

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 the 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 [3] are met). The fundamental issue in the MAC setup [4] 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 $\mathcal{E} : \mathcal{S}^m \rightarrow \mathcal{X}^n$. 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 $\mathcal{D} : \mathcal{Y}^n \rightarrow \hat{\mathcal{S}}^m$. Let $d : \mathcal{S} \times \hat{\mathcal{S}} \rightarrow \mathbb{R}$ 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

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

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

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

By the separation principle [5], we have that

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

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

Y. Kochman and O. Ordentlich are with the Rachel and Selim Benin School of Computer Science and Engineering, Hebrew University of Jerusalem, Israel (emails: yuvalko@cs.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, by the HUJI Cyber Security Research Center in conjunction with the Israel National Cyber Bureau in the Prime Minister’s Office and by the Ministry of Science & Technology, Israel, in conjunction with Ministry of Europe and Foreign Affairs and the Ministry of Higher Education, Research and Innovation of France. 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 and 2019 International Symposium on Information Theory [1], [2].

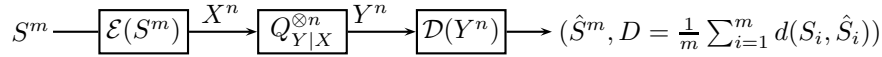


Fig. 1. The $(n, \rho, P_S, Q_{Y|X})$ joint source channel coding problem. It is assumed that $n = \rho m$.

to its asymptotic value. To that end, we define and study the mean distortion redundancy

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

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 [6]. To the contrary, for any $D < D_\infty^*$ it holds that $1 - \epsilon(D)$ is exponentially small [7]. The dispersion, i.e., convergence of D to D_∞^* for fixed excess-distortion probability, as well as finite-blocklength bounds, were derived in [8], [9].

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 [10], [11], [12], because these (mean-zero) Gaussian variations can be averaged out. If indeed the variations canceled out in a similar way for average distortion, it would suggest that the true behavior of the fundamental limit Δ_n^* should indeed be $o(1/\sqrt{n})$.

As further evidence that Δ_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 [13], [14] 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 [3] 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 there exist cases where $\Delta_n^* = \Omega\left(\frac{1}{\sqrt{n}}\right)$. 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 [15] for the quadratic-Gaussian case, and was recently generalized by Khezeli and Chen [16], [17]. Here, we generalize these bounds for our scenario of interest, and show that for the binary Hamming case, with bandwidth expansion ratio $\rho > 1$, 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_∞^* changes too slowly with the crossover probability of the BSC.

We note that it is still not clear whether the $n^{-1/2}$ scaling of Δ_n^* is achievable in general. In particular, a separation-based coding scheme can only achieve

$$\begin{aligned} \Delta_n &= D_n - D_\infty^* \\ &= \Omega\left(\sqrt{\frac{\log n}{n}}\right), \end{aligned}$$

and to the best of our knowledge no JSCC coding scheme that achieves better scaling in general is known. See Section VII. Thus, despite the progress made in this work, the exact correct scaling of Δ_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, i.e., noise uniform on an Hamming sphere, broadcast channel. The full proof of our main result is given in Section V. Some auxiliary results for the problem of sending a source over a broadcast channel are given in Section VI. The paper concludes with a discussion in Section VII.

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 BSC(δ) 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

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

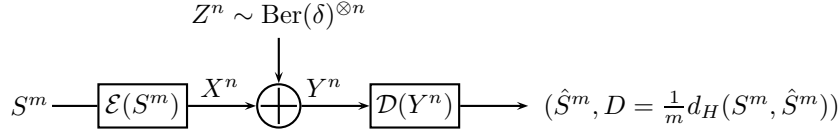


Fig. 2. The $n(\rho, \delta)$ binary joint source-channel coding problem. it is assumed that $n = \rho m$.

We use the binary entropy function¹

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

and its inverse restricted to the interval $[0, 1/2]$ as $h_b^{-1}(\cdot)$. It will be convenient to extend the domain of the function $h_b^{-1}(\cdot)$ to $(-\infty, \log 2]$, such that $h_b^{-1}(t) = 0$ for all $t \leq 0$. For two binary variables, or vectors, the notation \oplus is to be understood as addition modulo-2. 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.

Definition 1: Let $S^m \sim \text{Ber}(1/2)^{\otimes m}$, $Y^n = X^n \oplus Z^n$ 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

$$D^*(n, \rho, \delta) \triangleq \frac{1}{m} \min \mathbb{E} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) \oplus Z^n)), \quad (5)$$

where the minimum is over all $\mathcal{E} : \{0, 1\}^m \rightarrow \{0, 1\}^n$ and $\mathcal{D} : \{0, 1\}^n \rightarrow \{0, 1\}^m$, and its mean distortion redundancy is

$$\begin{aligned} \Delta_n^* &= \Delta^*(n, \rho, \delta) \\ &\triangleq D^*(n, \rho, \delta) - D(\rho, \delta), \end{aligned} \quad (6)$$

where

$$\begin{aligned} D(\rho, \delta) &\triangleq D_\infty^*(\rho, \delta) \\ &= h_b^{-1}(\log 2 - \rho(\log 2 - h_b(\delta))), \end{aligned} \quad (7)$$

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 $\rho > 1$.

Our main result is the following.

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

$$\Delta_n^* > \sqrt{\frac{\delta(1-\delta)}{2\pi n}} \eta(\rho, \delta) + \mathcal{O}(n^{-3/4} \log n), \quad (8)$$

¹Throughout, logarithms are taken to the natural base.

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

$$\Phi(\delta) \triangleq \frac{2}{(1-2\delta) \log(\frac{1-\delta}{\delta})} + \frac{1}{\delta(1-\delta) \log^2(\frac{1-\delta}{\delta})} \quad (9a)$$

$$f(\rho, \delta) \triangleq \frac{1}{\rho} \frac{\Phi(\delta)}{\Phi(D(\rho, \delta))}, \quad (9b)$$

$$\begin{aligned} \eta(\rho, \delta) &\triangleq 2\rho \frac{\log(\frac{1-\delta}{\delta})}{\log(\frac{1-D(\rho, \delta)}{D(\rho, \delta)})} \cdot \frac{1+f(\rho, \delta)}{f(\rho, \delta)} \\ &\frac{D(\rho, \delta) (1-f(\rho, \delta))^2}{2f(\rho, \delta) + 4D(\rho, \delta) (1-f(\rho, \delta))}. \end{aligned} \quad (9c)$$

In particular, this binary symmetric example serves to establish the following corollary.

Corollary 1: In the general point-to-point JSCC problem there exist choices of parameters $(\rho, P_S, Q_{Y|X})$, such that the mean distortion redundancy (4) satisfies:

$$\Delta^*(n, \rho, P_S, Q_{Y|X}) = \Omega\left(\frac{1}{\sqrt{n}}\right). \quad (10)$$

At this point, the reader may wonder why Theorem 1 is restricted to the case $\rho > 1$. The proof of Theorem 1, given in Section V, shows that the lower bound (8) holds as long as $f(\rho, \delta) < 1$, where the condition $f(\rho, \delta) < 1$ is needed in order to justify (66) in the derivation below. In Lemma 7 we show that the condition $f(\rho, \delta) < 1$ holds whenever $\rho > 1$ and $D(\rho, \delta) > 0$. Thus, Theorem 1 holds whenever $\rho > 1$.² In the regime $\rho \leq 1$, we have that $f(\rho, \delta) \geq 1$, and our bound is no longer valid.

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, where the full proof is given. 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 δ such that δn is an integer; this limitation 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

²There is no need to require $D(\rho, \delta) > 0$ in addition to $\rho > 1$. This follows since $D(\rho, \delta) = 0$ implies that $\eta(\rho, \delta) = 0$, in which case, the lower bound (8) holds trivially.

of all length- n binary sequences with Hamming weight x . For a given encoder/decoder pair $(\mathcal{E}, \mathcal{D})$ we define

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

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

$$\begin{aligned} &\mathbb{E} \left[\frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) \oplus Z^n)) \right] \\ &= \mathbb{E}[\Psi(K)] \\ &= \Pr(K \in \mathcal{K}_1) \mathbb{E}[\Psi(K) \mid |K| \in \mathcal{K}_1] + \\ &\quad \Pr(K \in \mathcal{K}_2) \mathbb{E}[\Psi(K) \mid |K| \in \mathcal{K}_2] \end{aligned} \quad (12)$$

where

$$\mathcal{K}_1 = \{K : |K| \leq \sqrt{n} \log^2(n)\} \quad (13a)$$

$$\mathcal{K}_2 = \{K : |K| > \sqrt{n} \log^2(n)\}. \quad (13b)$$

Standard concentration inequalities show that

$$\Pr(K \in \mathcal{K}_2) = \mathcal{O}\left(\frac{1}{n}\right), \quad (14)$$

and we can therefore neglect the second term in (12). Furthermore, since K is a shifted Binomial random variable with zero mean, and the interval \mathcal{K}_1 is sufficiently small, we have that the distribution of K is approximately symmetric within \mathcal{K}_1 . Thus, for any positive $k \in \mathcal{K}_1$ it holds that

$$\Pr(K = k \mid |K| = k) \approx \Pr(K = -k \mid |K| = k),$$

and, therefore, within this interval we have

$$\begin{aligned} \mathbb{E} \left[\Psi(K) \mid |K| = k \right] &= \Pr(K = k \mid |K| = k) \Psi(k) + \\ &\quad \Pr(K = -k \mid |K| = k) \Psi(-k) \\ &\approx \frac{1}{2} \Psi(k) + \frac{1}{2} \Psi(-k). \end{aligned} \quad (15)$$

Thus, combining (12), (14), and (15), we obtain

$$\begin{aligned} \mathbb{E}[\Psi(K)] &\approx \mathbb{E}[\Psi(K) \mid |K| \in \mathcal{K}_1] \\ &\approx \frac{1}{2} \mathbb{E} \left[\mathbb{E} \left[\frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) \oplus U_1^n)) \right] + \right. \\ &\quad \left. \mathbb{E} \left[\frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) \oplus U_2^n)) \right] \right. \\ &\quad \left. \mid |K| \in \mathcal{K}_1 \right], \end{aligned} \quad (16)$$

where

$$U_1^n \sim \text{Uniform}\left(\mathcal{S}_{n(\delta - \frac{|K|}{n}), n}\right), \quad (17a)$$

$$U_2^n \sim \text{Uniform}\left(\mathcal{S}_{n(\delta + \frac{|K|}{n}), n}\right), \quad (17b)$$

and the outer expectation in the right hand side is taken over $|K|$.

