

# A Lower Bound on the Expected Distortion of Joint Source-Channel Coding

Yuval Kochman  
School of CSE  
HUJI  
Jerusalem, Israel  
yuvalko@cs.huji.ac.il

Or Ordentlich  
School of CSE  
HUJI  
Jerusalem, Israel  
or.ordentlich@mail.huji.ac.il

Yury Polyanskiy  
Dept. EECS  
MIT  
Cambridge, MA, USA  
yp@mit.edu

**Abstract**—We consider the classic joint source-channel coding problem of transmitting a memoryless source over a memoryless channel. The focus of this work is on the rate of convergence of the smallest attainable expected distortion to its asymptotic value, as a function of blocklength  $n$ . Our main result is that in general the convergence rate is not faster than  $n^{-1/2}$ . In particular, we show that for the problem of transmitting i.i.d uniform bits over a binary symmetric channels with Hamming distortion, the smallest attainable distortion (bit error rate) is at least  $\Omega(n^{-1/2})$  above the asymptotic value, if the “bandwidth expansion ratio” is above 1.

## I. INTRODUCTION

We consider the classical point-to-point joint source-channel coding (JSCC) problem, depicted in Figure 1. In this setup, an encoder observes a sequence  $S^m = (S_1, \dots, S_m)$  of i.i.d. samples generated according to the distribution  $P_S$ , and would like to convey this sequence using  $n$  uses of a memoryless channel  $Q_{Y|X}$ . To that end, an encoding function  $\mathcal{E} : S^m \mapsto \mathcal{X}^n$  is used. The channel input  $X^n$  is transmitted through the channel  $Q_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n Q_{Y|X}(y_i|x_i)$  and the decoder that observes the channel output  $Y^n$ , generates an estimate  $\hat{S}^m = (\hat{S}_1, \dots, \hat{S}_m)$  of the source sequence, using a decoding function  $\mathcal{D} : \mathcal{Y}^n \mapsto \hat{S}^m$ . Let  $d : S \times \hat{S} \mapsto \mathbb{R}$  be some distortion measure, and define

$$d(S^m, \hat{S}^m) = \sum_{i=1}^m d(S_i, \hat{S}_i). \quad (1)$$

For a given source-channel pair, one is interested in the statistics of the distortion  $d(S^m, \hat{S}^m)$  that may be obtained, as a function of the blocklengths  $m$  and  $n$ . It is convenient to think of the *bandwidth expansion ratio*  $\rho = n/m$  as fixed (ignoring rounding effects), and then

The work of Y. Kochman was supported by the ISF under Grant 1555/18 and by the HUJI Cyber Security Research Center in conjunction with the Israel National Cyber Bureau in the Prime Minister’s Office. The work of O. Ordentlich was supported by the ISF under Grant 1791/17. The work of Y. Polyanskiy was supported, in part, by the Center for Science of Information (CSoI), and the NSF Science and Technology Center, under grant agreement CCF-09-39370.

consider the performance as a function of  $n$ . As the full statistics of the distortion are complicated, usually one of two figures of merit is considered: the *expected* distortion

$$D = \frac{1}{m} \mathbb{E}d(S^m, \hat{S}^m),$$

or the excess-distortion probability, which for any threshold  $D$  is given by

$$\epsilon(D) = \Pr \left[ \frac{1}{m} d(S^m, \hat{S}^m) > D \right].$$

For expected distortion, we define the fundamental limit for the JSCC problem by the function

$$D_n^* \triangleq D^*(n, \rho, P_S, Q_{Y|X}) \triangleq \frac{1}{m} \min_{\substack{\mathcal{E}: S^m \mapsto \mathcal{X}^n \\ \mathcal{D}: \mathcal{Y}^n \mapsto \hat{S}^m}} \mathbb{E}d(S^m, \hat{S}^m).$$

By the separation principle [1], we have that

$$D_\infty^* \triangleq \lim_{n \rightarrow \infty} D^*(n, \rho, P_S, Q_{Y|X}) = D_{P_S}(\rho C(Q_{Y|X})),$$

where  $D_{P_S}(R)$  is the distortion-rate function of a source with distribution  $P_S$ , and  $C(Q_{Y|X})$  is the capacity of the channel  $Q_{Y|X}$ .

