

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

Yuval Kochman, Or Ordentlich and Yury Polyanskiy

## 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 long-standing open problem of finding 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

Over the last decade there has been a great progress in understanding the rate of convergence to the asymptotic fundamental limits in various communication and compression setups. Yet, there remain some setups where although the asymptotic limits are known, the rate of convergence to those is not known. Two prominent examples are to show that the Gaussian-type  $1/\sqrt{n}$  backoff, common to many settings, also arises in the joint source-channel coding (JSCC) setup under expected distortion, and in the multiple access channel (MAC). The main difficulty in the JSCC setup is that due to averaging, the Gaussian variations in source/channel quality may possibly be canceled out (as in fact happens when the ‘not to code’ conditions of [2] are met). The fundamental issue in the MAC setup [3] is that the multi-user interference, unless randomly-coded, should not in general satisfy the central limit theorem and hence there is no reason to believe the back-off should be of  $1/\sqrt{n}$  order. This paper makes progress on the first of these open problems.

Specifically, we consider the classical point-to-point joint source-channel coding problem, depicted in Figure 1. 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  $n$  channel uses of the memoryless channel  $Q_{Y|X}$ . 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  $Q_{Y^n|X^n}^{\otimes n}(y^n|x^n) = \prod_{i=1}^n Q_{Y|X}(y_i|x_i)$  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  $\mathcal{D} : \mathcal{Y}^n \mapsto \hat{\mathcal{S}}^m$ . Let  $d : \mathcal{S} \times \hat{\mathcal{S}} \mapsto \mathbb{R}$  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 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

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

Y. Kochman and O. Ordentlich are with the Hebrew University of Jerusalem, Israel (emails: yuval.kochman@mail.huji.ac.il, or.ordentlich@mail.huji.ac.il). Y. Polyanskiy is with the Massachusetts Institute of Technology, MA, USA (email: yp@mit.edu)

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. The material in this paper was presented in part at the 2018 International Symposium on Information Theory [1].

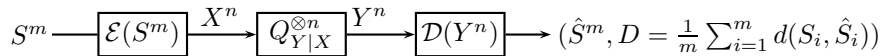


Fig. 1. The  $(n, \rho, P_S, Q_{Y|X})$  joint source channel coding problem. It is assumed that  $n = \rho m$ .

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

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

The focus of this work is the expected distortion, for which we define the fundamental limit for the JSCC problem by the function

$$D_n^* = D^*(n, \rho, P_S, Q_{Y|X}) \triangleq \frac{1}{m} \min_{\substack{\mathcal{E}: \mathcal{S}^m \mapsto \mathcal{X}^n \\ \mathcal{D}: \mathcal{Y}^n \mapsto \hat{\mathcal{S}}^m}} \mathbb{E} d(S^m, \hat{S}^m). \quad (2)$$

By the separation principle [4], we have that

$$D_\infty^* \triangleq \lim_{n \rightarrow \infty} D_n^*(n, \rho, P_S, Q_{Y|X}) = D_{P_S}(\rho C(Q_{Y|X})), \quad (3)$$

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  $Q_{Y|X}$ . We study the convergence of the expected distortion to its asymptotic value. To that end, we define and study the quantity

$$\Delta_n^* = \Delta^*(n, \rho, P_S, Q_{Y|X}) \triangleq D_n^* - D_\infty^*. \quad (4)$$

In terms of excess distortion, many facts are known about the convergence of the distortion to the infinite-blocklength limit. For any threshold  $D > D_\infty^*$  we have that  $\epsilon(D)$  is exponentially small, with upper and lower bounds on the exponents (which agree for low enough  $D$ ) given in [5]. To the contrary, for any  $D < D_\infty^*$  it holds that  $1 - \epsilon(D)$  is exponentially small [6]. The dispersion, i.e., convergence of  $D$  to  $D_\infty^*$  for fixed excess-distortion probability, as well as finite-blocklength bounds, were derived in [7], [8].

For the average distortion case, however, one may argue that  $\Delta_n^* = o(1/\sqrt{n})$  as follows. In problems where the error criterion is defined as a hard-constrained 0/1-loss (such as excess-distortion above), the Gaussian variations in channel quality cannot be leveraged and the fundamental limit experiences a  $1/\sqrt{n}$  back-off from its asymptotic values. However, in problems with averaging, the dispersion term disappears, for example [9], [10], [11], because these (mean-zero) Gaussian variations can be averaged out. This suggests that the true behavior of the fundamental limit  $\Delta_n^*$  should indeed be  $o(1/\sqrt{n})$ .

As further evidence that  $\Delta_n^*$  may be  $o(1/\sqrt{n})$ , consider the special case of the JSCC where  $Q_{Y|X}$  is a clean bit-pipe of rate  $R$ , for which the problem reduces to lossy source coding. In this case, it is known [12], [13] that for any discrete source and rate  $R > 0$ ,

$$\Delta_n^* = \mathcal{O} \left( \frac{\log n}{n} \right).$$

Furthermore, for some source-channel pairs the optimal asymptotic distortion is already achievable using a scalar scheme. See [2] for necessary and sufficient conditions. For example, this is the case for the problem of sending a binary symmetric source (BSS) over a binary symmetric channel (BSC) under expected Hamming distortion with  $\rho = 1$ . In light of this, one might hope that a low redundancy is possible in general.

Despite all this evidence for sub- $\sqrt{n}$  convergence, this work proves that it is not the case, by showing that in general  $\Delta_n^* = \Omega \left( \frac{1}{\sqrt{n}} \right)$ . More concretely, we study the very same symmetric binary-Hamming problem mentioned above, but with  $\rho > 1$ , and derive a lower bound on  $D_n^*$ .

Our approach to proving this result goes through a reduction to a JSCC broadcast problem. Let  $\hat{Q}_{Y|X}$  be the empirical channel realization in the point-to-point JSCC problem. The main observation in our distortion lower bound is that a good JSCC code must achieve distortions close to  $D_\infty(\rho, P_S, \hat{Q}_{Y|X})$  simultaneously for all “probable” channel realizations. To show that this is impossible, we reduce the problem to that of broadcasting a source to two users with different channel conditions, corresponding to one empirical channel that is better than  $Q_{Y|X}$ , and one that is worse. An outer bound on the distortions, in the infinite-blocklength limit, was derived in [14] for the

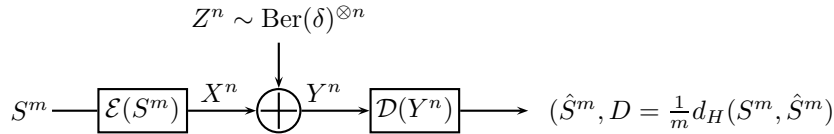


Fig. 2. The  $n(\rho, \delta)$  binary joint source-channel coding problem. it is assumed that  $n = \rho m$ .

quadratic-Gaussian case, and was recently generalized by Khezeli and Chen [15], [16]. Here, we generalize these bounds for our scenario of interest, and show that for the binary Hamming case, with  $\rho > 1$ , it is impossible to design a JSCC code that will be optimal simultaneously for both channel conditions. Unfortunately, for the case  $\rho < 1$  this technique falls short of providing similar bounds, mainly because in this regime  $D_\infty$  changes too slowly with  $\delta$ .

We note that it is still not clear whether the  $n^{-1/2}$  scaling of  $\Delta_n^*$  is achievable in general. In particular, a separation-based coding scheme can only achieve  $\Delta_n^* = \Omega\left(\sqrt{\frac{\log n}{n}}\right)$ , and to the best of our knowledge no JSCC coding scheme that achieves better scaling in general is known. See Section VI. Thus, despite the progress made in this work, the exact correct scaling of  $\Delta_n^*$  remains an open question.

The structure of the paper is as follows. In Section II, we define the binary instance of the JSCC problem we analyze, state the main result, and give a high-level sketch of proof. Section III develops outer bounds on the distortions that can be achieved when sending a source over a broadcast channel. The bounds from Section III are then specialized in Section IV, for the problem of sending a binary source over a binary additive spherical noise broadcast channel. The full proof of our main result is given in Section V. The paper concludes with a discussion in Section VI.

## II. MAIN RESULT AND MAIN TECHNICAL CONTRIBUTION

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

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

We use 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

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

The expected distortion for this problem is formally defined below. For two binary variables, or vectors, the notation  $+$  is to be understood as addition modulo-2.

*Definition 1:* Let  $S^m \sim \text{Ber}(1/2)^{\otimes m}$ ,  $Y^n = X^n + Z^n$  with  $Z^n \sim \text{Ber}(\delta)^{\otimes n}$  independent of  $X^n$ , and  $\rho = n/m$ . The minimum expected Hamming distortion for transmitting  $S^m$  over the channel from  $X^n$  to  $Y^n$  is defined as

$$D^*(n, \rho, \delta) \triangleq \frac{1}{m} \min_{\substack{\mathcal{E}: \{0,1\}^m \mapsto \{0,1\}^n \\ \mathcal{D}: \{0,1\}^n \mapsto \{0,1\}^m}} \mathbb{E} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + Z^n)). \quad (5)$$

and its convergence rate function is

$$\Delta_n^* = \Delta^*(n, \rho, \delta) \triangleq D^*(n, \rho, \delta) - D(\rho, \delta), \quad (6)$$

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

where

$$D(\rho, \delta) \triangleq h_b^{-1}(\log 2 - \rho(\log 2 - h_b(\delta))), \quad (7)$$

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  $\rho > 1$ . We will express our result in terms of the following functions:

$$\Phi(\delta) \triangleq \frac{2}{(1-2\delta)\log\left(\frac{1-\delta}{\delta}\right)} + \frac{1}{\delta(1-\delta)\log^2\left(\frac{1-\delta}{\delta}\right)} \quad (8a)$$

$$f(\rho, \delta) \triangleq \frac{1}{\rho} \frac{\Phi(\delta)}{\Phi(D(\rho, \delta))}, \quad (8b)$$

$$\eta(\rho, \delta) \triangleq 2\rho \frac{\log\left(\frac{1-\delta}{\delta}\right)}{\log\left(\frac{1-D(\rho, \delta)}{D(\rho, \delta)}\right)} \cdot \frac{D(\rho, \delta)(1-f(\rho, \delta))^2}{2f(\rho, \delta) + 4D(\rho, \delta)(1-f(\rho, \delta))} \cdot \frac{1+f(\rho, \delta)}{f(\rho, \delta)}. \quad (8c)$$

Our main result is the following.

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

$$\Delta_n^* \triangleq D^*(n, \rho, \delta) - D(\rho, \delta) \geq \sqrt{\frac{\delta(1-\delta)}{2\pi n}} \eta(\rho, \delta) + \mathcal{O}(n^{-3/4} \log n), \quad (9)$$

where  $\eta(\rho, \delta)$  is as defined in (8c) and is strictly positive for  $D(\rho, \delta) > 0$ .

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

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

#### A. Outline of 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, thus they are relegated to Section V. 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. Let  $\mathcal{S}_{x,n}$  be the set of all length- $n$  binary sequences with weight  $0 \leq x \leq n$ . For a given encoder/decoder pair  $(\mathcal{E}, \mathcal{D})$  we define

$$\begin{aligned} \Psi(k) &= \Psi_{\mathcal{E}, \mathcal{D}}(k) \triangleq \mathbb{E} \left[ \frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + Z^n)) \mid K = k \right] \\ &= \mathbb{E} \left[ \frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + U^n)) \right], \end{aligned} \quad (10)$$

where  $U^n \sim \text{Uniform}(\mathcal{S}_{\delta n + k, n})$ , although its dependence on  $k$  is not made explicit. We can then use iterated expectation to assert

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + Z^n)) \right] &= \mathbb{E}[\Psi(K)] \\ &= \mathbb{E} \left[ \mathbb{E}[\Psi(K) \mid |K|] \right]. \end{aligned} \quad (11)$$

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,

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

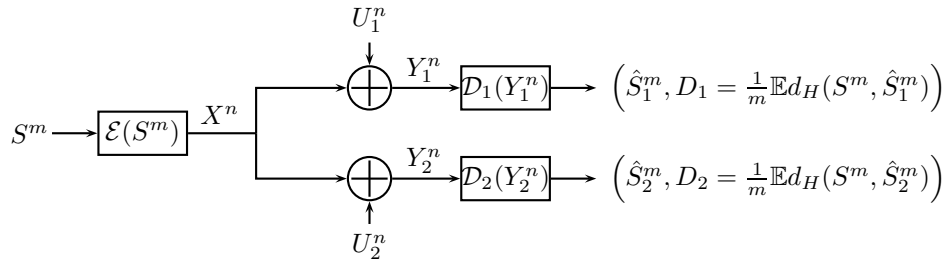


Fig. 3. Binary spherical noise JSCC broadcast problem.

where  $U_1^n \sim \text{Uniform}\left(\mathcal{S}_{n(\delta - \frac{A}{\sqrt{n}}), n}\right)$  and  $U_2^n \sim \text{Uniform}\left(\mathcal{S}_{n(\delta + \frac{A}{\sqrt{n}}), n}\right)$ . The outer expectation in the right hand side is taken over  $A$ .

