{m-1} = X^{m-1} + U^{m-1}$ to lighten notation, we get

$$\begin{aligned}&H(X^m + Z^m) - H(X^m + U^m) \\ &\le \tilde{c}_2(\delta) a_{m-1} + H(V^{m-1} | X_m + Z_m) - H(V^{m-1} | X_m + U_m).\end{aligned}$$

Using $H(X_m + Z_m) = H(X_m + U_m)$ and the chain rule once again, we have that

$$\begin{aligned}&H(V^{m-1} | X_m + Z_m) - H(V^{m-1} | X_m + U_m) \\ &= H(X_m + Z_m | V^{m-1}) - H(X_m + U_m | V^{m-1}).\end{aligned} \tag{31}$$

Note that Z_m can be thought of as an independent copy of U_m, and hence from Lemma 2 with $f(\cdot,\cdot) = \cdot + \cdot \mod 2$ we can bound the right-hand side of (31) by $\sqrt{\frac{I(U_m;X_m,V^{m-1})}{2}}\log\frac{1}{\delta}$ (we used the fact that $A + Z_m$ for independent $A \perp\!\!\!\perp Z_m$ has distribution with atoms in $[\delta, 1-\delta]$). By $I(U_m; X_m, V^{m-1}) \le I(U_m; U^{m-1})$, together with (18) we get

$$\begin{aligned} &H(X^m + Z^m) - H(X^m + U^m) \\ &\qquad \le \tilde{c}_2(\delta) a_{m-1} + \frac{\tilde{c}_2(\delta)}{\sqrt{n-(m-1)}} \end{aligned} \tag{32}$$

as required. ■

IV. Lower bound on $\Psi^-_{n,\delta}$ via Random Coding

In this section, we prove the following lemma.

Lemma 5: Let $U^n \sim \mathrm{Uniform}(\mathcal{S}_{\delta n,n})$, and $Z^n \sim \mathrm{Bernoulli}(\delta)^{\otimes n}$ for some $0 < \delta < 1/2$. There exists a random vector $X^n \perp\!\!\!\perp (Z^n, U^n)$, supported on $\{0,1\}^n$, such that

$$H(X^n + U^n) - H(X^n + Z^n) \ge c_1(\delta)\sqrt{n} + o(\sqrt{n}), \tag{33}$$

where $c_1(\delta)$ is as defined in (7).

The proof of Lemma 5 is based on the simple observation that the dispersion of the $x^n \mapsto x^n + U^n$ channel is zero.

Proposition 3: Let $0 < \delta < 1/2$ and let $U^n \sim \mathrm{Uniform}(\mathcal{S}_{\delta n,n})$. There exists a code $\mathcal{C} \subset \{0,1\}^n$ with rate $R = \log 2 - h(\delta) - \frac{\log n}{n}$, that achieves maximal error probability $P_e \le n^{-1}$ over the n-dimensional additive noise channel $x^n \mapsto x^n + U^n$.

Proof of Proposition 3. Let $\mathcal{C} \subset \{0,1\}^n$ be a rate R code, $X^n \sim \mathrm{Uniform}(\mathcal{C})$, and consider the maximum-likelihood decoder for X^n from the output of the channel $Y^n = X^n + U^n$: it constructs $\mathcal{L}(Y^n) = \{x^n \in \mathcal{C} \ : \ |Y^n - x^n| = \delta n\}$ and outputs $\hat{X}^n \sim \mathrm{Uniform}(\mathcal{L}(Y^n))$ (note that $\mathcal{L}(Y^n)$ is never empty). Enumerate the codewords in $\mathcal{C}$ as $x_1^n, \ldots, x_{e^{nR}}^n$. The error probability, given that $X^n = x_1^n$ is

$$\begin{aligned} \Pr(\hat{X}^n \ne x_1^n | X^n = x_1^n) &\le \Pr\left(|\mathcal{L}(x_1^n + U^n)| > 1\right) \\ &= \Pr\left((\mathcal{C} \setminus \{x_1^n\}) \cap (x_1^n + U^n + \mathcal{S}_{\delta n,n}) \ne \emptyset\right) \\ &\le \sum_{i=2}^{e^{nR}} \Pr(x_i^n \in (x_1^n + U^n + \mathcal{S}_{\delta n,n})). \end{aligned} \tag{34}$$