In terms of excess distortion, many facts are known about the convergence of the distortion, in terms of large-deviation and second-order asymptotics, as well as finite-blocklength bounds [2]–[4]. This work, on the contrary, studies the convergence of the expected distortion, which is a much less understood quantity. To that end, we define and study the quantity

$$\Delta_n^* \triangleq \Delta^*(n, \rho, P_S, Q_{Y|X}) \triangleq D_n^* - D_\infty^*.$$

A special case of the JSCC problem is the lossy source coding problem, obtained by taking the channel  $Q_{Y|X}$  as a clean bit-pipe of rate  $R$ . In this case, it is known [5], [6] that  $\Delta_n^* = O\left(\frac{\log n}{n}\right)$  for any discrete source and rate  $R > 0$ . It is therefore natural to ask whether this holds in general for the JSCC problem.

For some source-channel pairs, e.g. for the problem of sending a binary symmetric source (BSS) over a

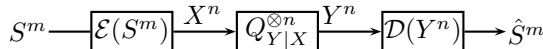


Fig. 1. The  $(n, \rho, P_S, Q_{Y|X})$  joint source channel coding problem. It is assumed that  $n = \rho m$ .

binary symmetric channel (BSC) under expected Hamming distortion with  $\rho = 1$ , the optimal asymptotic distortion is already achievable using a scalar scheme. In light of this, one might hope that a low redundancy is possible in general. However, this work proves that it is not the case, by showing that there exist distributions  $P_S$  and channels  $Q_{Y|X}$  for which  $\Delta_n^* = \Omega\left(\frac{1}{\sqrt{n}}\right)$ . In particular, we study the very same symmetric binary-Hamming problem mentioned above, but with  $\rho > 1$ , and derive a lower bound on  $D_n^*$ . We note that it is still not clear whether this is the correct scaling in general. A straightforward separation approach only suffices to show that  $\Delta_n^* = \mathcal{O}\left(\sqrt{\frac{\log n}{n}}\right)$ , and to the best of our knowledge no general JSCC coding scheme that achieves better scaling in general is known. See [7] for a more comprehensive discussion.

## II. MAIN RESULT

We study the binary symmetric joint source-channel coding problem, depicted in Figure 2, which is a special case of the general problem introduced above. The source is binary symmetric  $S \sim \text{Ber}(1/2)$ , the channel  $Q_{Y|X}$  is  $\text{BSC}(\delta)$  with  $\delta < 1/2$ , the reconstruction alphabet is  $\hat{\mathcal{S}} = \{0, 1\}$  and the distortion measure is Hamming, i.e.,  $d(S, \hat{S}) = \mathbb{1}_{\{S \neq \hat{S}\}}$ , such that

$$d(S^m, \hat{S}^m) = \frac{1}{m} \sum_{i=1}^m \mathbb{1}_{\{S_i \neq \hat{S}_i\}}.$$

We define the binary entropy function<sup>1</sup>

$$h_b(x) = -x \log x - (1-x) \log(1-x)$$

and its inverse restricted to the interval  $[0, 1/2]$  as  $h_b^{-1}(\cdot)$ . For  $0 \leq a, b \leq 1$  we also define the binary convolution

$$a * b = a(1-b) + b(1-a).$$

For two binary variables, or vectors, the notation  $+$  is to be understood as addition modulo-2.

Consider the BSS-BSC above and let  $Z^n \sim \text{Ber}(\delta)^{\otimes n}$  independent of  $X^n$ . The minimum expected Hamming distortion is given by

$$D^*(n, \rho, \delta) \triangleq \frac{1}{m} \min_{\mathcal{E}, \mathcal{D}} \mathbb{E} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + Z^n)),$$

<sup>1</sup>Throughout, logarithms are taken to the natural base.