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

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

where

$$\underline{D}(a) \triangleq \frac{1}{m} \min \mathbb{E}[d_H(S^m, \mathcal{D}_1(\mathcal{E}(S^m) + U_1^n))] + \mathbb{E}[d_H(S^m, \mathcal{D}_2(\mathcal{E}(S^m) + U_2^n))], \quad (12)$$

and the minimization is over all encoders  $\mathcal{E} : \{0, 1\}^m \mapsto \{0, 1\}^n$  and decoders  $\mathcal{D}_1 : \{0, 1\}^n \mapsto \{0, 1\}^m$ ,  $\mathcal{D}_2 : \{0, 1\}^n \mapsto \{0, 1\}^m$ . This last quantity is the sum-distortion in a problem of JSCC broadcast with spherical noise. We will next consider JSCC broadcast, and prove in Theorem 4 that

$$\underline{D}(a) \gtrsim 2D(\rho, \delta) + \frac{a}{\sqrt{n}}\eta(\rho, \delta). \quad (13)$$

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

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

which gives the stated result.

### B. Sending a Source Over a Broadcast Channel

We now outline the derivation of (13), stating the steps which we will prove in Section IV.

We 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  $\mathcal{E} : \mathcal{S}^m \mapsto \mathcal{X}^n$ . 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

$$D_i = \frac{1}{m} \mathbb{E}d(S^m, \hat{S}_i^m), \quad i = 1, 2, \quad (14)$$

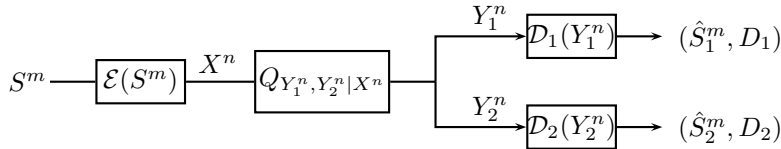


Fig. 4. JSCC over a Broadcast Channel.

but the results easily go through to the more general case as well. See Figure 4.

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  $i = 1, 2$ . It will be convenient to express results using the *bandwidth expansion factor*  $\rho = n/m$ . 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:

$$F_P(t) \triangleq \min_{V: U-S-V} \min_{I(S;V) \geq t} I(S;V|U), \quad (15)$$

$$\bar{R}_P(D) \triangleq \min_{\hat{S}: U-S-\hat{S}} \min_{\mathbb{E}d(S, \hat{S}) \leq D} I(U; \hat{S}). \quad (16)$$

Notice that when  $U = S$  the function  $\bar{R}_P(D)$  reduces to the rate-distortion function  $R_{P_S}(D)$  of the source  $S$ .<sup>2</sup> Furthermore, we define the following function of the channel  $Q^n = Q_{Y_1^n, Y_2^n | X^n}(y_1^n, y_2^n | x^n)$ ,

$$G_{Q^n}(t) \triangleq \max_{W, X^n: W-X^n-Y_1^n-Y_2^n} \max_{I(X^n; Y_1^n | W) \geq t} I(Y_2^n; W). \quad (17)$$

Note the relation to the capacity region of the broadcast channel: If  $nR_1$  and  $nR_2$  bits can be communicated reliably to the receivers  $Y_1^n$  and  $Y_2^n$ , respectively, then  $nR_2 \leq G_{Q^n}(nR_1)$  [18, Chapter 5.4.1].

Let  $R_{P_S}(D)$  be the rate-distortion function of the source  $P_S$ . We prove the following theorem.

**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

$$\bar{R}_P(D_2) \leq \frac{1}{m} \cdot G_{Q^n}(mF_P(R_{P_S}(D_1))), \quad (18)$$

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

The proof of this bound, which is based on generalizing a technique developed by Reznic, Feder and Zamir [14] for the Gaussian joint source-channel coding broadcast problem, will be given in Section III. We will also show that for the memoryless degraded case, the bound (43) single-letterizes, and can be obtained as a special case of [15, Theorem 5].

While the functions  $F_P(t)$ ,  $\bar{R}_P(D)$  and  $R_{P_S}(D)$ , involved in the evaluation of the bound from Theorem 2 only require solving a single-letter optimization, the function  $G_{Q^n}$  requires solving, or bounding the solution of, an  $n$ -letter optimization, which is in general a challenging task. Nevertheless, for our particular channel of interest, where  $Q^n$  is an additive spherical-noise broadcast channel (which is neither memoryless, nor degraded) 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,

$$\frac{1}{n} G_{Q^n}(nt) \leq \log 2 - h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + t)) + \Gamma(n, \delta_2), \quad (19)$$

where

$$\Gamma(n, \delta_2) \triangleq \sqrt{\frac{\delta_2}{n}} \log\left(\frac{n}{\delta_2}\right) + \frac{\log n + 1}{2n}. \quad (20)$$

<sup>2</sup>We note that the function  $\bar{R}_P(D)$  also arises as an upper bound on the communication rate required in order to perform a certain coordination task, see [17].

Indeed, if we replaced the spherical noises by an i.i.d. noise with the same marginals, the channel would become a  $\text{BSC}(\delta_1, \delta_1 * \delta_2)$  degraded memoryless broadcast channel. In that case, the corresponding  $G(t)$  function does single-letterize, and equals  $\log 2 - h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + t))$ , see Section III. Thus, Lemma 1 shows that the difference between these cases is upper bounded by  $\Gamma(n, \delta_2)$ . The proof of this lemma, which is given in Section IV, is based on replacing  $U_2^n$  with the noise  $U_1^n + \tilde{Z}^n$ , where  $\tilde{Z}^n \sim \text{Ber}(\delta_2)^{\otimes n}$ , such that the obtained broadcast channel  $\tilde{Q}^n$  is degraded (though not memoryless), and its corresponding function  $G_{\tilde{Q}^n}$  can be computed using Mrs. Gerber's Lemma [19]. The difference  $|G_{\tilde{Q}^n}(nt) - G_{Q^n}(nt)|$  is essentially bounded by  $\sup_{P_{X^n}} |H(X^n + U_1^n + \tilde{Z}^n) - H(X^n + U_2^n)|$ , which can be bounded via a coupling argument introduced in [20] (see also [21]). Using this technique, we show that  $|G_{\tilde{Q}^n}(nt) - G_{Q^n}(nt)| \lesssim n \cdot \Gamma(n, \delta_2)$ . Note that for the purpose of bounding  $\underline{D}(a)$ , we are interested in the regime where  $\delta_2 = \Theta(n^{-1/2})$ , and consequently  $\Gamma(n, \delta_2) = \mathcal{O}(n^{-3/4} \log n)$  is below the  $n^{-1/2}$  scale of approximation we are after.

Thus, in order to apply Theorem 2 for obtaining an outer bound on the achievable  $(D_1, D_2)$  pairs for the binary spherical noise JSCC broadcast problem, we need only choose an auxiliary channel  $P_{U|S}$  and evaluate the functions  $F_P(t)$ ,  $\bar{R}_P(D)$  and  $R_{P_S}(D)$ . For the particular choice  $U = S + N$ , where  $N \sim \text{Ber}(q)$  is statistically independent of  $X$ , evaluating these functions becomes particularly simple, and we obtain the following Theorem, proved in Section IV.

*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  $0 < q < 1/2$

$$\log 2 - h_b(q * D_2) \leq \rho \left[ \log 2 - h_b \left( \delta_2 * h_b^{-1} \left( h_b(\delta_1) + \frac{h_b(q * D_1) - h_b(D_1)}{\rho} \right) \right) \right] + \rho \Gamma(n, \delta_2), \quad (21)$$

where  $\Gamma(n, \delta_2)$  is as defined in (20).

Extracting a lower bound on  $\underline{D}(a)$  from Theorem 3 is now merely a matter of algebraic manipulations. This is done in Section IV-B, where we prove the following theorem.

*Theorem 4:* 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  $U_2^n \sim \text{Uniform}(\mathcal{S}_{n(\delta_1 * \delta_2), n})$ ,  $(U_1^n, U_2^n) \perp\!\!\!\perp X^n$ , and  $\rho = n/m > 1$ , where  $\delta_1 = \delta - \frac{a}{\sqrt{n}}$ ,  $\delta_1 * \delta_2 = \delta + \frac{a}{\sqrt{n}}$ , for some positive  $a < \log^2(n)$ . If  $(D_1, D_2)$  is achievable, then

$$D_1 + D_2 \geq 2D(\rho, \delta) + \frac{a\eta(\rho, \delta)}{\sqrt{n}} + \mathcal{O}(n^{-3/4} \log n), \quad (22)$$

where  $\eta(\rho, \delta)$  is as defined in (8c).

### III. OUTER BOUNDS FOR JOINT SOURCE-CHANNEL CODING OVER A BROADCAST CHANNEL

In this section we prove Theorem 2 by analyzing the source and channel functions (15)-(17). We also evaluate the functions for some memoryless degraded examples and recover some previously known results as special cases. We note that the setting of Theorem 2 is asymmetric: the source is assumed to be i.i.d., while the channel is neither memoryless nor degraded; this is the most general form that we need for this work, and the extension for sources with memory will become quite obvious in the sequel.

#### A. The Source Functions

First, we prove the following simple statements regarding properties of the functions  $F_P$  (15) and  $\bar{R}_P$  (16).

*Lemma 2:* The function  $F_P(t)$  is monotone non-decreasing and convex. Furthermore, it tensorizes, i.e.,

$$F_{P^m}(mt) = mF_P(t).$$

**Proof.** Monotonicity of the function  $F_P(t)$  follows by definition. For convexity, let  $V_0$  and  $V_1$  be random variables (induced by the channels  $P_{V_i|S}$ ) that attain  $F_P(t_0)$  and  $F_P(t_1)$  with equality. Let  $A \sim \text{Ber}(\alpha)$ , and define  $\tilde{V} = (A, V_A)$ . We have that

$$\begin{aligned} I(S; \tilde{V}) &= I(S; V_A|A) \\ &= (1 - \alpha)I(S; V_0) + \alpha I(S; V_1) \\ &= (1 - \alpha)t_0 + \alpha t_1, \end{aligned}$$

and by definition of  $F_P(t)$ ,

$$\begin{aligned} F_P((1 - \alpha)t_0 + \alpha t_1) &\leq I(S; \tilde{V}|U) \\ &= (1 - \alpha)I(S; V_0|U) + \alpha I(S; V_1|U) \\ &= (1 - \alpha)F_P(t_0) + \alpha F_P(t_1). \end{aligned} \tag{23}$$

We prove tensorization by induction. For any  $V$  that satisfies the Markov chain  $U^m - S^m - V$ , we have

$$\begin{aligned} F_P\left(\frac{I(S^m; V)}{m}\right) &= F_P\left(\frac{I(S^{m-1}; V) + I(S_m; V|S^{m-1})}{m}\right) \\ &= F_P\left(\frac{I(S^{m-1}; V) + I(S_m; V, S^{m-1})}{m}\right) \\ &= F_P\left(\frac{m-1}{m} \frac{I(S^{m-1}; V)}{m-1} + \frac{1}{m} I(S_m; V, S^{m-1})\right) \\ &\leq \frac{m-1}{m} F_P\left(\frac{I(S^{m-1}; V)}{m-1}\right) + \frac{1}{m} F_P(I(S_m; V, S^{m-1})), \end{aligned}$$

where we have used the convexity of  $t \mapsto F_P(t)$  in the last inequality. Invoking the induction hypothesis, we have

$$\begin{aligned} F_P\left(\frac{I(S^m; V)}{m}\right) &\leq \frac{1}{m} F_{P^{m-1}}(I(S^{m-1}; V)) + \frac{1}{m} F_P(I(S_m; V, S^{m-1})) \\ &\leq \frac{1}{m} [I(S^{m-1}; V|U^{m-1}) + I(S_m; V, S^{m-1}|U_m)], \end{aligned} \tag{24}$$

where the last inequality follows by definition of  $F_{P^{m-1}}$  and  $F_P$  and the fact  $U^{m-1} - S^{m-1} - V$  and  $U_m - S_m - (V, S^{m-1})$  are indeed Markov chains. Noting that

$$I(S^{m-1}; V|U^{m-1}) \leq I(S^{m-1}; V|U^m),$$

and

$$\begin{aligned} I(S_m; V, S^{m-1}|U_m) &= I(S_m; V, S^{m-1}|U^m) \\ &\leq I(S_m; V|S^{m-1}, U^m), \end{aligned}$$

which both follow since  $S^m$  is memoryless, we obtain

$$I(S^{m-1}; V|U^{m-1}) + I(S_m; V, S^{m-1}|U_m) \leq I(S^m; V|U^m). \tag{25}$$

Substituting (25) into (24), gives

$$I(S^m; V|U^m) \geq m F_P\left(\frac{I(S^m; V)}{m}\right), \tag{26}$$

as desired. ■

*Lemma 3:* Let  $\hat{S}^m$  be a random vector satisfying the Markov chain  $U^m - S^m - \hat{S}^m$  and  $\mathbb{E}d(S^m, \hat{S}^m) \leq D$ , then

$$I(U^m; \hat{S}^m) \geq m \bar{R}_P(D).$$

**Proof.** Since  $U^m$  is memoryless, we have that

$$I(U^m; \hat{S}^m) \geq \sum_{i=1}^m I(U_i; \hat{S}_i). \tag{27}$$



Note that  $\frac{1}{m} \sum_{i=1}^m \mathbb{E}d(S_i; \hat{S}_i) \leq D$  by separability of  $d(S^m; \hat{S}^m)$ , and that the Markov chain  $U^m - S^m - \hat{S}^m$  implies that  $U_i - S_i - \hat{S}_i$  is also a Markov chain. It is easy to see that the function  $D \mapsto \bar{R}_P(D)$  is convex. Thus, letting  $d_i = \mathbb{E}d(S_i; \hat{S}_i)$ , we have that

$$I(U^m; \hat{S}^m) \geq m \frac{1}{m} \sum_{i=1}^m \bar{R}_P(d_i) \geq m \bar{R}_P(D). \quad (28)$$

■

We now evaluate these functions for a binary example that will be important in this work: we take  $S$  to be a Bernoulli( $p$ ) source, and  $d(S_j, \hat{S}_j)$  to be the Hamming distortion measure. We define  $P_{U|S}$  by taking  $U = S + N$ , where  $N \sim \text{Ber}(q)$  is independent of  $S$ .

