How to Quantize n Outputs of a Binary Symmetric Channel to $n-1$ Bits?

Wasim Huleihel and Or Ordentlich

***Abstract*—Suppose that Y^n is obtained by observing a uniform Bernoulli random vector X^n through a binary symmetric channel with crossover probability α. The "most informative Boolean function" conjecture postulates that the maximal mutual information between Y^n and any Boolean function $\mathrm{b}(X^n)$ is attained by a dictator function. In this paper, we consider the "complementary" case in which the Boolean function is replaced by $f:\{0,1\}^n \to \{0,1\}^{n-1}$, namely, an $n-1$ bit quantizer, and show that $I(f(X^n);Y^n) \leq (n-1)\cdot(1-h(\alpha))$ for any such f. Thus, in this case, the optimal function is of the form $f(x^n)=(x_1,\ldots,x_{n-1})$.**

I. Introduction

Let X^n be an n-dimensional binary vector uniformly distributed over $\{0,1\}^n$, and let Y^n be the output of passing X^n through a binary symmetric channel (BSC) with crossover probability $\alpha \in [0,1/2]$. In other words, $Y^n = X^n \oplus Z^n$, where Z^n is a sequence of n independent and identically distributed (i.i.d.) Bernoulli(α) random variables, statistically independent of X^n. The following conjecture [1] have recently received considerable attention.

Conjecture 1: For any *Boolean* function $\mathrm{b}:\{0,1\}^n \to \{0,1\}$, we have $I(\mathrm{b}(X^n);Y^n) \leq 1-h(\alpha)$, where $h(\alpha) \triangleq -\alpha\log_2\alpha-(1-\alpha)\log_2(1-\alpha)$ is the binary entropy function.

Since the dictator function $\mathrm{b}(X^n)=X_i$ (for any $1 \leq i \leq n$) achieves this upper bound with equality, then intuitively Conjecture 1 postulates that the dictator function is the most "informative" one-bit quantization of X^n in terms of achieving the maximal $I(\mathrm{b}(X^n);Y^n)$. Clearly, by the symmetry of the pair (X^n,Y^n) we have that for any function $I(\mathrm{b}(X^n);Y^n) = I(X^n;\mathrm{b}(Y^n))$, so we can equivalently think of the problem at hand as seeking the optimal one-bit quantizer of n outputs of the channel. Despite attempts in various directions [1]–[7], Conjecture 1 remains open in general. However, for the "very noisy" case, where $\alpha > 1/2-\delta$, for some $\delta > 0$ independent of n, the validity of the conjecture was established by Samorodnitsky [8].

In this paper, we consider the "complementary" case in which the Boolean function in Conjecture 1 is replaced by an $n-1$ bit quantizer. Our main result is the following.

Theorem 1: For any function $f:\{0,1\}^n \to \{0,1\}^{n-1}$ we have

$$I(f(X^n);Y^n) \leq (n-1)\cdot(1-h(\alpha))\,, \tag{1}$$

and this bound is attained with equality by, e.g., $f(x^n) = (x_1,\ldots,x_{n-1})$.

One may wonder whether for any $f:\{0,1\}^n \to \{0,1\}^k$ we have $I(f(X^n);Y^n) \leq k\cdot(1-h(\alpha))$. However, for $k=Rn$ with $0<R<1$, the problem essentially reduces to remote source coding under log-loss distortion measure, for which the maximal value of $I(f(X^n);Y^n)/n$ (as a function of R,α) can be determined up to $o(n)$ terms. Indeed, [3], [9] characterizes this quantity which turns out to be greater than $R\cdot(1-h(\alpha))$. Conjecture 1 as well as Theorem 1 deal with the extreme cases of $k=1$ and $k=n-1$, respectively, where neglecting the $o(n)$ terms leads to non-informative characterization of the maximal $I(f(X^n);Y^n)$, and therefore [3], [9] do not suffice.

Theorem 1 can be generalized to a stronger statement concerning the entire class of binary-input memoryless output-symmetric (BMS) channels.

Definition 1 (BMS channels): A memoryless channel with binary input X and output Y is called *binary-input memoryless output-symmetric (BMS)* if there exists a sufficient statistic $g(Y)=(X\oplus Z_T,T)$ for X, where (T,Z_T) are statistically independent of X, and Z_T is a binary random variable with $\Pr(Z_T=1|T=t)=t$.