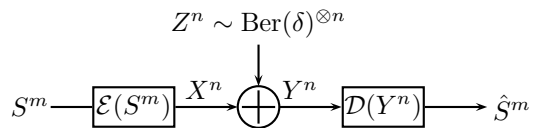


Fig. 2. The  $n(\rho, \delta)$  binary joint source-channel coding problem. It is assumed that  $n = \rho m$ .

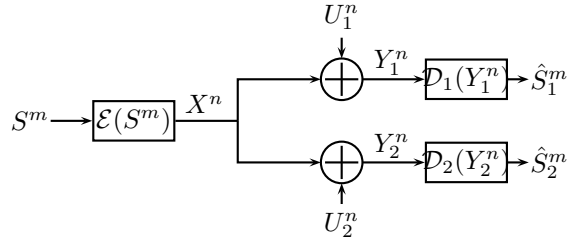


Fig. 3. Binary spherical noise JSCC broadcast problem.

where the minimization over all mappings  $\mathcal{E} : \{0, 1\}^m \mapsto \{0, 1\}^n$  and  $\mathcal{D} : \{0, 1\}^n \mapsto \{0, 1\}^m$ . The redundancy is

$$\Delta_n^* = \Delta^*(n, \rho, \delta) \triangleq D^*(n, \rho, \delta) - D(\rho, \delta),$$

where

$$D(\rho, \delta) \triangleq h_b^{-1}(\log 2 - \rho(\log 2 - h_b(\delta))), \quad (2)$$

is the asymptotic value of  $D^*(n, \rho, \delta)$ .

In this problem, it is well known that  $\Delta_n^* = 0$  for all  $n$ , when  $\rho = 1$ . Here, we will give a non-trivial lower bound for  $\rho > 1$ . The result is expressed in terms of the function  $\eta(\rho, \delta)$  which is explicitly stated in the sequel.

*Theorem 1:* In the binary JSCC problem defined above, for all  $\rho > 1$  we have that

$$\Delta_n^* \geq \sqrt{\frac{\delta(1-\delta)}{2\pi n}} \eta(\rho, \delta) + \mathcal{O}(n^{-3/4} \log n), \quad (3)$$

where  $\eta(\rho, \delta)$  is strictly positive whenever  $\rho > 1$ ,  $0 < \delta < 1/2$  and  $D(\rho, \delta) > 0$ .

In particular, this binary symmetric example serves to show that in general,

$$\Delta^*(n, \rho, P_S, Q_{Y|X}) = \Omega\left(\frac{1}{\sqrt{n}}\right).$$

## III. PROOF SKETCH: REDUCTION TO JSCC BROADCAST

The proof of Theorem 1 relies upon the reduction of the binary JSCC problem to the problem of sending a binary source over a broadcast channel, for which we then derive outer bounds on the achievable distortions region. We now outline this reduction. The details are straightforward but cumbersome, and are therefore omitted in this extended abstract, but can be found in [7]. Here we use approximated equality or inequality, to say

that the correction terms will be below the  $1/\sqrt{n}$  order of interest.

We restrict our attention to  $\delta$  such that  $\delta n$  is an integer; this reduction is insignificant in our scale of interest. We define the integer-valued random variable  $K = w_H(Z^n) - \delta n$ , where  $w_H(\cdot)$  is the Hamming weight of a vector. For a given encoder/decoder pair  $(\mathcal{E}, \mathcal{D})$  we define

$$\begin{aligned} \Psi(k) &\triangleq \mathbb{E} \left[ \frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + Z^n)) \mid K = k \right] \\ &= \mathbb{E} \left[ \frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + U^n)) \right], \end{aligned} \quad (4)$$

where  $U^n \sim \text{Uniform}(\mathcal{S}_{n\delta+k,n})$ , and where  $\mathcal{S}_{x,n}$  is the set of all length- $n$  binary sequences with weight  $0 \leq x \leq n$ , although the dependence is not made explicit. We can then use iterated expectation to assert