*Proposition 1:* For  $0 \leq t \leq h_b(p)$

$$F_P(t) \geq t - h_b(q * p) + h_b(q * h^{-1}(h_b(p) - t)), \quad (29)$$

with equality for  $p = 1/2$ .

**Proof.** By the Markov structure, we have that  $I(S; V) = I(U; V) + I(S; V|U)$ . Thus,

$$\begin{aligned} I(S; V|U) &= I(S; V) - H(U) + H(U|V) \\ &\geq I(S; V) - H(U) + h_b(q * h^{-1}(H(S|V))) \\ &= I(S; V) - H(U) + h_b(q * h^{-1}(H(S) - I(S; V))) \\ &= I(S; V) - h_b(q * p) + h_b(q * h^{-1}(h_b(p) - I(S; V))), \end{aligned}$$

where the inequality follows from Mrs. Gerber's Lemma [19]. Note that equality holds iff  $H(S|V = v) = H(S|V)$  for all  $v \in \mathcal{V}$ , which is the case for  $p = 1/2$  and  $V = S + A$ , where  $A \sim \text{Ber}(h_b^{-1}(1 - I(S; V)))$ . ■

*Proposition 2:* For  $0 \leq D \leq p$

$$\bar{R}_P(D) = h_b(q * p) - h_b(q * D). \quad (30)$$

**Proof.** For every  $P_{\hat{S}|S}$  satisfying the constraint  $\mathbb{E}(S + \hat{S}) \leq D$ , we must have that

$$\begin{aligned} I(U; \hat{S}) &= H(U) - H(U|\hat{S}) \\ &= H(U) - H(U + \hat{S}|\hat{S}) \\ &\geq H(U) - H(U + \hat{S}) \\ &= H(U) - H(N + S + \hat{S}) \\ &= h_b(q * p) - h_b(q * \mathbb{E}(S + \hat{S})) \\ &\geq h_b(q * p) - h_b(q * D). \end{aligned}$$

To see that this lower bound is tight, take the reverse test channel  $S = \hat{S} + V$  where  $V \sim \text{Ber}(D)$ . ■

## B. The Channel Function and Derivation of Theorem 2

In the general (non-memoryless, non-degraded) case the channel function  $G_{Q^n}$  (17) does not tensorise. However, we have the following basic properties.

*Lemma 4:* The function  $G_{Q^n}(t)$  is monotone non-increasing and concave.

**Proof.** Monotonicity of  $G_{Q^n}(t)$  follows by definition. For concavity, let  $(W_0, X_0)$  and  $(W_1, X_1)$  be random variables that attain  $G_{Q^n}(t_0)$  and  $G_{Q^n}(t_1)$  with equality. Let  $A \sim \text{Ber}(\alpha)$ , and define  $(\tilde{W}, \tilde{X}) = ((A, W_A), X_A)$ . We have that

$$\begin{aligned} I(\tilde{X}; Y_1|\tilde{W}) &= I(X_A; Y_1|W_A, A) \\ &= (1 - \alpha)I(X_0; Y_1|W_0) + \alpha I(X_1; Y_1|W_1) \\ &= (1 - \alpha)t_0 + \alpha t_1, \end{aligned} \quad (31)$$

and by definition of  $G_{Q^n}(t)$ ,

$$\begin{aligned}
G_{Q^n}(\alpha t_0 + (1 - \alpha)t_1) &= G_{Q^n}(I(\tilde{X}; Y_1 | \tilde{W})) \\
&\geq I(Y_2; \tilde{W}) \\
&= I(Y_2; A) + I(Y_2; W_A | A) \\
&\geq \alpha G_{Q^n}(t_0) + (1 - \alpha)G_{Q^n}(t_1).
\end{aligned} \tag{32}$$

■

We are now in a position to prove Theorem 2. The proof is essentially a generalization of the technique developed by Reznik, Feder and Zamir for the Gaussian joint source-channel coding problem [14]. Their proof relied heavily on the entropy power inequality, which is replaced by the functions  $F_P(t)$  and  $G_{Q^n}(t)$  in the proof below. We remark that although Theorem 2 is stated and proved for channels without a cost constraint, such a constraint can be included by constraining the distribution of  $X^n$  in the computation of  $G_{Q^n}(t)$ , in the obvious way.

**Proof of Theorem 2.** Let  $\hat{S}_1^m, \hat{S}_2^m$  be the estimates produced from the outputs  $Y_1^n$  and  $Y_2^n$ , respectively. We have

$$m\bar{R}_P(D_2) \leq I(U^m, \hat{S}_2^m) \tag{33}$$

$$\leq I(U^m; Y_2^n) \tag{34}$$

$$\leq G_{Q^n}(I(X^n; Y_1^n | U^m)) \tag{35}$$

$$\leq G_{Q^n}(I(S^m; Y_1^n | U^m)) \tag{36}$$

$$\leq G_{Q^n}(F_{P^m}(I(S^m; Y_1^n))) \tag{37}$$

$$= G_{Q^n}\left(mF_P\left(\frac{I(S^m; Y_1^n)}{m}\right)\right) \tag{38}$$

$$\leq G_{Q^n}\left(mF_P\left(\frac{I(S^m; \hat{S}_1^m)}{m}\right)\right) \tag{39}$$

$$\leq G_{Q^n}(mF_P(R(D_1))),$$

where (33) follows from Lemma 3, (34) follows from the data processing inequality (DPI), (35) from definition of  $G_{Q^n}$ , (36) from the DPI and monotonicity of  $G_{Q^n}$ , (37) from definition of  $F_{P^m}$ , (38) from tensorization of  $F_{P^m}$ , and (39) from the DPI. ■

Note that  $U^m$  plays a two-fold role here: in (35) we used the Markov chain  $U^m - X^n - Y_1^n - Y_2^n$ , whereas in (36) we used  $U^m - S^m - Y_1^n$ . Thus, the source function  $F_P(t)$ , and the broadcast function  $G_{Q^n}(t)$  are *coupled* via the same auxiliary variable  $U^m$ . This is also the main weakness of the bound above: Even though the same  $U^m$ , whose distribution is fixed once we choose the channel  $P_{U|S}$ , appears in both Markov chains, in the transition from (34) to (35), we have used the definition of  $G_{Q^n}$ , which involves a *maximization* with respect to  $U^m$ .

### C. Digression: Degraded Memoryless Channels

We now consider the special case where the channel  $Q^n$  is degraded and memoryless. i.e.,  $Q_{Y_1^n, Y_2^n | X^n}^n(y_1^n, y_2^n | x^n) = \prod_{i=1}^n Q_{Y_1 | X}(y_{1i} | x_i) Q_{Y_2 | Y_1}(y_{2i} | y_{1i})$ . Although this case is not required for our main result, we bring it as a demonstration of the power of Theorem 2.

We start with the the following lemma, which can be viewed as a restatement of the degraded broadcast channel converse theorem.

*Lemma 5:* If  $Q^n$  is a degraded memoryless broadcast channel, the function  $F_Q$  tensorizes, i.e.,

$$G_{Q^n}(nt) = nG_Q(t).$$

**Proof.** We use induction. For any  $(W, X^n)$  satisfying the Markov chain  $W - X^n - Y_1^n - Y_2^n$  we have

$$I(X^n; Y_1^n | W) = I(Y_1^{n-1}; X^{n-1} | W) + I(X_n; Y_{1,n} | W, Y_1^{n-1}).$$

Consequently,

$$\begin{aligned}
& G_Q \left( \frac{I(X^n; Y_1^n | W)}{n} \right) \\
&= G_Q \left( \frac{I(Y_1^{n-1}; X^{n-1} | W) + I(X_n; Y_{1,n} | W, Y_1^{n-1})}{n} \right) \\
&= G_Q \left( \frac{n-1}{n} \frac{I(Y_1^{n-1}; X^{n-1} | W)}{n-1} + \frac{I(X_n; Y_{1,n} | W, Y_1^{n-1})}{n} \right) \\
&\geq \frac{n-1}{n} G_Q \left( \frac{I(Y_1^{n-1}; X^{n-1} | W)}{n-1} \right) \\
&\quad + \frac{1}{n} G_Q (I(X_n; Y_{1,n} | W, Y_1^{n-1}))
\end{aligned} \tag{40}$$

where we have used the concavity of  $t \mapsto G_Q(t)$  in the last inequality. Invoking the induction hypothesis, we get

$$\begin{aligned}
& G_Q \left( \frac{I(X^n; Y_1^n | W)}{n} \right) \\
&= \frac{G_Q^{n-1} (I(Y_1^{n-1}; X^{n-1} | W)) + G_Q (I(X_n; Y_{1,n} | W, Y_1^{n-1}))}{n} \\
&\geq \frac{I(Y_2^{n-1}, W) + I(Y_{2,n}; W, Y_1^{n-1})}{n},
\end{aligned} \tag{41}$$

where the last inequality follows from the definition of  $G_{Q^{n-1}}(t)$  and  $G_Q(t)$  and the fact that  $W - X^n - Y_1^{n-1} - Y_2^{n-1}$  and  $(W, Y_1^{n-1}) - X_n - Y_{1,n} - Y_{2,n}$  are indeed Markov chains. Note that we have the Markov chain  $Y_{2,n} - (W, Y_1^{n-1}) - Y_2^{n-1}$ , and therefore

$$\begin{aligned}
I(Y_{2,n}; W, Y_1^{n-1}) &\geq I(Y_{2,n}; W, Y_2^{n-1}) \\
&\geq I(Y_{2,n}; W | Y_2^{n-1}).
\end{aligned} \tag{42}$$

Substituting (42) into (41) gives

$$\begin{aligned}
n G_Q \left( \frac{I(X^n; Y_1^n | W)}{n} \right) &\geq I(Y_2^{n-1}, W) + I(Y_{2,n}; W | Y_2^{n-1}) \\
&= I(Y_2^n; W),
\end{aligned}$$

as desired. ■

The following corollary is an immediate consequence of Theorem 2 and Lemma 5.

*Corollary 1:* Consider the degraded memoryless JSCC broadcast problem. If  $(D_1, D_2)$  is achievable, then for any  $P = P_S P_{U|S}$ , defined by a choice of an auxiliary channel  $P_{U|S}$ ,

$$\bar{R}_P(D_2) \leq \rho \cdot G_Q \left( \frac{F_P(R(D_1))}{\rho} \right). \tag{43}$$

This bound can be obtained as a special case of [15, Theorem 5] (see also [16]), by observing that the boundary of the degraded memoryless broadcast channel  $Q$  (without common message) is given by  $(C_1, G_Q(C_1))$  [18, Theorem 5.2].<sup>3</sup>

It is not difficult to see that the separation bounds for the extreme cases where only one distortion is of interest are obtained by setting  $U = \emptyset$  or  $U = S$  for  $D_1$  and  $D_2$ , respectively. When  $\rho = 1$  and the “not to code” conditions [22] hold, these choices give the tightest bound possible. Otherwise, other choices can give tighter bounds, as demonstrated in the examples below.

<sup>3</sup>In fact, the techniques developed in [15] should suffice to establish our Theorem 2. We nevertheless found it more convenient to prove the theorem using properties of the general functions  $F_P(t)$ ,  $\bar{R}_P(D)$ , and  $G_{Q^n}(t)$ , as those functions have a major role in other problems in network information theory, see [1] for more details.

1) *Quadratic Gaussian Case:* Let  $S \sim \mathcal{N}(0, \sigma^2)$ , and  $d(S_j, \hat{S}_j) = (S_j - \hat{S}_j)^2$ . We choose  $U$  that is the output of an AWGN channel with input  $S$  and noise that is Gaussian  $(0, \delta^2)$ . Using the EPI, one can verify that the corresponding source functions satisfy

$$F_P(t) = t - \frac{1}{2} \log \left( \frac{\delta^2 + \sigma^2}{\delta^2 + \sigma^2 e^{-t}} \right)$$

$$\bar{R}_P(D) = \frac{1}{2} \log \left( \frac{\delta^2 + \sigma^2}{\delta^2 + D} \right),$$

where  $F_P$  is attained by taking  $V$  that is the output of an AWGN with input  $S$ . Furthermore, let  $Q^n$  be the (memoryless degraded) AWGN broadcast channel,  $Y_1 = X + Z_1$ ,  $Y_2 = Y_1 + Z_2$ , where  $Z_1 \sim \mathcal{N}(0, N_1)$ ,  $Z_2 \sim \mathcal{N}(0, N_2)$ ,  $(X, Z_1, Z_2)$  mutually independent, where the channel input is subject to a quadratic cost constraint  $P$ . Using the EPI again, one can verify that

$$G_Q(t) = \frac{1}{2} \log \left( \frac{P + N_1 + N_2}{N_1 e^{2t} + N_2} \right),$$

where the function is attained by  $(W, X)$  that are jointly Gaussian. Combining with the source functions above and with the quadratic-Gaussian rate-distortion function, and applying Corollary 1, we recover the Reznic, Feder, Zamir outer bound [14, Theorem 1]: For all  $\delta$ ,

$$\frac{\delta^2 + \sigma^2}{\delta^2 + D_2} \leq \left( 1 + \frac{P}{N_1 + N_2} \right)^\rho \left[ \frac{N_1 + N_2}{N_1 \left( \frac{\sigma^2}{D_1} \cdot \frac{\delta^2 + D_1}{\delta^2 + \sigma^2} \right)^{\frac{1}{\rho}} + N_2} \right]^\rho.$$

2) *Binary-Hamming Case:* Consider first the case where  $Q^n$  is a (memoryless degraded) erasure broadcast channel, i.e.,  $Y_i$  is  $X$  w.p.  $1 - \epsilon_i$  and erased otherwise, for  $i = 1, 2$ , where  $\epsilon_2 \geq \epsilon_1$ , and the source is i.i.d. Bernoulli ( $p$ ), and the Hamming distortion measure is used. One can verify that:

$$G_Q(t) = \frac{1 - \epsilon_2}{1 - \epsilon_1} (\log 2 - \epsilon_1 - t).$$

Combining with propositions 1 and 2 and substituting in Corollary 1, one obtains the bound (for  $p = 1/2$ ):

$$\frac{\log 2 - h_b(D_2 * q)}{(1 - \epsilon_2) \log 2} + \frac{h_b(D_1 * q) - h_b(D_1)}{(1 - \epsilon_1) \log 2} \leq \rho,$$

which recovers the bound of [23] (which was also recovered in [16]).