In the point-to-point JSCC setup considered, the encoder \mathcal{E} and the decoder \mathcal{D} are not aware of $|K|$, and cannot adapt to it. In particular, the same decoder is used whether K is negative or positive. Nevertheless, we continue the analysis by making the following relaxations: we let the encoder and decoder vary as a function of $|K|$, and further for any value of $|K|$ we allow different decoders for the channel with additive noise U_1^n (negative K) and the channel with additive noise U_2^n (positive K). Thus we have that

$$D^*(n, \rho, \delta) \gtrsim \frac{1}{2} \mathbb{E}[\underline{D}(|K|) \mid |K| \in \mathcal{K}_1], \quad (18)$$

where

$$\underline{D}(k) \triangleq \frac{1}{m} \min \left(\mathbb{E}[d_H(S^m, \mathcal{D}_1(\mathcal{E}(S^m) \oplus U_1^n))] + \mathbb{E}[d_H(S^m, \mathcal{D}_2(\mathcal{E}(S^m) \oplus U_2^n))] \right), \quad (19)$$

U_1^n and U_2^n are as in (17a) and (17b), respectively, and the minimization is over all encoders $\mathcal{E} : \{0, 1\}^m \rightarrow \{0, 1\}^n$ and decoders $\mathcal{D}_1 : \{0, 1\}^n \rightarrow \{0, 1\}^m$, $\mathcal{D}_2 : \{0, 1\}^n \rightarrow \{0, 1\}^m$. This last quantity is the sum-distortion in a problem of JSCC broadcast with spherical noise, see Figure 3. Thus, in order to establish a lower bound on $\underline{D}(k)$, we will next study the JSCC broadcast problem. Specifically, in Theorem 4 (stated in Section II-B and proved in Section IV-B) we show that for positive $k \in \mathcal{K}_1$

$$\underline{D}(k) \gtrsim 2D(\rho, \delta) + \frac{k/\sqrt{n}}{\sqrt{n}} \eta(\rho, \delta). \quad (20)$$

Thus,

$$D^*(n, \rho, \delta) \gtrsim D(\rho, \delta) + \frac{\mathbb{E} \left[\frac{|K|}{\sqrt{n}} \mid |K| \in \mathcal{K}_1 \right]}{2\sqrt{n}} \eta(\rho, \delta).$$

The result of the theorem follows since $|K|/\sqrt{n}$ is approximately the absolute value of a normal variable with zero mean and variance $\delta(1 - \delta)$, thus

$$\mathbb{E} \left[\frac{|K|}{\sqrt{n}} \mid |K| \in \mathcal{K}_1 \right] \approx \sqrt{\frac{2\delta(1 - \delta)}{\pi}},$$

which gives the stated result.

B. Sending a Source Over a Broadcast Channel

We now outline the derivation of (20), 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, as follows, See Figure 4.

Definition 2: In the JSCC broadcast problem, an encoder observes a sequence $S^m = (S_1, \dots, S_m)$ of m i.i.d. samples generated according to the distribution P_S , and would like to send this sequence through the broadcast channel $Q_{Y_1^n, Y_2^n | X^n}$, which may not be memoryless, nor degraded. To that end, the encoder maps the source sequence S^m to the channel input X^n using an encoding function $\mathcal{E} : S^m \rightarrow \mathcal{X}^n$. 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 $\mathcal{D}_1 : \mathcal{Y}_1^n \rightarrow \hat{\mathcal{S}}^m$, whereas the second receiver, which observes the

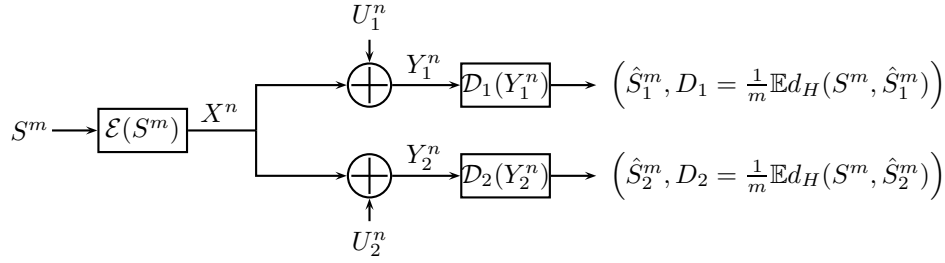


Fig. 3. Binary spherical noise JSCC broadcast problem.

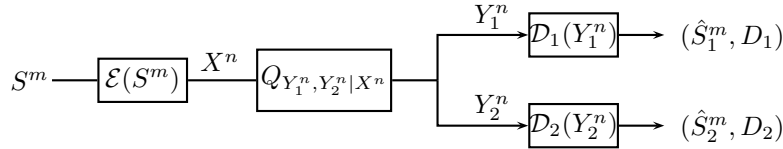


Fig. 4. JSCC over a Broadcast Channel.

channel output Y_2^n , generates an estimate \hat{S}_2^m of the source sequence, using a decoding function $\mathcal{D}_2 : \mathcal{Y}_2^n \rightarrow \hat{\mathcal{S}}^m$. 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 $d : \mathcal{S} \times \hat{\mathcal{S}} \rightarrow \mathbb{R}$.³

$$D_i = \frac{1}{m} \mathbb{E} d(S^m, \hat{S}_i^m), \quad i = 1, 2, \quad (21)$$

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

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

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 $P = P_{SU}$. With respect to this distribution, we define:

$$F_P(t) \triangleq \min_{V : \substack{U-S-V \\ I(S;V) \geq t}} I(S; V|U), \quad (22)$$

$$\bar{R}_P(D) \triangleq \min_{\substack{\hat{S} : U-S-\hat{S} \\ \mathbb{E} d(S, \hat{S}) \leq D}} I(U; \hat{S}). \quad (23)$$

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 .⁴ In Section VI we evaluate these functions for the source $S \sim \text{Ber}(p)$ under Hamming distortion, augmented with the

auxiliary random variable $U = S \oplus N$, where $N \sim \text{Ber}(q)$ is statistically independent of S . We also evaluate these functions for $S \sim \mathcal{N}(0, \sigma^2)$ under quadratic loss, augmented with the auxiliary random variable $U = S + N$, where $N \sim \mathcal{N}(0, \delta^2)$ is statistically independent of S .

Furthermore, we define the following function of the channel $Q^n = Q_{Y_1^n, Y_2^n | X^n}(y_1^n, y_2^n | x^n)$,

$$G_{Q^n}(t) \triangleq \max_{\substack{W, X^n : W-X^n-(Y_1^n, Y_2^n) \\ I(X^n; Y_1^n | W) \geq t}} I(Y_2^n; W). \quad (24)$$

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 $nR_2 \leq G_{Q^n}(nR_1)$ [19, 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

$$\bar{R}_P(D_2) \leq \frac{1}{m} \cdot G_{Q^n}(mF_P(R_{P_S}(D_1))), \quad (25)$$

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 [15] for the Gaussian joint source-channel coding broadcast problem, will be given in Section III. 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. However, if the broadcast channel is memoryless and degraded then G_{Q^n} single-letterizes; as a consequence the whole bound becomes single-letter as well, and in fact we will also show

³These assumptions are made for the sake of simplicity only, results easily go through to the more general case as well.

⁴We 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 [18].

that for this case it can be obtained as a special case of [16, Theorem 5].

We now specialize Theorem 2 to the *spherical-noise binary JSCC broadcast problem*. Recalling Figure 3, it is defined as follows.

Definition 3: The $(\rho, \delta_1, \delta_2)$ spherical-noise binary JSCC broadcast problem is a JSCC broadcast problem (as in Definition 2) with a memoryless symmetric binary source S^m , channels $Y_1^n = X^n \oplus U_1^n$, $Y_2^n = X^n \oplus U_2^n$, where $U_1^n \sim \text{Uniform}(\mathcal{S}_{n\delta_1, n})$ and $U_2^n \sim \text{Uniform}(\mathcal{S}_{n(\delta_1 * \delta_2), n})$, $(U_1^n, U_2^n) \perp\!\!\!\perp X^n$, and Hamming distortion.

In this problem the broadcast channel is neither memoryless, nor degraded. 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 channel function does single-letterize, and equals

$$\begin{aligned} \frac{1}{n}G_{Q^n}(nt) &= G_{\text{BSC}}(t) \\ &= \log 2 - h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + t)), \end{aligned} \quad (26)$$

see Section VI. The i.i.d. JSCC bound derived from this function would easily prove our result. Closing the gap between the i.i.d. case and the spherical-noise case, is one of the main technical challenges in this work.

For the spherical-noise broadcast channel, we prove that

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

where

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

The proof of this bound, which is given in Section IV, is based on replacing U_2^n with the noise $U_1^n \oplus \tilde{Z}^n$, 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 [20] (see Section IV). The difference $|G_{\tilde{Q}^n}(nt) - G_{Q^n}(nt)|$ is essentially bounded by $\sup_{P_{X^n}} |H(X^n \oplus U_1^n \oplus \tilde{Z}^n) - H(X^n \oplus U_2^n)|$, which can be bounded via a coupling argument introduced in [21] (see also [22]). Using this technique, we show that $|G_{\tilde{Q}^n}(nt) - G_{Q^n}(nt)| \lesssim n \cdot \Gamma(n, \delta_2)$.

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 $U = S \oplus N$, 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 $(\rho, \delta_1, \delta_2)$ spherical-noise binary JSCC broadcast problem. If (D_1, D_2) is achievable, then for

any $0 < q < 1/2$

$$\begin{aligned} \frac{\log 2 - h_b(q * D_2)}{\rho} &\leq \\ &\log 2 \\ &- h_b\left(\delta_2 * h_b^{-1}\left(h_b(\delta_1) + \frac{h_b(q * D_1) - h_b(D_1)}{\rho}\right)\right) \\ &+ \Gamma(n, \delta_2), \end{aligned} \quad (27)$$

where $\Gamma(n, \delta_2)$ is as defined in (40).

Recall now that our goal is to bound from $\underline{D}(k)$ (19). This is nothing but the sum $D_1 + D_2$ in the binary spherical noise JSCC broadcast problem, with appropriate δ_1 and δ_2 . The required result is as follows.