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + Z^n)) \right] &= \mathbb{E}[\Psi(K)] \\ &= \mathbb{E} \left[ \mathbb{E}[\Psi(K) \mid |K|] \right]. \end{aligned}$$

We then show that it is enough to consider a range of  $K$  around 0 that scales no faster than  $\sqrt{n}$ , and that within this range

$$\mathbb{E} \left[ \Psi(K) \mid |K| = k_0 \right] \approx \frac{1}{2} \Psi(k_0) + \frac{1}{2} \Psi(-k_0).$$

That is, we take  $A = |K|/\sqrt{n}$  and consider bounded  $A$ . Thus,

$$\begin{aligned} 2\mathbb{E}[\Psi(K) \mid |K|] &\gtrsim \mathbb{E} \left[ \frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + U_1^n)) \right] \\ &\quad + \mathbb{E} \left[ \frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + U_2^n)) \right], \end{aligned}$$

where  $U_1^n \sim \text{Uniform}(\mathcal{S}_{n(\delta - \frac{A}{\sqrt{n}}),n})$  and  $U_2^n \sim \text{Uniform}(\mathcal{S}_{n(\delta + \frac{A}{\sqrt{n}}),n})$ .

We now make the following relaxations: we let the encoder and decoder vary as a function of  $A$ , and further for any value of  $A$  we allow different decoders for the channel with additive noise  $U_1^n$  and the channel with additive noise  $U_2^n$ . Thus we have that

$$2D^*(n, \rho, \delta) \gtrsim \mathbb{E}[\underline{D}(A)],$$

where

$$\begin{aligned} \underline{D}(a) &\triangleq \frac{1}{m} \min_{\mathcal{E}, \mathcal{D}_1, \mathcal{D}_2} \mathbb{E} [d_H(S^m, \mathcal{D}_1(\mathcal{E}(S^m) + U_1^n))] \\ &\quad + \mathbb{E} [d_H(S^m, \mathcal{D}_2(\mathcal{E}(S^m) + U_2^n))]. \end{aligned}$$

This last quantity is the sum-distortion in a problem of JSCC broadcast with spherical noise. We study that JSCC broadcast problem, and prove that

$$\underline{D}(a) \gtrsim 2D(\rho, \delta) + \frac{a}{\sqrt{n}} \eta(\rho, \delta). \quad (5)$$

Thus,

$$D^*(n, \rho, \delta) \gtrsim D(\rho, \delta) + \frac{\mathbb{E}[A]}{2\sqrt{n}} \eta(\rho, \delta).$$

The result of the theorem follows since  $A$  is approximately the absolute value of a normal variable with zero mean and variance  $\delta(1 - \delta)$ , thus

$$\mathbb{E}[A] \approx \sqrt{\frac{2\delta(1 - \delta)}{\pi}}.$$

#### IV. BACKGROUND: JSCC BROADCAST

In this section we bring the general results we use regarding the problem of sending a source over a broadcast channel, from [8]. Then, we will specialize them for the additive spherical noise broadcast channel, in order to derive (5).