Next, we consider the (memoryless degraded) binary symmetric channel,  $Y_1 = X + Z_1$ , and  $Y_2 = Y_1 + Z_2$ , where  $Z_1 \sim \text{Ber}(\delta_1)$ ,  $Z_2 \sim \text{Ber}(\delta_2)$ , and  $(X, Z_1, Z_2)$  are mutually independent.

*Proposition 3:* For the binary symmetric degraded channel

$$G_Q(t) = \log 2 - h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + t)), \quad (44)$$

for  $0 \leq t \leq \log 2 - h_b(\delta_1)$ .

**Proof.** For any  $(W, X)$  satisfying the Markov chain  $W - X - (Y_1 = X + Z_1) - (Y_2 = Y_1 + Z_2)$ , we have

$$\begin{aligned} H(Y_2|W) &\geq h_b(\delta_2 * h_b^{-1}(H(Y_1|W))) \\ &= h_b(\delta_2 * h_b^{-1}(H(Y_1|X) + H(Y_1|W) - H(Y_1|X, W))) \\ &= h_b(\delta_2 * h_b^{-1}(H(Y_1|X) + I(X; Y_1|W))) \\ &= h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + I(X; Y_1|W))), \end{aligned}$$

where the inequality stems from Mrs. Gerber's Lemma and the fact that  $Y_2 = Y_1 + Z_2$ , with equality if  $X \sim \text{Ber}(1/2)$  and  $W = X + A$  for  $A \sim \text{Ber}(\eta)$ , where  $I(Y_2; W) = \log 2 - h_b(\eta * \delta_1 * \delta_2)$ . Noticing that  $I(Y_2; W) = H(Y_2) - H(Y_2|W) \leq \log 2 - H(Y_2|W)$ , with equality for  $X \sim \text{Ber}(1/2)$ , the proof is completed. ■

We can now combine this result with propositions 1 and 2 and substitute in Corollary 1, to obtain the following theorem.

*Theorem 5:* For the JSCC broadcast problem with a  $\text{Ber}(p)$  source, Hamming distortion and a binary symmetric channel, suppose that the pair  $(D_1, D_2)$  is achievable. Then, for any  $0 \leq q \leq 1/2$ , it holds that

$$h_b(q * p) - h_b(q * D_2) \leq \rho \left[ \log 2 - h_b(\delta_2 * h_b^{-1}(A_1)) \right],$$

where

$$A_1 = h_b(\delta_1) + \frac{1}{\rho} [h(q * D_1) - h(D_1) - h(q * p) + h(p)].$$

For  $p = 1/2$ , the bound significantly simplifies as on the left hand side  $h_b(q * p) = \log 2$ , while on the right hand side

$$A_1 = h_b(\delta_1) + \frac{h_b(q * D_1) - h_b(D_1)}{\rho}. \quad (45)$$

Following the treatment of the Gaussian-quadratic case in [14], we consider the case where the distortion of the “weak” user is optimal. That is, let  $D_2^* = D(\rho, \delta_1 * \delta_2)$ , where the function  $D(\rho, \delta)$  is as defined in (7). For the special case of  $D_2 = D_2^*$ . We can take  $q \rightarrow 0$  in Theorem 5, and applying some straightforward algebra, we obtain the following.

*Corollary 2:* For the JSCC broadcast problem with a binary source and a binary symmetric channel, suppose that the pair  $(D_1, D_2^*)$  is achievable, where  $D_2^* = D(\rho, \delta_1 * \delta_2)$ . Then,

$$g(D_1) \geq g(p) + \frac{g(\delta_1)}{g(\delta_1 * \delta_2)} [g(D_2^*) - g(p)], \quad (46)$$

where  $g(t) \triangleq (1 - 2t) \log\left(\frac{1-t}{t}\right)$ .

Similarly, for the special case of  $D_1 = D_1^* = D(\rho, \delta_1)$ , we can take  $q \rightarrow 1/2$  in Theorem 5, and after applying some straightforward algebra, obtain the following.

*Corollary 3:* For the JSCC broadcast problem with a binary source and a binary symmetric channel, suppose that the pair  $(D_1^*, D_2)$  is achievable, where  $D_1^* = D(\rho, \delta_1)$ . Then,

$$(1 - 2D_2)^2 \leq (1 - 2 \cdot \delta_2 * D_1^*)^2 + (1 - 2p)^2 (1 - (1 - 2 \cdot \delta_2 * D_1^*)^2). \quad (47)$$

In particular, for  $p = 1/2$ ,

$$D_2 \geq \delta_2 * D_1^*. \quad (48)$$

#### IV. OUTER BOUNDS FOR THE BINARY SPHERICAL NOISE JSCC BROADCAST PROBLEM

In this section we derive a lower bound on the quantity  $\bar{D}(a)$ , defined in (12), as outlined in Section II-B. To that end, we first establish Theorem 3, and then use it to prove the desired lower bound.

##### A. Proof of Theorem 3

Recall that the binary spherical noise JSCC broadcast problem consists of a symmetric binary source  $S$  and Hamming distortion, for which the corresponding functions  $F_P(t)$  and  $\bar{R}_P(D)$  were already evaluated in Proposition 1 and Proposition 2, under the choice  $U = S + N$ , for  $N \sim \text{Ber}(q)$ , statistically independent of  $S$ . The corresponding additive spherical noise broadcast channel is  $Q^n = Q_{Y_1^n, Y_2^n | X^n}(y_1^n, y_2^n | x^n)$  with  $Y_1^n = X^n + U_1^n$ ,  $Y_2^n = X^n + U_2^n$ , where  $U_1^n \sim \text{Uniform}(\mathcal{S}_{n\delta_1, n})$  and  $U_2^n \sim \text{Uniform}(\mathcal{S}_{n(\delta_1 * \delta_2), n})$ ,  $(U_1^n, U_2^n) \perp X^n$ . The function  $G_{Q^n}(t)$  for this channel is upper bounded in Lemma 1, which we next prove. Combined with Proposition 1, Proposition 2 and Theorem 2, Theorem 3 is immediately obtained.

**Proof of Lemma 1.** 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

$$H(Y_2^n | W) = H(X^n + U_1^n + Z_3^n | W) + [H(X^n + U_2^n | W) - H(X^n + U_1^n + Z_3^n | W)], \quad (49)$$

where  $Z_3^n \sim \text{Ber}(\delta_2)^{\otimes n}$ . We will upper bound the absolute value of the term in the square brackets 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

$$\begin{aligned} \mathbb{E}w_H(U_2^n + (U_1^n + Z_3^n)) &= \mathbb{E}|T - n(\delta_1 * \delta_2)| \\ &= \mathbb{E}\sqrt{(T - n(\delta_1 * \delta_2))^2} \\ &\leq \sqrt{\text{Var}(T)} \\ &= \sqrt{\text{Var}(T_1) + \text{Var}(T_2)} \\ &= \sqrt{n\delta_2(1 - \delta_2)} \\ &\leq \sqrt{n\delta_2}, \end{aligned} \quad (50)$$

where the first inequality follows from Jensen's inequality and the fact that  $\mathbb{E}(T) = n(\delta_1 * \delta_2)$ . Now, applying [20, Proposition 8] (see also [21]), we obtain for  $\delta_2 < 1/2$

$$|H(X^n + U_2^n|W) - H(X^n + U_1^n + Z_3^n|W)| \leq \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right). \quad (51)$$

Thus, we can use Mrs. Gerber's Lemma (MGL) to lower bound (49) as

$$\begin{aligned} H(Y_2^n|W) &\geq H(X^n + U_1^n + Z_3^n|W) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right) \\ &\geq n\varphi_{\delta_2}\left(\frac{H(X^n + U_1^n|W)}{n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right), \end{aligned} \quad (52)$$

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

$$\begin{aligned} H(Y_2^n|W) &= n\varphi_{\delta_2}\left(\frac{H(X^n + U_1^n|X^n) + H(X^n + U_1^n|W) - H(X^n + U_1^n|X^n, W)}{n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right) \\ &= n\varphi_{\delta_2}\left(\frac{H(U_1^n) + I(X^n; Y_1^n|W)}{n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right) \\ &\geq n\varphi_{\delta_2}\left(\frac{nh_b(\delta_1) - \frac{1}{2}(\log n + 1) + I(X^n; Y_1^n|W)}{n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right) \\ &= n\varphi_{\delta_2}\left(h_b(\delta_1) + \frac{I(X^n; Y_1^n|W)}{n} - \frac{\log n + 1}{2n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right) \\ &\geq n\varphi_{\delta_2}\left(h_b(\delta_1) + \frac{I(X^n; Y_1^n|W)}{n}\right) - \frac{\log n + 1}{2}\varphi'_{\delta_2}\left(h_b(\delta_1) + \frac{I(X^n; Y_1^n|W)}{n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right), \end{aligned} \quad (53)$$

where in (53) we substitute [24, Chapter 10, Lemma 7] to lower-bound  $H(U_1^n)$ , and in the last inequality we have defined the MGL derivative  $\varphi'_{\delta_2}(x) = \frac{d}{dx}\varphi_{\delta_2}(x)$ , and used the convexity of  $x \mapsto \varphi_{\delta_2}(x)$  [19]. Recalling that  $\varphi'_{\delta_2}(x) \leq 1$  due to [25, Theorem 2.6] (or alternatively, as can be seen directly from the expression for  $\varphi'_{\delta_2}(x)$  derived in [19]), we have obtained

$$H(Y_2^n|W) \geq nh_b\left(\delta_2 * h_b^{-1}\left(h_b(\delta_1) + \frac{I(X^n; Y_1^n|W)}{n}\right)\right) - n\Gamma(n, \delta_2), \quad (54)$$

and the claim now follows since  $H(Y_2^n) \leq n \log 2$ . ■

## B. Local Analysis

Our next goal is to manipulate the bound from Theorem 3 in order to obtain a lower bound on  $\bar{D}(a)$ . The first step is to derive a constraint on the achievable  $(D_1, D_2)$  pairs for the binary spherical noise JSCC broadcast problem. The proof of the following Lemma is brought in Appendix C, and is based on several auxiliary lemmas, which are stated and proved in Appendix B.

*Lemma 6:* 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  $U_1^n \sim$

Uniform  $(\mathcal{S}_{n\delta_1, n})$  and  $U_2^n \sim \text{Uniform}(\mathcal{S}_{n(\delta_1 * \delta_2), n})$ ,  $(U_1^n, U_2^n) \perp\!\!\!\perp X^n$ , and  $\rho = n/m$ . If  $(D_1, D_2)$  is achievable, then then for any  $\tau > 0$  we have

$$D_2 - D_1 \leq \frac{1 + 2D_2\tau}{2D_2\tau} \frac{\rho C(\delta_1 * \delta_2) - R(D_2) + \rho\Gamma(n, \delta_2)}{\log\left(\frac{1-D_2}{D_2}\right)} + (\delta_1 * \delta_2 - \delta_1) \frac{\log\left(\frac{1-\delta_1}{\delta_1}\right)}{\log\left(\frac{1-D_2}{D_2}\right)} \frac{\Phi(\delta_1)}{\Phi(D_2)} (1 + \tau) \frac{g(D_1)}{g(D_2)}, \quad (55)$$

where  $\Phi(t)$  is defined in (8a),  $\Gamma(n, \delta_2)$  is defined in (20), and

$$C(t) = R(t) \triangleq \log 2 - h_b(t), \quad (56)$$

$$g(t) \triangleq (1 - 2t) \log\left(\frac{1-t}{t}\right) = (1 - 2t)h'_b(t). \quad (57)$$

We will also need the following lemma, proved in Appendix D.

*Lemma 7:* Let  $\Phi(t)$  and  $D(\rho, \delta)$  be as defined in (8a) and (7), respectively. For every  $\rho > 1$  and  $0 < \delta < 1/2$  for which  $D(\rho, \delta) > 0$ , it holds that

$$f(\rho, \delta) \triangleq \frac{1}{\rho} \frac{\Phi(\delta)}{\Phi(D(\rho, \delta))} < 1. \quad (58)$$

Finally, we will also need the next proposition, which is a simple variation of the source-channel separation theorem for spherical noise.