Let us now assume $\mathcal{C}$ is a random codebook drawn from the standard (n, R) binary linear code ensemble [7, Chapter 6.2], and average (34), which gives

$$\begin{aligned} &\mathbb{E}_{\mathcal{C}} \Pr(\hat{X}^n \ne X_1^n | X^n = X_1^n) \\ &\le \mathbb{E}_{\mathcal{C}} \sum_{i=2}^{e^{nR}} \Pr(X_i^n \in (X_1^n + U^n + S_{\delta n,n})) \\ &= \left(e^{nR} - 1\right) \frac{|\mathcal{S}_{\delta n,n}|}{2^n} \\ &\le e^{-n(\log 2 - R - h(\delta))}. \end{aligned} \tag{35}$$

Taking $R = \log 2 - h(\delta) - \frac{\log n}{n}$, we see that the average error probability of the ensemble is at most n^{-1}, and there must exist a code with error probability no greater than the average. Since the code is linear, the error probability is the same for all codewords, and therefore the result holds for maximal error probability. ■

Proof of Lemma 5. Let $\mathcal{C}$ be a code satisfying the conditions of Proposition 3, and let $X^n \sim \mathrm{Uniform}(\mathcal{C})$. In order to establish a lower bound on $H(X^n + U^n) - H(X^n + Z^n)$, we will lower bound the first term and upper bound the second.

We have

$$\begin{aligned} H(X^n + U^n) &= I(X^n + U^n; X^n) + H(U^n) \\ &= H(X^n) + H(U^n) - H(X^n | X^n + U^n) \\ &\ge nR + nh(\delta) - \frac{1}{2}\log(8\delta(1-\delta)n) - nP_e - \log 2 \qquad (36) \\ &\ge nR + nh(\delta) - \frac{1}{2}\log(32\delta(1-\delta)n) - n \cdot \frac{1}{n} \\ &= n\log 2 - \frac{3}{2}\log n - \frac{1}{2}\log(32e\delta(1-\delta)). \qquad (37) \end{aligned}$$

where (36) follows from (3) and from Fano's inequality.

Next, we turn to upper bound $H(X^n + Z^n)$. Let $W = W(Z^n) = |Z^n|/n$ be the normalized Hamming weight of Z^n, and denote $a \wedge b = \min\{a, b\}$. For any X^n we have

$$\begin{aligned} H(X^n + Z^n) &= H(X^n + Z^n | W) + I(W; X^n + Z^n) \\ &\le H(X^n + Z^n | W) + \log n \\ &\le \mathbb{E}_W\left(H(X^n, Z^n | W = w) \bigwedge n\log 2\right) + \log n \\ &= \mathbb{E}_W\left(H(X^n) + H(Z^n | W = w) \bigwedge n\log 2\right) + \log n \\ &\le \mathbb{E}_W\left(H(X^n) + nh(W) \bigwedge n\log 2\right) + \log n. \end{aligned} \tag{38}$$

For $X^n \sim \mathrm{Uniform}(\mathcal{C})$, the bound (38) reads

$$\begin{aligned} H(X^n + Z^n) &\le n\Big(\log 2 - h(\delta) \\ &\qquad + \mathbb{E}_W\left(h(W) \wedge h(\delta)\right)\Big) + \log n \\ &\le n\Bigg(\log 2 - h(\delta) \\ &\qquad + \mathbb{E}_W\left(h(\delta) + h'(\delta)(W - \delta) \wedge h(\delta)\right)\Bigg) + \log n \\ &= n\log 2 - nh'(\delta) \cdot \mathbb{E}_W\left[\delta - W\right]^+ + \log n \\ &\le n\log 2 - \log\left(\frac{1-\delta}{\delta}\right)\sqrt{\frac{\delta(1-\delta)}{2\pi}}\sqrt{n} + \log n + c, \end{aligned} \tag{39}$$

where $[a]^+ = \max\{a, 0\}$, and the last inequality follows from the central limit theorem [8] for some universal constant c. The claim now follows from combining (37) and (39). ■