Consider the problem of sending a source over a broadcast channel, or simply, the *JSCC broadcast* problem. In this setup, an encoder observes a sequence  $S^m = (S_1, \dots, S_m)$  of  $m$  i.i.d. samples generated according to the distribution  $P_S$ , and would like to send this sequence through the broadcast channel  $Q_{Y_1^n, Y_2^n | X^n}$ , which may not be memoryless, nor degraded. To that end, the encoder maps the source sequence  $S^m$  to the channel input  $X^n$  using an encoding function  $\mathcal{E} : S^m \mapsto \mathcal{X}^n$ . The channel input  $X^n$  is transmitted through the channel and the first receiver, which observes the channel output  $Y_1^n$ , generates an estimate  $\hat{S}_1^m$  of the source sequence, using a decoding function  $\mathcal{D}_1 : \mathcal{Y}_1^n \mapsto \hat{\mathcal{S}}^m$ , whereas the second receiver, which observes the channel output  $Y_2^n$ , generates an estimate  $\hat{S}_2^m$  of the source sequence, using a decoding function  $\mathcal{D}_2 : \mathcal{Y}_2^n \mapsto \hat{\mathcal{S}}^m$ . For simplicity we assume that the reconstruction alphabets  $\hat{\mathcal{S}}_1, \hat{\mathcal{S}}_2$  are identical, and that the quality of the two reconstructions are measured with respect to the same separable distortion measure  $d : \mathcal{S} \times \hat{\mathcal{S}} \mapsto \mathbb{R}$  as

$$D_i = \frac{1}{m} \mathbb{E} d(S^m, \hat{S}_i^m), \quad i = 1, 2, \quad (6)$$

but the results easily extend to the more general case. For given  $m, n, P_S$ , and  $Q_{Y_1^n, Y_2^n | X^n}$ , a distortion pair  $(D_1, D_2)$  is said to be achievable if there exist  $(\mathcal{E}, \mathcal{D}_1, \mathcal{D}_2)$  such that  $\mathbb{E}[d(S^m, \hat{S}_i^m)] \leq D_i$  for  $i = 1, 2$ . We use  $\rho = n/m$  as in the point-to-point setting. Our goal is to establish an outer bound on the achievable pairs  $(D_1, D_2)$ .

For our results, we need the following functions of the source. We define an auxiliary variable  $U$  via a conditional distribution  $P_{U|S}$ . By combining with the

given  $P_S$  we obtain  $P = P_{S,U}$ . With respect to this distribution, we define:

$$F_P(t) \triangleq \min_{V: \substack{U-S-V \\ I(S;V) \geq t}} I(S;V|U), \quad (7)$$

$$\bar{R}_P(D) \triangleq \min_{\hat{S}: \substack{U-S-\hat{S} \\ \mathbb{E}d(S,\hat{S}) \leq D}} I(U;\hat{S}). \quad (8)$$

Notice that when  $U = S$  the function  $\bar{R}_P(D)$  reduces to the rate-distortion function  $R_{P_S}(D)$  of the source  $S$ . Furthermore, we define the following function of the channel  $Q^n = Q_{Y_1^n, Y_2^n | X^n}(y_1^n, y_2^n | x^n)$ ,

$$G_{Q^n}(t) \triangleq \max_{W, X^n: \substack{W-X^n-Y_1^n-Y_2^n \\ I(X^n; Y_1^n | W) \geq t}} I(Y_2^n; W). \quad (9)$$

The following theorem is proved in [8], and is based on generalizing a technique developed by Reznic, Feder and Zamir [9] for the Gaussian joint source-channel coding broadcast problem.

*Theorem 2:* Consider the problem of transmitting  $m$  realizations of the i.i.d. source  $S \sim P_S$ , over the  $n$ -letter broadcast channel  $Q^n$ . If  $(D_1, D_2)$  is achievable, then for any  $P_{U|S}$  we have that

$$\bar{R}_P(D_2) \leq \frac{1}{m} \cdot G_{Q^n}(mF_P(R_{P_S}(D_1))), \quad (10)$$

where  $P = P_S P_{U|S}$  is the joint distribution on  $(S, U)$  induced by the choice of  $P_{U|S}$ .