Corollary 1 ([10]): Let X^n be an n-dimensional binary vector uniformly distributed over $\{0,1\}^n$, and let Y^n be the output of passing X^n through a BMS with capacity C. Then for every $f:\{0,1\}^n \to \{0,1\}^{n-1}$, we have

$$I(f(X^n);Y^n) \leq (n-1)\cdot C,$$

and this bound is attained with equality by, e.g., $f(x^n) = (x_1,\ldots,x_{n-1})$.

Proof of Corollary 1: Let W be a BMS channel with capacity $C=1-h(\alpha)$. Let Y^n_W and Y^n_{BSC} be the outputs corresponding to the channel W and a BSC with crossover probability α, respectively, when the input to both channels is X^n. Define $[n] \triangleq \{1,2,\ldots,n\}$. For any function $f:\{0,1\}^n \to [M]$, we can write

$$\begin{aligned}
I(f(X^n);Y^n_W) &= I(f(X^n),X^n;Y^n_W) - I(X^n;Y^n_W|f(X^n))\\
&= I(X^n;Y^n_W) + I(f(X^n);Y^n_W|X^n) - I(X^n;Y^n_W|f(X^n))\\
&= I(X^n;Y^n_W)\\
&\quad - \sum_{m=1}^{M}\Pr(f(X^n)=m)I(X^n;Y^n_W|f(X^n)=m).
\end{aligned}$$

WH and OO are with the Department of Electrical Engineering and Computer Science, MIT. e-mail: {wasimh,ordent}@mit.edu. OO is also with the Department of Electrical and Computer Engineering, Boston University.

The work of WH and OO was supported by the MIT - Technion Postdoctoral Fellowship.

We proceed by noting that $I(X^n;Y^n_W) = I(X^n;Y^n_{\text{BSC}}) = nC$ as the capacity achieving input distribution of both channels is $\text{Bernoulli}(1/2)$. Furthermore, recall the fact that the BSC is the least capable among all BMS channels with the same capacity [11, page 116], [12, Lemma 7.1]. To wit, for any input U^n, the corresponding outputs of W and the BSC will satisfy $I(U^n;Y^n_{\text{BSC}}) \leq I(U^n;Y^n_W)$. This implies that

$$I(X^n;Y^n_W|f(X^n)=m) \geq I(X^n;Y^n_{\text{BSC}}|f(X^n)=m),$$

for all $m = 1,\dots,M$. Thus, we get that for any function f,

$$I(f(X^n);Y^n_W) \leq I(f(X^n);Y^n_{\text{BSC}}). \tag{2}$$

The corollary now follows by invoking Theorem 1. ■

II. Proof of Theorem 1

Since the vector Y^n is uniformly distributed over $\{0,1\}^n$, we have

$$I(f(X^n);Y^n) = n - H(Y^n|f(X^n)). \tag{3}$$

Our goal is therefore to lower bound $H(Y^n|f(X^n))$.

Consider the function $f : \{0,1\}^n \to [2^{n-1}]$, and define the sets

$$f^{-1}(j) \triangleq \{x^n \in \{0,1\}^n : \ f(x^n) = j\}\,,\ j = 1,\dots,2^{n-1},$$

which form a disjoint partition of $\{0,1\}^n$. Further, define the sizes of these sets as

$$m_j \triangleq \left|f^{-1}(j)\right| = \sum_{x^n\in\{0,1\}^n} \mathbb{1}\{f(x^n)=j\}\,,\ j = 1,\dots,2^{n-1},$$

and assume without loss of generality that $m_j > 0$, for all j. To see why this assumption is valid, first note that there must exist some i, for which $m_i \geq 2$. Let $f^{-1}(i) = \{x^n_{i_1},\dots,x^n_{i_{m_i}}\}$. Now if there exists some $j \neq i$, such that $m_j = 0$, we can define a new function $\tilde{f} : \{0,1\}^n \to [2^{n-1}]$ where $\tilde{f}^{-1}(i) = \{x^n_{i_1},\dots,x^n_{i_{m_i-1}}\}$, $\tilde{f}^{-1}(j) = \{x^n_{i_{m_i}}\}$, and $\tilde{f}^{-1}(t) = f^{-1}(t)$, for all $t \neq i,j$. For this function we must have

$$\begin{aligned} H\left(Y^n|f(X^n)\right) &\geq H\left(Y^n|f(X^n), \mathbb{1}\left\{X^n = x_{i_{m_i}}\right\}\right) \\ &= H(Y^n|\tilde{f}(X^n)), \end{aligned}$$

and consequently $I(f(X^n);Y^n) \leq I(\tilde{f}(X^n);Y^n)$.

