Ozarow-Type Outer Bounds for Memoryless Sources and Channels

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

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

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

***Abstract*—Two problems, namely multiple-description source coding and joint source-channel broadcasting of a common source, are addressed. For the multiple-description problem, we revisit Ozarow's technique for establishing impossibility results, and extend it to general sources and distortion measures. For the problem of sending a source over a broadcast channel, we revisit the bounding technique of Reznik, Feder and Zamir, and extend it to general sources, distortion measures and broadcast channels. Although the obtained bounds do not improve over existing results in the literature, they are relatively easy to evaluate, and their derivation reveals the similarities between the two bounding techniques.**

I. Introduction

The multiple-description (MD) problem [1] is one of the challenging settings in lossy source coding. An encoder translates a source sequence into two descriptions, but it may be that only one of them reaches the decoder, as might happen when the descriptions are sent separately over a packet-loss channel (e.g., the internet). There is an inherent tradeoff: for a good reproduction from an individual description, it must convey the "most important" information about the source string; however, that would mean that for the sake of reconstruction from both descriptions, the encoder has included unnecessary redundancy.

Translating this intuition into rigorous impossibility results turns out to be a challenging task. Ahlswede proved a tight converse for the "easier case" where the reproduction from both descriptions is optimal [2]. For the quadratic Gaussian case, Ozarow [1] proved a tight converse based upon augmenting the source with an auxiliary variable and using the entropy-power inequality (EPI). In both cases, the matching achievability result is that of El Gamal and Cover [3].

A somewhat related setting is joint source-channel coding (JSCC) of a common source, to be conveyed to two users over a degraded broadcast channel. Here too there is an inherent tradeoff: a highly redundant description that might help the "weak" user, is redundant for the "strong" one. It is known that source-channel separation is suboptimal, and indeed in some cases it is possible to achieve simultaneous optimality for both users using scalar ("analog") coding. One of these cases is of a Gaussian source with quadratic loss, over an additive white Gaussian noise (AWGN) broadcast channel, where the number of channel uses and source samples is equal. But what happens when these numbers differ ("bandwidth mismatch")? Many schemes (bot hybrid digital-analog and semi-analog) have been proposed, but the optimum tradeoff remains an open problem. In [4], Reznic, Feder and Zamir have put forward an outer bound that is inspired by Ozarow's technique for the MD problem. Most notably, their bound states that if the scheme is optimal for the weak user, the signal-to-distortion ratio of the strong user may only improve proportionally to the signal-to-noise ratio, even if the bandwidth expansion ratio is large.

In this paper, we revisit the approach of [1], [4] and extend it beyond the Gaussian-quadratic setting, to general memoryless MD and JSCC problems. For both settings we derive outer bounds that depend on the choice of an auxiliary variable. We stress that our MD (resp., JSCC) bound can also be obtained as a relaxation of a (possibly stronger) result of Song, Shao and Chen [5] (resp., of Khezeli and Chen [6], [7]). Our bounds, however, are easier to evaluate and agree with those prior results in all the examples we are aware of. Pedagogically, our derivation highlights the relation between Ozarow's technique, and the one used by Reznik, Feder and Zamir, by revealing the dual role the auxiliary channel plays in the JSCC outer bound.

II. Problem Definition

Let the source S^m be drawn according to an i.i.d. distribution P_S (we keep alphabets abstract in order to accommodate both discrete and continuous distributions). For a decoder index $i \in \{0,1,2\}$, let $\{\hat{S}_i^m\}$ be the i-th reconstructed source sequence. Consider some additive distortion measure

$$d(S^m, \hat{S}^m) = \frac{1}{m}\sum_{j=1}^{m} d(S_j, \hat{S}_j),$$

In the *multiple descriptions* problem we are given a pair of rates R_1 and R_2. The encoders $M_i = f_i(S^m)$, $i = 1,2$, produce indices in sets of size at most 2^{mR_i}. The "side decoders" produce reconstructions $\hat{S}_i^m = g_i(M_i)$, $i = 1,2$,