For the memoryless degraded case, the bound (10) single-letterizes [8], and can be obtained as a special case of [10, Theorem 5]. For example, consider transmission of a BSS over a degraded memoryless BSC, with crossover probabilities  $\delta_1$  and  $\delta_1 * \delta_2$ . Evaluating the quantities involved in the bound requires choosing an auxiliary test channel  $P_{U|S}$ . Choosing a binary symmetric test channel with crossover probability  $q$  yields:

$$F_P(t) = t - \log 2 + h_b(q * h^{-1}(\log 2 - t)), \quad (11a)$$

$$\bar{R}_P(D) = \log 2 - h_b(q * D), \quad (11b)$$

while the channel function single-letterizes as:

$$\begin{aligned} \frac{1}{n} G_{Q^n}(nt) &= \log 2 - h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + t)) \\ &\triangleq G_{\text{BSC}}(t). \end{aligned} \quad (12)$$

These quantities can be substituted in (10) to obtain an explicit bound.

## V. JSCC BROADCAST WITH SPHERICAL NOISE

Next, we would like to specialize the bound from Theorem 2 to the binary spherical noise JSCC broadcast problem. The functions  $F_P(t)$  and  $\bar{R}_P(D)$  tensorize and for a binary-symmetric  $S$  are given by (11) for  $P_{U|S} = \text{BSC}(q)$ . The function  $G_{Q^n}$ , on the other hand, requires solving, or bounding the solution of, an  $n$ -letter

optimization, which is in general a challenging task. Nevertheless, for the additive spherical noise broadcast channel (which is neither memoryless, nor degraded), we derive an upper bound on  $\frac{1}{n} G_{Q^n}(nt)$  that, up to some correction terms, agrees with the function  $G_{\text{BSC}}(t)$  corresponding to a memoryless additive noise broadcast channel with the same marginal noise distribution, i.e., a  $\text{BSC}(\delta_1, \delta_1 * \delta_2)$  broadcast channels. In particular, we prove the following.

*Lemma 1:* Let  $Q^n = Q_{Y_1^n, Y_2^n | X^n}(y_1^n, y_2^n | x^n)$  be the additive spherical-noise broadcast channel  $Y_1^n = X^n + U_1^n$ ,  $Y_2^n = X^n + U_2^n$ , where  $U_1^n \sim \text{Uniform}(\mathcal{S}_{n\delta_1, n})$  and  $U_2^n \sim \text{Uniform}(\mathcal{S}_{n(\delta_1 * \delta_2), n})$ ,  $(U_1^n, U_2^n) \perp\!\!\!\perp X^n$ . Then

$$\frac{1}{n} G_{Q^n}(nt) \leq G_{\text{BSC}}(t) + \Gamma(n, \delta_2),$$

where  $G_{\text{BSC}}(t)$  is given by (12), and

$$\Gamma(n, \delta_2) \triangleq \sqrt{\frac{\delta_2}{n}} \log \left( \frac{n}{\delta_2} \right) + \frac{\log n + 1}{2n}. \quad (13)$$

**Proof.** Let  $(W, X^n)$  satisfy the Markov chain  $W - X^n - (Y_1^n = X^n + U_1^n, Y_2^n = X^n + U_2^n)$ . We begin by writing

$$H(Y_2^n | W) = H(X^n + U_1^n + Z_3^n | W) + \kappa \quad (14)$$

where

$$\kappa = H(X^n + U_2^n | W) - H(X^n + U_1^n + Z_3^n | W),$$

and where  $Z_3^n \sim \text{Ber}(\delta_2)^{\otimes n}$ . We will upper bound the absolute value of  $\kappa$  via coupling. Consider the following joint distribution on  $(U_2^n, U_1^n + Z_3^n)$ :

- Let  $\Pi$  be a uniform random permutation on  $[n] = \{1, \dots, n\}$ .
- Let  $T = T_0 + T_1$  where  $T_0 \sim \text{Binomial}(n(1 - \delta_1), \delta_2)$  and  $T_1 \sim \text{Binomial}(n\delta_1, 1 - \delta_2)$  are independent.
- Set  $U_{2, \Pi(i)} = 1$  for  $i = 1, \dots, (\delta_1 * \delta_2)n$  and  $U_{\Pi(i)} = 0$  for  $i = (\delta_1 * \delta_2)n + 1, \dots, n$ .
- Set  $U_{1, \Pi(i)} + Z_{3, \Pi(i)} = 1$  for  $i = 1, \dots, T$  and  $U_{1, \Pi(i)} + Z_{3, \Pi(i)} = 0$  for  $i = T + 1, \dots, n$ .