Next, for every $m = 0,1,\dots,2^n$ define the quantity

$$\lambda(m) \triangleq \sum_{j=1}^{2^{n-1}} \mathbb{1}\{m_j = m\}\,, \tag{4}$$

which counts the number of sets $f^{-1}(j)$ with cardinality m, in the partition induced by the function f.[1] The next proposition expresses $\lambda(1)$ in terms of $\{\lambda(m)\}_{m\geq 2}$.

Proposition 1: For any $f : \{0,1\}^n \to [2^{n-1}]$ with $m_j > 0$ for all j, we have that

$$\lambda(1) = \sum_{m\geq 3} (m-2)\lambda(m). \tag{5}$$

Intuitively, this proposition states that since the average size of the sets $f^{-1}(j)$ is 2, then every set $f^{-1}(j)$ of cardinality $m > 2$, must be compensated for by $(m-2)$ sets of cardinality 1.

Proof: Using the definition of $\lambda(m)$ in (4), and the fact that $\{f^{-1}(j)\}$ forms a disjoint partition of $\{0,1\}^n$, we have

$$\sum_{m=0}^{2^n} \lambda(m) = 2^{n-1}, \tag{6}$$

$$\sum_{m=0}^{2^n} m\lambda(m) = 2^n. \tag{7}$$

Multiplying (6) by 2 and equating it with the left-hand side of (7), we get

$$\sum_{m=0}^{2^n} 2\lambda(m) = \sum_{m=0}^{2^n} m\lambda(m),$$

which implies

$$2\lambda(0) + \lambda(1) = \sum_{m\geq 3} (m-2)\lambda(m).$$

Invoking our assumption that $\lambda(0) = 0$ gives the desired result. ■

Definition 2 (Minimal entropy of a noisy subset): For a family of vectors $S \subset \{0,1\}^n$ let U_S be a random vector uniformly distributed over S, and let Z^n be a sequence of n i.i.d. $\text{Bernoulli}(\alpha)$ random variables, statistically independent of U_S. For $m = 1,\dots,2^n$, we define the quantity

$$H^n_m(\alpha) \triangleq \min_{S\subset\{0,1\}^n \,:\, |S|=m} H(U_S \oplus Z^n). \tag{8}$$

Some properties of $H^n_m(\alpha)$ will be studied in the next section. In particular, we will prove the following lemma.

Lemma 1: For any $2 < m < 2^n$,

$$\frac{m-2}{2m-2} H^n_1(\alpha) + \frac{m}{2m-2} H^n_m(\alpha) \geq H^n_2(\alpha). \tag{9}$$

We can now write

$$\begin{aligned} H(Y^n|f(X^n)) &= \sum_{j=1}^{2^{n-1}} \Pr\left(f(X^n)=j\right) H\left(Y^n|f(X^n)=j\right) \\ &= \sum_{j=1}^{2^{n-1}} \Pr\left(X^n \in f^{-1}(j)\right) H\left(Y^n|X^n \in f^{-1}(j)\right) \\ &= 2^{-n} \sum_{j=1}^{2^{n-1}} \left|f^{-1}(j)\right| H\left(U_{f^{-1}(j)} \oplus Z^n\right) \\ &\geq 2^{-n} \sum_{j=1}^{2^{n-1}} m_j H^n_{m_j}(\alpha) \\ &= 2^{-n} \sum_{m=1}^{2^n} m\lambda(m) H^n_m(\alpha) \end{aligned}$$

[1] In fact, since we have already assumed that $m_j > 0$ for all j, we have that $\lambda(0) = 0$ and $\lambda(m) = 0$ for $m > 2^n - (2^{n-1} - 1)$.

$$= 2^{-n}\left(\lambda(1)H_1^n(\alpha) + 2\lambda(2)H_2^n(\alpha) + \sum_{m\geq 3}^{2^n} m\lambda(m)H_m^n(\alpha)\right)$$

$$= 2^{-n}\Bigg(2\lambda(2)H_2^n(\alpha) + \sum_{m\geq 3}^{2^n}(m-2)\lambda(m)H_1^n(\alpha) + m\lambda(m)H_m^n(\alpha)\Bigg) \qquad (10)$$

$$= 2^{-n}\Bigg(2\lambda(2)H_2^n(\alpha) + \sum_{m\geq 3}^{2^n}(2m-2)\lambda(m)\left[\frac{m-2}{2m-2}H_1^n(\alpha) + \frac{m}{2m-2}H_m^n(\alpha)\right]\Bigg)$$

$$\geq H_2^n(\alpha)\cdot 2^{-n}\sum_{m=1}^{2^n}(2m-2)\lambda(m) \qquad (11)$$

$$= H_2^n(\alpha), \qquad (12)$$

where in (10) follows from Proposition 1, in (11) we have used Lemma 1, and (12) follows from (6) and (7). Proposition 4, stated and proved in the next section, shows that $H_2^n(\alpha) = 1+(n-1)h(\alpha)$. Combining this with (3) and (12) establishes the desired result.