The work of Y. Kochman was supported in part by the HUJI Cyber Security Research Center in conjunction with the Israel National Cyber Bureau in the Prime Minister's Office. The work of O. Ordentlich was supported by ISF under Grant 1791/17. The work of Y. Polyanskiy was supported in part by the National Science Foundation CAREER award under grant agreement CCF-12-53205, an NSF grant CCF-17-17842 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370

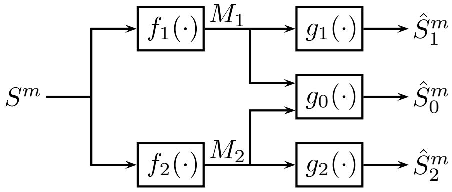

Fig. 1: The Multiple Descriptions Problem.

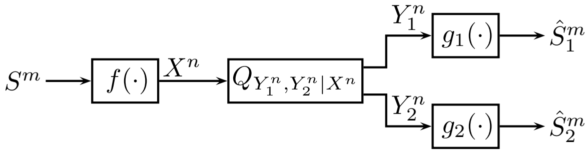

Fig. 2: JSCC over a Broadcast Channel.

while the "central decoder" produces $\hat{S}_0^m = g_0(M_1, M_2)$, see Fig. 1. For given P_S, R_1 and R_2, a distortion triplet (D_0, D_1, D_2) is said to be achievable if there exist two encoders and three decoders such that $\mathbb{E}[d(S^m, \hat{S}_i^m)] \leq D_i$ for $i = 0, 1, 2$ in the limit of large m.

In the *JSCC broadcast* problem there is no rate constraint, but we are also given an n-letter broadcast channel $Q_{Y_1^n,Y_2^n|X^n}$; the channel may be subject to a cost constraint in the usual manner. The source, channels and reconstructions are connected by an encoder $X^n = f(S^n)$ and decoders $\hat{S}_i^m = g_i(Y_i^m)$, see Fig. 2. 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 $f()$, $g_1()$ and $g_2()$ such that $\mathbb{E}[d(S^m, \hat{S}_i^m)] \leq D_i$ for $i = 1, 2$. It will be convenient to express results using the *bandwidth expansion factor* $\rho = n/m$. Special attention will be given to the memoryless degraded broadcast channel case, where $Q_{Y_1^n,Y_2^n|X^n}(y_1^n, y_2^n|x^n) = \prod_{i=1}^n Q_{Y_1|X}(y_{1i}|x_i) Q_{Y_2|Y_1}(y_{2i}|y_{1i})$.

III. Main Results

For our results, we need the following functions of the source. We define an auxiliary variable via a conditional distribution $P_{U|S}$. By combining with the given P_S we obtain $P = P_{S,U}$. With respect to this distribution, we define:

$$F_P(t) \triangleq \min_{\substack{V\,:\,U-S-V \\ I(S;V)\geq t}} I(S;V|U), \tag{1}$$

$$\bar{R}_P(D) \triangleq \min_{\substack{\hat{S}\,:\,U-S-\hat{S} \\ \mathbb{E}d(S,\hat{S})\leq D}} I(U;\hat{S}) \tag{2}$$

The function F_P satisfies the following key properties.

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

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

Furthermore, the following gives an operational rate-distortion meaning to $\bar{R}_P(D)$. Notice that both from the operational and analytical points of view, when $U = S$ it reduces to the rate-distortion function (RDF) of the source S.

Lemma 2: Let $(U_i, S_i) \overset{iid}{\sim} P_{U,S}$ and $\hat{S}^m$ be a random vector satisfying the Markov chain $U^m - S^m - \hat{S}^m$ and $\mathbb{E}d(S^m, \hat{S}^m) \leq D$, then

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

The proofs of these lemmas appear in Section V. In the sequel, $R(D)$ will be the standard RDF for the source P_S. The lemmas above can be used for establishing a lower bound on the sum-rate of the MD problem.