Theorem 4: Consider the $(\rho, \delta_1, \delta_2)$ spherical-noise binary JSCC broadcast problem, with $\rho > 1$, $\delta_1 = \delta - \frac{a}{\sqrt{n}}$, and $\delta_1 * \delta_2 = \delta + \frac{a}{\sqrt{n}}$, for some positive $a < \log^2(n)$. If (D_1, D_2) is achievable, then

$$D_1 + D_2 \geq 2D(\rho, \delta) + \frac{a\eta(\rho, \delta)}{\sqrt{n}} + \mathcal{O}\left(n^{-3/4} \log n\right), \quad (28)$$

where $\eta(\rho, \delta)$ is as defined in (9c).

The proof appears in Section IV-B. Given Theorem 3, it is merely a matter of algebraic manipulations and local approximations. Note that since we are interested in the regime where $\delta_2 = \Theta(n^{-1/2})$, the term $\Gamma(n, \delta_2)$ only contributes to the $\mathcal{O}(n^{-3/4} \log n)$ correction.

III. OUTER BOUND FOR THE GENERAL JSCC BROADCAST PROBLEM

In this section we address the general JSCC broadcast problem of Definition 2. We prove Theorem 2 by analyzing the source and channel functions (22)-(24). 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 two simple statements regarding properties of the functions F_P (22) and \bar{R}_P (23). Recall that for a given choice of auxiliary conditional distribution $P_{U|S}$, we have defined $P = P_S P_{U|S}$. The source function $F_{P^{\otimes m}}$ corresponding to m i.i.d. draws (U^m, S^m) from P is

$$F_{P^{\otimes m}}(t) \triangleq \min_{V: \substack{U^m - S^m - V \\ I(S^m; V) \geq t}} I(S^m; V|U^m).$$

Lemma 1: The function $F_P(t)$ is monotone non-decreasing and convex. Furthermore, it tensorizes, i.e.,

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

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 $A \sim \text{Ber}(\alpha)$ be statistically independent of (U, S) , 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 the definition of $F_P(t)$,

$$\begin{aligned} F_P((1 - \alpha)t_0 + \alpha t_1) &\leq I(S; \tilde{V} | U) \\ &= (1 - \alpha)I(S; V_0 | U) + \alpha I(S; V_1 | U) \\ &= (1 - \alpha)F_P(t_0) + \alpha F_P(t_1). \end{aligned} \quad (29)$$

We prove tensorization by induction. Let $(U^m, S^m) \sim P^{\otimes m}$. 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

$$\begin{aligned} F_P\left(\frac{I(S^m; V)}{m}\right) &\leq \frac{1}{m} F_{P^{\otimes(m-1)}}(I(S^{m-1}; V)) + \frac{1}{m} F_P(I(S_m; V, S^{m-1})) \\ &\leq \frac{1}{m} [I(S^{m-1}; V | U^{m-1}) + I(S_m; V, S^{m-1} | U_m)], \end{aligned} \quad (30)$$

where the last inequality follows by definition of $F_{P^{\otimes(m-1)}}$ 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

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

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

$$\begin{aligned} I(S^{m-1}; V | U^{m-1}) + I(S_m; V, S^{m-1} | U_m) \\ \leq I(S^m; V | U^m). \end{aligned} \quad (31)$$

Substituting (31) into (30), gives

$$I(S^m; V | U^m) \geq m F_P\left(\frac{I(S^m; V)}{m}\right). \quad (32)$$

Thus, we have shown that $F_{P^{\otimes m}}(mt) \geq m F_P(t)$. On the other hand, we have that

$$\begin{aligned} F_{P^{\otimes m}}(mt) &\triangleq \min_{V: U^m - S^m - V} \min_{I(S^m; V) \geq mt} I(S^m; V | U^m) \\ &\leq \min_{V^m: U^m - S^m - V^m} \min_{I(S^m; V^m) \geq mt} I(S^m; V^m | U^m) \\ &= m F_P(t), \end{aligned} \quad (33)$$

where the minimum in (33) is taken over the memoryless channels $P_{V^m | S^m} = \prod_{i=1}^m P_{V_i | S_i}$. ■

Lemma 2: Let $(U^m, S^m) \sim P^{\otimes m}$, and let \hat{S}^m be a random vector satisfying the Markov chain $U^m - S^m - \hat{S}^m$ and $\frac{1}{m} \mathbb{E}d(S^m, \hat{S}^m) \leq D$, then

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

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

$$I(U^m; \hat{S}^m) \geq \sum_{i=1}^m I(U_i; \hat{S}_i). \quad (34)$$

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

$$I(U^m; \hat{S}^m) \geq m \frac{1}{m} \sum_{i=1}^m \bar{R}_P(d_i) \geq m \bar{R}_P(D). \quad (35)$$

■

B. The Channel Function and Derivation of Theorem 2

In the general (non-memoryless, non-degraded) case the channel function G_{Q^n} (24) does not tensorize. However, we have the following basic properties.

Lemma 3: 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^n) and (W_1, X_1^n) be random variables that attain $G_{Q^n}(t_0)$ and $G_{Q^n}(t_1)$ with equality. Let $A \sim \text{Ber}(\alpha)$ be statistically independent of (W_0, X_0^n, W_1, X_1^n) , and define

$$(\tilde{W}, \tilde{X}^n) = ((A, W_A), X_A^n).$$

We have that

$$\begin{aligned} I(\tilde{X}^n; Y_1^n | \tilde{W}) &= I(X_A^n; Y_1^n | W_A, A) \\ &= (1 - \alpha)I(X_0^n; Y_1^n | W_0) \\ &\quad + \alpha I(X_1^n; Y_1^n | W_1) \\ &= (1 - \alpha)t_0 + \alpha t_1, \end{aligned} \quad (36)$$

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

$$\begin{aligned} G_{Q^n}(\alpha t_0 + (1 - \alpha)t_1) &= G_{Q^n}(I(\tilde{X}^n; Y_1^n | \tilde{W})) \\ &\geq I(Y_2^n; \tilde{W}) \\ &= I(Y_2^n; A) + I(Y_2^n; W_A | A) \\ &\geq \alpha G_{Q^n}(t_0) + (1 - \alpha)G_{Q^n}(t_1). \end{aligned} \quad (37)$$

■

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 [15]. Their proof relied heavily on the entropy-power inequality (EPI), 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

$$m\bar{R}_P(D_2) \leq I(U^m, \hat{S}_2^m) \quad (38a)$$

$$\leq I(U^m; Y_2^n) \quad (38b)$$

$$\leq G_{Q^n}(I(X^n; Y_1^n | U^m)) \quad (38c)$$

$$\leq G_{Q^n}(I(S^m; Y_1^n | U^m)) \quad (38d)$$

$$\leq G_{Q^n}(F_{P^{\otimes m}}(I(S^m; Y_1^n))) \quad (38e)$$

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

$$\leq G_{Q^n}\left(mF_P\left(\frac{I(S^m; \hat{S}_1^m)}{m}\right)\right) \quad (38g)$$

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

where (38a) follows from Lemma 2, (38b) follows from the data processing inequality (DPI), (38c) from definition of G_{Q^n} , (38d) from the DPI and monotonicity of G_{Q^n} , (38e) from definition of $F_{P^{\otimes m}}$, (38f) from tensorization of $F_{P^{\otimes m}}$, and (38g) from the DPI. ■

Note that U^m plays a two-fold role here: in (38c) we used the Markov chain $U^m - X^n - (Y_1^n, Y_2^n)$, whereas in (38a) and (38e) we used $U^m - S^m - Y_1^n$. Thus, the source functions $F_P(t)$ and $\bar{R}_P(D)$, 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 and memoryless once we choose the channel $P_{U|S}$, appears in both Markov chains, in the transition from (38b) to (38c), we have used the definition of G_{Q^n} , which involves a *maximization* with respect to U^m . As will be shown in the sequel, in the special case where Q^n is degraded and memoryless, the auxiliary random variables achieving the maximum in the definition of $G_{Q^n}(t)$ are of the form $(W^n, X^n) \sim P_{WX}^{\otimes n}$, i.e., n -letter memoryless distribution. We therefore see that for such Q^n , the random variables (U^m, X^n) , where U^m is m -letter memoryless, cannot achieve the maximum in the definition of $G_{Q^n}(t)$, unless $m = n$. Thus, the inequality (38c) must be strict in this case.

IV. OUTER BOUND FOR THE BINARY SPHERICAL-NOISE JSCC BROADCAST PROBLEM

In this section we derive an explicit bound for the binary spherical-noise JSCC broadcast problem (Definition 3), namely we prove Theorem 4. To that end, we first establish Theorem 3. As we make extensive use of Mrs. Gerber's

Lemma (MGL) [20], for completeness, before proceeding, we recall it here.

Lemma 4 (Mrs. Gerber's Lemma [20]): The MGL function $t \mapsto h_b(\delta * h_b^{-1}(t))$ is convex. Furthermore, for $Z^n \sim \text{Ber}(\delta)^{\otimes n}$ and any n -dimensional binary vector X^n , statistically independent of Z^n , we have that

$$H(X^n \oplus Z^n) \geq nh_b\left(\delta * h_b^{-1}\left(\frac{H(X^n)}{n}\right)\right).$$

A. Proof of Theorem 3

For specializing Theorem 2 to the binary case, we need to evaluate or at least bound the source and channel functions. The crucial part is a bound on the (non single-letter) channel function, as follows.