III. Properties of $H_m^n(\alpha)$

The main goal of this section is to prove Lemma 1. To this end, we establish some properties of the function $H_m^n(\alpha)$, which may be of independent interest.

Proposition 2 (Monotonicity in m): The function $H_m^n(\alpha)$ is monotonically non-decreasing as a function of m.

Proof: It is suffice to show that for any natural number $1 \leq m < 2^n$ it holds that $H_m^n(\alpha) \leq H_{m+1}^n(\alpha)$. To this end, let $S = \{s_1, \dots, s_{m+1}\} \subset \{0,1\}^n$ be a family of $m+1$ vectors, and let $S_{-i} \triangleq S \setminus \{s_i\}$, for $i = 1, \dots, m+1$. Clearly, $|S_{-i}| = m$ for all i. Furthermore, the random vector U_S can be generated by first drawing a random variable $A \sim \text{Uniform}([m+1])$ and then drawing a statistically independent random vector uniformly over S_{-A}. Thus, for any $S \subset \{0,1\}^n$ of size $m+1$ we have that

$$H(U_S \oplus Z^n) \geq H\left(U_S \oplus Z^n | A\right) = \frac{1}{m+1}\sum_{a=1}^{m+1} H\left(U_{S_{-a}} \oplus Z^n\right) \geq H_m^n(\alpha),$$

and in particular $H_m^n(\alpha) \leq H_{m+1}^n(\alpha)$. ■

We define the partial order "$\leq$" on the hypercube $\{0,1\}^n$ as $y \leq x$ iff $y_i \leq x_i$, for all $i = 1, \dots, n$.

Definition 3 (Monotone sets): A set $S \subset \{0,1\}^n$ is monotone if $x \in S$ implies $y \in S$, for all $y \leq x$.

Let $\mathcal{M}_m^n \triangleq \{S \subset \{0,1\}^n : \ |S| = m, \ S \text{ is monotone}\}$. We will prove the following result.

Lemma 2 (Sufficiency of monotone sets):

$$H_m^n(\alpha) = \min_{S \in \mathcal{M}_m^n} H(U_S \oplus Z^n).$$

Remark 1: Theorem 3 in [1] states that among all boolean functions, $I(\mathrm{b}(X^n); Y^n)$ is maximized by functions for which the induced set $\mathrm{b}^{-1}(0)$ is monotone.[2] While this statement is closely related to our Lemma 2, it does not imply it, although the proof technique is somewhat similar.

The proof of Lemma 2 is based on applying a procedure called *shifting* [13]–[15].

Definition 4 (Shifting): For a set of binary vectors $S \subset \{0,1\}^n$ the *shifting* procedure is defined as follows. For $i \in [n]$ and $x \in \{0,1\}^n$ write $x - i$ for the vector obtained by setting $x_i = 0$, and define

$$S_i \triangleq \{x \in S : \ x_i = 1, x - i \notin S\}.$$

Find the smallest i such that $S_i \neq \emptyset$. If there is no such i then we are done. Otherwise, replace S with the set $(S \setminus S_i) \cup (S_i - i)$, where $S_i - i \triangleq \{x - i \ : x \in S_i\}$, and repeat. The output of this process is a monotone set, denoted by S_{shifted}, with cardinality $|S_{\text{shifted}}| = |S|$.

The proof of Lemma 2 hinges on the following result.

Lemma 3: Let $S \subset \{0,1\}^n$ be some subset of vectors, and $\bar{S} \subset \{0,1\}^n$ be the result of applying one iteration of the shifting procedure, say, on the first coordinate. Let $P_{Y|X}$ be some discrete memoryless channel with binary input, and let Y^n be its output when the input is U_S and $\bar{Y}^n$ be its output when the input is $U_{\bar{S}}$. For every $\omega \in \mathcal{Y}^{n-1}$ we have that $\Pr(Y_2^n = \omega) = \Pr(\bar{Y}_2^n = \omega)$, and

$$\left|\Pr(U_{\bar{S},1} = 1 | \bar{Y}_2^n = \omega) - \frac{1}{2}\right| \geq \left|\Pr(U_{S,1} = 1 | Y_2^n = \omega) - \frac{1}{2}\right|.$$

Proof of Lemma 3: Let S_2^n be the projection of S onto the coordinates $\{2, \dots, n\}$, and note that the projection of $\bar{S}$ onto these coordinates is also S_2^n, as the shifting operations does not effect these coordinates. Consequently, $U_{S,2}^n$ and $\bar{U}_{S,2}^n$ have the same distribution, and therefore Y_2^n and $\bar{Y}_2^n$ have the same distribution.