*Proposition 4:* For any  $\mathcal{D}$ ,  $\mathcal{E}$  and  $k \in [-n\delta, n(1 - \delta)]$  it holds that

$$\Psi(k) \geq D\left(\rho, \delta + \frac{k}{n}\right) + \mathcal{O}\left(\frac{\log n}{n}\right), \quad (59)$$

where  $\Psi(\cdot)$  was defined in (10).

**Proof.** By the definition of  $\Psi(k)$  and the binary symmetric rate-distortion function, we have that for any encoder and decoder pair  $(\mathcal{E}, \mathcal{D})$  it holds that

$$I(S^m; \mathcal{D}(\mathcal{E}(S^m) + U_k^n)) \geq m(\log 2 - h_b(\Psi(k))). \quad (60)$$

Using the data processing inequality we obtain

$$I(S^m; \mathcal{D}(\mathcal{E}(S^m) + U_k^n)) \leq I(\mathcal{E}(S^m); \mathcal{E}(S^m) + U_k^n). \quad (61)$$

Recalling that [24, Chapter 10, Lemma 7]

$$\begin{aligned} H(U_k^n) &\geq nh_b\left(\delta + \frac{k}{n}\right) - \frac{\log n}{2} - \frac{1}{2} \log\left(2\pi\left(\delta + \frac{k}{n}\right)\left(1 - \delta - \frac{k}{n}\right)\right) \\ &\geq nh_b\left(\delta + \frac{k}{n}\right) - \frac{\log n}{2} - \frac{1}{2}, \end{aligned} \quad (62)$$

we have that for any random vector  $X^n$  in  $\{0, 1\}^n$

$$\begin{aligned} I(X^n; X^n + U_k^n) &\leq H(X^n + U_k^n) - H(U_k^n) \\ &\leq n \log 2 - h_b\left(\delta + \frac{k}{n}\right) + \frac{\log n}{2} + \frac{1}{2}. \end{aligned} \quad (63)$$

Thus, combining (60), (61), and (63), and recalling that  $\rho = n/m$ , we have

$$\log 2 - h_b(\Psi(k)) \leq \rho \left( \log 2 - h_b\left(\delta + \frac{k}{n}\right) + \frac{\log n + 1}{2n} \right), \quad (64)$$

which yields the desired result. ■

Using Lemma 6, Lemma 7, and Proposition 4, we can now prove Theorem 4, which is the main result of this subsection.

**Proof of Theorem 4.** First note that for  $\epsilon > 0$ , we can approximate  $D(\rho, \delta + \epsilon)$  as

$$D(\rho, \delta + \epsilon) = D(\rho, \delta) + \epsilon D'(\rho, \delta) + O(\epsilon^2), \quad (65)$$

where

$$D'(\rho, \delta) \triangleq \frac{\partial}{\partial \delta} D(\rho, \delta) = \rho \frac{h'_b(\delta)}{h'_b(D(\rho, \delta))} = \rho \frac{\log\left(\frac{1-\delta}{\delta}\right)}{\log\left(\frac{1-D(\rho, \delta)}{D(\rho, \delta)}\right)}. \quad (66)$$

Taking  $\epsilon = \frac{a\eta}{\sqrt{n}}$ , where  $\eta = \eta(\rho, \delta)$ , this implies that

$$D(\rho, \delta) - D(\rho, \delta_1) = \frac{a\eta}{\sqrt{n}} D'(\rho, \delta) + \mathcal{O}\left(\frac{1}{n}\right) \quad (67a)$$

$$D(\rho, \delta_1 * \delta_2) - D(\rho, \delta) = \frac{a\eta}{\sqrt{n}} D'(\rho, \delta) + \mathcal{O}\left(\frac{1}{n}\right). \quad (67b)$$

and thus

$$D(\rho, \delta_1 * \delta_2) - D(\rho, \delta_1) = \frac{2a\eta}{\sqrt{n}} D'(\rho, \delta) + \mathcal{O}\left(\frac{1}{n}\right), \quad (68)$$

Now, by Proposition 4 we have:

$$D_1 \geq D(\rho, \delta_1) + \mathcal{O}\left(\frac{\log n}{n}\right) \quad (69a)$$

$$D_2 \geq D(\rho, \delta_1 * \delta_2) + \mathcal{O}\left(\frac{\log n}{n}\right). \quad (69b)$$

Using (67b) and (69b), we can assert:

$$\begin{aligned} D_2 - D(\rho, \delta) &= (D_2 - D(\rho, \delta_1 * \delta_2)) + (D(\rho, \delta_1 * \delta_2) - D(\rho, \delta)) \\ &\geq \frac{a}{\sqrt{n}} D'(\rho, \delta) + \mathcal{O}\left(\frac{\log n}{n}\right). \end{aligned} \quad (70)$$

We now claim that we can assume without loss of generality:

$$D_1 - D(\rho, \delta_1) < \frac{a\eta}{\sqrt{n}} \quad (71a)$$

$$D_2 - D(\rho, \delta_1 * \delta_2) < \frac{a\eta}{\sqrt{n}}, \quad (71b)$$

To see why this is true, assume to the contrary that one of them, say the first, does not hold. Then, by (69b) and (68),

$$\begin{aligned} D_1 + D_2 &\geq D(\rho, \delta_1) + D_2 + \frac{a\eta}{\sqrt{n}} \\ &\geq D(\rho, \delta_1) + D(\rho, \delta_1 * \delta_2) + \frac{a\eta}{\sqrt{n}} + \mathcal{O}\left(\frac{\log n}{n}\right) \\ &= 2D(\rho, \delta) + \frac{a\eta}{\sqrt{n}} + \mathcal{O}\left(\frac{\log n}{n}\right), \end{aligned}$$

which is stronger than the desired bound.

We now proceed to bound the difference  $D_2 - D_1$  invoking Lemma 6 and using (71). By the concavity of  $t \mapsto h_b(t)$  and (71b) we have that

$$\begin{aligned} h_b(D_2) &= h_b(D(\rho, \delta_1 * \delta_2) + (D_2 - D(\rho, \delta_1 * \delta_2))) \\ &\leq h_b(D(\rho, \delta_1 * \delta_2)) + \frac{a\eta}{\sqrt{n}} h'_b(D(\rho, \delta_1 * \delta_2)), \end{aligned} \quad (72)$$

which implies that

$$\begin{aligned} \rho C(\delta_1 * \delta_2) - R(D_2) &= \rho C(\delta_1 * \delta_2) - \log 2 + h_b(D_2) \\ &\leq \frac{a\eta}{\sqrt{n}} h'_b(D(\rho, \delta_1 * \delta_2)) \\ &= \frac{a\eta}{\sqrt{n}} \log\left(\frac{1 - D(\rho, \delta_1 * \delta_2)}{D(\rho, \delta_1 * \delta_2)}\right). \end{aligned} \quad (73)$$



Substituting in Lemma 6, we have for any  $\tau > 0$ :

$$D_2 - D_1 \leq \frac{1 + 2D_2\tau}{2D_2\tau} \frac{\frac{a\eta}{\sqrt{n}} \log\left(\frac{1-D(\rho, \delta_1 * \delta_2)}{D(\rho, \delta_1 * \delta_2)}\right) + \rho\Gamma(n, \delta_2)}{\log\left(\frac{1-D_2}{D_2}\right)} + \frac{2a}{\sqrt{n}} \frac{\log\left(\frac{1-\delta_1}{\delta_1}\right)}{\log\left(\frac{1-D_2}{D_2}\right)} \frac{\Phi(\delta_1)}{\Phi(D_2)} (1 + \tau) \frac{g(D_1)}{g(D_2)}. \quad (74)$$

The functions  $t \mapsto \log\left(\frac{1-t}{t}\right)$ ,  $t \mapsto \Phi(t)$ , and  $t \mapsto g(t)$  are continuous at  $0 < t < 1/2$ . Thus, recalling that  $a < \log^2(n)$ , by the assumption that  $0 < D(\rho, \delta) < 1/2$ , (71a) and (71b), we have that

$$\begin{aligned} \frac{\log\left(\frac{1-D(\rho, \delta_1 * \delta_2)}{D(\rho, \delta_1 * \delta_2)}\right)}{\log\left(\frac{1-D_2}{D_2}\right)} &= 1 + \mathcal{O}\left(\frac{\log^2(n)}{\sqrt{n}}\right) \\ \frac{\log\left(\frac{1-\delta_1}{\delta_1}\right)}{\log\left(\frac{1-D_2}{D_2}\right)} &= \frac{\log\left(\frac{1-\delta}{\delta}\right)}{\log\left(\frac{1-D(\rho, \delta)}{D(\rho, \delta)}\right)} \left(1 + \mathcal{O}\left(\frac{\log^2(n)}{\sqrt{n}}\right)\right) \\ &= \frac{D'(\rho, \delta)}{\rho} \left(1 + \mathcal{O}\left(\frac{\log^2(n)}{\sqrt{n}}\right)\right) \\ \frac{\Phi(\delta_1)}{\Phi(D_2)} &= \frac{\Phi(\delta)}{\Phi(D(\rho, \delta))} \left(1 + \mathcal{O}\left(\frac{\log^2(n)}{\sqrt{n}}\right)\right) \\ &= \rho f(\rho, \delta) \left(1 + \mathcal{O}\left(\frac{\log^2(n)}{\sqrt{n}}\right)\right), \end{aligned}$$

where  $f(\cdot, \cdot)$  was defined in (58). In addition, under our assumptions on  $\delta_1$  and  $\delta_2$ , we have that  $\Gamma(n, \delta_2) = \mathcal{O}(n^{-3/4} \log n)$ . Thus, (74) amounts to the following upper bound on the difference  $D_2 - D_1$ :

$$D_2 - D_1 \leq \frac{1}{\sqrt{n}} \left( a\eta \frac{1 + 2D(\rho, \delta)\tau}{2D(\rho, \delta)\tau} + 2aD'(\rho, \delta)f(\rho, \delta)(1 + \tau) \right) + \mathcal{O}(n^{-3/4} \log n). \quad (75)$$

Combining with (70) now yields

$$\begin{aligned} D_1 + D_2 - 2D(\rho, \delta) &= 2(D_2 - D(\rho, \delta)) - (D_2 - D_1) \\ &\geq \frac{2a}{\sqrt{n}} D'(\rho, \delta) - \frac{1}{\sqrt{n}} \left( a\eta \frac{1 + 2D(\rho, \delta)\tau}{2D(\rho, \delta)\tau} + 2aD'(\rho, \delta)f(\rho, \delta)(1 + \tau) \right) + \mathcal{O}(n^{-3/4} \log n) \\ &= \frac{1}{\sqrt{n}} \left( 2aD'(\rho, \delta) (1 - f(\rho, \delta)(1 + \tau)) - a\eta \frac{1 + 2D(\rho, \delta)\tau}{2D(\rho, \delta)\tau} \right) + \mathcal{O}(n^{-3/4} \log n). \end{aligned} \quad (76)$$

Now, since  $0 < f(\rho, \delta) < 1$  by Lemma 7, we can take

$$\tau = \frac{1 - f(\rho, \delta)}{2f(\rho, \delta)} > 0, \quad (77)$$

for which (76) becomes

$$D_1 + D_2 - 2D(\rho, \delta) \geq \frac{a\eta(\rho, \delta)}{\sqrt{n}} + \mathcal{O}(n^{-3/4} \log n), \quad (78)$$

as desired. ■

## V. PROOF OF THEOREM 1

The outline of the proof was already given in Section II-A. Here we give the complete proof. First, without loss of generality we may restrict attention to  $\delta \in \{\frac{0}{n}, \frac{1}{n}, \dots, 1\}$  as

$$D^*(n, \rho, \delta) \geq D^*\left(n, \rho, \frac{\lfloor n\delta \rfloor}{n}\right), \quad (79)$$

and we may therefore write

$$\begin{aligned}
\Delta^*(n, \rho, \delta) &= D^*(n, \rho, \delta) - D(\rho, \delta) \\
&\geq D^*\left(n, \rho, \frac{\lfloor n\delta \rfloor}{n}\right) - D\left(\rho, \frac{\lfloor n\delta \rfloor}{n}\right) + \left[D\left(\rho, \frac{\lfloor n\delta \rfloor}{n}\right) - D(\rho, \delta)\right] \\
&= \Delta^*\left(n, \rho, \frac{\lfloor n\delta \rfloor}{n}\right) + \mathcal{O}\left(\frac{1}{n}\right),
\end{aligned} \tag{80}$$

where the last equality follows since  $D\left(\rho, \frac{\lfloor n\delta \rfloor}{n}\right) - D(\rho, \delta) = \mathcal{O}\left(\frac{1}{n}\right)$ .

As in Section II-A, we define the integer-valued random variable  $K = w_H(Z^n) - \delta n$ , and for a given encoder/decoder pair  $(\mathcal{E}, \mathcal{D})$  we have that

$$\mathbb{E}\left[\frac{1}{m}d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + Z^n))\right] = \mathbb{E}[\Psi(K)], \tag{81}$$

where

$$\Psi(k) = \Psi_{\mathcal{D}, \mathcal{E}}(k) \triangleq \mathbb{E}\left[\frac{1}{m}d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + U_k^n))\right], \tag{82}$$

and  $U_k^n$  is uniform over  $\mathcal{S}_{\delta n+k, n}$ . In terms of this function, we have

$$\Delta^*(n, \rho, \delta) = \mathbb{E}[\Psi(K) - D(\rho, \delta)].$$

We proceed by partitioning  $K$  into two regimes:

$$\begin{aligned}
\mathcal{K}_1 &= \{K : |K| \leq \sqrt{n} \log^2(n)\} \\
\mathcal{K}_2 &= \{K : \sqrt{n} \log^2(n) < |K|\},
\end{aligned}$$

and asserting:

$$\Delta^*(n, \rho, \delta) = \sum_{i=1}^2 \Pr(K \in \mathcal{K}_i) \mathbb{E}[\Psi(K) - D(\rho, \delta) | K \in \mathcal{K}_i]. \tag{83}$$

and lower bound each of the two terms.