Clearly  $U_2^n$  and  $U_1^n + Z_3^n$  have the correct marginal distributions. Moreover, the expected Hamming distance between these vectors satisfies

$$\begin{aligned} \mathbb{E}w_H(U_2^n + (U_1^n + Z_3^n)) &= \mathbb{E}|T - n(\delta_1 * \delta_2)| \\ &= \mathbb{E}\sqrt{(T - n(\delta_1 * \delta_2))^2} \\ &\leq \sqrt{\text{Var}(T)} \\ &= \sqrt{\text{Var}(T_1) + \text{Var}(T_2)} \\ &= \sqrt{n\delta_2(1 - \delta_2)} \leq \sqrt{n\delta_2}, \end{aligned}$$

where the first inequality follows from Jensen's inequality and the fact that  $\mathbb{E}(T) = n(\delta_1 * \delta_2)$ . Now, applying [11, Proposition 8], we obtain for  $\delta_2 < 1/2$

$$|\kappa| = |H(X^n + U_2^n | W) - H(X^n + U_1^n + Z_3^n | W)| \\ \leq \sqrt{n\delta_2} \log \left( \frac{n}{\delta_2} \right).$$

Thus, we can use Mrs. Gerber's Lemma (MGL) to lower bound (14) as

$$H(Y_2^n | W) \geq H(X^n + U_1^n + Z_3^n | W) - \sqrt{n\delta_2} \log \left( \frac{n}{\delta_2} \right) \\ \geq n\varphi_{\delta_2} \left( \frac{H(X^n + U_1^n | W)}{n} \right) - \sqrt{n\delta_2} \log \left( \frac{n}{\delta_2} \right),$$

where  $\varphi_{\delta_2}(x) = h_b(\delta_2 * h^{-1}(x))$  is the MGL function. Further bounding, we have

$$\varphi_{\delta_2} \left( \frac{H(X^n + U_1^n | W)}{n} \right) \\ = \varphi_{\delta_2} \left( \frac{H(U_1^n) + I(X^n; Y_1^n | W)}{n} \right) \\ \geq \varphi_{\delta_2} \left( \frac{nh_b(\delta_1) - \frac{1}{2}(\log n + 1) + I(X^n; Y_1^n | W)}{n} \right) \\ = \varphi_{\delta_2} \left( h_b(\delta_1) + \frac{I(X^n; Y_1^n | W)}{n} - \frac{\log n + 1}{2n} \right)$$

where for the inequality we substitute [12, Chapter 10, Lemma 7] to lower-bound  $H(U_1^n)$ . Therefore

$$H(Y_2^n | W) \geq n\varphi_{\delta_2} \left( h_b(\delta_1) + \frac{I(X^n; Y_1^n | W)}{n} \right) \\ - \frac{\log n + 1}{2} \varphi'_{\delta_2} \left( h_b(\delta_1) + \frac{I(X^n; Y_1^n | W)}{n} \right) \\ - \sqrt{n\delta_2} \log \left( \frac{n}{\delta_2} \right),$$

where we have defined the MGL derivative  $\varphi'_{\delta_2}(x) = \frac{d}{dx} \varphi_{\delta_2}(x)$ , and used the convexity of  $x \mapsto \varphi_{\delta_2}(x)$  [13]. Recalling that  $\varphi'_{\delta_2}(x) \leq 1$  due to [14, Theorem 2.6], the claim now follows since  $H(Y_2^n) \leq n \log 2$ . ■

Using Lemma 1, and (11), we can now apply Theorem 2 for obtaining an outer bound on the achievable  $(D_1, D_2)$  pairs for the binary spherical noise JSCC broadcast problem.