V. Proof of Spherical MGL

This section is devoted to proving Theorem 2. We define the function $f_\alpha(t) = h(\alpha * h^{-1}(t))$, where $a * b = a(1-b) + b(1-a)$. Recall that $t \mapsto f_\alpha(t)$ is convex [4].

Proof of Theorem 2. We have

$$\begin{aligned}
H(X^n+U^n) &= \sum_{m=1}^{n} H(X_m+U_m|X_1^{m-1}+U_1^{m-1}) \\
&\geq \sum_{m=1}^{n} H(X_m+U_m|X_1^{m-1},U_1^{m-1}) \\
&= \sum_{m=1}^{n} \mathbb{E}h\left(\Pr(X_m=1|X_1^{m-1}) * \Pr(U_m=1|U_1^{m-1})\right) \\
&= \sum_{m=1}^{n} \mathbb{E}h\left(h^{-1}\left(H(X_m|X_1^{m-1}=x)\right) * A_m)\right) \\
&\geq \sum_{m=1}^{n} \mathbb{E}h\left(h^{-1}\left(H(X_m|X_1^{m-1})\right) * A_m)\right), \qquad (40)
\end{aligned}$$

where the random variable A_m was defined in (13), and in the last inequality we have used the convexity of $f_\alpha(t)$ and the fact that the random variables A_m and $H(X_m|X_1^{m-1}=x)$ are statistically independent.

Let $\eta_m \triangleq H(X_m|X_1^{m-1})$. Recalling that $A_m = \delta - \frac{T_{m-1}}{n-(m-1)}$ and that $a*(b+c) = a*b+c(1-2a)$, we can now rewrite (40) as

$$\begin{aligned}
H(X^n+U^n) &\geq \sum_{m=1}^{n} \mathbb{E}h\left(h^{-1}(\eta_m) * \left(\delta - \frac{T_{m-1}}{n-(m-1)}\right)\right) \\
&= \sum_{m=1}^{n} \mathbb{E}h\left(h^{-1}(\eta_m) * \delta - \frac{T_{m-1}}{n-(m-1)}\left(1-2h^{-1}(\eta_m)\right)\right).
\end{aligned}$$

Applying Lemma 1 on each term in the sum, with $x_m = h^{-1}(\eta_m)*\delta$ and $\epsilon_m = \frac{T_{m-1}}{n-(m-1)}(1-2h^{-1}(\eta_m))$, we obtain

$$H(X^n+U^n) \geq \sum_{m=1}^{n} h\left(h^{-1}(\eta_m)*\delta\right) \qquad (41)$$

$$+ \sum_{m=1}^{n}\left(1-2h^{-1}(\eta_m)\right)\log\left(\frac{1-h^{-1}(\eta_m)*\delta}{h^{-1}(\eta_m)*\delta}\right)\cdot\frac{\mathbb{E}(T_{m-1})}{n-(m-1)} \qquad (42)$$

$$- \sum_{m=1}^{n} 4\left(1-2h^{-1}(\eta_m)\right)^2 \frac{h\left(h^{-1}(\eta_m)*\delta\right)}{\left(h^{-1}(\eta_m)*\delta\right)^2}\cdot\frac{\mathbb{E}(T_{m-1}^2)}{\left(n-(m-1)\right)^2} \qquad (43)$$