Next, for any vector $\omega \in \mathcal{Y}^{n-1}$, we have

$$\begin{aligned}\Pr(U_{S,1} = 1 | Y_2^n = \omega) &= \sum_{x \in S_2^n} \Pr(U_{S,1} = 1, U_{S,2}^n = x | Y_2^n = \omega) \\ &= \sum_{x \in S_2^n} \Pr(U_{S,1} = 1 | U_{S,2}^n = x)\Pr(U_{S,2}^n = x | Y_2^n = \omega).\end{aligned}$$

The fact that $U_{S,2}^n$ and $U_{\bar{S},2}^n$ have the same distribution, implies that $P_{U_{S,2}^n|Y_2^n} = P_{U_{\bar{S},2}^n|\bar{Y}_2^n}$, and therefore

$$\begin{aligned}\Pr(U_{\bar{S},1} = 1 | \bar{Y}_2^n = \omega) &= \sum_{x \in S_2^n} \Pr(U_{\bar{S},1} = 1 | U_{\bar{S},2}^n = x)\Pr(U_{S,2}^n = x | Y_2^n = \omega).\end{aligned}$$

We partition the set S_2^n into three subsets:

[2] In fact, [1, Theorem 3] provides a stronger statement about the structure of the induced $\mathrm{b}^{-1}(0)$.

- $A \triangleq \{x \in S_2^n : \ [0\ x] \in S, [1\ x] \in S\}$
- $B \triangleq \{x \in S_2^n : \ [0\ x] \notin S, [1\ x] \in S\}$
- $C \triangleq \{x \in S_2^n : \ [0\ x] \in S, [1\ x] \notin S\}$

and we note that

$$\Pr(U_{S,1} = 1 | U_{S,2}^n = x) = \begin{cases} 1/2 & x \in A \\ 1 & x \in B \\ 0 & x \in C \end{cases}.$$

Letting

$$\begin{aligned} a_\omega &\triangleq \Pr(U_{S,2}^n \in A | Y_2^n = \omega), \\ b_\omega &\triangleq \Pr(U_{S,2}^n \in B | Y_2^n = \omega), \\ c_\omega &\triangleq \Pr(U_{S,2}^n \in C | Y_2^n = \omega), \end{aligned}$$

we get

$$\Pr(U_{S,1} = 1 | Y_2^n = \omega) = \frac{a_\omega}{2} + b_\omega.$$

By the definition of the shifting procedure in Definition 4,

$$\Pr(U_{\bar{S},1} = 1 | U_{\bar{S},2}^n = x) = \begin{cases} 1/2 & x \in A \\ 0 & x \in B \\ 0 & x \in C \end{cases}.$$

Thus,

$$\Pr(U_{\bar{S},1} = 1 | \bar{Y}_2^n = \omega) = \frac{a_\omega}{2}.$$

We can use this to see that $\Pr(U_{\bar{S},1} = 1 | \bar{Y}_2^n = \omega)$ is more biased than $\Pr(U_{S,1} = 1 | Y_2^n = \omega)$. Indeed

$$\begin{aligned} &\left(\frac{1}{2} - \Pr(U_{\bar{S},1} = 1 | \bar{Y}_2^n = \omega)\right)^2 \\ &\qquad - \left(\frac{1}{2} - \Pr(U_{S,1} = 1 | Y_2^n = \omega)\right)^2 \\ &= \left(\frac{1}{2}(1 - a_\omega)\right)^2 - \left(\frac{1}{2}(1 - a_\omega) - b_\omega\right)^2 \\ &= b_\omega(1 - a_\omega) - b_\omega^2 = b_\omega c_\omega \geq 0, \end{aligned}$$

as desired. ■

Corollary 2 (Shifting decreases output entropy): Let $S \subset \{0,1\}^n$ be some subset of vectors, and $\bar{S} \subset \{0,1\}^n$ be the result of applying one iteration of the shifting procedure, say, on the first coordinate. Let Z^n be a sequence of n i.i.d. Bernoulli(α) random variables, statistically independent of U_S and $U_{\bar{S}}$. Then,