Lemma 5: 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 $Y_1^n = X^n \oplus U_1^n$, $Y_2^n = X^n \oplus U_2^n$, 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,

$$\begin{aligned} \frac{1}{n}G_{Q^n}(nt) &\leq \log 2 - h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + t)) \\ &\quad + \Gamma(n, \delta_2), \end{aligned} \quad (39)$$

where

$$\Gamma(n, \delta_2) \triangleq \sqrt{\frac{\delta_2}{n}} \log\left(\frac{n}{\delta_2}\right) + \frac{\log n + 1}{2n}. \quad (40)$$

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

$$H(Y_2^n | W) = H(X^n \oplus U_1^n \oplus Z_3^n | W) + \Delta H, \quad (41)$$

where

$$\Delta H \triangleq H(X^n \oplus U_2^n | W) - H(X^n \oplus U_1^n \oplus Z_3^n | W), \quad (42)$$

where $Z_3^n \sim \text{Ber}(\delta_2)^{\otimes n}$. We will upper-bound $|\Delta H|$ via coupling. That is, we define a joint distribution on $(U_2^n, U_1^n \oplus Z_3^n)$ which has the same marginals as the one in our problem, namely:

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

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

$$\begin{aligned} \mathbb{E}w_H(U_2^n \oplus (U_1^n \oplus Z_3^n)) &= \mathbb{E}|T - n(\delta_1 * \delta_2)| \\ &= \mathbb{E}\sqrt{(T - n(\delta_1 * \delta_2))^2} \\ &\leq \sqrt{\text{Var}(T)} \\ &= \sqrt{\text{Var}(T_0) + \text{Var}(T_1)} \\ &= \sqrt{n\delta_2(1 - \delta_2)} \\ &\leq \sqrt{n\delta_2}, \end{aligned} \quad (43)$$

where the first inequality follows from Jensen's inequality and the fact that $\mathbb{E}(T) = n(\delta_1 * \delta_2)$. Now, applying [21, Proposition 8] (see also [22]), we obtain for $\delta_2 < 1/2$

$$|\Delta H| \leq \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right). \quad (44)$$

Thus, we can use Mrs. Gerber's Lemma (MGL) to lower-bound (41) as

$$\begin{aligned} H(Y_2^n|W) &\geq H(X^n \oplus U_1^n \oplus Z_3^n|W) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right) \\ &\geq n\varphi_{\delta_2}\left(\frac{H(X^n \oplus U_1^n|W)}{n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right), \end{aligned} \quad (45)$$

where $\varphi_{\delta_2}(x) = h_b(\delta_2 * h_b^{-1}(x))$ is the MGL function. We further bound $H(Y_2^n|W)$ in (46) at the bottom of this page, where in (46a) we substitute [23, Chapter 10, Lemma 7] to lower-bound $H(U_1^n)$, and in (46b) we have defined the MGL derivative $\varphi'_{\delta_2}(x) = \frac{d}{dx}\varphi_{\delta_2}(x)$, and used the convexity of $x \mapsto \varphi_{\delta_2}(x)$ [20]. Recalling that $\varphi'_{\delta_2}(x) \leq 1$ due to [24, Theorem 2.6],⁵ we have obtained

$$\begin{aligned} H(Y_2^n|W) &\geq nh_b\left(\delta_2 * h_b^{-1}\left(h_b(\delta_1) + \frac{I(X^n; Y_1^n|W)}{n}\right)\right) \\ &\quad - n\Gamma(n, \delta_2), \end{aligned} \quad (47)$$

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

The treatment of the source functions is much simpler. As it is identical to the i.i.d. binary problem, we defer the derivation to Section VI-B, where we show that for $S \sim \text{Ber}(p)$ and the choice $U = S + N$, where $N \sim \text{Ber}(q)$ is statistically independent of S , we have:

$$F_P(t) \geq t - h_b(q * p) + h_b(q * h_b^{-1}(h_b(p) - t)) \quad (48)$$

$$\bar{R}_P(D) = h_b(q * p) - h_b(q * D), \quad (49)$$

for $0 \leq D \leq p$. Substituting these expressions for $p = 1/2$ and the channel-function bound of Lemma 5 in Theorem 2, Theorem 3 is immediately obtained.

⁵Alternatively, this can be seen directly from the expression for $\varphi'_{\delta_2}(x)$ derived in [20].

B. Local Analysis: From Theorem 3 to Theorem 4

Our next goal is to manipulate the bound from Theorem 3 in order to obtain a lower bound on the sum-distortion. The proof of the following Lemma is brought in Appendix D, and is based on several auxiliary lemmas, which are stated and proved in Appendix C.

Lemma 6: Consider the $(\rho, \delta_1, \delta_2)$ spherical-noise binary JSCC broadcast problem of Definition 3. If (D_1, D_2) is achievable, then for any $\tau > 0$ we have

$$\begin{aligned} D_2 - D_1 &\leq \frac{1 + 2D_2\tau}{2D_2\tau} \frac{\rho C(\delta_1 * \delta_2) - R(D_2) + \rho\Gamma(n, \delta_2)}{\log\left(\frac{1-D_2}{D_2}\right)} + \\ &\quad (\delta_1 * \delta_2 - \delta_1) \frac{\log\left(\frac{1-\delta_1}{\delta_1}\right)}{\log\left(\frac{1-D_2}{D_2}\right)} \frac{\Phi(\delta_1)}{\Phi(D_2)} (1 + \tau) \frac{g(D_1)}{g(D_2)}, \end{aligned} \quad (50)$$

where $\Phi(t)$ is defined in (9a), $\Gamma(n, \delta_2)$ is defined in (40), and

$$C(t) = R(t) \triangleq \log 2 - h_b(t), \quad (51)$$

$$g(t) \triangleq (1 - 2t) \log\left(\frac{1-t}{t}\right) = (1 - 2t)h'_b(t). \quad (52)$$

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

Lemma 7: Let $\Phi(t)$, $D(\rho, \delta)$, and $f(\rho, \delta)$ be as defined in (9a), (7), and (9b), respectively. For every $\rho > 1$ and $0 < \delta < 1/2$ for which $D(\rho, \delta) > 0$, it holds that

$$f(\rho, \delta) < 1.$$

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

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

$$\Psi(k) \geq D\left(\rho, \delta + \frac{k}{n}\right) + \mathcal{O}\left(\frac{\log n}{n}\right), \quad (53)$$

where $\Psi(\cdot)$ was defined in (11).

Using Lemma 6, Lemma 7, and Proposition 1, we can now prove Theorem 4, which is the main result of this subsection.

$$\begin{aligned} H(Y_2^n|W) &\geq n\varphi_{\delta_2}\left(\frac{H(X^n \oplus U_1^n|X^n) + H(X^n \oplus U_1^n|W) - H(X^n \oplus U_1^n|X^n, W)}{n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right) \\ &= n\varphi_{\delta_2}\left(\frac{H(U_1^n) + I(X^n; Y_1^n|W)}{n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right) \\ &\geq n\varphi_{\delta_2}\left(\frac{nh_b(\delta_1) - \frac{1}{2}(\log n + 1) + I(X^n; Y_1^n|W)}{n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right) \end{aligned} \quad (46a)$$

$$\begin{aligned} &= n\varphi_{\delta_2}\left(h_b(\delta_1) + \frac{I(X^n; Y_1^n|W)}{n} - \frac{\log n + 1}{2n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right) \\ &\geq n\varphi_{\delta_2}\left(h_b(\delta_1) + \frac{I(X^n; Y_1^n|W)}{n}\right) - \frac{\log n + 1}{2} \varphi'_{\delta_2}\left(h_b(\delta_1) + \frac{I(X^n; Y_1^n|W)}{n}\right) - \sqrt{n\delta_2} \log\left(\frac{n}{\delta_2}\right), \end{aligned} \quad (46b)$$

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

$$D(\rho, \delta + \epsilon) = D(\rho, \delta) + \epsilon D'(\rho, \delta) + \mathcal{O}(\epsilon^2), \quad (54)$$

where

$$D'(\rho, \delta) \triangleq \frac{\partial}{\partial \delta} D(\rho, \delta) = \rho \frac{h'_b(\delta)}{h'_b(D(\rho, \delta))} = \rho \frac{\log\left(\frac{1-\delta}{\delta}\right)}{\log\left(\frac{1-D(\rho, \delta)}{D(\rho, \delta)}\right)}. \quad (55)$$

Taking $\epsilon = \frac{a}{\sqrt{n}}$, this implies that

$$D(\rho, \delta) - D(\rho, \delta_1) = \frac{a}{\sqrt{n}} D'(\rho, \delta) + \mathcal{O}\left(\frac{1}{n}\right) \quad (56a)$$

$$D(\rho, \delta_1 * \delta_2) - D(\rho, \delta) = \frac{a}{\sqrt{n}} D'(\rho, \delta) + \mathcal{O}\left(\frac{1}{n}\right). \quad (56b)$$

and thus, taking the difference between the last two equations, we obtain

$$D(\rho, \delta_1 * \delta_2) + D(\rho, \delta_1) - 2D(\rho, \delta) = \mathcal{O}\left(\frac{1}{n}\right). \quad (57)$$

Now, by Proposition 1 we have:

$$D_1 \geq D(\rho, \delta_1) + \mathcal{O}\left(\frac{\log n}{n}\right) \quad (58a)$$

$$D_2 \geq D(\rho, \delta_1 * \delta_2) + \mathcal{O}\left(\frac{\log n}{n}\right). \quad (58b)$$

Using (56b) and (58b), we can assert:

$$\begin{aligned} D_2 - D(\rho, \delta) &= (D_2 - D(\rho, \delta_1 * \delta_2)) + \\ &\quad (D(\rho, \delta_1 * \delta_2) - D(\rho, \delta)) \\ &\geq \frac{a}{\sqrt{n}} D'(\rho, \delta) + \mathcal{O}\left(\frac{\log n}{n}\right). \end{aligned} \quad (59)$$

Let $\eta = \eta(\rho, \delta)$. We now claim that we can assume without loss of generality:

$$D_1 - D(\rho, \delta_1) < \frac{a\eta}{\sqrt{n}} \quad (60a)$$

$$D_2 - D(\rho, \delta_1 * \delta_2) < \frac{a\eta}{\sqrt{n}}, \quad (60b)$$

To see why this is true, assume to the contrary that one of them, say the first, does not hold. Then, by (58b) and (57),

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

which is stronger than the desired bound.