To lower bound the the second term it suffices to note that  $\Pr(K \in \mathcal{K}_2) = \mathcal{O}\left(\frac{1}{n}\right)$  and that  $\Psi(K) - D(\rho, \delta)$  is bounded, such that its the total contribution is at most  $\mathcal{O}\left(\frac{1}{n}\right)$ . For the first term, define for all natural  $k$

$$\gamma(k) = \frac{\Pr(K = k) - \Pr(K = -k)}{\Pr(K = k) + \Pr(K = -k)}, \tag{84}$$

and write:

$$\begin{aligned}
\mathbb{E}\left[\Psi(K) - D(\rho, \delta) \mid |K| = k_0\right] &= \Pr(K = k_0 \mid |K| = k_0) [\Psi(k_0) - D(\rho, \delta)] \\
&\quad + \Pr(K = -k_0 \mid |K| = k_0) [\Psi(-k_0) - D(\rho, \delta)] \\
&= \frac{1 + \gamma(k_0)}{2} [\Psi(k_0) - D(\rho, \delta)] + \frac{1 - \gamma(k_0)}{2} [\Psi(-k_0) - D(\rho, \delta)], \\
&= \frac{1 + \gamma(k_0)}{2} \left[ \left( \Psi(k_0) - D\left(\rho, \delta + \frac{k_0}{n}\right) \right) + \left( D\left(\rho, \delta + \frac{k_0}{n}\right) - D(\rho, \delta) \right) \right] \\
&\quad + \frac{1 - \gamma(k_0)}{2} \left[ \left( \Psi(-k_0) - D\left(\rho, \delta - \frac{k_0}{n}\right) \right) + \left( D\left(\rho, \delta - \frac{k_0}{n}\right) - D(\rho, \delta) \right) \right] \\
&= \frac{1 + \gamma(k_0)}{2} \left[ \left( \Psi(k_0) - D\left(\rho, \delta + \frac{k_0}{n}\right) \right) \right] + \frac{1 - \gamma(k_0)}{2} \left[ \left( \Psi(-k_0) - D\left(\rho, \delta - \frac{k_0}{n}\right) \right) \right] \\
&\quad + \frac{1}{2} \left[ \left( D\left(\rho, \delta + \frac{k_0}{n}\right) - D(\rho, \delta) \right) + \left( D\left(\rho, \delta - \frac{k_0}{n}\right) - D(\rho, \delta) \right) \right] \\
&\quad + \frac{\gamma(k_0)}{2} \left[ \left( D\left(\rho, \delta + \frac{k_0}{n}\right) - D(\rho, \delta) \right) - \left( D\left(\rho, \delta - \frac{k_0}{n}\right) - D(\rho, \delta) \right) \right].
\end{aligned} \tag{85}$$

In order to bound these quantities, note that

$$\left( D\left(\rho, \delta + \frac{k_0}{n}\right) - D(\rho, \delta) \right) + \left( D\left(\rho, \delta - \frac{k_0}{n}\right) - D(\rho, \delta) \right) = \mathcal{O}\left(\frac{k_0^2}{n^2}\right), \quad (86)$$

and that

$$\left( D\left(\rho, \delta + \frac{k_0}{n}\right) - D(\rho, \delta) \right) - \left( D\left(\rho, \delta - \frac{k_0}{n}\right) - D(\rho, \delta) \right) = \mathcal{O}\left(\frac{k_0}{n}\right). \quad (87)$$

In addition, in Appendix A we give Lemma 8, showing that for  $k_0 < n^{2/3}$ ,<sup>4</sup>

$$\gamma(k_0) = \mathcal{O}\left(\max\left(\frac{k_0}{n}, \frac{k_0^3}{n^2}\right)\right). \quad (88)$$

Applying (86)-(88) under the condition  $k_0 \leq \sqrt{n} \log^2(n)$ , and noting that by Proposition 4

$$\Psi(k_0) \geq D\left(\rho, \delta + \frac{k_0}{n}\right) + \mathcal{O}\left(\frac{\log n}{n}\right),$$

we see that (85) amounts to:

$$\begin{aligned} \mathbb{E} \left[ \Psi(K) - D(\rho, \delta) \mid |K| = k_0 \right] &\geq \frac{1 - |\gamma(k_0)|}{2} \left[ \left( \Psi(k_0) - D\left(\rho, \delta + \frac{k_0}{n}\right) \right) + \left( \Psi(-k_0) - D\left(\rho, \delta - \frac{k_0}{n}\right) \right) \right] \\ &\quad + \mathcal{O}\left(\frac{\log^8 n}{n}\right). \end{aligned} \quad (89)$$

Now, applying (86) again, and recalling that (88) amounts to  $\gamma(k_0) = \mathcal{O}\left(\frac{\log^6(n)}{\sqrt{n}}\right)$  in the regime  $k \in \mathcal{K}_1$ , we see that (89) can be further bounded as

$$\mathbb{E} \left[ \Psi(K) - D(\rho, \delta) \mid |K| = k_0 \right] \geq \frac{1}{2} [(\Psi(k_0) + \Psi(-k_0) - 2D(\rho, \delta))] \left( 1 + \mathcal{O}\left(\frac{\log^6(n)}{\sqrt{n}}\right) \right) + \mathcal{O}\left(\frac{\log^8 n}{n}\right). \quad (90)$$

Letting  $a = k_0/\sqrt{n}$  we have that by definition of  $\underline{D}(\cdot)$  (12),  $\Psi(k_0) + \Psi(-k_0) \geq \underline{D}(a)$ . Thus, the contribution of  $\mathcal{K}_1$  in (83) is at most:

$$\Pr(K \in \mathcal{K}_1) \mathbb{E} \left[ \underline{D}\left(\frac{K}{\sqrt{n}}\right) - 2D(\rho, \delta) \mid K \in \mathcal{K}_1 \right] \left( 1 + \mathcal{O}\left(\frac{\log^6(n)}{\sqrt{n}}\right) \right) + \mathcal{O}\left(\frac{\log^8 n}{n}\right). \quad (91)$$

Now, using Theorem 4, we have that for any fixed  $0 < a < \log^2(n)$ ,

$$\underline{D}(a) \geq 2D(\rho, \delta) + \frac{a}{\sqrt{n}} \eta(\rho, \delta) + \mathcal{O}(n^{-3/4} \log n) \quad (92)$$

where  $\eta(\rho, \delta)$  is as defined in (8c). Substituting in (91), we are left with:

$$\begin{aligned} &\frac{\eta(\rho, \delta)}{\sqrt{n}} \Pr(K \in \mathcal{K}_1) \mathbb{E} \left[ \frac{|K|}{\sqrt{n}} \mid K \in \mathcal{K}_1 \right] \left( 1 + \mathcal{O}\left(\frac{\log^6(n)}{\sqrt{n}}\right) \right) + \mathcal{O}\left(\frac{\log^8 n}{n}\right) \\ &= \frac{\sqrt{\delta(1-\delta)} \eta(\rho, \delta)}{\sqrt{n}} \Pr\left(|W| < \frac{\log^2(n)}{\sqrt{\delta(1-\delta)}}\right) \mathbb{E} \left[ |W| \mid |W| < \frac{\log^2(n)}{\sqrt{\delta(1-\delta)}} \right] \left( 1 + \mathcal{O}\left(\frac{\log^6(n)}{\sqrt{n}}\right) \right) + \mathcal{O}\left(\frac{\log^8 n}{n}\right), \end{aligned} \quad (93)$$

<sup>4</sup>A similar result can be shown for any distribution with a finite third moment using bounds on the (unsigned) Gaussian approximation error such as a theorem By Essen which appears in [26, Theorem 5.22]. However, we prefer to present the explicit calculation for the binomial distribution.

where the random variable  $W = \frac{1}{\sqrt{\delta(1-\delta)}} \frac{K}{\sqrt{n}}$  has zero mean and unit variance. Now let  $W_G$  be a standard Gaussian. By using non-uniform bounds on the rate of convergence in the central limit theorem such as the theorem of Bikelis (see [26, 5.10.4]), we have that

$$\begin{aligned} \Pr \left( |W| < \frac{\log^2(n)}{\sqrt{\delta(1-\delta)}} \right) \mathbb{E} \left[ |W| \middle| |W| < \frac{\log^2(n)}{\sqrt{\delta(1-\delta)}} \right] &= \Pr \left( |W_G| < \frac{\log^2(n)}{\sqrt{\delta(1-\delta)}} \right) \mathbb{E} \left[ |W_G| \middle| |W_G| < \frac{\log^2(n)}{\sqrt{\delta(1-\delta)}} \right] \\ &+ \mathcal{O} \left( \frac{1}{\sqrt{n}} \right) \\ &= \sqrt{\frac{2}{\pi}} + \mathcal{O} \left( \frac{1}{\sqrt{n}} \right). \end{aligned} \quad (94)$$

Thus we obtained

$$\Delta^*(n, \rho, \delta) \geq \sqrt{\frac{\delta(1-\delta)}{2\pi n}} \eta(\rho, \delta) + \mathcal{O}(n^{-3/4} \log n), \quad (95)$$

as desired.

## VI. DISCUSSION: THE REMAINING GAP TO ACHIEVABLE PERFORMANCE

In this work we have shown an example, where  $\Delta_n^* \geq \Omega(n^{-1/2})$ . It is natural, of course, to ask whether such performance is also achievable.

Consider a separation-based scheme: the source is quantized to a rate  $R_n$  with expected distortion  $D_{0,n}$ . This code is matched to a channel code with the same rate. Upon correct channel decoding we have distortion  $D_{0,n}$ , while incorrect decoding gives distortion that is trivially upper bounded by 1. If the channel error probability is  $p_n$ , this scheme yields

$$D_n \leq (1 - p_n)D_{0,n} + p_n. \quad (96)$$

Now we know that for the lossy source problem it is possible to achieve

$$D_{0,n} \leq D(\rho R_n) + \mathcal{O} \left( \frac{\log n}{n} \right),$$

Thus,

$$\begin{aligned} \Delta_n &\leq D_{0,n} - D_\infty^* + p_n [1 - D_{0,n}] \\ &= D(\rho R_n) - D(\rho C) + \mathcal{O}(p_n) + \mathcal{O} \left( \frac{\log n}{n} \right) \\ &= \mathcal{O}(C - R_n) + \mathcal{O}(p_n) + \mathcal{O} \left( \frac{\log n}{n} \right). \end{aligned} \quad (97)$$

In order for both first terms to decrease, we must choose  $R_n$  in the moderate-deviations regime. Using [27], [28] we have that:

$$\log p_n = \mathcal{O}(n(C - R_n)^2).$$

Substituting, we find that a separation-based scheme achieves

$$\Delta_n = \mathcal{O} \left( \sqrt{\frac{\log n}{n}} \right). \quad (98)$$

Next, one can consider the combination of successive-refinement (SR) source coding with a digital channel broadcast code, possibly with many layers to track well the channel quality, as done in the context of Gaussian channels in different formulations regarding the high signal-to-noise ratio regime [29], [30], [31]. However, one may verify that these techniques will not improve upon the order of convergence of a separation-based scheme due to the following consideration. The first layer of the SR code will have to be allocated a rate that is the same order below capacity as in the separation scheme. In order to reduce distortion, we will need a layer that will be correctly decoded when the empirical channel is above capacity in the same  $\sqrt{\log n/n}$  order. But substituting in the broadcast channel converse, it turns out that this refinement layer will be able to carry a very low rate, failing to reduce the distortion by the required amount.

We see that the limitation of digital schemes stems from two effects.

- 1) A broadcast code for two empirical channels that are symmetric around capacity, carries a sum-rate that is much lower than capacity.
- 2) The threshold effect: a digital JSCC scheme that performs well around capacity, cannot accommodate for bad channel conditions.

Indeed, recalling that our converse bound is based upon bounding the channel function  $G_Q$  that is intimately related to the converse for the digital BC problem, it can be seen as reflecting the first effect. However, it does not reflect the second, hence the remaining gap between the achievable and converse bounds, It remains to be seen, whether this threshold effect indeed applies to all relevant schemes.

One indication that the threshold effect might be unavoidable are the results of [32], which show that linear codes with any non-zero minimal distance necessarily admit medium-sized error vectors that result in maximal hamming distortion for the input bits. One approach for finding codes with graceful-degradation (or adaptation to channel conditions) was suggested in [33], see [34], [35] for more results.

As a possible coding scheme to improve upon separation, we can consider coding for the simplest case - a binary erasure channel (BEC). In this case, we only need to produce a good lossy source code which produces from  $m$  bits  $n$  coded bits with the following “fountain-code-like” property: if any subset of  $mR(D)$  of these coded bits is available (i.e. these positions are not erased by the BEC) then the source can be reconstructed with distortion  $mD$ , and this property should hold for a small range of  $D = D_\infty^* \pm c/\sqrt{n}$ . Since the channel returns  $nC + \sqrt{nV}Z$  unerased bits ( $Z \sim \mathcal{N}(0, 1)$ ), by averaging we would get distortion  $D_\infty^*(1 + o(1/\sqrt{n}))$ . So the only remaining task is to construct this “fountain-like” rate-distortion code. Whether such a code exists is an open problem, although known results about multiple-description problem (see, e.g., [36] in a Gaussian setting) suggest that the required property is not possible for all subsets of coded bits (without restriction on their size).