$$H(U_{\bar{S}} \oplus Z^n) \leq H(U_S \oplus Z^n). \tag{13}$$

Proof: By the chain rule,

$$H(U_S \oplus Z^n) = H(U_{S,2}^n \oplus Z_2^n) + H(U_{S,1} \oplus Z_1 | U_{S,2}^n \oplus Z_2^n),$$

and

$$\begin{aligned} H(U_{\bar{S}} \oplus Z^n) &= H(U_{\bar{S},2}^n \oplus Z_2^n) + H(U_{\bar{S},1} \oplus Z_1 | U_{\bar{S},2}^n \oplus Z_2^n) \\ &= H(U_{S,2}^n \oplus Z_2^n) + H(U_{\bar{S},1} \oplus Z_1 | U_{\bar{S},2}^n \oplus Z_2^n) \end{aligned}$$

where the last equality follows from the fact that $P_{U_{S,2}^n \oplus Z_2^n} = P_{U_{\bar{S},2}^n \oplus Z_2^n}$ due to Lemma 3. Thus, it suffices to show that

$$H(U_{\bar{S},1} \oplus Z_1 | U_{\bar{S},2}^n \oplus Z_2^n) \leq H(U_{S,1} \oplus Z_1 | U_{S,2}^n \oplus Z_2^n).$$

For any $\omega \in \{0,1\}^{n-1}$ let $\alpha_\omega \triangleq \Pr(U_{S,1} = 1 | U_{S,2}^n \oplus Z_2^n = \omega)$ and $\beta_\omega \triangleq \Pr(U_{\bar{S},1} = 1 | U_{\bar{S},2}^n \oplus Z_2^n = \omega)$. Then, we get

$$\begin{aligned} &H(U_{\bar{S},1} \oplus Z_1 | U_{\bar{S},2}^n \oplus Z_2^n) \\ &= \sum_{\omega \in \{0,1\}^{n-1}} \Pr(U_{\bar{S},2}^n \oplus Z_2^n = \omega) h\left(\alpha * \beta_\omega\right) \\ &= \sum_{\omega \in \{0,1\}^{n-1}} \Pr(U_{S,2}^n \oplus Z_2^n = \omega) h\left(\alpha * \beta_\omega\right) \\ &\leq \sum_{\omega \in \{0,1\}^{n-1}} \Pr(U_{S,2}^n \oplus Z_2^n = \omega) h\left(\alpha * \alpha_\omega\right) \\ &= H(U_{S,1} \oplus Z_1 | U_{S,2}^n \oplus Z_2^n), \end{aligned} \tag{14}$$

where $a * b \triangleq a \cdot (1 - b) + (1 - a) \cdot b$ for any $a, b \in [0,1]$, the second equality follows since $P_{U_{S,2}^n \oplus Z_2^n} = P_{U_{\bar{S},2}^n \oplus Z_2^n}$, and the inequality is because β_ω is more biased than α_ω, by Lemma 3. ■

Applying Corollary 2 recursively, we see that for any $S \subset \{0,1\}^n$ we have

$$H\left(U_{S_{\text{shifted}}} \oplus Z^n\right) \leq H(U_S \oplus Z^n). \tag{15}$$

In fact, it is easy to extend the above argument to show that for any BMS channel with inputs U_S and $U_{S_{\text{shifted}}}$ and corresponding outputs Y^n and $\tilde{Y}^n$, respectively, we get $H(\tilde{Y}^n) \leq H(Y^n)$. Inequality (15) immediately establishes Lemma 2.

We now turn to finding $H_m^n(\alpha)$ for $m = 1, 2, 3, 4$.

Proposition 3: $H_1^n(\alpha) = n \cdot h(\alpha)$.

Proof: For any vector $u \in \{0,1\}^n$ we have that $H(u \oplus Z^n) = H(Z^n) = n \cdot h(\alpha)$. ■

Proposition 4: $H_2^n(\alpha) = 1 + (n-1) \cdot h(\alpha)$.