Theorem 1: Consider the MD problem. If the distortions (D_0, D_1, D_2) are achievable at rates (R_1, R_2), then for any P defined by a choice of U,

$$R_1 + R_2 \geq F_P(R(D_0)) + \bar{R}_P(D_1) + \bar{R}_P(D_2). \tag{3}$$

The theorem can be obtained as a corollary from [5, Theorem 1], as follows: set $Z_2 = S$, $Z_1 = U$, $Z_0 = \emptyset$ there, and notice that $I(U; \hat{S}_i) \geq \bar{R}_P(D_i)$, $i = 1, 2$ and that

$$\begin{aligned} I(S; \hat{S}_0, \hat{S}_1, \hat{S}_2|U) &\geq I(S; \hat{S}_0|U) \\ &\geq F_P(I(S; \hat{S}_0)) \geq F_P(R(D_0)). \end{aligned} \tag{4}$$

Interestingly, even though this derivation shows that the bound of Theorem 1 may be weaker than that of [5, Theorem 1], in the two cases for which the optimization it involves was successfully solved in [5], the two bounds coincide. We bring the proof of the result, in order to demonstrate the usefulness of the lemmas above, and the resemblance to the JSCC setup.

Proof of Theorem 1: The first steps of the proof follow Ozarow's [1] proof for the Gaussian case. Specifically, from $I(S^m; M_1, M_2|U^m) \leq H(M_1|U^m) + H(M_2|U^m)$, we have

$$\begin{aligned} &m(R_1 + R_2) \geq \\ &I(S^m; M_1, M_2|U^m) + I(U^m; M_1) + I(U^m; M_2). \end{aligned} \tag{5}$$

Now, invoking Lemma 2 and the definition of F_{P^m}, we can further bound (5) as

$$\begin{aligned} R_1 + R_2 &\geq \frac{F_{P^m}(I(S^m; M_1, M_2))}{m} + \bar{R}_P(D_1) + \bar{R}_P(D_2) \\ &= F_P\left(\frac{I(S^m; M_1, M_2)}{m}\right) + \bar{R}_P(D_1) + \bar{R}_P(D_2) \quad (6) \\ &\geq F_P(R(D_0)) + \bar{R}_P(D_1) + \bar{R}_P(D_2), \quad (7) \end{aligned}$$

where (6) follows from tensorization of F_{P^m}, and (7) follows since $I(S^m; M_1, M_2) \geq I(S^m; \hat{S}_0^m) \geq mR(D_0)$. ■

For the JSCC problem, we need to also consider the channel law $Q^n = Q_{Y_1^n,Y_2^n|X^n}(y_1^n, y_2^n|x^n)$. Although our main interest will be in degraded memoryless broadcast channels, we will first prove our result for a general n-letter broadcast channel, and only then specialize to the degraded memoryless case. For the channel Q^n, we define:

$$G_{Q^n}(t) \triangleq \max_{\substack{W,X^n\,:\,W-X^n-Y_1^n-Y_2^n \\ I(X^n;Y_1^n|W)\geq t}} I(Y_2^n; W). \tag{8}$$

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 $nR_2 \leq G_{Q^n}(nR_1)$ [8, Chapter 5.4.1] .

We now state our result. Although it is stated 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.

Theorem 2: Consider the problem of transmitting m realizations of the i.i.d. source S, over the n-letter broadcast channel Q^n. If (D_1, D_2) is achievable, then for any P defined by a choice of U,

$$m\bar{R}_P(D_2) \leq G_{Q^n}\left(mF_P\left(R(D_1)\right)\right). \tag{9}$$

Proof: 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

$$\begin{aligned} m\bar{R}_P(D_2) &\leq I(U^m, \hat{S}_2^m) \\ &\leq I(U^m; Y_2^n) && (10)\\ &\leq G_{Q^n}(I(X^n; Y_1^n|U^m)) && (11)\\ &\leq G_{Q^n}\left(I(S^m; Y_1^n|U^m)\right) && (12)\\ &\leq G_{Q^n}\left(F_{P^m}\left(I(S^m; Y_1^n)\right)\right) && (13)\\ &= G_{Q^n}\left(mF_P\left(\frac{I(S^m; Y_1^n)}{m}\right)\right) && (14)\\ &\leq G_{Q^n}\left(mF_P\left(\frac{I(S^m; \hat{S}_1^m)}{m}\right)\right) && (15)\\ &\leq G_{Q^n}\left(mF_P\left(R(D_1)\right)\right), \end{aligned}$$

where (10) follows from the data processing inequality (DPI), (11) from definition of G_{Q^n}, (12) from the DPI and monotonicity of G_{Q^n}, (13) from definition of F_{P^m}, (14) from tensorization of F_{P^m}, and (15) from the DPI. ∎