We now proceed to bound the difference $D_2 - D_1$ invoking Lemma 6 and using (60). By the concavity of $t \mapsto h_b(t)$ and (60b) we have that

$$\begin{aligned} h_b(D_2) &= h_b(D(\rho, \delta_1 * \delta_2) + (D_2 - D(\rho, \delta_1 * \delta_2))) \\ &\leq h_b(D(\rho, \delta_1 * \delta_2)) + \frac{a\eta}{\sqrt{n}} h'_b(D(\rho, \delta_1 * \delta_2)), \end{aligned} \quad (61)$$

which implies that

$$\begin{aligned} \rho C(\delta_1 * \delta_2) - R(D_2) &= \rho C(\delta_1 * \delta_2) - \log 2 + h_b(D_2) \\ &\leq \frac{a\eta}{\sqrt{n}} h'_b(D(\rho, \delta_1 * \delta_2)) \\ &= \frac{a\eta}{\sqrt{n}} \log\left(\frac{1 - D(\rho, \delta_1 * \delta_2)}{D(\rho, \delta_1 * \delta_2)}\right). \end{aligned} \quad (62)$$

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

$$\begin{aligned} D_2 - D_1 &\leq \frac{1 + 2D_2\tau \frac{a\eta}{\sqrt{n}} \log\left(\frac{1 - D(\rho, \delta_1 * \delta_2)}{D(\rho, \delta_1 * \delta_2)}\right) + \rho\Gamma(n, \delta_2)}{2D_2\tau \log\left(\frac{1 - D_2}{D_2}\right)} \\ &\quad + \frac{2a \log\left(\frac{1 - \delta_1}{\delta_1}\right) \Phi(\delta_1)}{\sqrt{n} \log\left(\frac{1 - D_2}{D_2}\right) \Phi(D_2)} (1 + \tau) \frac{g(D_1)}{g(D_2)}. \end{aligned} \quad (63)$$

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 $0 < D(\rho, \delta) < 1/2$, (60a) and (60b), 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 $f(\cdot, \cdot)$ is as defined in (9b). In addition, under our assumptions on δ_1 and δ_2 , we have that $\Gamma(n, \delta_2) = \mathcal{O}(n^{-3/4} \log n)$. Thus, (63) amounts to the following upper bound on the distortion difference:

$$D_2 - D_1 \leq \frac{A}{\sqrt{n}} + \mathcal{O}(n^{-3/4} \log n), \quad (64)$$

where

$$A = a\eta \frac{1 + 2D(\rho, \delta)\tau}{2D(\rho, \delta)\tau} + 2aD'(\rho, \delta)f(\rho, \delta)(1 + \tau).$$

Combining with (59) now yields

$$\begin{aligned} D_1 + D_2 - 2D(\rho, \delta) &= 2(D_2 - D(\rho, \delta)) - (D_2 - D_1) \\ &\geq \frac{2a}{\sqrt{n}} D'(\rho, \delta) - \frac{A}{\sqrt{n}} + \mathcal{O}(n^{-3/4} \log n) \\ &= \frac{1}{\sqrt{n}} \left(2aD'(\rho, \delta)(1 - f(\rho, \delta)(1 + \tau)) \right. \\ &\quad \left. - a\eta \frac{1 + 2D(\rho, \delta)\tau}{2D(\rho, \delta)\tau}\right) + \mathcal{O}(n^{-3/4} \log n). \end{aligned} \quad (65)$$

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

$$\tau = \frac{1 - f(\rho, \delta)}{2f(\rho, \delta)} > 0, \quad (66)$$

for which (65) becomes

$$D_1 + D_2 - 2D(\rho, \delta) \geq \frac{a\eta(\rho, \delta)}{\sqrt{n}} + \mathcal{O}(n^{-3/4} \log n), \quad (67)$$

as desired. ■

V. PROOF OF THEOREM 1

We now turn back to the original point-to-point finite-blocklength JSCC problem. 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

$$D^*(n, \rho, \delta) \geq D^*\left(n, \rho, \frac{\lfloor n\delta \rfloor}{n}\right), \quad (68)$$

and we may therefore write

$$\begin{aligned} \Delta^*(n, \rho, \delta) &= D^*(n, \rho, \delta) - D(\rho, \delta) \\ &\geq D^*\left(n, \rho, \frac{\lfloor n\delta \rfloor}{n}\right) - D\left(\rho, \frac{\lfloor n\delta \rfloor}{n}\right) \\ &\quad + \left[D\left(\rho, \frac{\lfloor n\delta \rfloor}{n}\right) - D(\rho, \delta)\right] \\ &= \Delta^*\left(n, \rho, \frac{\lfloor n\delta \rfloor}{n}\right) + \mathcal{O}\left(\frac{1}{n}\right). \end{aligned} \quad (69)$$

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

$$\mathbb{E}\left[\frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) + Z^n))\right] = \mathbb{E}[\Psi(K)], \quad (70)$$

where

$$\Psi(k) = \Psi_{\mathcal{D}, \mathcal{E}}(k) \triangleq \mathbb{E}\left[\frac{1}{m} d_H(S^m, \mathcal{D}(\mathcal{E}(S^m) \oplus U_k^n))\right], \quad (71)$$

and U_k^n is uniform over $\mathcal{S}_{\delta n + k, n}$. In terms of this function, we have for the optimal encoder/decoder pair

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

We proceed by partitioning K into the two regimes \mathcal{K}_1 and \mathcal{K}_2 as defined in (13a) and (13b), respectively, asserting:

$$\Delta^*(n, \rho, \delta) = \sum_{i=1}^2 \Pr(K \in \mathcal{K}_i) \mathbb{E}[\Psi(K) - D(\rho, \delta) | K \in \mathcal{K}_i]. \quad (72)$$

and lower-bounding each of the two terms.

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

$$\gamma(k) = \frac{\Pr(K = k) - \Pr(K = -k)}{\Pr(K = k) + \Pr(K = -k)}, \quad (73)$$

and manipulate the conditional expectation of Ψ in (74) in the top of the next page. In order to bound the resulting quantities, we approximate

$$D(\rho, \delta + t) = D(\rho, \delta) + c_1 t + c_2 t^2 + \mathcal{O}(t^3),$$

for some $c_1, c_2 \in \mathbb{R}$, thus

$$(D(\rho, \delta + t) - D(\rho, \delta)) + (D(\rho, \delta - t) - D(\rho, \delta)) = \mathcal{O}(t^2) \quad (75a)$$

$$(D(\rho, \delta + t) - D(\rho, \delta)) - (D(\rho, \delta - t) - D(\rho, \delta)) = \mathcal{O}(t). \quad (75b)$$

In addition, in Appendix B we give Lemma 9, showing that⁶ for $k_0 < n^{2/3}$,

$$\gamma(k_0) = \mathcal{O}\left(\max\left(\frac{k_0}{n}, \frac{k_0^3}{n^2}\right)\right), \quad (76)$$

which amounts to

$$\gamma(k_0) = \mathcal{O}\left(\frac{\log^6(n)}{\sqrt{n}}\right), \quad \forall k_0 \leq \sqrt{n} \log^2(n). \quad (77)$$

Applying (75) with $t = k_0/n$ and (76)-(77) under the condition $k_0 \leq \sqrt{n} \log^2(n)$, and noting that by Proposition 1

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

we see that (74) amounts to:

$$\begin{aligned} \mathbb{E}\left[\Psi(K) - D(\rho, \delta) \mid |K| = k_0\right] &\geq \\ &\frac{1 - |\gamma(k_0)|}{2} \left[\left(\Psi(k_0) - D\left(\rho, \delta + \frac{k_0}{n}\right) \right) \right. \\ &\quad \left. + \left(\Psi(-k_0) - D\left(\rho, \delta - \frac{k_0}{n}\right) \right) \right] \\ &\quad + \mathcal{O}\left(\frac{\log^8 n}{n}\right). \end{aligned} \quad (78)$$

Now, applying (75a) and (77) again, we see that (78) can be further bounded as

$$\begin{aligned} \mathbb{E}\left[\Psi(K) - D(\rho, \delta) \mid |K| = k_0\right] &\geq \\ &\frac{1}{2} [(\Psi(k_0) + \Psi(-k_0) - 2D(\rho, \delta))] \left(1 + \mathcal{O}\left(\frac{\log^6(n)}{\sqrt{n}}\right)\right) \\ &\quad + \mathcal{O}\left(\frac{\log^8 n}{n}\right). \end{aligned} \quad (79)$$

By definition of $\underline{D}(\cdot)$ (19), we have that

$$\Psi(k_0) + \Psi(-k_0) \geq \underline{D}(k_0).$$

Thus, the contribution of \mathcal{K}_1 in (72) is at most:

$$\begin{aligned} \mathbb{E}\left[\underline{D}(|K|) - 2D(\rho, \delta) \mid |K| \in \mathcal{K}_1\right] &\left(1 + \mathcal{O}\left(\frac{\log^6(n)}{\sqrt{n}}\right)\right) \\ &\quad + \mathcal{O}\left(\frac{\log^8 n}{n}\right). \end{aligned} \quad (80)$$

⁶A 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 [25, Theorem 5.22]. However, we prefer to present the explicit calculation for the binomial distribution.

$$\begin{aligned}
\mathbb{E} \left[\Psi(K) - D(\rho, \delta) \mid |K| = k_0 \right] &= \Pr(K = k_0 \mid |K| = k_0) [\Psi(k_0) - D(\rho, \delta)] \\
&\quad + \Pr(K = -k_0 \mid |K| = k_0) [\Psi(-k_0) - D(\rho, \delta)] \\
&= \frac{1 + \gamma(k_0)}{2} [\Psi(k_0) - D(\rho, \delta)] + \frac{1 - \gamma(k_0)}{2} [\Psi(-k_0) - D(\rho, \delta)], \\
&= \frac{1 + \gamma(k_0)}{2} \left[\left(\Psi(k_0) - D \left(\rho, \delta + \frac{k_0}{n} \right) \right) + \left(D \left(\rho, \delta + \frac{k_0}{n} \right) - D(\rho, \delta) \right) \right] \\
&\quad + \frac{1 - \gamma(k_0)}{2} \left[\left(\Psi(-k_0) - D \left(\rho, \delta - \frac{k_0}{n} \right) \right) + \left(D \left(\rho, \delta - \frac{k_0}{n} \right) - D(\rho, \delta) \right) \right] \\
&= \frac{1 + \gamma(k_0)}{2} \left[\left(\Psi(k_0) - D \left(\rho, \delta + \frac{k_0}{n} \right) \right) \right] + \frac{1 - \gamma(k_0)}{2} \left[\left(\Psi(-k_0) - D \left(\rho, \delta - \frac{k_0}{n} \right) \right) \right] \\
&\quad + \frac{1}{2} \left[\left(D \left(\rho, \delta + \frac{k_0}{n} \right) - D(\rho, \delta) \right) + \left(D \left(\rho, \delta - \frac{k_0}{n} \right) - D(\rho, \delta) \right) \right] \\
&\quad + \frac{\gamma(k_0)}{2} \left[\left(D \left(\rho, \delta + \frac{k_0}{n} \right) - D(\rho, \delta) \right) - \left(D \left(\rho, \delta - \frac{k_0}{n} \right) - D(\rho, \delta) \right) \right]. \tag{74}
\end{aligned}$$