Proof: By Lemma 2, it is suffice to minimize $H(U_S \oplus Z^n)$ over $S \in \mathcal{M}_2^n$. It is easy to see that $\mathcal{M}_2^n$ consists of a single set $S^* = \{[1\ 0 \cdots\ 0], [0\ 0 \cdots\ 0]\}$, up to permuting the order of coordinates. Thus, direct calculation gives

$$H_2^n(\alpha) = H(U_{S^*} \oplus Z^n) = 1 + (n-1) \cdot h(\alpha). \tag{16}$$

■

Proposition 5:

$$\begin{aligned} H_3^n(\alpha) &= h\left(\frac{1}{3} * \alpha\right) + \left(\frac{2}{3} * \alpha\right) h\left(\frac{1 - \alpha^2}{2 - \alpha}\right) \\ &\quad + \left(\frac{1}{3} * \alpha\right) h\left(\frac{1 - \alpha + \alpha^2}{1 + \alpha}\right) + (n-2) h(\alpha) \qquad (17) \\ &\geq h\left(\frac{1}{3} * \alpha\right) + \frac{1}{3} h(\alpha) + \frac{2}{3} + (n-2) h(\alpha) \qquad (18) \end{aligned}$$

Proof: By Lemma 2, it is suffice to minimize $H(U_S \oplus Z^n)$ over $S \in \mathcal{M}_3^n$. It is easy to see that $\mathcal{M}_3^n$ consists of a single set $S^* = \{[1\ 0\ 0\ \cdots\ 0], [0\ 1\ 0\ \cdots\ 0], [0\ 0\ 0\ \cdots\ 0]\}$, up to permuting the order of coordinates. Thus, (17) is obtained by direct calculation of $H(U_{S^*} \oplus Z^n)$. To obtain the lower bound (18) we write

$$\begin{aligned}
&H_3^n(\alpha) = H(U_{S^*} \oplus Z^n) \\
&= H(U_{S_1^*} \oplus Z_1) + H(U_{S_2^*} \oplus Z_2 | U_{S_1^*} \oplus Z_1) + H\left(Z_3^n\right) \\
&\geq H(U_{S_1^*} \oplus Z_1) + H(U_{S_2^*} \oplus Z_2 | U_{S_1^*}) + H\left(Z_3^n\right) \\
&= h\left(\frac{1}{3} * \alpha\right) + \frac{1}{3} h(\alpha) + \frac{2}{3} + (n-2)h(\alpha).
\end{aligned}$$

■

Proposition 6: $H_4^n(\alpha) = 2 + (n-2) \cdot h(\alpha)$.

Proof: By Lemma 2, it is suffice to minimize $H(U_S \oplus Z^n)$ over $S \in \mathcal{M}_4^n$. It is easy to see that $\mathcal{M}_4^n$ consists of two sets

$$\begin{aligned}
\mathcal{C} \triangleq \{&[1\ 1\ 0\ \cdots\ 0], [1\ 0\ 0\ \cdots\ 0], \\
&[0\ 1\ 0\ \cdots\ 0], [0\ 0\ 0\ \cdots\ 0]\}, \\
\mathcal{B} \triangleq \{&[1\ 0\ 0\ 0\ \cdots\ 0], [0\ 1\ 0\ 0\ \cdots\ 0], \\
&[0\ 0\ 1\ 0\ \cdots\ 0], [0\ 0\ 0\ 0\ \cdots\ 0]\},
\end{aligned}$$

up to permuting the order of coordinates. In particular, $\mathcal{C}$ is the 2-dimensional cube padded by $(n-2)$ zeros, whereas $\mathcal{B}$ is the 3-dimensional Hamming ball of radius 1, padded by $n-3$ zeros. Thus,

$$H_4^n(\alpha) = \min\left\{H\left(U_{\mathcal{C}} \oplus Z^n\right), H\left(U_{\mathcal{B}} \oplus Z^n\right)\right\}.$$

It is easy to verify that $H\left(U_{\mathcal{C}} \oplus Z^n\right) = 2 + (n-2) \cdot h(\alpha)$. We show that $H\left(U_{\mathcal{B}} \oplus Z^n\right) \geq 2 + (n-2) \cdot h(\alpha)$. Indeed,

$$\begin{aligned}
&H\left(U_{\mathcal{B}} \oplus Z^n\right) \\
&= H\left(U_{\mathcal{B},1}^2 \oplus Z_1^2\right) + H\left(U_{\mathcal{B},3} \oplus Z_3 | U_{\mathcal{B},1}^2 \oplus Z_1^2\right) + H(Z_4^n) \\
&\geq H\left(U_{\mathcal{B},1}^2 \oplus Z_1^2\right) + H\left(U_{\mathcal{B},3} \oplus Z_3 | U_{\mathcal{B},1}^2\right) + (n-3) \cdot h(\alpha) \\
&= H\left(U_{\mathcal{B},1}^2 \oplus Z_1^2\right) + \frac{1}{2} + \frac{h(\alpha)}{2} + (n-3) \cdot h(\alpha). \qquad (19)
\end{aligned}$$

Direct calculation gives

$$H\left(U_{\mathcal{B},1}^2 \oplus Z_1^2\right) = \frac{3}{2} + \frac{h(\alpha)}{2}, \tag{20}$$

which together with (19) shows that $H\left(U_{\mathcal{B}} \oplus Z^n\right) \geq 2 + (n-2) \cdot h(\alpha)$. ■

We are now in a position to prove Lemma 1.