#### APPENDIX A THE ASYMMETRY OF THE BINOMIAL DISTRIBUTION

*Lemma 8:* For  $n\delta \in \mathbb{Z}$  define the integer random variable  $K = \text{Binomial}(n, \delta) - n\delta$ . For any integer  $0 \leq k < n^{2/3}$  it holds that

$$\frac{\Pr(K = k)}{\Pr(K = k) + \Pr(K = -k)} = \frac{1}{2} + \frac{1}{4} \frac{(1 - 2\delta)}{\delta(1 - \delta)} \left[ \frac{k^2}{n} \frac{1}{3\delta(1 - \delta)} - 1 \right] \frac{k}{n} + o\left(\frac{k^3}{n^2}\right).$$

**Proof.** Using the binomial distribution, we have

$$\begin{aligned} \Gamma &\triangleq \frac{\Pr(K = k)}{\Pr(K = -k)} \\ &= \frac{\binom{n}{n\delta+k} \left(\frac{\delta}{1-\delta}\right)^k}{\binom{n}{n\delta-k} \left(\frac{\delta}{1-\delta}\right)^{-k}} \\ &= \frac{(n\delta - k)! (n(1 - \delta) + k)!}{(n\delta + k)! (n(1 - \delta) - k)!} \left(\frac{\delta}{1 - \delta}\right)^{2k} \\ &= \frac{(n(1 - \delta))^{2k} \prod_{i=1}^{2k} \left(1 - \frac{k}{(1-\delta)n} + \frac{i}{n(1-\delta)}\right)}{(n\delta)^{2k} \prod_{i=1}^{2k} \left(1 - \frac{k}{\delta n} + \frac{i}{n\delta}\right)} \left(\frac{\delta}{1 - \delta}\right)^{2k} \\ &= \prod_{i=1}^{2k} \frac{1 - \frac{k}{(1-\delta)n} + \frac{i}{n(1-\delta)}}{1 - \frac{k}{\delta n} + \frac{i}{n\delta}} \end{aligned} \tag{99}$$

Letting  $b_i = \frac{i}{n} - \frac{k}{n}$ , we have

$$\log \Gamma = \sum_{i=1}^{2k} \log\left(1 + \frac{1}{1 - \delta} b_i\right) - \sum_{i=1}^{2k} \log\left(1 + \frac{1}{\delta} b_i\right). \tag{100}$$

Recalling that  $\log(1+x) = x - \frac{x^2}{2} + o(x^2)$ , we see that

$$\log \Gamma = \left( \frac{1}{1-\delta} - \frac{1}{\delta} \right) \sum_{i=1}^{2k} b_i - \frac{1}{2} \left( \frac{1}{(1-\delta)^2} - \frac{1}{\delta^2} \right) \sum_{i=1}^{2k} b_i^2 + o\left(\frac{k^3}{n^2}\right). \quad (101)$$

Furthermore, we have

$$\sum_{i=1}^{2k} b_i = \frac{k}{n} + \sum_{j=-(k-1)}^{k-1} \frac{j}{n} = \frac{k}{n} \quad (102a)$$

$$\sum_{i=1}^{2k} b_i^2 = \frac{k^2}{n^2} + 2 \sum_{j=0}^{k-1} \frac{j^2}{n^2} = \frac{k^2}{n^2} + \frac{2}{n^2} \frac{(k-1) \cdot k \cdot (2k-1)}{6} = \frac{2}{3} \frac{k^3}{n^2} + o\left(\frac{k^3}{n^2}\right) \quad (102b)$$

Thus,

$$\begin{aligned} \log \Gamma &= -\frac{1-2\delta}{\delta(1-\delta)} \frac{k}{n} + \frac{1-2\delta}{(\delta(1-\delta))^2} \frac{k^3}{3n^2} + o\left(\frac{k^3}{n^2}\right) \\ &= \frac{(1-2\delta)}{\delta(1-\delta)} \left[ \frac{k^2}{n} \frac{1}{3\delta(1-\delta)} - 1 \right] \frac{k}{n} + o\left(\frac{k^3}{n^2}\right). \end{aligned} \quad (103)$$

Since  $e^x = 1 + x + \mathcal{O}(x^2)$ , it follows that

$$\Gamma = 1 + \frac{(1-2\delta)}{\delta(1-\delta)} \left[ \frac{k^2}{n} \frac{1}{3\delta(1-\delta)} - 1 \right] \frac{k}{n} + o\left(\frac{k^3}{n^2}\right). \quad (104)$$

Now, since the required result is  $\Gamma/(1+\Gamma)$ , it follows easily. ■

#### APPENDIX B AUXILIARY LEMMAS

*Lemma 9 (MGL linearization):* For any  $0 \leq \delta_1, \delta_2 \leq 1/2$  and  $-h_b(\delta_1) < x < \log 2 - h_b(\delta_1)$  we have that

$$h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + x)) \geq h_b(\delta_1 * \delta_2) + \frac{g(\delta_1 * \delta_2)}{g(\delta_1)} x, \quad (105)$$

where  $g(t) = (1-2t) \log\left(\frac{1-t}{t}\right)$ , as defined in (57).

**Proof.** Let  $\varphi_{\delta_2}(t) = h_b(\delta_2 * h_b^{-1}(t))$  and recall that by [19] we have that  $t \mapsto \varphi_{\delta_2}(t)$  is convex, and

$$\varphi'_{\delta_2}(t) = \frac{d}{dt} \varphi_{\delta_2}(t) = (1-2\delta_2) \frac{h'_b(\delta_2 * h_b^{-1}(t))}{h'_b(h_b^{-1}(t))}. \quad (106)$$

Consequently,

$$\varphi'_{\delta_2}(h_b(\delta_1)) = (1-2\delta_2) \frac{h'_b(\delta_1 * \delta_2)}{h'_b(\delta_1)} \quad (107)$$

$$= \frac{(1-2\delta_2)(1-2\delta_1)}{(1-2\delta_1)} \frac{h'_b(\delta_1 * \delta_2)}{h'_b(\delta_1)} \quad (108)$$

$$= \frac{(1-2(\delta_1 * \delta_2))}{(1-2\delta_1)} \frac{h'_b(\delta_1 * \delta_2)}{h'_b(\delta_1)} \quad (109)$$

$$= \frac{g(\delta_1 * \delta_2)}{g(\delta_1)}. \quad (110)$$

Now, by convexity, we have

$$h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + x)) \geq \varphi_{\delta_2}(h_b(\delta_1)) + \varphi'_{\delta_2}(h_b(\delta_1))x, \quad (111)$$

and the statement follows by substituting (110) into (111). ■

*Lemma 10 (Properties of  $g(t)$ ):* The function  $t \mapsto g(t) = (1-2t) \log\left(\frac{1-t}{t}\right)$  is convex in  $[0, 1/2]$  and its derivative is given by

$$g'(t) = \frac{d}{dt} g(t) = -\kappa(t), \quad (112)$$

where

$$\kappa(t) \triangleq 2 \log \left( \frac{1-t}{t} \right) + \frac{1-2t}{t(1-t)}. \quad (113)$$

**Proof.** Calculating  $g'(t)$  is straightforward. Furthermore, all three functions  $t \mapsto 2 \log \left( \frac{1-t}{t} \right)$ ,  $t \mapsto 1-2t$ , and  $t \mapsto \frac{1}{t(1-t)}$  are decreasing in  $[0, 1/2]$ , so  $\kappa(t)$  is decreasing,  $g'(t)$  increasing, and  $g''(t) > 0$ . ■

*Lemma 11 (Properties of  $\beta_q(t)$ ):* Let  $\beta_q(t) \triangleq h_b(q * t) - h_b(t)$ . The function  $t \mapsto \beta_q(t)$  is convex, and its derivative satisfies

$$\beta'_q(t) = \frac{d}{dt} \beta_q(t) = -\phi(q, t), \quad (114)$$

where

$$\phi(q, t) \triangleq 2q \log \left( \frac{1-t}{t} \right) + (1-2q) \log \left( \frac{1+q \frac{1-2t}{t}}{1-q \frac{1-2t}{1-t}} \right). \quad (115)$$

Furthermore, for  $t \in [0, 1/2]$

$$\phi(q, t) \geq q \cdot \kappa(t) \cdot \nu(q, t), \quad (116)$$

where  $\kappa(t)$  is as in (113), and

$$\nu(q, t) \triangleq \frac{1-2q}{1+q \frac{1-2t}{t}}. \quad (117)$$

Moreover,

$$\beta_q(t) \leq qg(t). \quad (118)$$

**Proof.** Convexity follows by noting that  $\beta_q(t) = I(X; X+Z)$ , where  $X \sim \text{Ber}(q)$  and  $Z \sim \text{Ber}(t)$ . Calculation of  $f'_q(t)$  is straightforward. In order to lower bound  $\phi(q, t)$ , we note that

$$\log \left( 1 + q \frac{1-2t}{t} \right) \geq q \frac{\frac{1-2t}{t}}{1 + q \frac{1-2t}{t}} \quad (119)$$

$$-\log \left( 1 - q \frac{1-2t}{1-t} \right) \geq q \frac{1-2t}{1-t}, \quad (120)$$

and therefore

$$(1-2q) \log \left( \frac{1+q \frac{1-2t}{t}}{1-q \frac{1-2t}{1-t}} \right) \geq q \frac{1-2q}{1+q \frac{1-2t}{t}} \left( \frac{1-2t}{t} + \frac{1-2t}{1-t} \right) \quad (121)$$

$$= q \frac{1-2t}{t(1-t)} \frac{1-2q}{1+q \frac{1-2t}{t}}, \quad (122)$$

which gives

$$\begin{aligned} \phi(q, t) &\geq q \left( 2 \log \left( \frac{1-t}{t} \right) + \frac{1-2t}{t(1-t)} \frac{1-2q}{1+q \frac{1-2t}{t}} \right) \\ &\geq q \underbrace{\left( 2 \log \left( \frac{1-t}{t} \right) + \frac{1-2t}{t(1-t)} \right)}_{\kappa(t)} \underbrace{\left( \frac{1-2q}{1+q \frac{1-2t}{t}} \right)}_{\nu(q, t)}. \end{aligned} \quad (123)$$

To prove (118), we apply the concavity of  $t \mapsto h_b(t)$  to obtain

$$\begin{aligned} \beta_q(t) &= h_b(q * t) - h_b(t) \\ &= h_b(t + q(1-2t)) - h_b(t) \\ &\leq q(1-2t)h'_b(t) \\ &= qg(t). \end{aligned}$$

■

APPENDIX C  
PROOF OF LEMMA 6

By Theorem 3, we have that for any  $0 < q < 1/2$

$$\log 2 - h_b(q * D_2) \leq \rho \left[ \log 2 - h_b \left( \delta_2 * h_b^{-1} \left( h_b(\delta_1) + \frac{h_b(q * D_1) - h_b(D_1)}{\rho} \right) \right) \right] + \rho \Gamma(n, \delta_2). \quad (124)$$

Now, applying Lemma 9, we can write

$$h_b \left( \delta_2 * h_b^{-1} \left( h_b(\delta_1) + \frac{h_b(q * D_1) - h_b(D_1)}{\rho} \right) \right) \geq h_b(\delta_1 * \delta_2) + \frac{g(\delta_1 * \delta_2)}{g(\delta_1)} \frac{\beta_q(D_1)}{\rho}, \quad (125)$$

where  $\beta_q(\cdot)$  is as defined in Lemma 11. By combining (124) and (125), we obtain

$$\log 2 - h_b(D_2) - [h_b(q * D_2) - h_b(D_2)] \leq \rho(\log 2 - h(\delta_1 * \delta_2)) - \frac{g(\delta_1 * \delta_2)}{g(\delta_1)} \beta_q(D_1) + \rho \Gamma(n, \delta_2), \quad (126)$$

which, recalling the definition of  $C(t)$  and  $R(t)$ , reduces to

$$\rho C(\delta_1 * \delta_2) - R(D_2) \geq \frac{g(\delta_1 * \delta_2)}{g(\delta_1)} \beta_q(D_1) - \beta_q(D_2) - \rho \Gamma(n, \delta_2). \quad (127)$$

By Lemma 10, the function  $t \mapsto g(t)$  is convex, and consequently

$$g(\delta_1 * \delta_2) = g(\delta_1 + (1 - 2\delta_1)\delta_2) \geq g(\delta_1) + \delta_2 \cdot (1 - 2\delta_1)g'(\delta_1). \quad (128)$$