Now, using Theorem 4, with $a = |K|/\sqrt{n} < \log^2(n)$, we have that

$$\underline{D}(|K|) \geq 2D(\rho, \delta) + \frac{|K|}{n} \eta(\rho, \delta) + \mathcal{O}(n^{-3/4} \log n) \tag{81}$$

where $\eta(\rho, \delta)$ is as defined in (9c). Substituting in (80), we are left with:

$$\begin{aligned}
&\frac{\eta(\rho, \delta)}{\sqrt{n}} \mathbb{E} \left[\frac{|K|}{\sqrt{n}} \mid K \in \mathcal{K}_1 \right] \left(1 + \mathcal{O} \left(\frac{\log^6(n)}{\sqrt{n}} \right) \right) \\
&\quad + \mathcal{O} \left(\frac{\log^8 n}{n} \right) + \mathcal{O}(n^{-3/4} \log n) \\
&= \frac{\sqrt{\delta(1-\delta)} \eta(\rho, \delta)}{\sqrt{n}} \mathbb{E} \left[|W| \mid |W| < \frac{\log^2(n)}{\sqrt{\delta(1-\delta)}} \right] \\
&\quad \left(1 + \mathcal{O} \left(\frac{\log^6(n)}{\sqrt{n}} \right) \right) \\
&\quad + \mathcal{O}(n^{-3/4} \log n), \tag{82}
\end{aligned}$$

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 the theorem of Bikelis (see [25, 5.10.4]), we have that

$$\begin{aligned}
&\mathbb{E} \left[|W| \mid |W| < \frac{\log^2(n)}{\sqrt{\delta(1-\delta)}} \right] \\
&= \mathbb{E} \left[|W_G| \mid |W_G| < \frac{\log^2(n)}{\sqrt{\delta(1-\delta)}} \right] + \mathcal{O} \left(\frac{1}{\sqrt{n}} \right) \\
&= \sqrt{\frac{2}{\pi}} + \mathcal{O} \left(\frac{1}{\sqrt{n}} \right). \tag{83}
\end{aligned}$$

Thus we obtained

$$\Delta^*(n, \rho, \delta) \geq \sqrt{\frac{\delta(1-\delta)}{2\pi n}} \eta(\rho, \delta) + \mathcal{O}(n^{-3/4} \log n), \tag{84}$$

as desired.

VI. AUXILIARY RESULTS: 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 8: If Q^n is a degraded memoryless broadcast channel, the function $G_{Q^{\otimes n}}$ tensorizes, i.e.,

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

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,

$$\begin{aligned}
&G_Q \left(\frac{I(X^n; Y_1^n | W)}{n} \right) \\
&= G_Q \left(\frac{I(Y_1^{n-1}; X^{n-1} | W) + I(X_n; Y_{1,n} | W, Y_1^{n-1})}{n} \right) \\
&= G_Q \left(\frac{n-1}{n} \frac{I(Y_1^{n-1}; X^{n-1} | W)}{n-1} + \frac{I(X_n; Y_{1,n} | W, Y_1^{n-1})}{n} \right) \\
&\geq \frac{n-1}{n} G_Q \left(\frac{I(Y_1^{n-1}; X^{n-1} | W)}{n-1} \right) \\
&\quad + \frac{1}{n} G_Q(I(X_n; Y_{1,n} | W, Y_1^{n-1})) \tag{85}
\end{aligned}$$

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

$$\begin{aligned} G_Q &\left(\frac{I(X^n; Y_1^n | W)}{n} \right) \\ &= \frac{G_{Q^{\otimes(n-1)}}(I(Y_1^{n-1}; X^{n-1} | W)) + G_Q(I(X_n; Y_{1,n} | W, Y_1^{n-1}))}{n} \\ &\geq \frac{I(Y_2^{n-1}, W) + I(Y_{2,n}; W, Y_1^{n-1})}{n}, \end{aligned} \quad (86)$$

where the last inequality follows from the definition of $G_{Q^{\otimes(n-1)}}(t)$ 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

$$\begin{aligned} I(Y_{2,n}; W, Y_1^{n-1}) &\geq I(Y_{2,n}; W, Y_2^{n-1}) \\ &\geq I(Y_{2,n}; W | Y_2^{n-1}). \end{aligned} \quad (87)$$

Substituting (87) into (86) gives

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

such that $G_{Q^{\otimes n}}(nt) \leq nG_Q(t)$. On the other hand,

$$\begin{aligned} G_{Q^{\otimes n}}(nt) &\triangleq \max_{W, X^n : W - X^n - (Y_1^n, Y_2^n)} \max_{I(X^n; Y_1^n | W) \geq nt} I(Y_2^n; W) \\ &\geq \max_{W^n, X^n : W^n - X^n - (Y_1^n, Y_2^n)} \max_{I(X^n; Y_1^n | W^n) \geq nt} I(Y_2^n; W^n) \end{aligned} \quad (88a)$$

$$= nG_Q(t), \quad (88b)$$

where the maximization in (88a) is with respect to all i.i.d. $(W^n, X^n) \sim P_{WX}^{\otimes n}$. ■

The following corollary is an immediate consequence of Theorem 2 and Lemma 8.

Corollary 2: 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 $P_{U|S}$,

$$\bar{R}_P(D_2) \leq \rho \cdot G_Q \left(\frac{F_P(R(D_1))}{\rho} \right). \quad (89)$$

This bound can be obtained as a special case of [16, Theorem 5] (see also [17]), by observing that the boundary of the degraded memoryless broadcast channel Q (without common message) is given by $(C_1, G_Q(C_1))$ [19, Theorem 5.2].⁷

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 [26] hold, these choices give the tightest bound possible. Otherwise, other choices can give tighter bounds, as demonstrated in the examples below.

⁷In fact, the techniques developed in [16] 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 $G_Q(t)$, as those functions have a major role in other problems in network information theory, see [1] for more details.

A. 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 2, we recover the Reznic, Feder, Zamir outer bound [15, Theorem 1]:⁸ For all δ ,

$$\frac{\delta^2 + \sigma^2}{\delta^2 + D_2} \leq \left[\frac{N_1 + N_2 + P}{N_1 \left(\frac{\sigma^2}{D_1} \cdot \frac{\delta^2 + D_1}{\delta^2 + \sigma^2} \right)^{\frac{1}{\rho}} + N_2} \right]^{\rho} \quad (90)$$

B. Binary-Hamming Case

We now address the case where S is a Bernoulli(p) source, and $d(S_j, \hat{S}_j)$ is the Hamming distortion measure.

1) *The Source Functions:* We define $P_{U|S}$ by taking $U = S \oplus N$, where $N \sim \text{Ber}(q)$ is independent of S .

Proposition 2: For $0 \leq t \leq h_b(p)$

$$F_P(t) \geq t - h_b(q * p) + h_b(q * h^{-1}(h_b(p) - t)), \quad (91)$$

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 [20]. 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 $V = S \oplus A$, where $A \sim \text{Ber}(h_b^{-1}(1 - I(S; V)))$. ■

Proposition 3: For $0 \leq D \leq p$

$$\bar{R}_P(D) = h_b(q * p) - h_b(q * D). \quad (92)$$

⁸The bound is given in a different form in [15, Theorem 1], but it is easy to verify that the two expressions are equivalent.

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

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

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

2) *Erasure Channel:* 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 2 and 3 and substituting in Corollary 2, 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 [27] (which was also recovered in [17]).

3) *Binary Symmetric Channel:* 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 4: For the binary symmetric degraded channel

$$G_Q(t) = \log 2 - h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + t)), \quad (93)$$

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

Proof. For any (W, X) satisfying the Markov chain $W - X - (Y_1 = X \oplus Z_1) - (Y_2 = Y_1 \oplus Z_2)$, 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 $Y_2 = Y_1 \oplus Z_2$, with equality if $X \sim \text{Ber}(1/2)$ and $W = X \oplus A$ 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 2 and 3 and substitute in Corollary 2, 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

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

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

$$A_1 = h_b(\delta_1) + \frac{h_b(q * D_1) - h_b(D_1)}{\rho}. \quad (94)$$

Following the treatment of the Gaussian-quadratic case in [15], 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 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 $D_2^* = D(\rho, \delta_1 * \delta_2)$. Then,

$$g(D_1) \geq g(p) + \frac{g(\delta_1)}{g(\delta_1 * \delta_2)} [g(D_2^*) - g(p)], \quad (95)$$

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

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 4: 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,

$$\begin{aligned} (1 - 2D_2)^2 &\leq (1 - 2 \cdot \delta_2 * D_1^*)^2 \\ &\quad + (1 - 2p)^2 (1 - (1 - 2 \cdot \delta_2 * D_1^*)^2). \end{aligned} \quad (96)$$

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

$$D_2 \geq \delta_2 * D_1^*. \quad (97)$$

VII. DISCUSSION: THE REMAINING GAP TO ACHIEVABLE PERFORMANCE

In this work we have shown an example, where $\Delta_n^* = \Omega(n^{-1/2})$. 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

$$D_n \leq (1 - p_n)D_{0,n} + p_n. \quad (98)$$

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,

$$\begin{aligned}\Delta_n &\leq D_{0,n} - D_\infty^* + p_n [1 - D_{0,n}] \\ &= D(\rho R_n) - D(\rho C) + \mathcal{O}(p_n) + \mathcal{O}\left(\frac{\log n}{n}\right) \\ &= \mathcal{O}(C - R_n) + \mathcal{O}(p_n) + \mathcal{O}\left(\frac{\log n}{n}\right).\end{aligned}\quad (99)$$

In order for both the first term and the second term to decrease, we must choose R_n in the moderate-deviations regime. Using [28], [29] we have that:

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

Substituting, we find that a separation-based scheme achieves

$$\Delta_n = \mathcal{O}\left(\sqrt{\frac{\log n}{n}}\right).\quad (100)$$

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 [30], [31], [32]. 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 [33], 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 [34], see [35], [36] 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 (source) 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., [37] in a Gaussian setting) suggest that the required property is not possible for all subsets of coded bits (without restriction on their size).

ACKNOWLEDGMENT

Helpful discussions with Jun Chen are greatly acknowledged.