Proof of Lemma 1: For any $m \geq 4$ we have that $H_m^n(\alpha) \geq H_4^n(\alpha) > H_1^n(\alpha)$, which implies that

$$\begin{aligned}
\frac{m-2}{2m-2} H_1^n(\alpha) + \frac{m}{2m-2} H_m^n(\alpha) &\geq \frac{H_1^n(\alpha) + H_m^n(\alpha)}{2} \\
\geq \frac{H_1^n(\alpha) + H_4^n(\alpha)}{2} &= 1 + (n-1) \cdot h(\alpha) = H_2^n(\alpha).
\end{aligned}$$

It then remains to verify (9) for $m = 3$. Using the lower bound (18) for $H_3^n(\alpha)$, it suffices to verify that

$$\begin{aligned}
&\frac{1}{4} n h(\alpha) + \frac{3}{4}\left[h\left(\frac{1}{3} * \alpha\right) + \frac{1}{3} h(\alpha) + \frac{2}{3} + (n-2)h(\alpha)\right] \\
&\qquad \geq 1 + (n-1)h(\alpha), \qquad (21)
\end{aligned}$$

which is equivalent to

$$3 \cdot h\left((1/3) * \alpha\right) - 2 - h(\alpha) \geq 0. \tag{22}$$

Let $g(\alpha) \triangleq 3 \cdot h\left(\frac{1}{3} * \alpha\right) - 2 - h(\alpha)$. It is easy to check that $g(0) > 0$ and that $g(1/2) = 0$. Thus, it suffices to show that $g(\alpha)$ is monotonically decreasing as a function of α, namely, that $\mathrm{d}g(\alpha)/\mathrm{d}\alpha < 0$, for any $\alpha \in (0, 1/2)$. We have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\alpha} g(\alpha) &= -\log_2\left(\frac{\frac{1}{3} * \alpha}{\frac{2}{3} * \alpha}\right) + \log_2\left(\frac{\alpha}{1-\alpha}\right) \qquad (23) \\
&= \log_2\left(\frac{2\alpha - \alpha^2}{1-\alpha^2}\right), \qquad (24)
\end{aligned}$$

which is negative for all $\alpha \in (0, 1/2)$. ■

Acknowledgment

The authors are grateful to Yury Polyanskiy, Shlomo Shamai (Shitz), Ofer Shayevitz, and Omri Weinstein, for many discussions that helped prompt this work.

References

[1] T. Courtade and G. Kumar, "Which Boolean functions maximize mutual information on noisy inputs?" *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4515–4525, Aug 2014.

[2] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On hypercontractivity and the mutual information between Boolean functions," in *51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2013, pp. 13–19.

[3] V. Chandar and A. Tchamkerten, "Most informative quantization functions," in *Proc. ITA Workshop, San Diego, CA, USA*, Feb. 2014, available online http://perso.telecom-paristech.fr/ tchamker/CTAT.pdf.

[4] O. Ordentlich, O. Shayevitz, and O. Weinstein, "An improved upper bound for the most informative boolean function conjecture," in *IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 500–504.

[5] G. Kindler, R. O'Donnell, and D. Witmer, "Continuous analogues of the most informative function problem," 2015. [Online]. Available: http://arxiv.org/abs/1506.03167

[6] N. Weinberger and O. Shayevitz, "On the optimal boolean function for prediction under quadratic loss," in *IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 495–499.

[7] G. Pichler, G. Matz, and P. Piantanida, "A tight upper bound on the mutual information of two boolean functions," in *IEEE Information Theory Workshop (ITW)*, Sept 2016, pp. 16–20.

[8] A. Samorodnitsky, "On the entropy of a noisy function," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5446–5464, Oct 2016.

[9] E. Erkip and T. M. Cover, "The efficiency of investment information," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1026–1040, May 1998.

[10] Y. Polyanskiy, private communicaton.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1982.

[12] E. Sasoglu, "Polar coding theorems for discrete systems," Ph.D. dissertation, 2011.

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

[14] N. Alon, "On the density of sets of vectors," *Discrete Mathematics*, vol. 46, no. 2, pp. 199–202, 1983.

[15] P. Frankl, "On the trace of finite sets," *Journal of Combinatorial Theory, Series A*, vol. 34, no. 1, pp. 41 – 45, 1983.