$$\geq nh\left(\delta * h^{-1}\left(\frac{H(X^n)}{n}\right)\right) - \sum_{m=1}^{n} 4\left(1-2h^{-1}(\eta_m)\right)^2 \frac{h\left(h^{-1}(\eta_m)*\delta\right)}{\left(h^{-1}(\eta_m)*\delta\right)^2}\cdot\frac{\delta(1-\delta)}{n-(m-1)}, \qquad (44)$$

where in the last inequality we have used the convexity of $f_\alpha(t)$ to lower bound (41), the fact that $\mathbb{E}T_m = 0$ for all $m = 0,\ldots,n-1$ in order to null the term in (42), and (17) in order to lower bound the term in (43). Noting further that

$$\left(1-2h^{-1}(\eta_m)\right)^2 \frac{h\left(h^{-1}(\eta_m)*\delta\right)}{\left(h^{-1}(\eta_m)*\delta\right)^2} \leq \frac{1}{\delta^2},$$

we can further bound (44) as

$$\begin{aligned}
H(X^n+U^n) &\geq nh\left(\delta * h^{-1}\left(\frac{H(X^n)}{n}\right)\right) - \frac{4(1-\delta)}{\delta}\sum_{m=1}^{n}\frac{1}{n-(m-1)} \\
&= nh\left(\delta * h^{-1}\left(\frac{H(X^n)}{n}\right)\right) - \frac{4(1-\delta)}{\delta}\sum_{m=1}^{n}\frac{1}{m} \\
&\geq nh\left(\delta * h^{-1}\left(\frac{H(X^n)}{n}\right)\right) - \frac{8(1-\delta)}{\delta}\log n.
\end{aligned}$$

as desired. ■

APPENDIX A
PROOF OF LEMMA 1

Consider two distributions P, Q on some alphabet $\mathcal{X}$, then we have

$$H(P) - H(Q) \leq \sqrt{2D(P\|Q)}\log|\mathcal{X}| \qquad (45)$$

$$H(Q) - H(P) \leq \sqrt{\frac{D(P\|Q)}{2}}\log\frac{1}{\min_x P(x)}, \qquad (46)$$

where the first inequality is from [5, (64)] and the second one from $H(Q) - H(P) \leq (\mathbb{E}_Q - \mathbb{E}_P)\log\frac{1}{P(X)} \leq \mathrm{TV}(P,Q)\log\frac{1}{\min_x P(x)}$ and Pinsker inequality.

Applying these with $P = P_{f(A,B)|C=c}$ and $Q = P_{f(A,\bar{B})|C=c}$ we get (after using Jensen's inequality on $\mathbb{E}_C[\sqrt{\cdot}]$):

$$H(f(A,B)|C) - H(f(A,\bar{B})|C) \leq \gamma_1\sqrt{D} \qquad (47)$$

$$H(f(A,\bar{B})|C) - H(f(A,B)|C) \leq \gamma_2\sqrt{D}, \qquad (48)$$

where by data-processing $D = D(P_{f(A,B)|C}\|P_{f(A,\bar{B})|C}|P_C) \leq D(P_{A,B|C}\|P_{A|C}P_{\bar{B}}|P_C) = I(A,C;B)$

REFERENCES

[1] Y. Polyanskiy, "Hypercontractivity of spherical averages in Hamming space," *Arxiv preprint arXiv:1309.3014*, 2013.

[2] A. W. Harrow, A. Kolla, and L. J. Schulman, "Dimension-free l_2 maximal inequality for spherical means in the hypercube," *arXiv preprint arXiv:1209.4148*, 2012.

[3] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*. Elsevier, 1977.

[4] A. 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, Nov 1973.

[5] Y. Polyanskiy and Y. Wu, "Wasserstein continuity of entropy and outer bounds for interference channels," *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 3992–4002, July 2016.

[6] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability," vol. 60, no. 1, pp. 5–21, Jan. 2014.

[7] R. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, Inc., 1968.

[8] V. V. Petrov, "Limit theorems of probability theory: sequences of independent random variables," 1995.