Thus, we obtain

$$\rho C(\delta_1 * \delta_2) - R(D_2) \geq \beta_q(D_1) - \beta_q(D_2) + \delta_2 \frac{(1 - 2\delta_1)g'(\delta_1)}{g(\delta_1)} \beta_q(D_1) - \rho \Gamma(n, \delta_2) \quad (129)$$

$$= \beta_q(D_1) - \beta_q(D_2) - \psi(\delta_1)\delta_2\beta_q(D_1) - \rho \Gamma(n, \delta_2), \quad (130)$$

where

$$\begin{aligned} \psi(t) &\triangleq -\frac{(1 - 2t)g'(t)}{g(t)} \\ &= (1 - 2t) \frac{\kappa(t)}{g(t)}, \end{aligned} \quad (131)$$

where  $\kappa(\cdot)$  is as defined in Lemma 10. By Lemma 11, the function  $t \mapsto \beta_q(t)$  is convex, and consequently

$$\begin{aligned} \beta_q(D_1) &= \beta_q(D_2 + (D_1 - D_2)) \\ &\geq \beta_q(D_2) + \beta'_q(D_2)(D_1 - D_2) \\ &= \beta_q(D_2) + \phi(q, D_2)(D_2 - D_1), \end{aligned} \quad (132)$$

where  $\phi(\cdot, \cdot)$  is as defined in Lemma 11. We have obtained

$$D_2 - D_1 \leq \frac{[\rho C(\delta_1 * \delta_2) - R(D_2)] + \psi(\delta_1)\delta_2\beta_q(D_1) + \rho \Gamma(n, \delta_2)}{\phi(q, D_2)}. \quad (133)$$

By Lemma 11, we have that  $\beta_q(D_1) \leq qg(D_1)$ , and consequently

$$D_2 - D_1 \leq \frac{\rho C(\delta_1 * \delta_2) - R(D_2) + \rho \Gamma(n, \delta_2)}{\phi(q, D_2)} + \frac{\psi(\delta_1)\delta_2g(D_1)}{\phi(q, D_2)/q} \quad (134)$$

$$\leq \frac{\rho C(\delta_1 * \delta_2) - R(D_2) + \rho \Gamma(n, \delta_2)}{2q \log \left( \frac{1 - D_2}{D_2} \right)} + (\delta_1 * \delta_2 - \delta_1) \frac{\psi(\delta_1)}{1 - 2\delta_1} \frac{g(D_2)}{\kappa(D_2)} \frac{1}{\nu(q, D_2)} \frac{g(D_1)}{g(D_2)}, \quad (135)$$

where in the last inequality we have used the identity  $\delta_2 = \frac{\delta_1 * \delta_2 - \delta_1}{1 - 2\delta_1}$ , the fact that  $\phi(q, D_2) > 2q \log \left( \frac{1 - D_2}{D_2} \right)$  by (115) and that  $\phi(q, D_2) > q \cdot \kappa(t) \cdot \nu(q, t)$  by (117). Let

$$\tau \triangleq \frac{q}{(1 - 2q)D_2}, \quad (136)$$



such that  $\frac{1}{\nu(q, D_2)} = 1 + \tau$ , and  $q = \frac{D_2 \tau}{1 + 2D_2 \tau}$ . Since (135) holds for any  $0 < q < 1/2$ , we have that for any  $\tau > 0$

$$D_2 - D_1 \leq \frac{\rho C(\delta_1 * \delta_2) - R(D_2) + \rho \Gamma(n, \delta_2)}{\frac{2D_2 \tau}{1 + 2D_2 \tau} \log\left(\frac{1 - D_2}{D_2}\right)} + (\delta_1 * \delta_2 - \delta_1) \frac{\psi(\delta_1)}{1 - 2\delta_1} \frac{g(D_2)}{\kappa(D_2)} (1 + \tau) \frac{g(D_1)}{g(D_2)}. \quad (137)$$

Note that

$$\begin{aligned} \frac{\psi(t)}{1 - 2t} &= \frac{\kappa(t)}{g(t)} \\ &= \frac{2}{1 - 2t} + \frac{1}{t(1 - t) \log\left(\frac{1 - t}{t}\right)} \\ &= \log\left(\frac{1 - t}{t}\right) \Phi(t), \end{aligned} \quad (138)$$

where  $\Phi(t)$  is defined in (8a). Thus, (135) can be written as

$$D_2 - D_1 \leq \frac{\rho C(\delta_1 * \delta_2) - R(D_2) + \rho \Gamma(n, \delta_2)}{\frac{2D_2 \tau}{1 + 2D_2 \tau} \log\left(\frac{1 - D_2}{D_2}\right)} + (\delta_1 * \delta_2 - \delta_1) \rho \frac{\log\left(\frac{1 - \delta_1}{\delta_1}\right)}{\log\left(\frac{1 - D_2}{D_2}\right)} \frac{1}{\rho} \frac{\Phi(\delta_1)}{\Phi(D_2)} (1 + \tau) \frac{g(D_1)}{g(D_2)}, \quad (139)$$

which establishes our claim.

#### APPENDIX D PROOF OF LEMMA 7

For any  $0 < \delta < 1/2$  we have that  $f(1, \delta) = 1$ . Thus, it suffices to show that  $\rho \mapsto f(\rho, \delta)$  is monotone decreasing. Let  $\rho = \rho(d)$  be such that  $D(\rho, \delta) = d > 0$ , which is well defined for  $d > 0$ . It is easy to see that

$$\rho(d) = \frac{\log 2 - h_b(d)}{\log 2 - h_b(\delta)} = \frac{R(d)}{R(\delta)}. \quad (140)$$

Recalling that  $d < \delta$  for  $\rho > 1$ , our claim is equivalent to the claim that for any  $0 < d < \delta$  it holds that

$$\frac{R(d)}{R(\delta)} \Phi(d) > \Phi(\delta), \quad (141)$$

which is equivalent to the claim that

$$\vartheta(t) \triangleq \Phi(t) \cdot R(t), \quad (142)$$

is monotone decreasing in  $t$ , which we now establish.

To this end, note that we can write

$$\vartheta(t) = \frac{\gamma(t)}{g^2(t)}, \quad (143)$$

where

$$\gamma(t) \triangleq (1 - 2t)R(t)\kappa(t), \quad (144)$$

and  $\kappa(t) = -g'(t)$  is as defined in (113). We therefore have that

$$\begin{aligned} \vartheta'(t) &= \frac{1}{g^4(t)} [g^2(t)\gamma'(t) - 2g(t)g'(t)\gamma(t)] \\ &= \frac{1}{g^3(t)} \underbrace{\left[ g(t)\gamma'(t) + 2\kappa(t)\gamma(t) \right]}_{\zeta(t)}. \end{aligned} \quad (145)$$

Since  $g(t) > 0$  for all  $0 < t < 1/2$ , we have to show that  $\zeta(t) \leq 0$ . We write

$$\begin{aligned} \zeta'(t) &= \kappa(t)(1 - 2t)R'(t) + \kappa(t)R(t)[(1 - 2t)'] + (1 - 2t)R(t)\kappa'(t) \\ &= -\kappa(t)g(t) - 2\kappa(t)R(t) + (1 - 2t)R(t)\kappa'(t), \end{aligned} \quad (146)$$

where the last equality follows since  $R'(t) = -\log\left(\frac{1-t}{t}\right)$ , and therefore  $(1-2t)R'(t) = -g(t)$ . Furthermore, as  $(1-2t)\kappa(t) = 2g(t) + \frac{(1-2t)^2}{t(1-t)}$ , we also that

$$\kappa(t)\gamma(t) = 2g(t)\kappa(t)R(t) + \frac{(1-2t)^2}{t(1-t)}\kappa(t)R(t). \quad (147)$$

Thus,

$$\begin{aligned} \zeta(t) &= -\kappa(t)g^2(t) - 2g(t)\kappa(t)R(t) + (1-2t)g(t)R(t)\kappa'(t) + 4g(t)\kappa(t)R(t) + 2\frac{(1-2t)^2}{t(1-t)}\kappa(t)R(t) \\ &= -\kappa(t)g^2(t) + 2g(t)\kappa(t)R(t) + (1-2t)g(t)R(t)\kappa'(t) + 2\frac{(1-2t)^2}{t(1-t)}\kappa(t)R(t). \end{aligned} \quad (148)$$

Evaluating  $\kappa'(t)$  gives

$$\begin{aligned} \kappa'(t) &= -\frac{2}{t(1-t)} - \frac{1-2t+2t^2}{(t(1-t))^2} \\ &= -\frac{1}{(t(1-t))^2}. \end{aligned} \quad (149)$$

Substituting (149) into (148), gives

$$\begin{aligned} \zeta(t) &= -\kappa(t)g^2(t) + 2g(t)\kappa(t)R(t) - \frac{R(t)g(t)(1-2t)}{(t(1-t))^2} + 2\frac{(1-2t)^2}{t(1-t)}\kappa(t)R(t) \\ &= -g(t) \left[ \kappa(t)g(t) - 2\kappa(t)R(t) + \frac{R(t)(1-2t)}{(t(1-t))^2} - 2\frac{(1-2t)}{t(1-t)\log\left(\frac{1-t}{t}\right)}\kappa(t)R(t) \right] \\ &= -g(t) \left[ \kappa(t)g(t) - 2\kappa(t)R(t) + \frac{R(t)(1-2t)}{(t(1-t))^2} - \frac{4R(t)(1-2t)}{t(1-t)} - \frac{2R(t)(1-2t)^2}{t^2(1-t)^2\log\left(\frac{1-t}{t}\right)} \right] \\ &= -g(t) \left[ \kappa(t)g(t) - 2\kappa(t)R(t) + \frac{R(t)(1-2t)^3}{(t(1-t))^2} - \frac{2R(t)(1-2t)^2}{t^2(1-t)^2\log\left(\frac{1-t}{t}\right)} \right] \\ &= -g(t) \left[ \underbrace{\kappa(t)g(t) - 2\kappa(t)R(t)}_{A_1(t)} + \underbrace{\frac{R(t)(1-2t)^2}{(t(1-t))^2}}_{A_2(t)} \underbrace{\left[ 1 - 2\left( t + \frac{1}{\log\left(\frac{1-t}{t}\right)} \right) \right]}_{A_3(t)} \right]. \end{aligned}$$

It remains to show that the function  $A(t) = A_1(t) + A_2(t)A_3(t)$  is positive in  $0 < t < 1/2$ . To this end, one can easily verify that  $A(0) = \infty$  and  $A(1/2) = 0$ , and therefore, it suffices to verify that  $t \mapsto A(t)$  is decreasing in  $0 < t < 1/2$ . Indeed, it is straightforward to verify that  $t \mapsto A_1(t)$ ,  $t \mapsto A_2(t)$  and  $t \mapsto A_3(t)$  are decreasing, and the details are omitted.

## REFERENCES

- [1] Y. Kochman, O. Ordentlich, and Y. Polyanskiy, "Ozarow-type outer bounds for memoryless sources and channels," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 1774–1778.
- [2] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] D. Wang, A. Ingber, and Y. Kochman, "A strong converse for joint source-channel coding," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, July 2012, pp. 2117–2121.
- [7] —, "The dispersion of joint source-channel coding," in *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept 2011, pp. 180–187.
- [8] 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.

- [9] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4903–4925, Aug 2011.
- [10] —, "Minimum energy to send  $k$  bits through the Gaussian channel with and without feedback," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4880–4902, Aug 2011.
- [11] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4232–4265, July 2014.
- [12] 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.
- [13] B. Yu and T. Speed, "A rate of convergence result for a universal D-semifair code," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 813–820, May 1993.
- [14] 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.
- [15] 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.
- [16] —, "A source-channel separation theorem with application to the source broadcast problem," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1764–1781, April 2016.
- [17] E. Shafieepoorfard and M. Raginsky, "Sequential empirical coordination under an output entropy constraint," *IEEE Transactions on Information Theory*, vol. 64, no. 10, pp. 6830–6841, Oct 2018.
- [18] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.
- [19] 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.
- [20] 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.
- [21] O. Ordentlich and Y. Polyanskiy, "Entropy under additive Bernoulli and spherical noises," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 521–525.
- [22] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [23] L. Tan, A. Khisti, and E. Soljanin, "Distortion bounds for broadcasting a binary source over binary erasure channels," in *13th Canadian Workshop on Information Theory (CWIT)*, June 2013, pp. 49–54.
- [24] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*. Elsevier, 1977.
- [25] 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.
- [26] V. V. Petrov, *Limit theorems of probability theory: sequences of independent random variables*. Clarendon Press, Oxford University Press, 1995.
- [27] Y. Altug and A. B. Wagner, "Moderate deviations in channel coding," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4417–4426, Aug 2014.
- [28] Y. Polyanskiy and S. Verdú, "Channel dispersion and moderate deviations limits for memoryless channels," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2010, pp. 1334–1339.
- [29] S. Shamai and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2617–2635, Oct 2003.
- [30] M. Taherzadeh and A. K. Khandani, "Single-sample robust joint source-channel coding: Achieving asymptotically optimum scaling of SDR versus SNR," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1565–1577, March 2012.
- [31] G. Caire and K. Narayanan, "On the distortion SNR exponent of hybrid digital-analog space-time coding," *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2867–2878, Aug 2007.
- [32] H. Roozbehani and Y. Polyanskiy, "Triangulation codes: a family of non-linear codes with graceful degradation," in *2018 Conf. Inform. Sciences and Syst.(CISS)*.
- [33] Y. Kochman, A. Mazumdar, and Y. Polyanskiy, "Results on combinatorial joint source-channel coding," in *2012 IEEE Information Theory Workshop*, Sep. 2012, pp. 10–14.
- [34] Y. Polyanskiy, "On metric properties of maps between Hamming spaces and related graph homomorphisms," *Journal of Combinatorial Theory, Series A*, vol. 145, pp. 227–251, 2017.
- [35] H. Roozbehani and Y. Polyanskiy, "Input-output distance properties of good linear codes," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 2510–2514.
- [36] J. Chen, "Rate region of gaussian multiple description coding with individual and central distortion constraints," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 3991–4005, Sep. 2009.