Note that U^m plays a two-fold role here: in (11) we used the Markov chain $U^m - X^n - Y_1^n - Y_2^n$, whereas in (12) we used $U^m - S^m - Y_1^n$. Thus, the source two-descriptions problem, and the broadcast channel problem are *coupled* via the same auxiliary variable U^m. This is also the main weakness of the bound above: Even though the same U^m, whose distribution is fixed once we choose the channel $P_{U|S}$, appears in both Markov chains, in the transition from (10) to (11), we have used the definition of G_{Q^n}, which involves a *maximization* with respect to U^m.

We now consider the special case where the channel Q^n is degraded and memoryless. i.e., $Q^n_{Y_1^n, Y_2^n|X^n}(y_1^n, y_2^n|x^n) = \prod_{i=1}^n Q_{Y_1|X}(y_{1i}|x_i) Q_{Y_2|Y_1}(y_{2i}|y_{1i})$. We have the following (for proof see Section V).

Lemma 3: The function $G_Q(t)$ is monotone non-increasing and concave. Furthermore, if Q^n is a degraded memoryless broadcast channel, it tensorizes, i.e.,

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

The following Theorem is an immediate corollary of Theorem 2 and Lemma 3.

Theorem 3: Consider the degraded memoryless JSCC broadcast problem. If (D_1, D_2) is achievable, then for any P defined by a choice of U,

$$\bar{R}_P(D_2) \leq \rho \cdot G_Q\left(\frac{F_P\left(R(D_1)\right)}{\rho}\right). \tag{16}$$

This bound can be obtained as a special case of [6, Theorem 5], by observing that the capacity region boundary of the degraded memoryless broadcast channel Q (without common message) is given by $(C_1, G_Q(C_1))$ [8, Theorem 5.2].

It is not difficult to see that for $U = \emptyset$, our bound reads $R(D_1) \leq \rho \max_X I(X; Y_1)$, whereas for the choice $U = S$ it reduces to $R(D_2) \leq \rho \max_W I(W; Y_2)$. Those are the two extreme cases, when only the distortion of the reconstruction at one terminal is of interest.

IV. Special Cases

A. Quadratic Gaussian Case

In this case, P_S is Gaussian $(0, \sigma^2)$ 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

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

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

where F_P is attained by taking V that is the output of an AWGN with input S. Substituting these quantities and the quadratic-Gaussian RDF in Theorem 1 yields that for all δ,

$$R_1 + R_2 \geq \frac{1}{2}\log\left(\frac{\sigma^2}{D_0}\right) + \frac{1}{2}\log\left(\frac{(\delta^2 + \sigma^2)(\delta^2 + D_0)}{(\delta^2 + D_1)(\delta^2 + D_2)}\right),$$

which is exactly Ozarow's quadratic-Gaussian MD sum-rate (tight) bound.

Now we combine the Gaussian source with an 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 independnet, and 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 RDF, we recover the Reznic et al. outer bound [4, Theorem 1]: For all δ,

$$\frac{\delta^2 + \sigma^2}{\delta^2 + D_2} \leq \left(1 + \frac{P}{N_1 + N_2}\right)^\rho \left[\frac{N_1 + N_2}{N_1\left(\frac{\sigma^2}{D1} \cdot \frac{\delta^2 + D_1}{\delta^2 + \sigma^2}\right)^{\frac{1}{\rho}} + N_2}\right]^\rho.$$

B. Binary-Hamming Case

We now take S to be a Bernoulli(p) source, and $d(S_j, \hat{S}_j)$ to be the Hamming distortion measure. We define the function $h_b(x) = -x\log x - (1-x)\log(1-x)$ and its inverse restricted to the interval $[0, 1/2]$ as $h_b^{-1}(\cdot)$. For $0 \leq a, b \leq 1$ we also define $a * b = a(1-b) + b(1-a)$. We define $P_{U|S}$ by taking $U = S \oplus N$, where $N \sim \text{Ber}(q)$, independent of S.