**Theorem 3:** Consider the binary spherical noise JSCC broadcast problem of transmitting a memoryless symmetric binary source  $S^m$  over the channels  $Y_1^n = X^n + U_1^n$ ,  $Y_2^n = X^n + U_2^n$  under Hamming distortion, where  $U_1^n \sim \text{Uniform}(\mathcal{S}_{n\delta_1, n})$  and  $U_2^n \sim \text{Uniform}(\mathcal{S}_{n(\delta_1 * \delta_2), n})$ ,  $(U_1^n, U_2^n) \perp\!\!\!\perp X^n$ . If  $(D_1, D_2)$  is achievable, then for any  $0 < q < 1/2$

$$\log 2 - h_b(q * D_2) \leq \rho \Gamma(n, \delta_2) \\ + \rho \left[ \log 2 - h_b \left( \delta_2 * h_b^{-1} \left( h_b(\delta_1) + \frac{\beta_q(D_1)}{\rho} \right) \right) \right],$$

where  $\Gamma(n, \delta_2)$  is as defined in (13), and  $\beta_q(t) \triangleq h_b(q * t) - h_b(t)$ .

Obtaining the lower bound (5) on  $\underline{D}(a)$  from Theorem 3 is now merely a matter of considering the local properties of the functions and performing algebraic manipulations. See [7] for details. The constant  $\eta(\rho, \delta)$  is given by:

$$\eta(\rho, \delta) = 2\rho \frac{\log \left( \frac{1-\delta}{\delta} \right)}{\log \left( \frac{1-D(\rho, \delta)}{D(\rho, \delta)} \right)} \cdot \frac{1 + f(\rho, \delta)}{f(\rho, \delta)} \\ \cdot \frac{D(\rho, \delta) (1 - f(\rho, \delta))^2}{2f(\rho, \delta) + 4D(\rho, \delta) (1 - f(\rho, \delta))},$$

where

$$\Phi(\delta) \triangleq \frac{2}{(1 - 2\delta) \log \left( \frac{1-\delta}{\delta} \right)} + \frac{1}{\delta(1 - \delta) \log^2 \left( \frac{1-\delta}{\delta} \right)} \\ f(\rho, \delta) \triangleq \frac{1}{\rho} \frac{\Phi(\delta)}{\Phi(D(\rho, \delta))}.$$

## REFERENCES

- [1] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the Institute of Radio Engineers*, vol. 37, no. 1, pp. 10–21, January 1949.
- [2] I. Csiszár, "On the error exponent of source-channel transmission with a distortion threshold," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 823–828, Nov 1982.
- [3] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept 2011, pp. 180–187.
- [4] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.
- [5] Z. Zhang, E. Yang, and V. Wei, "The redundancy of source coding with a fidelity criterion - Part one: Known statistics," *IEEE Transactions on Information Theory*, vol. 43, no. 1, pp. 71–91, Jan 1997.
- [6] B. Yu and T. Speed, "A rate of convergence result for a universal D-semifaithful code," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 813–820, May 1993.
- [7] Y. Kochman, O. Ordentlich, and Y. Polyanskiy, "A lower bound on the expected distortion of joint source-channel coding," *arXiv preprint arXiv:1902.07979*, 2019.
- [8] —, "Ozarow-type outer bounds for memoryless sources and channels," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 1774–1778.
- [9] Z. Reznic, M. Feder, and R. Zamir, "Distortion bounds for broadcasting with bandwidth expansion," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3778–3788, Aug 2006.
- [10] K. Khezeli and J. Chen, "Outer bounds on the admissible source region for broadcast channels with correlated sources," *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 4616–4629, Sept 2015.
- [11] 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.
- [12] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*. Elsevier, 1977.
- [13] 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.
- [14] H. Witsenhausen and A. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 5, pp. 493–501, Sep 1975.