APPENDIX A PROOF OF PROPOSITION 1

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

$$I(S^m; \mathcal{D}(\mathcal{E}(S^m) \oplus U_k^n)) \geq m(\log 2 - h_b(\Psi(k))).\quad (101)$$

Using the data processing inequality we obtain

$$I(S^m; \mathcal{D}(\mathcal{E}(S^m) \oplus U_k^n)) \leq I(\mathcal{E}(S^m); \mathcal{E}(S^m) \oplus U_k^n).\quad (102)$$

Recalling that [23, Chapter 10, Lemma 7]

$$\begin{aligned}H(U_k^n) &\geq nh_b\left(\delta + \frac{k}{n}\right) - \frac{\log n}{2} \\ &\quad - \frac{1}{2} \log\left(2\pi\left(\delta + \frac{k}{n}\right)\left(1 - \delta - \frac{k}{n}\right)\right) \\ &\geq nh_b\left(\delta + \frac{k}{n}\right) - \frac{\log n}{2} - \frac{1}{2},\end{aligned}\quad (103)$$

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

$$\begin{aligned}I(X^n; X^n \oplus U_k^n) &\leq H(X^n \oplus U_k^n) - H(U_k^n) \\ &\leq n \log 2 - nh_b\left(\delta + \frac{k}{n}\right) + \frac{\log n}{2} + \frac{1}{2}.\end{aligned}\quad (104)$$

Thus, combining (101), (102), and (104), and recalling that $\rho = n/m$, we have

$$\log 2 - h_b(\Psi(k)) \leq \rho \left(\log 2 - h_b\left(\delta + \frac{k}{n}\right) + \frac{\log n + 1}{2n} \right),\quad (105)$$

which yields the desired result.

APPENDIX B

THE ASYMMETRY OF THE BINOMIAL DISTRIBUTION

Lemma 9: 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

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

Proof. Using the binomial distribution, we have

$$\begin{aligned} \Gamma &\triangleq \frac{\Pr(K = k)}{\Pr(K = -k)} \\ &= \frac{\binom{n}{n\delta+k} \left(\frac{\delta}{1-\delta}\right)^k}{\binom{n}{n\delta-k} \left(\frac{\delta}{1-\delta}\right)^{-k}} \\ &= \frac{(n\delta-k)! (n(1-\delta)+k)!}{(n\delta+k)! (n(1-\delta)-k)!} \left(\frac{\delta}{1-\delta}\right)^{2k} \\ &= \frac{(n(1-\delta))^{2k} \prod_{i=1}^{2k} \left(1 - \frac{k}{(1-\delta)n} + \frac{i}{n(1-\delta)}\right)}{(n\delta)^{2k} \prod_{i=1}^{2k} \left(1 - \frac{k}{\delta n} + \frac{i}{n\delta}\right)} \left(\frac{\delta}{1-\delta}\right)^{2k} \\ &= \prod_{i=1}^{2k} \frac{1 - \frac{k}{(1-\delta)n} + \frac{i}{n(1-\delta)}}{1 - \frac{k}{\delta n} + \frac{i}{n\delta}} \end{aligned} \quad (106)$$

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

$$\log \Gamma = \sum_{i=1}^{2k} \log\left(1 + \frac{1}{1-\delta} b_i\right) - \sum_{i=1}^{2k} \log\left(1 + \frac{1}{\delta} b_i\right). \quad (107)$$

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

$$\begin{aligned} \log \Gamma &= \left(\frac{1}{1-\delta} - \frac{1}{\delta}\right) \sum_{i=1}^{2k} b_i - \frac{1}{2} \left(\frac{1}{(1-\delta)^2} - \frac{1}{\delta^2}\right) \sum_{i=1}^{2k} b_i^2 \\ &\quad + o\left(\frac{k^3}{n^2}\right). \end{aligned} \quad (108)$$

Furthermore, we have

$$\sum_{i=1}^{2k} b_i = \frac{k}{n} + \sum_{j=-(k-1)}^{k-1} \frac{j}{n} = \frac{k}{n} \quad (109a)$$

$$\begin{aligned} \sum_{i=1}^{2k} b_i^2 &= \frac{k^2}{n^2} + 2 \sum_{j=0}^{k-1} \frac{j^2}{n^2} \\ &= \frac{k^2}{n^2} + \frac{2}{n^2} \frac{(k-1) \cdot k \cdot (2k-1)}{6} \\ &= \frac{2}{3} \frac{k^3}{n^2} + o\left(\frac{k^3}{n^2}\right). \end{aligned} \quad (109b)$$

Thus,

$$\begin{aligned} \log \Gamma &= -\frac{1-2\delta}{\delta(1-\delta)} \frac{k}{n} + \frac{1-2\delta}{(\delta(1-\delta))^2} \frac{k^3}{3n^2} + o\left(\frac{k^3}{n^2}\right) \\ &= \frac{(1-2\delta)}{\delta(1-\delta)} \left[\frac{k^2}{n} \frac{1}{3\delta(1-\delta)} - 1 \right] \frac{k}{n} + o\left(\frac{k^3}{n^2}\right). \end{aligned} \quad (110)$$

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

$$\Gamma = 1 + \frac{(1-2\delta)}{\delta(1-\delta)} \left[\frac{k^2}{n} \frac{1}{3\delta(1-\delta)} - 1 \right] \frac{k}{n} + o\left(\frac{k^3}{n^2}\right). \quad (111)$$

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

APPENDIX C
AUXILIARY LEMMAS

Lemma 10 (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

$$h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + x)) \geq h_b(\delta_1 * \delta_2) + \frac{g(\delta_1 * \delta_2)}{g(\delta_1)} x, \quad (112)$$

where $g(t) = (1-2t) \log\left(\frac{1-t}{t}\right)$, as defined in (52).

Proof. Let $\varphi_{\delta_2}(t) = h_b(\delta_2 * h_b^{-1}(t))$ and recall that by [20] we have that $t \mapsto \varphi_{\delta_2}(t)$ is convex, and

$$\begin{aligned} \varphi'_{\delta_2}(t) &= \frac{d}{dt} \varphi_{\delta_2}(t) \\ &= (1-2\delta_2) \frac{h'_b(\delta_2 * h_b^{-1}(t))}{h'_b(h_b^{-1}(t))}. \end{aligned} \quad (113)$$

Consequently,

$$\begin{aligned} \varphi'_{\delta_2}(h_b(\delta_1)) &= (1-2\delta_2) \frac{h'_b(\delta_1 * \delta_2)}{h'_b(\delta_1)} \\ &= \frac{(1-2\delta_2)(1-2\delta_1)}{(1-2\delta_1)} \frac{h'_b(\delta_1 * \delta_2)}{h'_b(\delta_1)} \\ &= \frac{(1-2(\delta_1 * \delta_2))}{(1-2\delta_1)} \frac{h'_b(\delta_1 * \delta_2)}{h'_b(\delta_1)} \\ &= \frac{g(\delta_1 * \delta_2)}{g(\delta_1)}. \end{aligned} \quad (114)$$

Now, by convexity, we have

$$h_b(\delta_2 * h_b^{-1}(h_b(\delta_1) + x)) \geq \varphi_{\delta_2}(h_b(\delta_1)) + \varphi'_{\delta_2}(h_b(\delta_1))x, \quad (115)$$

and the statement follows by substituting (114) into (115). ■

Lemma 11 (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

$$g'(t) = \frac{d}{dt} g(t) = -\kappa(t), \quad (116)$$

where

$$\kappa(t) \triangleq 2 \log\left(\frac{1-t}{t}\right) + \frac{1-2t}{t(1-t)}. \quad (117)$$

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 12 (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

$$\beta'_q(t) = \frac{d}{dt}\beta_q(t) = -\phi(q, t), \quad (118)$$

where

$$\phi(q, t) \triangleq 2q \log\left(\frac{1-t}{t}\right) + (1-2q) \log\left(\frac{1+q\frac{1-2t}{t}}{1-q\frac{1-2t}{1-t}}\right). \quad (119)$$

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

$$\phi(q, t) \geq q \cdot \kappa(t) \cdot \nu(q, t), \quad (120)$$

where $\kappa(t)$ is as in (117), and

$$\nu(q, t) \triangleq \frac{1-2q}{1+q\frac{1-2t}{t}}. \quad (121)$$

Moreover,

$$\beta_q(t) \leq q \cdot g(t). \quad (122)$$

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

$$\begin{aligned} \log\left(1 + q\frac{1-2t}{t}\right) &\geq q\frac{\frac{1-2t}{t}}{1+q\frac{1-2t}{t}} \\ -\log\left(1 - q\frac{1-2t}{1-t}\right) &\geq q\frac{1-2t}{1-t}, \end{aligned}$$

and therefore

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

which gives

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

To prove (122), we apply the concavity of $t \mapsto h_b(t)$ to obtain

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

■

APPENDIX D PROOF OF LEMMA 6

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

$$\begin{aligned} &\frac{\log 2 - h_b(q * D_2)}{\rho} \\ &\leq \log 2 - h_b\left(\delta_2 * h_b^{-1}\left(h_b(\delta_1) + \frac{h_b(q * D_1) - h_b(D_1)}{\rho}\right)\right) \\ &\quad + \Gamma(n, \delta_2). \end{aligned} \quad (123)$$

Now, applying Lemma 10, we can write

$$\begin{aligned} &h_b\left(\delta_2 * h_b^{-1}\left(h_b(\delta_1) + \frac{h_b(q * D_1) - h_b(D_1)}{\rho}\right)\right) \\ &\geq h_b(\delta_1 * \delta_2) + \frac{g(\delta_1 * \delta_2)\beta_q(D_1)}{g(\delta_1)\rho}, \end{aligned} \quad (124)$$

where $\beta_q(\cdot)$ is as defined in Lemma 12. By combining (123) and (124) and invoking the definition of $C(t)$ and $R(t)$, we have:

$$\begin{aligned} &\rho C(\delta_1 * \delta_2) - R(D_2) \\ &\geq \frac{g(\delta_1 * \delta_2)}{g(\delta_1)}\beta_q(D_1) - \beta_q(D_2) - \rho\Gamma(n, \delta_2). \end{aligned} \quad (125)$$

By Lemma 11, the function $t \mapsto g(t)$ is convex, and consequently

$$\begin{aligned} g(\delta_1 * \delta_2) &= g(\delta_1 + (1-2\delta_1)\delta_2) \\ &\geq g(\delta_1) + \delta_2 \cdot (1-2\delta_1)g'(\delta_1). \end{aligned} \quad (126)$$