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

$$F_P(t) \geq t - h_b(q * p) + h_b\left(q * h^{-1}\left(h_b(p) - t\right)\right), \tag{17}$$

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 [9]. 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 2: For $0 \leq D \leq p$

$$\bar{R}_P(D) = h_b(q * p) - h_b(q * D). \tag{18}$$

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 $V \sim \text{Ber}(D)$. ■

Substituting these results in Theorem 1, we recover the bound of [5] for the binary symmetric MD problem:

$$\begin{aligned} R_1 + R_2 \geq &[h_b(q * D_0) - h_b(D_0)] - [h(q * p) - h(p)] \\ &+ [h(q * p) - h(q * D_1)] + [h(q * p) - h(q * D_2)]. \end{aligned} \tag{19}$$

We now combine the binary source with Hamming distortion, with a degraded broadcast channel. First, consider the case of symmetric erasures, i.e., Y_i is X w.p. $1-\epsilon_i$ and erased otherwise, for $i = 1, 2$, one can verify that:

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

Combining with propositions 1 and 2 and substituting in Theorem 3, 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 [10] (which was also recovered in [7]).

Finally we turn to a binary symmetric channel, i.e., $Y_1 = X \oplus Z_1$, and $Y_2 = Y_1 \oplus Z_2$, where $Z_1 \sim \text{Ber}(\delta_1)$, $Z_2 \sim \text{Ber}(\delta_2)$, and (X, Z_1, Z_2) are mutually independent.

Proposition 3: For a binary symmetric degraded channel

$$G_Q(t) = \log 2 - h_b\left(\delta_2 * h_b^{-1}\left(h_b(\delta_1) + t\right)\right), \tag{20}$$

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

Proof:

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

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 1 and 2 and substitute in Theorem 3, to obtain the following theorem.

Theorem 4: For the JSCC broadcast problem with a binary symmetric source, Hamming distortion and a binary symmetric channel, suppose that the pair (D_1, D_2) is achievable. Then, for any $0 \leq q \leq 1/2$, it holds that

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

where

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

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

$$A_1 = h_b(\delta_1) + \frac{h_b(q * D_1) - h_b(D_1)}{\rho}. \tag{21}$$

Following the treatment of the Gaussian-quadratic case in [4], we consider the case were the distortion of the "weak" user is optimal. That is, let D_2^* satisfy

$$R(D_2^*) = \rho(\log 2 - h_b(\delta_1 * \delta_2)) \tag{22}$$

For the special case of $D_2 = D_2^*$. We can take $q \to 0$ in Theorem 4, and applying some straightforward algebra, we obtain the following.

Corollary 1: 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^* satisfies (22). Then,

$$g(D_1) \geq g(p) + \frac{g(\delta_1)}{g(\delta_1 * \delta_2)}\left[g(D_2^*) - g(p)\right], \tag{23}$$

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

Similarly, for the special case of $D_1 = D_1^*$, where

$$R(D_1^*) = \rho(\log 2 - h_b(\delta_1), \tag{24}$$

we can take $q \to 1/2$ in Theorem 4, and after applying some straightforward algebra, obtain the following.

Corollary 2: For the JSCC broadcast problem with a binary source and a binary symmetric channel, suppose that the pair (D_1^*, D_2) is achievable, where D_1^* satisfies (24). Then,

$$\begin{aligned} (1 - 2D_2)^2 \leq &(1 - 2 \cdot \delta_2 * D_1^*)^2 \\ &+ (1-2p)^2\left(1 - (1 - 2 \cdot \delta_2)^2\right). \end{aligned} \tag{25}$$

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

$$D_2 \geq \delta_2 * D_1^*. \tag{26}$$

V. Proofs of information inequalities

In this section we prove lemmas 1-3, which serve as the main technical ingredient of our results.