Thus, we obtain

$$\begin{aligned} \rho C(\delta_1 * \delta_2) - R(D_2) &\geq \beta_q(D_1) - \beta_q(D_2) \\ &\quad + \delta_2 \frac{(1-2\delta_1)g'(\delta_1)}{g(\delta_1)}\beta_q(D_1) \\ &\quad - \rho\Gamma(n, \delta_2) \\ &= \beta_q(D_1) - \beta_q(D_2) - \psi(\delta_1)\delta_2\beta_q(D_1) \\ &\quad - \rho\Gamma(n, \delta_2), \end{aligned} \quad (127)$$

where

$$\begin{aligned} \psi(t) &\triangleq -\frac{(1-2t)g'(t)}{g(t)} \\ &= (1-2t)\frac{\kappa(t)}{g(t)}, \end{aligned} \quad (128)$$

where $\kappa(\cdot)$ is as defined in Lemma 11. By Lemma 12, the function $t \mapsto \beta_q(t)$ is convex, and consequently

$$\begin{aligned} \beta_q(D_1) &= \beta_q(D_2 + (D_1 - D_2)) \\ &\geq \beta_q(D_2) + \beta'_q(D_2)(D_1 - D_2) \\ &= \beta_q(D_2) + \phi(q, D_2)(D_2 - D_1), \end{aligned} \quad (129)$$

where $\phi(\cdot, \cdot)$ is as defined in Lemma 12. We can thus bound the distortion difference in (130) at the top of the next page, which is justified as follows. (130a) follows from combining (129) and (127). (130b) holds since by Lemma 12, we have that $\beta_q(D_1) \leq q \cdot g(D_1)$. For (130c) we have used the identity $\delta_2 =$

$$D_2 - D_1 \leq \frac{[\rho C(\delta_1 * \delta_2) - R(D_2)] + \psi(\delta_1)\delta_2\beta_q(D_1) + \rho\Gamma(n, \delta_2)}{\phi(q, D_2)} \quad (130a)$$

$$\leq \frac{\rho C(\delta_1 * \delta_2) - R(D_2) + \rho\Gamma(n, \delta_2)}{\phi(q, D_2)} + \frac{\psi(\delta_1)\delta_2 g(D_1)}{\phi(q, D_2)/q} \quad (130b)$$

$$\leq \frac{\rho C(\delta_1 * \delta_2) - R(D_2) + \rho\Gamma(n, \delta_2)}{2q \log\left(\frac{1-D_2}{D_2}\right)} + (\delta_1 * \delta_2 - \delta_1) \frac{\psi(\delta_1)}{1-2\delta_1} \frac{g(D_2)}{\kappa(D_2)} \frac{1}{\nu(q, D_2)} \frac{g(D_1)}{g(D_2)} \quad (130c)$$

$$= \frac{\rho C(\delta_1 * \delta_2) - R(D_2) + \rho\Gamma(n, \delta_2)}{\frac{2D_2\tau}{1+2D_2\tau} \log\left(\frac{1-D_2}{D_2}\right)} + (\delta_1 * \delta_2 - \delta_1) \frac{\psi(\delta_1)}{1-2\delta_1} \frac{g(D_2)}{\kappa(D_2)} (1+\tau) \frac{g(D_1)}{g(D_2)}. \quad (130d)$$

$\frac{\delta_1 * \delta_2 - \delta_1}{1-2\delta_1}$, the fact that $\phi(q, D_2) > 2q \log\left(\frac{1-D_2}{D_2}\right)$ by (119) and that $\phi(q, D_2) > q \cdot \kappa(t) \cdot \nu(q, t)$ by (121). Finally, let

$$\tau \triangleq \frac{q}{(1-2q)D_2}, \quad (131)$$

such that $\frac{1}{\nu(q, D_2)} = 1 + \tau$, and $q = \frac{D_2\tau}{1+2D_2\tau}$. Since (130c) holds for any $0 < q < 1/2$, we have that (130d) holds for any $\tau > 0$.

To complete the proof, note that

$$\begin{aligned} \frac{\psi(t)}{1-2t} &= \frac{\kappa(t)}{g(t)} \\ &= \frac{1}{1-2t} + \frac{1}{t(1-t) \log\left(\frac{1-t}{t}\right)} \\ &= \log\left(\frac{1-t}{t}\right) \Phi(t), \end{aligned} \quad (132)$$

where $\Phi(t)$ is defined in (9a). Thus, (130d) establishes our claim.

APPENDIX E 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

$$\rho(d) = \frac{\log 2 - h_b(d)}{\log 2 - h_b(\delta)} = \frac{R(d)}{R(\delta)}. \quad (133)$$

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

$$\frac{R(d)}{R(\delta)} \Phi(d) > \Phi(\delta), \quad (134)$$

which is equivalent to the claim that

$$\vartheta(t) \triangleq \Phi(t) \cdot R(t), \quad (135)$$

is monotone decreasing in t , which we now establish.

To this end, note that we can write

$$\vartheta(t) = \frac{\gamma(t)}{g^2(t)}, \quad (136)$$

where

$$\gamma(t) \triangleq (1-2t)R(t)\kappa(t), \quad (137)$$

and $\kappa(t) = -g'(t)$ is as defined in (117). We therefore have that

$$\begin{aligned} \vartheta'(t) &= \frac{1}{g^4(t)} [g^2(t)\gamma'(t) - 2g(t)g'(t)\gamma(t)] \\ &= \frac{1}{g^3(t)} \left[\underbrace{g(t)\gamma'(t) + 2\kappa(t)\gamma(t)}_{\zeta(t)} \right]. \end{aligned} \quad (138)$$

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

$$\begin{aligned} \gamma'(t) &= \kappa(t)(1-2t)R'(t) + \kappa(t)R(t)[(1-2t)'] \\ &\quad + (1-2t)R(t)\kappa'(t) \\ &= -\kappa(t)g(t) - 2\kappa(t)R(t) + (1-2t)R(t)\kappa'(t), \end{aligned} \quad (139)$$

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 have that

$$\kappa(t)\gamma(t) = 2g(t)\kappa(t)R(t) + \frac{(1-2t)^2}{t(1-t)}\kappa(t)R(t). \quad (140)$$

Thus we can evaluate $\zeta(t)$ as in (141) at the top of the next page, where for (141a) we substitute

$$\begin{aligned} \kappa'(t) &= -\frac{2}{t(1-t)} - \frac{1-2t+2t^2}{(t(1-t))^2} \\ &= -\frac{1}{(t(1-t))^2}. \end{aligned} \quad (142)$$

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.

$$\begin{aligned}
\zeta(t) &= -\kappa(t)g^2(t) - 2g(t)\kappa(t)R(t) + (1-2t)g(t)R(t)\kappa'(t) + 4g(t)\kappa(t)R(t) + 2\frac{(1-2t)^2}{t(1-t)}\kappa(t)R(t) \\
&= -\kappa(t)g^2(t) + 2g(t)\kappa(t)R(t) + (1-2t)g(t)R(t)\kappa'(t) + 2\frac{(1-2t)^2}{t(1-t)}\kappa(t)R(t). \tag{141a} \\
&= -\kappa(t)g^2(t) + 2g(t)\kappa(t)R(t) - \frac{R(t)g(t)(1-2t)}{(t(1-t))^2} + 2\frac{(1-2t)^2}{t(1-t)}\kappa(t)R(t) \\
&= -g(t) \left[\kappa(t)g(t) - 2\kappa(t)R(t) + \frac{R(t)(1-2t)}{(t(1-t))^2} - 2\frac{(1-2t)}{t(1-t)\log\left(\frac{1-t}{t}\right)}\kappa(t)R(t) \right] \\
&= -g(t) \left[\kappa(t)g(t) - 2\kappa(t)R(t) + \frac{R(t)(1-2t)}{(t(1-t))^2} - \frac{4R(t)(1-2t)}{t(1-t)} - \frac{2R(t)(1-2t)^2}{t^2(1-t)^2\log\left(\frac{1-t}{t}\right)} \right] \\
&= -g(t) \left[\kappa(t)g(t) - 2\kappa(t)R(t) + \frac{R(t)(1-2t)^3}{(t(1-t))^2} - \frac{2R(t)(1-2t)^2}{t^2(1-t)^2\log\left(\frac{1-t}{t}\right)} \right] \\
&= -g(t) \left[\underbrace{\kappa(t)g(t) - 2\kappa(t)R(t)}_{A_1(t)} + \underbrace{\frac{R(t)(1-2t)^2}{(t(1-t))^2}}_{A_2(t)} \underbrace{\left[1 - 2 \left(t + \frac{1}{\log\left(\frac{1-t}{t}\right)} \right) \right]}_{A_3(t)} \right]. \tag{141b}
\end{aligned}$$

- [2] —, “A lower bound on the expected distortion of joint source-channel coding,” in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019.
- [3] 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.
- [4] 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.
- [5] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley-Interscience, 2006.
- [6] 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.
- [7] 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.
- [8] —, “The dispersion of joint source-channel coding,” in *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept 2011, pp. 180–187.
- [9] 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.
- [10] 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.
- [11] —, “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.
- [12] 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.
- [13] 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.
- [14] B. Yu and T. Speed, “A rate of convergence result for a universal D-semifaithful code,” *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 813–820, May 1993.
- [15] 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.
- [16] 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.
- [17] —, “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.
- [18] 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.
- [19] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*. Elsevier, 1977.
- [24] 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.
- [25] V. V. Petrov, *Limit theorems of probability theory: sequences of independent random variables*. Clarendon Press, Oxford University Press, 1995.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.

- [32] 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.
- [33] H. Roozbehani and Y. Polyanskiy, "Triangulation codes: a family of non-linear codes with graceful degradation," in *2018 Conf. Inform. Sciences and Syst.(CISS)*.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.

Yuval Kochman received the B.Sc., M.Sc. and Ph.D. degrees from Tel Aviv University in 1993, 2003, and 2010, respectively, all in electrical engineering. During 2009–2011, he was a Postdoctoral Associate with the Signals, Information and Algorithms Laboratory, MIT. Since 2012, he has been with the School of Computer Science and Engineering, Hebrew University of Jerusalem. Outside academia, he has worked in the areas of radar, digital communications and video compression. His research interests include information theory, communications and signal processing.

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

Yury Polyanskiy (S'08-M'10-SM'14) is an Associate Professor of Electrical Engineering and Computer Science and a member of LIDS at MIT. Yury received M.S. degree in applied mathematics and physics from the Moscow Institute of Physics and Technology, Moscow, Russia in 2005 and Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ in 2010. His research interests span information theory, statistical learning, error-correcting codes, wireless communication and fault tolerance. Dr. Polyanskiy won the 2013 NSF CAREER award and 2011 IEEE Information Theory Society Paper Award.