Proof of Lemma 1: Monotonicity and of the function $F_P(t)$ follows by definition. Establishing convexity is straightforward, and the proof is omitted. We prove tensorization by induction. For any V that satisfies the Markov chain

$U^m - S^m - V$, we have

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

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

$$F_P\left(\frac{I(S^m;V)}{m}\right)$$
$$\leq \frac{1}{m}F_{P^{m-1}}\left(I(S^{m-1};V)\right)+\frac{1}{m}F_P\left(I(S_m;V,S^{m-1})\right)$$
$$\leq \frac{1}{m}\left[I(S^{m-1};V|U^{m-1})+I(S_m;V,S^{m-1}|U_m)\right] \quad (27)$$

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

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

and

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

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

$$I(S^{m-1};V|U^{m-1})+I(S_m;V,S^{m-1}|U_m) \leq I(S^m;V|U^m). \quad (28)$$

Substituting (28) into (27), gives

$$I(S^m;V|U^m) \geq mF_P\left(\frac{I(S^m;V)}{m}\right), \quad (29)$$

as desired. ■

Proof of Lemma 3: Monotonicity of $G_Q(t)$ follows by definition. Establishing concavity is straightforward, and the proof is omitted. We prove tensorization by 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,

$$G_Q\left(\frac{I(X^n;Y_1^n|W)}{n}\right)$$
$$= G_Q\left(\frac{I(Y_1^{n-1};X^{n-1}|W)+I(X_n;Y_{1,n}|W,Y_1^{n-1})}{n}\right)$$
$$= G_Q\left(\frac{n-1}{n}\frac{I(Y_1^{n-1};X^{n-1}|W)}{n-1}+\frac{I(X_n;Y_{1,n}|W,Y_1^{n-1})}{n}\right)$$
$$\geq \frac{n-1}{n}G_Q\left(\frac{I(Y_1^{n-1};X^{n-1}|W)}{n-1}\right)$$
$$+\frac{1}{n}G_Q\left(I(X_n;Y_{1,n}|W,Y_1^{n-1})\right) \quad (30)$$

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

$$G_Q\left(\frac{I(X^n;Y_1^n|W)}{n}\right)$$
$$= \frac{G_Q^{n-1}\left(I(Y_1^{n-1};X^{n-1}|W)\right)+G_Q\left(I(X_n;Y_{1,n}|W,Y_1^{n-1})\right)}{n}$$
$$\geq \frac{I(Y_2^{n-1},W)+I(Y_{2,n};W,Y_1^{n-1})}{n}, \quad (31)$$

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

$$I(Y_{2,n};W,Y_1^{n-1}) \geq I(Y_{2,n};W,Y_2^{n-1}) \geq I(Y_{2,n};W|Y_2^{n-1}). \quad (32)$$

Substituting into (31) gives

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

as desired. ■

Proof of Lemma 2: Since U^m is memoryless, we have that $I(U^m;\hat{S}^m) \geq \sum_{i=1}^m I(U_i;\hat{S}_i)$. 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

$$I(U^m;\hat{S}^m) \geq m\frac{1}{m}\sum_{i=1}^m \bar{R}_P(d_i) \geq m\bar{R}_P(D). \quad (33)$$

■

REFERENCES

[1] L. Ozarow, "On a source-coding problem with two channels and three receivers," *The Bell System Technical Journal*, vol. 59, no. 10, pp. 1909–1921, Dec 1980.

[2] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 721–726, Nov 1985.

[3] A. E. Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 851–857, November 1982.

[4] 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.

[5] L. Song, S. Shao, and J. Chen, "A lower bound on the sum rate of multiple description coding with symmetric distortion constraints," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7547–7567, Dec 2014.

[6] 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.

[7] ——, "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.

[8] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[9] 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.

[10] 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.